Advances in Experimental Medicine and Biology 1019

Sebastien Gagneux *Editor*

Strain Variation in the *Mycobacterium tuberculosis* Complex: Its Role in Biology, Epidemiology and Control

Springer

Advances in Experimental Medicine and Biology

Volume 1019

More information about this series at http://www.springer.com/series/5584

Sebastien Gagneux
Editor

Strain Variation in the *Mycobacterium tuberculosis* Complex: Its Role in Biology, Epidemiology and Control

Springer

Editor
Sebastien Gagneux
Swiss Tropical & Public Health Institute
University of Basel
Basel, Switzerland

ISSN 0065-2598 ISSN 2214-8019 (electronic)
Advances in Experimental Medicine and Biology
ISBN 978-3-319-64369-4 ISBN 978-3-319-64371-7 (eBook)
DOI 10.1007/978-3-319-64371-7

Library of Congress Control Number: 2017954359

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

When I started to work on tuberculosis (TB) 15 years ago, a widespread dogma in the field was that *Mycobacterium tuberculosis* was "a clone" with "negligible" genetic diversity across strains and that this limited strain-to-strain variation bore no phenotypic relevance. Since then, advances in genotyping, molecular epidemiology, and experimental investigation of the impact of strain variation on virulence and immune phenotypes have shed a very different light on the subject. In particular, recent findings based on large-scale whole-genome sequencing (WGS) of global collections of *M. tuberculosis* clinical strains and comparisons to other mycobacteria have generated many new insights into the evolution of the *M. tuberculosis* complex (MTBC). As a consequence, the field has experienced a paradigm shift in recent years, and most scientists involved in TB research today agree that bacterial diversity should be considered when studying the biology and epidemiology of TB, as well as for the development of new treatments and vaccines. The purpose of this book is to summarize some of the most important insights gained over the years with respect to the nature, evolution, and consequences of strain variation in the MTBC, including those related to antibiotic resistance.

In Chap. 1, Brites and Gagneux review the evolutionary forces governing strain variation in the MTBC, both across different patient populations and within individuals under treatment. This chapter also serves as an introduction to many of the subsequent chapters of the book. In Chap. 2, Supply and Brosch discuss the biology and epidemiology of *Mycobacterium canettii*, an organism thought to represent the closest living relative of the progenitor of the MTBC. In Chap. 3, Merker and colleagues give a comprehensive overview of all typing methods that have been applied to MTBC clinical strains during the last two decades, culminating in the large-scale application of next-generation sequencing. The latter is further developed in Chap. 4 by Comas, where the application of WGS to molecular epidemiological investigation and its limitations are discussed in detail. In Chap. 5, Coscolla summarizes some of the most important insights into the biological and epidemiological consequences of strain variation in the MTBC. In Chap. 6, Yeboah-Manu and colleagues pursue this subject with a particular focus on *Mycobacterium africanum* (today mainly referred to as MTBC lineages 5 and 6), which is an important cause of human TB in West Africa. In Chap. 7, Malone and Gordon

expand to the so-called animal-adapted members of the MTBC and their biological characteristics. In Chap. 8, Abdullah and Behr switch gears slightly with a discussion of the diversity and evolution of the various Bacillus Calmette-Guerin (BCG) sub-strains that have emerged following the generation of the original vaccine strain in 1921. In Chap. 9, Ernst pursues the related themes of vaccinology and immunology with a particular focus on antigen diversity in the MTBC. In Chap. 10, Delogu and colleagues review the state of the art with respect to the enigmatic strain families of PE and PPE genes. In Chap. 11, Dean and colleagues present the current epidemiological situation of drug-resistant TB in the world, which is followed by Chap. 12, in which Cirillo and colleagues describe the various phenotypic and molecular methods used for drug susceptibility testing. In Chap. 13, Warner and colleagues provide a detailed description of the particularities of the DNA replication machinery in the MTBC, while in Chap. 14, Borrell and Trauner review the latest findings in the evolution of antibiotic resistance in the MTBC. Finally in Chap. 15, Pečerska and colleagues present an overview of the latest developments in mathematical modeling of MTBC epidemiology and evolution.

My hope is that after reading this book, any remaining skeptics will be convinced of the relevance of strain variation in the MTBC and that new researchers will become interested in this fascinating topic. More importantly, I am convinced that by exploiting some of the insights and concepts presented in this book, and integrating them into the ongoing efforts in research and development for new tools and control strategies, we will be in a much better position to tackle TB globally.

Basel, Switzerland Sebastien Gagneux

Contents

Contributors

Abdallah M. Abdallah Bioscience Core Laboratory, King Abdullah University of Science and Technology, Jeddah, Kingdom of Saudi Arabia

Marcel A. Behr Department of Medicine, McGill University Health Centre, Montreal, QC, Canada

Sonia Borrell Swiss Tropical and Public Health Institute, Basel, Switzerland

University of Basel, Basel, Switzerland

Michael J. Brennan Aeras, Rockville, MD, USA

Daniela Brites Swiss Tropical and Public Health Institute, Basel, Switzerland

University of Basel, Basel, Switzerland

Roland Brosch Institut Pasteur, Unit for Integrated Mycobacterial Pathogenomics, Paris Cedex 15, France

Daniela M. Cirillo San Raffaele Scientific Institute, Milan Italy IRCCS, Milano, Italy

Iñaki Comas Institute of Biomedicine of Valencia (IBV-CSIC) and CIBER in Epidemiology and Public Health, Valencia, Spain

Mireia Coscolla Swiss Tropical and Public Health Institute, University of Basel, Basel, Switzerland

Helen Cox Division of Medical Microbiology and Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa

Bouke C. de Jong Institute for Tropical Medicine, Antwerp, Belgium

Anna S. Dean Global TB Programme, World Health Organization, Geneva, Switzerland

Giovanni Delogu Institute of Microbiology, Università Cattolica del Sacro Cuore, Rome, Italy

Joel D. Ernst Division of Infectious Diseases and Immunology, Departments of Medicine, Microbiology, and Pathology, New York University School of Medicine, New York, NY, USA

Sarah M. Fortune Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA, USA

Broad Institute of MIT and Harvard, Cambridge, MA, USA

Ragon Institute of Massachusetts General Hospital, Harvard and Massachusetts Institute of Technology, Cambridge, MA, USA

Sebastien Gagneux Swiss Tropical & Public Health Institute, University of Basel, Basel, Switzerland

Florian Gehre Institute for Tropical Medicine, Antwerp, Belgium

Medical Research Council (MRC) Unit, Serrekunda, Gambia

Stephen V. Gordon UCD School of Veterinary Medicine, University College Dublin, Dublin 4, Ireland

UCD School of Medicine, University College Dublin, Dublin 4, Ireland

UCD School of Biomolecular and Biomedical Science, University College Dublin, Dublin 4, Ireland

UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin 4, Ireland

Thomas A. Kohl Molecular and Experimental Mycobacteriology, Forschungszentrum Borstel, Leibniz-Zentrum für Medizin und Biowissenschaften, Borstel, Germany

German Center for Infection Research (DZIF), partner site Borstel, Borstel, Germany

Kerri M. Malone UCD School of Veterinary Medicine, University College Dublin, Dublin 4, Ireland

Riccardo Manganelli Department of Molecular Medicine, University of Padua, Padua, Italy

Matthias Merker Molecular and Experimental Mycobacteriology, Forschungszentrum Borstel, Leibniz-Zentrum für Medizin und Biowissenschaften, Borstel, Germany

German Center for Infection Research (DZIF), partner site Borstel, Borstel, Germany

Paolo Miotto San Raffaele Scientific Institute, Milan Italy IRCCS, Milano, Italy

Valerie Mizrahi MRC/NHLS/UCT Molecular Mycobacteriology Research Unit and DST/NRF Centre of Excellence for Biomedical TB Research, Department of Pathology and Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Rondebosch, South Africa

Stefan Niemann Molecular and Experimental Mycobacteriology, Forschungszentrum Borstel, Leibniz-Zentrum für Medizin und Biowissenschaften, Borstel, Germany

German Center for Infection Research (DZIF), partner site Borstel, Borstel, Germany

Jūlija Pečerska Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland

Jeremy M. Rock Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA, USA

Philip Supply Université de Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 8204 - CIIL - Centre d'Infection et d'Immunité de Lille, Lille, France

Tanja Stadler Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland

Mark M. Tanaka School of Biotechnology & Biomolecular Sciences, and Evolution & Ecology Research Centre, UNSW Sydney, Sydney, NSW, Australia

Enrico Tortoli San Raffaele Scientific Institute, Milan Italy IRCCS, Milano, Italy

Andrej Trauner Swiss Tropical and Public Health Institute, Basel, Switzerland

University of Basel, Basel, Switzerland

Digby F. Warner MRC/NHLS/UCT Molecular Mycobacteriology Research Unit and DST/NRF Centre of Excellence for Biomedical TB Research, Department of Pathology and Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Rondebosch, South Africa

James Wood School of Public Health and Community Medicine, UNSW Sydney, Sydney, NSW, Australia

Dorothy Yeboah-Manu Noguchi Memorial Institute for Medical Research, University of Ghana, Accra, Ghana

Matteo Zignol Global TB Programme, World Health Organization, Geneva, Switzerland

The Nature and Evolution of Genomic Diversity in the *Mycobacterium tuberculosis* Complex

1

Daniela Brites and Sebastien Gagneux

Abstract

The *Mycobacterium tuberculosis* Complex (MTBC) consists of a clonal group of several mycobacterial lineages pathogenic to a range of different mammalian hosts. In this chapter, we discuss the origins and the evolutionary forces shaping the genomic diversity of the human-adapted MTBC. Advances in whole-genome sequencing have brought invaluable insights into the macro-evolution of the MTBC, and the biogeographical distribution of the different MTBC lineages, the phylogenetic relationships between these lineages. Moreover, micro-evolutionary processes start to be better understood, including those influencing bacterial mutation rates and those governing the fate of new mutations emerging within patients during treatment. Current genomic and epidemiological evidence reflect the fact that, through ecological specialization, the MTBC affecting humans became an obligate and extremely well-adapted human pathogen. Identifying the adaptive traits of human-adapted MTBC and unraveling the bacterial loci that interact with human genomic variation might help identify new targets for developing better vaccines and designing more effective treatments.

Keywords

co-evolution • virulence • adaptation • mutation

D. Brites (✉)
Swiss Tropical and Public Health Institute, Basel, Switzerland

University of Basel, Basel, Switzerland
e-mail: d.brites@swisstph.ch;

S. Gagneux
Swiss Tropical & Public Health Institute, University of Basel, Basel, Switzerland
e-mail: sebastien.gagneux@swisstph.ch

1.1 Introduction

Tuberculosis (TB) has affected human populations since ancient times. In the nineteenth century, TB was recognized as a contagious illness and its etiologic agent, the bacillus *Mycobac-*

S. Gagneux (ed.), *Strain Variation in the* Mycobacterium tuberculosis *Complex: Its Role in Biology, Epidemiology and Control*, Advances in Experimental Medicine and Biology 1019,
DOI 10.1007/978-3-319-64371-7_1

terium tuberculosis, was identified by Robert Koch in 1882. At that time, TB was a common cause of mortality in Europe and North America, causing 20–25% of all adult deaths (Stead 2001). Since the mid-1920s, TB rates have decreased dramatically in the developed world. However, TB is still among the ten top causes of death in many parts of the developing world (WHO 2016). In 2015, the World Health Organization (WHO) estimated 10.4 million new TB cases globally and 1.8 million deaths, making TB the number one cause of human death due to a single infectious agent (WHO 2016). The number of people carrying latent infections is estimated at 1.7 billion worldwide, urging for new diagnostic tools, better vaccines and new treatment regimens to control TB (Houben and Dodd 2016). Much of the TB epidemic, especially in sub-Saharan Africa, is driven by co-infection with the human immunodeficiency virus (HIV), which accounted for 1.2 million of all new TB cases in 2015 (WHO 2016). Moreover, 480,000 cases were estimated to be caused by multidrug-resistant TB (MDR-TB), posing additional challenges for diagnosis and treatment (see Chaps. 12 and 14 of this book).

M. tuberculosis is one of the most successful human pathogens, having killed an estimated one billion people over the last 200 years (Paulson 2013). It is well recognized that the social characteristics of human populations (Lonnroth et al. 2009), host genetics (Abel et al. 2014), co-morbidities and environmental conditions such as the quality of TB control programs are crucial determinants of TB (Dye et al. 2009). Yet, despite years of research, the mechanisms underlying protective immune responses remain poorly understood, and an effective vaccine against pulmonary TB is still lacking (Ernst 2012). During recent years, it has become increasingly evident that in addition to host and environmental determinants, the genetic diversity of the pathogen may also play a role in the outcome of TB infection and disease (Comas and Gagneux 2009) (see Chap. 5). In this chapter, we review the nature of the genetic diversity in the tubercle bacilli as well as the various evolutionary forces that have been driving this diversity. This chapter also serves as an introduction to some of the other chapters of this book.

1.2 The Phylogeography of the *Mycobacterium tuberculosis* Complex

TB in humans and animals is caused by a group of acid-fast bacteria collectively known as the *Mycobacterium tuberculosis* Complex (MTBC). The MTBC consists of a group of highly related mycobacterial lineages (99.9% nucleotide identity) that infect a wide range of mammalian host species. The MTBC comprises the obligate human pathogens *Mycobacterium tuberculosis sensu stricto* and *Mycobacterium africanum* (see Chap. 6), several lineages mainly affecting wild and domestic mammalian hosts (Chap. 7), and the so-called "smooth tubercle bacilli" that include *Mycobacterium canettii* (Chap. 2) (Coscolla and Gagneux 2014). We will use the term MTBC throughout this chapter to refer to all MTBC lineages except *M. canettii*, which differs markedly from the remaining members of the Complex as discussed in detail in Chap. 2. The present chapter focuses on the latest insights gained mainly from whole-genome sequencing (WGS) data. For a detailed overview of the different genotyping methods used for the MTBC over the last two decades, please refer to Chap. 3.

The MTBC comprises seven phylogenetic lineages, for which humans are the only known host (Fig. 1.1a). This is to say, humans are the only hosts, in which these seven lineages are able to maintain complete cycles of infection, disease and airborne transmission (Ernst 2012); we thus refer to these lineages as "human-adapted" MTBC. This is in contrast to the "animal-adapted" MTBC lineages, which are discussed in more detail in Chap. 7. The seven human-adapted lineages comprise *M. tuberculosis sensu strictu* Lineages 1–4, Lineage 7, as well as *M. africanum* Lineages 5 and 6 (Fig. 1.1a). The geographical spread of these lineages differs markedly, with some lineages exhibiting a global distribution and others a strong geographical restriction

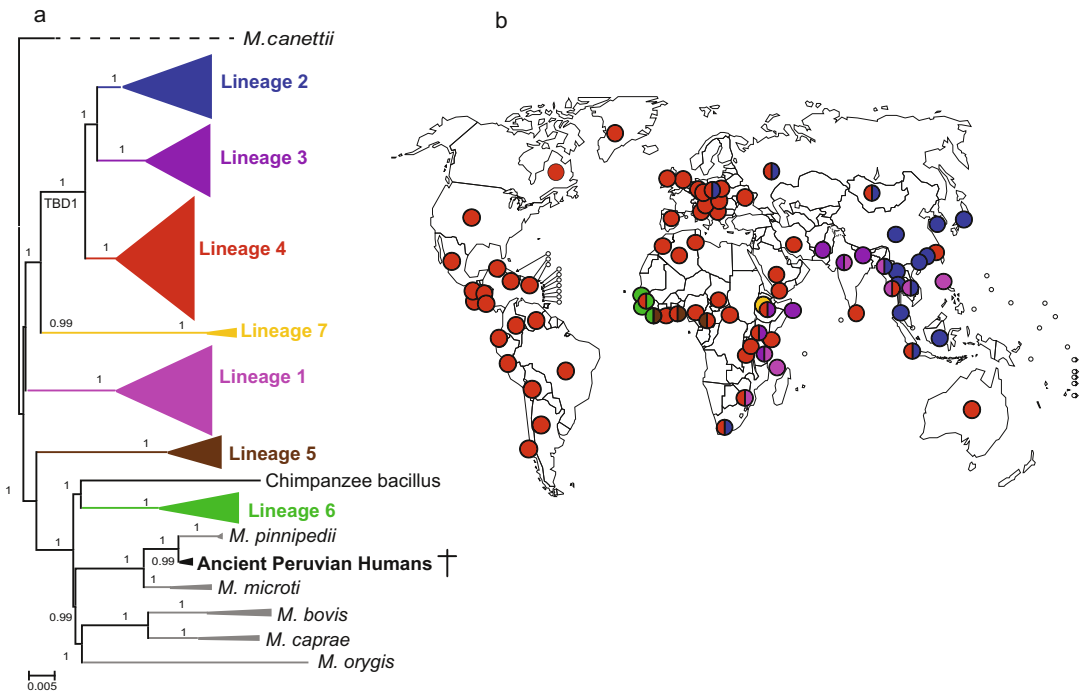

Fig. 1.1 Global phylogeography of the human-adapted MTBC. (**a**) Phylogeny of 220 MTBC strains obtained by Comas et al. (2013) based on a maximum-likelihood method. Support values were obtained by bootstrap. Branch lengths are proportional to nucleotide substitutions and the tips of the tree were collapsed for simplicity. The tree is rooted using *M. canettii*. (**b**) Global distribution of the seven main human-adapted MTBC lineages. Lineages were coloured as in **a**, and coloured dots represent the dominant MTBC lineage(s) in each country (Adapted from Gagneux and Small 2007)

(Fig. 1.1b) (Gagneux et al. 2006b; Gagneux and Small 2007). Lineage 2 (also known as East-Asian lineage; includes the Beijing family of strains) and Lineage 4 (Euro-American lineage) occur worldwide. Lineage 5 (*M. africanum* West Africa 1) and Lineage 6 (*M. africanum* West Africa 2) are highly restricted to West-Africa (de Jong et al. 2010b), while Lineage 7 occurs almost exclusively in Ethiopia (Firdessa et al. 2013). The remaining two lineages show an intermediate distribution, with Lineage 1 (also known as Indo-Oceanic or EAI) occurring all around the Indian Ocean and Lineage 3 (Central Asian Strain, CAS) dominating in parts of East Africa, Central- and South-Asia (Hershberg et al. 2008).

Lineages 2, 3 and 4 are sometimes also referred to as "modern", whereas the remaining lineages are called "ancient". This nomenclature has its roots in work by Brosch and colleagues which confirmed, based on phylogenetic analyses of genomic deletions, that human-adapted strains did not derive from a bovine form of TB as postulated previously (Brosch et al. 2002; Gordon et al. 1999). More specifically, the authors showed that MTBC strains responsible for major outbreaks in humans such as Beijing strains (Lineage 2) and Haarlem strains (Lineage 4) shared a deletion of the genomic region known as TbD1, which is intact in *M. africanum*, *M. bovis* and all other animal-adapted lineages. The terms "modern" and "ancient", despite their temporal connotation, refer thus to the absence or presence of the TbD1 region, respectively, and should not be interpreted necessarily as reflecting an evolutionary time dimension (Smith et al. 2009). By definition, all extant MTBC lineages have been evolving for the same amount of time since their most recent

common ancestor (Smith et al. 2009). Some lineages are closer to the root of the MTBC phylogeny indicating an earlier divergence from the MTBC common ancestor like Lineage 1 (Fig. 1.1). By contrast the divergence of *Mycobacterium bovis* (an "ancient" lineage according to the presence of the TbD1 region) for instance, has occurred comparatively later in the evolutionary history of the MTBC.

In the following sections, we will discuss the evolutionary forces driving the phylogeographical structure of the MTBC, as well as the potential implications for the biology and epidemiology of human TB. Additional aspects of the consequences of the MTBC diversity are discussed in Chaps. 3, 5, 6, and 14 of this book.

1.3 The Origin of the MTBC as a Professional Pathogen

1.3.1 The Transition from the Environment to an Obligate Pathogen

Given the characteristics of *M. canettii* and its high nucleotide similarity (97.3%) with the other members of the MTBC (Supply et al. 2013), it seems plausible that the ancestor of the extant MTBC lineages was an organism similar to *M. canettii*. *M. canettii* is highly restricted to the Horn of Africa (mainly Djibouti), and current evidence suggests it is a facultative pathogen with a likely reservoir in the environment (Koeck et al. 2011; Blouin et al. 2014) (this is further discussed in Chap. 2). The transition of this *M. canettii*-like MTBC ancestor to a professional pathogen ultimately enabled the MTBC to spread globally. In the case of the human-adapted lineages of the MTBC and those adapted to domesticated animals, this transition was the key to a great evolutionary success. Hence, comparative analyses of *M. canettii* with the other members of the MTBC and other closely related mycobacteria can shed light on some of the key traits and genes underlying the MTBC's evolutionary transition from an environmental organism to an obligate pathogen. Such studies have highlighted several genomic regions in the MTBC that are important for the pathogenesis of TB (Jang et al. 2008; Veyrier et al. 2011; Behr 2015). For example, some Type VII Secretion Systems like ESX-1 contribute to the survival of the MTBC inside macrophages (Groschel et al. 2016). However, many of those regions are also conserved in other mycobacteria (Groschel et al. 2016; Behr 2015). Comparative studies have revealed around 50 genes that are present in the MTBC but absent from *M. canettii* strains (Supply et al. 2013). Those were probably acquired via horizontal gene transfer preceding the clonal expansion of the MTBC (Supply et al. 2013). Yet, the relevance of these genes for the process of adaptation of the MTBC to the human host remains unclear. The results from these comparative studies suggest that the evolution towards an obligate pathogenic lifestyle in humans is not just the result of the presence or absence of specific "virulence factors" (Behr 2015; Wang et al. 2015; Jang et al. 2008; Supply et al. 2013). Instead, many individual genomic features, including particular point mutations in key genes, as well as epistatic interactions between these features might have contributed to this adaptation (Gonzalo-Asensio et al. 2014). The human-adapted MTBC lineages are considered obligate human pathogens because (i) humans are the only hosts, in which these pathogens successfully maintain complete life cycles (Ernst 2012), and (ii) because these microbes have to cause pulmonary disease to successfully spread from human-to-human (Brites and Gagneux 2012). In other words, in human TB, and contrary to many other pathogens (Woolhouse and Gaunt 2007), transmission is tightly linked to virulence, which has important evolutionary and epidemiological consequences.

1.3.2 Immune Subversion Instead of Immune Escape

After inhalation, MTBC is initially recognized and taken up by alveolar macrophages (O'Garra et al. 2013). The bacterium has evolved multiple ways of escaping macrophage killing, which allows the bacterium to replicate inside this intracellular niche (O'Garra et al. 2013). Increased bacterial loads and antigen

presentation by antigen-presenting cells trigger T cell responses, which lead to the recruitment of more macrophages and to the formation of lung granulomas that contain infected macrophages (O'Garra et al. 2013). Adequate T cell responses are essential for containing the infection inside granulomas as revealed by the systemic infections that the MTBC causes in patients co-infected with HIV (for review see Kwan and Ernst 2011). Most immunocompetent individuals are able to keep the replication of the bacterium at check inside the lung granulomas. In 5–10% of cases, uncontrolled growth of the pathogen causes necrosis, promoting lung damage and enhancing aerosol production (O'Garra et al. 2013). Between humans, the MTBC is transmitted exclusively through aerosols, and therefore virulence (here meant as the damage that the infection causes to the host) is a pre-requisite for efficient pathogen transmission. Host T cell responses are thus essential for a protective immune response against TB, but at the same time contribute to efficient transmission of the pathogen (Brites and Gagneux 2012). What triggers the uncontrolled bacterial growth inside the granuloma? Comparing the molecular evolution of genomic regions encoding proteins that are recognized by human T cells (hereafter called T cell epitopes) of representatives of the human-adapted MTBC has revealed that these T cell epitopes are among the most conserved regions of the MTBC genomes, exhibiting lower frequencies of amino acid changes than essential genes or non-epitope regions of antigens (Comas et al. 2010; Coscolla et al. 2015; Pepperell et al. 2013; Yruela et al. 2016). These results suggest that most amino acid mutations in T cell epitopes have a deleterious effect for the bacteria and are therefore removed from the population by natural selection. One possible underlying selective pressure is the need of recognition by T cells for promoting inflammatory immune responses, leading to enhanced lung damage and increased transmission. Hence following such a scenario, the strategy of MTBC might be to subvert the host immune system instead of evading immune responses through antigenic variation as observed in many other pathogens (see also Chap. 9). However, in *M. canettii,* for which human-to-human transmission via aerosols has not been reported (Koeck et al. 2011; Supply et al. 2013) (Chap. 2), in the animal-adapted lineages of MTBC (Yruela et al. 2016) (Chap. 7) and in BCG-strains (Copin et al. 2014) (Chap. 8), T cell epitope sequences are also more conserved than other regions of the genome. Hence, one alternative explanation might be that T cell epitopes in the MTBC are hyperconserved because they are part of enzymes or proteins with essential roles in vivo, which are independent of the interaction with T cells. As more genome sequences of MTBC strains become available, it is becoming evident that although the majority of T cell epitopes in the MTBC are hyperconserved, some epitopes do exhibit amino acid diversity (Coscolla et al. 2015; Stucki et al. 2016). This suggests that some T cell epitopes might be involved in antigenic variation and/or specific adaption to different human populations (further discussed below) (Coscolla et al. 2015). A study comparing individuals latently infected with MTBC to healthy uninfected controls found that the latently infected individuals primarily recognized T cell epitopes that were hyperconserved, whereas the un-infected individuals mainly recognized T cell epitopes that were more diverse (Lindestam Arlehamn et al. 2015). Moreover, this study also found that the epitopes that were hyperconserved within the MTBC and more strongly recognized in latently infected individuals tended to be less conserved in non-tuberculous mycobacteria (Lindestam Arlehamn et al. 2015). Both of these observations support the link between sequence hyperconservation of T cell epitopes and transmission in the context of human-adapted MTBC.

1.4 The Evolutionary Forces Driving MTBC Diversity

1.4.1 Impact of Clonality and Genetic Drift

The transition to an obligate pathogenic lifestyle was also associated with profound changes in the way genomic variation has been shaped through-

out the evolution of the MTBC. Unlike *M. canettii* and many other bacteria, the human- and animal-adapted members of the MTBC do not undergo horizontal gene transfer (Boritsch et al. 2016). An important consequence of this is that genes lost through deletion cannot be reacquired. The fact that genomic deletions can be used as robust phylogenetic markers to classify MTBC strains into discrete lineages is a consequence of this (Gagneux et al. 2006b) (Brosch et al. 2002; Hirsh et al. 2004), and further illustrates the strong clonality of the MTBC (Gagneux and Small 2007) (see also Chap. 3). Comparative analyses of clonal and non-clonal bacteria including the MTBC suggest that variation in gene content through gene loss is a major source of genetic variation in clonal species (Bolotin and Hershberg 2015). A lack of horizontal gene exchange also means that the negative fitness effects of mutations can only be alleviated by reversion or compensatory evolution as deleteriously mutated loci cannot be replaced through transfer of intact loci from other strains (Rice 2002). In recent years, fascinating mechanisms of compensation for deleterious mutations affecting phenotypes involved in virulence and resistance to antibiotics have been elucidated.

One of such examples is the PhoPR virulence regulation system that controls important MTBC virulence determinants such as ESAT-6 and others (Gonzalo-Asensio et al. 2014). The common ancestor of *M. africanum* Lineage 6 and the animal-adapted lineages (Fig. 1.1) has three mutations affecting the PhoPR genes. By transferring these mutations into human-adapted *M. tuberculosis* reference strains CG1237 (a Lineage 2 strain) and H37Rv (a Lineage 4 strain), the secretion of ESAT-6 antigen was decreased. These strains also became less virulent in macrophages and mice (Gonzalo-Asensio et al. 2014). Interestingly, *M. africanum* Lineage 6 and the animal-adapted lineages compensate the deleterious effects of the PhoPR mutations by secreting ESAT-6 in a PhoPR-independent way mediated by the deletion of the Region of Difference (RD) 8 (Gonzalo-Asensio et al. 2014) (Chap. 6).

Compensatory evolution for the deleterious fitness effects of mutations that confer drug resistance have also been described (Trauner et al. 2014). The best known examples are the emergence of secondary mutations in the RNA polymerase of rifampicin-resistant strains, which have been shown to compensate for the deleterious effects of RpoB mutations in the absence of rifampicin (Comas et al. 2012; de Vos et al. 2013; Gygli et al. 2017). Similarly in isoniazid resistance, promotor mutations increasing the expression of AhpC have been proposed to compensate for high-cost mutations in KatG (Sherman et al. 1996; Gagneux et al. 2006a) (see Chap. 14).

Strong clonality and the lack of horizontal gene transfer also cause strong linkage between sites in the MTBC genome. As a consequence, the fate of new mutations, being structural variants or point mutations, are not independent of each other (Kaiser and Charlesworth 2009). For instance, mutations with a strong selective advantage, such as those conferring antibiotic resistance in TB patients undergoing drug treatment, will be selected for and will drive all other mutations present in the strain genetic background, in which the new mutation arose, a process known as genetic hitchhiking (Smith and Haigh 1974; Eldholm et al. 2015). Similarly, if a mutation has a strong deleterious effect, all other genetic variation (be it deleterious, neutral or beneficial) linked to such a mutation will also be removed from the population, a process known as background selection (Kaiser and Charlesworth 2009; Pepperell et al. 2013). These evolutionary processes, together with genetic drift caused by the tight transmission bottlenecks characteristic of TB (Lin et al. 2014), likely contribute to the low genomic diversity in the MTBC compared to other bacteria (Achtman 2008). In average, MTBC clinical isolates harbour about one single nucleotide polymorphism (SNP) every 3000 base pairs, and the maximum distance between two most distant MTBC strains is about 2000 SNPs (Coscolla and Gagneux 2014). This level of genomic diversity is orders of magnitude lower than e.g. in *M. canettii* strains, which have genomes

10–114 kb larger than the MTBC, and differ on average by about 42,000 SNPs (Supply et al. 2013). Furthermore, a high proportion of the SNPs in MTBC populations is present in very low frequencies, a phenomena usually thought to be the result of purifying selection removing from the population mutations that are deleterious (Hershberg et al. 2008; Pepperell et al. 2013). An excess of mutations in very low frequencies can also be explained by population expansions and in the case of the human adapted MTBC such expansions have probably occurred together with those of modern human populations (Comas et al. 2013; Merker et al. 2015; Luo et al. 2015). In addition, MTBC populations exhibit a high ratio of nonsynonymous to synonymous changes compared to other bacteria (Rose et al. 2013), which has been proposed to reflect the effect of strong genetic drift linked to transmission bottlenecks (Hershberg et al. 2008). Studies in macaques have shown that new infections can be initiated by only a single or a few bacterial cells (Lin et al. 2014). However, as in other bacteria (Worby et al. 2014), not much is known about transmission bottlenecks in the MTBC in clinical settings. Importantly, the high ratio of nonsynonymous mutations might incur a substantial genetic mutational load as a high proportion of nonsynonymous mutations in the MTBC (around 44%) have been predicted to have deleterious functional consequences (Rose et al. 2013). Interestingly, recent studies have revealed that the patterns of genetic diversity found within hosts are similar to those found between hosts (O'Neill et al. 2015; Trauner et al. 2017; Lieberman et al. 2016), suggesting that both intra and inter-host processes account for the low sequence diversity in MTBC strains.

1.4.2 Mutation Rates and Within-Host Evolution

In vitro mutation rates (here meant as the number of mutations per cell division per generation) in the MTBC, like for instance those determined under rifampicin selection, are comparable to those in *Escherichia coli* (see Chap. 13) (Ford et al. 2011, 2013). The rapid emergence of drug resistance in clinical settings also suggests that mutation supply is not limiting in the MTBC. Studies that have followed the evolution of MTBC strains within individual patients during treatment using serial sampling suggest that significant genetic diversity is generated as the bacteria replicate inside the host, and that the presence of mutations conferring resistance to one or more drugs can be already detected in bacterial sub-populations within the first weeks of treatment (Eldholm et al. 2014; Trauner et al. 2017). In addition, multiple independent resistance mutations can emerge in parallel, suggesting again that mutation rates in the MTBC do not limit adaptation, at least not in the context of drug resistance acquisition (Eldholm et al. 2014; Bloemberg et al. 2015). Eldholm and colleagues demonstrated that the expansion of certain drug-resistant clones within patients was clearly due to an increased fitness of those clones compared to other less frequent clones (Eldholm et al. 2014). In a study by Trauner et al. by using very deep sequencing of sequential sputum samples, the authors could show that there is a continuous turn-over of minor alleles, which are constantly purged by purifying selection over time as patients undergo effective treatment (Trauner et al. 2017). A snapshot of how genetic diversity of MTBC strains within the host may look like in patients who have undergone minimal antibiotic treatment and who died of TB was recently provided by Lieberman et al. (2016). These authors analysed post-mortem MTBC samples isolated from the lungs and extra-pulmonary sites of 44 individuals, most of whom were HIV-coinfected. These results also suggested that purifying selection is the major process shaping genetic diversity within the host, even in the absence of antibiotic treatment. Despite that, MTBC diversified within these patients, producing different genotypes that co-existed for years (Lieberman et al. 2016). The latter confirmed what most other studies of intra-host diversity had already suggested; the genetic diversity of MTBC inside patients can be as high as the genetic diversity sampled across pa-

tients during outbreaks (Perez-Lago et al. 2014; O'Neill et al. 2015). As a consequence, using a single MTBC genome per host to infer transmission events in the MTBC can lead to spurious inferences (see Chap. 4). Moreover, these within-host studies illustrate the fact that lungs from TB patients harbour genetically heterogeneous MTBC populations. Hence, single sputum samples do not necessarily represent the full spectrum of genetic diversity present within a host, which is an important aspect to consider in the context of diagnosis of drug resistance conferring mutations (see Chap. 12).

To date, few studies have reported genetic signatures of adaptive evolution in the MTBC (Osorio et al. 2013; Gonzalo-Asensio et al. 2014; Koch et al. 2017), except in the context of drug resistance or compensatory evolution. This is in contrast to other pathogens such as *Burkholderia dolosa* and *Pseudomonas aeruginosa* that exhibit strong de novo adaptation to the human host (Didelot et al. 2016; Lieberman et al. 2014; Marvig et al. 2015). Perhaps the lack of horizontal gene transfer and having to rely merely on point mutations and deletions limits the capacity of MTBC strains to adapt to new host environments. A short-term experimental evolution experiment of the laboratory strain H37Rv infecting mice reported purifying selection as the main process purging genetic diversity inside these hosts (Copin et al. 2016). The same was reported for disseminated infections with *M. bovis* BCG in humans (Copin et al. 2016). These results could also be interpreted as reflecting a limited capacity of the MTBC to adapt to new environments. However, a low adaptability of the human-adapted MTBC is difficult to reconcile with its long-term persistence and global success in human populations (Paulson 2013). Hence alternatively, it could be argued that MTBC is already extremely well adapted to its human host, and therefore most new mutations emerging within patients are likely to be deleterious and therefore selected against. Under this scenario, the major adaptive steps of the MTBC to its human host would have occurred in the remote past.

1.4.3 Mutation Rates During Latency

As referred above, around 90% of the individuals who become infected with the human-adapted MTBC develop latent TB and never progress to active disease. Latent TB can be defined as a persistent infection by MTBC without clinical symptoms. However, latent TB infection represents a broad spectrum, in which the bacteria can exhibit a range of physiological states (Barry et al. 2009). Little is known about the replication ability and consequently about mutation rates of the MTBC during latency (Gill et al. 2009). It is a notoriously difficult subject of study because, except for nonhuman primates, animal models do not usually develop latent TB infection (Flynn et al. 2015). However, for improving drug regimens to treat latent TB, a better understanding of latency is of paramount importance (Gill et al. 2009; Esmail et al. 2014). Furthermore, the impact of latency on MTBC mutation rates has implications for our understanding of drug resistance evolution (see Chaps. 13 and 14) and on the so-called "molecular clock", which is at the basis of the models used for understanding the evolution and the epidemiology of the MTBC (see section below and Chaps. 4 and 15). A study performed in cynomolgus macaques infected with the laboratory strain Erdman derived from human MTBC concluded that the bacterial mutation rates during the latent stages of the infection were similar to those during active disease (Ford et al. 2011). The polymorphic changes observed suggested that oxidative DNA damage, putatively induced by the host immune responses during latency, was a main mechanism driving the emergence of mutations during latency (Ford et al. 2011). By contrast, another study sequenced four MTBC clinical strains isolated two and 20 year after exposure to the same index case and found that mutation rates were lower in the strain that had undergone a longer period of latency when compared to the strain that reactivated more quickly (Colangeli et al. 2014). The authors also did not find any evidence of mutations induced by oxidative DNA

damage (Colangeli et al. 2014). Possibly, latent TB in macaques and humans differs. Moreover as described above, genetic diversity of the MTBC inside the host can be influenced by selective processes, which we are only starting to understand.

1.5 The Origin and Evolutionary History of the Human-Adapted MTBC

1.5.1 From Promiscuity to Host Specialisation

The seven human-adapted members of the MTBC can be classified into independent lineages. However, these lineages do not form a monophyletic unit (i.e. they do not all share the same most recent common ancestor (MRCA)). In particular, the *M. africanum* Lineages 5 and 6 share a common ancestor with the animal-adapted lineages that is not shared by the other human-adapted lineages (Fig. 1.1a) (Brites and Gagneux 2015). The latest whole-genome based phylogenetic reconstructions of the MTBC also show that human TB did not evolve from a bovine form of the disease as previously believed (Coscolla and Gagneux 2014). From all the extant MTBC lineages, the first ones to have diverged from a common ancestor were the human-adapted Lineage 1 and the ancestor, from which Lineages 5 and 6, as well as the animal-adapted lineages evolved (Brites and Gagneux 2015). Based on these phylogenetic relationships, a most parsimonious scenario would postulate that this ancestor was a promiscuous pathogen with a broad host-spectrum that subsequently diverged into separate lineages that became more host-specialized over time (Smith et al. 2009; Brites and Gagneux 2015; Coscolla et al. 2013). The same scenario might apply for the MRCA of the whole MTBC, after the initial transition of an *M. canettii*-like organism into an obligate pathogen.

1.5.2 The MTBC Originated in Africa

Just as the African origin of humans is strongly supported (Mallick et al. 2016), most evidence to date also point to Africa, and possibly the Horn of Africa, as the most likely geographical origin for the MRCA of the MTBC. Specifically, as outlined above, (i) *M. canettii* is the most closely related organism to the human and animal MTBC members and is strongly associated with the Horn of Africa (Koeck et al. 2011; Supply et al. 2013; Gutierrez et al. 2005), (ii) Africa is the only continent where all extant MTBC lineages are present (Fig. 1.1) (Hershberg et al. 2008; Comas et al. 2013, 2015; Gagneux et al. 2006b; Wirth et al. 2008), (iii) phylogeographical analyses based on 220 global MTBC strains revealed Africa as the most likely origin of the MTBC (Comas et al. 2013), (iv) the genetic diversity of the MTBC is highest in Africa and decreases with increasing distance from Africa (Comas et al. 2015), (v) the most basal lineages of the MTBC are present in Africa, and in the case of Lineage 5 and 6, almost exclusively on that continent (Comas et al. 2013; de Jong et al. 2010b). Finally, the distribution of the animal-adapted MTBC lineages also support an African origin for the whole MTBC, as most known lineages affecting wild animals are restricted to African mammal host species (Brites and Gagneux 2015) (see Chap. 7). These include *M. suricattae* (infects mercats), *M. mungi* (mongooses), the dassy bacillus (hyraxes) and the chimpanzee bacillus (chimpanzees).

Recently, it has been hypothesized that the controlled use of fire by human populations in the Pleistocene may have created the ideal conditions for the transition of the MTBC ancestor from an environmental organism to a pathogen, by potentiating social gatherings and by increasing smoke-induced susceptibility to lung infection (Chisholm et al. 2016). The authors suggest that "humans were the only species able to support opportunistic infection at a sufficient rate for the MTBC precursor to evolve transmissibility and

strategies to evade the host immune response". According to this scenario, humans are considered as the breeding ground species, in which the MTBC evolved its ability to cause persistent infections, enabling it to subsequently also infect wild- and domesticated animals. It has also been suggested that infection with the MTBC might have been beneficial in early human populations by providing nutrients important for the development of the nervous system during times of shortage (Williams and Dunbar 2014).

The fact that the MTBC most likely originated in Africa appears at odds with early colonial reports, according to which TB did not exist on the continent before European colonization (Stead 1998). Indeed at the time, Africa was considered a "virgin soil" for the disease because colonial medical reports from before the twentieth century did not recognize TB as a health problem in Africa; a time when TB was a major cause of mortality and morbidity in Europe (Stead 1998, 2001). A detailed phylogeographic analysis of the MTBC diversity in Ethiopia by Comas et al. contradicts this "virgin soil" hypothesis by demonstrating that several genotypes of MTBC were present on the African continent before the first European contact (Comas et al. 2015). Notwithstanding, several other genotypes have probably also been introduced by Europeans (Comas et al. 2015) (see below). Currently, Ethiopia is a high-burden TB country, and TB is caused by a range of different genotypes from both pre- and post-European contact (Comas et al. 2015). As we shall discuss later, certain MTBC genotypes circulating in the crowded cities of industrializing Europe at the time may have evolved increased virulence that upon introduction by Europeans, allowed them to expand in Africa and outcompete some of the native strains. However, irrespective of this more recent European influence, how and when did the main human-adapted MTBC lineages initially expand out of Africa to the rest of the world? Several hypotheses have been put forward to explain the extant distribution of the main human-adapted MTBC lineages. Likewise, different approaches have been taken to estimate the age of the main lineage splits in the MTBC.

1.5.3 Human and MTBC Co-phylogenies

Because the human-adapted MTBC lineages are obligate human pathogens, it is reasonable to assume that human migrations influence the population structure of the MTBC. This notion has been tested by comparing the phylogenetic structure of globally representative collections of MTBC clinical strains with that of global human populations based on human mitochondrial haplogroups in one study (Comas et al. 2013), and in another study based on the human Y chromosome (Pepperell et al. 2013). The study by Comas et al. (2013) found significant qualitative and quantitative links between the MTBC- and human mitochondrial phylogenies, suggesting co-divergence between the MTBC main lineages and human continental populations (Fig. 1.2a, b). By contrast, the study by Pepperell et al. (2013) observed no equivalent association between the MTBC and human Y chromosome data (Fig. 1.2c–e). Nevertheless, interesting commonalities between both studies are also apparent (reviewed in Brites and Gagneux 2015). In particular, (i) a strong topological concordance exists between the split of basal branches from which *M. africanum* Lineage 5 and Lineage 6 evolved and the splits of both the Y chromosome haplotype A and the mitochondrial macrohaplogroup L, which are exclusively present in African human populations (Fig. 1.2), (ii) there is concordance between the split that originated MTBC Lineage 1 and the Y chromosome haplogroup C and the mitochondrial macrohaplogroup M that are found in East Asia, South Asia and Oceania, and (iii) the split of both the Y chromosome haplogroup F and the mitochondrial macrohaplogroup N that dispersed outside Africa is in concordance with the split of the "modern" Eurasian MTBC Lineages 2, 3 and 4.

1.5.4 Molecular Dating

In recent years, some of the hypotheses about the origin and evolutionary history of the MTBC

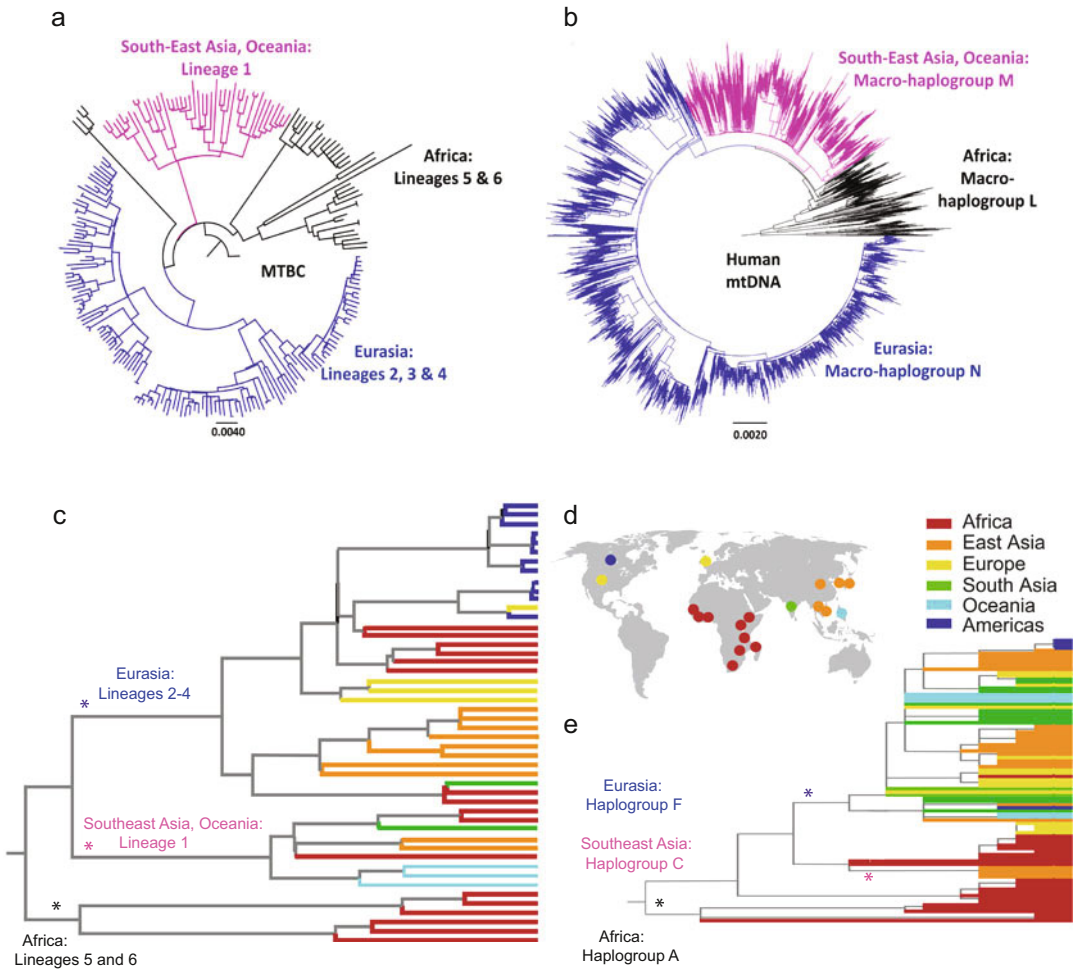

Fig. 1.2 Comparison of the global MTBC and global human populations phylogenies. (**a**) MTBC Maximum-likelihood phylogeny derived from WGS of 220 strains with a global distribution (From Comas et al. 2013). (**b**) Phylogeny derived from 4955 human mitochondrial haplotypes (From Comas et al. 2013). The *colour codes* of **a** and **b** highlight the similarities in tree topology and geographic distribution between MTBC strains and the main human mitochondrial macrohaplogroups (*black*, African clades: MTBC Lineages 5 and 6, human mitochondrial macrohaplogroups L0–L3; *pink*, Southeast Asian and Oceania clades: MTBC Lineage 1, human mitochondrial macrohaplogroup M; *blue*, Eurasian clades: MTBC Lineages 2, 3 and 4, human mitochondrial macrohaplogroup N). MTBC Lineage 7 has only been found in Ethiopia, and its correlation with any of the three main human haplogroups remains unclear. (**c**) MTBC Maximum clade credibility phylogeny inferred from WGS of 48 MTBC globally distributed strains. Tips are coloured by the geographic origin of the MTBC isolate (see key) (Adapted from Pepperell et al. 2013). Modifications to the original figures: the lineage nomenclature (indicated with an *asterisk*) and colour codes used by Comas et al. (2013) were added to the original tree. (**d**) Countries of origin for MTBC isolates used in this study are shown as *coloured dots* on the global map. One *dot* is shown per country but some countries were represented by >1 MTBC isolate. *Colours* correspond to global regions (see key) (From Pepperell et al. 2013). (**e**) Phylogeny of global human populations based on Y chromosome data. Tips are coloured according to the same scheme as the MTBC phylogeny in **c** (From Pepperell et al. 2013). Modifications: the names of the relevant human Y haplogroups (Underhill and Kivisild 2007) were added (indicated with an *asterisk*) and colour coded as in Comas et al. (2013). Figure from Brites and Gagneux 2015

have been tested using time-calibrated phylogenies (Comas et al. 2013; Pepperell et al. 2013; Bos et al. 2014; Wirth et al. 2008; Kay et al. 2015). The results of these studies differed substantially. Here we will focus on the most contrasting time estimates and discuss the factors

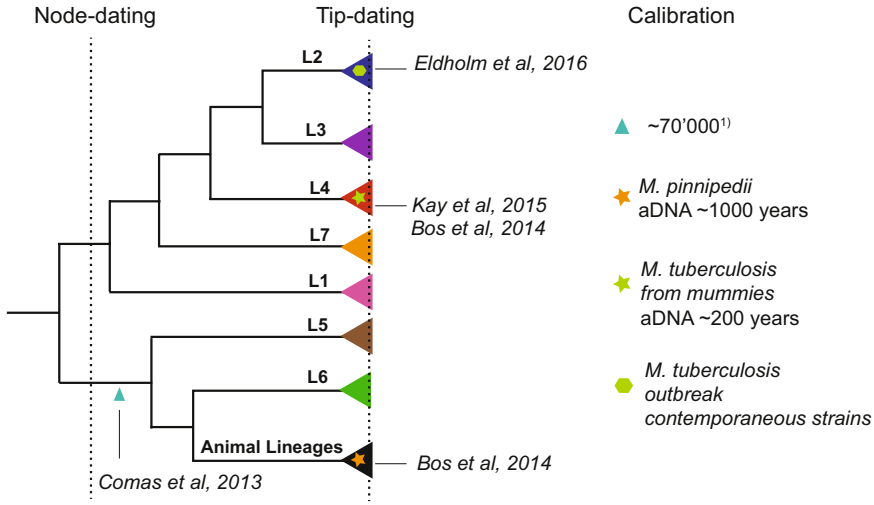

Fig. 1.3 Schematic depiction of some of the approaches used to date the MTBC phylogeny. The phylogenetic relationships depicted are based on the phylogenetic trees obtained by (Comas et al. 2013; Bos et al. 2014) and the branch length is not informative. The node-dating performed by Comas et al. (2013) used four different calibration points for the coalescence time of Lineages 5 and 6 based on the assumption of co-divergence with human populations. The four calibration points were 185,000, 70,000, 65,000 and 10,000 years based on the emergence of anatomically modern humans, the emergence of mitochondrial haplogroup L3, the emergence of mitochondrial haplogroup M and the beginning of the Neolith Demographic Transition, respectively. The estimates based on the 70,000 hypothesis were considered the most likely given known human migration events and archaeological evidence

that might account for the discrepant findings. For a more comprehensive discussion of the different studies and analytical approaches taken, see (Brites and Gagneux 2015; Eldholm et al. 2016) (Fig. 1.3).

A study by Comas et al. (2013) proposed that the MRCA of all members of the MTBC existed around 70,000 years ago (range 50,000–96,000) (Comas et al. 2013). This estimate was obtained by attributing known split times in humans to internal nodes in the MTBC phylogeny based on the assumption of co-divergence with human populations. This age estimate contrasts with more recent estimates obtained by calibrating the MTBC phylogeny with the age of sequenced MTBC ancient DNA (aDNA). Two independent studies that have used such an approach (Bos et al. 2014; Kay et al. 2015) point to the origin of the MRCA of the MTBC to less than 6000 years ago. In the study by Bos et al. three aDNA samples of MTBC were isolated from human remains in Peru dated to approximately 1000 AD. Interestingly, whole genome analysis of these samples revealed that the infecting agents were closely related to *M. pinnipedii*, an animal-adapted member of the MTBC which today is known to mainly infect seals and sea-lions (Chap. 7). In the study by Kay et al. the aDNA was isolated from human remains from the eighteenth-century in Hungary (Kay et al. 2015). In these cases, infections were caused by MTBC Lineage 4. The considerably younger age obtained by tip-dating the MTBC phylogeny with aDNA implies that substitutions require considerably less time (approximately tenfold difference) to appear than in the estimates by Comas et al. (2013). In fact, the substitution rate per time unit obtained by tip-dating MTBC with aDNA is similar to that obtained by tip-dating MTBC with contemporaneous samples (Eldholm et al. 2015, 2016; Roetzer et al. 2013) and to rates obtained by directly counting the number of mutations accumulated between pairs of related samples separated in time (Walker et al. 2013; Eldholm et al. 2016).

It is well recognized that applying a molecular clock to phylogenies has several pit-falls, and

that not necessarily all estimates are valid. For example, it is crucial that the genetic data exhibits a temporal signal. A lack of temporal signal can be caused by short sampling periods, during which too little measurable molecular evolution has occurred, or by the fact that the set of samples analysed have evolutionary rates which are too variable, meaning that rates validated in one part of the tree cannot be extrapolated to other parts of the tree (Drummond et al. 2003; Murray et al. 2016). As shown by Duchene et al. (2016), not all datasets are equally suitable for molecular dating in MTBC, calling for careful examination of such heterogeneity in order to obtain reliable estimates of substitution rates and divergence times.

The age estimates based on aDNA and contemporaneous samples are similar, and all suggest a common ancestor of the MTBC much younger than previously thought (Eldholm et al. 2016). Yet, an origin of only around 6000 years for MTBC is difficult to reconcile with the African origin of MTBC, for which the evidence is very strong. And how to explain the concordance of the phylogenetic splits that lead to the most basal groups in both MTBC and *Homo sapiens*? Moreover, such a young age also implies that the MTBC host specialization to quite disparate mammal species must have occurred during a very short period of only 4000 years. Finally, several pieces of aDNA evidence from archaeological remains also predate that timing (reviewed in Brites and Gagneux 2015). The authentication of ancient sequence data is notoriously difficult, and the validity of some of the oldest aDNA findings has been debated (Wilbur et al. 2009; Donoghue et al. 2009). It is now technically possible to use enrichment techniques to increase DNA amounts and to generate DNA damage profiles which confirm degradation typical of ancient molecules (Orlando et al. 2015). Hopefully, these and other technical advances in next generation sequencing can contribute to position phylogenetically some of the oldest molecular evidence of MTBC such as the one found in the 17,000 year old bison bone in Wyoming (Rothschild et al. 2001), in the 9000 years old humans skeletons of Atlit-Yam in Israel (Hershkovitz et al. 2008) or in the similarly aged skeletons from Syria (Baker et al. 2015). However, the genetic diversity landscape of the MTBC that can be accessed with WGS has been shaped by lineage extinctions, population bottlenecks and genetic hitchhiking, all of which can lead to a much younger estimate of the MTBC ancestor than the age of the first MTBC ancestors that caused TB disease in humans (Smith et al. 2009). There are thus intrinsic limitations in the usage of molecular dating to study co-evolution of the MTBC with humans that we might not be able to overcome.

1.6 The Ecology of the Human-Adapted MTBC

1.6.1 Local Adaptation

Irrespective of the age of the human-MTBC association, the strong phylogeographic structure of the human-adapted MTBC lineages (Fig. 1.1) has led to the hypothesis that particular MTBC variants might be locally adapted to specific human populations (Gagneux 2012). Local adaptation refers to the phenomenon, in which a given pathogen genotype has a higher fitness in a host population with which it has co-evolved (sympatric association) than in a different host population (allopatric association) (Kawecki and Ebert 2004). The notion of local adaptation in the human-adapted MTBC originally stemmed from the observation that in metropolitan settings such as San Francisco, Montreal and London, where diverse human and pathogen genotypes mix, the sympatric host-pathogen associations remained stable (Hirsh et al. 2004; Gagneux et al. 2006b; Reed et al. 2009; Baker et al. 2004; Fenner et al. 2013). Moreover, studies in San Francisco and Switzerland showed that MTBC lineages preferentially transmitted among sympatric hosts (Fenner et al. 2013; Gagneux et al. 2006b).

The sympatry patterns in MTBC transmission could also be driven by the preferential social mixing among people of the same ethnicity. However, a biological effect is suggested by the

observation that in TB patients with impaired immune system caused by HIV co-infection, these sympatric host-pathogen associations are perturbed (Gagneux et al. 2006b; Fenner et al. 2013). Strikingly, a nation-wide study from Switzerland found that not only were allopatric host-pathogen combinations associated with HIV co-infection, but this association became stronger with increasing immune suppression as measured by decreasing $CD4^+$ T cell counts, and remained statistically significant after allowing for social factors such as frequent travelling to TB-endemic countries (Fenner et al. 2013).

M. africanum Lineage 5 makes a compelling case for locally adapted MTBC. In addition to being restricted to West-Africa (Fig. 1.1, see also Chap. 6), two separate studies reported a strong independent association between Lineage 5 and the Ewe ethnic group in Ghana (Asante-Poku et al. 2015, 2016) (see Chap. 6). Lineage 5 has mainly been reported in the West-African countries of Ghana, Benin, and Nigeria (de Jong et al. 2010b), places where Ewes and other related human populations live. Moreover, despite centuries of trans-Atlantic slave trade between West-Africa and the New World, *M. africanum* is hardly found in the Americas, except in recent immigrants from West-Africa or in US-born individuals with African ancestry (Sharma et al. 2016).

M. africanum Lineage 6 is also largely restricted to West-Africa (Fig. 1.1). However, its relatively close phylogenetic relationship with the animal-adapted lineages of the MTBC has led to the hypothesis that an as yet unknown animal reservoir might exist for this lineage (further discussed in Chap. 6). The comparably low virulence of Lineage 6 in infection models (de Jong et al. 2010b), and the observation that both in The Gambia and in Ghana Lineage 6 was independently associated with HIV co-infection (Asante-Poku et al. 2016; de Jong et al. 2010a) support the view that this lineage might behave like an opportunistic human pathogen (de Jong et al. 2005).

1.6.2 Specialists and Generalists

Local adaption of a pathogen to a particular host is a form of ecological specialisation. In other words, locally adapted MTBC strains can be considered specialists. By contrast, generalists are organisms that thrive across a wider ecological niche (Futuyma and Moreno 1988). As mentioned above, MTBC Lineage 4 is geographically the most widespread cause of human TB (Fig. 1.1). Hence, Lineage 4 could be considered a generalist compared to e.g. MTBC Lineages 5, 6 and 7. Based on the findings of many experimental and molecular epidemiological studies (for detailed reviews see Coscolla and Gagneux 2014, 2010) (see also Chap. 5), Lineage 4 (together with Lineage 2) appears to be more virulent and more transmissible in average than other MTBC lineages. However, this is not universally true, and much variation exists among Lineage 4 strains. Several studies have used WGS to demonstrate that Lineage 4 can be further subdivided into several sublineages (Coll et al. 2014; Stucki et al. 2016; Malm et al. 2017) (Fig. 1.4). These sublineages partially reflected strain families previously defined based on various genotyping techniques (Demay et al. 2012) (see Chap. 3). In a recent study, Stucki et al. screened 3366 Lineage 4 clinical isolates from 100 countries to determine the global frequency and geographical distribution of ten Lineage 4 sublineages (Stucki et al. 2016). This work revealed that four of these ten sublineages were geographically restricted, much like MTBC Lineage 5, 6 and 7 discussed above, suggesting that within the globally most successful Lineage 4, several specialist sublineages exist (Stucki et al. 2016). These included sublineages L4.6.1/Uganda and L4.6.2/Cameroon, which are restricted to parts of Africa (Stucki et al. 2016) (Fig. 1.5). L4.6.1/Uganda is the most common cause of TB in Uganda (Wampande et al. 2013), and has a distribution largely restricted to Uganda and neighbouring countries (Stucki et al. 2016). Moreover, its presence in the region dates back to at least several centuries (Comas et al. 2015),

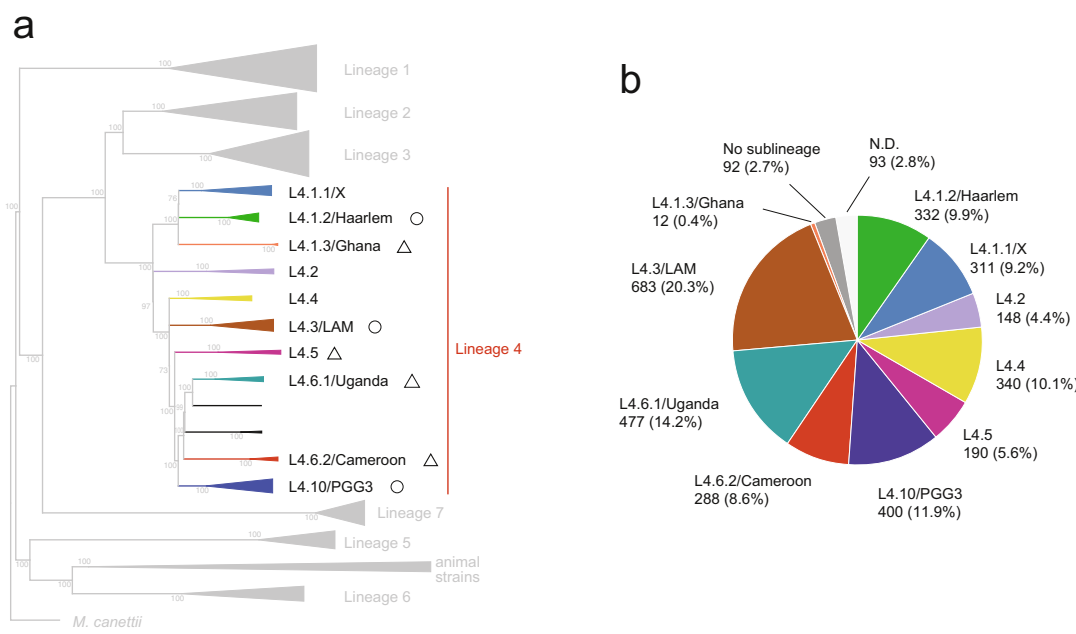

Fig. 1.4 Figure adapted from Stucki et al. (2016) representing the phylogenetic substructure of Lineage 4 and the frequency of Lineage 4 sub-lineages; (**a**) Ten sublineages were defined based on 72 MTBC Lineage 4 whole genome sequences. Sublineages were labelled according to Coll et al. (2014) when possible and previous designations based on spoligotyping. Black triangles represent sublineages identified as specialists, and black circles represent generalists; (**b**) Occurrence of each sub-lineage from a total of 3366 Lineage 4 isolates screened with sublineage specific SNPs defined from **a** (*ND* sub-lineage not defined)

which provides additional evidence against the notion of an African "virgin soil" for TB as discussed above. By contrast, three of the ten Lineage 4 sublineages exhibited a broad geographic distribution (Fig. 1.5), indicating that these sublineages might represent generalists capable of infecting and causing disease in many different human populations (Stucki et al. 2016). In support of this notion, these generalist sublineages occurred in more than 50 countries each, compared to less than 15 countries in the case of specialist sublineages (Fig. 1.5).

The most frequent and globally widespread Lineage 4 sublineage is L4.3/LAM. The broad distribution of L4.3/LAM can be attributed to migratory and colonization processes led by Europeans, which started in the fifteenth century with Portuguese and Spanish colonizers and lasted several centuries (Mokrousov et al. 2016). This notion is supported by the fact that Europe is the most likely geographical origin of the MRCA of L4.3/LAM and that the genetic diversity of this sublineage is significantly higher in Europe than on all other continents (Stucki et al. 2016). Opportunities for dispersal brought about by human movement have thus been crucial to the evolutionary success of generalist sublineages such as L4.3/LAM. However, as referred above, an estimated 11 million Africans were shipped to the Americas during the Atlantic slave trade (Lovejoy 1989), providing opportunities for what we known today as extant African MTBC groups to disperse to the Americas. The niche restriction of, for example L4.6.1/Uganda and L4.6.2/Cameroon, is possibly not only caused by a lack of opportunity of dispersal via human migration, trade and conquest, but biological determinants may have also played a role. These sublineages might have specialized in certain human population at the cost of decreased

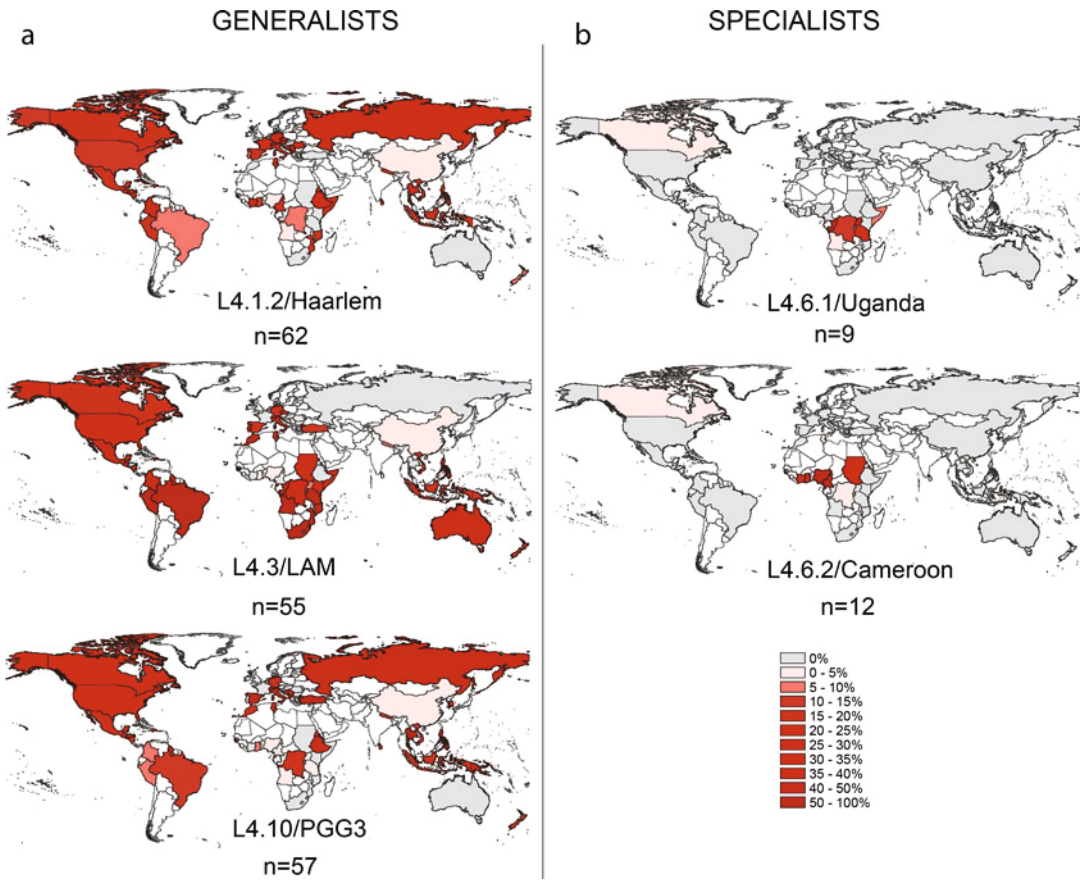

Fig. 1.5 Figure adapted from Stucki et al. (2016) representing the country specific proportions of each sublineage which underlie the definitions of generalist (**a**) and specialist (**b**) sub-lineages of Lineage 4. Intensity of red on the scale corresponds to the proportion of a sublineage among all Lineage 4 isolates in each country. 3366 Lineage 4 isolates from 100 countries were screened with sub-lineage specific SNPs defined from Fig. 1.4a. The number of countries where each sub-lineage was found is indicated below each sub-lineage map

performance in other populations, especially in the presence of other more competitive MTBC strains. Stucki et al. also found that the generalist Lineage 4 sublineages had more variable T cell epitopes than the specialist L4.6/Uganda, which could reflect contact with a broader population of hosts (Stucki et al. 2016). However, an understanding of the biological underpinnings of what makes an MTBC strain behave as specialist or generalist is still lacking.

1.6.3 Evolution Towards Higher Virulence

One way by which generalist MTBC genotypes might differ from specialists is in their overall virulence and transmission potential. As mentioned above, and discussed in more detail in Chaps. 5 and 6, MTBC strains have been shown to differ in immunogenicity and virulence in human macrophages and multiple animal models

of infection (Coscolla and Gagneux 2014). MTBC strains also differ in their transmissibility (Coscolla and Gagneux 2014) (Chap. 5). It has been hypothesized that during hunter-gatherer times, the MTBC might have benefitted from reduced virulence and a prolonged latency period to avoid exhausting the limited pool of susceptible hosts by jumping generations and reaching new birth cohorts (Blaser and Kirschner 2007; Comas and Gagneux 2011). Recent mathematical models support such a scenario (Zheng et al. 2014). If at that time, active TB primarily affected older individuals with a weakening immune system, TB disease is likely to have exerted minimal selection pressure on human populations because TB disease would have affected most individuals past their reproductive age. Later, increases in human population densities during the Agricultural and Industrial Revolutions would then have selected for increased virulence and shortened latency in some MTBC lineages (Comas and Gagneux 2011; Comas et al. 2013; Hershberg et al. 2008; Merker et al. 2015; Luo et al. 2015; de Jong et al. 2008). Because the human-adapted MTBC is an obligate pathogen as defined above, such changes in virulence and latency are expected in the context of an expanding host population (May and Anderson 1983). Moreover, such an increased virulence and shortened latency period would have enhanced the selective pressure on humans. Indirect support for this notion comes from a study that found human populations with a long history of urbanization to have higher frequencies of a protective allele of the gene *SLC11A1,* which is a gene of the innate immune system with polymorphisms frequently associated with resistance to intracellular infections (Barnes et al. 2011).

Another study however concluded that the sharp declines in TB deaths in European and North American cities preceding the start of antimicrobial therapies were unlikely due to increased genetic resistance induced by natural selection (Lipsitch and Sousa 2002). Instead, it was recently hypothesised that increasing meat consumption during that time might have led to a reduction of TB in these settings (Williams and Dunbar 2014). Correlations of historical and ancient events with extant human allele frequencies are however difficult to pursue because, except for rare Mendelian disorders, the genetic basis of susceptibility to TB remains poorly understood (Abel et al. 2014).

1.7 MTBC and *Homo sapiens* Gene-to-Gene Interactions

In order to transmit, human-adapted MTBC strains have to harm their hosts. Generally, more than 50% of TB patients die if left untreated (Tiemersma et al. 2011). Human populations have therefore probably evolved increased resistance to the MTBC. This is supported by a clear genetic basis for susceptibility to MTBC infection and disease (reviewed in Abel et al. 2014). Some of the strongest evidence for the heritability of TB susceptibility stems from early twin studies (Abel et al. 2014), but is also reflected by the apparent genetic resistance exhibited by individuals with a high exposure to TB who never become infected, and by the fact the human populations differ in their susceptibilities to TB (Abel et al. 2014; Stead et al. 1990). Still, 80–90% of immune-competent people exposed to MTBC do become infected and contribute to the pool of latent TB. The majority of individuals remain latently infected with no clinical symptoms, whereas 5–10% will developed TB at some point during their lives (Ernst 2012). Some key components and molecules fundamental for a protective human immune response are well known, as for example the importance of T cell mediated immunity, the roles of interferon-gamma gene (*IFNG*) and of Tumor Necrosis Factor (*TNF*) among others (Flynn and Chan 2005). However, the identification of the genetic variants underlying

the variation in TB susceptibility has had limited success (Qu et al. 2011; Abel et al. 2014). Studies based on candidate gene approaches have highlighted numerous polymorphisms in genes associated with TB susceptibility (Moller and Hoal 2010; Azad et al. 2012). However, genome-wide association studies (GWAS) have largely failed to replicate the results of candidate studies and identified only a few new loci with small effects (Abel et al. 2014; Thye et al. 2010; Grant et al. 2016). Several factors account for the difficulties in uncovering genetic variants associated with TB susceptibility and in replicating findings across different studies; a strong influence of the environment in the outcome of an infection (Young et al. 2008), the possible influence of many small effect alleles, difficulties in defining and quantifying the relevant phenotype (O'Garra et al. 2013; Esmail et al. 2014), and finally, and more important in the context of this chapter, co-evolution between the host and pathogen (Woolhouse et al. 2002; Karlsson et al. 2014).

If human populations and the human-adapted members of the MTBC have co-evolved for some time, one would expect that the outcome of the interaction between both partners depends on their respective genotypes (Woolhouse et al. 2002). The local adaptation patterns described above would be a manifestation of such interactions. If the patterns of human variation to TB susceptibility have been in part shaped by diversity in the pathogen, taking the variation of the pathogen into account might increase the statistical power of human genetic association studies (Karlsson et al. 2014). A few studies based on candidate genes where both host and pathogen genotypes were analyzed suggest that such interactions can influence clinical phenotypes. In most of these studies, particular variants of candidate genes were associated with TB caused by particular MTBC genotypes. For example in Vietnam, individuals carrying the T597C change of *TLR*-2 gene (Toll-like receptor 2) were more likely to be infected with Lineage 2 strains from the Beijing family than with other strains (Caws et al. 2008). In Ghana, the variant G57E of *MBL2* (*Mannose-binding Lectin)* gene was associated with protection against TB caused by *M. africanum* as opposed to *M. tuberculosis sensu stricto* (Thye et al. 2011). In South Africa, different HLA class I types variants were associated with different strains of human-adapted MTBC (Salie et al. 2014).

So far, MTBC strain diversity has not been considered in GWAS studies of human susceptibility to TB. Such an approach has however been successfully applied to uncover associations between viral variants and host DNA polymorphisms (Bartha et al. 2013; Ansari et al. 2017). Such "genome-to-genome" analyses have a great potential to uncover relevant interactions between the pathogen and host genome and their joint effects on phenotypes such as e.g. viral load. In TB, given the difficulties in defining relevant clinical phenotypes, and the long co-evolutionary relationship with human populations, genome-to-genome approaches offer a promising avenue for deciphering the genetic loci involved in that relationship (Kodaman et al. 2014).

1.8 Conclusions

A detailed understanding of the various factors that determine whether an immunocompetent person infected with the MTBC will develop active TB is lacking (Russell et al. 2010). On the host side, the molecules and mechanisms that make up a protective immune response remain largely unknown despite many past and ongoing efforts (Orme et al. 2015). The genetic diversity of the pathogen is increasingly being recognized as playing a role in the outcome of TB infection and disease (Coscolla and Gagneux 2014), but here too, the underlying mechanisms remain unknown. The human-adapted members of the MTBC have co-evolved with human populations for at least several thousands, perhaps even tens of thousands of years (Brites and Gagneux 2015). As a likely consequence, human populations and MTBC strains differ in their susceptibility/resistance and virulence characteristics, respectively. Hence, the outcome of TB infection and disease will be partially

determined by the interaction between the host and pathogen genotypes (Kodaman et al. 2014). Identifying the specific human and bacterial genomic loci involved in this interaction could shed new light onto the molecular determinants of host immune protection and pathogen virulence. Moreover, these interacting loci could provide attractive targets for the development of new TB treatments and vaccines. The interaction between human and MTBC diversity will also be relevant for the development of host-directed therapies that are increasingly being discussed as possible adjunct treatments for TB (Zumla et al. 2015). In today's era of personalized medicine, TB control of the future will likely benefit from a detailed understanding of the relevant human and MTBC diversity, as well as their interaction.

Acknowledgements We thank all the members of our group for the stimulating discussions. Work in our group is supported by the Swiss National Science Foundation (grants 310030_166687, IZRJZ3_164171 and IZLSZ3_170834), the European Research Council (309540-EVODRTB), SystemsX.ch, and the Novartis Foundation.

References

Abel L, El-Baghdadi J, Bousfiha AA, Casanova JL, Schurr E (2014) Human genetics of tuberculosis: a long and winding road. Philos Trans R Soc B 369(1645). doi:ARTN 20130428. 10.1098/rstb.2013.0428

Achtman M (2008) Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. Annu Rev Microbiol 62:53–70. doi:10.1146/annurev.micro.62.081307.162832

Ansari MA, Pedergnana V, C LCI, Magri A, Von Delft A, Bonsall D, Chaturvedi N, Bartha I, Smith D, Nicholson G, McVean G, Trebes A, Piazza P, Fellay J, Cooke G, Foster GR, Consortium S-H, Hudson E, McLauchlan J, Simmonds P, Bowden R, Klenerman P, Barnes E, Spencer CCA (2017) Genome-to-genome analysis highlights the effect of the human innate and adaptive immune systems on the hepatitis C virus. Nat Genet 49(5):666–673. doi:10.1038/ng.3835

Asante-Poku A, Yeboah-Manu D, Otchere ID, Aboagye SY, Stucki D, Hattendorf J, Borrell S, Feldmann J, Danso E, Gagneux S (2015) Mycobacterium africanum is associated with patient ethnicity in Ghana. PLoS Negl Trop Dis 9(1):e3370. doi:10.1371/journal.pntd.0003370

Asante-Poku A, Otchere ID, Osei-Wusu S, Sarpong E, Baddoo A, Forson A, Laryea C, Borrell S, Bonsu F, Hattendorf J, Ahorlu C, Koram KA, Gagneux S, Yeboah-Manu D (2016) Molecular epidemiology of Mycobacterium africanum in Ghana. BMC Infect Dis 16:385. doi:10.1186/s12879-016-1725-6

Azad AK, Sadee W, Schlesinger LS (2012) Innate immune gene polymorphisms in tuberculosis. Infect Immun 80(10):3343–3359. doi:10.1128/IAI.00443-12

Baker L, Brown T, Maiden MC, Drobniewski F (2004) Silent nucleotide polymorphisms and a phylogeny for Mycobacterium tuberculosis. Emerg Infect Dis 10(9):1568–1577. doi:10.3201/eid1009.040046

Baker O, Lee OY, Wu HH, Besra GS, Minnikin DE, Llewellyn G, Williams CM, Maixner F, O'Sullivan N, Zink A, Chamel B, Khawam R, Coqueugniot E, Helmer D, Le Mort F, Perrin P, Gourichon L, Dutailly B, Palfi G, Coqueugniot H, Dutour O (2015) Human tuberculosis predates domestication in ancient Syria. Tuberculosis (Edinb) 95(Suppl 1):S4–S12. doi:10.1016/j.tube.2015.02.001

Barnes I, Duda A, Pybus OG, Thomas MG (2011) Ancient urbanization predicts genetic resistance to tuberculosis. Evolution 65(3):842–848. doi:10.1111/j.1558-5646.2010.01132.x

Barry CE 3rd, Boshoff HI, Dartois V, Dick T, Ehrt S, Flynn J, Schnappinger D, Wilkinson RJ, Young D (2009) The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. Nat Rev Microbiol 7(12):845–855. doi:10.1038/nrmicro2236

Bartha I, Carlson JM, Brumme CJ, McLaren PJ, Brumme ZL, John M, Haas DW, Martinez-Picado J, Dalmau J, Lopez-Galindez C, Casado C, Rauch A, Gunthard HF, Bernasconi E, Vernazza P, Klimkait T, Yerly S, O'Brien SJ, Listgarten J, Pfeifer N, Lippert C, Fusi N, Kutalik Z, Allen TM, Muller V, Harrigan PR, Heckerman D, Telenti A, Fellay J (2013) A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. elife 2:e01123. doi:10.7554/eLife.01123

Behr MA (2015) Comparative genomics of mycobacteria: some answers, yet more new questions. Cold Spring Harb Perspect Med 5(2):a021204. doi:10.1101/cshperspect.a021204

Blaser MJ, Kirschner D (2007) The equilibria that allow bacterial persistence in human hosts. Nature 449(7164):843–849. doi:10.1038/nature06198

Bloemberg GV, Keller PM, Stucki D, Trauner A, Borrell S, Latshang T, Coscolla M, Rothe T, Homke R, Ritter C, Feldmann J, Schulthess B, Gagneux S, Bottger EC (2015) Acquired resistance to bedaquiline and delamanid in therapy for tuberculosis. N Engl J Med 373(20):1986–1988. doi:10.1056/NEJMc1505196

Blouin Y, Cazajous G, Dehan C, Soler C, Vong R, Hassan MO, Hauck Y, Boulais C, Andriamanantena D, Martinaud C, Martin E, Pourcel C, Vergnaud G (2014) Progenitor "Mycobacterium canettii" clone responsible for lymph node tuberculosis epidemic, Djibouti. Emerg Infect Dis 20(1):21–28. doi:10.3201/eid2001.130652

Bolotin E, Hershberg R (2015) Gene loss dominates as a source of genetic variation within clonal pathogenic bacterial species. Genome Biol Evol 7(8):2173–2187. doi:10.1093/gbe/evv135

Boritsch EC, Khanna V, Pawlik A, Honore N, Navas VH, Ma L, Bouchier C, Seemann T, Supply P, Stinear TP, Brosch R (2016) Key experimental evidence of chromosomal DNA transfer among selected tuberculosis-causing mycobacteria. Proc Natl Acad Sci U S A 113(35):9876–9881. doi:10.1073/pnas.1604921113

Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, Forrest SA, Bryant JM, Harris SR, Schuenemann VJ, Campbell TJ, Majander K, Wilbur AK, Guichon RA, Steadman DLW, Cook DC, Niemann S, Behr MA, Zumarraga M, Bastida R, Huson D, Nieselt K, Young D, Parkhill J, Buikstra JE, Gagneux S, Stone AC, Krause J (2014) Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature 514(7523):494–497. doi:10.1038/nature13591

Brites D, Gagneux S (2012) Old and new selective pressures on Mycobacterium tuberculosis. Infect Genet Evol 12(4):678–685. doi:10.1016/j.meegid.2011.08.010

Brites D, Gagneux S (2015) Co-evolution of Mycobacterium tuberculosis and *Homo sapiens*. Immunol Rev 264(1):6–24. doi:10.1111/imr.12264

Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, Garnier T, Gutierrez C, Hewinson G, Kremer K, Parsons LM, Pym AS, Samper S, van Soolingen D, Cole ST (2002) A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci U S A 99(6):3684–3689. doi:10.1073/pnas.052548299

Caws M, Thwaites G, Dunstan S, Hawn TR, Lan NT, Thuong NT, Stepniewska K, Huyen MN, Bang ND, Loc TH, Gagneux S, van Soolingen D, Kremer K, van der Sande M, Small P, Anh PT, Chinh NT, Quy HT, Duyen NT, Tho DQ, Hieu NT, Torok E, Hien TT, Dung NH, Nhu NT, Duy PM, van Vinh CN, Farrar J (2008) The influence of host and bacterial genotype on the development of disseminated disease with Mycobacterium tuberculosis. PLoS Pathog 4(3):e1000034. doi:10.1371/journal.ppat.1000034

Chisholm RH, Trauer JM, Curnoe D, Tanaka MM (2016) Controlled fire use in early humans might have triggered the evolutionary emergence of tuberculosis. Proc Natl Acad Sci U S A 113(32):9051–9056. doi:10.1073/pnas.1603224113

Colangeli R, Arcus VL, Cursons RT, Ruthe A, Karalus N, Coley K, Manning SD, Kim S, Marchiano E, Alland D (2014) Whole genome sequencing of Mycobacterium tuberculosis reveals slow growth and low mutation rates during latent infections in humans. PLoS One 9(3):e91024. doi:10.1371/journal.pone.0091024

Coll F, McNerney R, Guerra-Assuncao JA, Glynn JR, Perdigao J, Viveiros M, Portugal I, Pain A, Martin N, Clark TG (2014) A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. Nat Commun 5:4812. doi:10.1038/ncomms5812

Comas I, Gagneux S (2009) The past and future of tuberculosis research. PLoS Pathog 5(10):e1000600. doi:10.1371/journal.ppat.1000600

Comas I, Gagneux S (2011) A role for systems epidemiology in tuberculosis research. Trends Microbiol 19(10):492–500. doi:10.1016/j.tim.2011.07.002

Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, Kremer K, Ernst JD, Gagneux S (2010) Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved. Nat Genet 42(6):498–503. doi:10.1038/ng.590

Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, Galagan J, Niemann S, Gagneux S (2012) Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. Nat Genet 44(1):106–110. doi:10.1038/ng.1038

Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, Parkhill J, Malla B, Berg S, Thwaites G, Yeboah-Manu D, Bothamley G, Mei J, Wei LH, Bentley S, Harris SR, Niemann S, Diel R, Aseffa A, Gao Q, Young D, Gagneux S (2013) Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. Nat Genet 45(10):1176–1182. doi:10.1038/ng.2744

Comas I, Hailu E, Kiros T, Bekele S, Mekonnen W, Gumi B, Tschopp R, Ameni G, Hewinson RG, Robertson BD, Goig GA, Stucki D, Gagneux S, Aseffa A, Young D, Berg S (2015) Population genomics of Mycobacterium tuberculosis in Ethiopia contradicts the virgin soil hypothesis for human tuberculosis in Sub-Saharan Africa. Curr Biol 25(24):3260–3266. doi:10.1016/j.cub.2015.10.061

Copin R, Coscolla M, Efstathiadis E, Gagneux S, Ernst JD (2014) Impact of in vitro evolution on antigenic diversity of Mycobacterium bovis bacillus Calmette-Guerin (BCG). Vaccine 32(45):5998–6004. doi:10.1016/j.vaccine.2014.07.113

Copin R, Wang XY, Louie E, Escuyer V, Coscolla M, Gagneux S, Palmer GH, Ernst JD (2016) Within host evolution selects for a dominant genotype of Mycobacterium tuberculosis while T cells increase pathogen genetic diversity. PLoS Pathog 12(12). doi:ARTN e1006111. 10.1371/journal.ppat.1006111

Coscolla M, Gagneux S (2010) Does M. tuberculosis genomic diversity explain disease diversity? Drug Discov Today Dis Mech 7(1):e43–e59. doi:10.1016/j.ddmec.2010.09.004

Coscolla M, Gagneux S (2014) Consequences of genomic diversity in Mycobacterium tuberculosis. Semin Immunol 26(6):431–444. doi:10.1016/j.smim.2014.09.012

Coscolla M, Lewin A, Metzger S, Maetz-Rennsing K, Calvignac-Spencer S, Nitsche A, Dabrowski PW, Radonic A, Niemann S, Parkhill J, Couacy-Hymann E, Feldman J, Comas I, Boesch C, Gagneux S,

Leendertz FH (2013) Novel Mycobacterium tuberculosis complex isolate from a wild chimpanzee. Emerg Infect Dis 19(6):969–976. doi:10.3201/eid1906.121012

Coscolla M, Copin R, Sutherland J, Gehre F, de Jong B, Owolabi O, Mbayo G, Giardina F, Ernst JD, Gagneux S (2015) M. tuberculosis T cell epitope analysis reveals paucity of antigenic variation and identifies rare variable TB antigens. Cell Host Microbe 18(5):538–548. doi:10.1016/j.chom.2015.10.008

de Jong BC, Hill PC, Brookes RH, Otu JK, Peterson KL, Small PM, Adegbola RA (2005) Mycobacterium africanum: a new opportunistic pathogen in HIV infection? AIDS 19(15):1714–1715

de Jong BC, Hill PC, Aiken A, Awine T, Antonio M, Adetifa IM, Jackson-Sillah DJ, Fox A, Deriemer K, Gagneux S, Borgdorff MW, McAdam KP, Corrah T, Small PM, Adegbola RA (2008) Progression to active tuberculosis, but not transmission, varies by Mycobacterium tuberculosis lineage in The Gambia. J Infect Dis 198(7):1037–1043. doi:10.1086/591504

de Jong BC, Adetifa I, Walther B, Hill PC, Antonio M, Ota M, Adegbola RA (2010a) Differences between tuberculosis cases infected with Mycobacterium africanum, West African type 2, relative to Euro-American Mycobacterium tuberculosis: an update. FEMS Immunol Med Microbiol 58(1):102–105. doi:10.1111/j.1574-695X.2009.00628.x

de Jong BC, Antonio M, Gagneux S (2010b) Mycobacterium africanum – review of an important cause of human tuberculosis in West Africa. PLoS Negl Trop Dis 4(9):e744. doi:10.1371/journal.pntd.0000744

de Vos M, Muller B, Borrell S, Black PA, van Helden PD, Warren RM, Gagneux S, Victor TC (2013) Putative compensatory mutations in the rpoC gene of rifampin-resistant Mycobacterium tuberculosis are associated with ongoing transmission. Antimicrob Agents Chemother 57(2):827–832. doi:10.1128/Aac.01541-12

Demay C, Liens B, Burguiere T, Hill V, Couvin D, Millet J, Mokrousov I, Sola C, Zozio T, Rastogi N (2012) SITVITWEB – a publicly available international multimarker database for studying Mycobacterium tuberculosis genetic diversity and molecular epidemiology. Infect Genet Evol 12(4):755–766. doi:10.1016/j.meegid.2012.02.004

Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ (2016) Within-host evolution of bacterial pathogens. Nat Rev Microbiol 14(3):150–162. doi:10.1038/nrmicro.2015.13

Donoghue HD, Hershkovitz I, Minnikin DE, Besra GS, Lee OYC, Galili E, Greenblatt CL, Lemma E, Spigelman M, Bar-Gal GK (2009) Biomolecular archaeology of ancient tuberculosis: response to "Deficiencies and challenges in the study of ancient tuberculosis DNA" by Wilbur et al. (2009). J Archaeol Sci 36(12):2797–2804. doi:10.1016/j.jas.2009.09.007

Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG (2003) Measurably evolving populations. Trends Ecol Evol 18(9):481–488. doi:10.1016/S0169-5347(03)00216-7

Duchene S, Holt KE, Weill FX, Le Hello S, Hawkey J, Edwards DJ, Fourment M, Holmes EC (2016) Genome-scale rates of evolutionary change in bacteria. Microb Genom 2(11):e000094. doi:10.1099/mgen.0.000094

Dye C, Lonnroth K, Jaramillo E, Williams BG, Raviglione M (2009) Trends in tuberculosis incidence and their determinants in 134 countries. Bull World Health Organ 87(9):683–691. doi:10.2471/Blt.08.058453

Eldholm V, Norheim G, von der Lippe B, Kinander W, Dahle UR, Caugant DA, Mannsaker T, Mengshoel AT, Dyrhol-Riise AM, Balloux F (2014) Evolution of extensively drug-resistant Mycobacterium tuberculosis from a susceptible ancestor in a single patient. Genome Biol 15(11):490. doi:10.1186/s13059-014-0490-3

Eldholm V, Monteserin J, Rieux A, Lopez B, Sobkowiak B, Ritacco V, Balloux F (2015) Four decades of transmission of a multidrug-resistant Mycobacterium tuberculosis outbreak strain. Nat Commun 6:7119. doi:10.1038/ncomms8119

Eldholm V, Pettersson JH, Brynildsrud OB, Kitchen A, Rasmussen EM, Lillebaek T, Ronning JO, Crudu V, Mengshoel AT, Debech N, Alfsnes K, Bohlin J, Pepperell CS, Balloux F (2016) Armed conflict and population displacement as drivers of the evolution and dispersal of Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 113(48):13881–13886. doi:10.1073/pnas.1611283113

Ernst JD (2012) The immunological life cycle of tuberculosis. Nat Rev Immunol 12(8):581–591. doi:10.1038/nri3259

Esmail H, Barry CE 3rd, Young DB, Wilkinson RJ (2014) The ongoing challenge of latent tuberculosis. Philos Trans R Soc Lond Ser B Biol Sci 369(1645):20130437. doi:10.1098/rstb.2013.0437

Fenner L, Egger M, Bodmer T, Furrer H, Ballif M, Battegay M, Helbling P, Fehr J, Gsponer T, Rieder HL, Zwahlen M, Hoffmann M, Bernasconi E, Cavassini M, Calmy A, Dolina M, Frei R, Janssens JP, Borrell S, Stucki D, Schrenzel J, Bottger EC, Gagneux S, Swiss HIVC, Molecular Epidemiology of Tuberculosis Study G (2013) HIV infection disrupts the sympatric host-pathogen relationship in human tuberculosis. PLoS Genet 9(3):e1003318. doi:10.1371/journal.pgen.1003318

Firdessa R, Berg S, Hailu E, Schelling E, Gumi B, Erenso G, Gadisa E, Kiros T, Habtamu M, Hussein J, Zinsstag J, Robertson BD, Ameni G, Lohan AJ, Loftus B, Comas I, Gagneux S, Tschopp R, Yamuah L, Hewinson G, Gordon SV, Young DB, Aseffa A (2013) Mycobacterial lineages causing pulmonary and extrapulmonary tuberculosis, Ethiopia. Emerg Infect Dis 19(3):460–463. doi:10.3201/eid1903.120256

Flynn J, Chan J (2005) What's good for the host is good for the bug. Trends Microbiol 13:98–102

Flynn JL, Gideon HP, Mattila JT, Lin PL (2015) Immunology studies in non-human primate models of tuberculosis. Immunol Rev 264(1):60–73. doi:10.1111/imr.12258

Ford CB, Lin PL, Chase MR, Shah RR, Iartchouk O, Galagan J, Mohaideen N, Ioerger TR, Sacchettini JC, Lipsitch M, Flynn JL, Fortune SM (2011) Use of whole genome sequencing to estimate the mutation rate of Mycobacterium tuberculosis during latent infection. Nat Genet 43(5):482–486. doi:10.1038/ng.811

Ford CB, Shah RR, Maeda MK, Gagneux S, Murray MB, Cohen T, Johnston JC, Gardy J, Lipsitch M, Fortune SM (2013) Mycobacterium tuberculosis mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. Nat Genet 45(7):784–790. doi:10.1038/ng.2656

Futuyma DJ, Moreno G (1988) The evolution of ecological specialization. Annu Rev Ecol Syst 19:207–233. doi:10.1146/annurev.ecolsys.19.1.207

Gagneux S (2012) Host-pathogen coevolution in human tuberculosis. Philos Trans R Soc Lond Ser B Biol Sci 367(1590):850–859. doi:10.1098/rstb.2011.0316

Gagneux S, Small PM (2007) Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development. Lancet Infect Dis 7(5):328–337. doi:10.1016/S1473-3099(07)70108-1

Gagneux S, Burgos MV, DeRiemer K, Encisco A, Munoz S, Hopewell PC, Small PM, Pym AS (2006a) Impact of bacterial genetics on the transmission of isoniazid-resistant Mycobacterium tuberculosis. PLoS Pathog 2(6):e61. doi:10.1371/journal.ppat.0020061

Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, Nicol M, Niemann S, Kremer K, Gutierrez MC, Hilty M, Hopewell PC, Small PM (2006b) Variable host-pathogen compatibility in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 103(8):2869–2873. doi:10.1073/pnas.0511240103

Gill WP, Harik NS, Whiddon MR, Liao RP, Mittler JE, Sherman DR (2009) A replication clock for Mycobacterium tuberculosis. Nat Med 15(2):211–214. doi:10.1038/nm.1915

Gonzalo-Asensio J, Malaga W, Pawlik A, Astarie-Dequeker C, Passemar C, Moreau F, Laval F, Daffe M, Martin C, Brosch R, Guilhot C (2014) Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence regulator. Proc Natl Acad Sci U S A 111(31):11491–11496. doi:10.1073/pnas.1406693111

Gordon SV, Brosch R, Billault A, Garnier T, Eiglmeier K, Cole ST (1999) Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. Mol Microbiol 32(3):643–655

Grant AV, Sabri A, Abid A, Abderrahmani Rhorfi I, Benkirane M, Souhi H, Naji Amrani H, Alaoui-Tahiri K, Gharbaoui Y, Lazrak F, Sentissi I, Manessouri M, Belkheiri S, Zaid S, Bouraqadi A, El Amraoui N, Hakam M, Belkadi A, Orlova M, Boland A, Deswarte C, Amar L, Bustamante J, Boisson-Dupuis S, Casanova JL, Schurr E, El Baghdadi J, Abel L (2016) A genome-wide association study of pulmonary tuberculosis in Morocco. Hum Genet 135(3):299–307. doi:10.1007/s00439-016-1633-2

Groschel MI, Sayes F, Simeone R, Majlessi L, Brosch R (2016) ESX secretion systems: mycobacterial evolution to counter host immunity. Nat Rev Microbiol 14(11):677–691. doi:10.1038/nrmicro.2016.131

Gutierrez MC, Brisse S, Brosch R, Fabre M, Omais B, Marmiesse M, Supply P, Vincent V (2005) Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis. PLoS Pathog 1(1):e5. doi:10.1371/journal.ppat.0010005

Gygli SM, Borrell S, Trauner A, Gagneux S (2017) Antimicrobial resistance in Mycobacterium tuberculosis: mechanistic and evolutionary perspectives. FEMS Microbiol Rev 41(3):354–373. doi:10.1093/femsre/fux011

Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, Roach JC, Kremer K, Petrov DA, Feldman MW, Gagneux S (2008) High functional diversity in Mycobacterium tuberculosis driven by genetic drift and human demography. PLoS Biol 6(12):e311. doi:10.1371/journal.pbio.0060311

Hershkovitz I, Donoghue HD, Minnikin DE, Besra GS, Lee OY, Gernaey AM, Galili E, Eshed V, Greenblatt CL, Lemma E, Bar-Gal GK, Spigelman M (2008) Detection and molecular characterization of 9000-year-old Mycobacterium tuberculosis from a Neolithic settlement in the Eastern Mediterranean. PLoS One 3(10):e3426. doi:10.1371/journal.pone.0003426

Hirsh AE, Tsolaki AG, DeRiemer K, Feldman MW, Small PM (2004) Stable association between strains of Mycobacterium tuberculosis and their human host populations. Proc Natl Acad Sci U S A 101(14):4871–4876. doi:10.1073/pnas.0305627101

Houben RM, Dodd PJ (2016) The global burden of latent tuberculosis infection: a re-estimation using mathematical modelling. PLoS Med 13(10):e1002152. doi:10.1371/journal.pmed.1002152

Jang J, Becq J, Gicquel B, Deschavanne P, Neyrolles O (2008) Horizontally acquired genomic islands in the tubercle bacilli. Trends Microbiol 16(7):303–308. doi:10.1016/j.tim.2008.04.005

Kaiser VB, Charlesworth B (2009) The effects of deleterious mutations on evolution in non-recombining genomes. Trends Genet 25(1):9–12. doi:10.1016/j.tig.2008.10.009

Karlsson EK, Kwiatkowski DP, Sabeti PC (2014) Natural selection and infectious disease in human populations. Nat Rev Genet 15(6):379–393. doi:10.1038/nrg3734

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. Ecol Lett 7(12):1225–1241. doi:10.1111/j.1461-0248.2004.00684.x

Kay GL, Sergeant MJ, Zhou Z, Chan JZ, Millard A, Quick J, Szikossy I, Pap I, Spigelman M, Loman NJ, Achtman M, Donoghue HD, Pallen MJ (2015) Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe. Nat Commun 6:6717. doi:10.1038/ncomms7717

Koch A, Brites D, Stucki D, Evans JC, Seldon R, Heekes A, Mulder N, Nicol M, Oni T, Warner DF, Mizrahi V, Parkhill J, Gagneux S, Martin DP, Wilkinson RJ (2017) The influence of HIV on the evolution of Mycobacterium tuberculosis. Mol Biol Evol. doi:10.1093/molbev/msx107

Kodaman N, Sobota RS, Mera R, Schneider BG, Williams SM (2014) Disrupted human-pathogen co-evolution: a model for disease. Front Genet 5. doi:UNSP 290. doi:10.3389/fgene.2014.00290

Koeck JL, Fabre M, Simon F, Daffe M, Garnotel E, Matan AB, Gerome P, Bernatas JJ, Buisson Y, Pourcel C (2011) Clinical characteristics of the smooth tubercle bacilli 'Mycobacterium canettii' infection suggest the existence of an environmental reservoir. Clin Microbiol Infect 17(7):1013–1019. doi:10.1111/j.1469-0691.2010.03347.x

Kwan CK, Ernst JD (2011) HIV and tuberculosis: a deadly human syndemic. Clin Microbiol Rev 24(2):351–378. doi:10.1128/Cmr.00042-10

Lieberman TD, Flett KB, Yelin I, Martin TR, McAdam AJ, Priebe GP, Kishony R (2014) Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. Nat Genet 46(1):82–87. doi:10.1038/ng.2848

Lieberman TD, Wilson D, Misra R, Xiong LL, Moodley P, Cohen T, Kishony R (2016) Genomic diversity in autopsy samples reveals within-host dissemination of HIV-associated Mycobacterium tuberculosis. Nat Med. doi:10.1038/nm.4205

Lin PL, Ford CB, Coleman MT, Myers AJ, Gawande R, Ioerger T, Sacchettini J, Fortune SM, Flynn JL (2014) Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing. Nat Med 20(1):75. doi:10.1038/nm.3412

Lindestam Arlehamn CS, Paul S, Mele F, Huang C, Greenbaum JA, Vita R, Sidney J, Peters B, Sallusto F, Sette A (2015) Immunological consequences of intragenus conservation of Mycobacterium tuberculosis T-cell epitopes. Proc Natl Acad Sci U S A 112(2):E147–E155. doi:10.1073/pnas.1416537112

Lipsitch M, Sousa AO (2002) Historical intensity of natural selection for resistance to tuberculosis. Genetics 161(4):1599–1607

Lonnroth K, Jaramillo E, Williams BG, Dye C, Raviglione M (2009) Drivers of tuberculosis epidemics: the role of risk factors and social determinants. Soc Sci Med 68(12):2240–2246. doi:10.1016/j.socscimed.2009.03.041

Lovejoy PE (1989) The impact of the Atlantic slave-trade on Africa – a review of the literature. J Afr Hist 30(3):365–394

Luo T, Comas I, Luo D, Lu B, Wu J, Wei L, Yang C, Liu Q, Gan M, Sun G, Shen X, Liu F, Gagneux S, Mei J, Lan R, Wan K, Gao Q (2015) Southern East Asian origin and coexpansion of Mycobacterium tuberculosis Beijing family with Han Chinese. Proc Natl Acad Sci U S A 112(26):8136–8141. doi:10.1073/pnas.1424063112

Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N, Nordenfelt S, Tandon A, Skoglund P, Lazaridis I, Sankararaman S, Fu Q, Rohland N, Renaud G, Erlich Y, Willems T, Gallo C, Spence JP, Song YS, Poletti G, Balloux F, van Driem G, de Knijff P, Romero IG, Jha AR, Behar DM, Bravi CM, Capelli C, Hervig T, Moreno-Estrada A, Posukh OL, Balanovska E, Balanovsky O, Karachanak-Yankova S, Sahakyan H, Toncheva D, Yepiskoposyan L, Tyler-Smith C, Xue Y, Abdullah MS, Ruiz-Linares A, Beall CM, Di Rienzo A, Jeong C, Starikovskaya EB, Metspalu E, Parik J, Villems R, Henn BM, Hodoglugil U, Mahley R, Sajantila A, Stamatoyannopoulos G, Wee JT, Khusainova R, Khusnutdinova E, Litvinov S, Ayodo G, Comas D, Hammer MF, Kivisild T, Klitz W, Winkler CA, Labuda D, Bamshad M, Jorde LB, Tishkoff SA, Watkins WS, Metspalu M, Dryomov S, Sukernik R, Singh L, Thangaraj K, Paabo S, Kelso J, Patterson N, Reich D (2016) The Simons genome diversity project: 300 genomes from 142 diverse populations. Nature 538(7624):201–206. doi:10.1038/nature18964

Malm S, Linguissi LS, Tekwu EM, Vouvoungui JC, Kohl TA, Beckert P, Sidibe A, Rusch-Gerdes S, Madzou-Laboum IK, Kwedi S, Penlap Beng V, Frank M, Ntoumi F, Niemann S (2017) New Mycobacterium tuberculosis complex sublineage, Brazzaville, Congo. Emerg Infect Dis 23(3):423–429. doi:10.3201/eid2303.160679

Marvig RL, Sommer LM, Molin S, Johansen HK (2015) Convergent evolution and adaptation of Pseudomonas aeruginosa within patients with cystic fibrosis. Nat Genet 47(1):57–64. doi:10.1038/ng.3148

May RM, Anderson RM (1983) Epidemiology and genetics in the coevolution of parasites and hosts. P Roy Soc Lond a Mat 390(1798):219–219

Merker M, Blin C, Mona S, Duforet-Frebourg N, Lecher S, Willery E, Blum MG, Rusch-Gerdes S, Mokrousov I, Aleksic E, Allix-Beguec C, Antierens A, Augustynowicz-Kopec E, Ballif M, Barletta F, Beck HP, Barry CE 3rd, Bonnet M, Borroni E, Campos-Herrero I, Cirillo D, Cox H, Crowe S, Crudu V, Diel R, Drobniewski F, Fauville-Dufaux M, Gagneux S, Ghebremichael S, Hanekom M, Hoffner S, Jiao WW, Kalon S, Kohl TA, Kontsevaya I, Lillebaek T, Maeda S, Nikolayevskyy V, Rasmussen M, Rastogi N, Samper S, Sanchez-Padilla E, Savic B, Shamputa IC, Shen A, Sng LH, Stakenas P, Toit K, Varaine F, Vukovic D, Wahl C, Warren R, Supply P, Niemann S, Wirth T (2015) Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage. Nat Genet 47(3):242–249. doi:10.1038/ng.3195

Mokrousov I, Vyazovaya A, Iwamoto T, Skiba Y, Pole I, Zhdanova S, Arikawa K, Sinkov V, Umpeleva T, Valcheva V, Alvarez Figueroa M, Ranka R, Jansone I, Ogarkov O, Zhuravlev V, Narvskaya O (2016) Latin-American-Mediterranean lineage of Mycobacterium tuberculosis: human traces across pathogen's phylogeography. Mol Phylogenet Evol 99:133–143. doi:10.1016/j.ympev.2016.03.020

Moller M, Hoal EG (2010) Current findings, challenges and novel approaches in human genetic susceptibility to tuberculosis. Tuberculosis (Edinb) 90(2):71–83. doi:10.1016/j.tube.2010.02.002

Murray GG, Wang F, Harrison EM, Paterson GK, Mather AE, Harris SR, Holmes MA, Rambaut A, Welch JJ (2016) The effect of genetic structure on molecular dating and tests for temporal signal. Methods Ecol Evol 7(1):80–89. doi:10.1111/2041-210X.12466

O'Garra A, Redford PS, McNab FW, Bloom CI, Wilkinson RJ, Berry MP (2013) The immune response in tuberculosis. Annu Rev Immunol 31:475–527. doi:10.1146/annurev-immunol-032712-095939

O'Neill MB, Mortimer TD, Pepperell CS (2015) Diversity of Mycobacterium tuberculosis across evolutionary scales. PLoS Pathog 11(11):e1005257. doi:10.1371/journal.ppat.1005257

Orlando L, Gilbert MT, Willerslev E (2015) Reconstructing ancient genomes and epigenomes. Nat Rev Genet 16(7):395–408. doi:10.1038/nrg3935

Orme IM, Robinson RT, Cooper AM (2015) The balance between protective and pathogenic immune responses in the TB-infected lung. Nat Immunol 16(1):57–63. doi:10.1038/ni.3048

Osorio NS, Rodrigues F, Gagneux S, Pedrosa J, Pinto-Carbo M, Castro AG, Young D, Comas I, Saraiva M (2013) Evidence for diversifying selection in a set of Mycobacterium tuberculosis genes in response to antibiotic- and nonantibiotic-related pressure. Mol Biol Evol 30(6):1326–1336. doi:10.1093/molbev/mst038

Paulson T (2013) Epidemiology: a mortal foe. Nature 502(7470):S2–S3. doi:10.1038/502S2a

Pepperell CS, Casto AM, Kitchen A, Granka JM, Cornejo OE, Holmes EC, Birren B, Galagan J, Feldman MW (2013) The role of selection in shaping diversity of natural M. tuberculosis populations. PLoS Pathog 9(8):e1003543. doi:10.1371/journal.ppat.1003543

Perez-Lago L, Comas I, Navarro Y, Gonzalez-Candelas F, Herranz M, Bouza E, Garcia-de-Viedma D (2014) Whole genome sequencing analysis of intrapatient microevolution in Mycobacterium tuberculosis: potential impact on the inference of tuberculosis transmission. J Infect Dis 209(1):98–108. doi:10.1093/infdis/jit439

Qu HQ, Li Q, McCormick JB, Fisher-Hoch SP (2011) What did we learn from the genome-wide association study for tuberculosis susceptibility? J Med Genet 48(4):217–218. doi:10.1136/jmg.2010.087361

Reed MB, Pichler VK, McIntosh F, Mattia A, Fallow A, Masala S, Domenech P, Zwerling A, Thibert L, Menzies D, Schwartzman K, Behr MA (2009) Major Mycobacterium tuberculosis lineages associate with patient country of origin. J Clin Microbiol 47(4):1119–1128. doi:10.1128/JCM.02142-08

Rice WR (2002) Experimental tests of the adaptive significance of sexual recombination. Nat Rev Genet 3(4):241–251. doi:10.1038/nrg760

Roetzer A, Diel R, Kohl TA, Ruckert C, Nubel U, Blom J, Wirth T, Jaenicke S, Schuback S, Rusch-Gerdes S, Supply P, Kalinowski J, Niemann S (2013) Whole genome sequencing versus traditional genotyping for investigation of a Mycobacterium tuberculosis outbreak: a longitudinal molecular epidemiological study. PLoS Med 10(2):e1001387. doi:10.1371/journal.pmed.1001387

Rose G, Cortes T, Comas I, Coscolla M, Gagneux S, Young DB (2013) Mapping of genotype-phenotype diversity among clinical isolates of mycobacterium tuberculosis by sequence-based transcriptional profiling. Genome Biol Evol 5(10):1849–1862. doi:10.1093/gbe/evt138

Rothschild BM, Martin LD, Lev G, Bercovier H, Bar-Gal GK, Greenblatt C, Donoghue H, Spigelman M, Brittain D (2001) Mycobacterium tuberculosis complex DNA from an extinct bison dated 17,000 years before the present. Clin Infect Dis 33(3):305–311. doi:10.1086/321886

Russell DG, Barry CE 3rd, Flynn JL (2010) Tuberculosis: what we don't know can, and does, hurt us. Science 328(5980):852–856. doi:10.1126/science.1184784

Salie M, van der Merwe L, Moller M, Daya M, van der Spuy G, van Helden PD, Martin MP, Gao XJ, Warren RM, Carrington M, Hoal EG (2014) Associations between human leukocyte antigen class I variants and the Mycobacterium tuberculosis subtypes causing disease. Int J Infect Dis 21:300–300. doi:10.1016/j.ijid.2014.03.1043

Sharma A, Bloss E, Heilig CM, Click ES (2016) Tuberculosis caused by Mycobacterium africanum, United States, 2004–2013. Emerg Infect Dis 22(3):396–403. doi:10.3201/eid2203.151505

Sherman DR, Mdluli K, Hickey MJ, Arain TM, Morris SL, Barry CE 3rd, Stover CK (1996) Compensatory ahpC gene expression in isoniazid-resistant Mycobacterium tuberculosis. Science 272(5268):1641–1643

Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. Genet Res 23(1):23–35

Smith NH, Hewinson RG, Kremer K, Brosch R, Gordon SV (2009) Myths and misconceptions: the origin and evolution of Mycobacterium tuberculosis. Nat Rev Microbiol 7(7):537–544. doi:10.1038/nrmicro2165

Stead WW (1998) Tuberculosis in Africa. Int J Tuberc Lung Dis 2(10):791–792

Stead WW (2001) Variation in vulnerability to tuberculosis in America today: random, or legacies of different ancestral epidemics? Int J Tuberc Lung Dis 5(9):807–814

Stead WW, Senner JW, Reddick WT, Lofgren JP (1990) Racial differences in susceptibility to infection by Mycobacterium tuberculosis. N Engl J Med 322(7):422–427. doi:10.1056/NEJM199002153220702

Stucki D, Brites D, Jeljeli L, Coscolla M, Liu Q, Trauner A, Fenner L, Rutaihwa L, Borrell S, Luo T, Gao Q, Kato-Maeda M, Ballif M, Egger M, Macedo R, Mardassi H, Moreno M, Vilanova GT, Fyfe J, Globan M, Thomas J, Jamieson F, Guthrie JL, Asante-Poku A, Yeboah-Manu D, Wampande E, Ssengooba W, Joloba M, Boom WH, Basu I, Bower J, Saraiva M, Vasconcellos SE, Suffys P, Koch A, Wilkinson R, Gail-Bekker L, Malla B, Ley SD, Beck HP, de Jong BC, Toit K, Sanchez-Padilla E, Bonnet M, Gil-Brusola A, Frank M, Penlap Beng VN, Eisenach K, Alani I, Ndung'u PW, Revathi G, Gehre F, Akter S, Ntoumi F, Stewart-Isherwood L, Ntinginya NE, Rachow A, Hoelscher M, Cirillo DM, Skenders G, Hoffner S, Bakonyte D,

Stakenas P, Diel R, Crudu V, Moldovan O, Al-Hajoj S, Otero L, Barletta F, Carter EJ, Diero L, Supply P, Comas I, Niemann S, Gagneux S (2016) Mycobacterium tuberculosis lineage 4 comprises globally distributed and geographically restricted sublineages. Nat Genet. doi:10.1038/ng.3704

Supply P, Marceau M, Mangenot S, Roche D, Rouanet C, Khanna V, Majlessi L, Criscuolo A, Tap J, Pawlik A, Fiette L, Orgeur M, Fabre M, Parmentier C, Frigui W, Simeone R, Boritsch EC, Debrie AS, Willery E, Walker D, Quail MA, Ma L, Bouchier C, Salvignol G, Sayes F, Cascioferro A, Seemann T, Barbe V, Locht C, Gutierrez MC, Leclerc C, Bentley SD, Stinear TP, Brisse S, Medigue C, Parkhill J, Cruveiller S, Brosch R (2013) Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of Mycobacterium tuberculosis. Nat Genet 45(2):172–179. doi:10.1038/ng.2517

Thye T, Vannberg FO, Wong SH, Owusu-Dabo E, Osei I, Gyapong J, Sirugo G, Sisay-Joof F, Enimil A, Chinbuah MA, Floyd S, Warndorff DK, Sichali L, Malema S, Crampin AC, Ngwira B, Teo YY, Small K, Rockett K, Kwiatkowski D, Fine PE, Hill PC, Newport M, Lienhardt C, Adegbola RA, Corrah T, Ziegler A, African TBGC, Wellcome Trust Case Control C, Morris AP, Meyer CG, Horstmann RD, Hill AV (2010) Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. Nat Genet 42(9):739–741. doi:10.1038/ng.639

Thye T, Niemann S, Walter K, Homolka S, Intemann CD, Chinbuah MA, Enimil A, Gyapong J, Osei I, Owusu-Dabo E, Rusch-Gerdes S, Horstmann RD, Ehlers S, Meyer CG (2011) Variant G57E of mannose binding lectin associated with protection against tuberculosis caused by Mycobacterium africanum but not by M. tuberculosis. PLoS One 6(6):e20908. doi:10.1371/journal.pone.0020908

Tiemersma EW, van der Werf MJ, Borgdorff MW, Williams BG, Nagelkerke NJ (2011) Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review. PLoS One 6(4):e17601. doi:10.1371/journal.pone.0017601

Trauner A, Borrell S, Reither K, Gagneux S (2014) Evolution of drug resistance in tuberculosis: recent progress and implications for diagnosis and therapy. Drugs 74(10):1063–1072. doi:10.1007/s40265-014-0248-y

Trauner A, Liu Q, Via LE, Liu X, Ruan X, Liang L, Shi H, Chen Y, Wang Z, Liang R, Zhang W, Wei W, Gao J, Sun G, Brites D, England K, Zhang G, Gagneux S, Barry CE 3rd, Gao Q (2017) The within-host population dynamics of Mycobacterium tuberculosis vary with treatment efficacy. Genome Biol 18(1):71. doi:10.1186/s13059-017-1196-0

Underhill PA, Kivisild T (2007) Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. Annu Rev Genet 41:539–564. doi:10.1146/annurev.genet.41.110306.130407

Veyrier FJ, Dufort A, Behr MA (2011) The rise and fall of the Mycobacterium tuberculosis genome. Trends Microbiol 19(4):156–161. doi:10.1016/j.tim.2010.12.008

Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, Eyre DW, Wilson DJ, Hawkey PM, Crook DW, Parkhill J, Harris D, Walker AS, Bowden R, Monk P, Smith EG, Peto TE (2013) Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. Lancet Infect Dis 13(2):137–146. doi:10.1016/S1473-3099(12)70277-3

Wampande EM, Mupere E, Debanne SM, Asiimwe BB, Nsereko M, Mayanja H, Eisenach K, Kaplan G, Boom HW, Gagneux S, Joloba ML (2013) Long-term dominance of Mycobacterium tuberculosis Uganda family in peri-urban Kampala-Uganda is not associated with cavitary disease. BMC Infect Dis 13:484. doi:10.1186/1471-2334-13-484

Wang J, McIntosh F, Radomski N, Dewar K, Simeone R, Enninga J, Brosch R, Rocha EP, Veyrier FJ, Behr MA (2015) Insights on the emergence of Mycobacterium tuberculosis from the analysis of Mycobacterium kansasii. Genome Biol Evol 7(3):856–870. doi:10.1093/gbe/evv035

WHO (2016) Global tuberculosis report 2016. World Health Organization, Geneva

Wilbur AK, Bouwman AS, Stone AC, Roberts CA, Pfister LA, Buikstra JE, Brown TA (2009) Deficiencies and challenges in the study of ancient tuberculosis DNA. J Archaeol Sci 36(9):1990–1997. doi:10.1016/j.jas.2009.05.020

Wilkipedia Ewe People (n.d.) https://en.wikipedia.org/wiki/Ewe_people

Williams AC, Dunbar RI (2014) Big brains, meat, tuberculosis and the nicotinamide switches: co-evolutionary relationships with modern repercussions on longevity and disease? Med Hypotheses 83(1):79–87. doi:10.1016/j.mehy.2014.04.003

Wirth T, Hildebrand F, Allix-Beguec C, Wolbeling F, Kubica T, Kremer K, van Soolingen D, Rusch-Gerdes S, Locht C, Brisse S, Meyer A, Supply P, Niemann S (2008) Origin, spread and demography of the Mycobacterium tuberculosis complex. PLoS Pathog 4(9):e1000160. doi:10.1371/journal.ppat.1000160

Woolhouse M, Gaunt E (2007) Ecological origins of novel human pathogens. Crit Rev Microbiol 33(4):231–242. doi:10.1080/10408410701647560

Woolhouse ME, Webster JP, Domingo E, Charlesworth B, Levin BR (2002) Biological and biomedical implications of the co-evolution of pathogens and their hosts. Nat Genet 32(4):569–577. doi:10.1038/ng1202-569

Worby CJ, Lipsitch M, Hanage WP (2014) Within-host bacterial diversity hinders accurate reconstruction of transmission networks from genomic distance data. PLoS Comput Biol 10(3):e1003549. doi:10.1371/journal.pcbi.1003549

Young DB, Perkins MD, Duncan K, Barry CE 3rd (2008) Confronting the scientific obstacles to global control of tuberculosis. J Clin Invest 118(4):1255–1265. doi:10.1172/JCI34614

Yruela I, Contreras-Moreira B, Magalhaes C, Osorio NS, Gonzalo-Asensio J (2016) Mycobacterium tuberculosis complex exhibits lineage-specific variations affecting protein ductility and epitope recognition. Genome Biol Evol 8(12):3751–3764. doi:10.1093/gbe/evw279

Zheng NB, Whalen CC, Handel A (2014) Modeling the potential impact of host population survival on the evolution of M. tuberculosis latency. PLoS One 9 (8). doi:ARTN e105721. 10.1371/journal.pone.0105721

Zumla A, Maeurer M, Host-Directed Therapies N, Chakaya J, Hoelscher M, Ntoumi F, Rustomjee R, Vilaplana C, Yeboah-Manu D, Rasolof V, Munderi P, Singh N, Aklillu E, Padayatchi N, Macete E, Kapata N, Mulenga M, Kibiki G, Mfinanga S, Nyirenda T, Maboko L, Garcia-Basteiro A, Rakotosamimanana N, Bates M, Mwaba P, Reither K, Gagneux S, Edwards S, Mfinanga E, Abdulla S, Cardona PJ, Russell JB, Gant V, Noursadeghi M, Elkington P, Bonnet M, Menendez C, Dieye TN, Diarra B, Maiga A, Aseffa A, Parida S, Wejse C, Petersen E, Kaleebu P, Oliver M, Craig G, Corrah T, Tientcheu L, Antonio M, Rao M, McHugh TD, Sheikh A, Ippolito G, Ramjee G, Kaufmann SH, Churchyard G, Steyn A, Grobusch M, Sanne I, Martinson N, Madansein R, Wilkinson RJ, Mayosi B, Schito M, Wallis RS (2015) Towards host-directed therapies for tuberculosis. Nat Rev Drug Discov 14(8):511–512. doi:10.1038/nrd4696

The Biology and Epidemiology of *Mycobacterium canettii*

2

Philip Supply and Roland Brosch

Abstract

Genome-based insights into the evolution of *Mycobacterium tuberculosis* and other tuberculosis-causing mycobacteria are constantly increasing. In particular, the recent genomic and functional characterization of several *Myocbacterium canettii* strains, which are thought to resemble in many aspects the putative common ancestor of the members of the *M. tuberculosis* complex (MTBC), has consolidated a plausible scenario of the early evolution of tuberculosis-causing mycobacteria, in which the clonal MTBC, comprising numerous key pathogens of mammalian hosts, has evolved from a generalist mycobacterium living in the environment. These studies also have considerably enriched our knowledge on selected molecular events that likely have contributed to the incursion, maintenance and spread of the MTBC members in diverse mammalian hosts. Here, we summarize and discuss recently revealed molecular and evolutionary aspects and emphasize the vast utility of *M. canettii* strains for identifying the mechanisms that contributed to the global emergence of *M. tuberculosis* as one of the most important human pathogens.

Keywords

Mycobacterium tuberculosis • Progenitor • Evolution • Phylogeny • Smooth colony morphology • Virulence

P. Supply
Université de Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 8204 - CIIL - Centre d'Infection et d'Immunité de Lille, F-59000 Lille, France
e-mail: philip.supply@ibl.cnrs.fr

R. Brosch (✉)
Institut Pasteur, Unit for Integrated Mycobacterial Pathogenomics, 75724, Paris Cedex 15, France
e-mail: roland.brosch@pasteur.fr

S. Gagneux (ed.), *Strain Variation in the* Mycobacterium tuberculosis *Complex: Its Role in Biology, Epidemiology and Control*, Advances in Experimental Medicine and Biology 1019,
DOI 10.1007/978-3-319-64371-7_2

2.1 Introduction

Among the mycobacterial species, the tuberculosis-causing mycobacteria, also known as tubercle bacilli, constitute a particular group of mycobacterial strains than can cause tuberculosis in different mammalian species (Magee and Ward 2012; Brosch et al. 2002; Boritsch et al. 2014).

According to recent genome-based phylogenetic analyses, the tubercle bacilli can be divided into a clonal cluster of strains, the so-called *Mycobacterium tuberculosis* complex (MTBC) and a group of highly recombinogenic and diversified tubercle bacilli, named *Mycobacterium canettii* cluster, the latter showing numerous traces of inter-strain recombination events and genome signatures that suggest close genomic relationship with the most recent common ancestor (MCRA) of the tubercle bacilli (Supply et al. 2013; Boritsch and Brosch 2016). While strains belonging to the MTBC members are responsible for almost all of the globally distributed tuberculosis cases in humans and in animals, the *M. canettii* strains are very rare and have mainly been isolated from human tuberculosis cases in countries around the Horn of Africa. Due to their close phylogenetic relationship to the MRCA of the tubercle bacilli, the extant population of *M. canettii* strains represents a particularly interesting group of strains for elucidating the key evolutionary events that have led to the emergence of the MTBC, and *M. tuberculosis* in particular.

However, before describing in more detail the characteristics of the *M. canettii* strains and their differences with the MTBC members, a brief description shall be given on the MTBC, which contains a range of different variants of tubercle bacilli that share more than 99.99% similarity at the genome sequence level but display characteristic differences in host preference. *M. tuberculosis* represents the variant transmitted efficiently among humans, still causing about ten million active human tuberculosis cases per year. Five main phylogenetic lineages of *M. tuberculosis* strains have been defined, named L1, L2, L3, L4 and L7 (Comas et al. 2013, 2015; Blouin et al. 2012; Nebenzahl-Guimaraes et al. 2016), lineages L1 and L7 comprising strains defined as "ancestral" because of the preservation of the *M. tuberculosis*-specific deletion region TbD1 (Brosch et al. 2002) in their genomes, and lineages L2, L3 and L4 representing "modern" TbD1-deleted strains (Gagneux and Small 2007; Smith et al. 2009b). Besides *M. tuberculosis*, other MTBC members can also cause disease in humans and animals (see also Chap. 7 of this book). This is the case for members of a particular sublineage of the MTBC that is characterized by the absence of the genomic region of difference RD9 (Brosch et al. 2002), comprising the two lineages (L5 and L6) of *Mycobacterium africanum*, which are prevalent in different countries of West-Africa (de Jong et al. 2010; Koro Koro et al. 2013) and numerous variants that preferentially infect different animal species (Fig. 2.1). A limited number of human tuberculosis cases are reported to be caused by *Mycobacterium orygis* (van Ingen et al. 2012) and by *Mycobacterium bovis*, which are mainly transmitted to humans by contaminated milk and/or close contact to animals (Muller et al. 2013). A close relative to *M. bovis* is *Mycobacterium caprae,* causing tuberculosis in goats and deer (Aranaz et al. 1999; Domogalla et al. 2013). *Mycobacterium microti,* another MTBC member, was originally found as a disease agent in voles (Wells 1937), but may also be found in cats (Smith et al. 2009a) and even as the cause of rare tuberculosis cases in humans (van Soolingen et al. 1998; Horstkotte et al. 2001). The closest relative of *M. microti* is *Mycobacterium pinnipedii* (Cousins et al. 2003; Smith et al. 2009b), which is an MTBC member that preferentially infects seals and sea-lions. Zoonotic contact between seals and man can lead to human *M. pinnipedii* infections (Kiers et al. 2008), a situation that is thought to have played a role in human tuberculosis cases in pre-Columbian South America. In a recent study, using a sophisticated whole-genome array capture approach to obtain sufficient DNA for whole genome sequencing, the presence of *M. pinnipedii*-like tubercle bacilli was demonstrated in 1000 year-old Peruvian skeletons (Bos et al. 2014), suggesting a zoonotic transmission of tubercle bacilli from seals to indige-

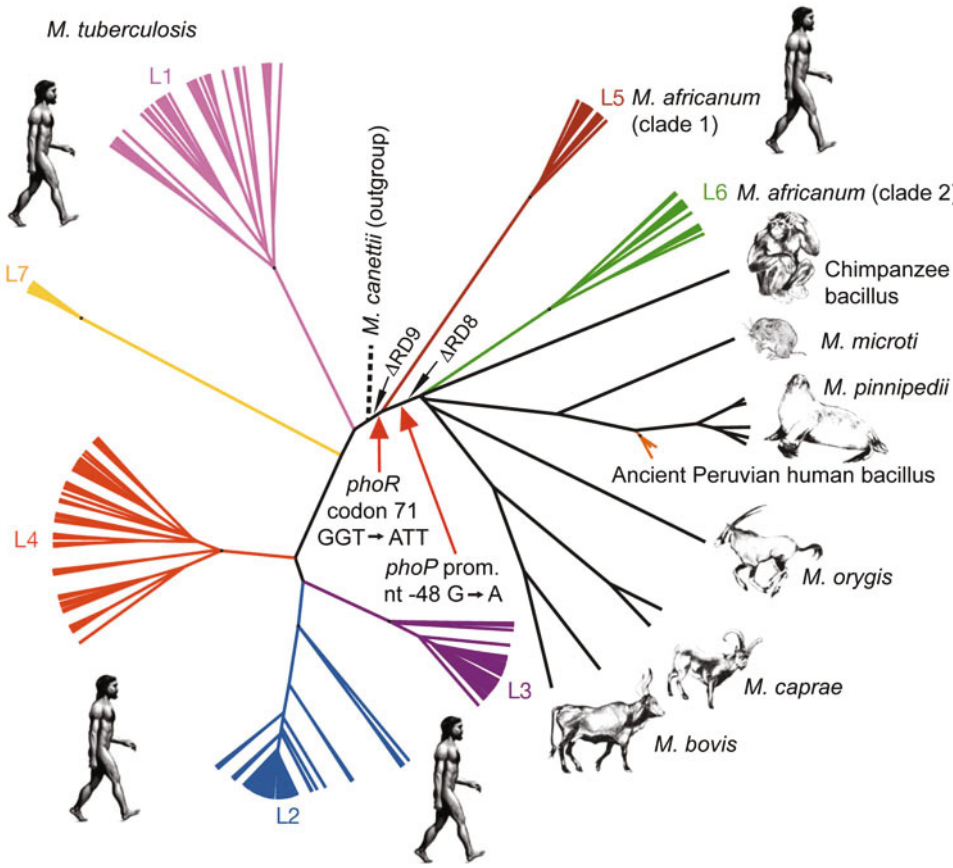

Fig. 2.1 Phylogenic analysis of the different MTBC lineages, based on SNPs of 261 mycobacterial genomes (Adapted from references Bos et al. 2014 with image from Brosch and Guilhot 2015), including additional information on deletion points of selected regions of difference (RDs) and mutations in the PhoPR mycobacterial two component system involved in virulence (Perez et al. 2001; Gonzalo-Asensio et al. 2014)

nous human populations at the time. Moreover, the use of molecular typing and genome sequencing recently has helped to describe a wide range of different MTBC members in dassies (Mostowy et al. 2004), in meercats (Parsons et al. 2013), mongooses (Alexander et al. 2010, 2016) as well as in a chimpanzee (Coscolla et al. 2013). The latter strains are all closely related to the *M. africanum* lineage L6 strains (Boritsch et al. 2014; Bos et al. 2014), also characterized by the absence of region of difference RD8, which was shown to impact type VII system-mediated secretion of EsxA and EsxB, representing important virulence factors of *M. tuberculosis* (Gonzalo-Asensio et al. 2014; Groschel et al. 2016). Genome analyses of the different members of the MTBC have confirmed the compact phylogenetic population structure with a maximum of about 2200–2400 single nucleotide polymorphisms (SNPs) between *M. tuberculosis* and the most distantly related members of the MTBC, *M. africanum* (Comas et al. 2013) and *M. bovis* (Garnier et al. 2003), respectively (Boritsch et al. 2014).

In contrast to the genomically monomorphic MTBC members, the *M. canettii* strains show significantly more genomic variability and a – for tubercle bacilli – unusual colony morphology, which has led in the past to them being referred to as "smooth tubercle bacilli" (STM). The *M. canettii* strains belong to a taxon that has been named by van Soolingen and co-workers

"Canetti" in honour of Professor Georges Canetti, who was a pioneer in research on tuberculosis drug treatment (Canetti 1965), and who first isolated such a strain from a patient with pulmonary tuberculosis in the 1960s with his team at the Institut Pasteur in Paris (van Soolingen et al. 1997). Since this first isolation of *M. canettii,* fewer than 100 strains of tubercle bacilli with smooth colony morphology have been identified worldwide, whereby the great majority of these *M. canettii* strains have been isolated from patients that lived in, or had contact to, the geographical region of East-Africa / the Horn of Africa (van Soolingen et al. 1997; Pfyffer et al. 1998; Koeck et al. 2002, 2011; Fabre et al. 2004, 2010; Somoskovi et al. 2009; Blouin et al. 2014).

While in most of these previous reports, *M. canettii* strains were described as members of the MTBC, we suggest that *M. canettii* strains should be considered as a closely related, but separate clade of tubercle bacilli, based on their specific phenotypic and genomic characteristics, which are clearly different from those of the classical members of the MTBC, as explained in detail further below. Indeed, the insights obtained from genome sequencing showed that *M. canettii* strains represent a genomically-diversified and non-clonal clade of strains of putative environmental origin that likely resembles the pool of strains from which the common ancestor of the MTBC has evolved by clonal expansion into the MTBC. This feature makes *M. canettii* strains exceptional and interesting study objects that have attracted substantial scientific interest in recent years. Due to their presumed ancestral origin, *M. canettii* strains were previously suggested to be named *Mycobacterium prototuberculosis* (Gutierrez et al. 2005). Although their evolutionary early branching status has subsequently been confirmed by whole genome sequence analyses (Supply et al. 2013; Blouin et al. 2014), this suggestion was not retained due to nomenclatorial incompatibilities (Smith 2006; Brisse et al. 2006). Thus, the *M. canettii* taxon name is used now for all extant smooth tubercle bacilli, but the name *M. prototuberculosis* might still be appropriate for describing the putative common ancestor that existed at the branching point of the MTBC and *M. canettii* strains.

Independent of their name, *M. canettii* strains represent interesting mycobacteria, for which we will describe the different aspects in this chapter. We will focus on the history of research on *M. canettii* strains, the genomic and phenotypic characteristics of these strains, as well as the epidemiological characteristics linked to clinical cases of *M. canettii* infection. Finally, we will provide an overview of the various insights that have been gained recently from in depth studies that have used *M. canettii* as a model in comparative experimental studies together with *M. tuberculosis*. The results from these studies are of major interest for questions related to the evolution and pathogenicity of tuberculosis.

2.2 The Phylogeny and Epidemiology of *M. canettii* Strains

In the pre-genomic era, for *M. tuberculosis* and other members of the MTBC, IS*6110*-restriction fragment length polymorphisms and spoligotyping were used as first reliable molecular typing methods that could efficiently separate different groups within the MTBC members (Kremer et al. 1999) (see Chap. 3). However, for *M. canettii* strains very different patterns were obtained with these methods (van Soolingen et al. 1997; Pfyffer et al. 1998; Gutierrez et al. 2005), which contributed to the description of *M. canettii* strains as a separate taxon of tubercle bacilli (van Soolingen et al. 1997). Further insights into the genetic diversity of *M. canettii* strains were then obtained by additional use of MIRU-VNTR typing as well as multilocus sequence typing (MLST) (Gutierrez et al. 2005; Fabre et al. 2004), which indicated a higher degree of genetic variation relative to the MTBC, as well as traces of recombination – not observed in the MTBC – in some house-keeping genes. However, the full extent of the diversity within the *M. canettii* taxon was only understood when selected *M. canettii* strains were subjected to whole genome sequence analyses

(Supply et al. 2013; Blouin et al. 2014), which showed that *M. canettii* strains were clear outliers relative to the MTBC strains. Despite the strong geographic confinement of the *M. canettii* strains selected for genome analysis, isolated mainly from patients at the Bouffard French Military Hospital and the Paul Faure Anti-tuberculosis Centre in Djibouti (Koeck et al. 2002, 2011; Fabre et al. 2004; Gutierrez et al. 2005; Blouin et al. 2014), the numbers of SNPs found among their genomes and between them and MTBC genomes ranged from around 9000 to 65,000, thus by far exceeding the diversity of the entire MTBC strain population found worldwide.

In contrast to the highly clonal MTBC genomes, *M. canettii* genomes also showed numerous traces of recent inter-strain horizontal gene transfer (HGT), beyond the few housekeeping genes analysed initially (see above), suggesting that *M. canettii* strains might have retained capacities for HGT and might be in frequent contact with a yet unknown environmental reservoir (Blouin et al. 2014; Supply et al. 2013; Boritsch et al. 2014). This hypothesis was recently strengthened by the experimental demonstration of inter-strain transfer of large chromosomal DNA fragments between two *M. canettii* strains (Boritsch et al. 2016b), described in more detail below. Together with their larger genomes, this diversified and non-clonal population structure positions *M. canettii* strains in evolutionarily lineages of tubercle bacilli that branched before the clonal emergence of the MTBC (Supply et al. 2013; Blouin et al. 2014; Derbyshire and Gray 2014; Mortimer and Pepperell 2014). The isolation of these early-branching tubercle bacilli from patients in East Africa and the regular association of the deepest branches within the MTBC, i.e. *M. africanum* lineages L5 and L6 and *M. tuberculosis* L1, L7, with patients in Africa point conjointly to an African origin of the MTBC and the *M. tuberculosis* ancestor (Fabre et al. 2004; Gagneux et al. 2006; Supply et al. 2013; Gutierrez et al. 2005). From these observations, it seems highly plausible that the common ancestor of the clonal MTBC lineage comprising the human lineages L1–L7 and the animal lineages evolved from an *M. canettii*-like progenitor that likely has shared many of the genomic and phenotypic characteristics that may still be present in the extant *M. canettii* strains, such as the ones further described below.

As a contemporary reflection of how a particular ancestral *M. canettii*-like strain or cluster of strains may have evolved into a pathogen that shows enhanced spread and disease-causing potential, the observations around a recently recognized outbreak of lymph-node tuberculosis in the area of Djibouti should be mentioned here. As described in their publication, Blouin and coworkers (2014) observed an elevated number of lymph node infections with a particular *M. canettii* strain type and generated genome sequence data of the 12 strains involved. This analysis showed that the *M. canettii* strains formed a compact group within the otherwise highly divergent *M. canettii* strains (Blouin et al. 2014) (Fig. 2.2), which is consistent with the identical MLST profile (D) of the strains from this cluster observed previously (Gutierrez et al. 2005; Supply et al. 2013). While the close genomic relationship among the strains suggest a common source and/or transmission of infection, epidemiological investigations did not find evidence for human-to-human transmission that could explain this outbreak of lymphnode-tuberculosis. Similarly, attempts to identify a common source, such as for example a water source used by all patients, were unsuccessful, and although other potential transmission modes were studied, no potential transmission vector could be identified (Blouin et al. 2014). Despite the lack of information on the exact route of transmission, this example shows that under yet unknown conditions, selected *M. canettii* strain clusters may spread and cause disease in humans more frequently than others. These particular strains represent thus interesting study objects to determine the molecular or phenotypic reasons for this increased virulence. Interestingly, *M. canettii* strains of the MLST profile (D) were precisely found as the most closely related to *M. tuberculosis* in the genomic comparisons (Gutierrez et al. 2005; Supply et al. 2013). Moreover, *M. canettii* strains of this genotype were also found as being the most virulent

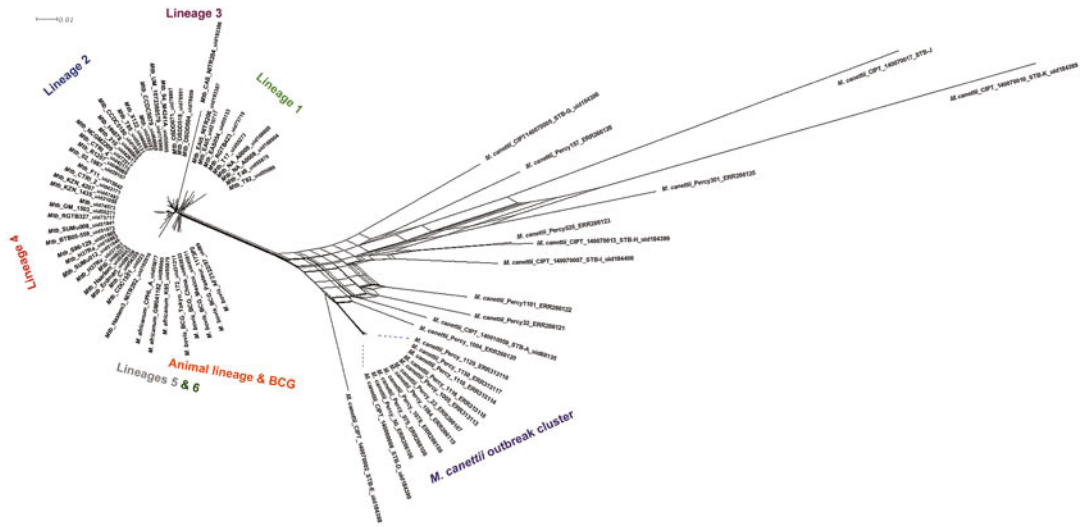

Fig. 2.2 Network phylogeny inferred among 24 *M. canettii* genomes and 53 selected genome sequences from members of the *M. tuberculosis* complex by NeighborNet analysis. Note that some of the exceptional long branches found within the otherwise genomically homogeneous MTBC members are caused by poor sequence data, as discussed in reference (Boritsch et al. 2014) (Figure reproduced from reference Boritsch et al. 2014)

M. canettii strains in the comparative virulence analysis in different mouse infection assays (Supply et al. 2013). This example of a particular *M. canettii* cluster showing higher virulence in experimental models and higher prevalence in human cases may also serve as an interesting model to better understand how in the past, the common ancestor of the MTBC members evolved into a lineage that has become a globally distributed complex with a clonal population structure and outstanding potential to infect and spread among humans and various animal species.

2.3 The Smooth Morphotype of *M. canettii*

Microbiologists who are familiar with the usual colony morphology of *M. tuberculosis* strains on agar plates might be very surprised if they saw the colonies of *M. canettii* strains, because in contrast to the *M. tuberculosis* colonies, which show a dry and rough morphotype, the colonies of *M. canettii* strains are characterized by a smooth, glossy phenotype (Fig. 2.3). This morphotype also justifies the alternative designation of the *M. canettii* strains as smooth tubercle bacilli (STB). While it was known that selected *M. canettii* strains produced lipooligosaccharides (LOS) that were not present in *M. tuberculosis* (Daffe et al. 1991; van Soolingen et al. 1997), the underlying mechanism for the smooth and rough colony variation in *M. canettii* remained ambiguous, since earlier studies could not confirm the exclusive presence of LOS in the smooth morphotypes (Lemassu et al. 1992), nor could a genetic mechanism for the phenomenon be clearly defined (van Soolingen et al. 1997).

However, recent genomic analyses and genetic complementation experiments formally demonstrated that the phenotypic differences between smooth and spontaneously converted rough variants of *M. canettii* were caused by recombination events in the polyketide synthase 5 (Pks5)-encoding locus (Boritsch et al. 2016a). In smooth variants, this locus contains two proximal *pks5* homologues separated by an associated *pap* gene, similarly to the LOS-producing non-tuberculous mycobacteria such as *M. marinum* and *M. kansasii*. Interestingly, the single *pks5* configuration resulting from recombination in the rough variant was similar to the genomic configuration that became fixed

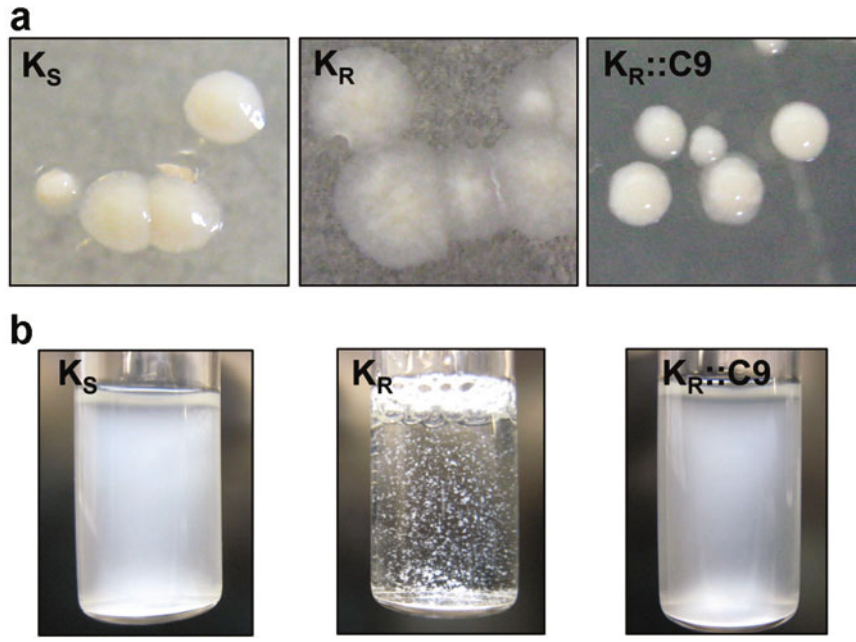

Fig. 2.3 Colony morphology of smooth (S) (**a**, *left panel*) and rough (R) (**a**, *middle panel*) variants of *M. canettii* strain K (CIPT 140070010). The *left* panel of **a** shows the recovered smooth morphotype after complementation of a rough variant with an integrating vector (*C9*) containing the pks5 locus of a smooth *M. canettii* strain. Scale bars, 2.5 mm (Image reproduced from reference Boritsch et al. 2016a). **b** shows the corresponding *M. canettii* variants *K(S)*, *K(R)* and *K(R::C9)* grown in liquid medium the different aggregation characteristics of the strains linked to their surface structure (Image reproduced from reference Boritsch et al. 2016a)

in *M. tuberculosis* and the other members of the MTBC, suggesting that during the course of evolution, LOS production was lost by the ancestor of the MTBC (Boritsch et al. 2016a). Importantly, the rough *M. canettii* variants showed an altered host–pathogen interaction and increased virulence in cellular- and animal-infection models. As described by Boritsch and colleagues, the LOS deficient rough variant showed increased replication rates in human macrophages and in lungs of infected guinea pigs. Moreover, the rough strain induced higher levels of inflammatory cytokines IL-6 and IL-12p40 in infected phagocytes, suggesting that the modification of the bacterial surface composition also altered the recognition by the host immune system. These phenotypes were all reverted upon complementation with the *pks5* locus-containing allele from a smooth *M. canettii* strain. These results thus suggest that *pks5*-recombination-mediated bacterial surface remodelling of the ancestor of the MTBC has driven evolution from putative generalist mycobacteria towards professional pathogens of mammalian hosts (Boritsch et al. 2016a).

2.4 Virulence Characteristics of *M. canettii* Strains

From an epidemiological point of view, the *M. canettii* strains are much less successful in causing human disease than do *M. tuberculosis* strains, although susceptible individuals may experience very severe infections with *M. canettii* strains that can even lead to death of patients (Koeck et al. 2002, 2011; Fabre et al. 2010). Indeed, when comparing the strong geographic restriction of *M. canettii* isolates to East Africa relative to the worldwide distribution of other tubercle bacilli, it seems plausible that immunocompetent individuals may control *M. canettii* infections better than infections by *M. tuberculosis*, as otherwise we would be

confronted with a much higher number of *M. canettii* infections. These presumptions also suggest that *M. canettii* might not efficiently spread by human-to-human transmission, which could mean that infection by *M. canettii* strains might not cause sufficient tissue destruction in the lungs for efficient transmission compared to *M. tuberculosis*.

In agreement with these predictions based on epidemiological findings, recent experiments in different mouse infection models confirmed the lower virulence and lower persistence of *M. canettii* strains relative to *M. tuberculosis* strains. Indeed, results from two different models, including intra-nasally infected BALB/c mice and aerosol-infected C57Bl/6 mice, showed that *M. canettii* strains were able to effectively multiply in lungs and disseminate to spleens during the acute infection phase in both models, but to a lower extent compared to *M. tuberculosis* H37Rv in C57Bl/6 mice (Supply et al. 2013). The differential phenotypes were most dramatic at the chronic infection phase in BALB/c mice, where *M. tuberculosis* was able to persist in the lungs for up to 30 weeks at infection levels close to those of the acute phase, while the amount of viable *M. canettii* decreased by at least 1 log at the different time points in lungs and spleens, with the most dramatic effect observed for *M. canettii* strain K, which is the most phylogenomically distant from MTBC strains and became undetectable in BALB/c mice infection levels after 30 weeks. In addition, histopathological analyses showed less intense lung lesions and inflammation after infection with the *M. canettii* strains compared to *M. tuberculosis* infection (Supply et al. 2013). These results are in agreement with some older results, where a *M. canettii* strain was found to be less virulent in a comparative virulence screen in mice that was focused on the comparison of a wide range of different *M. tuberculosis* strains (Dormans et al. 2004). Taken together, it seems that *M. canettii* strains, despite being able to cause infection and disease in humans and selected experimental animal models, seem less well adapted to cause persistent infection and disease necessary to spread in mammalian hosts, consistent with a putative environmental origin of these strains. A putative environmental origin of *M. canettii* strains is also supported by the numerous recombination traces in the genomes of *M. canettii* strains that most likely have occurred by different inter-strain mating events and by other HGT-related specificities of *M. canettii* genomes, as described in the next paragraphs.

2.5 The Recombination Cluster

Following first hints obtained by MLST, comparative genomics revealed the existence of numerous mosaic structures in the genomes of different *M. canettii* strains. From these observations, it became clear that within this clade of strains, numerous opportunities for inter-strain recombination events might have existed or still exist, which have led to the hypothesis that the genomes of *M. canettii* and the MRCA of the MTBC are composite sequence assemblies shaped by extensive inter-strain recombination events (Gutierrez et al. 2005; Supply et al. 2013). While this hypothesis was also supported by results from dedicated in silico analyses (Mortimer and Pepperell 2014), experimental evidence of chromosomal gene transfer was only available for the quite distantly-related, fast-growing mycobacterial model organism *Mycobacterium smegmatis* (Coros et al. 2008) In this species, laboratory-based mating experiments revealed that large DNA fragments could be transferred from a donor to a recipient strain in an unconventional conjugation process, named distributive conjugal transfer. The obtained transconjugants displayed highly mosaic genomes that were reminiscent of products of eukaryotic meiosis, and the genome mosaicism observed among *M. canettii* strains (Gray et al. 2013). Very recently, mating experiments involving different *M. canettii* strains generated unique evidence of donor-to-recipient transfer of large chromosomal DNA fragments (Boritsch et al. 2016b), a situation which resembles the distributive conjugal transfer mode in *M. smegmatis*. The genomes of the *M. canettii* recipient

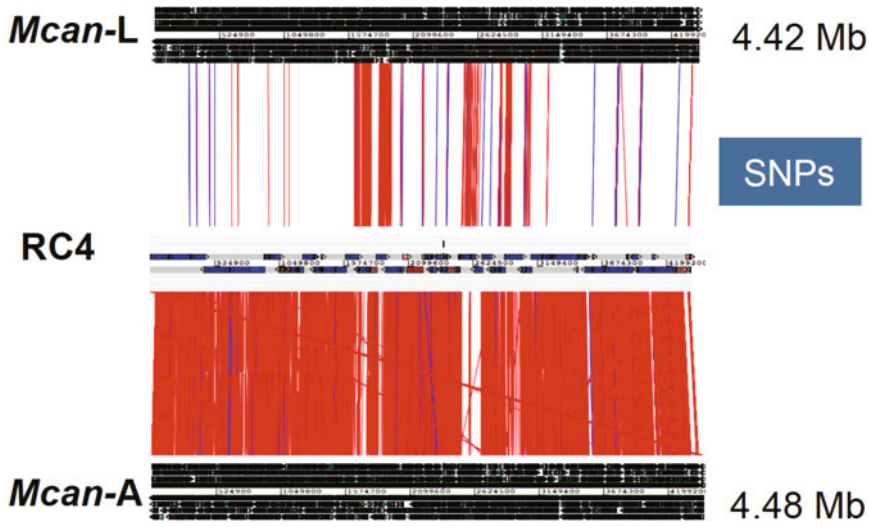

Fig. 2.4 Genome visualization by using the software Artemis Comparison Tool (ACT) (Carver et al. 2008) of SNPs identified between genomes of donor strain *M. canettii* A (*bottom* genome), recipient *M. canettii* L (*top* genome) and recombinant *M. canettii* (RC) (*middle* genome). Note that the whole genomes of the strains are shown (4.42 and 4.48 Mb). SNPs are represented by *red lines* and Indels by *blue lines*. In the recombinant strain RC, 13 donor-derived fragments in the size of 328 kb were identified (Boritsch et al. 2016b) (Image reproduced from reference Boritsch et al. 2016b). *Mcan M. canettii*, *RC* recombinant

contained donor DNA fragments ranging from 100 bp to over 100 kb in different sites of the recipient chromosome (Fig. 2.4). The distinction of donor DNA fragments from recipient DNA was possible due to the high number of SNPs (>22,000) that differentiate the genomes of the donor and recipient strains (STB/*M. canettii* A and STB/*M. canetii* L, respectively) (Supply et al. 2013). However, compared to the findings for *M. smegmatis* strains, the mating frequency observed between *M. canettii* strains was rather low (Boritsch et al. 2016b). It should also be mentioned that despite numerous attempts and mating combinations, no recombinants/transconjugants were obtained for different *M. tuberculosis* strain couples, suggesting that *M. tuberculosis* and the clonal MTBC members might have lost the ability of inter-strain HGT since their branching from an *M. canettii*-like MRCA (Boritsch et al. 2016b). Together with the observed clonal population structure of the MTBC, these data strongly argue against a recent postulate that frequent inter-strain transfer of small DNA fragments of about 50 bp between *M. tuberculosis* strains exists (Namouchi et al. 2012). This hypothesis was based on results from an *in silico* analysis searching homoplastic SNP regions in a given reading frame, which however also counted adjacent base pair changes caused by single tandem-base mutations as putative recombination regions and thus likely strongly over-estimated the presence of putative homoplastic SNPs (Boritsch et al. 2014). As the question of recombination among MTBC strains is very important for the risk evaluation of potential transfer of drug resistance and fitness mutations among clinically relevant mycobacterial strains, continued attention should be paid to this point. In any case, the finding that the closely related *M. canettii* strains can transfer large fragments of chromosomal DNA from donor to recipient strains may serve as proof of principle that this kind of transfer may have played an important role during the evolution of the MTBC from an *M. canettii*-like progenitor (described further

below). Genomic and phenotypic comparative experiments shall help identify the factors that contribute to the apparently different faculties of horizontal transfer among *M. canettii* and MTBC strains.

2.6 Other Genotypic Differences Between *M. canettii* Strains and the MTBC

Analysis of the genomes of five sequenced *M. canettii* strains representing lineages/strains A, D, J, K, and L revealed that these strains exhibited a pan-genome that was about 20% (900 genes) larger than that found for MTBC members. As described above, these extra genes are likely the result of individual mating encounters and chromosomal gene transfer, rather than acquisition by common ancestors, as most of these extra genes were restricted to individual strains. Indeed, the experimentally demonstrated large sizes of transferred DNA in single mating events between donor and recipient strains might explain the existence of strain-specific gene clusters, such as the presence of an extra *mce5* operon in strain *M. canettii* J (also known as STB-J) (Supply et al. 2013). Similarly, *M. canettii* J also contains a unique *eptABCD* operon orthologous to the *eptABCD* operon of *M. smegmatis* (Panas et al. 2014). Interestingly, the presence of this operon in *M. smegmatis* is linked to an inability to be transformed with *Mycobacterium fortuitum* pAL5000 plasmids, as EptC was shown to stabilize negative supercoils in plasmid DNA and to prevent segregation of plasmid DNA to daughter cells (Panas et al. 2014). The presence of *eptABCD* in strain *M. canettii* J thus seems to have the same consequences, as observed in a recent study where this strain could not be transformed with pMRF1-DsRed during the preparation of mycobacterial mating experiments, whereas all other *M. canettii* strains were transformable with the DsRed encoding plasmid (Boritsch et al. 2016b). Moreover, genome analysis of individual *M. canettii* strains also revealed the presence of a 55 kb-sized prophage in the genome of *M. canettii* strain I (STB-I) that might encode a potentially complete virion (Supply et al. 2013). Although currently no information is available on the functionality of this putative phage, the presence of such a large prophage in a strain of tuberculosis-causing mycobacteria is unique and different from the ubiquitous 10 kb-sized prophage-like elements *phiRv1* and *phiRv2* (Cole et al. 1998; Hendrix et al. 1999) that show variable presence among the members of the MTBC.

Finally, genome analyses also highlighted the presence of clustered regularly interspaced short palindromic repeat (CRISPR) loci and CRISPR-associated (cas) genes (CRISPR-Cas) in genomes of *M. canettii* strains, differing from those of the MTBC. In the MTBC, variation of so-called spacers within the CRISPR-regions is used for spoligotyping, a widely applied method for molecular differentiation of MTBC isolates (Kamerbeek et al. 1997). The potential function of mycobacterial CRISPR-Cas systems as elements for adaptive immunity against phages and plasmids via repeat-spacer–derived RNAs, as repeatedly demonstrated for other bacteria (see review e.g. by Makarova et al. 2011), awaits further confirmation. If one assumes similar functions, however, the differences observed among CRISPR-Cas systems of *M. canettii* and MTBC strains suggest exposure of *M. canettii* and MTBC strains to different environments. Strikingly, the genomes of STB/*M.canettii*-A and -D, – which are phylogenetically closest to the MTBC – contain a CRISPR-Cas locus encoding a system of a same major type (III-A) as that of MTBC genomes. However, the locus contains only a few *crispr* spacers in common with MTBC strains. The same locus in the other more distant *M. canettii* strains is occupied by completely different CRISPR-Cas systems of other types (I-C variant or I-E), most closely related to those of environmental actinobacteria. Furthermore, the strain STB/*M. canettii*-K has a second locus, encoding a CRISPR-Cas module of yet another type (I-C) with Cas proteins that were most similar to those of other environmental bacteria (such as *Thiorhodovibrio* species). Again, the identified *crispr* spacers were strain-specific and distinct from the MTBC. These structural

divergences in CRIPR-Cas modules and distinct composition in spacer sequence indicate that the corresponding systems might indeed show different targets (Supply et al. 2013), a finding that thus also points to different habitats for *M. canettii* and MTBC strains.

Interestingly, the genomes of *M. canettii* strains also differ by their intein-encoding sequences and their insertion points in their respective *recA* genes, relative to the MTBC (Boritsch et al. 2016b). Inteins are self-splicing protein elements that are mobile at the DNA level and thought to be horizontally transferred through invasion by mycobacteriophages (Kelley et al. 2016). These variations thus argue for independent invasion events, perhaps in distinct environments, as presumed for the CRIPR-Cas module differences (see above).

In addition to these extra genomic regions in selected individual *M. canettii* strains, a few other genes are nevertheless specifically conserved throughout all *M. canettii* strains as a result of shared phylogenetic ancestry, preceding that of the MTBC. This is the case for example for the gene *cobF*, encoding a precorrin-6a synthase required for vitamin B12 synthesis, which is present in *M. canettii* and absent from the MTBC, apparently lost during evolution from an *M. canettii*-like ancestor (Boritsch et al. 2014; Blouin et al. 2014). Although it remains unclear whether the deletion of *cobF* has entirely ablated B12 biosynthesis, because alternative methyltransferases might partially compensate for the loss of CobF (Gopinath et al. 2013), it seems plausible that the absence of CobF from the MCRA of the MTBC might have forced previously environmental bacteria to rely stronger on the mammalian host as a source of vitamin B12 (Boritsch et al. 2014; Young et al. 2015), thereby enhancing their specialisation as pathogens. *M. canettii* strains thus represent very interesting study objects to explore this hypothesis in more detail.

Apart from genes that are specifically present in *M. canettii* and absent from the MTBC, the opposite configuration also exists, involving for example the gene *pe_pgrs33* (*rv1818c*), which is absent from all *M. canetti* but is found in all MTBC strains suggesting a lateral transfer of *pe_pgrs33* to the common ancestor of the MTBC at the beginning of its clonal emergence and after its divergence from *M. canettii*. Interestingly, PE_PGRS33 represents a unique member that has been associated with pathogenicity (Cadieux et al. 2011), within the large family of Pro-Glu (PE) motive-containing proteins of *M. tuberculosis* that are surface exposed (Delogu et al. 2004; Banu et al. 2002) and which are often associated with ESX/type VII secretion systems (Bottai and Brosch 2009; Groschel et al. 2016).

Finally, from the literature and from confirmative experiments it is known that the *M. canettii* strains do not exhibit a special antibiotic resistance phenotype and resemble in this aspect very closely *M. tuberculosis*, with one exception, which is linked to the resistance to pyrazinamide. Indeed *M. canettii* strains are intrinsically resistant to this first line anti-tuberculosis drug for reasons that are still unclear, but possibly linked to mutations in *rpsA* and/or *panD* (Somoskovi et al. 2007; Feuerriegel et al. 2013, 2014).

In conclusion, the *M. canettii* strains represent an interesting group of strains that are phylogenetically positioned at the interphase between the key obligate mycobacterial pathogens of the MTBC and environmental mycobacteria. Although the *M. canettii* strains isolated so far are all human isolates and no environmental isolation of *M. canettii* has yet been possible, the genomic and phenotypic characteristics of these strains point to a potential environmental habitat. Indeed, given the mosaic genome organisation of *M. canettii* strains that shows numerous traces of recent inter-strain recombination and other lateral gene acquisition events, it seems reasonable to suppose that they have occurred in an environmental niche, possibly involving protozooic hosts, which allow frequent contact of different *M. canettii* and other (myco)bacterial strains. Given the geographical restriction of patients with *M. canettii* infections to countries in the region of the Horn of Africa, it is tempting to speculate that adequate environmental habitats might exist in that area. Finally, *M. canettii* strains still hold on to many secrets, whose elucidation have the potential to teach us important

details of the origin of, and the molecular events contributing to, the emergence of the MTBC, one of the most devastating pathogens in human history.

Acknowledgements We are grateful to our colleagues who were involved in the different research projects that are summarized and reviewed in this chapter. Support by grants from the European Union (H2020-PHC-643381), the Agence National de Recherche (ANR-14-JAMR-001-02), the Fondation pour la Recherche Médicale (DEQ20130326471), the Institut Pasteur (PTR 314 and PTR 383), and Genoscope collaborative grant 114 are gratefully acknowledged. P.S. is a consultant for Genoscreen. R.B. is a member of the LabEx consortium IBEID at the Institut Pasteur.

References

Alexander KA, Laver PN, Michel AL, Williams M, van Helden PD, Warren RM, Gey van Pittius NC (2010) Novel Mycobacterium tuberculosis complex pathogen, M. mungi. Emerg Infect Dis 16(8):1296–1299

Alexander KA, Sanderson CE, Larsen MH, Robbe-Austerman S, Williams MC, Palmer MV (2016) Emerging tuberculosis pathogen hijacks social communication behavior in the group-living banded mongoose (*Mungos mungo*). MBio 7(3):e00281–e00216

Aranaz A, Liebana E, Gomez-Mampaso E, Galan JC, Cousins D, Ortega A, Blazquez J, Baquero F, Mateos A, Suarez G, Dominguez L (1999) Mycobacterium tuberculosis subsp. caprae subsp. nov.: a taxonomic study of a new member of the Mycobacterium tuberculosis complex isolated from goats in Spain. Int J Syst Bacteriol 49(Pt 3):1263–1273

Banu S, Honore N, Saint-Joanis B, Philpott D, Prevost MC, Cole ST (2002) Are the PE-PGRS proteins of Mycobacterium tuberculosis variable surface antigens? Mol Microbiol 44(1):9–19

Blouin Y, Hauck Y, Soler C, Fabre M, Vong R, Dehan C, Cazajous G, Massoure PL, Kraemer P, Jenkins A, Garnotel E, Pourcel C, Vergnaud G (2012) Significance of the identification in the Horn of Africa of an exceptionally deep branching Mycobacterium tuberculosis clade. PLoS One 7(12):e52841

Blouin Y, Cazajous G, Dehan C, Soler C, Vong R, Hassan MO, Hauck Y, Boulais C, Andriamanantena D, Martinaud C, Martin E, Pourcel C, Vergnaud G (2014) Progenitor "Mycobacterium canettii" clone responsible for lymph node tuberculosis epidemic, Djibouti. Emerg Infect Dis 20(1):21–28

Boritsch EC, Brosch R (2016) Evolution of Mycobacterium tuberculosis: new insights into pathogenicity and drug resistance. Microbiol Spectr 4(5):TBTB2-0020-2016

Boritsch EC, Supply P, Honore N, Seemann T, Stinear TP, Brosch R (2014) A glimpse into the past and predictions for the future: the molecular evolution of the tuberculosis agent. Mol Microbiol 93(5):835–852

Boritsch EC, Frigui W, Cascioferro A, Malaga W, Etienne G, Laval F, Pawlik A, Le Chevalier F, Orgeur M, Ma L, Bouchier C, Stinear TP, Supply P, Majlessi L, Daffe M, Guilhot C, Brosch R (2016a) pks5-recombination-mediated surface remodelling in Mycobacterium tuberculosis emergence. Nat Microbiol 1(2):15019

Boritsch EC, Khanna V, Pawlik A, Honore N, Navas VH, Ma L, Bouchier C, Seemann T, Supply P, Stinear TP, Brosch R (2016b) Key experimental evidence of chromosomal DNA transfer among selected tuberculosis-causing mycobacteria. Proc Natl Acad Sci U S A 113(35):9876–9881

Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, Forrest SA, Bryant JM, Harris SR, Schuenemann VJ, Campbell TJ, Majander K, Wilbur AK, Guichon RA, Wolfe Steadman DL, Cook DC, Niemann S, Behr MA, Zumarraga M, Bastida R, Huson D, Nieselt K, Young D, Parkhill J, Buikstra JE, Gagneux S, Stone AC, Krause J (2014) Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature 514(7523):494–497

Bottai D, Brosch R (2009) Mycobacterial PE, PPE and ESX clusters: novel insights into the secretion of these most unusual protein families. Mol Microbiol 73(3):325–328

Brisse S, Supply P, Brosch R, Vincent V, Gutierrez MC (2006) "A re-evaluation of M. prototuberculosis": continuing the debate. PLoS Pathog 2(9):e95

Brosch R, Guilhot C (2015) Bovine tubercle bacilli: evolution associated with loss of transmission capacity in humans. Med Sci (Paris) 31(2):123–126

Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, Garnier T, Gutierrez C, Hewinson G, Kremer K, Parsons LM, Pym AS, Samper S, van Soolingen D, Cole ST (2002) A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci U S A 99(6):3684–3689

Cadieux N, Parra M, Cohen H, Maric D, Morris SL, Brennan MJ (2011) Induction of cell death after localization to the host cell mitochondria by the Mycobacterium tuberculosis PE_PGRS33 protein. Microbiology 157(Pt 3):793–804

Canetti G (1965) The J. Burns Amberson lecture. Present aspects of bacterial resistance in tuberculosis. Am Rev Respir Dis 92:687

Carver T, Berriman M, Tivey A, Patel C, Bohme U, Barrell BG, Parkhill J, Rajandream MA (2008) Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. Bioinformatics 24(23):2672–2676

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K,

Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 393(6685): 537–544

Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, Parkhill J, Malla B, Berg S, Thwaites G, Yeboah-Manu D, Bothamley G, Mei J, Wei L, Bentley S, Harris SR, Niemann S, Diel R, Aseffa A, Gao Q, Young D, Gagneux S (2013) Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. Nat Genet 45(10): 1176–1182

Comas I, Hailu E, Kiros T, Bekele S, Mekonnen W, Gumi B, Tschopp R, Ameni G, Hewinson RG, Robertson BD, Goig GA, Stucki D, Gagneux S, Aseffa A, Young D, Berg S (2015) Population genomics of Mycobacterium tuberculosis in Ethiopia contradicts the virgin soil hypothesis for human tuberculosis in Sub-Saharan Africa. Curr Biol 25(24):3260–3266

Coros A, Callahan B, Battaglioli E, Derbyshire KM (2008) The specialized secretory apparatus ESX-1 is essential for DNA transfer in Mycobacterium smegmatis. Mol Microbiol 69(4):794–808

Coscolla M, Lewin A, Metzger S, Maetz-Rennsing K, Calvignac-Spencer S, Nitsche A, Dabrowski PW, Radonic A, Niemann S, Parkhill J, Couacy-Hymann E, Feldman J, Comas I, Boesch C, Gagneux S, Leendertz FH (2013) Novel Mycobacterium tuberculosis complex isolate from a wild chimpanzee. Emerg Infect Dis 19(6):969–976

Cousins DV, Bastida R, Cataldi A, Quse V, Redrobe S, Dow S, Duignan P, Murray A, Dupont C, Ahmed N, Collins DM, Butler WR, Dawson D, Rodriguez D, Loureiro J, Romano MI, Alito A, Zumarraga M, Bernardelli A (2003) Tuberculosis in seals caused by a novel member of the Mycobacterium tuberculosis complex: Mycobacterium pinnipedii sp. nov. Int J Syst Evol Microbiol 53(Pt 5):1305–1314

Daffe M, McNeil M, Brennan PJ (1991) Novel type-specific lipooligosaccharides from Mycobacterium tuberculosis. Biochemistry 30(2):378–388

de Jong BC, Antonio M, Gagneux S (2010) Mycobacterium africanum – review of an important cause of human tuberculosis in West Africa. PLoS Negl Trop Dis 4(9):e744

Delogu G, Pusceddu C, Bua A, Fadda G, Brennan MJ, Zanetti S (2004) Rv1818c-encoded PE_PGRS protein of Mycobacterium tuberculosis is surface exposed and influences bacterial cell structure. Mol Microbiol 52(3):725–733

Derbyshire KM, Gray TA (2014) Distributive conjugal transfer: new insights into horizontal gene transfer and genetic exchange in Mycobacteria. Microbiol Spectr 2(1):04

Domogalla J, Prodinger WM, Blum H, Krebs S, Gellert S, Muller M, Neuendorf E, Sedlmaier F, Buttner M (2013) Region of difference 4 in alpine Mycobacterium caprae isolates indicates three variants. J Clin Microbiol 51(5):1381–1388

Dormans J, Burger M, Aguilar D, Hernandez-Pando R, Kremer K, Roholl P, Arend SM, van Soolingen D (2004) Correlation of virulence, lung pathology, bacterial load and delayed type hypersensitivity responses after infection with different Mycobacterium tuberculosis genotypes in a BALB/c mouse model. Clin Exp Immunol 137(3):460–468

Fabre M, Koeck JL, Le Fleche P, Simon F, Herve V, Vergnaud G, Pourcel C (2004) High genetic diversity revealed by variable-number tandem repeat genotyping and analysis of hsp65 gene polymorphism in a large collection of "Mycobacterium canettii" strains indicates that the M. tuberculosis complex is a recently emerged clone of "M. canettii". J Clin Microbiol 42(7):3248–3255

Fabre M, Hauck Y, Soler C, Koeck JL, van Ingen J, van Soolingen D, Vergnaud G, Pourcel C (2010) Molecular characteristics of "Mycobacterium canettii" the smooth Mycobacterium tuberculosis bacilli. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis 10(8):1165–1173

Feuerriegel S, Koser CU, Richter E, Niemann S (2013) Mycobacterium canettii is intrinsically resistant to both pyrazinamide and pyrazinoic acid. J Antimicrob Chemother 68(6):1439–1440

Feuerriegel S, Koser CU, Niemann S (2014) Phylogenetic polymorphisms in antibiotic resistance genes of the Mycobacterium tuberculosis complex. J Antimicrob Chemother 69(5):1205–1210

Gagneux S, Small PM (2007) Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development. Lancet Infect Dis 7(5):328–337

Gagneux S, Deriemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, Nicol M, Niemann S, Kremer K, Gutierrez MC, Hilty M, Hopewell PC, Small PM (2006) Variable host-pathogen compatibility in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 103:2869–2873

Garnier T, Eiglmeier K, Camus JC, Medina N, Mansoor H, Pryor M, Duthoy S, Grondin S, Lacroix C, Monsempe C, Simon S, Harris B, Atkin R, Doggett J, Mayes R, Keating L, Wheeler PR, Parkhill J, Barrell BG, Cole ST, Gordon SV, Hewinson RG (2003) The complete genome sequence of Mycobacterium bovis. Proc Natl Acad Sci U S A 100(13):7877–7882

Gonzalo-Asensio J, Malaga W, Pawlik A, Astarie-Dequeker C, Passemar C, Moreau F, Laval F, Daffe M, Martin C, Brosch R, Guilhot C (2014) Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence regulator. Proc Natl Acad Sci U S A 111(31):11491–11496

Gopinath K, Moosa A, Mizrahi V, Warner DF (2013) Vitamin B(12) metabolism in Mycobacterium tuberculosis. Future Microbiol 8(11):1405–1418

Gray TA, Krywy JA, Harold J, Palumbo MJ, Derbyshire KM (2013) Distributive conjugal transfer in mycobacteria generates progeny with meiotic-like genome-wide mosaicism, allowing mapping of a mating identity locus. PLoS Biol 11(7):e1001602

Groschel MI, Sayes F, Simeone R, Majlessi L, Brosch R (2016) ESX secretion systems: mycobacterial evolution to counter host immunity. Nat Rev Microbiol 14(11):677–691

Gutierrez MC, Brisse S, Brosch R, Fabre M, Omais B, Marmiesse M, Supply P, Vincent V (2005) Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis. PLoS Pathog 1(1):e5

Hendrix RW, Smith MC, Burns RN, Ford ME, Hatfull GF (1999) Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc Natl Acad Sci U S A 96(5):2192–2197

Horstkotte MA, Sobottka I, Schewe CK, Schafer P, Laufs R, Rusch-Gerdes S, Niemann S (2001) Mycobacterium microti llama-type infection presenting as pulmonary tuberculosis in a human immunodeficiency virus-positive patient. J Clin Microbiol 39(1):406–407

Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, van Embden J (1997) Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. J Clin Microbiol 35(4):907–914

Kelley DS, Lennon CW, Belfort M, Novikova O (2016) Mycobacteriophages as incubators for intein dissemination and evolution. MBio 7(5):e01537–e01516

Kiers A, Klarenbeek A, Mendelts B, Van Soolingen D, Koeter G (2008) Transmission of Mycobacterium pinnipedii to humans in a zoo with marine mammals. Int J Tuberc Lung Dis 12(12):1469–1473

Koeck JL, Bernatas JJ, Gerome P, Fabre M, Houmed A, Herve V, Teyssou R (2002) Epidemiology of resistance to antituberculosis drugs in Mycobacterium tuberculosis complex strains isolated from adenopathies in Djibouti. Prospective study carried out in 1999. Med Trop (Mars) 62(1):70–72

Koeck JL, Fabre M, Simon F, Daffe M, Garnotel E, Matan AB, Gerome P, Bernatas JJ, Buisson Y, Pourcel C (2011) Clinical characteristics of the smooth tubercle bacilli 'Mycobacterium canettii' infection suggest the existence of an environmental reservoir. Clin Microbiol Infect 17(7):1013–1019

Koro Koro F, Kamdem Simo Y, Piam FF, Noeske J, Gutierrez C, Kuaban C, Eyangoh SI (2013) Population dynamics of tuberculous Bacilli in Cameroon as assessed by spoligotyping. J Clin Microbiol 51(1):299–302

Kremer K, van Soolingen D, Frothingham R, Haas WH, Hermans PW, Martin C, Palittapongarnpim P, Plikaytis BB, Riley LW, Yakrus MA, Musser JM, van Embden JD (1999) Comparison of methods based on different molecular epidemiological markers for typing of Mycobacterium tuberculosis complex strains: interlaboratory study of discriminatory power and reproducibility. J Clin Microbiol 37(8):2607–2618

Lemassu A, Levy-Frebault VV, Laneelle MA, Daffe M (1992) Lack of correlation between colony morphology and lipooligosaccharide content in the Mycobacterium tuberculosis complex. J Gen Microbiol 138(7):1535–1541

Magee JG, Ward AC (2012) Genus I. Mycobacterium. In: Goodfellow M, Kämpfer P, Busse HJ et al (eds) Bergey's manual of systematic bacteriology, the actinobacteria, vol 5, 2nd edn. Springer, New York, pp 312–375

Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV (2011) Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol 9(6):467–477

Mortimer TD, Pepperell CS (2014) Genomic signatures of distributive conjugal transfer among mycobacteria. Genome Biol Evol 6(9):2489–2500

Mostowy S, Cousins D, Behr MA (2004) Genomic interrogation of the dassie bacillus reveals it as a unique RD1 mutant within the Mycobacterium tuberculosis complex. J Bacteriol 186(1):104–109

Muller B, Durr S, Alonso S, Hattendorf J, Laisse CJ, Parsons SD, van Helden PD, Zinsstag J (2013) Zoonotic Mycobacterium bovis-induced tuberculosis in humans. Emerg Infect Dis 19(6):899–908

Namouchi A, Didelot X, Schock U, Gicquel B, Rocha EP (2012) After the bottleneck: genome-wide diversification of the Mycobacterium tuberculosis complex by mutation, recombination, and natural selection. Genome Res 22(4):721–734

Nebenzahl-Guimaraes H, Yimer SA, Holm-Hansen C, de Beer J, Brosch R, van Soolingen D (2016) Genomic characterization of *Mycobacterium tuberculosis* lineage 7 and a proposed name: "Aethiops vetus". Microbial Genomics 2:e000063. doi:10.1099/mgen.0.000063

Panas MW, Jain P, Yang H, Mitra S, Biswas D, Wattam AR, Letvin NL, Jacobs WR Jr (2014) Noncanonical SMC protein in Mycobacterium smegmatis restricts maintenance of Mycobacterium fortuitum plasmids. Proc Natl Acad Sci U S A 111(37):13264–13271

Parsons SD, Drewe JA, Gey van Pittius NC, Warren RM, van Helden PD (2013) Novel cause of tuberculosis in meerkats, South Africa. Emerg Infect Dis 19(12):2004–2007

Perez E, Samper S, Bordas Y, Guilhot C, Gicquel B, Martin C (2001) An essential role for phoP in Mycobacterium tuberculosis virulence. Mol Microbiol 41(1):179–187

Pfyffer GE, Auckenthaler R, van Embden JD, van Soolingen D (1998) Mycobacterium canettii, the smooth variant of M. tuberculosis, isolated from a Swiss patient exposed in Africa. Emerg Infect Dis 4(4):631–634

Smith NH (2006) A re-evaluation of M. prototuberculosis. PLoS Pathog 2(9):e98

Smith NH, Hewinson RG, Kremer K, Brosch R, Gordon SV (2009a) Myths and misconceptions: the origin and evolution of Mycobacterium tuberculosis. Nat Rev Microbiol 7(7):537–544

Smith NH, Crawshaw T, Parry J, Birtles RJ (2009b) Mycobacterium microti: More diverse than previously thought. J Clin Microbiol 47(8):2551–2559

Somoskovi A, Dormandy J, Parsons LM, Kaswa M, Goh KS, Rastogi N, Salfinger M (2007) Sequencing of the

pncA gene in members of the Mycobacterium tuberculosis complex has important diagnostic applications: Identification of a species-specific pncA mutation in "Mycobacterium canettii" and the reliable and rapid predictor of pyrazinamide resistance. J Clin Microbiol 45(2):595–599.

Somoskovi A, Dormandy J, Mayrer AR, Carter M, Hooper N, Salfinger M (2009) "Mycobacterium canettii" isolated from a human immunodeficiency virus-positive patient: first case recognized in the United States. J Clin Microbiol 47(1):255–257.

Supply P, Marceau M, Mangenot S, Roche D, Rouanet C, Khanna V, Majlessi L, Criscuolo A, Tap J, Pawlik A, Fiette L, Orgeur M, Fabre M, Parmentier C, Frigui W, Simeone R, Boritsch EC, Debrie AS, Willery E, Walker D, Quail MA, Ma L, Bouchier C, Salvignol G, Sayes F, Cascioferro A, Seemann T, Barbe V, Locht C, Gutierrez MC, Leclerc C, Bentley SD, Stinear TP, Brisse S, Médigue C, Parkhill J, Cruveiller S, Brosch R (2013) Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of Mycobacterium tuberculosis. Nat Genet 45:172–179

van Ingen J, Rahim Z, Mulder A, Boeree MJ, Simeone R, Brosch R, van Soolingen D (2012) Characterization of Mycobacterium orygis as M. tuberculosis complex subspecies. Emerg Infect Dis 18(4):653–655

van Soolingen D, Hoogenboezem T, de Haas PE, Hermans PW, Koedam MA, Teppema KS, Brennan PJ, Besra GS, Portaels F, Top J, Schouls LM, van Embden JD (1997) A novel pathogenic taxon of the Mycobacterium tuberculosis complex, Canetti: characterization of an exceptional isolate from Africa. Int J Syst Bacteriol 47(4):1236–1245

van Soolingen D, van der Zanden AG, de Haas PE, Noordhoek GT, Kiers A, Foudraine NA, Portaels F, Kolk AH, Kremer K, van Embden JD (1998) Diagnosis of Mycobacterium microti infections among humans by using novel genetic markers. J Clin Microbiol 36(7):1840–1845

Wells AQ (1937) Tuberculosis in wild voles. Lancet 1221

Young DB, Comas I, de Carvalho LP (2015) Phylogenetic analysis of vitamin B12-related metabolism in Mycobacterium tuberculosis. Front Mol Biosci 2:6

The Evolution of Strain Typing in the *Mycobacterium tuberculosis* Complex

3

Matthias Merker, Thomas A. Kohl, Stefan Niemann, and Philip Supply

Abstract

Tuberculosis (TB) is a contagious disease with a complex epidemiology. Therefore, molecular typing (genotyping) of *Mycobacterium tuberculosis* complex (MTBC) strains is of primary importance to effectively guide outbreak investigations, define transmission dynamics and assist global epidemiological surveillance of the disease. Large-scale genotyping is also needed to get better insights into the biological diversity and the evolution of the pathogen. Thanks to its shorter turnaround and simple numerical nomenclature system, mycobacterial interspersed repetitive unit–variable-number tandem repeat (MIRU-VNTR) typing, based on 24 standardized plus 4 hypervariable loci, optionally combined with spoligotyping, has replaced IS*6110* DNA fingerprinting over the last decade as a gold standard among classical strain typing methods for many applications. With the continuous progress and decreasing costs of next-generation sequencing (NGS) technologies, typing based on whole genome sequencing (WGS) is now increasingly performed for near complete exploitation of the available genetic information. However, some important challenges remain

Stefan Niemann and Philip Supply contributed equally to this work.

M. Merker • T.A. Kohl • S. Niemann (✉)
Molecular and Experimental Mycobacteriology, Forschungszentrum Borstel, Leibniz-Zentrum für Medizin und Biowissenschaften, Parkallee 1, 23845, Borstel, Germany

German Center for Infection Research (DZIF), partner site Borstel, Borstel, Germany
e-mail: mmerker@fz-borstel.de; tkohl@fz-borstel.de; sniemann@fz-borstel.de

P. Supply (✉)
Université de Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 8204 - CIIL - Centre d'Infection et d'Immunité de Lille, F-59019, Lille, France
e-mail: philip.supply@ibl.cnrs.fr

S. Gagneux (ed.), *Strain Variation in the* Mycobacterium tuberculosis *Complex: Its Role in Biology, Epidemiology and Control*, Advances in Experimental Medicine and Biology 1019,
DOI 10.1007/978-3-319-64371-7_3

such as the lack of standardization of WGS analysis pipelines, the need of databases for sharing WGS data at a global level, and a better understanding of the relevant genomic distances for defining clusters of recent TB transmission in different epidemiological contexts. This chapter provides an overview of the evolution of genotyping methods over the last three decades, which culminated with the development of WGS-based methods. It addresses the relative advantages and limitations of these techniques, indicates current challenges and potential directions for facilitating standardization of WGS-based typing, and provides suggestions on what method to use depending on the specific research question.

Keywords
Tuberculosis • Mycobacterium tuberculosis • Genotyping • Whole genome sequencing • Transmission • Phylogeny

3.1 Introduction: Need for Effective Genotyping of the MTBC

The *Mycobacterium tuberculosis* complex (MTBC) comprises *Mycobacterium tuberculosis* and *Mycobacterium africanum*, the main causative agents of human TB, as well as several members causing TB in various wild and domestic animal species (e.g. *Mycobacterium bovis*, *Mycobacterium microti* and *Mycobacterium pinnipedii*). During several decades, typing of the MTBC has been the subject of intensive research in numerous fields and applications. These include local and global surveillance, as well as the evolution of the pathogen and its interaction with the human host. This wide range of applications reflects the global importance, the complex epidemiology and the multiple transmission patterns of TB, as well as the numerous questions and challenges that remain to fully understand the pathobiology and transmissibility of the organism, and for improving the diagnostics, the treatment and the prophylaxis of the disease.

According to the last World Health Organization (WHO) global TB report, this disease caused 10.4 million new cases and 1.8 million deaths in 2015 (WHO 2016). Moreover, although this estimation is uncertain, approximately two billion individuals are thought to be infected with MTBC (Corbett et al. 2003; Getahun et al. 2015). The majority of active TB cases occur within the first 2–5 years after the initial infection, although risk of progression to active disease is distributed over the complete lifetime. As clinical signs of active TB may not be directly evocative and are often non-specific, diagnosis may be delayed, resulting in prolonged patient illness and infectiousness. Delays in the diagnosis and in the application of control measures can result in further spread of TB, even among individuals thought to be at low risk (Coitinho et al. 2013). Transmission is not restricted to household members or other frequent contacts, but can extend to social and casual contacts as well, which is more difficult to trace by classical epidemiological investigation (Gardy et al. 2011; Lambregts-van Weezenbeek et al. 2003; Wang et al. 2014). Furthermore, the comprehensiveness and potential side effects of the standard treatment regimen, combining four drugs for 2 months followed by two drugs for 4 months, can result in patient non-compliance. When going undetected for a prolonged period of time, non-compliant individuals can remain contagious and be the source of large, unsuspected TB outbreaks (Walker et al. 2013a; Small et al. 1994). Such insidious modes and undetected sources of TB infection and transmission strongly support the use of strain typing methods for effectively guiding epidemiological investigations, identifying transmission hotspots in particular settings or searching for the causes of the spread of particular strain types at regional or global levels.

The interest in typing of MTBC is additionally fuelled by the globally raising threat represented by multi- (MDR; defined as resistant to the main two first-line drugs isoniazid and rifampicin) and extensively drug resistant (XDR; defined as MDR plus additional resistance to also resistant to an aminoglycoside and at least one fluoroquinolone) TB (Dye and Williams 2010; Gandhi et al. 2010; Marais 2016). MDR and, even more so, XDR-TB cases suffer increased treatment failure and higher mortality rates, and represent considerably higher costs for health systems (Diel et al. 2014). Drug resistance, MDR or XDR can be acquired de novo and also further increase during treatment, for instance as a result of patient non-adherence, or when applying standardized treatment regimens without adjusting for possible baseline drug resistance (Cegielski et al. 2014; Daley and Horsburgh 2014). However, globally, most MDR-TB cases occur among new TB patients, thus without any antecedent of previous anti-TB treatment (WHO 2016), pointing to the importance of direct transmission of MTBC strains that are already drug-resistant. As a prominent example, more than 30% of new TB patients were diagnosed with MDR-TB in a recent nationwide survey in Belarus (Skrahina et al. 2013) (see Chap. 11). Transmission of MDR-TB has been repeatedly documented in nosocomial contexts (Allix et al. 2004), as well as in the context of large outbreaks at regional (Bifani et al. 1996; Cohen et al. 2015; Eldholm et al. 2015; Cox et al. 2005; Niemann et al. 1997) and intercontinental scales (Cooke et al. 2011; Williams et al. 2015; Coscolla et al. 2015; Merker et al. 2015). Even more worrisome, solid evidence exists on large, province-wide epidemic transmission of XDR-TB in South Africa (Cohen et al. 2015). Direct transmission of XDR-TB in the community is also beginning to be seen, even in low TB incidence settings such as the UK (Arnold et al. 2016) and France (Lafeuille et al. 2016). Exogenous re-infection by XDR strains has also been shown to contribute to emergence of XDR-TB in patients under treatment for MDR-TB in regions like Uzbekistan (Cox et al. 2008). Given the major adverse consequences of MDR- and XDR-TB for individual patients and public health overall, tracing the modes of acquisition and routes of transmission of these most dangerous forms of TB should be a high priority.

The need for effective MTBC genotyping methods is also driven by more basic research questions, such as defining the biological and geographical diversity of the pathogen and its association with host populations (Hirsh et al. 2004; Tsolaki et al. 2004; Gagneux et al. 2006; Comas et al. 2013; Brudey et al. 2006; Hershberg et al. 2008), understanding the origin of the MTBC (Gutierrez et al. 2005, 2006; Supply et al. 2013; Fabre et al. 2004, 2010; Boritsch et al. 2014, 2016a), and retracing the evolutionary history and the demography of the MTBC (Pepperell et al. 2013; Brosch et al. 2002; Mostowy et al. 2002; Comas et al. 2013, 2015; Wirth et al. 2008; Mokrousov et al. 2016; Merker et al. 2015; Luo et al. 2015) and of some of its major clones in association with multidrug resistance (Mokrousov 2013; Merker et al. 2015).

Investigation of some of these questions, as well as of some questions related to the public health issues listed above, was already pioneered by the use of phage typing, which is based on defining the susceptibility or resistance of MTBC strains to lytic infection by different mycobacteriophages. Using this approach, Bates and Mitchinson (Bates and Mitchison 1969) were the first to show differences in (phage) types among *M. tuberculosis* strains from different geographic origins. Phage typing was also used to provide evidence for the contribution of re-infection in cases of TB relapse (Snider et al. 1984; Raleigh and Wichelhausen 1973), for the occurrence of simultaneous infection with more than one *M. tuberculosis* strains in some patients (Mankiewicz and Liivak 1975), and to document TB transmission between individuals (Stead and Bates 1969; Kitahara 1973; Snider et al. 1984). However, the low level of resolution linked to the relatively restricted number of phage types identified, the technical difficulty of the method and its lack of standardization restricted its use to few laboratories.

These hurdles were gradually resolved by the development of molecular, DNA-based

typing – also termed genotyping – methods of MTBC strains, now culminating with whole genome sequencing (WGS). The following section will provide an overview of different "classical" genotyping methods – defined as methods that interrogate markers in a targeted manner, in contrast to WGS – that were developed over more than a decade, with an emphasis on the three methods that reached global application, namely IS*6110* DNA fingerprinting based on restriction fragment length polymorphism (RFLP) analysis, spoligotyping, and mycobacterial interspersed repetitive-unit–variable-number tandem-repeat (MIRU-VNTR) typing. A concise overview will also be provided on genotyping methods based on the analysis of genomic deletions (also termed regions of differences (RDs) or large sequence polymorphisms (LSPs)) and pre-defined single nucleotide polymorphisms (SNPs), which are more reserved for phylogenetic applications. As WGS will be more extensively treated in Chap. 4 (i.e. on genomic epidemiology), some key technical aspects will be introduced here, with a focused comparison of classical reference-based SNP analysis and genome-wide multilocus sequence typing (MLST) for future standardized WGS-based genotyping. Final sections will discuss definitions of molecular clusters based on classical genotyping and WGS, and discuss practical implications on the choice of the methods depending on the research contexts and the questions that are prioritized.

3.2 Classical Genotyping

3.2.1 General Overview

The tubercle bacilli are slow growing bacteria, with a relatively low level of genetic diversity of about 1 SNP per 2 kbp (Sreevatsan et al. 1997; Comas et al. 2010) and a high degree of clonality (Hirsh et al. 2004; Supply et al. 2003), with the exception of the rare *Mycobacterium canettii* strains (Fabre et al. 2004; Gutierrez et al. 2005; Supply et al. 2013) (see Chap. 2). These specific properties have largely prevented the application of genotyping techniques classically used for other bacterial species, such as pulse-field gel electrophoresis (Arbeit et al. 1990) or multi-locus sequence typing (MLST) (Maiden et al. 1998), requiring substantial chromosomal variation and/or diversity of house-keeping gene sequences for molecular discrimination at strain level, respectively. These constraints have prompted the development of a variety of alternative genotyping assays for the MTBC.

First attempts relied on the analysis of RFLP profiles in gels after digestion of chromosomal DNA purified from MTBC strains with restriction endonucleases, but they produced very complex patterns with a low level of differentiation between strains (Collins and De Lisle 1984). More interpretable patterns and greater polymorphism were obtained when specific restriction fragments were detected by hybridization, especially with the use of IS*6110* insertion sequence as a labeled DNA probe (Fig. 3.1) (reviewed in (Crawford 1993; van Embden et al. 1992)). Other types of repetitive DNA – identified by cloning or by analysis of the genome sequence of *M. tuberculosis* H37Rv gradually available at the time (Cole et al. 1998) – were also used as alternative probes, including for instance other insertion sequences, the polymorphic GC-rich tandem repeat sequence (PGRS) (Poulet and Cole 1995), or a poly GTG repeat (Wiid et al. 1994).

Because IS*6110* RFLP and other RFLP methods are labor intensive and require weeks of culturing to obtain sufficient amounts of purified DNA, a range of PCR-based methods targeting different polymorphic markers were also developed over the years 1990–2000, comprising spoligotyping (Fig. 3.1), MIRU-VNTR typing (Fig. 3.1) based on different sets and numbers of markers, plus various other techniques such as ligation-mediated PCR (Prod'hom et al. 1997), and fluorescent amplified fragment length polymorphism (FAFLP) (Ahmed et al. 2003; Wiid et al. 1994). Opportunely, a total of 20 RFLP-based or PCR-based techniques, representing most of the methods developed primarily for epidemiological investigation, were subjected to comparisons of their relative performances by using a same standard collection of MBTC

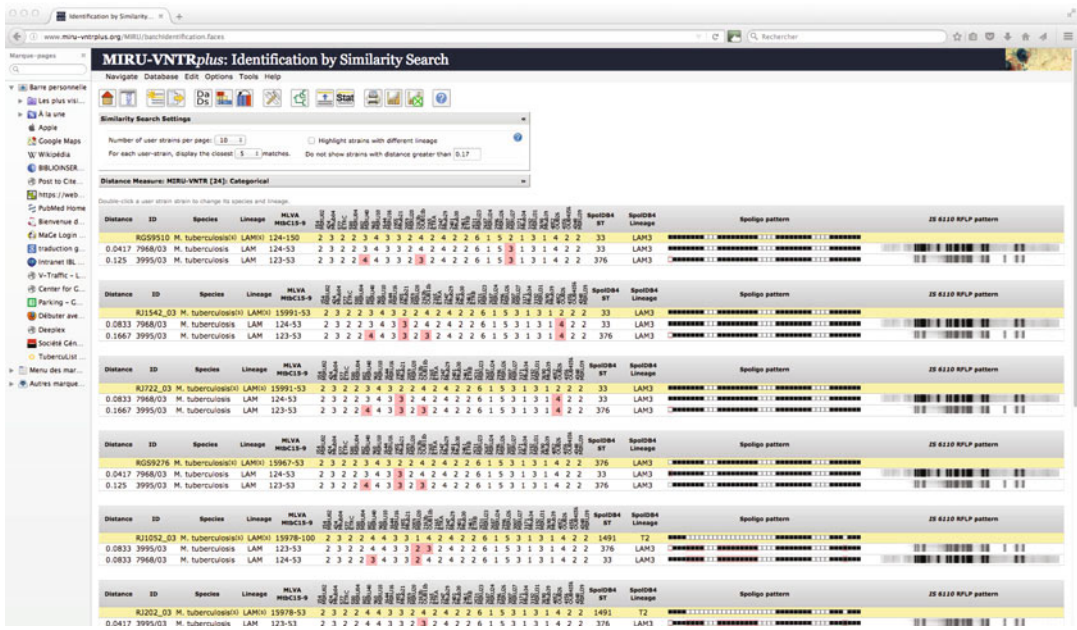

Fig. 3.1 Standard 24-locus MIRU-VNTR genotypes, spoligotypes and IS*6110*-RFLP types of MTBC clinical isolates, visualized under the MIRU-VNTRplus database. MIRU-VNTR types are represented by 24-number strings, spoligotypes by digitalised 43-spacer profiles (individual spacers present or absent are shown as *black* or *white* boxes, respectively), IS*6110* RFLP profiles by banding patterns. Genotypes of isolates as submitted by a database user (shown with a *yellow* background) are analyzed for similarity with genotypes of reference isolates of the database. Allelic differences in MIRU-VNTR genotypes and spacer differences in spoligotypes relative to genotypes of reference isolates are shown with a *red* background and red squares, respectively. Distance on the extreme *left* refers to allelic distances between user and reference isolates based on 24-locus MIRU-VNTR genotypes (e.g. 1 allele difference among the 24 loci equals to a distance of 0.0417). Other information shown from *left* to *right* includes strain identification number (ID), species, genetic lineage, standard 24-locus MIRU-VNTR nomenclature (MLVA MtbC15-9), and spoligotype designation according to SpolDB4. MLVA MtbC15-9 refers to nomenclature codes specifically identifying each known MIRU-VNTR haplotype based on the combination of the 15 most discriminatory loci and 9 auxiliary loci, respectively (see text). Standard MIRU-VNTR locus names are shown according to locus position on the *M. tuberculosis* H37Rv chromosome (in kbp; i.e. 154, 424, etc.), in addition to alias names according to historical designation (i.e. MIRU02, Mtub04, ETRC, etc.). IS*6110*-RFLP profiles are only shown for reference isolates

strains set up by Kremer and Colleagues (Kremer et al. 1999), in different studies (Kremer et al. 1999, 2005a; Supply et al. 2001, 2006). A hierarchy was established on this normalized basis among the methods according to three key parameters. The discriminatory or resolution power, that is the ability of the markers to distinguish epidemiologically unrelated strains, was evaluated by determining the number of different types detected among DNA samples from 90 MTBC strains selected from diverse geographical origins. The number of typeable strains was also determined among these strains, while the reproducibility was assessed by analyzing 31 duplicated DNA samples. The results demonstrated a clear superiority of MIRU-VNTR typing based on 15 or 24 loci and IS*6110* RFLP among the PCR- and RFLP-based methods, with 89 and 84 types distinguished in this collection of 90 MTBC strains, respectively, and 100% reproducibility and 98–100% typeability in both cases (Table 3.1). Both methods are described more extensively below, together with the other major "classical" methods still in use today, i.e. spoligotyping, SNP typing and typing based on regions of difference.

3.2.2 IS6110 RFLP

IS*6110* RFLP was developed following the identification of copies of the corresponding

Table 3.1 Resolution power and reproducibility by using MTBC genotyping methods for differentiating 90 MTBC strains

Method	PCR or RFLP	DNA targets	N° types obtained	Reproducibility (%)	N° non typeable strains
24-locus MIRU-VNTR	Multiplex PCR	MIRU+ETR+QUB	89	100[a]	0[b]
15-locus MIRU-VNTR	Multiplex PCR	MIRU+ETR+QUB	89	100[a]	0[b]
IS*6110* RFLP (PvuII)	RFLP	IS*6110*	84	100	2
5-locus MIRU-VNTR	PCR	Five QUBs	82	87	0
IS*6110* mixed-linker PCR	PCR	IS*6110*	81	100	2
FLiP	PCR	IS*6110*	81	97	2
12-locus MIRU-VNTR	Multiplex PCR	12 MIRUs	78	100	0
LM-PCR	PCR	IS*6110*	73	81	2
APPCR	PCR	Total genome	71	71	0
PGRS RFLP (*Alu*I)	RFLP	PGRS	70	100	0
DRE-PCR	PCR	IS*6110*/PGRS	63	58	0
Spoligotyping	PCR	DR locus	61	94	0
Second-generation spoligotyping	PCR	DR locus	61	90	0
5-locus MIRU-VNTR	PCR	ETRs A–E	56	97	0
DR RFLP (*Alu*I)	RFLP	DR locus	48	100	0
(GTG)5 RFLP (*Hinf*I)	RFLP	(GTG)5	30	94	0
Amadio PCR	PCR	Four loci	13	74	0
IS*1081* RFLP (*Pvu*II)	RFLP	IS*1081*	12	100	0
mtp40 RFLP (*Pvu*II)	RFLP	*mtp40*	12	100	0
mbp64 RFLP (*Pvu*II)	RFLP	*mbp64*	8	100	1
katG RFLP (*Pvu*II)	RFLP	*katG*	5	100	0
16rDNA RFLP (*Pvu*II)	RFLP	16S rRNA	5	100	0
FAFLP typing (+selective primers)	PCR	EcoRI/MseI sites	ND*d*	7	0
IS*6110* inverse PCR	PCR	IS*6110*	ND*b*	6	1
IS*6110* ampliprinting	PCR	IS*6110*/MPTR	ND	39	5
FAFLP typing	PCR	EcoRI/MseI sites	ND	0	1
FAFLP typing	PCR	BamHI/PstI sites	ND	0	0

Adapted from Kremer et al. (1999, 2005b) and Supply et al. (2001, 2006)

[a]According to results obtained by laboratory 36 (Supply's team) when testing of inter- and intralaboratory reproducibility based on 30 MTBC DNA samples from Kremer's collection (Kremer et al. 1999, 2005a), as published in (de Beer et al. (2012)

[b]One locus (i.e. 2163b) out of 15 or out of 24 could not be amplified for three MTBC strains; for these cases, genotype comparison was done based on the alleles identified in the 14 or 23 other loci

insertion sequence in MTBC DNA by different researchers (Thierry et al. 1990a, b; McAdam et al. 1990; Hermans et al. 1991). The methodology was standardized in 1993 according to recommendations of Van Embden et al. (1993). Briefly, after DNA restriction with *Pvu*II endonuclease and electrophoretic separation of the obtained fragments on agarose gels, hybridization by Southern blotting is performed with an IS*6110* probe to generate strain-specific

DNA fingerprints (Fig. 3.1), which reflect the variability in the numbers and positions of IS*6110* elements in genomes of MTBC isolates. To enable comparison of DNA fingerprints obtained in different experiments and across different laboratories, a combination of external and internal molecular size standards is used to normalize the banding patterns obtained from test strains, through the use of image analysis software.

When compared to the other techniques available at the time in the first study using the Kremer's strain collection (Kremer et al. 1999, see above), this technique showed the highest discriminatory power, combined with a high reproducibility (Table 3.1). This relatively high-resolution power was reflected by detection of distinct fingerprints among isolates from the majority of unrelated TB cases in diverse studies (Van Soolingen 2001). At the same time, DNA profiles were found to be relatively stable in serial isolates from individual patients or in isolates from patients involved in a same chain of transmission (Niemann et al. 2000b; Yeh et al. 1998). On this basis, as for the other classical genotyping techniques (see below), detection of identical or distinct fingerprints among patient isolates is assumed to reflect infection by the same or different clone(s), and is used as a proxy to identify or exclude clusters of active TB transmission in patient populations, respectively. This approach was used for more than a decade in hundreds of studies, e.g., for population-based studies of the TB epidemiology at metropolitan or national levels (Small et al. 1994; Diel et al. 2002; van Soolingen et al. 1999), for investigating large outbreaks in the community (Bifani et al. 1996) or nosocomial transmission (Michele et al. 1997; Allix et al. 2004), for distinguishing exogenous reinfection from unsuccessful treatment of initial infection in relapse or persistent TB cases (Small et al. 1993b), and identification of laboratory cross-contamination events (Small et al. 1993a; CRM et al. 1999), (for review see e.g. Van Soolingen 2001).

In addition to these public health- and diagnostics-related applications, IS*6110* fingerprinting was also used for studying the population structure of MTBC strains and exploring some evolutionary questions. Indeed, limited changes in banding patterns from longitudinal isolates from individual patients – assumed to reflect clonal variants derived from a single original infecting strain – were occasionally observed, suggesting that the degree of genetic relatedness or clonal relationships among MTBC strains could be inferred by analysis of similarities of their IS*6110* RFLP patterns (Yeh et al. 1998). This approach, combined with parallel detection of similarities in other genetic markers, was used to show the existence of so-called families of related MTBC strains sharing a same common ancestor, such as the Beijing or the Haarlem families (van Soolingen et al. 1995; Kremer et al. 1999, 2004; Sola et al. 2001b; Supply et al. 2001). Mapping and comparison of the IS*6110* insertion sites in MTBC genomes, as done by sequence analysis of flanking regions, was also used for identifying the genetic relatedness of clinical strains and for phylogenetic reconstruction of the evolution of some lineages (Warren et al. 2000, 2004; Fomukong et al. 1997; Kurepina et al. 1998.

Although it has been extensively used for multiple applications, IS*6110* RFLP suffered from significant limitations. It could only be used as a retrospective tool for epidemiological analyses because it is slow and labor intensive. Despite standardization, the obtained banding patterns, which can be relatively complex, remain difficult to compare. As a result, efforts to develop and maintain an internationally centralized database of profiles (Suffys et al. 2000) were largely unsuccessful. In addition, a fraction of the MTBC strain population – such as many *M. bovis* strains and some sublineages of *M. tuberculosis* – contain a low number of copies or are devoid of IS*6110* elements, which thus reduces the resolution power of IS*6110* RFLP in these strain groups (Fomukong et al. 1994; Kremer et al. 1999; Cowan et al. 2002; Warren et al. 2004). All these limitations stimulated the search for alternative, PCR-based genotyping methods.

3.2.3 First- and Second-Generation Spoligotyping

Spoligotyping was the first PCR-based genotyping method that was used at large scale for the MTBC (Kamerbeek et al. 1997). This method targets the so-called direct repeat locus (Hermans et al. 1991) present in virtually all MTBC strains. This locus is a member of the notorious, large bacterial family of clustered, regularly interspaced short palindromic repeat (CRISPR) loci (Makarova et al. 2011). The locus contains a series of different spacer sequences, variable among strains, separated by conserved repeated motives, called direct repeats. The set of spacers present in an isolate is amplified by PCR with a primer pair against the direct repeats flanking each spacer. The obtained amplicons are then hybridized on a reference set of 43 spacers – originating from *M. tuberculosis* H37Rv and *M. bovis* BCG – immobilized on a membrane, to reveal the presence or absence of each reference spacer in the tested strain. The corresponding presence/absence signals can then be encoded as a portable 43-bit (i.e. 1/0, or black/white) barcode (Fig. 3.1). Through an ingenious design of an ad hoc cassette system, 43 strains can be analyzed at a time (Barnes and Cave 2003; Kamerbeek et al. 1997).

Classical spoligotyping is relatively cheap, but require significant hands-on time due to the hybridization steps. In addition, the quality of the spoligotyping membranes is variable, because of contaminations due to carry over or switches of oligonucleotide probes during production, which can lead to systematic errors in spoligotyping patterns. Weak hybridization signals are also difficult to interpret, which can lead to ambiguous reading of spoligotypes of related strains. Careful quality control of the membranes, best performed with a set of strains representing the phylogenetic diversity of the MTBC, is therefore recommended. In order to help overcoming some of these problems, Cowan and Colleagues (2004) transposed the principle of interrogating the reference spacer sequences onto the Luminex multianalyte profiling system, which enables faster and more automated analysis. Such a system has been adapted later to additionally include parallel interrogation of common rifampicin resistance-associated mutations in the *rpoB* hotspot region. Commercially available kits (Beamedex) were developed to facilitate use (Gomgnimbou et al. 2012).

Despite its advantages, spoligotyping does not provide resolution at the strain level, but rather provides information at sub-lineage or lineage levels based on conservation of some typical spacer signatures (Kremer et al. 1999; Brosch et al. 2002; Brudey et al. 2006). As an extreme example, most strains of the Beijing lineage, a major MTBC branch with a wide geographic distribution, share a single, dominant spoligotype composed of the nine last spacers (Kremer et al. 2004). As an attempt to improve the resolution power, second-generation of spoligotyping was introduced by including additional sets of 25–51 novel reference spacers, originating from other MTBC strains or *M. canettii* (van der Zanden et al. 2002; Brudey et al. 2004). However, although resolution of Beijing strains and strains of other branches of the MTBC (i.e. *M. bovis*, *M. microti*, *M. caprae* and *M. africanum*) was somewhat enhanced, no significant improvement was found for other *M. tuberculosis* strains (van der Zanden et al. 2002; Kremer et al. 2005a; Brudey et al. 2004). As a result, no increase in discriminatory power relative to first-generation spoligotyping was detected upon analysis of the Kremer's collection (Table 3.1, and Kremer et al. 2005a), and as a consequence the use of second-generation spoligotyping has been very limited.

Because of its limited resolution power, spoligotyping cannot be used as a sole method to ascertain epidemiological links between TB cases (Barnes and Cave 2003). Instead, particular spacer combinations, termed prototypes, recurrently occurring in isolates from diverse geographic origins were detected and systematically classified as (potential, see below) signatures of MTBC lineages and sublineages, to constitute a large, global spoligotype database, called SpolDB (Brudey et al. 2006; Sola et al. 2001a; Filliol et al. 2003), which later evolved into the SITVIT database (Demay et al. 2012). Detection and analysis of such prototypes and

their variants was used in multiple studies – and is still widely used nowadays (e.g. Nogueira et al. 2016; Balcells et al. 2015; Uzoewulu et al. 2016; Mbugi et al. 2016) – to explore the structure of MTBC strain populations circulating in numerous countries over the five continents. As of October 2016, the accessible database included more than 7000 spoligotypes from more than 50,000 isolates from more than 100 countries. Such compilation of spoligotyping data contributed to provide first insights into the global phylogeographic diversity of MTBC strains (Brudey et al. 2006; Filliol et al. 2003; Demay et al. 2012).

It has to be noted, however, that the information for phylogenetic classification and inference conveyed by spoligotyping is only an approximation. In contrast to more canonical phylogenetic markers such as SNPs (further discussed below), and as could be expected for a mono-locus marker of this kind, spoligotyping has been shown to exhibit a significant level of homoplasy and convergent evolution that blurs its phylogenetic signal (Comas et al. 2009). As a prominent example, certain *M. tuberculosis* strains from European and Asian countries that displayed the prototypic nine-spacer signature of Beijing spoligotype (part of lineage 2, according to Gagneux's classification (Coscolla and Gagneux 2014)) were demonstrated to be actually part of a distinct main MTBC lineage, namely lineage 3 (alias Delhi/CAS by spoligotyping), on the basis of concordant results based on SNPs, genomic deletions and 24-locus MIRU-VNTR typing. This misleading pseudo-Beijing spoligotype signature resulted from an evolutionary independent deletion from the RD207 deletion in the direct repeat region, associated with the usual Beijing spoligotype (Fenner et al. 2011).

3.2.4 MIRU-VNTR Typing

Because of the limited resolution of spoligotyping mentioned above, PCR-based molecular epidemiological typing of MTBC strains is mostly performed by using MIRU-VNTR typing. This method interrogates multiple genomic loci, mostly in intergenic regions, which contain variable numbers of DNA tandem repeats. Repeat unit sizes range from about 50 to 110 bp; the loci thus show a similar structure as minisatellite regions in higher eukaryote genomes (Supply et al. 2000). MIRU-VNTR typing shares the common principles of multilocus VNTR analysis (MLVA) (Keim et al. 1999) techniques, also widely used for typing and molecular surveillance of other bacteria (Nadon et al. 2013). This typing is performed by amplification of the target regions, using primers against the unique flanking regions, followed by sizing of the amplicons. As the lengths of the repeat units in the different markers are known, the numbers of repeat units in each target can be simply deduced from the amplicon sizes, based on the use of a standard allele calling table (www.miru-vntrplus.org; see background/protocols). This results in a numerical type code (24-number, when the standard format is used, see below), which enables easy comparison of results between experiments and laboratories and facilitates the establishment of standard web-based analysis and nomenclature systems such as www.miru-vntrplus.org (Fig. 3.1).

Following initial identification of different VNTR and MIRU loci in MTBC genomes (Frothingham 1995; Supply et al. 1997; Zhang and Young 1994), different numbers and sets of VNTR loci have been proposed and used (Magdalena et al. 1998; Le Fleche et al. 2002; Smittipat et al. 2005; Smittipat and Palittapongarnpim 2000; Skuce et al. 2002; Roring et al. 2002; Mazars et al. 2001; Supply et al. 2000; Warren et al. 2004; Gutierrez et al. 2006; Frothingham and Meeker-O'Connell 1998). To facilitate its uniform use for global molecular surveillance, a 24-locus format, including a discriminatory subset of 15 loci with the highest evolutionary rates, has been internationally standardized in 2006 (Supply et al. 2006). This standard set of loci was selected based on the analysis of a global sample of 824 isolates selected to test for technical robustness, reproducibility, and appropriate balance between the power to distinguish strains from the main MTBC lineages and clonal stability in different

sets of longitudinal isolates from chronically infected patients or from confirmed human transmission chains, as well as among single colonies from individually cultured strains (Supply et al. 2006).

MIRU-VNTR typing has been adapted to different platforms for sizing of the PCR fragments with different throughputs. Although sizing of individual PCR fragments has also been described based on non-denaturing high performance liquid chromatography (Evans et al. 2004), the most commonly used platforms rely on electrophoretic separation. Using agarose gels, PCRs and sizing of the amplicons are performed individually for each locus (Mazars et al. 2001; Frothingham and Meeker-O'Connell 1998). Although simple and usable at low throughput, this method becomes rapidly cumbersome upon increased fluxes of strain samples. Single DNA dye-based multicapillary systems, such as the QiaXcel platform, offer an intermediate throughput, as they enable relatively quick and semi-automated size analysis of amplicons from each individually amplified marker (Gauthier et al. 2014; Matsumoto et al. 2013). This method recently proved reproducible and accurate, except for some alleles with large numbers of repeat (Nikolayevskyy et al. 2016). Commercially available kits (Genoscreen) are expected to further facilitate implementation and robust use. The highest throughput are achieved by using automated DNA analyzers with up to 96 capillaries and multiplexed PCRs based on multidye chemistry, with semi-automated size analysis and allele calling (Cowan et al. 2005; Allix et al. 2004; Supply et al. 2006). Commercially available kits (Genoscreen) adapted for Applied Biosystems platforms include ready-to-go mixes (and controls), for amplification of the 24 standard loci in six quadruplex PCRs (Supply et al. 2014) or of the four hypervariable loci (Allix-Beguec et al. 2014; see below) in a single quadruplex PCR assay. Calibration kits including allelic ladders are also available for initial calibration of allele calling, to correct deviations between the expected sizes based on the allele sequence lengths and the actually observed sizes after an electrophoretic run, which vary according to each marker, the alleles of each marker, and individual DNA analyzers (Allix et al. 2004; Supply et al. 2006). These kits also include bioinformatics modules as well as validation DNA samples, covering wide ranges of allelic variations across the loci, for internal quality assurance of the correct implementation of the whole process, from PCR to final interpretation. Together with the systematic approach used for the definition and international standardization of the assay, the availability of these tools for internal quality assurance combined with the regular external quality assurance rounds (see below) of typing organized at worldwide level make MIRU-VNTR typing a leading example for harmonization of a MLVA method for international surveillance of pathogens, fitting the recommendations of international public health institutions for standardization of such methods (Nadon et al. 2013).

While MIRU-VNTR typing with agarose gels or automated DNA analyzers using in-house PCR conditions can also be highly reproducible (Cowan et al. 2002, 2012; Savine et al. 2002; Kremer et al. 2005a; Supply et al. 2001), international proficiency studies have repeatedly shown that the use of automated DNA analysers with commercial kits provided the highest reproducibility of typing results (de Beer et al. 2012, 2014b). In these studies, one major cause of lower reproducibility in some laboratories, especially in those not using kits, was the utilization of a non-standard allele calling system. Technical optimization of 24-locus MIRU-VNTR typing compared to initial standard conditions and commercial kits has been proclaimed (de Beer et al. 2014a), but this assertion has been contradicted (Supply et al. 2014).

Before the use of WGS for such purposes, standard 15- or 24-locus MIRU-VNTR typing has been evaluated in different population-based studies for its discriminatory power versus IS*6110* RFLP and/or its capacity to identify epidemiologically relevant molecular clusters, such as defined by agreement with classical epidemiological investigation or known risk factors for TB transmission (e.g. Allix-Beguec et al. 2008a, 2008c; Alonso-Rodriguez et al. 2009; de Beer et al. 2013; Bidovec-Stojkovic et al. 2011; Roetzer et al. 2011; Oelemann et al.

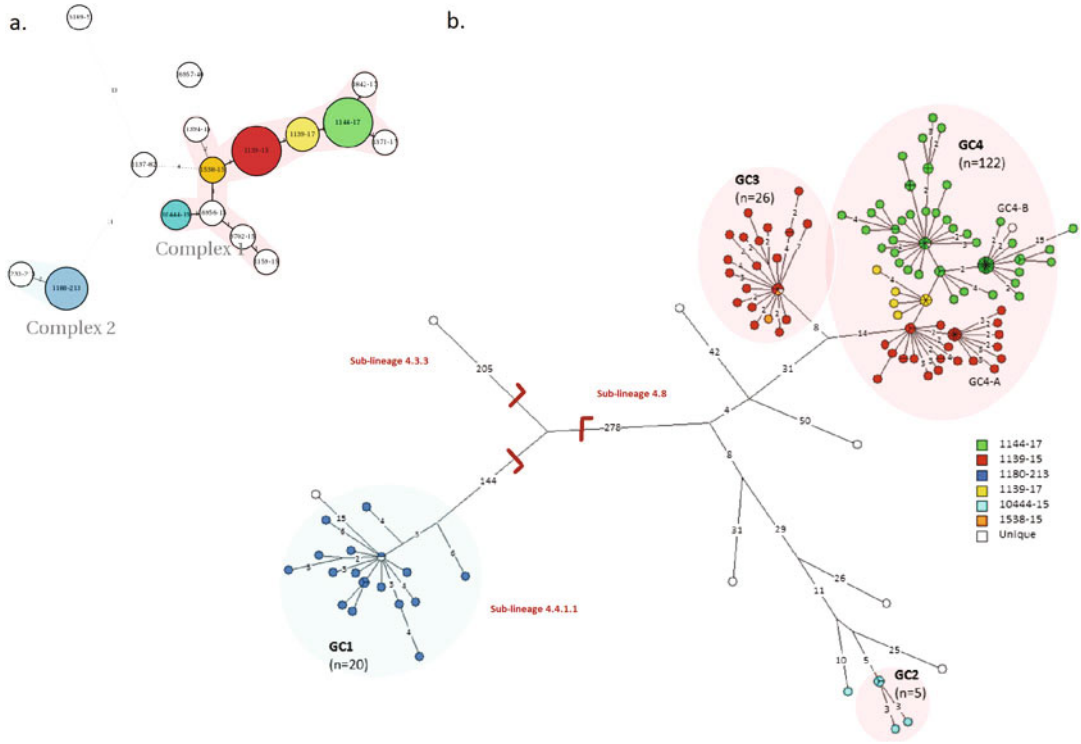

Fig. 3.2 MIRU-VNTR and whole genome sequencing-based phylogenies of isolates from all (98%) culture-positive TB patients 1992–2012 from East Greenland (n = 182). (**a**) Minimum spanning tree based on MIRU-VNTR genotypes created using MIRU-VNTRplus functionalities. The colours represent five clusters including isolates sharing a same genotype and ten unique strains grouped into two clonal complexes, as defined by groups of genotypes separated by a maximal distance of one-locus change. (**b**) Maximum parsimony tree built from 1385 single nucleotide polymorphism (SNP) positions. *Numbers* on branches indicate SNP distances. Branches with just a one SNP distance are represented without a number. Four genomic clusters (GCs) indicate isolates grouped within maximal mutual distances of 12 SNPs. Colours in the WGS-based tree correspond to clustered and unique genotypes as identified by MIRU-VNTR typing in A. Note the clearly higher resolution obtained with WGS analysis, although groupings based on genomic clusters correlate overall well with those based on MIRU-VNTR genotypes and clonal complexes (Adapted from Bjorn-Mortensen et al. 2016)

2007; van Deutekom et al. 2005). The overall picture was that standard MIRU-VNTR typing, used alone or in combination with spoligotyping, has a similar discriminatory power to IS*6110* RFLP, and is generally in similar agreement with findings from cluster investigation or known risk factors, making it a suitable method for TB surveillance systems, at least as an alternative to IS*6110* RFLP. As a result, this triggered a general shift towards the general adoption of this newer method in progressive replacement of the former gold standard, for instance for universal molecular-guided surveillance of TB in the United States by the US Center for Disease Control and Prevention (http://www.cdc.gov/tb/publications/factsheets/statistics/genotyping.htm), and the surveillance of drug resistant TB in Europe by the European Center for Disease Control (http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20742).

However, standard 15- or 24-locus MIRU-VNTR typing lacks resolution power for accurately discriminating closely related clones that can predominate in high incidence settings, such as those circulating in highly confined human communities residing in Artic regions (Bjorn-Mortensen et al. 2016; Lee et al. 2015a, b) (Fig. 3.2), or those of the Beijing strain lineage

that prevail in Russia or East Asia (Roetzer et al. 2011; Iwamoto et al. 2007; Mokrousov et al. 2008; Nikolayevskyy et al. 2006; Wada et al. 2007; Kremer et al. 2005b; Luo et al. 2014a). Therefore, based on an evaluation of more than 600 Beijing isolates from different world regions, an international consortium proposed an additional consensus set of four so-called hypervariable MIRU-VNTR loci for subtyping Beijing clonal complexes and clusters identified by standard 15- or 24-locus typing (see Allix-Beguec et al. 2014 and references therein). This set provides an additional resolution power relative to the standard loci, while the stability of the markers proved to be still sufficient to follow the clonal expansion of a Beijing strain introduced on the Gran Canaria island over more than 10 years (Allix-Beguec et al. 2014). The additional use of this four-locus set was also shown to contribute to an improvement of the epidemiologically true cluster definition for non-Beijing strains (Trovato et al. 2016). The hierarchical use of alternative subsets, partly or totally included among the standard and hypervariable loci, has been proposed for screening strain population specific to some countries with smaller numbers of markers (Luo et al. 2014a; Maeda et al. 2008; Mokrousov et al. 2008).

Consistent with its multilocus basis, and even if loci taken separately can show homoplastic evolution patterns (Supply et al. 2000; Comas et al. 2009), the phylogenetic information conveyed by 24-locus MIRU-VNTR types is less prone to homoplasy than spoligotyping (Cardoso Oelemann et al. 2011; Comas et al. 2009; Wirth et al. 2008). As such, quantitative analysis of the linkage disequilibrium between MIRU-VNTR loci has been used to provide evidence for the clonal evolution of *M. tuberculosis* (Supply et al. 2003). Standard MIRU-VNTR typing has also been used for identifying the population structure and phylogenetic lineages of isolates from multiple settings (e.g. Cardoso Oelemann et al. 2011; Allix-Beguec et al. 2008a; Aleksic et al. 2013; Biadglegne et al. 2015; Maes et al. 2008). Such typing has recently been utilized to screen a collection of 5000 isolates from 99 countries to identify main clonal complexes of the Beijing lineage, and select representatives for WGS analysis, which allowed the reconstruction of the evolutionary history of the lineage (Merker et al. 2015). Bayesian and coalescence analyses of MIRU-VNTR data have been used to detect recent expansions of strain populations of main MTBC lineages during the industrial revolution, and estimate the age of the most recent common ancestor of the complex at around 40,000 years before present (Wirth et al. 2008), an estimate that is not so far from the 70,000 year-estimate obtained by at least one leading WGS-based study published subsequently on this question (Comas et al. 2013).

Databases freely accessible via the Internet, such as MIRU-VNTR*plus* (www.miru-vntrplus.org; Allix-Beguec et al. 2008b; Weniger et al. 2010) and SIT-VIT (http://www.pasteur-guadeloupe.fr:8081/SITVIT_ONLINE/; Demay et al. 2012), have been developed, which exploit the numerical format of MIRU-VNTR types for genetic identification and analysis. In contrast to SITVIT, MIRU-VNTR*plus* is fully designed for the standard 24-locus format, and integrates multiple functionalities for easy uploading and analysis of MIRU-VNTR types (and spoligotypes) of user strains, in comparison with reference strains well characterised by classical phylogenetic markers such as genomic deletions/regions of differences (Fig. 3.1) (Allix-Beguec et al. 2008b). Also included is a universal MIRU-VNTR nomenclature system, called MLVA MtbC 15–9, reflecting the allelic combinations over the 15 most discriminatory loci and the 9 less discriminatory loci (Weniger et al. 2010). New universal genotypes can be automatically generated for user isolates, in case of allelic combinations as yet unknown. As of October 2016, more than 18,000 different types based on the sole 15-locus portion of the standard 24-locus types were registered by worldwide users (http://www.miru-vntrplus.org/MIRU/types15.faces), which gives an appreciable overview of the global genetic diversity of the MTBC.

3.2.5 Regions of Difference and Single Nucleotide Polymorphism Typing

Two other main categories of genetic markers have been used for typing MTBC strains. Because these methods have a relatively low discriminatory power, they are primarily used for phylogenetic purposes and classification of MTBC isolates into phylogenetic lineages and sub-lineages. Early studies of comparative genomics, using microarrays or bacterial artificial chromosomes, identified variable regions, called regions of differences (RDs), large sequence polymorphisms (LSPs) or indels, among the MTBC genomes which consist of genomic deletions compared to reference strains (e.g. H37Rv), spanning from several kbp to more than 10 kbp (Brosch et al. 1998, 1999; Gordon et al. 1999; Behr et al. 1999). While prophage regions or regions flanked by insertion sequence elements turned out to be phylogenetically uninformative (because of homologous recombination-mediated deletion events that are strain-specific), other deletion regions that were well conserved in different strain groups were used to reconstruct a new scenario of reductive clonal evolution for the complex. With the help of PCR using primers against flanking regions and Sanger sequencing of the amplicons, the study of the distribution of RDs and of their junction regions in different strains showed that the different MTBC members represent different clonal lineages originating from a common ancestor, which resembled more *M. tuberculosis* or *M. canettii* than *M. bovis*, thus contradicting the traditional hypothesis of zoonotic origin of human TB (Brosch et al. 2002; Mostowy et al. 2002). Using similar approaches, additional RDs were identified among *M. tuberculosis* and *M. africanum* strains and used to define a more refined phylogeny, with six main lineages of human TB-causing strains (four composed of *M. tuberculosis* strains and two of *M. africanum* or West African strains) in addition to a lineage of animal TB-causing strains (*M. bovis* and others) (note that a seventh lineage of human TB-causing strains and an additional lineage of animal TB-causing strains have now been discovered, based on WGS analysis (Coscolla and Gagneux 2014)). These phylogenetically robust markers were used to provide solid evidence for the global phylogeographic structure of the pathogen and the sympatric association between MTBC strains and their human host populations (Hirsh et al. 2004; Tsolaki et al. 2004; Gagneux et al. 2006). Other RDs specific to more regional strain groups clustered by related spoligotyping and/or MIRU-VNTR typing were discovered, which led for instance to identify a clonal sublineage (named RDRio) of *M. tuberculosis* predominating in Brazil (Cardoso Oelemann et al. 2011; Lazzarini et al. 2007), and *M. bovis* sublineages (named Af1 and Af2, respectively) epidemiologically important in West (Muller et al. 2009) and East Africa (Berg et al. 2011).

Targeted interrogation of SNP positions, pre-identified by comparison of reference genomes (Filliol et al. 2006; Gutacker et al. 2002, 2006; Homolka et al. 2012; Niemann et al. 2000a; Coll et al. 2014) has also been used to analyze the population structure of the pathogen, and has been employed in the first commercial assay using SNPs for discrimination of MTBC members among clinical isolates in the routine diagnostic laboratory (Hain MTBC Genotype assay) (Richter et al. 2003). Minimal reference sets of 13–62 phylogenetically informative SNPs were defined based on strain collections representing the main branches of the MTBC for robust assignation of isolates into different main (sub)lineages (Coll et al. 2014; Homolka et al. 2012). Based on a 62 SNP set, a convenient numerical system reflecting the hierarchical positions of branches and sub-branches within the main lineages, numbered according to Gagneux's nomenclature, was proposed by Coll et al. (2014). Similar to RDs and LSPs, such approaches define genetic clusters of strains at a broad taxonomic level. However, their usefulness for defining the true phylogeny of the MTBC is limited because of the SNP selection bias necessarily resulting from the choice of the reference genomes used to initially identify the SNPs used for typing (Smith et al.

2009). Multilocus sequence analysis (MLSA, i.e. the targeted sequencing of several genes, irrespectively of nucleotide variant positions predefined by comparison of genome references) was used as an alternative (Baker et al. 2004; Sreevatsan et al. 1997; Homolka et al. 2012). As a culminant example, Hershberg and colleagues determined the sequences of as many as 89 genes (corresponding to 66,000 bp per strain) in 108 strains, to establish the most detailed phylogeny of the MTBC available at the time (Hershberg et al. 2008). The results obtained confirmed the main lineages identified by RD/LSP-based studies, as well as the richness of non-synonymous variation in the coding sequences, also seen by early comparison of the genome sequences of *M. tuberculosis* H37RV and CDC1551 (Fleischmann et al. 2002), and taken as evidence for reduced purifying selection acting on the organisms. However, such large efforts of targeted sequencing were re-considered with the advent of next-generation sequencing opening the way to genome-wide analyses.

3.3 Whole-Genome-Based Typing

As outlined above, classical genotyping of MTBC bacteria can help to discern recent chains of transmission and address clinical questions, e.g. by distinguishing re-infection/mixed infections from cases of treatment failure or relapse (Barnes and Cave 2003). Genotyping can focus classical contact tracing investigations by local health authorities on patients related to ongoing or emerging MTBC outbreaks (i.e. increased incidence of TB cases infected with bacteria exhibiting identical DNA-fingerprints, 'clustered' cases). As such, it has been used as a tool to evaluate TB control programs by pointing out possible risk factors for recent transmission events (Barnes and Cave 2003). However, even the gold-standard method for classical MTBC genotyping, 24-locus MIRU-VNTR typing, interrogates just a selected set of highly polymorphic regions that only serves as surrogate marker for the entire genomic profiles of clinical isolates and their relationships. Furthermore, classical genotyping provides no information on possible mutations in specific genes such as drug resistance-associated genes, thus, clearly limiting its clinical usability.

In contrast, WGS provides the ultimate, most comprehensive source of information on the genome content of a given clinical isolate, which can thus be used to most precisely define its set of variable sequence positions, e.g. when compared to the sequence of a reference strain (e.g. H37Rv). This variation can then be interrogated for numerous questions, ranging from resistance predictions for virtually all resistance markers (resistome) (see Chap. 12) (Pankhurst et al. 2016; Walker et al. 2015; Niemann and Supply 2014) to high-resolution definition of clusters of TB transmission (see Chap. 4). Such higher resolution frequently allows to further discriminate strains grouped in clusters by classical genotyping, thus allowing more precise resolution of outbreaks, particularly in places or conditions where the genetic diversity of strains is low (Fig. 3.2). In addition, data from WGS analysis can also allow to estimate temporal scales of TB epidemics after calculation of genome evolution rates and dating of ancestries of the studied isolate populations, based on reconstruction of SNP-based phylogenies (Roetzer et al. 2013; Walker et al. 2013a, 2014; Niemann and Supply 2014; Merker et al. 2015; Luo et al. 2014b; Cohen et al. 2015; Eldholm et al. 2015). Finally, WGS can be further exploited for a wide range of evolutionary and comparative genomic applications as discussed in several other chapters of this book.

Important steps in WGS analysis for molecular epidemiological investigations are as follows: (1) DNA extraction and sequencing library preparation, (2) massively parallel sequencing, analysis of the sequencing results which can be done by (3) mapping and SNP analysis or by (4) genome wide multi-locus sequence typing (MLST), and (5) interpretation and practical im-

plications of the results. In the following, we will discuss currently applied methodologies for each step, as well as benefits and some limitations for epidemiological inferences.

3.3.1 DNA Extraction and Library Preparation

In the early days of NGS, a main limitation for WGS of MTBC bacteria was the time needed to obtain sufficient DNA quantities for subsequent library preparation protocols, which required several weeks of culturing starting from clinical specimens due to the slow growth of the organisms. A breakthrough came from newly developed library preparation methods that require as little as 1–10 ng of DNA (e.g., Illumina Nextera XT [1 ng], New England BioLabs NEBNext Ultra [5–1000 ng], and Bioo Scientific NEXTflex ChIP-Seq [1–10 ng]), resulting in considerable reduction in the turn-around time and overall costs. Using the Illumina Nextera XT kit, renders it possible to perform WGS from a starting material containing only a limited amount of bacterial cells, such as early positive MGIT960 cultures. When tested prospectively by an international group of clinical laboratories for complete mycobacterial diagnostics, such an approach allowed to generate and interpret WGS results within 9 days (IQR 6–10) at an average cost of £481 per culture positive specimen including staff time, consumables and equipment, which was 7% lower than the total cost of classical, routine diagnostic workflow (Pankhurst et al. 2016).

However, even amounts as low as a few nanograms of DNA are still well above the amount obtainable from 10^4 mycobacteria in 1 ml of clinical sample, corresponding to the classical threshold of detection by microscopic examination. Contamination with of human DNA and from bacterial flora present in clinical specimens represents an additional obstacle, as it minimizes the representation – and thus the coverage – of specific, mycobacterial DNA in the analysis. Direct, culture-independent WGS analysis for diagnostic detection of MTBC bacilli in sputum samples, after application of a differential lysis step to reduce contamination by human DNA, has been published (Doughty et al. 2014). However, the coverage depths of reference genome of 0.002X to 0.7X that were reached were still far insufficient for reliable SNP analysis and epidemiological inferences. More specific enrichment strategies aiming to increase the representation of mycobacterial DNA in the sequencing step, such as that using hybridization on large arrays of biotinylated RNA baits spanning the length of the H37Rv *M. tuberculosis* reference genome, are in principle more promising (Brown et al. 2015). However, as such enrichment is relatively laborious, requires specific additional equipment and increases consumable costs to more than 350$ per sample, cost effectiveness and workability in a routine workflow remains to be demonstrated.

Another crucial part of the WGS workflow resides in the quality control of the resulting sequencing libraries, which includes measuring the concentration and size distribution of the obtained DNA fragments, by using capillary electrophoresis systems (e.g. Fragment Analyzer, from Advanced Analytical, or QIAxcel, from Qiagen) or microfluidics chips (e.g. Bioanalyzer, from Agilent). Some WGS strategies require tightly controlled size distributions, which might necessitate gel excision.

3.3.2 Massively Parallel Sequencing

The relative performances of the different NGS technologies and platforms currently available, with respect to DNA sequence output, cost per Gb obtained, read lengths, or rates and types of sequencing errors, are now well established (Loman et al. 2012; Harris et al. 2013; Junemann et al. 2013; Miyamoto et al. 2014; Laver et al. 2015). As a probable reflection, most studies published hitherto have used short read sequencing, mostly based on Illumina technology with 50–300 bp read length ranges (Niemann et al.

2009; Roetzer et al. 2013; Gardy et al. 2011; Walker et al. 2013a, 2014; Smit et al. 2014; Kato-Maeda et al. 2013; Bjorn-Mortensen et al. 2016). Only a few initial studies used GS FLX (Roche/454) platforms with slightly longer reads (Schurch et al. 2010; Sandegren et al. 2011), and a single study used a SOLiD 5500XL instrument (Applied Biosystems) (Colangeli et al. 2014). Such short read sequencing approaches are well adapted to classical SNP analysis after mapping of reads on a reference genome, as explained in the next section.

Longer reads are to be preferred for de novo assembly approaches, especially when completely circularized genomes and generation of new genome references are aimed for. In such cases, single molecule, real-time (SMRT) sequencing (Pacific Bioscience) (Chin et al. 2013), producing read lengths that can currently exceed 60 kbp, can be used. Because of this and because errors are randomly distributed over sequence positions of individual reads – although error levels are higher than for Illumina sequencing at individual read levels -, this technology can generate highly accurate consensus sequence assemblies with sufficient coverage depths (Miyamoto et al. 2014; Pouseele and Supply 2015). However, given its relatively high cost, this method is currently more appropriate for research-oriented questions, rather than for routine WGS-based typing. Oxford Nanopore sequencing, also based on sequencing of native DNA in real time conditions- but going through nanopore channels in a miniaturized, so-called MiniION sequencing device – generates even longer read lengths of up to 200 kbp. However, these advantages come with the current limitation of a higher error rate at the sequencing read level (Laver et al. 2015; Lu et al. 2016). Therefore, proofs of principle of this technology have been established for other applications, such as rapid identification of drug resistance based on presence/absence of specific genes for other bacterial pathogens (Bradley et al. 2015; Lu et al. 2016).

3.3.3 Analysis of Sequencing Results by Reference Mapping and SNP Analysis

The method of choice for analysis of WGS data for molecular epidemiological investigations and for many evolutionary studies of MTBC strains has been based on mapping of sequencing reads and detection of sequence variants relative to a given reference genome sequence (Roetzer et al. 2013; Walker et al. 2013a, 2014; Niemann and Supply 2014; Merker et al. 2015; Luo et al. 2014b; Cohen et al. 2015; Eldholm et al. 2015; Supply et al. 2013; Biek et al. 2012; Gardy et al. 2011; Comas et al. 2010, 2013; Stucki et al. 2015, 2016b; Boritsch et al. 2016a). The highly clonal and conserved backbone of the MTBC genomes (Boritsch et al. 2016b) favors this approach of comparison to a common reference over approaches of de novo assemblies that are more computationally demanding. The first historically obtained and well-annotated complete sequence, the *M. tuberculosis* H37Rv genome (Cole et al. 1998), and to a much lesser extent the *M. tuberculosis* CDC 1551 genome (Fleischmann et al. 2002), are usually used as references. With such mapping approaches, and subsequent exclusion of repetitive regions (see below), sequence variation can be captured for 92–95% of the reference genome.

Despite their wide use and advantages for a variety of research questions, such mapping approaches also have limitations. As they are restricted by design to the actual reference sequence used, DNA regions not present in the reference sequence are not included in the analysis, and rare genomic re-arrangements are not easily detected. However, some of these limitations are minor given the overall conservation of the MTBC genomes and might be further reduced by the use of a chimaeric reference composed of non-redundant parts of different genomes. Furthermore, results from mapping-based SNP analyses are specifically dependent upon a defined set of genomes from a particular

study, as variant positions depend on the strain set considered. Thus, adding new isolates later on, e.g. in the case of prospective molecular surveillance, usually implies a new comparative SNP analysis. In addition, because SNP positions depend on the reference used, results obtained by using distinct references can not be directly compared.

More importantly, key parameters for mapping and SNP analysis have not been harmonized across different laboratories and a standardized analysis pipeline has not been established so far. Published studies have used quite variable analysis parameters using different pipelines combining diverse sets of software solutions for read mapping and variant detection. This lack of standardization has sometimes resulted in very divergent results e.g. demonstrated by the relatively high amount of SNPs among outbreak isolates in some studies (e.g. Gardy et al. 2011) compared to other studies (e.g. Roetzer et al. 2013; Smit et al. 2014; Walker et al. 2013a). These factors render direct comparisons of findings from different laboratories difficult, and hamper the development of a standardized application for international WGS-based molecular TB surveillance and global strain comparisons. Concerted efforts are ongoing in order to address this problem and construct web-based platforms for exchanging and analyzing WGS data on more normalized bases (Starks et al. 2015).

3.3.4 Analysis of Sequencing Results by Genome-Wide MLST

Limitations indicated above for classical reference mapping and SNP analysis can be largely overcome by using a gene-by-gene approach (Maiden et al. 2013), extending the multilocus sequence typing (MLST) concept, initially developed for six or seven housekeeping genes, to a genome-wide typing scheme (Maiden et al. 1998). Such an approach transfers genome wide SNP diversity into a more easily processed and standardizable allele numbering system. The analysis is based on a predefined scheme of genetic loci and their alleles present in strains of a given species. Core genome MLST (cgMLST) and/or pan-genome MLST (pgMLST) schemes are usually defined on the basis of the set of loci/genes shared (core genome) by a well-chosen set of representative reference strain genomes, or on the accumulation of non-redundant genes detected in any of such reference strain genome (pan-genome). A locus can be any genomic region; it usually represents a coding sequence (CDS), but can also be created for other genomic features such as promoter regions or spoligotyping spacers. For each locus, the specific sequence variant represents an allele that is assigned a unique, arbitrary allele number. The combination of allele numbers of the complete scheme obtained for a given strain defines a sequence type (ST), which constitutes a portable and universal genetic identifier of strains and clonal complexes. Allele sequences are optimally stored in central, publicly accessible web-based databases such as BIGSdb (Jolley and Maiden 2010; Maiden et al. 2013).

Genetic relationships based on allelic profiles can be easily calculated based on the comparison of allelic identity or differences, regardless of the number of distinct variant positions between two alleles of a gene. As a secondary benefit, genome wide MLST can also buffer against potential skewing effects caused by importation events of multiple polymorphisms into a genomic region resulting from horizontal gene transfer events from a distant strain, while it preserves the phylogenetic signal conveyed by vertically transmitted genetic variation in the other target genes. Although this is important for bacterial species prone to frequent inter-strain recombination, this aspect is of lower importance for classical strains of tubercle bacilli. Indeed, there is a well-documented overall consensus that the classical MTBC shows a highly clonal evolution with

no significant ongoing inter-strain recombination (Hirsh et al. 2004; Supply et al. 2003, 2013; Brosch et al. 2002, 2016b; Wirth et al. 2008; Sreevatsan et al. 1997; Mostowy et al. 2002; Baker et al. 2004; Comas et al. 2013), despite few studies suggesting some genetic exchanges among such strains (Namouchi et al. 2012; Liu et al. 2006). Only *M. canettii* strains, representing rare human TB isolates mostly obtained from patients from the Djibouti region in East Africa, show abundant traces of inter-strain recombination as inferred from numerous mosaic structures in their genomes (see Chap. 2) (Gutierrez et al. 2005; Supply et al. 2013; Blouin et al. 2014; Boritsch et al. 2014, 2016b), probably mediated by distributive conjugal transfer (Boritsch et al. 2016b; Mortimer and Pepperell 2014; Derbyshire and Gray 2014).

A first cgMLST scheme has been developed for the MTBC (Kohl et al. 2014). It is available through the publicly available BIGSdb platform, and the commercial software packages SeqSphere+ (Ridom) and BioNumerics (Applied Maths). The suggested cgMLST scheme encompasses 3257 CDS loci uniformly detected in any of a set of seven representative MTBC genomes, discarding genes associated with repetitive regions (Comas et al. 2010) due to the current limitations in correctly resolving these regions with short read sequencing. This CDS set represents about 80% of the coding capacity of *M. tuberculosis* H37Rv. When evaluated for resolving a set of 26 isolates from a selected MTBC outbreak, this cgMLST scheme fully agreed with a classical reference mapping and SNP analysis approach in reconstructing the overall phylogenetic relationships of the isolates. As expected, cgMLST had a slightly lower discriminatory power in this first evaluation, due to SNPs detected by the classical approach in intergenic regions and in genes not part of the selected core set. This proof of principle study thus supports the notion that a genome-wide MLST approach is well suited for continuous, self-consistent definition of the genetic relationships and broad genomic clusters in (growing) strain collections and generating universal strain identifiers such as needed in prospective surveillance. Where needed, the reference mapping method can then be applied for isolates in cgMLST-defined clusters to maximize genomic resolution for fine-tuned outbreak investigations and possibly better resolving individual transmission chains. A similar two-step strategy, combining in this case pgMLST with reference mapping-based SNP analysis, has been recently used to resolve two outbreaks caused by *Staphylococcus aureus* at ultra-high resolution in a neonatal context (Roisin et al. 2016).

Of note, a pgMLST scheme available through the BioNumerics software suite has also been developed for the MTBC, based on the cumulative compilation of genes from additional selected MTBC reference genomes and *M. canettii* genomes to represent ancestral lineages of the tubercle bacilli (Supply et al. 2013). Although perhaps less directly useful for standardized typing, this extended scheme provides more resolution power and allows for more comprehensive analysis of the gene content in the genomes of the diverse tubercle bacilli.

3.4 Molecular Cluster Definition by Classical Typing and WGS

The relevance of epidemiological inference based on a particular molecular typing technique depends on a properly established significance and definition of molecular clusters based on that method. Such significance and definition evidently depend on the actual mutation rates of the markers involved- and thus, in the case of the WGS-based typing, on the genome mutation rate – during infection and human to human transmission, which is also correlated with the discriminatory power of the markers. Various estimates of mutation rates for markers of classical typing methods and for the *M. tuberculosis* genome are listed in Table 3.2.

While spoligotypes are known to almost never change among longitudinal isolates from transmission chains or individual patients, their variability among unrelated isolates within a lineage is also limited, reflecting an overall low mutation rate (Kremer et al. 1999). Thus, molecular clus-

Table 3.2 Estimates of mutation rates and levels of homoplasy associated with different typing methods of MTBC isolates, and usual range of applications

MTBC markers	Mutation rate estimates	Homoplasy index	Applications
Spoligotype	2.0×10^{-2}–9.0×10^{-2} per year[a]	Yes/relatively high	Preliminary screen of genetic diversity, excluding possible laboratory contaminations
Regions of difference/targeted interrogation of phylogenetic SNPs	Not determined	No	(Sub-)lineage classification
IS*6110* sequence	0.0135 changes per copy per year[b]	Yes/low to moderate (genotypes with low numbers of IS6110 copies)	(Local) molecular epidemiological investigations, differentiation between relapse/re-infection
	0.0161 changes per copy per year[c]		
MIRU-VNTR loci	7.0×10^{-4}–1.5×10^{-2} per locus per year[a]	Yes/moderate	(Global) molecular epidemiological investigations, differentiation between relapse/re-infection, screening for potential TB transmission clusters, screening for lineage identification
	3.3×10^{-4}–9.8×10^{-3} per locus per year[d]		
	10^{-4} per locus per year[e]		
	2.5×10^{-3}–2.6×10^{-2} per locus per year[f]		
	1.2×10^{-3}–2.6×10^{-3} per locus per year[g]		
WGS/genome wide SNP analysis	0.24–0.34 SNPs per genome per year[h]	No/very low	Molecular epidemiological investigations, differentiation between relapse/re-infection, high resolution outbreak investigation, drug resistance prediction, robust phylogenetic analysis
	0.26–0.66 SNPs per genome per year[i]		
	0.3–0.5 SNPs per genome per year[j,k]		
	0.3–0.7 SNPs per genome per year[l]		
	0.93–1.56 SNPs per genome per year[g]		
	0.13–0.27 SNPs per genome per year[m]		
	0.0073–0.013 SNPs per genome per year[n] (long-term rate)		

References are as follows: [a]Reyes and Tanaka (2010), [b]Tanaka and Rosenberg (2001), [c]Rosenberg et al. (2003), [d]Ragheb et al. (2013), [e]Wirth et al. (2008), [f]Aandahl et al. (2012), [g]Eldholm et al. (2016), [h]Eldholm et al. (2015), [i]Roetzer et al. (2013), [j]Ford et al. (2011), [k]Ford et al. (2013), [l]Walker et al. (2013a),[m]Bos et al. (2014), and [n]Comas et al. (2013)

ters based on spoligotyping alone are generally considered of weak significance, if any, for inferring TB transmission (Barnes and Cave 2003).

The mutation rates of standard and hypervariable MIRU-VNTR genotypes have been evaluated in more details by measuring the actual rates of genotypic changes in large sets of longitudinal isolates from chronically infected patients, from well documented transmission chains, and from a large outbreak spanning more than 10 years originating from the introduction of a single clone in the Gran Canaria island (Allix-Beguec et al. 2014; Savine et al. 2002; Supply et al. 2006; Velji et al. 2009). Genotypes were fully conserved in almost all cases within clonal sets, and only rare changes – always limited to single-locus variations – indicating development of a clonal subpopulation e.g. during longitudinal treatment or delayed diagnosis (Al-Hajoj et al. 2010; Garcia de Viedma et al. 2005) – were detected among clonally allied isolates. Therefore, cluster analysis should be based on strict identity of MIRU-VNTR alleles over the 24 standard loci (and over the 4 hypervariable loci, if used in addition), as a proxy to ongoing TB transmission in populations (Allix-Beguec et al. 2014; Savine et al. 2002; Supply et al. 2006). As an additional consequence of such limited variation among isolates descending from a common clonal ancestor over periods extending to at least 10 years, differences in at least two MIRU-VNTR loci or simultaneous detection of double alleles in two loci or more in an isolate can be safely concluded to reflect infections from independent strains/sources or mixed infection with two independent strains, respectively (Allix et al. 2004; Garcia de Viedma et al. 2005; Shamputa et al. 2006; Merker et al. 2013; Supply et al. 2006). Using this rule of minimal genotypic changes, MIRU-VNTR typing based on standard and hypervariable loci is thus best used as a reliable exclusion method (Supply et al. 2006; Allix-Beguec et al. 2014).

Based on these measures of actual genotypic changes in clonal isolates during human infection and human to human transmission, and taking into account the differences in rates of changes between loci as inferred by relative frequencies of single locus variations in a large population-based strain sample, a mean mutation rate of 10^{-4} per standard MIRU-VNTR locus per year was estimated, which was further supported by posterior Bayesian analysis of different datasets (Wirth et al. 2008; Merker et al. 2015). This results in a combined mutation rate of 2.4 10^{-3} per year per 24-locus genotype, which translates in a mean of 0.05 pairwise (single-locus) changes expected in sets of isolates descending from a same original clone over 10 years. This prediction matches well the range of 4–12 single-locus changes actually observed among sets of 86 or 92 clonal isolates from two studied large clonal outbreaks, each spanning 15 years (Allix-Beguec et al. 2014; Roetzer et al. 2013). Hence, this mutation rate is well supported by actual variation detected during longitudinal transmission, and appears more likely than divergent estimations obtained through modelization or other means (Aandahl et al. 2012; Grant et al. 2008; Ragheb et al. 2013; Reyes and Tanaka 2010), as these have not taken into account e.g. significant differences in mutation rates between loci or using much less defined/smaller sets of isolates for their estimation (see discussion in Supply et al. 2011a, b).

However, as a consequence of the relatively low mutation rate of MIRU-VNTR genotypes (see above), clustering based on MIRU-VNTR typing without epidemiological or clinical independent support should be considered only as a potential indication for TB transmission or a true relapse episode. Typically, only about half of the molecular clusters detected based on 24-locus MIRU-VNTR typing in populations of low incidence settings are confirmed by epidemiological investigation (de Beer et al. 2013), and a recent WGS-based study found that about half of the 24-locus- and spoligotyping-based clusters including patient isolates in Switzerland were further distinguished by WGS with a separation of >12 SNPs between the isolates involved (Stucki et al. 2016a). In high incidence settings where more homogenous strain populations predominate, such as those of the Beijing lineage or those circulating in Artic regions (Fig. 3.2), the proportion of isolates clustered by MIRU-VNTR typing and/or other classical methods but

distinguished by WGS may even be higher (Luo et al. 2014b; Lee et al. 2015a, b; Bjorn-Mortensen et al. 2016). Hence, patient isolates clustered by MIRU-VNTR typing that are not independently supported by classical epidemiological investigations should be prioritized for WGS analysis.

Like for the definition of a threshold for MIRU-VNTR typing, the above threshold of genomic distances of >12 SNPs used to exclude recent transmission based on WGS was defined based on pilot studies systematically measuring genome-wide mutation rates in large sets of longitudinal isolates obtained from chronically infected patients, household outbreaks, confirmed transmission chains in a large community outbreak, or other sets of epidemiologically linked isolates confirmed by contact investigation in low TB incidence settings in Western Europe. Based on the observed distributions of SNPs and time spans separating clonal isolates, mutation rates of 0.3–0.5 SNP per genome per year have been estimated (Bryant et al. 2013; Roetzer et al. 2013; Walker et al. 2013a), and maximal thresholds ranging from 5 (stringent, most likely) to 12 SNPs (more putative) have been defined, beyond which recent transmission is considered unlikely (Walker et al. 2013a) (Fig. 3.3). This range of short-term mutation rates has been confirmed by WGS analyses of other large longitudinal outbreaks in medium (Argentina, Eldholm et al. 2015) and high TB incidence settings (South Africa, Cohen et al. 2015), and is supported by data from experimental infection of simian models (Ford et al. 2011). This genome wide mutation range, higher by two orders of magnitudes relative to the rate of 24-locus MIRU-VNTR genotypes, thus illustrates well the higher resolution power of WGS relative to classical typing.

However, this limit of maximum 5–12 SNPs to support or exclude recent transmission based on WGS may not be universally applicable, and needs to take into account possible specificities of the studied settings, strain and/or patient populations. For instance, some studies found as many as 14 SNPs to 24 SNPs between isolates from (presumed) directly epidemiologically linked patients, reflecting possibly extensive microevolution within some patients with potential differential transmission of subpopulations to secondary cases (Perez-Lago et al. 2014; Luo et al. 2014b). Conversely, Lee and Colleagues found that in the Canadian Artic, where the genetic diversity of *M. tuberculosis* strains is limited, a significant number of isolates from patient pairs residing in different villages in the Canadian Arctic and classified as improbable transmission cases were separated by as few as 2–12 SNPs (Lee et al. 2015a, b). On this basis, they defined a very stringent threshold of 0 to maximum 1 SNP to define transmission clusters within a village and compared pairwise SNP differences between isolates within villages and between villages to infer most frequent routes of transmission (Lee et al. 2015a). In a similar area in East Greenland with remote and poorly populated human settlements and a very low diversity of *M. tuberculosis* strains, a group of researchers used a different algorithm taking into account both SNP distance and year of isolation to identify subgroups with active transmission and guide targeted public health interventions (Bjorn-Mortensen et al. 2016).

Similar strategies of WGS cluster analysis might be relevant in other world regions dominated by homogenous MTBC strain populations such as those of particular branches of the Beijing lineage in China and Eurasia (Luo et al. 2014b; Merker et al. 2015). Importantly, as mutation rate estimates and SNP thresholds for defining epidemiologically relevant clusters also depend on the parameters and methods employed for WGS data analysis, they should be established and validated using generally acknowledged algorithms.

3.5 Practical Implications: What Method to Use in What Context and for What Question

The mutation rates linked to the resolution power of the typing methods determine their potential range of applications (Table 3.2). However, the choice of what typing method to use also depends on several additional parameters, including the levels of available resources, the

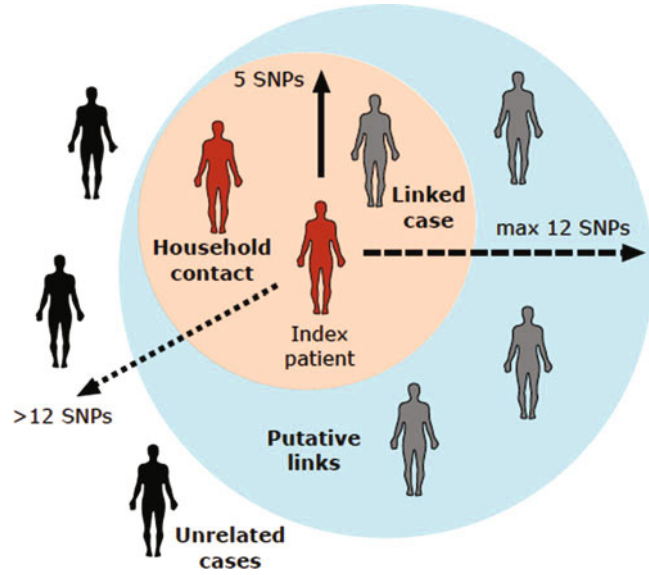

Fig. 3.3 Definition of molecular epidemiological clusters using WGS analysis based on proposed cut-off values from a reference study in the UK (Walker et al. 2013a). This study systematically measured levels of genomic variation within relatively large sets of longitudinal isolates from individual TB patients from household outbreaks with various time spans between the isolates in the various sets. Thresholds of SNP distances for defining linked and likely linked cases around an index patient were set based on the observation of maximal variations of 5 SNPs (*orange circle*) for most cases and of 12 SNPs (*blue circle*) for few exceptional cases. Cases with SNP distances exceeding the latter threshold are considered unlinked. These thresholds may need to be evaluated and possibly adapted depending upon local epidemiological contexts and local strain populations

main question(s) to address, and the epidemiological or strain population context that is considered.

Estimation of typing costs per strain is not straightforward, as these costs are influenced by the number of isolates to be analyzed in a given study or per time period, and by whether a method is already established in the laboratory or outsourcing of the typing work is preferred. Nevertheless, as an example for 68 outbreak isolates screened out of 1,642 isolates, Stucki and colleagues reported estimated costs per isolate of 3US$ for a SNP assay established in-house (requiring preliminary WGS analysis), and 26US$ for spoligotyping, 49US$ for standard MIRU-VNTR typing and 330 US$ for WGS, based on commercial services as of August 2014 (Stucki et al. 2015). Costs of the three latter methods when performed in-house are expected to be about two to three times lower. This relative hierarchy remains largely valid as of early 2017.

The choice of a method also depends on the technical expertise that is available. WGS is again the most demanding approach in this respect, given the expertise in bioinformatics needed for relevant analysis and interpretation of NGS data. If adequate resources are available, WGS will undoubtedly be the method of choice if most precise/comprehensive information is desired for public health-related, clinical and/or basic research questions. For instance, WGS can uniquely provide potential information on the directionality of transmission, as a reflection of the vertical inheritance and accumulation of mutations without any detectable reversions. Thanks to its highest resolution power, it can also uniquely help to determine so-called super-spreaders (corresponding to particularly contagious cases) in the population investigated. The presence of super-spreader cases is suggested by detection of star-like patterns in WGS-based trees, – by using algorithms allowing for potential hypothetical nodes such

as minimum spanning and maximum likelihood trees or median joining networks – indicating multiple secondary cases originating from a same common node/index case (Walker et al. 2013a, b, 2014; Gardy et al. 2011; Perez-Lago et al. 2014; Stucki et al. 2015). However, inferences on super-spreaders can be blurred by potential co-existence and differential transmission of clonal subpopulations in some patients, particularly in cases of persistent disease with levels of diversity within a patient that can be as high as that observed between patients along a transmission chain (Perez-Lago et al. 2014).

WGS should also be prioritized for best delineation and estimation of starting dates of outbreaks, based on estimates of genome evolution rates (Cohen et al. 2015; Eldholm et al. 2015; Roetzer et al. 2013; Bryant et al. 2013). The same choice should be made if drug resistance-related questions are also to be addressed, for instance in order to reconstruct the sequential accumulation of mutations leading to epidemic MDR- and XDR-TB strains (Cohen et al. 2015; Eldholm et al. 2015). If performed rapidly from newly positive cultures or potentially even directly from clinical samples (Pankhurst et al. 2016; Walker et al. 2015; Brown et al. 2015), WGS analysis can also be of direct diagnostic benefit. Further aspects on drug resistance and genomic epidemiology will be discussed in the following chapters.

Although less informative and precise than WGS, standard MIRU-VNTR typing can be used for initial screening or more routine/focal purposes, such as (approximate) strain lineage identification, ruling out/in of reinfection or laboratory contamination episodes, or exclusion or inclusion of from/in potential TB transmission clusters. As indicated in the previous section, isolates that are clustered based on standard (and hypervariable) MIRU-VNTR markers can then be prioritized for subsequent WGS analysis, especially in the absence of supportive epidemiological data or when the genetic diversity of the strain population is known to be restricted (as in Artic regions, see above).

Such a strategy combining different methods in a staggered manner can also be an efficient, cost-effective option, especially when dealing with large strain collections from multiple sources and global questions, such as the evolution of major lineages of the MTBC (Luo et al. 2015; Merker et al. 2015; Stucki et al. 2016b). In these studies, MIRU-VNTR, spoligotyping, SNP and/or RD-based typing were first used as screens in large datasets to identify isolates of the respective lineages, and to select representatives of main clonal complexes for subsequent WGS analysis and reconstruction of the overall genetic structure and bacterial population dynamics.

Finally, typing using reference RDs or using well defined reference phylogenetic SNPs can be used alone if the priority is to analyze potential links between phenotypes and main strain lineages in a given strain set, or to know the broad distribution of main MTBC (sub)lineages in a geographic region (Hirsh et al. 2004; Reiling et al. 2013). Spoligotyping can also be used alone as a relatively easy guide for the latter purpose, although the precision for (sub)lineage identification will be substantially lower because of the levels of homoplasy seen with this typing method (see Sect. 2.3). The latter method can also be utilized for possible exclusion of laboratory contamination episodes.

3.6 Concluding Remarks

WGS undoubtedly represents the culminant point in the evolution of the MTBC genotyping methods. Its intrinsically superior resolution power combined with decreasing prices of NGS and the many additional advantages described above argue for its increased adoption for TB surveillance and control, particularly in higher resourced settings. However, methodologies of WGS-based typing are still evolving and standardization is still awaited for facilitating its use in global TB surveillance. Even more so than for classical typing, some important analysis parameters in-

fluencing epidemiological inferences are particularly sensitive to the algorithms used for identification of genome variation and the particular epidemiological and biological contexts, which results in additional complexity in the interpretation. In addition to pointing out such complexities, this chapter described potential solutions and options for adapting WGS analysis depending on local contexts and facilitating standardization, such as development of cgMLST for universal description of the genomic diversity and more comprehensive understanding of the global transmission of MTBC strains.

Acknowledgements Parts of this chapter result from research supported by a grant from the European Union's Seventh Framework Program (FP7/2007-2013) under Grant Agreement N° 278864 in the framework of the PathoNGenTrace project.

References

Aandahl RZ, Reyes JF, Sisson SA, Tanaka MM (2012) A model-based Bayesian estimation of the rate of evolution of VNTR loci in Mycobacterium tuberculosis. PLoS Comput Biol 8(6):e1002573. doi:10.1371/journal.pcbi.1002573

Ahmed N, Alam M, Abdul Majeed A, Asad Rahman S, Cataldi A, Cousins D, Hasnain SE (2003) Genome sequence based, comparative analysis of the fluorescent amplified fragment length polymorphisms (FAFLP) of tubercle bacilli from seals provides molecular evidence for a new species within the Mycobacterium tuberculosis complex. Infect Genet Evol 2(3):193–199

Aleksic E, Merker M, Cox H, Reiher B, Sekawi Z, Hearps AC, Ryan CE, Lee AV, Goursaud R, Malau C, O'Connor J, Cherry CL, Niemann S, Crowe SM (2013) First molecular epidemiology study of Mycobacterium tuberculosis in Kiribati. PLoS One 8(1):e55423. doi:10.1371/journal.pone.0055423

Al-Hajoj SA, Akkerman O, Parwati I, Al-Gamdi S, Rahim Z, van Soolingen D, van Ingen J, Supply P, Van der Zanden AG (2010) Microevolution of Mycobacterium tuberculosis in a tuberculosis patient. J Clin Microbiol 48(10):3813–3816. JCM.00556-10 [pii]. doi:10.1128/JCM.00556-10

Allix C, Supply P, Fauville-Dufaux M (2004) Utility of fast mycobacterial interspersed repetitive unit-variable number tandem repeat genotyping in clinical mycobacteriological analysis. Clin Infect Dis 39(6):783–789

Allix-Beguec C, Fauville-Dufaux M, Supply P (2008a) Three-year population-based evaluation of standardized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing of Mycobacterium tuberculosis. J Clin Microbiol 46(4):1398–1406. doi:10.1128/JCM.02089-07

Allix-Beguec C, Harmsen D, Weniger T, Supply P, Niemann S (2008b) Evaluation and strategy for use of MIRU-VNTRplus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of Mycobacterium tuberculosis complex isolates. J Clin Microbiol 46(8):2692–2699. JCM.00540-08 [pii]. doi:10.1128/JCM.00540-08

Allix-Beguec C, Supply P, Wanlin M, Bifani P, Fauville-Dufaux M (2008c) Standardised PCR-based molecular epidemiology of tuberculosis. Eur Respir J 31(5):1077–1084

Allix-Beguec C, Wahl C, Hanekom M, Nikolayevskyy V, Drobniewski F, Maeda S, Campos-Herrero I, Mokrousov I, Niemann S, Kontsevaya I, Rastogi N, Samper S, Sng LH, Warren RM, Supply P (2014) Proposal of a consensus set of hypervariable mycobacterial interspersed repetitive-unit-variable-number tandem-repeat loci for subtyping of Mycobacterium tuberculosis Beijing isolates. J Clin Microbiol 52(1):164–172. doi:10.1128/JCM.02519-13

Alonso-Rodriguez N, Martinez-Lirola M, Sanchez ML, Herranz M, Penafiel T, Bonillo Mdel C, Gonzalez-Rivera M, Martinez J, Cabezas T, Diez-Garcia LF, Bouza E, Garcia de Viedma D (2009) Prospective universal application of mycobacterial interspersed repetitive-unit-variable-number tandem-repeat genotyping to characterize Mycobacterium tuberculosis isolates for fast identification of clustered and orphan cases. J Clin Microbiol 47(7):2026–2032. JCM.02308-08 [pii]. doi:10.1128/JCM.02308-08

Arbeit RD, Arthur M, Dunn R, Kim C, Selander RK, Goldstein R (1990) Resolution of recent evolutionary divergence among Escherichia coli from related lineages: the application of pulsed field electrophoresis to molecular epidemiology. J Infect Dis 161(2):230–235

Arnold A, Witney AA, Vergnano S, Roche A, Cosgrove CA, Houston A, Gould KA, Hinds J, Riley P, Macallan D, Butcher PD, Harrison TS (2016) XDR-TB transmission in London: case management and contact tracing investigation assisted by early whole genome sequencing. J Infect. doi:10.1016/j.jinf.2016.04.037

Baker L, Brown T, Maiden MC, Drobniewski F (2004) Silent nucleotide polymorphisms and a phylogeny for Mycobacterium tuberculosis. Emerg Infect Dis 10(9):1568–1577

Balcells ME, Garcia P, Meza P, Pena C, Cifuentes M, Couvin D, Rastogi N (2015) A first insight on the population structure of Mycobacterium tuberculosis complex as studied by spoligotyping and MIRU-VNTRs in Santiago, Chile. PLoS One 10(2):e0118007. doi:10.1371/journal.pone.0118007

Barnes PF, Cave MD (2003) Molecular epidemiology of tuberculosis. N Engl J Med 349(12):1149–1156. doi:10.1056/NEJMra021964

Bates JH, Mitchison DA (1969) Geographic distribution of bacteriophage types of Mycobacterium tuberculosis. Am Rev Respir Dis 100(2):189–193. doi:10.1164/arrd.1969.100.2.189

de Beer JL, Kremer K, Kodmon C, Supply P, van Soolingen D (2012) First worldwide proficiency study on variable-number tandem-repeat typing of Mycobacterium tuberculosis complex strains. J Clin Microbiol 50(3):662–669. doi:10.1128/JCM.00607-11

de Beer JL, van Ingen J, de Vries G, Erkens C, Sebek M, Mulder A, Sloot R, van den Brandt AM, Enaimi M, Kremer K, Supply P, van Soolingen D (2013) Comparative study of IS6110 restriction fragment length polymorphism and variable-number tandem-repeat typing of Mycobacterium tuberculosis isolates in the Netherlands, based on a 5-year nationwide survey. J Clin Microbiol 51(4):1193–1198. doi:10.1128/JCM.03061-12

de Beer JL, Akkerman OW, Schurch AC, Mulder A, van der Werf TS, van der Zanden AG, van Ingen J, van Soolingen D (2014a) Optimization of standard in-house 24-locus variable-number tandem-repeat typing for Mycobacterium tuberculosis and its direct application to clinical material. J Clin Microbiol 52(5):1338–1342. doi:10.1128/JCM.03436-13

de Beer JL, Kodmon C, van Ingen J, Supply P, van Soolingen D (2014b) Second worldwide proficiency study on variable number of tandem repeats typing of Mycobacterium tuberculosis complex. Int J Tuberc Lung Dis 18(5):594–600. doi:10.5588/ijtld.13.0531

Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, Small PM (1999) Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science 284(5419):1520–1523

Berg S, Garcia-Pelayo MC, Muller B, Hailu E, Asiimwe B, Kremer K, Dale J, Boniotti MB, Rodriguez S, Hilty M, Rigouts L, Firdessa R, Machado A, Mucavele C, Ngandolo BN, Bruchfeld J, Boschiroli L, Muller A, Sahraoui N, Pacciarini M, Cadmus S, Joloba M, van Soolingen D, Michel AL, Djonne B, Aranaz A, Zinsstag J, van Helden P, Portaels F, Kazwala R, Kallenius G, Hewinson RG, Aseffa A, Gordon SV, Smith NH (2011) African 2, a clonal complex of Mycobacterium bovis epidemiologically important in East Africa. J Bacteriol 193(3):670–678. doi:10.1128/JB.00750-10

Biadglegne F, Merker M, Sack U, Rodloff AC, Niemann S (2015) Tuberculous lymphadenitis in Ethiopia predominantly caused by strains belonging to the Delhi/CAS lineage and newly identified Ethiopian clades of the Mycobacterium tuberculosis complex. PLoS One 10(9):e0137865. doi:10.1371/journal.pone.0137865

Bidovec-Stojkovic U, Zolnir-Dovc M, Supply P (2011) One year nationwide evaluation of 24-locus MIRU-VNTR genotyping on Slovenian Mycobacterium tuberculosis isolates. Respir Med 105(Suppl 1):S67–S73. doi:10.1016/S0954-6111(11)70014-2

Biek R, O'Hare A, Wright D, Mallon T, McCormick C, Orton RJ, McDowell S, Trewby H, Skuce RA, Kao RR (2012) Whole genome sequencing reveals local transmission patterns of Mycobacterium bovis in sympatric cattle and badger populations. PLoS Pathog 8(11):e1003008. doi:10.1371/journal.ppat.1003008

Bifani PJ, Plikaytis BB, Kapur V, Stockbauer K, Pan X, Lutfey ML, Moghazeh SL, Eisner W, Daniel TM, Kaplan MH, Crawford JT, Musser JM, Kreiswirth BN (1996) Origin and interstate spread of a New York City multidrug-resistant Mycobacterium tuberculosis clone family. JAMA 275(6):452–457

Bjorn-Mortensen K, Soborg B, Koch A, Ladefoged K, Merker M, Lillebaek T, Andersen AB, Niemann S, Kohl TA (2016) Tracing Mycobacterium tuberculosis transmission by whole genome sequencing in a high incidence setting: a retrospective population-based study in East Greenland. Sci Rep 6:33180. doi:10.1038/srep33180

Blouin Y, Cazajous G, Dehan C, Soler C, Vong R, Hassan MO, Hauck Y, Boulais C, Andriamanantena D, Martinaud C, Martin E, Pourcel C, Vergnaud G (2014) Progenitor "Mycobacterium canettii" clone responsible for lymph node tuberculosis epidemic, Djibouti. Emerg Infect Dis 20(1):21–28. doi:10.3201/eid2001.130652

Boritsch EC, Supply P, Honore N, Seemann T, Stinear TP, Brosch R (2014) A glimpse into the past and predictions for the future: the molecular evolution of the tuberculosis agent. Mol Microbiol 93(5):835–852. doi:10.1111/mmi.12720

Boritsch EC, Frigui W, Cascioferro A, Malaga W, Etienne G, Laval F, Pawlik A, Le Chevalier F, Orgeur M, Ma L, Bouchier C, Stinear TP, Supply P, Majlessi L, Daffé M, Guilhot C, Brosch R (2016a) pks5-recombination-mediated surface remodelling in Mycobacterium tuberculosis emergence. Nat Microbiol 1:15019

Boritsch EC, Khanna V, Pawlik A, Honore N, Navas VH, Ma L, Bouchier C, Seemann T, Supply P, Stinear TP, Brosch R (2016b) Key experimental evidence of chromosomal DNA transfer among selected tuberculosis-causing mycobacteria. Proc Natl Acad Sci U S A 113(35):9876–9881. doi:10.1073/pnas.1604921113

Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, Forrest SA, Bryant JM, Harris SR, Schuenemann VJ, Campbell TJ, Majander K, Wilbur AK, Guichon RA, Wolfe Steadman DL, Cook DC, Niemann S, Behr MA, Zumarraga M, Bastida R, Huson D, Nieselt K, Young D, Parkhill J, Buikstra JE, Gagneux S, Stone AC, Krause J (2014) Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature 514(7523):494–497

Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, Earle S, Pankhurst LJ, Anson L, de Cesare M, Piazza P, Votintseva AA, Golubchik T, Wilson DJ, Wyllie DH, Diel R, Niemann S, Feuerriegel S, Kohl TA, Ismail N, Omar SV, Smith EG, Buck D, McVean G, Walker AS, Peto TE, Crook DW, Iqbal Z (2015) Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis. Nat Commun 6:10063. doi:10.1038/ncomms10063

Brosch R, Gordon SV, Billault A, Garnier T, Eiglmeier K, Soravito C, Barrell BG, Cole ST (1998) Use of a Mycobacterium tuberculosis H37Rv bacterial artificial chromosome library for genome mapping, sequencing, and comparative genomics. Infect Immun 66(5):2221–2229

Brosch R, Philipp WJ, Stavropoulos E, Colston MJ, Cole ST, Gordon SV (1999) Genomic analysis reveals variation between Mycobacterium tuberculosis H37Rv and the attenuated M. tuberculosis H37Ra strain. Infect Immun 67(11): 5768–5774

Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, Garnier T, Gutierrez C, Hewinson G, Kremer K, Parsons LM, Pym AS, Samper S, van Soolingen D, Cole ST (2002) A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci U S A 99(6):3684–3689. doi:10.1073/pnas.052548299

Brown AC, Bryant JM, Einer-Jensen K, Holdstock J, Houniet DT, Chan JZ, Depledge DP, Nikolayevskyy V, Broda A, Stone MJ, Christiansen MT, Williams R, McAndrew MB, Tutill H, Brown J, Melzer M, Rosmarin C, McHugh TD, Shorten RJ, Drobniewski F, Speight G, Breuer J (2015) Rapid whole-genome sequencing of mycobacterium tuberculosis isolates directly from clinical samples. J Clin Microbiol 53(7):2230–2237. doi:10.1128/JCM.00486-15

Brudey K, Gutierrez MC, Vincent V, Parsons LM, Salfinger M, Rastogi N, Sola C (2004) Mycobacterium africanum genotyping using novel spacer oligonucleotides in the direct repeat locus. J Clin Microbiol 42(11):5053–5057. doi:10.1128/JCM.42.11.5053-5057.2004

Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajoj SA, Allix C, Aristimuno L, Arora J, Baumanis V, Binder L, Cafrune P, Cataldi A, Cheong S, Diel R, Ellermeier C, Evans JT, Fauville-Dufaux M, Ferdinand S, Garcia de Viedma D, Garzelli C, Gazzola L, Gomes HM, Guttierez MC, Hawkey PM, van Helden PD, Kadival GV, Kreiswirth BN, Kremer K, Kubin M, Kulkarni SP, Liens B, Lillebaek T, Ho ML, Martin C, Mokrousov I, Narvskaia O, Ngeow YF, Naumann L, Niemann S, Parwati I, Rahim Z, Rasolofo-Razanamparany V, Rasolonavalona T, Rossetti ML, Rusch-Gerdes S, Sajduda A, Samper S, Shemyakin IG, Singh UB, Somoskovi A, Skuce RA, van Soolingen D, Streicher EM, Suffys PN, Tortoli E, Tracevska T, Vincent V, Victor TC, Warren RM, Yap SF, Zaman K, Portaels F, Rastogi N, Sola C (2006) Mycobacterium tuberculosis complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol 6:23. 1471-2180-6-23 [pii]. doi:10.1186/1471-2180-6-23

Bryant JM, Schurch AC, van Deutekom H, Harris SR, de Beer JL, de Jager V, Kremer K, van Hijum SA, Siezen RJ, Borgdorff M, Bentley SD, Parkhill J, van Soolingen D (2013) Inferring patient to patient transmission of Mycobacterium tuberculosis from whole genome sequencing data. BMC Infect Dis 13(1):110. doi:10.1186/1471-2334-13-110

Cardoso Oelemann M, Gomes HM, Willery E, Possuelo L, Batista Lima KV, Allix-Beguec C, Locht C, Goguet de la Salmoniere YO, Gutierrez MC, Suffys P, Supply P (2011) The forest behind the tree: phylogenetic exploration of a dominant Mycobacterium tuberculosis strain lineage from a high tuberculosis burden country. PLoS One 6(3):e18256. doi:10.1371/journal.pone.0018256

Cegielski JP, Dalton T, Yagui M, Wattanaamornkiet W, Volchenkov GV, Via LE, Van Der Walt M, Tupasi T, Smith SE, Odendaal R, Leimane V, Kvasnovsky C, Kuznetsova T, Kurbatova E, Kummik T, Kuksa L, Kliiman K, Kiryanova EV, Kim H, Kim CK, Kazennyy BY, Jou R, Huang WL, Ershova J, Erokhin VV, Diem L, Contreras C, Cho SN, Chernousova LN, Chen MP, Caoili JC, Bayona J, Akksilp S (2014) Extensive drug resistance acquired during treatment of multidrug-resistant tuberculosis. Clin Infect Dis 59(8):1049–1063. doi:10.1093/cid/ciu572

Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10(6):563–569. doi:10.1038/nmeth.2474

Cohen KA, Abeel T, Manson McGuire A, Desjardins CA, Munsamy V, Shea TP, Walker BJ, Bantubani N, Almeida DV, Alvarado L, Chapman SB, Mvelase NR, Duffy EY, Fitzgerald MG, Govender P, Gujja S, Hamilton S, Howarth C, Larimer JD, Maharaj K, Pearson MD, Priest ME, Zeng Q, Padayatchi N, Grosset J, Young SK, Wortman J, Mlisana KP, O'Donnell MR, Birren BW, Bishai WR, Pym AS, Earl AM (2015) Evolution of extensively drug-resistant tuberculosis over four decades: whole genome sequencing and dating analysis of Mycobacterium tuberculosis isolates from KwaZulu-Natal. PLoS Med 12(9):e1001880. doi:10.1371/journal.pmed.1001880

Coitinho C, Greif G, Robello C, Laserra P, Willery E, Supply P (2013) Rapidly progressing tuberculosis outbreak in a very low risk group. Eur Respir J. doi:10.1183/09031936.00150413

Colangeli R, Arcus VL, Cursons RT, Ruthe A, Karalus N, Coley K, Manning SD, Kim S, Marchiano E, Alland D (2014) Whole genome sequencing of Mycobacterium tuberculosis reveals slow growth and low mutation rates during latent infections in humans. PLoS One 9(3):e91024. doi:10.1371/journal.pone.0091024

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 393(6685):537–544. doi:10.1038/31159

Coll F, McNerney R, Guerra-Assuncao JA, Glynn JR, Perdigao J, Viveiros M, Portugal I, Pain A, Martin N, Clark TG (2014) A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. Nat Commun 5:4812. doi:10.1038/ncomms5812

Collins DM, De Lisle GW (1984) DNA restriction endonuclease analysis of Mycobacterium tuberculosis and Mycobacterium bovis BCG. J Gen Microbiol 130(4):1019–1021. doi:10.1099/00221287-130-4-1019

Comas I, Homolka S, Niemann S, Gagneux S (2009) Genotyping of genetically monomorphic bacteria: DNA sequencing in Mycobacterium tuberculosis highlights the limitations of current methodologies. PLoS One 4(11):e7815. doi:10.1371/journal.pone.0007815

Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, Kremer K, Ernst JD, Gagneux S (2010) Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved. Nat Genet 42(6):498–503. doi:10.1038/ng.590

Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, Parkhill J, Malla B, Berg S, Thwaites G, Yeboah-Manu D, Bothamley G, Mei J, Wei L, Bentley S, Harris SR, Niemann S, Diel R, Aseffa A, Gao Q, Young D, Gagneux S (2013) Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. Nat Genet 45(10):1176–1182. doi:10.1038/ng.2744

Comas I, Hailu E, Kiros T, Bekele S, Mekonnen W, Gumi B, Tschopp R, Ameni G, Hewinson RG, Robertson BD, Goig GA, Stucki D, Gagneux S, Aseffa A, Young D, Berg S (2015) Population genomics of Mycobacterium tuberculosis in Ethiopia contradicts the virgin soil hypothesis for human tuberculosis in Sub-Saharan Africa. Curr Biol 25(24):3260–3266. doi:10.1016/j.cub.2015.10.061

Cooke GS, Beaton RK, Lessells RJ, John L, Ashworth S, Kon OM, Williams OM, Supply P, Moodley P, Pym AS (2011) International spread of MDR TB from Tugela Ferry, South Africa. Emerg Infect Dis 17(11):2035–2037. doi:10.3201/eid1711.110291

Corbett EL, Watt CJ, Walker N, Maher D, Williams BG, Raviglione MC, Dye C (2003) The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. Arch Intern Med 163(9):1009–1021. doi:10.1001/archinte.163.9.1009

Coscolla M, Gagneux S (2014) Consequences of genomic diversity in Mycobacterium tuberculosis. Semin Immunol 26(6):431–444. doi:10.1016/j.smim.2014.09.012

Coscolla M, Barry PM, Oeltmann JE, Koshinsky H, Shaw T, Cilnis M, Posey J, Rose J, Weber T, Fofanov VY, Gagneux S, Kato-Maeda M, Metcalfe JZ (2015) Genomic epidemiology of multidrug-resistant Mycobacterium tuberculosis during transcontinental spread. J Infect Dis 212(2):302–310. doi:10.1093/infdis/jiv025

Cowan LS, Mosher L, Diem L, Massey JP, Crawford JT (2002) Variable-number tandem repeat typing of Mycobacterium tuberculosis isolates with low copy numbers of IS6110 by using mycobacterial interspersed repetitive units. J Clin Microbiol 40(5):1592–1602

Cowan LS, Diem L, Brake MC, Crawford JT (2004) Transfer of a Mycobacterium tuberculosis genotyping method, Spoligotyping, from a reverse line-blot hybridization, membrane-based assay to the Luminex multianalyte profiling system. J Clin Microbiol 42(1):474–477

Cowan LS, Diem L, Monson T, Wand P, Temporado D, Oemig TV, Crawford JT (2005) Evaluation of a two-step approach for large-scale, prospective genotyping of Mycobacterium tuberculosis isolates in the United States. J Clin Microbiol 43(2):688–695. 43/2/688 [pii]. doi:10.1128/JCM.43.2.688-695.2005

Cowan LS, Hooks DP, Christianson S, Sharma MK, Alexander DC, Guthrie JL, Jamieson FB, Supply P, Allix-Beguec C, Cruz L, Desmond E, Kramer R, Lugo S, Rudrik J (2012) Evaluation of mycobacterial interspersed repetitive-unit-variable-number tandem-repeat genotyping as performed in laboratories in Canada, France, and the United States. J Clin Microbiol 50(5):1830–1831.; author reply 1832. doi:10.1128/JCM.00168-12

Cox HS, Kubica T, Doshetov D, Kebede Y, Rusch-Gerdess S, Niemann S (2005) The Beijing genotype and drug resistant tuberculosis in the Aral Sea region of Central Asia. Respir Res 6:134. doi:10.1186/1465-9921-6-134

Cox HS, Sibilia K, Feuerriegel S, Kalon S, Polonsky J, Khamraev AK, Rusch-Gerdes S, Mills C, Niemann S (2008) Emergence of extensive drug resistance during treatment for multidrug-resistant tuberculosis. N Engl J Med 359(22):2398–2400. doi:10.1056/NEJMc0805644

Crawford JT (1993) Applications of molecular methods to epidemiology of tuberculosis. Res Microbiol 144(2):111–116

CRM D, Soini H, Roscanni GC, Jaques M, Villares MC, Musser JM (1999) Extensive cross-contamination of specimens with Mycobacterium tuberculosis in a reference laboratory. J Clin Microbiol 37(4):916–919

Daley CL, Horsburgh CR Jr (2014) Editorial commentary: treatment for multidrug-resistant tuberculosis: it's worse than we thought! Clin Infect Dis 59(8):1064–1065. doi:10.1093/cid/ciu578

Demay C, Liens B, Burguiere T, Hill V, Couvin D, Millet J, Mokrousov I, Sola C, Zozio T, Rastogi N (2012) SITVITWEB – a publicly available international multimarker database for studying Mycobacterium tuberculosis genetic diversity and molecular epidemiology. Infect Genet Evol 12(4):755–766. doi:10.1016/j.meegid.2012.02.004

Derbyshire KM, Gray TA (2014) Distributive conjugal transfer: new insights into horizontal gene transfer and genetic exchange in mycobacteria. Microbiol Spectrum 2(1):MGM2-0022-2013. doi:10.1128/microbiolspec.MGM2-0022-2013

Diel R, Schneider S, Meywald-Walter K, Ruf CM, Rusch-Gerdes S, Niemann S (2002) Epidemiology of tuberculosis in Hamburg, Germany: long-term population-based analysis applying classical and molecular epidemiological techniques. J Clin Microbiol 40(2):532–539

Diel R, Vandeputte J, de Vries G, Stillo J, Wanlin M, Nienhaus A (2014) Costs of tuberculosis disease in the European Union: a systematic analysis

and cost calculation. Eur Respir J 43(2):554–565. doi:10.1183/09031936.00079413

Doughty EL, Sergeant MJ, Adetifa I, Antonio M, Pallen MJ (2014) Culture-independent detection and characterisation of Mycobacterium tuberculosis and M. africanum in sputum samples using shotgun metagenomics on a benchtop sequencer. Peer J 2:e585. doi:10.7717/peerj.585

Dye C, Williams BG (2010) The population dynamics and control of tuberculosis. Science 328(5980):856–861. doi:10.1126/science.1185449

Eldholm V, Monteserin J, Rieux A, Lopez B, Sobkowiak B, Ritacco V, Balloux F (2015) Four decades of transmission of a multidrug-resistant Mycobacterium tuberculosis outbreak strain. Nat Commun 6:7119. doi:10.1038/ncomms8119

Evans JT, Hawkey PM, Smith EG, Boese KA, Warren RE, Hong G (2004) Automated high-throughput mycobacterial interspersed repetitive unit typing of Mycobacterium tuberculosis strains by a combination of PCR and nondenaturing high-performance liquid chromatography. J Clin Microbiol 42(9):4175–4180. doi:10.1128/JCM.42.9.4175-4180.2004

Fabre M, Koeck JL, Le Fleche P, Simon F, Herve V, Vergnaud G, Pourcel C (2004) High genetic diversity revealed by variable-number tandem repeat genotyping and analysis of hsp65 gene polymorphism in a large collection of "Mycobacterium canettii" strains indicates that the M. tuberculosis complex is a recently emerged clone of "M. canettii". J Clin Microbiol 42(7):3248–3255. doi:10.1128/JCM.42.7.3248-3255.2004

Fabre M, Hauck Y, Soler C, Koeck JL, van Ingen J, van Soolingen D, Vergnaud G, Pourcel C (2010) Molecular characteristics of "Mycobacterium canettii" the smooth Mycobacterium tuberculosis bacilli. Infect Genet Evol 10(8):1165–1173. doi:10.1016/j.meegid.2010.07.016

Fenner L, Malla B, Ninet B, Dubuis O, Stucki D, Borrell S, Huna T, Bodmer T, Egger M, Gagneux S (2011) "Pseudo-Beijing": evidence for convergent evolution in the direct repeat region of Mycobacterium tuberculosis. PLoS One 6(9):e24737. doi:10.1371/journal.pone.0024737

Filliol I, Driscoll JR, van Soolingen D, Kreiswirth BN, Kremer K, Valetudie G, Dang DA, Barlow R, Banerjee D, Bifani PJ, Brudey K, Cataldi A, Cooksey RC, Cousins DV, Dale JW, Dellagostin OA, Drobniewski F, Engelmann G, Ferdinand S, Gascoyne-Binzi D, Gordon M, Gutierrez MC, Haas WH, Heersma H, Kassa-Kelembho E, Ho ML, Makristathis A, Mammina C, Martin G, Mostrom P, Mokrousov I, Narbonne V, Narvskaya O, Nastasi A, Niobe-Eyangoh SN, Pape JW, Rasolofo-Razanamparany V, Ridell M, Rossetti ML, Stauffer F, Suffys PN, Takiff H, Texier-Maugein J, Vincent V, de Waard JH, Sola C, Rastogi N (2003) Snapshot of moving and expanding clones of Mycobacterium tuberculosis and their global distribution assessed by spoligotyping in an international study. J Clin Microbiol 41(5):1963–1970

Filliol I, Motiwala AS, Cavatore M, Qi W, Hazbon MH, Bobadilla del Valle M, Fyfe J, Garcia-Garcia L, Rastogi N, Sola C, Zozio T, Guerrero MI, Leon CI, Crabtree J, Angiuoli S, Eisenach KD, Durmaz R, Joloba ML, Rendon A, Sifuentes-Osornio J, Ponce de Leon A, Cave MD, Fleischmann R, Whittam TS, Alland D (2006) Global phylogeny of Mycobacterium tuberculosis based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. J Bacteriol 188(2):759–772. doi:10.1128/JB.188.2.759-772.2006

Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, Peterson J, DeBoy R, Dodson R, Gwinn M, Haft D, Hickey E, Kolonay JF, Nelson WC, Umayam LA, Ermolaeva M, Salzberg SL, Delcher A, Utterback T, Weidman J, Khouri H, Gill J, Mikula A, Bishai W, Jacobs WR Jr, Venter JC Jr, Fraser CM (2002) Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains. J Bacteriol 184(19):5479–5490

Fomukong NG, Tang TH, Al-Maamary S, Ibrahim WA, Ramayah S, Yates M, Zainuddin ZF, Dale JW (1994) Insertion sequence typing of Mycobacterium tuberculosis: characterization of a widespread subtype with a single copy of IS6110. Tuber Lung Dis 75(6):435–440. doi:10.1016/0962-8479(94)90117-1

Fomukong N, Beggs M, el Hajj H, Templeton G, Eisenach K, Cave MD (1997) Differences in the prevalence of IS6110 insertion sites in Mycobacterium tuberculosis strains: low and high copy number of IS6110. Tuber Lung Dis 78(2):109–116

Ford CB, Lin PL, Chase MR, Shah RR, Iartchouk O, Galagan J, Mohaideen N, Ioerger TR, Sacchettini JC, Lipsitch M, Flynn JL, Fortune SM (2011) Use of whole genome sequencing to estimate the mutation rate of Mycobacterium tuberculosis during latent infection. Nat Genet 43(5):482–486

Frothingham R (1995) Differentiation of strains in Mycobacterium tuberculosis complex by DNA sequence polymorphisms, including rapid identification of M. bovis BCG. J Clin Microbiol 33(4):840–844

Frothingham R, Meeker-O'Connell WA (1998) Genetic diversity in the Mycobacterium tuberculosis complex based on variable numbers of tandem DNA repeats. Microbiology 144(Pt 5):1189–1196

Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, Nicol M, Niemann S, Kremer K, Gutierrez MC, Hilty M, Hopewell PC, Small PM (2006) Variable host-pathogen compatibility in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 103(8):2869–2873. doi:10.1073/pnas.0511240103

Gandhi NR, Nunn P, Dheda K, Schaaf HS, Zignol M, van Soolingen D, Jensen P, Bayona J (2010) Multidrug-resistant and extensively drug-resistant tuberculosis: a threat to global control of tuberculosis. Lancet 375(9728):1830–1843. doi:10.1016/S0140-6736(10)60410-2

Garcia de Viedma D, Alonso Rodriguez N, Andres S, Ruiz Serrano MJ, Bouza E (2005) Characterization of clonal complexity in tuberculosis by mycobacterial interspersed repetitive unit-variable-number tandem repeat typing. J Clin Microbiol 43(11): 5660–5664

Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodkin E, Rempel S, Moore R, Zhao Y, Holt R, Varhol R, Birol I, Lem M, Sharma MK, Elwood K, Jones SJ, Brinkman FS, Brunham RC, Tang P (2011) Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N Engl J Med 364(8):730–739. doi:10.1056/NEJMoa1003176

Gauthier M, Bidault F, Mosnier A, Bablishvili N, Tukvadze N, Somphavong S, Paboriboune P, Ocheretina O, Pape JW, Paranhos-Baccala G, Berland JL (2014) High-throughput MIRU-VNTR genotyping for Mycobacterium tuberculosis epidemiological studies. J Clin Microbiol. doi:10.1128/JCM.01611-14

Getahun H, Matteelli A, Chaisson RE, Raviglione M (2015) Latent Mycobacterium tuberculosis infection. N Engl J Med 372(22):2127–2135. doi:10.1056/NEJMra1405427

Gomgnimbou MK, Abadia E, Zhang J, Refregier G, Panaiotov S, Bachiyska E, Sola C (2012) "Spoligoriftyping," a dual-priming-oligonucleotide-based direct-hybridization assay for tuberculosis control with a multianalyte microbead-based hybridization system. J Clin Microbiol 50(10):3172–3179. doi:10.1128/JCM.00976-12

Gordon SV, Brosch R, Billault A, Garnier T, Eiglmeier K, Cole ST (1999) Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. Mol Microbiol 32(3): 643–655

Grant A, Arnold C, Thorne N, Gharbia S, Underwood A (2008) Mathematical modelling of Mycobacterium tuberculosis VNTR loci estimates a very slow mutation rate for the repeats. J Mol Evol 66(6):565–574. doi:10.1007/s00239-008-9104-6

Gutacker MM, Smoot JC, Migliaccio CA, Ricklefs SM, Hua S, Cousins DV, Graviss EA, Shashkina E, Kreiswirth BN, Musser JM (2002) Genome-wide analysis of synonymous single nucleotide polymorphisms in Mycobacterium tuberculosis complex organisms: resolution of genetic relationships among closely related microbial strains. Genetics 162(4): 1533–1543

Gutacker MM, Mathema B, Soini H, Shashkina E, Kreiswirth BN, Graviss EA, Musser JM (2006) Single-nucleotide polymorphism-based population genetic analysis of Mycobacterium tuberculosis strains from 4 geographic sites. J Infect Dis 193(1):121–128. doi:JID34975 [pii]941086/498574

Gutierrez MC, Brisse S, Brosch R, Fabre M, Omais B, Marmiesse M, Supply P, Vincent V (2005) Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. PLoS Pathog 1(1):e5

Gutierrez MC, Ahmed N, Willery E, Narayanan S, Hasnain SE, Chauhan DS, Katoch VM, Vincent V, Locht C, Supply P (2006) Predominance of ancestral lineages of Mycobacterium tuberculosis in India. Emerg Infect Dis 12(9):1367–1374

Harris SR, Torok ME, Cartwright EJ, Quail MA, Peacock SJ, Parkhill J (2013) Read and assembly metrics inconsequential for clinical utility of whole-genome sequencing in mapping outbreaks. Nat Biotechnol 31(7):592–594. doi:10.1038/nbt.2616

Hermans PW, van Soolingen D, Bik EM, de Haas PE, Dale JW, van Embden JD (1991) Insertion element IS987 from Mycobacterium bovis BCG is located in a hot-spot integration region for insertion elements in Mycobacterium tuberculosis complex strains. Infect Immun 59(8):2695–2705

Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, Roach JC, Kremer K, Petrov DA, Feldman MW, Gagneux S (2008) High functional diversity in Mycobacterium tuberculosis driven by genetic drift and human demography. PLoS Biol 6(12):e311. 08-PLBI-RA-2811 [pii]. doi:10.1371/journal.pbio.0060311

Hirsh AE, Tsolaki AG, DeRiemer K, Feldman MW, Small PM (2004) Stable association between strains of Mycobacterium tuberculosis and their human host populations. Proc Natl Acad Sci U S A 101(14): 4871–4876

Homolka S, Projahn M, Feuerriegel S, Ubben T, Diel R, Nubel U, Niemann S (2012) High resolution discrimination of clinical Mycobacterium tuberculosis complex strains based on single nucleotide polymorphisms. PLoS One 7(7):e39855. doi:10.1371/journal.pone.0039855

Iwamoto T, Yoshida S, Suzuki K, Tomita M, Fujiyama R, Tanaka N, Kawakami Y, Ito M (2007) Hypervariable loci that enhance the discriminatory ability of newly proposed 15-loci and 24-loci variable-number tandem repeat typing method on Mycobacterium tuberculosis strains predominated by the Beijing family. FEMS Microbiol Lett 270(1):67–74. doi:10.1111/j.1574-6968.2007.00658.x

Jolley KA, Maiden MC (2010) BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinforma 11:595. doi:10.1186/1471-2105-11-595

Junemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U, Kalinowski J, Mellmann A, Goesmann A, von Haeseler A, Stoye J, Harmsen D (2013) Updating benchtop sequencing performance comparison. Nat Biotechnol 31(4):294–296. doi:10.1038/nbt.2522

Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, van Embden J (1997) Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. J Clin Microbiol 35(4):907–914

Kato-Maeda M, Ho C, Passarelli B, Banaei N, Grinsdale J, Flores L, Anderson J, Murray M, Rose G, Kawamura LM, Pourmand N, Tariq MA, Gagneux S, Hopewell PC (2013) Use of whole genome sequencing to determine the microevolution of Mycobacterium tuber-

culosis during an outbreak. PLoS One 8(3):e58235. doi:10.1371/journal.pone.0058235

Keim P, Klevytska AM, Price LB, Schupp JM, Zinser G, Smith KL, Hugh-Jones ME, Okinaka R, Hill KK, Jackson PJ (1999) Molecular diversity in Bacillus anthracis. J Appl Microbiol 87(2): 215–217

Kitahara K (1973) Studies on the identification of infection source for pulmonary tuberculosis in a family using bacteriophage types as a marker. 2. Studies on the source of infection in family. Kekkaku 48(3): 61–69

Kohl TA, Diel R, Harmsen D, Rothganger J, Walter KM, Merker M, Weniger T, Niemann S (2014) Whole-genome-based Mycobacterium tuberculosis surveillance: a standardized, portable, and expandable approach. J Clin Microbiol 52(7):2479–2486. doi:10.1128/JCM.00567-14

Kremer K, van Soolingen D, Frothingham R, Haas WH, Hermans PW, Martin C, Palittapongarnpim P, Plikaytis BB, Riley LW, Yakrus MA, Musser JM, van Embden JD (1999) Comparison of methods based on different molecular epidemiological markers for typing of Mycobacterium tuberculosis complex strains: interlaboratory study of discriminatory power and reproducibility. J Clin Microbiol 37(8):2607–2618

Kremer K, Glynn JR, Lillebaek T, Niemann S, Kurepina NE, Kreiswirth BN, Bifani PJ, van Soolingen D (2004) Definition of the Beijing/W lineage of Mycobacterium tuberculosis on the basis of genetic markers. J Clin Microbiol 42(9):4040–4049. doi:10.1128/JCM.42.9.4040-4049.2004

Kremer K, Arnold C, Cataldi A, Gutierrez MC, Haas WH, Panaiotov S, Skuce RA, Supply P, van der Zanden AG, van Soolingen D (2005a) Discriminatory power and reproducibility of novel DNA typing methods for Mycobacterium tuberculosis complex strains. J Clin Microbiol 43(11):5628–5638. 3/11/5628 [pii]. doi:410.1128/JCM.43.11.5628-5638.2005

Kremer K, Au BK, Yip PC, Skuce R, Supply P, Kam KM, van Soolingen D (2005b) Use of variable-number tandem-repeat typing to differentiate Mycobacterium tuberculosis Beijing family isolates from Hong Kong and comparison with IS6110 restriction fragment length polymorphism typing and spoligotyping. J Clin Microbiol 43(1):314–320. doi:10.1128/jcm.43.1.314-320.2005

Kurepina NE, Sreevatsan S, Plikaytis BB, Bifani PJ, Connell ND, Donnelly RJ, van Sooligen D, Musser JM, Kreiswirth BN (1998) Characterization of the phylogenetic distribution and chromosomal insertion sites of five IS6110 elements in Mycobacterium tuberculosis: non-random integration in the dnaA-dnaN region. Tuber Lung Dis 79(1):31–42. doi:10.1054/tuld.1998.0003

Lafeuille E, Veziris N, Sougakoff W, Roure F, Le Du D, Dournon N, Caumes E, Jarlier V, Aubry A, Robert J, Bernard C (2016) XDR-tuberculosis in France: community transmission due to non-compliance with isolation precautions. Med Mal Infect 46(1):52–55. doi:10.1016/j.medmal.2015.12.008

Lambregts-van Weezenbeek CS, Sebek MM, van Gerven PJ, de Vries G, Verver S, Kalisvaart NA, van Soolingen D (2003) Tuberculosis contact investigation and DNA fingerprint surveillance in The Netherlands: 6 years' experience with nation-wide cluster feedback and cluster monitoring. Int J Tuberc Lung Dis 7(12 Suppl 3):S463–S470

Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, Studholme DJ (2015) Assessing the performance of the Oxford Nanopore Technologies MinION. Biomol Detect Quantif 3:1–8. doi:10.1016/j.bdq.2015.02.001

Lazzarini LC, Huard RC, Boechat NL, Gomes HM, Oelemann MC, Kurepina N, Shashkina E, Mello FC, Gibson AL, Virginio MJ, Marsico AG, Butler WR, Kreiswirth BN, Suffys PN, Lapa ESJR, Ho JL (2007) Discovery of a novel Mycobacterium tuberculosis lineage that is a major cause of tuberculosis in Rio de Janeiro, Brazil. J Clin Microbiol 45(12):3891–3902. JCM.01394-07 [pii]. doi:10.1128/JCM.01394-07

Le Fleche P, Fabre M, Denoeud F, Koeck JL, Vergnaud G (2002) High resolution, on-line identification of strains from the Mycobacterium tuberculosis complex based on tandem repeat typing. BMC Microbiol 2:37

Lee RS, Radomski N, Proulx JF, Levade I, Shapiro BJ, McIntosh F, Soualhine H, Menzies D, Behr MA (2015a) Population genomics of Mycobacterium tuberculosis in the Inuit. Proc Natl Acad Sci U S A 112(44):13609–13614. doi:10.1073/pnas.1507071112

Lee RS, Radomski N, Proulx JF, Manry J, McIntosh F, Desjardins F, Soualhine H, Domenech P, Reed MB, Menzies D, Behr MA (2015b) Reemergence and amplification of tuberculosis in the Canadian arctic. J Infect Dis 211(12):1905–1914. doi:10.1093/infdis/jiv011

Liu X, Gutacker MM, Musser JM, Fu YX (2006) Evidence for recombination in Mycobacterium tuberculosis. J Bacteriol 188:8169

Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ (2012) Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 30(5):434–439. doi:10.1038/nbt.2198

Lu H, Giordano F, Ning Z (2016) Oxford Nanopore MinION sequencing and genome assembly. Genomics Proteomics Bioinformatics 14(5):265–279. doi:10.1016/j.gpb.2016.05.004

Luo T, Yang C, Pang Y, Zhao Y, Mei J, Gao Q (2014a) Development of a hierarchical variable-number tandem repeat typing scheme for Mycobacterium tuberculosis in China. PLoS One 9(2):e89726. doi:10.1371/journal.pone.0089726

Luo T, Yang C, Peng Y, Lu L, Sun G, Wu J, Jin X, Hong J, Li F, Mei J, DeRiemer K, Gao Q (2014b) Whole-genome sequencing to detect recent transmission of Mycobacterium tuberculosis in settings with a high burden of tuberculosis. Tuberculosis (Edinb) 94(4):434–440. doi:10.1016/j.tube.2014.04.005

Luo T, Comas I, Luo D, Lu B, Wu J, Wei L, Yang C, Liu Q, Gan M, Sun G, Shen X, Liu F, Gagneux S, Mei J, Lan R, Wan K, Gao Q (2015) Southern east Asian origin

and coexpansion of Mycobacterium tuberculosis Beijing family with Han Chinese. Proc Natl Acad Sci U S A 112(26):8136–8141. doi:10.1073/pnas.1424063112

Maeda S, Murase Y, Mitarai S, Sugawara I, Kato S (2008) Rapid, simple genotyping method by the variable numbers of tandem repeats (VNTR) for Mycobacterium tuberculosis isolates in Japan–analytical procedure of JATA (12)-VNTR. Kekkaku 83(10):673–678

Maes M, Kremer K, van Soolingen D, Takiff H, de Waard JH (2008) 24-locus MIRU-VNTR genotyping is a useful tool to study the molecular epidemiology of tuberculosis among Warao Amerindians in Venezuela. Tuberculosis (Edinb) 88(5):490–494. doi:10.1016/j.tube.2008.04.003

Magdalena J, Supply P, Locht C (1998) Specific differentiation between Mycobacterium bovis BCG and virulent strains of the Mycobacterium tuberculosis complex. J Clin Microbiol 36(9):2471–2476

Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A 95(6):3140–3145

Maiden MC, Jansen van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, McCarthy ND (2013) MLST revisited: the gene-by-gene approach to bacterial genomics. Nat Rev Microbiol 11(10):728–736. doi:10.1038/nrmicro3093

Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV (2011) Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol 9(6):467–477. doi:10.1038/nrmicro2577

Mankiewicz E, Liivak M (1975) Phage types of mycobacterium tuberculosis in cultures isolated from Eskimo patients. Am Rev Respir Dis 111(3):307–312. doi:10.1164/arrd.1975.111.3.307

Marais BJ (2016) The global tuberculosis situation and the inexorable rise of drug-resistant disease. Adv Drug Deliv Rev. doi:10.1016/j.addr.2016.01.021

Matsumoto T, Koshii Y, Sakane K, Murakawa T, Hirayama Y, Yoshida H, Kurokawa M, Tamura Y, Nagai T, Kawase I (2013) A novel approach to automated genotyping of Mycobacterium tuberculosis using a panel of 15 MIRU VNTRs. J Microbiol Methods 93(3):239–241. doi:10.1016/j.mimet.2013.03.022

Mazars E, Lesjean S, Banuls AL, Gilbert M, Vincent V, Gicquel B, Tibayrenc M, Locht C, Supply P (2001) High-resolution minisatellite-based typing as a portable approach to global analysis of Mycobacterium tuberculosis molecular epidemiology. Proc Natl Acad Sci U S A 98(4):1901–1906

Mbugi EV, Katale BZ, Streicher EM, Keyyu JD, Kendall SL, Dockrell HM, Michel AL, Rweyemamu MM, Warren RM, Matee MI, van Helden PD, Couvin D, Rastogi N (2016) Mapping of mycobacterium tuberculosis complex genetic diversity profiles in Tanzania and other African countries. PLoS One 11(5):e0154571. doi:10.1371/journal.pone.0154571

McAdam RA, Hermans PW, van Soolingen D, Zainuddin ZF, Catty D, van Embden JD, Dale JW (1990) Characterization of a Mycobacterium tuberculosis insertion sequence belonging to the IS3 family. Mol Microbiol 4(9):1607–1613

Merker M, Kohl TA, Roetzer A, Truebe L, Richter E, Rusch-Gerdes S, Fattorini L, Oggioni MR, Cox H, Varaine F, Niemann S (2013) Whole genome sequencing reveals complex evolution patterns of multidrug-resistant Mycobacterium tuberculosis Beijing strains in patients. PLoS One 8(12):e82551. doi:10.1371/journal.pone.0082551

Merker M, Blin C, Mona S, Duforet-Frebourg N, Lecher S, Willery E, Blum M, Rusch-Gerdes S, Mokrousov I, Aleksic E, Allix-Beguec C, Antierens A, Augustynowicz-Kopec E, Ballif M, Barletta F, Beck HP, Barry CE 3rd, Bonnet M, Borroni E, Campos-Herrero I, Cirillo D, Cox H, Crowe S, Crudu V, Diel R, Drobniewski F, Fauville-Dufaux M, Gagneux S, Ghebremichael S, Hanekom M, Hoffner S, Jiao WW, Kalon S, Kohl TA, Kontsevaya I, Lillebaek T, Maeda S, Nikolayevskyy V, Rasmussen M, Rastogi N, Samper S, Sanchez-Padilla E, Savic B, Shamputa IC, Shen A, Sng LH, Stakenas P, Toit K, Varaine F, Vukovic D, Wahl C, Warren R, Supply P, Niemann S, Wirth T (2015) Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage. Nat Genet. doi:10.1038/ng.3195

Michele TM, Cronin WA, Graham NM, Dwyer DM, Pope DS, Harrington S, Chaisson RE, Bishai WR (1997) Transmission of Mycobacterium tuberculosis by a fiberoptic bronchoscope. Identification by DNA fingerprinting. JAMA 278(13):1093–1095

Miyamoto M, Motooka D, Gotoh K, Imai T, Yoshitake K, Goto N, Iida T, Yasunaga T, Horii T, Arakawa K, Kasahara M, Nakamura S (2014) Performance comparison of second- and third-generation sequencers using a bacterial genome with two chromosomes. BMC Genomics 15:699. doi:10.1186/1471-2164-15-699

Mokrousov I (2013) Insights into the origin, emergence, and current spread of a successful Russian clone of Mycobacterium tuberculosis. Clin Microbiol Rev 26(2):342–360. doi:10.1128/CMR.00087-12

Mokrousov I, Narvskaya O, Vyazovaya A, Millet J, Otten T, Vishnevsky B, Rastogi N (2008) Mycobacterium tuberculosis Beijing genotype in Russia: in search of informative variable-number tandem-repeat loci. J Clin Microbiol 46(11):3576–3584. doi:10.1128/JCM.00414-08

Mokrousov I, Vyazovaya A, Iwamoto T, Skiba Y, Pole I, Zhdanova S, Arikawa K, Sinkov V, Umpeleva T, Valcheva V, Alvarez Figueroa M, Ranka R, Jansone I, Ogarkov O, Zhuravlev V, Narvskaya O (2016) Latin-American-Mediterranean lineage of Mycobacterium tuberculosis: human traces across pathogen's phylogeography. Mol Phylogenet Evol 99:133–143. doi:10.1016/j.ympev.2016.03.020

Mortimer TD, Pepperell CS (2014) Genomic signatures of distributive conjugal transfer among mycobacteria. Genome Biol Evol 6(9):2489–2500. doi:10.1093/gbe/evu175

Mostowy S, Cousins D, Brinkman J, Aranaz A, Behr MA (2002) Genomic deletions suggest a phylogeny for the Mycobacterium tuberculosis complex. J Infect Dis 186(1):74–80

Muller B, Hilty M, Berg S, Garcia-Pelayo MC, Dale J, Boschiroli ML, Cadmus S, Ngandolo BN, Godreuil S, Diguimbaye-Djaibe C, Kazwala R, Bonfoh B, Njanpop-Lafourcade BM, Sahraoui N, Guetarni D, Aseffa A, Mekonnen MH, Razanamparany VR, Ramarokoto H, Djonne B, Oloya J, Machado A, Mucavele C, Skjerve E, Portaels F, Rigouts L, Michel A, Muller A, Kallenius G, van Helden PD, Hewinson RG, Zinsstag J, Gordon SV, Smith NH (2009) African 1, an epidemiologically important clonal complex of Mycobacterium bovis dominant in Mali, Nigeria, Cameroon, and Chad. J Bacteriol 191(6):1951–1960. doi:10.1128/JB.01590-08

Nadon CA, Trees E, Ng LK, Moller Nielsen E, Reimer A, Maxwell N, Kubota KA, Gerner-Smidt P (2013) Development and application of MLVA methods as a tool for inter-laboratory surveillance. Euro surveillance: bulletin Europeen sur les maladies transmissibles. Eur Commun Dis Bull 18(35):20565

Namouchi A, Didelot X, Schock U, Gicquel B, Rocha EP (2012) After the bottleneck: genome-wide diversification of the Mycobacterium tuberculosis complex by mutation, recombination, and natural selection. Genome Res 22(4):721–734. doi:10.1101/gr.129544.111

Niemann S, Supply P (2014) Diversity and evolution of Mycobacterium tuberculosis: moving to whole-genome-based approaches. Cold Spring Harb Perspect Med 4(12):a021188. doi:10.1101/cshperspect.a021188

Niemann S, Rusch-Gerdes S, Richter E (1997) IS6110 fingerprinting of drug-resistant Mycobacterium tuberculosis strains isolated in Germany during 1995. J Clin Microbiol 35(12):3015–3020

Niemann S, Harmsen D, Rusch-Gerdes S, Richter E (2000a) Differentiation of clinical Mycobacterium tuberculosis complex isolates by gyrB DNA sequence polymorphism analysis. J Clin Microbiol 38(9):3231–3234

Niemann S, Rusch-Gerdes S, Richter E, Thielen H, Heykes-Uden H, Diel R (2000b) Stability of IS6110 restriction fragment length polymorphism patterns of Mycobacterium tuberculosis strains in actual chains of transmission. J Clin Microbiol 38(7):2563–2567

Niemann S, Koser CU, Gagneux S, Plinke C, Homolka S, Bignell H, Carter RJ, Cheetham RK, Cox A, Gormley NA, Kokko-Gonzales P, Murray LJ, Rigatti R, Smith VP, Arends FP, Cox HS, Smith G, Archer JA (2009) Genomic diversity among drug sensitive and multidrug resistant isolates of Mycobacterium tuberculosis with identical DNA fingerprints. PLoS One 4(10):e7407. doi:10.1371/journal.pone.0007407

Nikolayevskyy V, Gopaul K, Balabanova Y, Brown T, Fedorin I, Drobniewski F (2006) Differentiation of tuberculosis strains in a population with mainly Beijing-family strains. Emerg Infect Dis 12(9):1406–1413

Nikolayevskyy V, Trovato A, Broda A, Borroni E, Cirillo D, Drobniewski F (2016) MIRU-VNTR genotyping of mycobacterium tuberculosis strains using QIAxcel technology: a multicentre evaluation study. PLoS One 11(3):e0149435. doi:10.1371/journal.pone.0149435

Nogueira CL, Prim RI, Senna SG, Rovaris DB, Maurici R, Rossetti ML, Couvin D, Rastogi N, Bazzo ML (2016) First insight into the molecular epidemiology of Mycobacterium tuberculosis in Santa Catarina, southern Brazil. Tuberculosis (Edinb) 97:57–64. doi:10.1016/j.tube.2015.12.005

Oelemann MC, Diel R, Vatin V, Haas W, Rusch-Gerdes S, Locht C, Niemann S, Supply P (2007) Assessment of an optimized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiology studies of tuberculosis. J Clin Microbiol 45(3):691–697. doi:10.1128/JCM.01393-06

Pankhurst LJ, Del Ojo EC, Votintseva AA, Walker TM, Cole K, Davies J, Fermont JM, Gascoyne-Binzi DM, Kohl TA, Kong C, Lemaitre N, Niemann S, Paul J, Rogers TR, Roycroft E, Smith EG, Supply P, Tang P, Wilcox MH, Wordsworth S, Wyllie D, Xu L, Crook DW (2016) Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. Lancet Respir Med 4(1):49–58. doi:10.1016/S2213-2600(15)00466-X

Pepperell CS, Casto AM, Kitchen A, Granka JM, Cornejo OE, Holmes EC, Birren B, Galagan J, Feldman MW (2013) The role of selection in shaping diversity of natural M. tuberculosis populations. PLoS Pathog 9(8):e1003543. doi:10.1371/journal.ppat.1003543

Perez-Lago L, Comas I, Navarro Y, Gonzalez-Candelas F, Herranz M, Bouza E, Garcia-de-Viedma D (2014) Whole genome sequencing analysis of intrapatient microevolution in Mycobacterium tuberculosis: potential impact on the inference of tuberculosis transmission. J Infect Dis 209(1):98–108. doi:10.1093/infdis/jit439

Poulet S, Cole ST (1995) Characterization of the highly abundant polymorphic GC-rich-repetitive sequence (PGRS) present in Mycobacterium tuberculosis. Arch Microbiol 163(2):87–95

Pouseele H, Supply P (2015) Accurate whole-genome sequencing-based epidemiological surveillance of mycobacterium tuberculosis. In: Sails A, Tang YW (eds) Current and emerging technologies for the diagnosis of microbial infections, vol 45. Methods in microbiology, pp 359–394

Prod'hom G, Guilhot C, Gutierrez MC, Varnerot A, Gicquel B, Vincent V (1997) Rapid discrimination of Mycobacterium tuberculosis complex strains by ligation-mediated PCR fingerprint analysis. J Clin Microbiol 35(12):3331–3334

Ragheb MN, Ford CB, Chase MR, Lin PL, Flynn JL, Fortune SM (2013) The mutation rate of mycobacte-

rial repetitive unit loci in strains of M. tuberculosis from cynomolgus macaque infection. BMC Genomics 14:145. doi:10.1186/1471-2164-14-145

Raleigh JW, Wichelhausen R (1973) Exogenous reinfection with Mycobacterium tuberculosis confirmed by phage typing. Am Rev Respir Dis 108(3):639–642. doi:10.1164/arrd.1973.108.3.639

Reiling N, Homolka S, Walter K, Brandenburg J, Niwinski L, Ernst M, Herzmann C, Lange C, Diel R, Ehlers S, Niemann S (2013) Clade-specific virulence patterns of Mycobacterium tuberculosis complex strains in human primary macrophages and aerogenically infected mice. mBio 4 (4):10.1128/mBio.00250-13

Reyes JF, Tanaka MM (2010) Mutation rates of spoligotypes and variable numbers of tandem repeat loci in Mycobacterium tuberculosis. Infect Genet Evol 10(7):1046–1051. doi:10.1016/j.meegid.2010.06.016

Richter E, Weizenegger M, Rusch-Gerdes S, Niemann S (2003) Evaluation of genotype MTBC assay for differentiation of clinical Mycobacterium tuberculosis complex isolates. J Clin Microbiol 41(6): 2672–2675

Roetzer A, Schuback S, Diel R, Gasau F, Ubben T, di Nauta A, Richter E, Rusch-Gerdes S, Niemann S (2011) Evaluation of Mycobacterium tuberculosis typing methods in a 4-year study in Schleswig-Holstein, northern Germany. J Clin Microbiol 49(12):4173–4178. doi:10.1128/jcm.05293-11

Roetzer A, Diel R, Kohl TA, Ruckert C, Nubel U, Blom J, Wirth T, Jaenicke S, Schuback S, Rusch-Gerdes S, Supply P, Kalinowski J, Niemann S (2013) Whole genome sequencing versus traditional genotyping for investigation of a Mycobacterium tuberculosis outbreak: a longitudinal molecular epidemiological study. PLoS Med 10(2):e1001387. doi:10.1371/journal.pmed.1001387

Roisin S, Gaudin C, De Mendonca R, Bellon J, Van Vaerenbergh K, De Bruyne K, Byl B, Pouseele H, Denis O, Supply P (2016) Pan-genome multilocus sequence typing and outbreak-specific reference-based single nucleotide polymorphism analysis to resolve two concurrent Staphylococcus aureus outbreaks in neonatal services. Clin Microbiol Infect 22(6):520–526. doi:10.1016/j.cmi.2016.01.024

Roring S, Scott A, Brittain D, Walker I, Hewinson G, Neill S, Skuce R (2002) Development of variable-number tandem repeat typing of Mycobacterium bovis: comparison of results with those obtained by using existing exact tandem repeats and spoligotyping. J Clin Microbiol 40(6):2126–2133

Rosenberg NA, Tsolaki AG, Tanaka MM (2003) Estimating change rates of genetic markers using serial samples: applications to the transposon IS6110 in Mycobacterium tuberculosis. Theor Popul Biol 63(4):347–363

Sandegren L, Groenheit R, Koivula T, Ghebremichael S, Advani A, Castro E, Pennhag A, Hoffner S, Mazurek J, Pawlowski A, Kan B, Bruchfeld J, Melefors O, Kallenius G (2011) Genomic stability over 9 years of an isoniazid resistant Mycobacterium tuberculosis outbreak strain in Sweden. PLoS One 6(1):e16647. doi:10.1371/journal.pone.0016647

Savine E, Warren RM, van der Spuy GD, Beyers N, van Helden PD, Locht C, Supply P (2002) Stability of variable-number tandem repeats of mycobacterial interspersed repetitive units from 12 loci in serial isolates of Mycobacterium tuberculosis. J Clin Microbiol 40(12):4561–4566

Schurch AC, Kremer K, Kiers A, Daviena O, Boeree MJ, Siezen RJ, Smith NH, van Soolingen D (2010) The tempo and mode of molecular evolution of Mycobacterium tuberculosis at patient-to-patient scale. Infect Genet Evol 10(1):108–114. doi:10.1016/j.meegid.2009.10.002

Shamputa IC, Jugheli L, Sadradze N, Willery E, Portaels F, Supply P, Rigouts L (2006) Mixed infection and clonal representativeness of a single sputum sample in tuberculosis patients from a penitentiary hospital in Georgia. Respir Res 7:99. doi:10.1186/1465-9921-7-99

Skrahina A, Hurevich H, Zalutskaya A, Sahalchyk E, Astrauko A, Hoffner S, Rusovich V, Dadu A, de Colombani P, Dara M, van Gemert W, Zignol M (2013) Multidrug-resistant tuberculosis in Belarus: the size of the problem and associated risk factors. Bull World Health Organ 91(1):36–45. doi:10.2471/BLT.12.104588

Skuce RA, McCorry TP, McCarroll JF, Roring SM, Scott AN, Brittain D, Hughes SL, Hewinson RG, Neill SD (2002) Discrimination of Mycobacterium tuberculosis complex bacteria using novel VNTR-PCR targets. Microbiology 148(Pt 2):519–528

Small PM, McClenny NB, Singh SP, Schoolnik GK, Tompkins LS, Mickelsen PA (1993a) Molecular strain typing of Mycobacterium tuberculosis to confirm cross-contamination in the mycobacteriology laboratory and modification of procedures to minimize occurrence of false-positive cultures. J Clin Microbiol 31(7):1677–1682

Small PM, Shafer RW, Hopewell PC, Singh SP, Murphy MJ, Desmond E, Sierra MF, Schoolnik GK (1993b) Exogenous reinfection with multidrug-resistant Mycobacterium tuberculosis in patients with advanced HIV infection. N Engl J Med 328(16):1137–1144. doi:10.1056/NEJM199304223281601

Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, Schecter GF, Daley CL, Schoolnik GK (1994) The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. N Engl J Med 330(24):1703–1709. doi:10.1056/NEJM199406163302402

Smit PW, Vasankari T, Aaltonen H, Haanpera M, Casali N, Marttila H, Marttila J, Ojanen P, Ruohola A, Ruutu P, Drobniewski F, Lyytikainen O, Soini H (2014) Enhanced tuberculosis outbreak investigation using whole genome sequencing and IGRA. Eur Respir J. doi:10.1183/09031936.00125914

Smith NH, Hewinson RG, Kremer K, Brosch R, Gordon SV (2009) Myths and misconceptions: the origin and

evolution of Mycobacterium tuberculosis. Nat Rev Microbiol 7(7):537–544. doi:10.1038/nrmicro2165

Smittipat N, Palittapongarnpim P (2000) Identification of possible loci of variable number of tandem repeats in Mycobacterium tuberculosis. Tuber Lung Dis 80(2):69–74. doi:10.1054/tuld.2000.0236

Smittipat N, Billamas P, Palittapongarnpim M, Thong-On A, Temu MM, Thanakijcharoen P, Karnkawinpong O, Palittapongarnpim P (2005) Polymorphism of variable-number tandem repeats at multiple loci in Mycobacterium tuberculosis. J Clin Microbiol 43(10):5034–5043. doi:10.1128/JCM.43.10.5034-5043.2005

Snider DE Jr, Jones WD, Good RC (1984) The usefulness of phage typing Mycobacterium tuberculosis isolates. Am Rev Respir Dis 130(6):1095–1099. doi:10.1164/arrd.1984.130.6.1095

Sola C, Filliol I, Legrand E, Mokrousov I, Rastogi N (2001a) Mycobacterium tuberculosis phylogeny reconstruction based on combined numerical analysis with IS1081, IS6110, VNTR, and DR-based spoligotyping suggests the existence of two new phylogeographical clades. J Mol Evol 53(6):680–689. doi:10.1007/s002390010255

Sola C, Filliol I, Gutierrez MC, Mokrousov I, Vincent V, Rastogi N (2001b) Spoligotype database of Mycobacterium tuberculosis: biogeographic distribution of shared types and epidemiologic and phylogenetic perspectives. Emerg Infect Dis 7(3):390–396. doi:10.3201/eid0703.010304

Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, Musser JM (1997) Restricted structural gene polymorphism in the Mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination. Proc Natl Acad Sci U S A 94(18):9869–9874

Starks AM, Aviles E, Cirillo DM, Denkinger CM, Dolinger DL, Emerson C, Gallarda J, Hanna D, Kim PS, Liwski R, Miotto P, Schito M, Zignol M (2015) Collaborative effort for a centralized worldwide tuberculosis relational sequencing data platform. Clin Infect Dis 61(Suppl 3):S141–S146. doi:10.1093/cid/civ610

Stead WW, Bates JH (1969) Primary tuberculosis from the Far East. Transmission by a veteran to two civilians. Ann Intern Med 70(4):707–711

Stucki D, Ballif M, Bodmer T, Coscolla M, Maurer AM, Droz S, Butz C, Borrell S, Langle C, Feldmann J, Furrer H, Mordasini C, Helbling P, Rieder HL, Egger M, Gagneux S, Fenner L (2015) Tracking a tuberculosis outbreak over 21 years: strain-specific single-nucleotide polymorphism typing combined with targeted whole-genome sequencing. J Infect Dis 211(8):1306–1316. doi:10.1093/infdis/jiu601

Stucki D, Ballif M, Egger M, Furrer H, Altpeter E, Battegay M, Droz S, Bruderer T, Coscolla M, Borrell S, Zurcher K, Janssens JP, Calmy A, Mazza Stalder J, Jaton K, Rieder HL, Pfyffer GE, Siegrist HH, Hoffmann M, Fehr J, Dolina M, Frei R, Schrenzel J, Bottger EC, Gagneux S, Fenner L (2016a) Standard genotyping overestimates transmission of mycobacterium tuberculosis among immigrants in a low-incidence country. J Clin Microbiol 54(7):1862–1870. doi:10.1128/jcm.00126-16

Stucki D, Brites D, Jeljeli L, Coscolla M, Liu Q, Trauner A, Fenner L, Rutaihwa L, Borrell S, Luo T, Gao Q, Kato-Maeda M, Ballif M, Egger M, Macedo R, Mardassi H, Moreno M, Vilanova GT, Fyfe J, Globan M, Thomas J, Jamieson F, Guthrie JL, Asante-Poku A, Yeboah-Manu D, Wampande E, Ssengooba W, Joloba M, Boom WH, Basu I, Bower J, Saraiva M, Vasconcellos SE, Suffys P, Koch A, Wilkinson R, Gail-Bekker L, Malla B, Ley SD, Beck HP, de Jong BC, Toit K, Sanchez-Padilla E, Bonnet M, Gil-Brusola A, Frank M, Penlap Beng VN, Eisenach K, Alani I, Ndung'u PW, Revathi G, Gehre F, Akter S, Ntoumi F, Stewart-Isherwood L, Ntinginya NE, Rachow A, Hoelscher M, Cirillo DM, Skenders G, Hoffner S, Bakonyte D, Stakenas P, Diel R, Crudu V, Moldovan O, Al-Hajoj S, Otero L, Barletta F, Carter EJ, Diero L, Supply P, Comas I, Niemann S, Gagneux S (2016b) Mycobacterium tuberculosis lineage 4 comprises globally distributed and geographically restricted sublineages. Nat Genet 48(12):1535–1543. doi:10.1038/ng.3704

Suffys PN, Ivens de Araujo ME, Rossetti ML, Zahab A, Barroso EW, Barreto AM, Campos E, van Soolingen D, Kremer K, Heersma H, Degrave WM (2000) Usefulness of IS6110-restriction fragment length polymorphism typing of Brazilian strains of Mycobacterium tuberculosis and comparison with an international fingerprint database. Res Microbiol 151(5):343–351

Supply P, Magdalena J, Himpens S, Locht C (1997) Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. Mol Microbiol 26(5):991–1003

Supply P, Mazars E, Lesjean S, Vincent V, Gicquel B, Locht C (2000) Variable human minisatellite-like regions in the Mycobacterium tuberculosis genome. Mol Microbiol 36(3):762–771

Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D, Locht C (2001) Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. J Clin Microbiol 39(10):3563–3571

Supply P, Warren RM, Banuls AL, Lesjean S, Van Der Spuy GD, Lewis LA, Tibayrenc M, Van Helden PD, Locht C (2003) Linkage disequilibrium between minisatellite loci supports clonal evolution of *Mycobacterium tuberculosis* in a high tuberculosis incidence area. Mol Microbiol 47(2):529–538

Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rusch-Gerdes S, Willery E, Savine E, de Haas P, van Deutekom H, Roring S, Bifani P, Kurepina N, Kreiswirth B, Sola C, Rastogi N, Vatin V, Gutierrez MC, Fauville M, Niemann S, Skuce R, Kremer K, Locht C, van Soolingen D (2006) Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of Mycobacterium tuberculosis. J Clin Microbiol 44(12):4498–4510. doi:10.1128/JCM.01392-06

Supply P, Niemann S, Wirth T (2011a) On the mutation rates of spoligotypes and variable numbers of tandem repeat loci of Mycobacterium tuberculosis. Infect Genet Evol 11(2):251–252. S1567-1348(10)00347-3 [pii]. doi:10.1016/j.meegid.2010.12.009

Supply P, Niemann S, Wirth T (2011b) On the mutation rates of spoligotypes and variable numbers of tandem repeat loci of Mycobacterium tuberculosis: continued-when tuning matters. Infect Genet Evol. S1567-1348(11)00093-1 [pii]. doi:10.1016/j.meegid.2011.03.012

Supply P, Marceau M, Mangenot S, Roche D, Rouanet C, Khanna V, Majlessi L, Criscuolo A, Tap J, Pawlik A, Fiette L, Orgeur M, Fabre M, Parmentier C, Frigui W, Simeone R, Boritsch EC, Debrie AS, Willery E, Walker D, Quail MA, Ma L, Bouchier C, Salvignol G, Sayes F, Cascioferro A, Seemann T, Barbe V, Locht C, Gutierrez MC, Leclerc C, Bentley SD, Stinear TP, Brisse S, Medigue C, Parkhill J, Cruveiller S, Brosch R (2013) Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of Mycobacterium tuberculosis. Nat Genet 45(2):172–179. doi:10.1038/ng.2517

Supply P, Gaudin C, Raze D (2014) Optimization of standard 24-locus variable-number tandem-repeat typing of Mycobacterium tuberculosis isolates: a multicenter perspective. J Clin Microbiol 52(9):3518–3519. doi:10.1128/JCM.01790-14

Tanaka MM, Rosenberg NA (2001) Optimal estimation of transposition rates of insertion sequences for molecular epidemiology. Stat Med 20(16):2409–2420

Thierry D, Brisson-Noel A, Vincent-Levy-Frebault V, Nguyen S, Guesdon JL, Gicquel B (1990a) Characterization of a Mycobacterium tuberculosis insertion sequence, IS6110, and its application in diagnosis. J Clin Microbiol 28(12): 2668–2673

Thierry D, Cave MD, Eisenach KD, Crawford JT, Bates JH, Gicquel B, Guesdon JL (1990b) IS6110, an IS-like element of Mycobacterium tuberculosis complex. Nucleic Acids Res 18(1):188

Trovato A, Tafaj S, Battaglia S, Alagna R, Bardhi D, Kapisyzi P, Bala S, Haldeda M, Borroni E, Hafizi H, Cirillo DM (2016) Implementation of a consensus set of hypervariable mycobacterial interspersed repetitive-unit-variable-number tandem-repeat loci in mycobacterium tuberculosis molecular epidemiology. J Clin Microbiol 54(2):478–482. doi:10.1128/JCM.02945-15

Tsolaki AG, Hirsh AE, DeRiemer K, Enciso JA, Wong MZ, Hannan M, Goguet de la Salmoniere YO, Aman K, Kato-Maeda M, Small PM (2004) Functional and evolutionary genomics of Mycobacterium tuberculosis: insights from genomic deletions in 100 strains. Proc Natl Acad Sci U S A 101(14):4865–4870

Uzoewulu GN, Lawson L, Nnanna IS, Rastogi N, Goyal M (2016) Genetic diversity of Mycobacterium tuberculosis complex strains isolated from patients with pulmonary tuberculosis in Anambra State, Nigeria. Inte J Mycobacteriol 5(1):74–79. doi:10.1016/j.ijmyco.2015.06.008

van Deutekom H, Supply P, de Haas PE, Willery E, Hoijng SP, Locht C, Coutinho RA, van Soolingen D (2005) Molecular typing of Mycobacterium tuberculosis by mycobacterial interspersed repetitive unit-variable-number tandem repeat analysis, a more accurate method for identifying epidemiological links between patients with tuberculosis. J Clin Microbiol 43(9): 4473–4479

van Embden JD, van Soolingen D, Small PM, Hermans PW (1992) Genetic markers for the epidemiology of tuberculosis. Res Microbiol 143(4):385–391

van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, Hermans P, Martin C, McAdam R, Shinnick TM et al (1993) Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 31(2):406–409

Van Soolingen D (2001) Molecular epidemiology of tuberculosis and other mycobacterial infections: main methodologies and achievements. J Intern Med 249(1):1–26

van Soolingen D, Qian L, de Haas PE, Douglas JT, Traore H, Portaels F, Qing HZ, Enkhsaikan D, Nymadawa P, van Embden JD (1995) Predominance of a single genotype of Mycobacterium tuberculosis in countries of east Asia. J Clin Microbiol 33(12):3234–3238

van Soolingen D, Borgdorff MW, de Haas PE, Sebek MM, Veen J, Dessens M, Kremer K, van Embden JD (1999) Molecular epidemiology of tuberculosis in the Netherlands: a nationwide study from 1993 through 1997. J Infect Dis 180(3):726–736. doi:10.1086/314930

van der Zanden AG, Kremer K, Schouls LM, Caimi K, Cataldi A, Hulleman A, Nagelkerke NJ, van Soolingen D (2002a) Improvement of differentiation and interpretability of spoligotyping for Mycobacterium tuberculosis complex isolates by introduction of new spacer oligonucleotides. J Clin Microbiol 40(12): 4628–4639

Velji P, Nikolayevskyy V, Brown T, Drobniewski F (2009) Discriminatory ability of hypervariable variable number tandem repeat loci in population-based analysis of Mycobacterium tuberculosis strains, London, UK. Emerg Infect Dis 15(10):1609–1616. doi:10.3201/eid1510.090463

Wada T, Maeda S, Hase A, Kobayashi K (2007) Evaluation of variable numbers of tandem repeat as molecular epidemiological markers of Mycobacterium tuberculosis in Japan. J Med Microbiol 56(Pt 8):1052–1057. doi:10.1099/jmm.0.46990-0

Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, Eyre DW, Wilson DJ, Hawkey PM, Crook DW, Parkhill J, Harris D, Walker AS, Bowden R, Monk P, Smith EG, Peto TE (2013a) Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. Lancet Infect Dis 13(2):137–146. doi:10.1016/S1473-3099(12)70277-3

Walker TM, Monk P, Grace Smith E, Peto TE (2013b) Contact investigations for outbreaks of Mycobacterium tuberculosis: advances through

whole genome sequencing. Clin Microbiol Infect. doi:10.1111/1469-0691.12183

Walker TM, Lalor MK, Broda A, Saldana Ortega L, Morgan M, Parker L, Churchill S, Bennett K, Golubchik T, Giess AP, Del Ojo EC, Jeffery KJ, Bowler IC, Laurenson IF, Barrett A, Drobniewski F, McCarthy ND, Anderson LF, Abubakar I, Thomas HL, Monk P, Smith EG, Walker AS, Crook DW, Peto TE, Conlon CP (2014) Assessment of Mycobacterium tuberculosis transmission in Oxfordshire, UK, 2007-12, with whole pathogen genome sequences: an observational study. Lancet Respir Med 2(4):285–292. doi:10.1016/S2213-2600(14)70027-X

Walker TM, Kohl TA, Omar SV, Hedge J, Del Ojo EC, Bradley P, Iqbal Z, Feuerriegel S, Niehaus KE, Wilson DJ, Clifton DA, Kapatai G, Ip CL, Bowden R, Drobniewski FA, Allix-Beguec C, Gaudin C, Parkhill J, Diel R, Supply P, Crook DW, Smith EG, Walker AS, Ismail N, Niemann S, Peto TE (2015) Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. Lancet Infect Dis. doi:10.1016/S1473-3099(15)00062-6

Wang W, Mathema B, Hu Y, Zhao Q, Jiang W, Xu B (2014) Role of casual contacts in the recent transmission of tuberculosis in settings with high disease burden. Clin Microbiol Infect 20(11):1140–1145. doi:10.1111/1469-0691.12726

Warren RM, Sampson SL, Richardson M, Van Der Spuy GD, Lombard CJ, Victor TC, van Helden PD (2000) Mapping of IS6110 flanking regions in clinical isolates of Mycobacterium tuberculosis demonstrates genome plasticity. Mol Microbiol 37(6): 1405–1416

Warren RM, Victor TC, Streicher EM, Richardson M, van der Spuy GD, Johnson R, Chihota VN, Locht C, Supply P, van Helden PD (2004) Clonal expansion of a globally disseminated lineage of Mycobacterium tuberculosis with low IS6110 copy numbers. J Clin Microbiol 42(12):5774–5782. 42/12/5774 [pii]. doi:10.1128/JCM.42.12.5774-5782.2004

Weniger T, Krawczyk J, Supply P, Niemann S, Harmsen D (2010) MIRU-VNTRplus: a web tool for polyphasic genotyping of Mycobacterium tuberculosis complex bacteria. Nucleic Acids Res 38(Suppl):W326–W331. doi:10.1093/nar/gkq351

WHO (2016) Global tuberculosis report. World Health Organization, Geneva

Wiid IJ, Werely C, Beyers N, Donald P, van Helden PD (1994) Oligonucleotide (GTG)5 as a marker for Mycobacterium tuberculosis strain identification. J Clin Microbiol 32(5):1318–1321

Williams OM, Abeel T, Casali N, Cohen K, Pym AS, Mungall SB, Desjardins CA, Banerjee A, Drobniewski F, Earl AM, Cooke GS (2015) Fatal nosocomial MDR TB identified through routine genetic analysis and whole-genome sequencing. Emerg Infect Dis 21(6):1082–1084. doi:10.3201/eid2106.141903

Wirth T, Hildebrand F, Allix-Beguec C, Wolbeling F, Kubica T, Kremer K, van Soolingen D, Rusch-Gerdes S, Locht C, Brisse S, Meyer A, Supply P, Niemann S (2008) Origin, spread and demography of the Mycobacterium tuberculosis complex. PLoS Pathog 4(9):e1000160. doi:10.1371/journal.ppat.1000160

Yeh RW, Ponce de Leon A, Agasino CB, Hahn JA, Daley CL, Hopewell PC, Small PM (1998) Stability of Mycobacterium tuberculosis DNA genotypes. J Infect Dis 177(4):1107–1111

van der Zanden AG, Kremer K, Schouls LM, Caimi K, Cataldi A, Hulleman A, Nagelkerke NJ, van Soolingen D (2002b) Improvement of differentiation and interpretability of spoligotyping for Mycobacterium tuberculosis complex isolates by introduction of new spacer oligonucleotides. J Clin Microbiol 40(12): 4628–4639

Zhang Y, Young D (1994) Strain variation in the katG region of Mycobacterium tuberculosis. Mol Microbiol 14(2):301–308

Genomic Epidemiology of Tuberculosis

4

Iñaki Comas

Abstract

The application of next generation sequencing technologies has opened the door to a new molecular epidemiology of tuberculosis, in which we can now look at transmission at a resolution not possible before. At the same time, new technical and analytical challenges have appeared, and we are still exploring the wider potential of this new technology. Whole genome sequencing in tuberculosis still requires bacterial cultures. Thus, although whole genome sequencing has revolutionized the interpretation of transmission patterns, it is not yet ready to be applied at the point-of-care. In this chapter, I will review the promises and challenges of genomic epidemiology, as well as some of the new questions that have arisen from the use of this new technology. In addition, I will examine the role of molecular epidemiology within the general frame of global tuberculosis control and how genomic epidemiology can contribute towards the elimination of the disease.

Keywords

Transmission • genome sequencing • outbreak • mutation rate

4.1 Introduction

The tuberculosis (TB) epidemic has changed since 1993 when TB was declared a "global health emergency" by the World Health Organization (WHO). At that time, TB control was insufficient in most countries of the world, and large outbreaks occurred even in high-income countries (Comas and Gagneux 2009). Since 2015, elimination of TB is back in the public health agenda, but at different speeds across the globe (Dye et al. 2013). Some countries like the United States are close to eliminating TB among US-born patients, while TB incidence remains significantly higher among immigrants. By contrast, in 6 of the 22

I. Comas (✉)
Institute of Biomedicine of Valencia (IBV-CSIC) and CIBER in Epidemiology and Public Health, Jaime Roig 11, 46010, Valencia, Spain
e-mail: icomas@ibv.csic.es

S. Gagneux (ed.), *Strain Variation in the* Mycobacterium tuberculosis *Complex: Its Role in Biology, Epidemiology and Control*, Advances in Experimental Medicine and Biology 1019,
DOI 10.1007/978-3-319-64371-7_4

high burden countries such as Mozambique or Pakistan TB is on the rise, and the growing incidence of TB already multidrug-resistant (MDR) is hampering control programs. Human immunodeficiency virus (HIV) is no longer the only co-morbidity focusing attention of the public health community. Diabetes, alcohol abuse and smoking are also driving the disease in developing countries. Globally, TB incidence has only started to decline but at a pace insufficient to eliminate the disease in the next century. The DOTS program set up by WHO in 1996 barely can cope with this changing epidemic, and now that elimination is back on the agenda, there are concerns that we are reaching the limits of DOTS (Dye et al. 2013). Efforts like DOTS-plus have been developed to address MDR-TB in resource-limited countries, but DOTS-plus relies on culture facilities and prompt diagnosis (Sterling et al. 2003). Even though DOTS has saved six million lives, we will need to adapt global TB control to the specific epidemics in each country or geographic region. The "one-size-fits-all" approach cannot be applied anymore to a situation, in which many countries are close to reaching elimination, while in others the epidemic is still ongoing (WHO 2015).

The good news is that in the last 5 years, we have witnessed the development of new tools against TB that are slowly being integrated into control programs. Rapid molecular tests, in particular Xpert MTB/RIF, have dramatically improved and accelerated the diagnosis of TB and are being used for drug resistance surveillance (Pai and Schito 2015). However, despite of these major achievements, many problems around the global rollout of these tests remain. For example, these tests are mainly used at referral centers and are yet to be further developed into point-of-care diagnostics. Moreover, increased detection of TB cases does not always lead to an increase of cases starting or completing treatment. Two new drugs, bedaquiline and delamanid, have recently been licensed for the treatment of TB for the first time in 40 years, but so far, they have primarily been used on a compassionate basis for MDR and XDR-TB cases (Zumla et al. 2014). While still evaluating their future role on global TB control, we are far from being able to substitute the current four drug regimen lasting 6 months by a new and shorter regimen. Moreover, we have already witnessed the emergence of resistance to these new drugs (Bloemberg et al. 2015) and even cross-resistance to some of them (Andries et al. 2014). In 2013, the first novel vaccine candidate for prevention of TB infection has entered clinical trials (Tameris et al. 2013). The vaccine failed to enhanced protection compared to the current, almost century old BCG vaccine. However, the work showed that it is possible to carry out controlled TB vaccine clinical trials at a large scale. Many additional new vaccine candidates are in the pipeline, but none of them are expected to be licensed for clinical use anytime soon.

An additional dimension of global tuberculosis control is how to tackle on-going transmission. Part of the problem is that DOTS failed to deliver one of its primary aims. By treating patients early and providing full follow-up, DOTS was expected to limit both the emergence of drug resistance and transmission. But even under the umbrella of the most advanced public health systems in low-endemic countries, direct transmission of TB still accounts for up to 20% of cases, while in high-burden countries, it is the main contributor to disease incidence (up to 75% of cases) (Yates et al. 2016). As a result, even the success of DOTS limiting the number of new drug resistance cases is in danger as lack of control of transmission will leave room for MDR-TB to spread. Therefore, new epidemiological tools to inform new intervention strategies including advances towards real-time molecular epidemiology are needed to limit TB transmission. However, real-time molecular epidemiology in TB is a challenge. This is because the nature of an outbreak in TB is very special. One could even argue whether outbreaks in TB exists at all. First, most individuals infected do not develop TB disease. For the remaining, it can take months, years or even decades for active TB to develop. Contrary to most acute infections,

outbreaks of TB may span years before being noticed. For those cases, in which progression to active disease occurs within months there is room for infection control, but because the TB bacteria take between 2 and 4 weeks to grow in culture, by the time molecular epidemiological data is generated, it is already too late for an intervention. This is the reason why molecular epidemiology has been usually confined to retrospective studies rather than represent a tool for infection control in real-time. Yet, we badly need such a tool to cut transmission.

Whole genome sequencing (WGS) shows great potential for becoming the ideal tool to tackle TB transmission. WGS has been shown to resolve with greater resolution transmission clusters and overcome the limitations of conventional genotyping techniques by avoiding false assignments to transmission clusters (Walker et al. 2013). It is still not an ideal marker due to the slow pace of mutation accumulation, but when combined with classical epidemiological analysis, it becomes a powerful tool. Furthermore, due to the comprehensive nature of genomic data, it has already given important insights into different aspects of the biology of TB bacteria, including the emergence of drug resistance (Comas et al. 2012; Cohen et al. 2015), within-host bacterial variation (Eldholm et al. 2014), and responses to immune pressure (Comas et al. 2010). However, until now, the use of WGS has largely been confined to research environments, and in most cases, it has been applied retrospectively. In this chapter, I will review different aspects on the application of WGS to better understand the epidemiology of TB. In particular, I will review the use of WGS to resolve large TB outbreaks, define transmission clusters, monitor new interventions, and differentiate relapse from re-infection. Moreover, I will review recent insights gained from WGS into the biological factors driving the bacterial diversity within patients and between transmission cases, and how this diversity may impact epidemiological inference. I shall end by discussing the future role of WGS in the global control of TB.

4.2 Next-Generation DNA Sequencing Applied to the Tubercle Bacilli

In 1905, a strain of *Mycobacterium tuberculosis* known as H37 was isolated from an outbreak in the United States. That strain is today known as H37Rv and is the reference strain used worldwide for experimental work on *M. tuberculosis*. In 1998, H37Rv was among the first bacterial genomes to be fully sequenced, opening the era of TB genomics research (Cole et al. 1998). The genome of H37Rv and its experimental manipulation has represented a quantum leap on the development of new antibiotics, vaccines and diagnostics. It has paved the way for an improved understanding of the genetic determinants of clinical resistance to antibiotics, and helped defining regions appropriate for molecular typing or identifying virulence factors that now are the basis for many of the vaccine candidates under development.

In 1998, sequencing a bacterial genome required 2 years of work and cost around 4–5 million US dollars. Today, we could sequence *M. tuberculosis* H37Rv hundreds of times in less than a week and at price hundreds of thousands of times cheaper than in 1998. It is therefore not surprising that the whole genome sequence of bacterial pathogens is becoming the new gold standard used as an epidemiological marker (Loman and Pallen 2015). The first next-generation sequencing technology (also known as deep sequencing, massive sequencing and next generation sequencing) was introduced in 2005. Contrary to the previous techniques used for whole genome sequencing, known as shotgun sequencing, next generation sequencing technologies do not require a cloning step in *Escherichia coli*. The other main feature of these new technologies was the high throughput in terms of sequencing yield in a single run. Illumina platforms are the most widely used for bacterial genome sequencing today. The mean genome size of *M. tuberculosis* is around 4.4 million base pairs (bp). If one were to run a single strain of *M. tuberculosis* on an Illumina MiSeq platform with v3 kits, one

would read each nucleotide of the genome 3400 times, i.e. the coverage of the strain would be 3400×. With the NextSeq "rapid run" platform, the coverage would be around 27,200×, and on an Illumina HiSeq platform at full capacity 136,300×. On an Ion Torrent platform, one can read each base up to 455 times. Reading the same position of the genome hundreds of times is typically not necessary as for many studies we are interested in a coverage around 80×. Thus, all of these platforms generate much more data than what we really need to analyze a single genome. To take advantage of the high throughput of these new sequencing machines, several strains are usually multiplexed in a single run. The amount of multiplexing depends on the targeted coverage, but in theory one can sequence at 80× up to 1700 strain in a single run of HiSeq 1500/2500. To allow multiplexing for each strain, the genomic DNA is sheared into fragments of 500–1000 bp and specific nucleotide tags are added that can be retrieved during the bioinformatics analyses and used to assign each sequenced read to its corresponding strain. By multiplexing, one not only maximizes the number samples analyzed in one run, but one also reduces the time and the cost per sample. Multiplexing allows sequencing one strain for 80–100 euros if it is done in-house with recommended reagents. If customized reagents are used, the price can drop to below 50 euros, which is not more expensive than other methods for the molecular characterization of pathogens used in clinical microbiology units of hospitals in high-income countries.

The other main feature of next generation sequencing technologies that has to be taken into account is the length of the sequencing reads. Depending on the organism, one usually needs to choose a platform based on a trade-off between read length and overall throughput. In TB epidemiology, short read length technologies are mostly used that generate reads between 100 and 300 bp. There are several reasons for that. The throughput is much higher, and as we have seen above, this is essential for high multiplexing and lower price. In addition, for *M. tuberculosis,* the bioinformatic analysis does usually not reconstruct the strain genome (approach known as "de novo assembly") but maps the sequencing reads to a reference genome. This is because it is known that in terms of gene content and genome structure, all strains are very similar. Indeed, the average nucleotide identity is above 99% and thus many stretches of the genome are almost identical across strains. This reference mapping is known as re-sequencing and it is very useful to identify single nucleotide polymorphism (SNPs) among strains. The drawback is that short reads are difficult to map to repetitive regions and those regions must therefore be excluded from the analysis.

In more genetically diverse bacteria, the strategy of mapping to a reference can be misleading unless there is *a priori* knowledge that they have a very recent common ancestor (for example if they belong to the same outbreak than the reference genome). In the case of *E.coli,* up to 60% of the gene content may not be shared across strains, and a mapping to reference will therefore likely miss important information on the origin and other genomic characteristics of the strain of interest (Gordienko et al. 2013). In these cases, it is better to build a new genome from the sequencing reads based on those reads that are overlapping. In general, technologies that produce longer reads are more likely to generate a better assembly of the genome. These so-called third generation technologies like PacBio (Pacific Biosciences) not only generate reads between 10 and 15 kb, but also work with the DNA molecule directly, thus avoiding the PCR steps necessary when using the Illumina or Ion Torrent platforms. This is why PacBio and other technologies like Oxford Nanopore are considered single molecule technologies (Loman and Pallen 2015). In addition, single molecule sequencing can identify simultaneously the methylation pattern of the DNA molecule. The drawback is that because the throughput of third-generation technologies based on long reads is comparably low, one cannot multiplex many strains and thus the price per strain becomes too high for routine public health and diagnostics purposes. One common strategy is to combine both technologies so short-read sequenced strains can be mapped to accu-

rate assemblies of representative strains generated with long-read technologies. In summary, short-read, highly multiplexed technologies are currently preferred for molecular epidemiological studies or diagnostics of bacterial diseases including TB. In the future, technologies that offer simultaneously high multiplexing, long reads and cheap sequencing will likely replace the current combination of different technologies.

4.3 The Genome as an Epidemiological Marker

In TB, the first evidence that the whole genome sequencing can lead to a higher epidemiological resolution was the analysis of two strains from Uzbekistan (Niemann et al. 2009). Those strains had almost exactly the same MIRU profile and RFLP pattern, the two most commonly used epidemiological markers (reviewed in Chap. 3). The only known difference between these two strains was the drug sensitivity profile, as one was pan-susceptible and the other multidrug-resistant. Whole genome sequence data revealed a more complex picture with dozens of SNP differences between the strains. For the first time, it was shown that the genome can provide a greater resolution than other molecular markers. Moreover, this study provided first evidence that the *M. tuberculosis* complex is more diverse than previously anticipated, even at an epidemiological scale. This analysis was performed using one of the first next-generation sequencing platforms, the Solexa Genome Analyzer (known as Illumina today). Shortly thereafter, another technology, Roche 454, was used to analyze genomes from the first and last case of a suspected transmission cluster (Schurch et al. 2010). Again the genome showed a higher resolution compared to conventional methods.

In 2011, the first large-scale genome analysis of a TB outbreak was published, starting the era of genomic epidemiology for TB (Gardy et al. 2011). In this landmark publication, Gardy et al. analysed 36 cases belonging to an outbreak in British Columbia previously identified by MIRU and/or RFLP typing, and spanning several years. WGS was able to trace a clear picture about how the outbreak started. In particular, this analysis pointed to the importance of a superspreader in the outbreak. Superspeaders are patients that generate a disproportionally high number of secondary cases. Superspreaders rather than chains of transmission seem to be a common transmission topology in TB. Following this initial publication, other large outbreaks have been described, all of them have in common the complexity of the transmission network. In Hamburg, a large outbreak identified by MIRU-VNTR included 86 cases between 1996 and 2011 (Roetzer et al. 2013). However, not all cases were linked by epidemiological investigations, and before whole genome sequencing, the true source of the outbreak remained unknown. WGS analyses closed the gap between molecular and epidemiological data. Two different outbreaks with two closely related but distinct strains were involved. One was linked to many of the cases of the first years and the other was still on-going in 2010. This explained the lack of epidemiological links between the patients involved in this outbreak, and why the index case was so difficult to trace. Thus, with the high resolution obtained from whole genome sequences we can apply different evolutionary tools, including phylodynamics and phylogeography, to dissect the tempo and mode of transmission and drug resistance acquisition of successful *M. tuberculosis* clones (Eldholm et al. 2015).

Genomic analysis applied to specific strains of interest has also allowed to identify the true extend of an outbreak. This information can be of vital importance for public health authorities. Recent examples of that include TB outbreaks in Bern, Switzerland (Stucki et al. 2015a) and in Almeria, Spain (Perez-Lago et al. 2015). In the Bernese outbreak, Stucki et al. (2015a) applied WGS to three strains of an outbreak that started in Bern among homeless people, a typical high risk population for TB in high-income countries. These strains were thought to be representative of the diversity within the outbreak as they were chosen from different periods during the outbreak. By comparing to control strains, Stucki et al. (2015a) were able to identify single

nucleotide polymorphism (SNPs) that could be used to assign patients to the outbreak. Real-time PCR analyses designed based on these SNPs was then developed and applied to the retrospective collection of 1642 TB cases in the canton of Bern between 1991 and 2011. The analyses allowed assigning 68 new cases to the Bernese outbreak. The RFLP pattern between all these strains was almost identical. In contrast, WGS comparison revealed the true complexity of the outbreak. Transmission network reconstruction based on genetic data detected three central nodes in the topology of transmission that combined with epidemiological data allowed to detect two index cases that had infected many others, i.e. two superspreaders. The Bernese outbreak is an example of how WGS data can illuminate epidemiological investigations in TB. A similar approach was used by Pérez-Lago et al. (2015) to identify retrospectively and prospectively new cases due to an *M. tuberculosis* strain that had already led to many secondary cases in Almeria, Spain. The authors developed a SNP-typing approach similar to Stucki et al. (2015a, b) but based on a low-cost, low-tech, decentralized protocol with the aim to use it at the point-of-care or the closest referral center. The typing assay is called TRAP and can be run on a gel to quickly scan a large number of strain both from cultures and sputum samples.

These different works are examples of how genomic information can be used to identify transmission patterns, and how a technology that is not universally accessible can help to design other rapid and low-cost molecular assays. In summary, the application of WGS to specific outbreaks have shown its superiority when compared to conventional typing tools and have led to important new epidemiological insights (Walker et al. 2013).

4.4 Population Scale Analysis of TB Transmission Using WGS

Public Health England has led the way to implement WGS as both an epidemiological marker and as a diagnostic for public health systems. In a series of publications from 2012 to 2015, they have demonstrated the potential impact of WGS on the control of TB. In a landmark paper, Walker and collaborators (Walker et al. 2013a) published the first large-scale, population-based study of TB transmission based on WGS. In that study, the superiority WGS over previous strain typing methods was corroborated by identifying more accurately those cases belonging to an outbreak. At the same time, the authors sequenced serial isolates from the same patient and isolates from different body parts of the same patient. Based on these data, the authors proposed a threshold to identify a transmission link between two TB cases. Specifically, a genetic distance separating patient isolates of five or fewer SNPs was used to define high-confident transmission clusters, as most of the time, these clusters were also supported by epidemiological links. A genetic distance of between 5 and 12 SNPs was considered a cluster in which recent transmission was very likely but often not supported by epidemiological data. A genetic distance of more than 12 SNPs was defined to classify epidemiologically unrelated cases. Using those thresholds, a study in Switzerland has shown that standard genotyping (MIRU) overestimates the rate of transmission among immigrant (Stucki et al. 2015a). This is likely due to the high genetic similarity of strains circulating in high-burden countries. In addition, it is important to remember that identification of transmission clusters based on WGS data does not necessarily imply that transmission has occurred at the place of study. For example, transmission between immigrants may have occurred at their country of origin, followed by progression to active disease at the country of residence. This is best exemplified by the detection of transcontinental spread of *M. tuberculosis* clones like in the published case of Thai refugees in California (Coscolla et al. 2015).

The SNP thresholds described above have been corroborated by other studies, but most examples are from low-burden countries (Bryant et al. 2013b; Hatherell et al. 2016). The complexity of the global TB epidemic suggests that the same threshold may not apply equally to all

epidemiological settings. For example, in high-burden countries where there is a high rate of on-going transmission, several genetically similar clones may act as index cases of different transmission clusters, and thus differentiating between those clusters will be difficult (Yates et al. 2016). The only example from a high-burden country we have until now is the analysis of 1687 TB patient isolates collected in rural Malawi between 1995 and 2010 (Guerra-Assunção et al. 2015a). Genetic distances and genetic network analyses showed consistency with the thresholds described by Walker et al. (2012). However, this is just one case, and it is likely that the situation may change in large urbanized African settings. Similarly, areas with a high burden of MDR-TB might also show different patterns. Due to the selection of drug resistance mutations, we can expect a higher number of mutations between epidemiologically linked strains as a result of genetic hitchhiking effects (further discussed in the following section) (Sun et al. 2012; Eldholm et al. 2014; Liu et al. 2015). For example, by looking at serial isolates of a single patient Eldholm et al. (2015) identifies more SNPs separating the isolates than the number expected between two transmission cases. To date, only one large population-based study has been published from a high- MDR-TB burden setting in Russia (Casali et al. 2014). The study included more than one thousand patient isolates, 50% of which were MDR. Most of these MDR isolates carried the mutation S450L in the *rpoB* gene. This mutation is generally the most common mutation conferring resistance to rifampicin. In addition, large transmission clusters also carried additional fitness compensatory mutations, mainly in the *rpoC* gene (Comas 2012), partially explaining the high transmissibility of these strains in the region. Unfortunately, because the associated epidemiological data was not published, it is difficult to evaluate SNP thresholds levels to delineate transmission clusters in this Russian dataset.

The epidemiological significance of SNP thresholds remains a matter of ongoing research. On the one hand, we lack data from different epidemiological settings. On the other hand, calling SNPs from next generation sequencing data is not straightforward (O'Rawe et al. 2013). There are multiple steps in the bioinformatics analyses that can introduce false positive or false negative SNPs. It is thus very important to implement and follow multiple quality control checkpoints. In particular, mobile elements and repetitive regions of the *M. tuberculosis* genome are difficult to interrogate with short read length technologies such as Illumina. In a typical analysis, most of these loci are excluded. In addition, other loci can also be problematic. For example, scars from insertion elements that led to incorrect mapping of reads and unknown deletions and/or insertions can complicate the analysis. Importantly, the parameters used for mapping and SNP calling are critical. Finally, initial quality of the sequencing data is key. Given that epidemiological inferences are based in small number of SNPs, it is recommended to corroborate results independently with a different analysis pipeline and/or by laboratory confirmation of a subset of SNPs. In addition, Illumina technology is currently the main platform used in genomic epidemiology of TB, but we do not yet know the potential epidemiological impact of using long-read sequencing platforms (Quail et al. 2012). These platforms, although expensive, should allow identifying SNPs in regions of the genome that are not accessible by short-read sequencing technologies. While some of these regions, like the PE/PPE genes, are clearly the most variable of the *M. tuberculosis* genome (Copin et al. 2014), they are also likely involved in gene conversion events and/or recombination with external sources (Phelan et al. 2016). Thus, even if interrogated with the appropriate technology, we first will need to understand if and how these loci can be exploited for epidemiological purposes.

4.5 Role of Within-Host Diversity in Transmission Inference

As in any other pathogen, the bacteria belonging to the *M. tuberculosis* complex are in constant evolution, and it is therefore not surprising that

the bacterial population infecting a patient is not genetically homogeneous (Pérez-Lago et al. 2013). Understanding this within-host diversity is crucial, as it can have important consequences for drug resistance diagnostics, epidemiology and disease outcomes (Didelot et al. 2016). One important open question is what is the magnitude of within-host diversity during TB infection? This question has been mainly studied in the context of drug resistance because of its clinical importance. Historically, clinical microbiologists have recognized the phenomenon of heteroresistance, which manifests in discordant results of drug susceptible testing using repeat testing or different isolates from the same patient (Van Rie et al. 2005). Heteroresistance suggests that there are two co-existing populations in the patient, one drug-resistant and the other susceptible to a particular drug. WGS has the potential to identify the genotypes of these co-existing populations. Moreover, we can follow how the relative frequency of these genotypes changes over time, based on serial samples from the same patient (Sun et al. 2012, Eldholm et al. 2015). The precision of this technique is such that we can detect drug resistance minority variants in an otherwise homogenous population at the 10% or even 5% level when the sequencing coverage is high enough. The drawback is that many of these minority variants associated with drug resistance are identified based on cultured specimens, and thus we cannot easily trace back their origin in the lung. However, PET-CT scans have been able to correlate different lesions in the lung with different *M. tuberculosis* genotypes obtained from cultured isolates, suggesting that micro-geography of the lungs and the lung lesions are important for the selection of genetically distinct sub-populations (Liu et al. 2015).

The co-existence of several bacterial clones within an individual patient has been demonstrated beyond the case of heteroresistance (Pérez-Lago et al. 2013). But what is the origin of such genetic variation? Recent efforts using whole genome data have explored this variation both in vivo and in vitro (see Chap. 13). The mean whole genome mutation rate as derived from experiments with cynomolgus macaques is around 0.39 (0.16–0.80 95% CI) SNPs per genome per year (Ford et al. 2011). This rate was found to be similar between macaques that developed active disease and those that remained latent although the number of SNPs analyzed were too low to allow for strong statistical conclusions. In fact, given that the generation time of the bacteria is thought to be longer during latency, the authors concluded that the mutation rate was higher during latency than during active disease (Ford et al. 2011). The same research group was later able to define the in vitro rate of mutation of different strains belonging to different lineages in the presence and absence of different antibiotics (Ford et al. 2013). This in vitro rate was fairly similar to the rate seen in vivo in macaques, and slightly higher for Lineage 2 compared to Lineage 4.

However, the mutation rate of bacteria measured in short-term experiments is not necessarily the same than the substitution rate we observe in a clinical setting. This is because in addition to the rate of mutation generation, we have to consider the action of evolutionary forces. Most of the mutations arising de novo are either neutral or deleterious. Neutral mutations can increase in frequency unnoticed, whereas deleterious mutations will be removed, more or less efficiently, by natural purifying selection. Thus, the action of evolutionary forces uncouples the intrinsic bacterial mutation rate from the substitution rate. Importantly, the substitution rate is time-dependent because natural selection needs time to act. Hence, the shorter natural selection can act, the closer the substitution rate will be to the intrinsic mutation rate (Rocha et al. 2006; Biek et al. 2015). This explains why the long-term substitution rate of many pathogens is much lower than the mutation rate or the rate measured over short periods of time. At the epidemiological level, the substitution rate can be derived by comparing the number of differences between two transmission cases and the difference in time of diagnosis between both cases. This approach has been used to estimate the short-term substitution rate of *M. tuberculosis* in clinical settings. Findings from several studies converged in a substitution rate of 0.3–0.5 SNPs

per genome per year, thus very close to the in vivo and in vitro mutation rate discussed above. However, the variance around that estimate is so high that it has to be interpreted cautiously (Bryant et al. 2013b). One of the main "known unknowns" likely to impact our inferences of mutation- and substitution rates is latency. To date, only one study has used time of infection rather than time of diagnosis to measure the substitution rate in *M. tuberculosis*. The study showed that the substitution rate during latency was much lower than during active TB (Colangeli et al. 2014). Much more data is needed to draw conclusions about the substitution rate throughout the life cycle of *M. tuberculosis*. In the meantime, we are limited to model the substitution rate based on estimates during infection and transmission.

How is the mutation rate modulated by different evolutionary forces within the host? Again, this is best exemplified in the context of antibiotic treatment. As antibiotics are a strong selective force, the mutations causing drug resistance are positively selected. Because *M. tuberculosis* is clonal, other mutations present in a particular genetic background experiencing such positive selection are also selected, even when they have nothing to do with antibiotic resistance. This phenomenon is called genetic hitchhiking and have been elegantly described in several publications (Sun et al. 2012; Eldholm et al. 2014). In this context, the drug resistance conferring mutation can also be referred to as "driver" mutations, and the hitchhiking mutations as "passenger" mutations. However, it is also important to understand the bacterial variation expected in a patient that does not develop drug resistance. In these cases, purifying selection is mainly acting in the form of antibiotic purification (Black et al. 2015). However, variation maybe present because of neutral processes (e.g. genetic drift) or because of selection due to an unidentified evolutionary force. Thus, as a consequence of the equilibrium between different evolutionary forces, the amount of bacterial variation seen in different patients may vary substantially. For example, a recent study has shown that bacterial variation can be observed in most consecutive sputum samples obtained from a given patient (Pérez-Lago et al. 2013). By contrast, other studies have reported no genetic differences between bacterial samples from the same patient, even when the cultures were separated by more than 1 year (Pérez-Lago et al. 2015). The causes of these discrepancies are not well understood, but they might be linked to differences in treatment efficacy, variation in the site of infection, or the differential control of the bacterial load by the immune system. However, what is clear is that within-host bacterial diversity exists both inside and outside the context of drug resistance. From a practical standpoint, the potential overlap between within-patient diversity and the diversity seen between transmission cases will complicate the epidemiological interpretation (Pérez-Lago et al. 2013). From a biological standpoint it is important to identify the small fraction of SNPs detected that are linked to natural selection pressures apart from antibiotics.

4.6 Special Cases of Within-Host Diversity: Relapse, Re-infection and Co-infection

One major outcome measured by clinical trials of new drugs and drug regimens is the number of relapses occurring after successful completion of the treatment (Johnston et al. 2015). In high-burden countries, it is difficult to differentiate between a true case of relapse, indicating failure of the drug/regimen, or reinfection after treatment completion. Thus, molecular tools that can easily identify these two situations are key to evaluate new interventions. Several recent examples have highlighted the potential of WGS data to distinguish between relapses and secondary infections (Bryant et al. 2013a; Guerra-Assunção et al. 2015b). However, these studies also highlight the complexity associated to the interpretation of the data.

As discussed before, a difference of 10–12 SNPs is often used as the cutoff to define a transmission link. This same cutoff can be used to determine if two TB episodes in the same

patient are due to the same strain or not. Applying this logic, Bryant et al. (2013a) identified the number of true relapses *versus* the number of re-infections in the context of a clinical trial of a new treatment regimen (ReMOX-TB) (Gillespie et al. 2014). Forty-seven patients had a second episode of TB after completing treatment. Using the SNP-threshold approach, the authors found that 33 of the 47 patients had less than 12 SNPs between the first and the second isolate, indicating cases of true relapse. Only three cases harboured very different strains, indicated by a pairwise difference of more than 1000 SNPs; accordingly, these cases were classified as re-infections. In addition, the authors also reported a mix of two strains during the second episode of one patient. One of the isolate was almost identical to the isolate of the first episode while the other was clearly different. This single patient therefore likely represents case of relapse due to the presence of the original strain, plus a re-infection as a new strain was detected in the second episode. Similar results were obtained after analysis of relapse cases in a long-term epidemiological study carried out in Malawi (Guerra-Assunção et al. 2015b). However, there are limitations detecting such mixed infections. Identifying a re-infection when the second strain belongs to a different lineage is straightforward, as the strains will be separated by hundreds or thousands of SNPs. However, in high-burden countries, where many of the strains circulating are closely related, classifying strains as "the same" or "different" is much more challenging (Guerra-Assunção et al. 2015b).

WGS data also gives us the opportunity to detect co-infection in a single sample, i.e. cases in which two genetically distinct *M. tuberculosis* strains (more than 12 SNPs) co-exist in the same patient. These cases where difficult to identify until now as one had to rely on e.g. identifying evidence of mixed number of alleles in MIRU loci (see Chap. 3). However, this approach tends to miss many instances of co-infection, as single locus data has low resolution. By contrast, because WGS achieves high coverage, the presence of two different isolates in the same culture becomes evident when looking at SNP positions where the reference allele and the variant allele coexist as low frequency variants (i.e. none of these alleles are fixed in the population). SNP positions showing this mix of reference and variant allele are known as "heterozygous calls", and are indicative of two different sub-populations present in the culture. Heterozygous calls maybe related to clonal diversification from a single infecting strain within the lung of a patient. In this case, the number of heterozygous calls is expected to be low given the low mutation rate of *M. tuberculosis*. On the other hand, if two different strains infected the same patient, the number of heterozygous calls will be higher. Thus, as in the case of relapse and re-infection, the number of heterozygous calls detected can be used to differentiate between these two scenarios. To be compatible with clonal diversification, the number of heterozygous calls must be below the 12 SNP threshold used to identify transmission clusters. If more than 12 SNPs are seen, one can safely assume that two different strains are co-infecting the same patient. However, although in theory differentiating between clonal diversification and co-infection is straight forward, interpreting heterozygous calls is very challenging (Hatherell et al. 2016). Due to errors in mapping and SNP calling, much noise can be introduced during the bioinformatics analyses. As a result, it is often difficult to differentiate between true and false heterozygous calls. This is particularly relevant when the sequencing is done directly from the diagnostic sample, in which DNA from unrelated organisms (i.e. contaminants and/or commensals) might be present. This contaminating DNA maybe sequenced together with the DNA of interest, which can contribute to a percentage of the heterozygous calls. In practice, bioinformatics analyses can easily detect a co-infection when the number of heterozygous calls is larger than 100, and clonal diversification when the number is below 12; however, anything in between, is difficult to interpret (Guerra-Assunção et al. 2015b). In summary, high-throughput WGS is an ideal tool to identify co-existing variants, but more work is needed to develop analytical tools than can reveal co-infection when the two infecting strains are evolutionary close (between 20 and

100 SNPs). This is especially relevant in high-burden countries where transmission is often due to highly similar strains (Kay et al. 2015).

4.7 Reconstructing Transmission

The high variance in terms of genetic changes accumulated over time seen during both between host transmission and within a single patient leads to a very weak correlation between time and accumulated sequence diversity in *M. tuberculosis* (Bryant et al. 2013b). This weak correlation between time and accumulated number of SNPs is known as overdispersion of the molecular clock. A practical consequence is that the low correlation limits our ability determine the exact time of infection using genetic data (Hatherell et al. 2016). Only when large transmission clusters spanning several years or even decades are analyzed, researchers have been able to correlate the dating based on genomic substitution data with the epidemiological records (Roetzer et al. 2013). Thus, new analytical approaches are needed to model the genetic diversity seen during infection and transmission, and can use the information to build accurate genealogies on how and when transmission has occurred. Several such approaches have recently been developed, although not all of them have been tested in large-scale, population-based studies. Most approaches, like Outbreaker (Jombart et al. 2014), use the genetic data to infer the most likely network of transmission between isolates of a cluster, minimizing the number of genetic changes between isolates. However, as usually happens for this kind of approximations, clusters must be defined a priori, meaning that the researcher needs to select those strains suspected to be involved in a transmission group. As a consequence, transmission can only be defined among suspected cases, limiting the de novo discovery of cases linked to transmission in population-based samples. In addition, the "carriage" state represented by the latency period is usually not explicitly modeled. Hence, these approximations are better suited for acute diseases where infection, disease and transmission are directly linked (Jombart et al. 2014). Didelot et al. (2013) have developed a new way of interpreting transmission in bacterial infections including TB. Their approximation is based on using a dated phylogenetic topology, i.e. one that incorporates epidemiological data like date of diagnosis, and convert it into transmission events in a Bayesian framework. Furthermore, other epidemiological data like geography can be incorporated to delineate the most likely epidemiological scenario. The approach takes into account the amount of genetic diversity that may have accumulated within the patient during latency. It may be the case that within-host diversity can be neglected in most of the cases, particularly if the first isolate of each patient is analyzed. However overall, there remains a general lack of data to evaluate the true role of within-patient diversity. Certainly in the context of drug resistance, one expects a potentially misleading effect of this diversity on epidemiological inference (Eldholm et al. 2015).

In parallel, tools to understand the epidemic dynamics from WGS data are being applied to *M. tuberculosis* (see Chap. 15 for a review). Based on the principles of Bayesian phylodynamics, those approximations allow to infer important epidemiological parameters like incidence, prevalence, and changes in the basic reproductive number (R0) over time (Grenfell et al. 2004; Stadler et al. 2012). The integration of the phylodynamic and epidemiological frameworks is a research field in constant development, and thus different approaches are available, ranging from classical coalescence to birth-death models (Stadler et al. 2012). Compartmental epidemiological models like SIR (susceptible–infected–removed) (Rasmussen et al. 2011; Kühnert et al. 2014) can be also incorporated as well as geographical data (du Plessis and Stadler 2015). These different approaches are based on different assumptions, and it remains to be seen which one will perform best in the case of TB (Stadler et al. 2012). Again, the effect of latency and how best to model it remains a "black box" that needs to be addressed in the future. Nevertheless, these efforts will play an important role in the future, as we will be able to evaluate the impact of new interventions using

genetic data as well as predict epidemic trends based on the evolution of R0 over time.

4.8 Challenges of Genomic Epidemiology

WGS is becoming the new gold standard for molecular epidemiology of TB. It provides higher resolution than previous molecular markers, which allows for both a better delineation of transmission clusters and the potential to establish the direction of transmission. However, we still have many knowledge gaps to resolve (van Soolingen 2014).

4.8.1 Understanding the Biology

To properly model what happens within a patient, we need to understand how bacterial genetic diversity is generated within a patient and how much of it is transmitted. Linked to that, we need to determine the contribution of latency to the diversity seen, and how to account for it in transmission models. Interpretation of transmission network is also not straightforward, and while algorithms have been developed, they are not always easy to implement. Moreover, they still have to be evaluated at a larger scale. This is particularly true if we want to move towards routine WGS in the clinical environment.

4.8.2 Beyond Distance Thresholds

Since the publication of Walker et al. (2012), using SNP threshold has become the rule to define transmission clusters. However, these thresholds are likely not universal, and have to be tested in a wider range of epidemiological settings. For example, due to the action of natural selection, it is very likely that in high MDR-TB burden settings, we will find epidemiologically linked cases with more than 5 or 12 SNPs. Furthermore, distance thresholds are useful but can be misleading. The absence of an intermediate transmission link, an un-sampled index case for example, may separate two strains that are in fact epidemiologically related. As we have seen, algorithms that capture the diversity and complexity of transmission trees are being developed. These methods should be tested at a larger scale to understand how to interpret transmission at the population level.

4.8.3 Towards Routine Genomic Epidemiology

Interpretation of WGS data is not straightforward, and remains mostly confined to the research settings. Efforts have been done in certain places, most notably Public Health England/NHS and the US CDC, but we are still far from democratizing WGS among medical institutions and across countries. Implementation of WGS must also be integrated in local, regional, and national health systems, and the results shared across the globe to accelerate diagnostic and epidemiological research (Yozwiak et al. 2015). However, the reality in high-burden countries is very different. This is best illustrated by the low number of genomic epidemiological studies published to date from high-burden regions. In most of these countries, the only routine diagnostic remains sputum microscopy, and there is usually no access to bacterial cultures. Thus, the genomic revolution will likely take time to get to these high-burden, low-income countries given that the benefits remain limited.

4.8.4 Genomic Sequencing from Diagnostic Samples

The real genomic revolution in TB clinical practice will be the direct sequencing from complex sputum samples. The advantages of such an approach are multiple. It will eliminate the need for culture, generate a positive diagnostic result in less than a week or perhaps even hours, and will be used simultaneously for infection control. However, the challenges of obtaining an *M. tuberculosis* genome from a sputum samples are enormous. Some advances have been achieved recently, for example by performing

diagnostic metagenomics (Doughty et al. 2014). This approach consists in isolating total DNA from sputum and subject it to direct sequencing. However, for samples with low bacterial load, *M. tuberculosis* enrichment strategies must be applied (Brown et al. 2015, Votintseva et al. 2017). An alternative is targeted sequencing of certain regions of the genome, including known drug resistance loci, an approach also referred to as high-throughput amplicon sequencing (Colman et al. 2015). While still in early days, there is no doubt that new approaches will be developed, some probably linked to real-time, portable genomic technologies like those based on nanopores (Bradley et al. 2015, Votintseva et al. 2017) that have been successfully applied during the recent Ebola epidemic (Quick et al. 2016).

4.9 Practical Implications of Genomic Epidemiology

WGS has the potential to, and to some extent already is, revolutionizing molecular epidemiology of TB. Once more powerful analytical tools are developed, we will be able to establish a high resolution picture of TB transmission in different epidemiological settings, which will help developed tailored control strategies. Such a high resolution picture will allow determining the role of different host, bacterial and environmental factors in the transmission of TB. In addition, together with geopositioning data, we will be able to develop new tools for spatial epidemiology of TB. Thus, genomic epidemiology will improve our view not only of the various drivers of TB transmission, but also of the specific foci of transmission. If factors driving TB transmission can be established, particularly those involving human and bacterial variation, new avenues of research will open and contribute to a better understanding of the complex interplay between the host and the pathogen. In addition, WGS offers the opportunity to identify the genetic changes and the evolutionary forces behind infection, disease and transmission. Indeed, the study of mutations associated with drug resistance is currently leading the way, but we need to identify the genetic loci of both the host and the pathogen that are under different selective pressures. If we are able to identify such loci, a whole new opportunity for research will be created for the development of new drugs, vaccines and host directed therapies. Finally, if we are able to position WGS analysis close to point-of-care and to the diagnostic sample, we will transition from a tool used in retrospective studies to a tool used for infection control and monitoring treatment efficacy in real-time. While there is still a lot of work to do in terms of standardization and interpretation, genomic epidemiology can have a central role on the new strategies for global TB control and help pave the way towards TB elimination.

Acknowledgements I thank the members of my group for stimulating discussions. Work in my laboratory is supported by the Spanish National Foundation (MINECO SAF2013-43521-R) and the European Research Council (638553-TB-ACCELERATE).

References

Andries K, Villellas C, Coeck N et al (2014) Acquired resistance of *Mycobacterium tuberculosis* to bedaquiline. PLoS One 9:e102135

Biek R, Pybus OG, Lloyd-Smith JO, Didelot X (2015) Measurably evolving pathogens in the genomic era. Trends Ecol Evol 30:306–313

Black PA, de Vos M, Louw GE et al (2015) Whole genome sequencing reveals genomic heterogeneity and antibiotic purification in *Mycobacterium tuberculosis* isolates. BMC Genomics 16:857

Bloemberg GV, Keller PM, Stucki D et al (2015) Acquired resistance to bedaquiline and delamanid int for tuberculosis. N Engl J Med 373:1986–1988

Bradley P, Gordon NC, Walker TM et al (2015) Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and Mycobacterium tuberculosis. Nat Commun 6:10063

Brown AC, Bryant JM, Einer-Jensen K et al (2015) Rapid whole-genome sequencing of *Mycobacterium tuberculosis* isolates directly from clinical samples. J Clin Microbiol 53:2230–2237

Bryant JM, Harris SR, Parkhill J et al (2013a) Whole-genome sequencing to establish relapse or re-infection with Mycobacterium tuberculosis: a retrospective observational study. Lancet Respir Med 1:786–792

Bryant JM, Schürch AC, van Deutekom H et al (2013b) Inferring patient to patient transmission of *Mycobac-*

terium tuberculosis from whole genome sequencing data. BMC Infect Dis 13:110
Casali N, Nikolayevskyy V, Balabanova Y et al (2014) Evolution and transmission of drug-resistant tuberculosis in a Russian population. Nat Genet 46:279–286
Cohen K, Abeel T, Manson McGuire A et al (2015) Evolution of extensively drug-resistant tuberculosis over four decades: whole genome sequencing and dating analysis of *Mycobacterium tuberculosis* isolates from KwaZulu-Natal. PLoS Med 12:1–22
Colangeli R, Arcus VL, Cursons RT et al (2014) Whole genome sequencing of *Mycobacterium tuberculosis* reveals slow growth and low mutation rates during latent infections in humans. PLoS One 9:e91024
Cole ST, Brosch R, Parkhill J et al (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393:537–544
Colman RE, Schupp JM, Hicks ND et al (2015) Detection of low-level mixed-population drug resistance in *Mycobacterium tuberculosis* using high fidelity amplicon sequencing. PLoS One 10:e0126626
Comas I, Gagneux S (2009) The past and future of tuberculosis research. PLoS Pathog 5:e1000600
Comas I, Chakravartti J, Small PM et al (2010) Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. Nat Genet 42: 498–503
Comas I, Borrell S, Roetzer A et al (2012) Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. Nat Genet 44:106–110
Copin R, Coscolla M, Seiffert SN et al (2014) Sequence diversity in the pe_pgrs genes of *Mycobacterium tuberculosis* is independent of human T cell recognition. MBio 5:e00960–e00913
Coscolla M, Barry PM, Oeltmann JE et al (2015) Genomic epidemiology of multidrug-resistant *Mycobacterium tuberculosis* during transcontinental spread. J Infect Dis 15:383–388
Didelot X, Gardy J, Colijn C (2013) Bayesian inference of infectious disease transmission from whole genome sequence data. Mol Biol Evol 31:1869–1879
Didelot X, Walker AS, Peto TE et al (2016) Within-host evolution of bacterial pathogens. Nat Rev Microbiol 14:150–162
Doughty EL, Sergeant MJ, Adetifa I et al (2014) Culture-independent detection and characterisation of Mycobacterium tuberculosis and *M. africanum* in sputum samples using shotgun metagenomics on a benchtop sequencer. PeerJ 2:e585
du Plessis L, Stadler T (2015) Getting to the root of epidemic spread with phylodynamic analysis of genomic data. Trends Microbiol 23:383–386
Dye C, Glaziou P, Floyd K, Raviglione M (2013) Prospects for tuberculosis elimination. Annu Rev Public Health 34:271–286
Eldholm V, Norheim G, von der Lippe B et al (2014) Evolution of extensively drug-resistant *Mycobacterium tuberculosis* from a susceptible ancestor in a single patient. Genome Biol 15:490
Eldholm V, Monteserin J, Rieux A et al (2015) Four decades of transmission of a multidrug-resistant *Mycobacterium tuberculosis* outbreak strain. Nat Commun 6:7119
Ford CB, Lin PL, Chase MR et al (2011) Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. Nat Genet 43:482–486
Ford CB, Shah RR, Maeda MK et al (2013) *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. Nat Genet 45:784–790
Gardy JL, Johnston JC, Sui SJH et al (2011) Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N Engl J Med 364:730–739
Gillespie SH, Crook AM, McHugh TD et al (2014) Four-month moxifloxacin-based regimens for drug-sensitive tuberculosis. N Engl J Med 371:1577–1587
Gordienko EN, Kazanov MD, Gelfand MS (2013) Evolution of pan-genomes of *Escherichia coli*, Shigella spp., and Salmonella enterica. J Bacteriol 195:2786–2792
Grenfell BT, Pybus OG, Gog JR et al (2004) Unifying the epidemiological and evolutionary dynamics of pathogens. Science 303(80):327–332
Guerra-Assunção JA, Crampin AC, Houben RMGJ et al (2015a) Large-scale whole genome sequencing of M. tuberculosis provides insights into transmission in a high prevalence area. elife 4:1–17
Guerra-Assunção JA, Houben RMGJ, Crampin AC et al (2015b) Recurrence due to relapse or reinfection with *Mycobacterium tuberculosis*: a whole-genome sequencing approach in a large, population-based cohort with a high HIV infection prevalence and active follow-up. J Infect Dis 211:1154–1163
Hatherell H-A, Colijn C, Stagg HR et al (2016) Interpreting whole genome sequencing for investigating tuberculosis transmission: a systematic review. BMC Med 14:21
Johnston JC, Khan FA, Dowdy DW (2015) Reducing relapse in tuberculosis treatment: is it time to reassess WHO treatment guidelines? Int J Tuberc Lung Dis 19:624
Jombart T, Cori A, Didelot X et al (2014) Bayesian eeconstruction of disease outbreaks by combining epidemiologic and genomic data. PLoS Comput Biol 10:e1003457
Kay GL, Sergeant MJ, Zhou Z et al (2015) Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe. Nat Commun 6:6717
Kühnert D, Stadler T, Vaughan TG, Drummond AJ (2014) Simultaneous reconstruction of evolutionary history and epidemiological dynamics from viral sequences with the birth-death SIR model. J R Soc Interface 11:20131106
Liu Q, Via LE, Luo T et al (2015) Within patient microevolution of *Mycobacterium tuberculosis* correlates with heterogeneous responses to treatment. Sci Rep 5:17507

Loman NJ, Pallen MJ (2015) Twenty years of bacterial genome sequencing. Nat Rev Microbiol:1–9

Niemann S, Köser CU, Gagneux S et al (2009) Genomic diversity among drug sensitive and multidrug resistant isolates of *Mycobacterium tuberculosis* with identical DNA fingerprints. PLoS One 4:e7407

O'Rawe J, Jiang T, Sun G et al (2013) Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. Genome Med 5:28

Pai M, Schito M (2015) Tuberculosis diagnostics in 2015: landscape, priorities, needs, and prospects. J Infect Dis 211(Suppl):S21–S28

Pérez-Lago L, Comas I, Navarro Y et al (2013) Whole genome sequencing analysis of intrapatient microevolution in *Mycobacterium tuberculosis*: potential impact on the inference of tuberculosis transmission. J Infect Dis:1–11

Perez-Lago L, Martinez Lirola M, Herranz M et al (2015) Fast and low-cost decentralized surveillance of transmission of tuberculosis based on strain-specific PCRs tailored from whole genome sequencing data: a pilot study. Clin Microbiol Infect 21: 249.e1–249.e9

Pérez-Lago L, Navarro Y, Montilla P et al (2015) Persistent infection by a *Mycobacterium tuberculosis* strain that was theorized to have advantageous properties, as it was responsible for a massive outbreak. J Clin Microbiol 53:3423–3429

Phelan JE, Coll F, Bergval I et al (2016) Recombination in pe/ppe genes contributes to genetic variation in *Mycobacterium tuberculosis* lineages. BMC Genomics 17:151

Quail M, Smith ME, Coupland P et al (2012) A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers. BMC Genomics 13:1

Quick J, Loman NJ, Duraffour S et al (2016) Real-time, portable genome sequencing for Ebola surveillance. Nature 530:228–232

Rasmussen DA, Ratmann O, Koelle K (2011) Inference for nonlinear epidemiological models using genealogies and time series. PLoS Comput Biol 7:e1002136

Rocha EPC, Smith JM, Hurst LD et al (2006) Comparisons of dN/dS are time dependent for closely related bacterial genomes. J Theor Biol 239: 226–235

Roetzer A, Diel R, Kohl TA et al (2013) Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study. PLoS Med 10:e1001387

Schurch AC, Kremer K, Daviena O et al (2010) High resolution typing by integration of genome sequencing data in a large tuberculosis cluster. J Clin Microbiol 48:3403–3406

Stadler T, Kühnert D, Bonhoeffer S, Drummond AJ (2012) Birth – death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proc Natl Acad Sci U S A 110:228–233

Sterling TR, Lehmann HP, Frieden TR (2003) Impact of DOTS compared with DOTS-plus on multidrug resistant tuberculosis and tuberculosis deaths: decision analysis. BMJ 326:574

Stucki D, Ballif M, Bodmer T et al (2015a) Tracking a tuberculosis outbreak over 21 years: strain-specific single-nucleotide polymorphism typing combined with targeted whole-genome sequencing. J Infect Dis 211:1306–1316

Stucki D, Ballif M, Egger M et al (2015b) Standard genotyping overestimates transmission of Mycobacterium tuberculosis among immigrants in a low-incidence country. J Clin Microbiol 7:1862–1870

Sun G, Luo T, Yang C et al (2012) Dynamic population changes in *Mycobacterium tuberculosis* during acquisition and fixation of drug resistance in patients. J Infect Dis 206:1724–1733

Tameris MD, Hatherill M, Landry BS et al (2013) Safety and efficacy of MVA85A, a new tuberculosis vaccine, in infants previously vaccinated with BCG: a randomised, placebo-controlled phase 2b trial. Lancet 6736:1–8

Van Rie A, Victor TC, Richardson M et al (2005) Reinfection and mixed infection cause changing *Mycobacterium tuberculosis* drug-resistance patterns. Am J Respir Crit Care Med 172:636–642

Van Soolingen D (2014) Whole-genome sequencing of *Mycobacterium tuberculosis* as an epidemiological marker. Lancet Respir Med 4:251–252

Votintseva AA, Bradley P, Pankhurst L, Del Ojo Elias C, Loose M, Nilgiriwala K, Chatterjee A, Smith EG, Sanderson N, Walker TM, Morgan MR, Wyllie DH, Walker AS, Peto TEA, Crook DW, Iqbal Z (2017) Same-day diagnostic and surveillance data for tuberculosis via whole-genome sequencing of direct respiratory samples. J Clin Microbiol 55(5): 1285–1298

Walker TM, Ip CL, Harrell RH et al (2013a) Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. Lancet Infect Dis 13:137–146

Walker TM, Monk P, Smith EG, Peto TE a (2013b) Contact investigations for outbreaks of Mycobacterium tuberculosis: advances through whole genome sequencing. Clin Microbiol Infect 19:796–802

World Health Organization (2015). Global tuberculosis report

Yates TA, Khan PY, Knight GM et al (2016) The transmission of *Mycobacterium tuberculosis* in high burden settings. Lancet Infect Dis 16:227–238

Yozwiak NL, Schaffner SF, Sabeti PC (2015) Data sharing: make outbreak research open access. Nature 518:477–479

Zumla A, Memish ZA, Maeurer M et al (2014) Emerging novel and antimicrobial-resistant respiratory tract infections: new drug development and therapeutic options. Lancet Infect Dis 14:1136–1149

Biological and Epidemiological Consequences of MTBC Diversity

5

Mireia Coscolla

Abstract

Tuberculosis is caused by different groups of bacteria belonging to the *Mycobacterium tuberculosis* complex (MTBC). The combined action of human factors, environmental conditions and bacterial virulence determine the extent and form of human disease. MTBC virulence is a composite of different clinical phenotypes such as transmission rate and disease severity among others. Clinical phenotypes are also influenced by cellular and immunological phenotypes. MTBC phenotypes are determined by the genotype, therefore finding genotypes responsible for clinical phenotypes would allow discovering MTBC virulence factors. Different MTBC strains display different cellular and clinical phenotypes. Strains from Lineage 5 and Lineage 6 are metabolically different, grow slower, and are less virulent. Also, at least certain groups of Lineage 2 and Lineage 4 strains are more virulent in terms of disease severity and human-to-human transmission. Because phenotypic differences are ultimately caused by genotypic differences, different genomic loci have been related to various cellular and clinical phenotypes. However, defining the impact of specific bacterial genomic loci on virulence when other bacterial determinants, human and environmental factors are also impacting the phenotype would contribute to a better knowledge of tuberculosis virulence and ultimately benefit tuberculosis control.

Keywords

MTBC diversity • Polymorphisms • Virulence • Cellular phenotypes

5.1 Introduction

Tuberculosis (TB) in humans is mostly caused by the members of the *Mycobacterium tuberculosis* complex (MTBC) known as *Mycobacterium tuberculosis* sensu stricto and *Mycobacterium*

M. Coscolla (✉)
Swiss Tropical and Public Health Institute, University of Basel, Socinstrasse 57, 4002, Basel, Switzerland
e-mail: mireia.coscolla@unibas.ch

S. Gagneux (ed.), *Strain Variation in the* Mycobacterium tuberculosis *Complex: Its Role in Biology, Epidemiology and Control*, Advances in Experimental Medicine and Biology 1019,
DOI 10.1007/978-3-319-64371-7_5

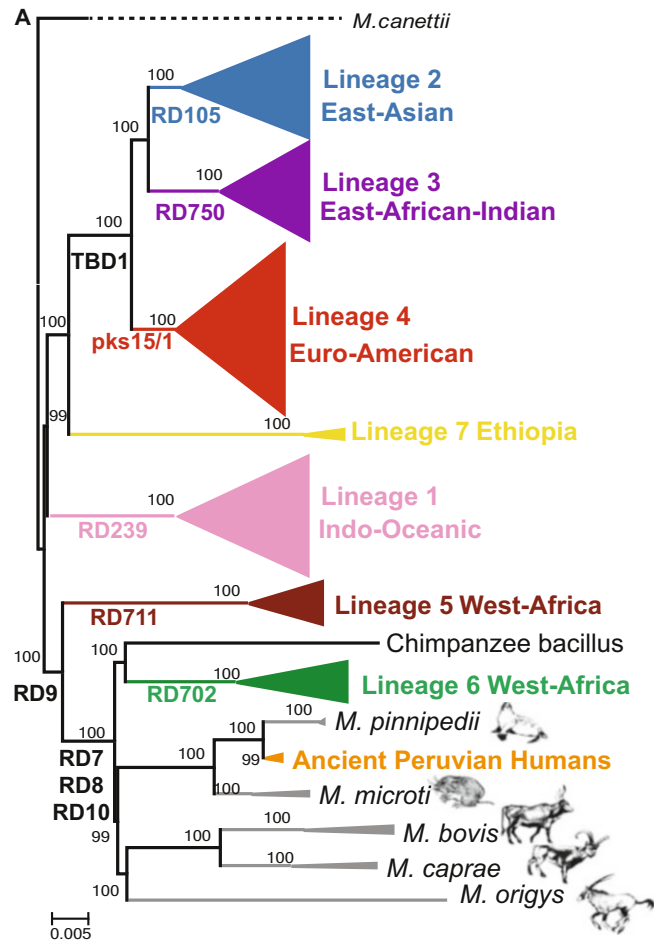

Fig. 5.1 Maximum likelihood phylogeny (Coscolla and Gagneux 2014). Node support after 1000 bootstrap replications is shown on branches and the tree is rooted by the outgroup *M. canettii*. Large Sequence Polymorphisms (LSPs) described in (Brosch et al. 2002) are indicated along branches. *Scale bar* indicates the number of nucleotide substitutions per site

africanum. Today, based on whole genome sequence (WGS) analyses, we know that MTBC includes seven human-adapted lineages (Lineage 1 to Lineage 7 in Fig. 5.1). The seven lineages can be grouped in "modern" and "ancestral" based on the presence or absence of a genomic region named TbD1. The so called "modern" strains belong to Lineages 2, 3 and 4 forming a monophyletic clade in which TbD1 is deleted. By contrast, the "ancestral" strains form a paraphyletic clade in which TbD1 is intact.

Apart from Lineages 1 to 7, the MTBC also includes several animal-adapted ecotypes (Smith et al. 2006) i.e. *M. bovis, M. caprae, M. microti, M. pinnipedii, M. origys, M. mungi*, *M. suricattae,* the dassie bacillus, the chimpanzee bacillus (Cousins et al. 1994; Alexander et al. 2010; Coscolla et al. 2013; Parsons et al. 2013), and a group of bacilli that display a smooth colony morphology and only occasionally infect humans; the latter are known as *M. canettii* and "smooth tuberculosis bacilli" (STB, see Chap. 2) (Canetti 1970; van Soolingen et al. 1997; Supply et al. 2013).

This chapter discusses the role of bacterial factors in human TB, therefore I will focus on the group of MTBC that predominantly causes disease in humans: *M. tuberculosis* sensu stricto (MTBC Lineage 1–4 and Lineage 7) and *M. africanum* (Lineage 5 and 6). A more thorough discussion of the smooth bacilli, *M. africanum* and the animal adapted ecotypes can be found in Chaps. 2, 6, and 7 of this book.

In human TB, the MTBC is transmitted from a person with the respiratory form of the disease to another person via air droplets generated

through coughing and sneezing. After exposure to the MTBC, the outcome can vary. Following initial infection (1) the bacteria can be cleared quickly through the action of innate immunity, (2) the infection can progress rapidly to active disease, or (3) be contained in a latent form with no clinical signs of disease. This latent TB infection may or may not re-activate up to several decades following initial exposure. Additionally, not only the outcome of infection is variable, but the course of disease can also vary. Active TB disease comprises a range of presentations, including pulmonary TB, and various forms of extrapulmonary disease such as TB meningitis and miliary TB. Each of these forms of disease feature a variety of symptoms that are associated with diverse host responses to the pathogen (O'Garra et al. 2013).

A particularity of TB is that the disease in different patient populations are often caused by different members of the MTBC (Hirsh et al. 2004; Gagneux et al. 2006a; Filliol et al. 2006; Baker et al. 2004; Hershberg et al. 2008; Wirth et al. 2008; Reed et al. 2009) (Figs. 5.1 and 5.2). Moreover, not all MTBC lineages are equally abundant and geographically distributed. Some MTBC lineages are restricted to one or a few countries and other lineages occur across a broader geographical range, even worldwide (Fig. 5.2). Lineage 5, 6 and 7 are good examples of the former as they occur almost exclusively in certain regions of Africa or in recent immigrants from those regions. The broadly distributed lineages are Lineage 2 and Lineage 4, also known as East-Asian and Euro-American lineages, respectively, which occur frequently in populations from Asia, Europe, Africa and the Americas. Finally, the remaining lineages, Lineage 1 and Lineage 3 show an intermediate geographical range. Interestingly, the specific bacteria-human combinations in disease can be disrupted in HIV co-infected patients (Fenner et al. 2013), indicating a role for immunosupression in the host-pathogen relationship. Indeed, human factors, including immunity, are known to impact the outcome of TB. For example, T cell defective immunity, such as in HIV seropositive individuals, is a strong risk factor for rapid progression from TB infection to active disease (Kwan and Ernst 2011). Similarly, human genetic determinants influence the susceptibility to TB (Casanova and Abel 2002). Environmental factors also play a role in TB epidemiology. For example, increased ventilation reduces TB transmission by diminishing exposure to infectious particles (Clark et al. 2002; Lygizos et al. 2013). In addition to these human and environmental factors, bacterial factors have increasingly been acknowledged as playing a role in the heterogeneity of TB infection and disease.

Human-adapted MTBC is an obligate human pathogen, and "virulence" in TB can be conceptualized as a composite of (i) the ability of the bacteria to survive in face of the host immune responses, (ii) their capacity to cause lung damage, (iii) to survive the aerosolisation process outside of the host, and (iv) successfully transmit to and infect a new host. The purpose of this chapter is to summarize and discuss the evidence that bacterial variation can result in different cellular, immunological or epidemiological phenotypes. Ultimately, differences among MTBC strains are determined by their genomic differences. Therefore, when possible, I will also present evidence for the bacterial genetic determinants that underlie these phenotypic differences. This chapter specifically excludes a particularly relevant virulence phenotype, which is drug resistance, as it is the subject of Chap. 14 of this book.

5.2 Genomic Differences Among MTBC Strains and Lineages

Strains of the MTBC differ genetically in their content of single nucleotide polymorphisms (SNPs), small insertion and deletions (indels), mobile and repetitive elements, large genomic deletions and large duplications. There are various typing methods to classify MTBC strains based in these sources of genetic diversity (see Chap. 3 of this book for details). However, the more robust and discriminatory way of classifying the MTBC and asses its genetic makeup is whole genome sequencing (WGS),

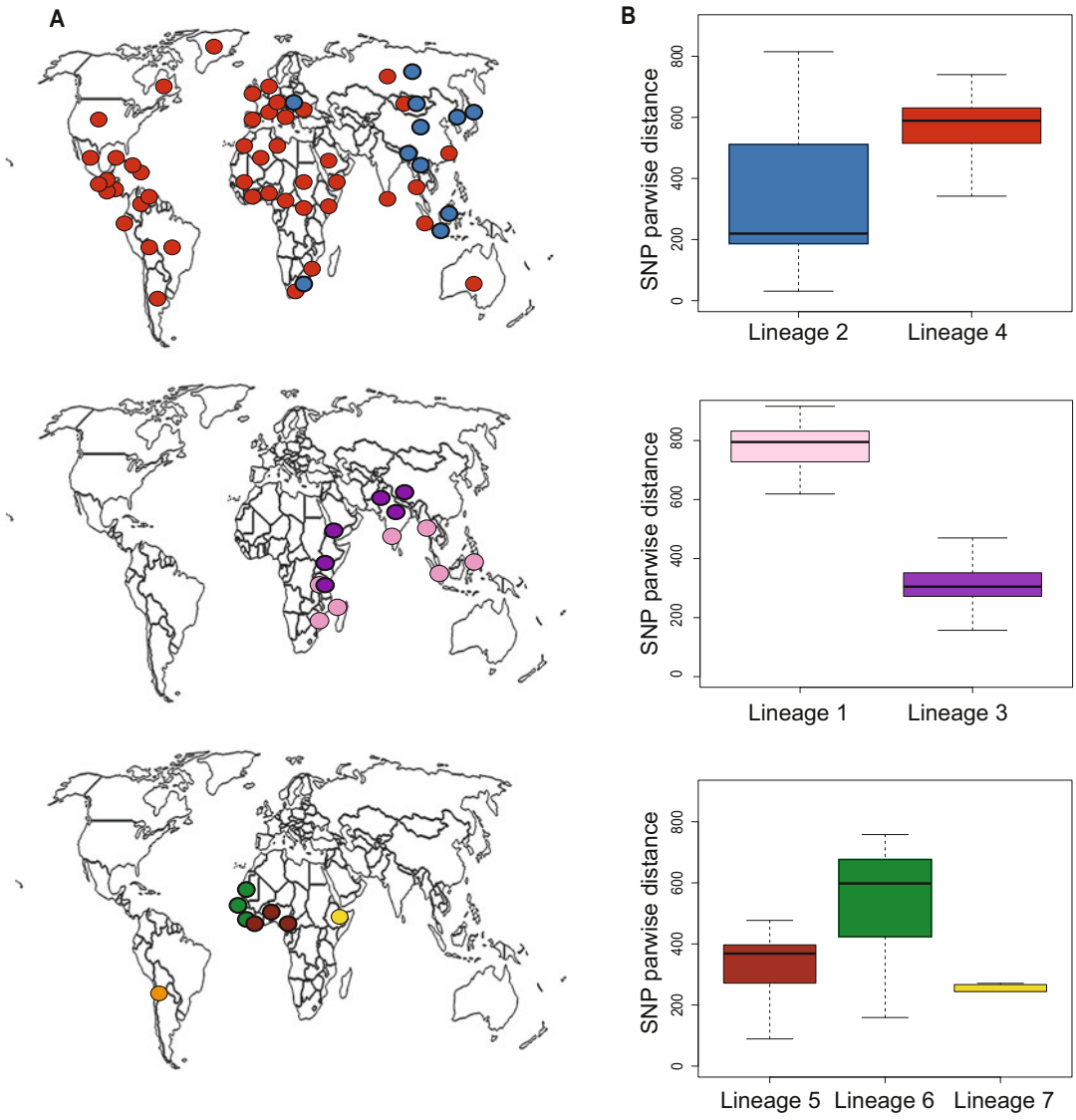

Fig. 5.2 Panel **a** shows the dominant MTBC Lineages by country (Coscolla and Gagneux 2014). Colors correspond to lineages in Fig. 5.1 and panel (**b**). Each dot corresponds to 1 of 80 countries represented in the 875 MTBC strains from the global strain collection analysed by Gagneux et al. (2006a). The *yellow* and an *orange dot* represent Lineage 7 in Ethiopia (Firdessa et al. 2013) and the extinct MTBC strains from Peru, respectively (Bos et al. 2014). Panel **b** shows the number of pairwise differences between MTBC strains for each lineage. The alignment of 217 human-adapted MTBC clinical strains published previously (Comas et al. 2013) were used to calculate the number of SNPs (i.e. the SNP-distance) (Adapted from Coscolla and Gagneux 2014)

which can reveal all sources of genomic variation.

Based on WGS analyses of all MTBC lineages, we know that on average, two human-adapted MTBC strains differ by about 1200 SNPs, which corresponds to 0.03 % of the genome when excluding repetitive sequences (Fig. 5.2b). Lineage 1 appears as the most diverse lineage, harbouring the largest genetic diversity with an average of 730 SNPs between any two strains belonging to this lineage (Fig. 5.2b). On the contrary, Lineage 7 is the most conserved

with only 230 SNPs separating any two strains (Fig. 5.2b). In terms of between-lineage diversity, the strains belonging to the "modern" Lineages 2, 3 and 4 differ by 970 SNPs on average. Strains belonging to the "ancestral" Lineages 1, 5, and 6 are more distantly related with an average of 1500 SNPs between them. The maximum SNP distance of 1800 SNPs is observed between strains of Lineage 7 and either Lineage 1, 5, or 6.

Apart from SNPs, there are other sources of genomic diversity. Repetitive and mobile elements in the MTBC genome can be classified as: Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), Variable Number of Tandem Repeats (VNTR) and Insertion Sequences (IS). All three elements are polymorphic, and they have been used for strain typing but their functional role has only rarely been determined. For example, variation of specific IS and VNTR elements in promotor regions can have phenotypic consequences by altering promotor activity and the differential expression of the downstream genes (Olsen et al. 2009; Phares et al. 2007; Tantivitayakul et al. 2010; Perez-Lago et al. 2013; Soto et al. 2004; Safi et al. 2004; Alonso et al. 2011). In addition, VNTR variability can impact bacterial phenotypes, as intragenic VNTR diversity can modify the structure and function of the proteins affected (Yindeeyoungyeon et al. 2009).

Although duplications are not an extensive source of genomic diversity among contemporary MTBC strains, large duplications have played a role in the expansion of gene families within mycobacteria. A large gene family known as the PE/PPE gene family, appeared by duplications and expansion of the ESAT-6 gene clusters in an ancestor of the MTBC (Gey van Pittius et al. 2006). Large-scale duplications might also have played a role during the in vitro evolution of the BCG vaccine strains. Two tandem duplications termed DU1 and DU2, of 29,668 bp and 36,161 bp, respectively, are present in *M. bovis* BCG Pasteur compared to a reference Lineage 4 MTBC strain, H37Rv (Brosch et al. 2000). Intriguingly, DU2 showed four alternative forms evolved in the different laboratories where the different BCG strains were passaged, leading to independent duplications of similar genomic regions (Brosch et al. 2007). The evolution of the BCG vaccine strains is discussed in detail in Chap. 8 of this book.

Large genomic deletions are regions from tens to thousands of base pairs that affect up to least 5.5 % of the MTBC genome (Tsolaki et al. 2004). Although large genomic deletions are robust phylogenetic markers that can be used to differentiate the main human MTBC lineages (Gagneux et al. 2006a), comparisons with other organisms suggested that the impact of large genomic deletions on the overall MTBC diversity is low (Tsolaki et al. 2004). However, deletions are known to modify relevant phenotypes such as the mycobacterial cell wall (Constant et al. 2002) and drug susceptibility (Rouse and Morris 1995).

5.3 Molecular and Cellular Consequences of Genomic Differences Between MTBC Strains

Biological differences among unicellular organisms comprise any different observable characteristics of cells. Bacterial phenotypes can be broadly applied to any cell property, including molecular phenotypes. In the MTBC, differences among strains have been reported for mRNA, protein and metabolite levels, methylation profiles, cellular growth, and cell wall structure.

5.3.1 Differences in mRNAs, Proteins and Metabolites

Deciphering the biochemical and metabolic networks of the MTBC is fundamental for the development of new diagnostic, drugs and vaccines. Differences in those networks between MTBC strains might result in heterogeneous efficacy of such interventions. Strain differences in transcriptomes (Mostowy et al. 2004; Golby et al. 2007; Homolka et al. 2010; Rose et al. 2013; Shell et al. 2013) and proteomes (Jungblut et al. 1999; Betts et al. 2000; He et al. 2003; Pheiffer

et al. 2005; Jhingan et al. 2016) have been reported. Both strain-specific and lineage-specific transcriptomes have been described for different MTBC clinical strains (Homolka et al. 2010; Rose et al. 2013). The genomic determinants of these transcriptional differences have been attributed to point mutations including synonymous and non-synonymous SNPs, deletions and duplications (Mostowy et al. 2004; Domenech et al. 2010; Weiner et al. 2012; Rose et al. 2013). Because the implication on the biology of the mycobacterial cell, one of the most widely studied phenotype at the transcription level is the dormancy survival regulator DosR. The DosR regulon is composed of 48 co-regulated genes and is essential for MTBC persistence (Sherman et al. 2001; Voskuil et al. 2003). Some Lineage 2/Beijing strains show a higher expression of DosR. Both, a gene duplication and a synonymous SNP within the gene immediately upstream of DosR have been suggested to be partially responsible for increased expression of the DosR regulon in Lineage 2/Beijing (Domenech et al. 2010; Weiner et al. 2012; Rose et al. 2013).

Other genetic determinants have been found to be responsible for transcriptional differences between isolates. One synonymous SNP upstream of malQ in Lineage 1 creates a new TSS associated with increased expression of malQ in strains from Lineage 1 (Rose et al. 2013). Moreover, deletions have been proven to affect transcription patterns of genes outside the deleted region (Mostowy et al. 2004). Interestingly, point mutations in methyltransferases can affect transcriptional patterns by altering the methylation patterns in the genome (Shell et al. 2013). One non-synonymous SNP in the methyltransferase *mamA* in Lineage 2 and another non-synonymous SNP in the active site of *hsdM* in Lineage 4 are responsible for the different adenine methylation between these two lineages (Shell et al. 2013). Similarly, mutations resulting in different methylation patterns have been shown at supra-lineage level between "ancestral" versus "modern" strains, at lineage level between Lineage 2 and lineage 3, but also within Lineage 4 (Zhu et al. 2016).

The final consequences of different transcription patterns in a cell are differences in metabolism. Large scale metabolic data is currently not available for different MTBC strains, but predictions can be made by modeling metabolic states using whole cell measurements of gene expression (Colijn et al. 2009). However, specific metabolite differences have been shown between MTBC strains and lineages. Lineage 5 and Lineage 6 differ at the metabolic level from the other human-adapted lineages; a feature which has been exploited for the traditional method to differentiate those lineages based on biochemical testing (see Chap. 6 in this book). Lineage 5 and Lineage 6 are also deficient for vitamin B12 synthesis. The explanation could be that a gene involved in biosynthesis of vitamin B12 is missing from Lineage 5 and Lineage 6 as a consequence of a deletion of the RD9 region (Brosch et al. 2002). Additionally, genes involved in Vitamin B12 metabolism appear as pseudogenes in both Lineage 5 and Lineage 6 genomes (Bentley et al. 2012).

One of the reasons of the slow in vitro growth of Lineage 5 and 6 is the pyruvate deficiency, and strains for Lineage 5 and 6 grow better in the presence of pyruvate. A non-synonymous SNP in the *pykA* gene has been found to be responsible for the specific pyruvate requirements for growth of Lineage 5 and 6 strains (Chavadi et al. 2009).

5.3.2 Differences in Bacterial Growth

Intracellular growth of the MTBC is used as a proxy for bacterial fitness in vitro (Koch et al. 2014). In vivo, bacterial fitness can be approximated by highly transmissible strains that can be identified as molecularly clustered strains using discriminatory genotyping methods (see Chap. 3). One study showed that such clustered MTBC strains grew significantly faster in a monocyte cell-line compared to less transmissible strains (Theus et al. 2005, 2006), indicating that bacterial growth can also be a proxy for fitness in clinical settings. Several studies have shown that different strains show different growth rates

in vitro and in human and murine phagocytes (McDonough et al. 1993; Laochumroonvorapong et al. 1997; Zhang et al. 1999; Manca et al. 1999; Li et al. 2002; Sarkar et al. 2012; Homolka et al. 2010), as well as in animal models (Ordway et al. 1995; North et al. 1999; Collins and Smith 1969; Valway et al. 1998; Bishai et al. 1999; Barczak et al. 2005). Most of the current evidence comes from experiments using reference strains like H37Rv, H37Ra, CDC1551, Erdman and HN878. Reference strains originated from clinical strains but have been passaged for a long time and have thus become laboratory-adapted. Comparisons between the laboratory strain H37Rv and its non-virulent derived form, known as H37Ra, showed they grew differently in murine but not in human macrophages (Paul et al. 1996). Clinical and reference strains also grew differently in liquid culture (Sarkar et al. 2012), human macrophages (Newton et al. 2006; Tanveer et al. 2009) and mice (Valway et al. 1998). Interestingly, Lineage 3 strains showed consistently slower growth than reference strains (Sarkar et al. 2012; Newton et al. 2006; Tanveer et al. 2009). Studying clinical strains has the advantage of better representing the phenotype of the virulent MTBC strains. However, the heterogeneous nature of clinical strains makes comparisons across studies challenging. In liquid culture, Sarkar et al. showed that Lineage 4 grew the fastest, Lineage 3 the slowest and Lineage 2 displayed an intermediate phenotype (Sarkar et al. 2012). In contrast, Homolka et al. reported that Lineage 4 strains grew slower than Lineage 1, 2 and 5 (Homolka et al. 2010). Such discrepancies do not only appear across studies, but also within studies where strains that grew faster in liquid culture grew slower in macrophages (Homolka et al. 2010; Sarkar et al. 2012). Reiling et al. showed that "modern" strains replicated faster in vitro, in human monocyte-derived macrophages and in aerosol infected mice (Reiling et al. 2013). The different experimental conditions across studies together with the different growth requirements of the different MTBC strains might result in apparent contradicting results across experiments. Therefore, to gain a better knowledge of the differences among clinical strains, it is necessary to take into account the experimental conditions as well as the qualitative and quantitative genetic differences among strains belonging to the same lineage.

Specific genetic differences among strains have been shown to be responsible for the different growth dynamics. As explained in the previous section and in Chap. 6 of this book, the deficient growth patterns of Lineage 5 and 6 in standard media can be explained by mutations in genes coding for biosynthesis of metabolites such as pyruvate and vitamin B12. Additionally, because drug resistance-conferring mutations affect genes encoding essential functions of the cell, most drug resistant mutations confer a fitness defect (Melnyk et al. 2015). Consequently, MTBC strains with drug resistance-conferring mutations grow slower (Gagneux et al. 2006b). Moreover, a disruption of a great number of genes implicated in the biogenesis of the mycobacterial cell wall has also been shown to result in slower growth in vitro and in different tissues of infected animal models (reviewed in Forrellad et al. 2013).

5.3.3 Differences in the Mycobacterial Cell Wall

The mycobacterial cell wall consists of an inner and an outer layer. The inner layer is the "core" of the membrane and includes peptidoglycan, arabinogalactan and mycolic acids. The upper segment is composed of free lipids and fatty acids. Interspersed are the cell-wall proteins, the phosphatidylinositol mannosides (PIMs), the phthiocerolcontaining lipids, lipomannan (LM), and lipoarabinomannan (LAM). Modification of the composition and structure of the cell wall have various consequences for the cell, ranging from the basic viability of the cells (Hett and Rubin 2008) to the interaction with host cells (Brennan 2003). There is strong evidence for differences in cell wall among MTBC strains. Even closely related strains differing in a small number of SNPs, like the laboratory strain H37Rv and its avirulent variant H37Ra, exhibited a fivefold dif-

ference in abundance of 29 membrane-associated proteins (Målen et al. 2011). One important component of the cell wall is the manosilated lipoarabinomannan (manLam), which is expressed at the MTBC cell surface. ManLam variations among strains translate in differences in the surface and most likely also in immune-pathogenesis (Torrelles and Schlesinger 2010). Natural mutants in manLam among clinical Lineage 2 grew slower than the Lineage 4 reference strains H37Rv and Erdman (Torrelles et al. 2008).

Lipids are important constituents of the mycobacterial cell wall, representing 60 % of the dry mass and impacting cytokine responses and antigen presentation (Astarie-Dequeker et al. 2010). Mycolic acids play an important role in modulating the host immune response and helping the bacteria resist oxidative stress (Vander Beken et al. 2011). Portevin et al. reported significant differences in mycolic acid profiles between different MTBC strains and lineages (Portevin et al. 2014). Mutants in the genes *pcaA* and *hma* have been shown to affect mycolic acids and their effect on virulence in mouse model (Glickman et al. 2000; Dubnau et al. 2000). Gene *pcaA* is required for mycolic acid cyclopropane ring synthesis in the cell wall of both BCG and *M. tuberculosis* (Glickman et al. 2000). *PcaA* mutants showed decreased bacterial load in lung, spleen and liver. *Hma* mutant strains no longer synthesizes oxygenated mycolic acids, causing profound alterations in envelope permeability and attenuation in mouse infections (Dubnau et al. 2000).

Complex lipids such as phthiocerol dimycocerosate (PDIM) help mycobacteria evade the immune system (Cambier et al. 2014). PDIM is prone to being lost during extended periods of *in vitro* culture (Domenech and Reed 2009). Mutations in *psk1*, *pks5, pks7, pks 10, pks 12, pks 15, mmpL4, mmpL7, fadD26* and *fadD28* have been demonstrated to alter the synthesis of PDIM, leading to differences in growth and/or attenuation in mice (Sirakova et al. 2003a, b; Reed et al. 2004; Tsenova et al. 2005; Cox et al. 1999; Camacho et al. 1999; Rousseau et al. 2004; Domenech et al. 2005; Rousseau et al. 2003).

Sulfatides and phospholipids can also vary across MTBC strains, which is often linked to differences in virulence (Goren et al. 1974a). MTBC clinical isolates showed differences in the production of sulfolipids in culture, which was correlated with the rank order of virulence in a guinea pig model of infection (Gangadharam et al. 1963; Goren et al. 1974a, 1982; Grange et al. 1978). Genes such as *pks2* (Sirakova et al. 2001), and *mmpL8* gene (Domenech et al. 2004; Converse et al. 2003) have been shown to determine the variable sulfolipid composition of MTBC strains. Moreover, an amino acid change at position 71 of PhoR that occurs naturally in all MTBC strains belonging to Lineage 5 and 6, and in all animal-adapted MTBC strains, could be a genetic determinant that affects the synthesis of sulfolipid and lipids of the polyacyltrehalose (PAT). Another cell wall constituent that showed differences among strains was phthiocerol dimycocerosate (DIM) (Goren et al. 1974b). Lineage 2/Beijing strains exhibited structural variants of DIM and phenolic glycolipid (PGL) compared to other MTBC strains (Huet et al. 2009). These structural variants were due to a point mutation in the gene Rv2952 (Huet et al. 2009). But other genetic determinants do also affect PGL. A deletion disrupting the locus of the polyketide synthase (Pks)15/1 in strains belonging to Lineage 4 resulted in a defective production of phenolic glycolipid (PGL) (Constant et al. 2002), thereby affecting the bacterial interaction with host immune cells (Sinsimer et al. 2008). Indeed, the absence of PGL has been linked to reduced MTBC virulence in infection models (Reed et al. 2004; Cambier et al. 2014).

5.4 Immunological Consequences of MTBC Diversity

Innate and adaptive immunity play an important role during TB, and these processes can be affected by genetic differences among MTBC strains. Different strains might interact differently with host cell receptors, which will have

consequences for bacterial entry into the host cell, binding to MHC that will interact with T cells, the activation of host cells to kill *M. tuberculosis*, the death of the infected cells, and the ability to propagate to other cells. The key immunological outcome for controlling MTBC is bacterial killing. There is strong experimental evidence that MTBC strains differ in their susceptibility to phagocytes because bacterial strains vary in the inflammatory responses they induce, in their susceptibility to reactive oxygen and nitrogen intermediate (ROIs and RNIs, respectively), and in their capacity to control phagosome maturation. While there is large experimental evidence for different susceptibility to phagocytes among MTBC strains, immunological differences among MTBC strains in clinical settings are mostly restricted to a different inflammatory response.

5.4.1 Differences in Inflammatory Response

Transmission of the MTBC occurs from persons with active pulmonary disease, therefore MTBC transmission might benefit from a more severe inflammatory host response causing more lung damage (further discussed in Ernst 2012). However, many studies have found increased MTBC virulence associated with reduced and/or delayed inflammatory responses, perhaps by allowing a stronger bacterial proliferation early during the infection process, leading to increased virulence at a later stage.

Different MTBC strains and lineages have been shown to induce distinct immune responses. At the supra-lineage level, "modern" strains (Lineages 2, 3 and 4) caused a lower early inflammatory response compared to other lineages (Portevin et al. 2011; Chen et al. 2014). Similarly, Krishnan et al. (2011) showed increased inflammatory responses in Lineage 1 strains compared to Lineage 4, but these were not significantly higher than in Lineage 2 strains. The lower inflammatory response of "modern" lineages has also been demonstrated in clinical setting, where Lineage 2 and 3 tended to induce lower IFN-γ than Lineage 1 (Rakotosamimanana et al. 2010). Lineage 6, belonging to the "ancestral" group, showed a different inflammatory phenotype than "modern" lineages in clinical studies (de Jong et al. 2006; Tientcheu et al. 2014). De Jong et al. showed that Lineage 6 was less likely to trigger an IFN-γ response to the dominant antigen ESAT-6 than patients and household contacts infected by other lineages (de Jong et al. 2006). Further differences in host immune responses between TB patients infected with Lineage 6 have been reported (Tientcheu et al. 2014). Some genetic determinants could explain the different inflammatory phenotypes of Lineage 6. Lineage 5 and Lineage 6 have a base pair insertion that truncates the gene product of the ortholog of Rv2958c (Bentley et al. 2012) that could be responsible for lower survival in human macrophages (Miller and Shinnick 2000). Moreover, Lineage 6 has a frameshift mutations in the ortholog of Rv0386 (Bentley et al. 2012) that could result in a reduction of TNF-alpha production (Agarwal et al. 2009).

Different strains within the "modern" lineages also differed in their host immune profile (Newton et al. 2006; Mihret et al. 2012; Tanveer et al. 2009; Sarkar et al. 2012; Wang et al. 2010; van Laarhoven et al. 2013). For example, studies using different infection models found that strain NH878 (belongs to Lineage 2/Beijing) was consistently associated with a delayed inflammatory immune response and increased virulence (Manca et al. 2001, 2004, 2005; Li et al. 2002; Tsenova et al. 2005; Chacon-Salinas et al. 2005; Rocha-Ramirez et al. 2008; Subbian et al. 2013). Lineage 3 strains exhibited a higher inflammatory phenotype compared to strains of Lineage 4 in macrophages (Newton et al. 2006) and in plasma samples from the patients (Mihret et al. 2012). This could be partially explained by the deletion of Rv1519 in Lineage 3 strains that has been linked to a decrease in the production of the anti-inflammatory cytokine IL-10 by in infected hosts (Newton et al. 2006). Wang et al. (2010) reported that differ-

ent Lineage 2/Beijing strains commonly induced lower levels of TNF-α, IL-6, IL-10 and GRO-α compared to strain H37Rv in monocyte-derived macrophages and dendritic cells. Similarly, other studies have reported a lower proinflammatory phenotype for Lineage 2 compared to Lineage 4 in in monocyte-derived macrophages (Sarkar et al. 2012) and in peripheral blood mononuclear cells (van Laarhoven et al. 2013). Conversely, Krishnan found lower concentrations of TNF-α elicited by Lineage 4 than by Lineage 2 in bone-marrow-derived murine macrophages (Krishnan et al. 2011). These contradictory findings with respect to the host immune responses elicited by Lineage 2 and Lineage 4 could on the one hand be due to differences in experimental conditions, including the different cell types or the antigens used to stimulate the immune response. On the other hand, sub-lineages within the main lineages, as well as individual strains within any given lineage might vary in the inflammatory responses they elicit.

Because the main MTBC lineages are not genetically homogeneous (Figs. 5.1 and 5.2), differences in immune regulation among strains are expected at sub-lineage level (Portevin et al. 2011). Such differences have been most widely studied in Lineage 2. Although one study did not find differences between sublineages within Lineage 2/Beijing (Wang et al. 2010), two independent studies showed that Lineage 2/"modern Beijing" showed a lower inflammatory response compared to so-called Lineage 2/"ancient Beijing" (Chen et al. 2014; van Laarhoven et al. 2013). Some of the variable cytokine profiles elicited by the distinct groups of strains within Lineage 2 have been linked to differential Toll-like receptor recognition (Carmona et al. 2013). Deletions in PE_PGRS33 were associated with reduced induction of TNF-α by the host (Basu et al. 2007). Because PE_PGRS 33 is polymorphic within lineages (Copin et al. 2014), it is possible that deletions in PE_PGRS33 might be partially responsible for the different immune phenotypes at sublineage level.

Increased "virulence" surely entails more than a mere delay in early pro-inflammatory response. Indeed, Reiling et al. (2013) reported lineage-specific differences in virulence profiles based on variable bacterial uptake by host cells, differences in cytokine induction, and intracellular growth. In agreement with other studies (discussed above), the "modern" Lineages 2 and 4 showed high replicative potential compared to "ancestral" Lineages 1 and 6. However, Lineage 2 was characterized by low uptake, and low cytokine induction, whereas Lineage 1 and Lineage 4 exhibited high uptake and higher cytokine induction.

5.4.2 Differences in Susceptibility to Macrophages

Reactive oxygen and nitrogen species are part of the antimicrobial repertoire of phagocytes. Therefore differences in hydrogen peroxide susceptibility across MTBC strains are relevant to tuberculosis infection and pathology. Starting in the 1960s, results from in vitro cultures have shown differences among MTBC strains in the susceptibility to hydroxen peroxide (Subbaiah and Ramachandran 1961; Subbaiah et al. 1960; Grange et al. 1977; Grange et al. 1978; Rhoades and Orme 1997; Firmani and Riley 2002). Jackett et al. showed that strains showing a high virulence phenotype in guinea pigs were associated with lower susceptibility to hydrogen peroxide (Jackett et al. 1978). Similarly, resistance to reactive nitrogen species was positively correlated with bacterial virulence in guinea pigs (O'Brien et al. 1994). More recently, Lineage 3 clinical strains appeared to be less susceptible to hydroxen peroxide than reference strains (Newton et al. 2006). However, using only reference strains, Laochumroonvorapong et al. found that virulent and attenuated strains did not differ in susceptibility to exogenous hydrogen-peroxide (Laochumroonvorapong et al. 1997). Repetitive genomic elements might be implicated in macrophage resistance. Increased number of repeats in the VNTR3239 locus of MTBC was reported to lead to a higher expression of DNA glycosylase, which is involved in the

protection against oxidative DNA damage (Olsen et al. 2009). Therefore, the genetic basis of the susceptibility to reactive nitrogen species could be partially due to VNTR polymorphisms.

MTBC can also resist macrophage killing by arresting phagosome maturation. The bacteria can modify their intracellular milieu by blocking the acquisition of late endosomal/lysosomal constituents and facilitating fusion with early endosomal compartments (Vergne et al. 2004). Important constituents of the mycobacterial cell wall such as LAM and PIM are known to play an important role in bacterial survival within infected macrophages (Fratti 2003 and Vergne et al. 2004). Consequently, several mutations affecting the production of cell wall components described in the section above will impact MTBC resistance to macrophages. Moreover, additional genes were found to be responsible for a defective arrest of phagosome maturation in *M. tuberculosis* (Beatty et al. 2000).

5.4.3 Differences in Host Cell Death

The way infected cells die is an important mechanism to control pathogenesis during infection (Ruckdeschel et al. 1997; Wickstrum et al. 2009; Zhou et al. 2000). In the MTBC, apoptosis is a defense mechanism that initiates both innate and adaptive immunity (Behar et al. 2010), and necrosis allows the bacteria to evade host defense mechanisms by inducing cellular lysis and spreading the infection (Chen et al. 2006). Cells infected with the attenuated MTBC laboratory strain H37Ra undergo apoptosis, while cells infected with the virulent H37Rv undergo necrosis in both human and mouse macrophages (Gan et al. 2008; Divangahi et al. 2009; Duan et al. 2001; Lee et al. 2006; Oddo et al. 1998; Brookes et al. 2003; Chen et al. 2006, 2008). Different genes such as *secA2* and *nuoG* have been demonstrated to be involved in controlling cell death (Hinchey et al. 2007; Velmurugan et al. 2007). Although mutants of these genes altered either the virulence or the immune profile in mice and guinea pigs, there is no evidence that natural mutants of these genes present a more attenuated virulence phenotype during human disease.

5.5 Epidemiological Consequences of Genomic Differences Between MTBC Strains

Virulence in TB is a composite of different clinical or epidemiological phenotypes such as the severity of the disease, the ranges of disease presentation and the transmission rate.

5.5.1 Disease Severity

The virulence of a pathogen depends on how much it damages the host. Therefore, one way to measure virulence in TB is by probing the severity of the disease. Many studies have reported differences in disease severity between MTBC strains in animal models. But investigating the impact of bacterial differences in human disease is more challenging, because social, environmental and biological factors contribute to the phenotype. However, two independent studies in human TB concluded that strains from the "modern" lineages were more virulent than strains from the "ancestral" clades (Stavrum et al. 2014; de Jong et al. 2008). Stavrum et al. (2014) showed that TB patients from Tanzania infected with "modern" Lineage 4 strains showed more c1-acid glycoprotein and C reactive protein, higher neutrophils counts, and a lower body mass index than those infected with Lineage 1. De Jong et al. (2008) showed that in the Gambia, individuals infected with modern Lineages 2 and 4 were more likely to progress to active disease compared to individuals infected with Lineage 6. The higher virulence of the modern Lineage 2 compared to "ancestral" Lineage 6 was corroborated in a novel marmoset model of infection (Via et al. 2013) where infection with Lineage 2 induced more rapid weight loss, and led to a higher bacterial load in liver, spleen and lymph nodes. Deciphering which of the natural sequence vari-

ants that appeared during MTBC evolution affect disease severity in human populations is challenging. Mutations in *phoP* have been shown to play a role in virulence (Chesne-Seck et al. 2008; Gonzalo-Asensio et al. 2014). Lineage 6 and animal strains have a non-synonymous SNP in *phoP* that reduces transcription of the PhoP regulon, and thus might be partially responsible for the reduced virulence of Lineage 6 compared to "modern" strains.

Clinical studies have also demonstrated a higher virulence of Lineage 2, generally compared to Lineage 4. For example, Ogarkov and colleagues (2012) found that patients infected with Lineage 2 and carrying a human genetic polymorphism previously associated with TB (CD209_336 A/G) were more likely to die of TB compared to patients infected with other strains. Nahid et al. (2010) found during a clinical trial that patients infected with Lineage 2 strains and one sub-group of Lineage 4 were more likely to yield a positive culture at week eight after treatment initiation when compared to other Lineage 4 strains. Other experimental studies in mice and rabbits have supported the higher virulence of Lineage 2 (Barczak et al. 2005; Ordway et al. 2007; Tsenova et al. 2005; Manca et al. 2005). Even within Lineage 2, differences of virulence have been demonstrated in guinea pigs and mice (Kato-Maeda et al. 2012; Aguilar et al. 2010), demonstrating the role of strain genetic background on the virulence phenotype of TB.

Apart from genetic variation of *phoP* mentioned above, there is no clear genetic determinant of strain- or lineage-specific virulence in MTBC. However, nucleotide variation and differences in gene expression in some PE/PPE genes have been associated with virulence (Dheenadhayalan et al. 2005; Yu et al. 2011), indicating a possible role of this gene family in TB virulence.

5.5.2 Transmissibility

The ultimate virulence phenotype that impacts the extent of TB epidemic is transmissibility. MTBC transmissibility can be estimated by comparing genotypic clustering between patient isolates (van der Spuy et al. 2003), comparing the prevalence of household contacts with a positive TST (de Jong et al. 2008), measuring the frequency of genotypes over time, and by association with younger patient age, because TB in older patients is more likely to reflect reactivation rather than ongoing transmission (Borgdorff and van Soolingen 2013). The majority of epidemiological studies have supported the view that overall, strains from "modern" lineages are more transmissible than other MTBC strains. On the one hand, three studies have reported a decreasing prevalence of *M. africanum* (i.e. Lineage 5 and 6) in Cameroon, Guinea-Bissau and Burkina Faso (Niobe-Eyangoh et al. 2003; Godreuil et al. 2007; Groenheit et al. 2011), supporting the notion of a lower transmission potential of these "ancestral" lineages. Similarly, Buu et al. (2012) reported higher genotypic clustering of Lineage 2 compared Lineage 1 in Vietnam. However, (de Jong et al. 2008) reported no differences in transmission between "ancestral" and "modern" lineages by comparing contacts with positive TST. Many studies have noted a higher transmissibility of Lineage 2/Beijing strains based on higher clustering and increases in frequency over time (Wirth et al. 2008; Chen et al. 2014; Anh et al. 2000; European Concerted Action on New Genetic et al. 2006; Hanekom et al. 2007; Cowley et al. 2008; van der Spuy et al. 2009; Tuite et al. 2013; Buu et al. 2012). In some cases, the increased transmission of Lineage 2/Beijing was associated with drug resistance (Buu et al. 2012; Kubica et al. 2004). Yet, several other studies observed no such increases and no enhanced transmissibility of Lineage 2 strains compared to other strains (de Jong et al. 2008; Lillebaek et al. 2003; Albanna et al. 2006; Marais et al. 2009; Langlois-Klassen et al. 2013). Because transmission is shaped by human and social factors, some of these contradictory findings could be due to differences in the study populations. Another explanation for discrepant results is that the different measures for approximating transmission are measuring different things. For example, by comparing the positive TST of household contacts you can measure transmission independently of

progression to active disease. By contrast, when comparing isolate clustering, because you need to genotype the infecting bacterial strains, only the cases that progressed to active disease are considered. Transmission studies that depend on isolating the bacteria can also be affected by culture bias due to MTBC variability in growth and metabolism. Additionally, if sublineages differ in transmissibility, the genetic heterogeneity within lineages and different prevalence of sublineages across geographical locations would lead to inconsistent results. In support of this notion, Kato-Maeda et al. showed that one sublineage within Lineage 2 showed higher genotypic clustering in San Francisco (Kato-Maeda et al. 2010). This sublineage was also more virulent in guinea pigs, supporting again an intrinsic bacterial role in increased pathogenicity linked to enhanced transmission (Kato-Maeda et al. 2012).

Taken together, although de Jong et al. (2008) did not find different transmission rates between Lineage 6 and other lineages in The Gambia, several studies support the view that on average, strains belonging to "modern" MTBC lineages are more transmissible than other strains.

5.5.3 Disease Presentation

MTBC can cause a wide variety of disease manifestations affecting many organs of the human body (Zumla et al. 2013). However, because efficient MTBC transmission relies on lung damage, only pulmonary disease has a role in TB transmission. Therefore, bacterial genotypes that are more prone to cause pulmonary and cavitary disease will be associated with an increased transmission potential.

In humans, epidemiological studies have been rather inconsistent with regards to MTBC strain or lineage effects on disease presentations. At least five studies reported such effects but there is no consistency of one group of strains being associated with a particular disease presentation in different clinical settings (Kong et al. 2006, 2007; Pareek et al. 2013; Click et al. 2012; Caws et al. 2008). On the other hand, at least another five studies failed to find any association between MTBC lineage and disease presentation (Borgdorff et al. 2004; Nicol et al. 2005; Wampande et al. 2013; van Crevel et al. 2001; Firdessa et al. 2013).

Experiments in the marmoset model revealed more extra-pulmonary spread to the lymph node, liver and spleen in animals infected with a Lineage 6 clinical strain compared to animals infected with the Lineage 4 strain CDC1551 (Via et al. 2013). Unlike clinical studies, where human factors might obscure associations, these experiments were conducted using genetically identical individuals (i.e. marmosets always give birth to twins or triplets), suggesting that the different disease presentation associated with the different MTBC strains are due to bacterial rather host genetic differences. In support of this notion, the disruption of a bacterial gene coding for a possible phospholipase (*plcD*) has been associated with extrathoracic TB (Yang et al. 2005; Kong et al. 2005).

5.6 Conclusions

MTBC strains belonging to Lineage 2 and Lineage 4 are more globally wide-spread than other strains. If the global success of Lineage 2 and Lineage 4 is linked to virulence, we would expect these lineages to show a more virulent phenotype. Different clinical phenotypes such as transmission rate, disease severity and presentation contribute to virulence in TB. However, clinical phenotypes are probably influenced by cellular and immunological phenotypes. The cell wall for the MTBC has a key role in the interaction with the host immune system. There is a long list of genomic loci responsible for difference in mycobacterial cell wall (reviewed in Forrellad et al. 2013). As a consequence, natural bacterial mutants in these genes are likely to impact MTBC pathogenicity. Growth rate is a cellular phenotype used as a proxy of bacterial fitness, and MTBC strains differ in growth rate in different experimental conditions. Similarly, different MTBC strains have shown differences in susceptibility to host immune cells. However, because cellular phenotypes are measured in dif-

ferent experimental conditions, it is difficult to assess the role of bacterial genetics during human infections. Differences between strains in clinical settings have been demonstrated with respect to their inflammatory response, the severity of the disease, transmission rate and disease presentation. Although the variety of strains and the different approximations to measure complex phenotypes poses a challenge when comparing studies, the observation that highly transmissible strains in humans tend to cause more severe disease in animal models is consistent with intrinsic bacterial features linking "virulence" with pathogenicity and strain fitness in epidemiological settings. Moreover, the growing evidence of bacterial genetic determinants affecting different cellular and clinical phenotypes strongly supports a role for bacterial factors in the outcome of TB infection and disease. In particular, evidence supports that strains from Lineage 5 and Lineage 6 are metabolically different, grow slower, and are less virulent. Similarly, at least certain groups of Lineage 2 and Lineage 4 strains are more virulent in terms of disease severity and human-to-human transmission. In summary, there is growing evidence that, together with human and environmental factors, MTBC strain diversity contributes to the variable outcome of infection and disease in human TB.

References

Agarwal N, Lamichhane G, Gupta R, Nolan S, Bishai WR (2009) Cyclic AMP intoxication of macrophages by a *Mycobacterium tuberculosis* adenylate cyclase. Nature 460:98–102

Aguilar D, Hanekom M, Mata D, Van Pittius NCG, Van Helden PD, Warren RM, Hernandez-Pando R (2010) *Mycobacterium tuberculosis* strains with the Beijing genotype demonstrate variability in virulence associated with transmission. Tuberculosis 90:319–325

Albanna AS, Reed MB, Kotar KV, Fallow A, Mcintosh FA, Behr MA, Menzies D (2006) Reduced transmissibility of East African Indian strains of *Mycobacterium tuberculosis*. PLoS One 2011:e25075

Alexander KA, Laver PN, Michel AL, Williams M, Van Helden PD, Warren RM, Gey Van Pittius NC (2010) Novel *Mycobacterium tuberculosis* complex pathogen, M. mungi. Emerg Infect Dis 16:1296–1299

Alonso H, Aguilo JI, Samper SA, Caminero JA, Campos-Herrero MAI, Gicquel B, Brosch R, Martín C, Otal I (2011) Deciphering the role of IS6110 in a highly transmissible *Mycobacterium tuberculosis* Beijing strain, GC1237. Tuberculosis 91:117–126

Anh DD, Borgdorff MW, Van LN, Lan NT, Van Gorkom T, Kremer K, Van Soolingen D (2000) *Mycobacterium tuberculosis* Beijing genotype emerging in Vietnam. Emerg Infect Dis 6:302–305

Astarie-Dequeker C, Nigou J, Passemar C, Guilhot C (2010) The role of mycobacterial lipids in host pathogenesis. Drug Discov Today Dis Mech 7:e33–e41

Baker L, Brown T, Maiden MC, Drobniewski F (2004) Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. Emerg Infect Dis 10:1568–1577

Barczak A, Domenech P, Boshoff H, Reed MB, Manca C, Kaplan G, Barry CE III (2005) In vivo phenotypic dominance in mouse mixed infections with *Mycobacterium tuberculosis* clinical isolates. J Infect Dis 192:600–606

Basu S, Pathak SK, Banerjee A, Pathak S, Bhattacharyya A, Yang Z, Talarico S, Kundu M, Basu J (2007) Execution of macrophage apoptosis by PE_PGRS33 of *Mycobacterium tuberculosis* is mediated by toll-like receptor 2-dependent release of tumor necrosis factor-α. J Biol Chem 282:1039–1050

Beatty WL, Rhoades ER, Ullrich HJ, Chatterjee D, Heuser JE, Russell DG (2000) Trafficking and release of mycobacterial lipids from infected macrophages. Traffic 1:235–247

Behar SM, Divangahi M, Remold HG (2010) Evasion of innate immunity by *Mycobacterium tuberculosis*: is death an exit strategy? Nat Rev Microbiol 8:668–674

Bentley SD, Comas I, Bryant JM, Walker D, Smith NH, Harris SR, Thurston S, Gagneux S, Wood J, Antonio M, Quail MA, Gehre F, Adegbola RA, Parkhill J, De Jong BC (2012) The genome of *Mycobacterium africanum* West African 2 reveals a lineage-specific locus and genome erosion common to the M. tuberculosis complex. PLoS Negl Trop Dis 6:e1552

Betts JC, Dodson P, Quan S, Lewis AP, Thomas PJ, Duncan K, Mcadam RA (2000) Comparison of the proteome of *Mycobacterium tuberculosis* strain H37Rv with clinical isolate CDC 1551. Microbiology 146:3205–3216

Bishai WR, Dannenberg AM Jr, Parrish N, Ruiz R, Chen P, Zook BC, Johnson W, Boles JW, Pitt ML (1999) Virulence of *Mycobacterium tuberculosis* CDC1551 and H37Rv in rabbits evaluated by Lurie's pulmonary tubercle count method. Infect Immun 67:4931–4934

Borgdorff MW, Van Soolingen D (2013) The re-emergence of tuberculosis: what have we learnt from molecular epidemiology? Clin Microbiol Infect 19:889–901

Borgdorff MW, Van Deutekom H, De Haas PEWP, Kremer K, Van Soolingen D (2004) *Mycobacterium tuberculosis*, Beijing genotype strains not associated with radiological presentation of pulmonary tuberculosis. Tuberculosis 84:337–340

Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, Forrest SA, Bryant JM, Harris SR, Schuenemann VJ, Campbell TJ, Majander K, Wilbur AK, Gui-

chon RA, Wolfe Steadman DL, Cook DC, Niemann S, Behr MA, Zumarraga M, Bastida R, Huson D, Nieselt K, Young D, Parkhill J, Buikstra JE, Gagneux S, Stone AC, Krause J (2014) Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature 514:494–497

Brennan PJ (2003) Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*. Tuberculosis 83:91–97

Brookes RH, Pathan AA, Mcshane H, Hensmann M, Price DA, Hill AV (2003) CD8+ T cell-mediated suppression of intracellular *Mycobacterium tuberculosis* growth in activated human macrophages. Eur J Immunol 33:3293–3302

Brosch R, Gordon SV, Buchrieser C, Pym AS, Garnier T, Cole ST (2000) Comparative genomics uncovers large tandem chromosomal duplications in Mycobacterium bovis BCG Pasteur. Yeast 17:111–123

Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, Garnier T, Gutierrez C, Hewinson G, Kremer K, Parsons LM, Pym AS, Samper S, Van Soolingen D, Cole ST (2002) A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci U S A 99:3684–3689

Brosch R, Gordon SV, Garnier T, Eiglmeier K, Frigui W, Valenti P, Dos Santos S, Duthoy SP, Lacroix C, Garcia-Pelayo C, Inwald JK, Golby P, Garcia JNÝ, Hewinson RG, Behr MA, Quail MA, Churcher C, Barrell BG, Parkhill J, Cole ST (2007) Genome plasticity of BCG and impact on vaccine efficacy. Proc Natl Acad Sci 104:5596–5601

Buu TN, Van Soolingen D, Huyen MNT, Lan NTN, Quy HT, Tiemersma EW, Kremer K, Borgdorff MW, Cobelens FGJ (2012) Increased transmission of *Mycobacterium tuberculosis* Beijing genotype strains associated with resistance to streptomycin: a population-based study. PLoS One 7:e42323

Camacho LR, Ensergueix D, Perez E, Gicquel B, Guilhot C (1999) Identification of a virulence gene cluster of *Mycobacterium tuberculosis* by signature-tagged transposon mutagenesis. Mol Microbiol 34:257–267

Cambier CJ, Takaki KK, Larson RP, Hernandez RE, Tobin DM, Urdahl KB, Cosma CL, Ramakrishnan L (2014) Mycobacteria manipulate macrophage recruitment through coordinated use of membrane lipids. Nature 505:218–222

Canetti G (1970) Infection caused by atypical mycobacteria and antituberculous immunity. Lille Med 15:280–282

Carmona J, Cruz A, Moreira-Teixeira L, Sousa C, Sousa J, Osorio NS, Saraiva AL, Svenson S, Kallenius G, Pedrosa J, Rodrigues F, Castro AG, Saraiva M (2013) *Mycobacterium tuberculosis* strains are differentially recognized by TLRs with an impact on the immune response. PLoS ONE 8:e67277

Casanova JL, Abel L (2002) Genetic dissection of immunity to mycobacteria: the human model. Annu Rev Immunol 20:581–620

Caws M, Thwaites G, Dunstan S, Hawn TR, Thi Ngoc Lan N, Thuong NTT, Stepniewska K, Huyen MNT, Bang ND, Huu Loc T, Gagneux S, Van Soolingen D, Kremer K, Van Der Sande M, Small P, Thi Hoang Anh P, Chinh NT, Thi Quy H, Thi Hong Duyen N, Quang Tho D, Hieu NT, Torok E, Hien TT, Dung NH, Thi Quynh Nhu N, Duy PM, Vinh Chau N, Farrar J (2008) The influence of host and bacterial genotype on the development of disseminated disease with *Mycobacterium tuberculosis*. PLoS Pathog 4:e1000034

Chacon-Salinas R, Serafin-Lopez J, Ramos-Payan R, Mendez-AragoN P, Hernndez-Pando R, Van Soolingen D, Flores-Romo L, Estrada-Parra S, Estrada-Garcia I (2005) Differential pattern of cytokine expression by macrophages infected in vitro with different *Mycobacterium tuberculosis* genotypes. Clin Exp Immunol 140:443–449

Chavadi S, Wooff E, Coldham NG, Sritharan M, Hewinson RG, Gordon SV, Wheeler PR (2009) Global effects of inactivation of the pyruvate kinase gene in the *Mycobacterium tuberculosis* complex. J Bacteriol 191:7545–7553

Chen M, Gan H, Remold HG (2006) A mechanism of virulence: virulent *Mycobacterium tuberculosis* strain H37Rv, but not attenuated H37Ra, causes significant mitochondrial inner membrane disruption in macrophages leading to necrosis. J Immunol 176:3707–3716

Chen M, Divangahi M, Gan H, Shin DS, Hong S, Lee DM, Serhan CN, Behar SM, Remold HG (2008) Lipid mediators in innate immunity against tuberculosis: opposing roles of PGE2 and LXA4 in the induction of macrophage death. J Exp Med 205:2791–2801

Chen YY, Chang JR, Huang WF, Hsu SC, Kuo SC, Sun JR, Dou HY (2014) The pattern of cytokine production in vitro induced by ancient and modern Beijing *Mycobacterium tuberculosis* strains. PLoS One 9:e94296

Chesne-Seck ML, Barilone N, Boudou F, Asensio JG, Kolattukudy PE, Martin C, Cole ST, Gicquel B, Gopaul DN, Jackson M (2008) A point mutation in the two-component regulator PhoP-PhoR accounts for the absence of polyketide-derived acyltrehaloses but not that of phthiocerol dimycocerosates in *Mycobacterium tuberculosis* H37Ra. J Bacteriol 190:1329–1334

Clark M, Riben P, Nowgesic E (2002) The association of housing density, isolation and tuberculosis in Canadian First Nations communities. Int J Epidemiol 31:940–945

Click ES, Moonan PK, Winston CA, Cowan LS, Oeltmann JE (2012) Relationship between *Mycobacterium tuberculosis* phylogenetic lineage and clinical site of tuberculosis. Clin Infect Dis 54:211–219

Colijn C, Brandes A, Zucker J, Lun DS, Weiner B, Farhat MR, Cheng TY, Moody DB, Murray M, Galagan JE (2009) Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. PLoS Comput Biol 5:e1000489

Collins FM, Smith MM (1969) Comparative study of virulence of *Mycobacterium tuberculosis* measured in mice and guinea pigs. Am Rev Respir Dis 100:631

Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, Parkhill J, Malla B, Berg S, Thwaites

G, Yeboah-Manu D, Bothamley G, Mei J, Wei LH, Bentley S, Harris SR, Niemann S, Diel R, Aseffa A, Gao Q, Young D, Gagneux S (2013) Out-of-Africa migration and neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. Nat Genet 45:1176–U311

Constant P, Perez E, Malaga W, Laneelle MA, Saurel O, Daffe M, Guilhot C (2002) Role of the pks15/1 gene in the biosynthesis of phenolglycolipids in the *Mycobacterium tuberculosis* complex. J Biol Chem 277:38148–38158

Converse SE, Mougous JD, Leavell MD, Leary JA, Bertozzi CR, Cox JS (2003) MmpL8 is required for sulfolipid-1 biosynthesis and *Mycobacterium tuberculosis* virulence. Proc Natl Acad Sci U S A 100:6121–6126

Copin R, Coscolla M, Seiffert SN, Bothamley G, Sutherland J, Mbayo G, Gagneux S, Ernst JD (2014) Sequence diversity in the pe_pgrs genes of *Mycobacterium tuberculosis* is independent of human T cell recognition. mBio 5:e00960

Coscolla M, Gagneux S (2014) Consequences of genomic diversity in *Mycobacterium tuberculosis*. Semin Immunol 26:431–444

Coscolla M, Lewin A, Metzger S, Maetz-Rennsing K, Calvignac-Spencer S, Nitsche A, Dabrowski PW, Radonic A, Niemann S, Parkhill J, Couacy-Hymann E, Feldman J, Comas I, Boesch C, Gagneux S, Leendertz FH (2013) Novel *Mycobacterium tuberculosis* complex isolate from a wild chimpanzee. Emerg Infect Dis 19:969–976

Cousins DV, Peet RL, Gaynor WT, Williams SN, Gow BL (1994) Tuberculosis in imported hyrax (Procavia capensis) caused by an ususual variant belonging to the *Mycobacterium tuberculosis* complex. Vet Microbiol 42:135–145

Cowley D, Govender D, February B, Wolfe M, Steyn L, Evans J, Wilkinson R-J, Nicol M-P (2008) Recent and rapid emergence of W-Beijing strains of *Mycobacterium tuberculosis* in Cape Town, South Africa. Clin Infect Dis 47:1252–1259

Cox JS, Chen B, Mcneil M, Jacobs WR Jr (1999) Complex lipid determines tissue-specific replication of *Mycobacterium tuberculosis* in mice. Nature 402:79–83

De Jong BC, Hill PC, Brookes RH, Gagneux S, Jeffries DJ, Otu JK, Donkor SA, Fox A, Mcadam KP, Small PM, Adegbola RA (2006) *Mycobacterium africanum* elicits an attenuated T cell response to early secreted antigenic target, 6 kDa, in patients with tuberculosis and their household contacts. J Infect Dis 193:1279–1286

De Jong B-C, Hill P-C, Aiken A, Awine T, Antonio M, Adetifa I-M, Jacksonâ-Sillah D-J, Fox A, Deriemer K, Gagneux S, Borgdorff M-W, Mcadam K-P, Corrah T, Small P-M, Adegbola R-A (2008) Progression to active tuberculosis, but not transmission, varies by *Mycobacterium tuberculosis* lineage in the Gambia. J Infect Dis 198:1037–1043

Dheenadhayalan V, Delogu G, Brennan MJ (2004) Expression of the PE_PGRS 33 protein in Mycobacterium smegmatis triggers necrosis in macrophages and enhanced mycobacterial survival. Microbes Infect

Dheenadhayalan V, Delogu G, Brennan MJ. Expression of the PE_PGRS 33 protein in Mycobacterium smegmatis triggers necrosis in macrophages and enhanced mycobacterial survival. Microbes and Infection. 2006;8(1):262–72.

Divangahi M, Chen M, Gan H, Desjardins D, Hickman TT, Lee DM, Fortune S, Behar SM, Remold HG (2009) *Mycobacterium tuberculosis* evades macrophage defenses by inhibiting plasma membrane repair. Nat Immunol 10:899–906

Domenech P, Reed MB (2009) Rapid and spontaneous loss of phthiocerol dimycocerosate (PDIM) from *Mycobacterium tuberculosis* grown in vitro: implications for virulence studies. Microbiology 155:3532–3543

Domenech P, Reed MB, Dowd CS, Manca C, Kaplan G, Barry CE 3rd (2004) The role of MmpL8 in sulfatide biogenesis and virulence of *Mycobacterium tuberculosis*. J Biol Chem 279:21257–21265

Domenech P, Reed MB, Barry CE 3rd (2005) Contribution of the *Mycobacterium tuberculosis* MmpL protein family to virulence and drug resistance. Infect Immun 73:3492–3501

Domenech P, Kolly GS, Leon-Solis L, Fallow A, Reed MB (2010) Massive gene duplication event among clinical isolates of the *Mycobacterium tuberculosis* W/Beijing family. J Bacteriol 192:4562–4570

Duan L, Gan H, Arm J, Remold HG (2001) Cytosolic phospholipase A2 participates with TNF-alpha in the induction of apoptosis of human macrophages infected with *Mycobacterium tuberculosis* H37Ra. J Immunol 166:7469–7476

Dubnau E, Chan J, Raynaud C, Mohan VP, Laneelle MA, Yu K, Quemard A, Smith I, Daffe M (2000) Oxygenated mycolic acids are necessary for virulence of *Mycobacterium tuberculosis* in mice. Mol Microbiol 36:630–637

Ernst JD (2012) The immunological life cycle of tuberculosis. Nat Rev Immunol 12:581–591

European Concerted Action on New Genetic, M, Techniques for The, E, Control Of, T (2006) Beijing/W genotype *Mycobacterium tuberculosis* and drug resistance. Emerg Infect Dis 12:736–743

Fenner L, Egger M, Bodmer T, Furrer H, Ballif M, Battegay M, Helbling P, Fehr J, Gsponer T, Rieder HL, Zwahlen M, Hoffmann M, Bernasconi E, Cavassini M, Calmy A, Dolina M, Frei R, Janssens JP, Borrell S, Stucki D, Schrenzel J, Bçôttger EC, Gagneux S, For the Swiss, H. I. V. C. & Molecular Epidemiology of Tuberculosis Study, G (2013) HIV infection disrupts the sympatric host-pathogen relationship in human tuberculosis. PLoS Genet 9:e1003318

Filliol I, Motiwala AS, Cavatore M, Qi W, Hernando Hazbon M, Bobadilla Del Valle M, Fyfe J, Garcia-Garcia L, Rastogi N, Sola C, Zozio T, Guerrero MI, Leon CI, Crabtree J, Angiuoli S, Eisenach KD, Durmaz R, Joloba ML, Rendon A, Sifuentes-Osornio J, Ponce De Leon A, Cave MD, Fleischmann R, Whittam TS, Alland D (2006) Global phylogeny of *Mycobacterium*

tuberculosis based on Single Nucleotide Polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. J Bacteriol 188:759–772

Firdessa R, Berg S, Hailu E, Schelling E, Gumi B, Erenso G, Gadisa E, Kiros T, Habtamu M, Hussein J, Zinsstag J, Robertson BD, Ameni G, Lohan AJ, Loftus B, Comas I, Gagneux S, Tschopp R, Yamuah L, Hewinson G, Gordon SV, Young DB, Aseffa A (2013) Mycobacterial lineages causing pulmonary and extrapulmonary tuberculosis, Ethiopia. Emerg Infect Dis 19: 460–463

Firmani MA, Riley LW (2002) *Mycobacterium tuberculosis* CDC1551 is resistant to reactive nitrogen and oxygen intermediates in vitro. Infect Immun 70:3965–3968

Forrellad MA, Klepp LI, Gioffré A, Sabio Y, García J, Morbidoni HR, Santangelo MDLP, Cataldi AA, Bigi F (2013) Virulence factors of the *Mycobacterium tuberculosis* complex. Virulence 4:3–66

Fratti RA, Chua J, Vergne I, Deretic V. Mycobacterium tuberculosis glycosylated phosphatidylinositol causes phagosome maturation arrest. Proceedings of the National Academy of Sciences. 2003;100(9):5437–42.

Gagneux S, Deriemer K, Van T, Kato-Maeda M, De Jong BC, Narayanan S, Nicol M, Niemann S, Kremer K, Gutierrez MC, Hilty M, Hopewell PC, Small PM (2006a) Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. Proc Natl Acad Sci 103:2869–2873

Gagneux S, Long CD, Small PM, Van T, Schoolnik GK, Bohannan BJM (2006b) The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. Science 312:1944–1946

Gan H, Lee J, Ren F, Chen M, Kornfeld H, Remold HG (2008) *Mycobacterium tuberculosis* blocks crosslinking of annexin-1 and apoptotic envelope formation on infected macrophages to maintain virulence. Nat Immunol 9:1189–1197

Gangadharam PR, Cohn ML, Middlebrook G (1963) Infectivity, pathogenicity and sulpholipid fraction of some Indians and British strains of tubercle bacilli. Tubercle 44:452–455

Gey Van Pittius N, Sampson S, Lee H, Kim Y, Van Helden P, Warren R (2006) Evolution and expansion of the *Mycobacterium tuberculosis* PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. BMC Evol Biol 6:95

Glickman MS, Cox JS, Jacobs WR Jr (2000) A novel mycolic acid cyclopropane synthetase is required for cording, persistence, and virulence of *Mycobacterium tuberculosis*. Mol Cell 5:717–727

Godreuil S, Torrea G, Terru D, Chevenet F, Diagbouga S, Supply P, Van De Perre P, Carriere C, Bañuls AL (2007) First molecular epidemiology study of *Mycobacterium tuberculosis* in Burkina Faso. J Clin Microbiol 45:921–927

Golby P, Hatch KA, Bacon J, Cooney R, Riley P, Allnutt J, Hinds J, Nunez J, Marsh PD, Hewinson RG, Gordon SV (2007) Comparative transcriptomics reveals key gene expression differences between the human and bovine pathogens of the *Mycobacterium tuberculosis* complex. Microbiology 153:3323–3336

Gonzalo-Asensio J, Malaga W, Pawlik A, Astarie-Dequeker C, Passemar C, Moreau F, Laval FO, Daffe M, Martin C, Brosch R, Guilhot C (2014) Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence regulator. Proc Natl Acad Sci 111:11491–11496

Goren MB, Brokl O, Schaefer WB (1974a) Lipids of putative relevance to virulence in *Mycobacterium tuberculosis*: correlation of virulence with elaboration of sulfatides and strongly acidic lipids. Infect Immun 9:142–149

Goren MB, Brokl O, Schaefer WB (1974b) Lipids of putative relevance to virulence in *Mycobacterium tuberculosis*: phthiocerol dimycocerosate and the attenuation indicator lipid. Infect Immun 9:150–158

Goren MB, Grange JM, Aber VR, Allen BW, Mitchison DA (1982) Role of lipid content and hydrogen peroxide susceptibility in determining the guinea-pig virulence of *Mycobacterium tuberculosis*. Br J Exp Pathol 63:693–700

Grange JM, Aber VR, Allen BW, Mitchison DA, Mikhail JR, Mcswiggan DA, Collins CH (1977) Comparison of strains of *Mycobacterium tuberculosis* from british, ugandan and asian immigrant patients: a study in bacteriophage typing, susceptibility to hydrogen peroxide and sensitivity to thiophen-2-carbonic acid hydrazide. Tubercle 58:207–215

Grange JM, Aber VR, Allen BW, Mitchison DA, Goren MB (1978) The correlation of bacteriophage types of *Mycobacterium tuberculosis* with guinea-pig virulence and in vitro-indicators of virulence. J Gen Microbiol 108:1–7

Groenheit R, Ghebremichael S, Svensson J, Rabna P, Colombatti R, Riccardi F, Couvin D, Hill VÇ, Rastogi N, Koivula T, Kçïllenius G (2011) The guinea-bissau family of *Mycobacterium tuberculosis* complex revisited. PLoS One 6:e18601

Hanekom M, Van Der Spuy GD, Streicher E, Ndabambi SL, Mcevoy CRE, Kidd M, Beyers N, Victor TC, Van Helden PD, Warren RM (2007) A recently evolved sublineage of the Mycobacterium tuberculosis Beijing strain family was associated with an increased ability to spread and cause disease. J Clin Microbiol JCM 45:1483–1490

He XY, Zhuang YH, Zhang XG, Li GL (2003) Comparative proteome analysis of culture supernatant proteins of *Mycobacterium tuberculosis* H37Rv and H37Ra. Microbes Infect 5:851–856

Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, Roach JC, Kremer K, Petrov DA, Feldman MW, Gagneux S (2008) High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. PLoS Biol 6:e311

Hett EC, Rubin EJ (2008) Bacterial growth and cell division: a mycobacterial perspective. Microbiol Mol Biol Rev 72:126–156. table of contents

Hinchey J, Lee S, Jeon BY, Basaraba RJ, Venkataswamy MM, Chen B, Chan J, Braunstein M, Orme IM, Derrick SC, Morris SL, Jacobs WR Jr, Porcelli SA (2007) Enhanced priming of adaptive immunity by a proapoptotic mutant of *Mycobacterium tuberculosis*. J Clin Invest 117:2279–2288

Hirsh AE, Tsolaki AG, Deriemer K, Feldman MW, Small PM (2004) Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. Proc Natl Acad Sci U S A 101:4871–4876

Homolka S, Niemann S, Russell DG, Rohde KH (2010) Functional genetic diversity among *Mycobacterium tuberculosis* complex clinical isolates: delineation of conserved core and lineage-specific transcriptomes during intracellular survival. PLoS Pathog 6:e1000988

Huet G, Constant P, Malaga W, Laneelle MA, Kremer K, Van Soolingen D, Daffe M, Guilhot C (2009) A lipid profile typifies the Beijing strains of *Mycobacterium tuberculosis*: identification of a mutation responsible for a modification of the structures of phthiocerol dimycocerosates and phenolic glycolipids. J Biol Chem 284:27101–27113

Jackett PS, Aber VR, Lowrie DB (1978) Virulence and resistance to superoxide, low pH and hydrogen peroxide among strains of *Mycobacterium tuberculosis*. J Gen Microbiol 104:37–45

Jhingan GD, Kumari S, Jamwal SV, Kalam H, Arora D, Jain N, Krishnakumaar L, Samal A, Rao KVS, Kumar D, Nandicoori VK (2016) Comparative proteomic analyses of avirulent, virulent and clinical strains of Mycobacterium tuberculosis identifies strain-specific patterns. J Biol Chem 291: 14257–14273

Jungblut PR, Schaible UE, Mollenkopf HJ, Zimny-Arndt U, Raupach B, Mattow J, Halada P, Lamer S, Hagens K, Kaufmann SH (1999) Comparative proteome analysis of *Mycobacterium tuberculosis* and Mycobacterium bovis BCG strains: towards functional genomics of microbial pathogens. Mol Microbiol 33:1103–1117

Kato-Maeda M, Kim EY, Flores L, Jarlsberg LG, Osmond D, Hopewell PC (2010) Differences among sublineages of the East-Asian lineage of *Mycobacterium tuberculosis* in genotypic clustering. Int J Tuberc Lung Dis 14:538–544

Kato-Maeda M, Shanley CA, Ackart D, Jarlsberg LG, Shang S, Obregon-Henao A, Harton M, Basaraba RJ, Henao-Tamayo M, Barrozo JC, Rose J, Kawamura LM, Coscolla M, Fofanov VY, Koshinsky H, Gagneux S, Hopewell PC, Ordway DJ, Orme IM (2012) Beijing sublineages of *Mycobacterium tuberculosis* differ in pathogenicity in the guinea pig. Clin Vaccine Immunol 19:1227–1237

Koch A, Mizrahi V, Warner DF (2014) The impact of drug resistance on *Mycobacterium tuberculosis* physiology: what can we learn from rifampicin? Emerg Microbes Infect 3:e17

Kong Y, Cave MD, Yang D, Zhang L, Marrs CF, Foxman B, Bates JH, Wilson F, Mukasa LN, Yang ZH (2005) Distribution of insertion- and deletion-associated genetic polymorphisms among four *Mycobacterium tuberculosis* phospholipase C genes and associations with extrathoracic tuberculosis: a population-based study. J Clin Microbiol 43:6048–6053

Kong Y, Cave MD, Zhang L, Foxman B, Marrs CF, Bates JH, Yang ZH (2006) Population-based study of deletions in five different genomic regions of *Mycobacterium tuberculosis* and possible clinical relevance of the deletions. J Clin Microbiol 44:3940–3946

Kong Y, Cave MD, Zhang L, Foxman B, Marrs CF, Bates JH, Yang ZH (2007) Association between *Mycobacterium tuberculosis* Beijing/W lineage strain infection and extrathoracic tuberculosis: insights from epidemiologic and clinical characterization of the three principal genetic groups of M. tuberculosis clinical isolates. J Clin Microbiol 45:409–414

Krishnan N, Malaga W, Constant P, Caws M, Thi Hoang Chau T, Salmons J, Thi Ngoc Lan N, Bang ND, Daffç M, Young DB, Robertson BD, Guilhot C, Thwaites GE (2011) *Mycobacterium tuberculosis* lineage influences innate immune response and virulence and is associated with distinct cell envelope lipid profiles. PLoS One 6:e23870

Kubica T, Rusch-Gerdes S, Niemann S (2004) The Beijing genotype is emerging among multidrug-resistant *Mycobacterium tuberculosis* strains from Germany. Int J Tuberc Lung Dis 8:1107–1113

Kwan CK, Ernst JD (2011) HIV and tuberculosis: a deadly human syndemic. Clin Microbiol Rev 24:351–376

Langlois-Klassen D, Senthilselvan A, Chui L, Kunimoto D, Saunders LD, Menzies D, Long R (2013) Transmission of *Mycobacterium tuberculosis* Beijing strains, Alberta, Canada, 1991–2007. Emerg Infect Dis 19: 701–711

Laochumroonvorapong P, Paul S, Manca C, Freedman VH, Kaplan G (1997) Mycobacterial growth and sensitivity to H2O2 killing in human monocytes in vitro. Infect Immun 65:4850–4857

Lee J, Remold HG, Ieong MH, Kornfeld H (2006) Macrophage apoptosis in response to high intracellular burden of *Mycobacterium tuberculosis* is mediated by a novel caspase-independent pathway. J Immunol 176:4267–4274

Li Q, Whalen CC, Albert JM, Larkin R, Zukowski L, Cave MD, Silver RF (2002) Differences in rate and variability of intracellular growth of a panel of *Mycobacterium tuberculosis* clinical isolates within a human monocyte model. Infect Immun 70:6489–6493

Lillebaek T, Andersen AB, Dirksen A, Glynn JR, Kremer K (2003) *Mycobacterium tuberculosis* Beijing genotype. Emerg Infect Dis 9:1553–1557

Lygizos M, Shenoi S, Brooks R, Bhushan A, Brust J, Zelterman D, Deng Y, Northrup V, Moll A, Friedland G (2013) Natural ventilation reduces high TB transmission risk in traditional homes in rural KwaZulu-Natal, South Africa. BMC Infect Dis 13:300

Målen H, De Souza GA, Pathak S, Søfteland T, Wiker HG (2011) Comparison of membrane proteins of *Mycobacterium tuberculosis* H37Rv and H37Ra strains. BMC Microbiol 11:1–10

Manca C, Paul S, Barry CE III, Freedman VH, Kaplan G (1999) *Mycobacterium tuberculosis* catalase and peroxidase activities and resistance to oxidative killing in human monocytes in vitro. Infect Immun 67:74–79

Manca C, Tsenova L, Bergtold A, Freeman S, Tovey M, Musser JM, Barry CE, Freedman VH, Kaplan G (2001) Virulence of a *Mycobacterium tuberculosis* clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFN g. Proc Natl Acad Sci U S A 98:5752–5757

Manca C, Reed MB, Freeman S, Mathema B, Kreiswirth B, Barry CE III, Kaplan G (2004) Differential monocyte activation underlies strain-specific *Mycobacterium tuberculosis* pathogenesis. Infect Immun:5511–5514

Manca C, Tsenova L, Freeman S, Barczak AK, Tovey M, Murray PJ, Barry C, Kaplan G (2005) Hypervirulent M. tuberculosis W/Beijing strains upregulate type I IFNs and increase expression of negative regulators of the Jak-Stat pathway. J Interf Cytokine Res 25:694–701

Marais BJ, Hesseling AC, Schaaf HS, Gie RP, Van Helden PD, Warren RM (2009) *Mycobacterium tuberculosis* transmission is not related to household genotype in a setting of high endemicity. J Clin Microbiol 47: 1338–1343

Mcdonough KA, Kress Y, Bloom BR (1993) Pathogenesis of tuberculosis: interaction of *Mycobacterium tuberculosis* with macrophages. Infect Immun 61:2763–2773

Melnyk AH, Wong A, Kassen R (2015) The fitness costs of antibiotic resistance mutations. Evolutionary applications. 8(3):273–283

Mihret A, Bekele Y, Loxton AG, Aseffa A, Howe R, Walzl G (2012) Plasma level of IL-4 differs in patients infected with different modern lineages of M. tuberculosis. J Trop Med 2012:518564

Miller BH, Shinnick TM (2000) Evaluation of *Mycobacterium tuberculosis* genes involved in resistance to killing by human macrophages. Infect Immun 68: 387–390

Mostowy S, Cleto C, Sherman DR, Behr MA (2004) The *Mycobacterium tuberculosis* complex transcriptome of attenuation. Tuberculosis (Edinb) 84:197–204

Nahid P, Bliven EE, Kim EY, Mac Kenzie WR, Stout JE, Diem L, Johnson JL, Gagneux S, Hopewell PC, Kato-Maeda M, The Tuberculosis Trials, C (2010) Influence of M. tuberculosis lineage variability within a clinical trial for pulmonary tuberculosis. PLoS One 5:e10753

Newton SM, Smith RJ, Wilkinson KA, Nicol MP, Garton NJ, Staples KJ, Stewart GR, Wain JR, Martineau AR, Fandrich S, Smallie T, Foxwell B, Al Obaidi A, Shafi J, Rajakumar K, Kampmann B, Andrew PW, Ziegler-Heitbrock L, Barer MR, Wilkinson RJ (2006) A deletion defining a common Asian lineage of *Mycobacterium tuberculosis* associates with immune subversion. Proc Natl Acad Sci 103:15594–15598

Nicol MP, Sola C, February B, Rastogi N, Steyn L, Wilkinson RJ (2005) Distribution of strain families of *Mycobacterium tuberculosis* causing pulmonary and extrapulmonary disease in hospitalized children in Cape Town, South Africa. J Clin Microbiol 43: 5779–5781

Niobe-Eyangoh SN, Kuaban C, Sorlin P, Cunin P, Thonnon J, Sola C, Rastogi N, Vincent V, Gutierrez MC (2003) Genetic biodiversity of *Mycobacterium tuberculosis* complex strains from patients with pulmonary tuberculosis in Cameroon. J Clin Microbiol 41: 2547–2553

North RJ, Ryan L, Lacource R, Mogues T, Goodrich ME (1999) Growth rate of Mycobacteria in mice as an unreliable indicator of Mycobacterial virulence. Infect Immun 67:5483–5485

O'Brien L, Carmichael J, Lowrie DB, Andrew PW (1994) Strains of *Mycobacterium tuberculosis* differ in susceptibility to reactive nitrogen intermediates in vitro. Infect Immun 62:5187–5190

O'Garra A, Redford PS, Mcnab FW, Bloom CI, Wilkinson RJ, Berry MPR (2013) The immune response in tuberculosis. Annu Rev Immunol 31:475–527

Oddo M, Renno T, Attinger A, Bakker T, Macdonald HR, Meylan PR (1998) Fas ligand-induced apoptosis of infected human macrophages reduces the viability of intracellular *Mycobacterium tuberculosis*. J Immunol 160:5448–5454

Ogarkov O, Mokrousov I, Sinkov V, Zhdanova S, Antipina S, Savilov E (2012) Lethal combination of *Mycobacterium tuberculosis* Beijing genotype and human CD209 -336G allele in Russian male population. Infect Genet Evol 12:732–736

Olsen I, Balasingham SV, Davidsen T, Debebe E, Rodland EA, Van Soolingen D, Kremer K, Alseth I, Tønjum T (2009) Characterization of the major formamidopyrimidine-DNA glycosylase homolog in *Mycobacterium tuberculosis* and its linkage to variable tandem repeats. FEMS Immunol Med Microbiol 56:151–161

Ordway DJ, Sonnenberg MG, Donahue SA, Belisle JT, Orme IM (1995) Drug-resistant strains of *Mycobacterium tuberculosis* exhibit a range of virulence for mice. Infect Immun 63:741–743

Ordway D, Henao-Tamayo M, Harton M, Palanisamy G, Troudt J, Shanley C, Basaraba RJ, Orme IM (2007) The hypervirulent *Mycobacterium tuberculosis* strain HN878 induces a potent TH1 response followed by rapid down-regulation. J Immunol 179: 522–531

Pareek M, Evans J, Innes J, Smith G, Hingley-Wilson S, Lougheed KE, Sridhar S, Dedicoat M, Hawkey P, Lalvani A (2013) Ethnicity and mycobacterial lineage as determinants of tuberculosis disease phenotype. Thorax 68:221–229

Parsons SDC, Drewe JA, Van Pittius NCG, Warren RM, Van Heiden PD (2013) Novel cause of tuberculosis in meerkats, South Africa. Emerg Infect Dis 19: 2004–2007

Paul S, Laochumroonvorapong P, Kaplan G (1996) Comparable growth of virulent and avirulent *Mycobacterium tuberculosis* in human macrophages in vitro. J Infect Dis 174:105–112

Perez-Lago L, Navarro Y, Herranz M, Bouza E, Garcia-De-Viedma D (2013) Differences in gene expression between clonal variants of *Mycobacterium tuberculosis*

emerging as a result of microevolution. Int J Med Microbiol 303:674–677

Phares C, Wangroongsarb P, Chantra S, Paveenkitiporn W, Tondella ML, Benson R, Thacker WL, Fields B, Moore M, Fischer J, Dowell S, Olsen S (2007) Epidemiology of severe pneumonia caused by Legionella longbeachae, Mycoplasma pneumoniae, and Chlamydia pneumoniae: 1 year, population based surveillance for severe pneumonia in Thailand. Clin Infect Dis 45:e147–e155

Pheiffer C, Betts JC, Flynn HR, Lukey PT, Van Helden P (2005) Protein expression by a Beijing strain differs from that of another clinical isolate and *Mycobacterium tuberculosis* H37Rv. Microbiology 151:1139–1150

Portevin D, Gagneux S, Comas I, Young D (2011) Human macrophage responses to clinical isolates from the *Mycobacterium tuberculosis* complex discriminate between ancient and modern lineages. PLoS Pathog 7:e1001307

Portevin D, Sukumar S, Coscolla M, Shui G, Li B, Guan XL, Bendt AK, Young D, Gagneux S, Wenk MR (2014) Lipidomics and genomics of *Mycobacterium tuberculosis* reveal lineage-specific trends in mycolic acid biosynthesis. Microbiol Open 3:823–835

Rakotosamimanana N, Raharimanga V, Andriamandimby SF, Soares JL, Doherty TM, Ratsitorahina M, Ramarokoto H, Zumla A, Huggett J, Rook G, Richard V, Gicquel B, Rasolofo-Razanamparany V, The VVSG (2010) Variation in IFN-{gamma} responses to different infecting strains of *Mycobacterium tuberculosis* in AFB smear positive patients and household contacts in Antananarivo, Madagascar. Clin Vaccine Immunol CVI

Reed MB, Domenech P, Manca C, Su H, Barczak AK, Kreiswirth BN, Kaplan G, Barry CE (2004) A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. Nature 431:84–87

Reed MB, Pichler VK, Mcintosh F, Mattia A, Fallow A, Masala S, Domenech P, Zwerling A, Thibert L, Menzies D, Schwartzman K, Behr MA (2009) Major *Mycobacterium tuberculosis* lineages associate with patient country of Origin. J Clin Microbiol 47:1119–1128

Reiling N, Homolka S, Walter K, Brandenburg J, Niwinski L, Ernst M, Herzmann C, Lange C, Diel R, Ehlers S, Niemann S (2013) Clade-specific virulence patterns of *Mycobacterium tuberculosis* complex strains in human primary macrophages and aerogenically infected mice. mBio 4

Rhoades ER, Orme IM (1997) Susceptibility of a panel of virulent strains of *Mycobacterium tuberculosis* to reactive nitrogen intermediates. Infect Immun 65:1189–1195

Rocha-Ramirez LM, Estrada-Garcia I, Lopez-Marin LM, Segura-Salinas E, Mendez-Aragon P, Van Soolingen D, Torres-Gonzalez R, Chacon-Salinas R, Estrada-Parra S, Maldonado-Bernal C, Lopez-Macias C, Isibasi A (2008) *Mycobacterium tuberculosis* lipids regulate cytokines, TLR-2/4 and MHC class II expression in human macrophages. Tuberculosis 88:212–220

Rose G, Cortes T, Comas I, Coscolla M, Gagneux S, Young DB (2013) Mapping of genotype-phenotype diversity among clinical isolates of *Mycobacterium tuberculosis* by sequence-based transcriptional profiling. Genome Biol Evol 5:1849–1862

Rouse DA, Morris SL (1995) Molecular mechanisms of isoniazid resistance in *Mycobacterium tuberculosis* and Mycobacterium bovis. Infect Immun 63:1427–1433

Rousseau C, Sirakova TD, Dubey VS, Bordat Y, Kolattukudy PE, Gicquel B, Jackson M (2003) Virulence attenuation of two Mas-like polyketide synthase mutants of *Mycobacterium tuberculosis*. Microbiology 149:1837–1847

Rousseau C, Winter N, Pivert E, Bordat Y, Neyrolles O, Ave P, Huerre M, Gicquel B, Jackson M (2004) Production of phthiocerol dimycocerosates protects *Mycobacterium tuberculosis* from the cidal activity of reactive nitrogen intermediates produced by macrophages and modulates the early immune response to infection. Cell Microbiol 6:277–287

Ruckdeschel K, Roggenkamp A, Lafont V, Mangeat P, Heesemann J, Rouot B (1997) Interaction of Yersinia enterocolitica with macrophages leads to macrophage cell death through apoptosis. Infect Immun 65:4813–4821

Safi H, Barnes PF, Lakey DL, Shams H, Samten B, Vankayalapati R, Howard ST (2004) IS6110 functions as a mobile, monocyte-activated promoter in *Mycobacterium tuberculosis*. Mol Microbiol 52:999–1012

Sarkar R, Lenders L, Wilkinson KA, Wilkinson RJ, Nicol MP (2012) Modern lineages of *Mycobacterium tuberculosis* exhibit lineage-specific patterns of growth and cytokine induction in human monocyte-derived macrophages. PLoS One 7:e43170

Shell SS, Prestwich EG, Baek SH, Shah RR, Sassetti CM, Dedon PC, Fortune SM (2013) DNA methylation impacts gene expression and ensures hypoxic survival of *Mycobacterium tuberculosis*. PLoS Pathog 9:e1003419

Sherman DR, Voskuil M, Schnappinger D, Liao R, Harrell MI, Schoolnik GK (2001) Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding α-crystallin. Proc Natl Acad Sci 98:7534–7539

Sinsimer D, Huet G, Manca C, Tsenova L, Koo MS, Kurepina N, Kana B, Mathema B, Marras SAE, Kreiswirth BN, Guilhot C, Kaplan G (2008) The phenolic glycolipid of *Mycobacterium tuberculosis* differentially modulates the early host cytokine response but does not in itself confer hypervirulence. Infect Immun 76:3027–3036

Sirakova TD, Thirumala AK, Dubey VS, Sprecher H, Kolattukudy PE (2001) The *Mycobacterium tuberculosis pks2* gene encodes the synthase for the hepta- and octamethyl-branched fatty acids required for sulfolipid synthesis. J Biol Chem 276:16833–16839

Sirakova TD, Dubey VS, Cynamon MH, Kolattukudy PE (2003a) Attenuation of *Mycobacterium tuberculosis* by disruption of a mas-like gene or a chalcone synthase-like gene, which causes deficiency in dimycocerosyl phthiocerol synthesis. J Bacteriol 185:2999–3008

Sirakova TD, Dubey VS, Kim HJ, Cynamon MH, Kolattukudy PE (2003b) The largest open reading frame (pks12) in the *Mycobacterium tuberculosis* genome is involved in pathogenesis and dimycocerosyl phthiocerol synthesis. Infect Immun 71:3794–3801

Smith NH, Kremer K, Inwald J, Dale J, Driscoll JR, Gordon SV, Van Soolingen D, Glyn Hewinson R, Maynard Smith J (2006) Ecotypes of the *Mycobacterium tuberculosis* complex. J Theor Biol 239:220–225

Soto CY, Menendez MC, Perez E, Samper S, Gomez AB, Garcia MJ, Martin C (2004) IS6110 mediates increased transcription of the phoP virulence gene in a multidrug-resistant clinical isolate responsible for tuberculosis outbreaks. J Clin Microbiol 42:212–219

Stavrum R, Praygod G, Range N, Faurholt-Jepsen D, Jeremiah K, Faurholt-Jepsen M, Krarup H, Aabye M, Changalucha J, Friis H, Andersen A, Grewal H (2014) Increased level of acute phase reactants in patients infected with modern *Mycobacterium tuberculosis* genotypes in Mwanza, Tanzania. BMC Infect Dis 14:309

Subbaiah TV, Ramachandran K (1961) The consistency of the susceptibility to hydrogen peroxide of tubercle bacilli isolated from South Indian patients before treatment. Tubercle 42:322–324

Subbaiah TV, Mitchison DA, Selkon JB (1960) The susceptibility to hydrogen peroxide of Indian and British isoniazid-sensitive and isoniazid-resistant tubercle bacilli. Tubercle 41:323–333

Subbian S, Bandyopadhyay N, Tsenova L, O'Brien P, Khetani V, Kushner N, Peixoto B, Soteropoulos P, Bader J, Karakousis P, Fallows D, Kaplan G (2013) Early innate immunity determines outcome of *Mycobacterium tuberculosis* pulmonary infection in rabbits. Cell Commun Signal 11:60

Supply P, Marceau M, Mangenot S, Roche D, Rouanet C, Khanna V, Majlessi L, Criscuolo A, Tap J, Pawlik A, Fiette L, Orgeur M, Fabre M, Parmentier C, Frigui W, Simeone R, Boritsch EC, Debrie AS, Willery E, Walker D, Quail MA, Ma L, Bouchier C, Salvignol G, Sayes F, Cascioferro A, Seemann T, Barbe V, Locht C, Gutierrez MC, Leclerc C, Bentley SD, Stinear TP, Brisse S, Medigue C, Parkhill J, Cruveiller S, Brosch R (2013) Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of *Mycobacterium tuberculosis*. Nat Genet 45:172–179

Tantivitayakul P, Panapruksachat S, Billamas P, Palittapongarnpim P (2010) Variable number of tandem repeat sequences act as regulatory elements in *Mycobacterium tuberculosis*. Tuberculosis 90:311–318

Tanveer M, Hasan Z, Kanji A, Hussain R, Hasan R (2009) Reduced TNF-[alpha] and IFN-[gamma] responses to Central Asian strain 1 and Beijing isolates of *Mycobacterium tuberculosis* in comparison with H37Rv strain. Trans R Soc Trop Med Hyg 103:581–587

Theus S-A, Cave MD, Eisenach K-D (2005) Intracellular macrophage growth rates and cytokine profiles of *Mycobacterium tuberculosis* strains with different transmission dynamics. J Infect Dis 191:453–460

Theus SA, Cave MD, Eisenach K, Walrath J, Lee H, Mackay W, Whalen C, Silver RF (2006) Differences in the growth of paired Ugandan Isolates of *Mycobacterium tuberculosis* within human mononuclear phagocytes correlate with epidemiological evidence of strain virulence. Infect Immun 74:6865–6876

Tientcheu LD, Sutherland JS, De Jong BC, Kampmann B, Jafali J, Adetifa IM, Antonio M, Dockrell HM, Ota MO (2014) Differences in T-cell responses between *Mycobacterium tuberculosis* and *Mycobacterium africanum*-infected patients. Eur J Immunol 44:1387–1398

Torrelles JB, Schlesinger LS (2010) Diversity in *Mycobacterium tuberculosis* mannosylated cell wall determinants impacts adaptation to the host. Tuberculosis 90:84–93

Torrelles JB, Knaup R, Kolareth A, Slepushkina T, Kaufman TM, Kang P, Hill PJ, Brennan PJ, Chatterjee D, Belisle JT, Musser JM, Schlesinger LS (2008) Identification of *Mycobacterium tuberculosis* clinical isolates with altered phagocytosis by human macrophages due to a truncated lipoarabinomannan. J Biol Chem 283:31417–31428

Tsenova L, Ellison E, Harbacheuski R, Moreira A-L, Kurepina N, Reed M-B, Mathema B, Barry C-E III, Kaplan G (2005) Virulence of selected *Mycobacterium tuberculosis* clinical isolates in the rabbit model of meningitis is dependent on phenolic glycolipid produced by the Bacilli. J Infect Dis 192:98–106

Tsolaki AG, Hirsh AE, Deriemer K, Enciso JA, Wong MZ, Hannan M, De La Salmoniere Y-O, Aman K, Kato-Maeda M, Small PM (2004) Functional and evolutionary genomics of *Mycobacterium tuberculosis* : Insights from genomic deletions in 100 strains. Proc Natl Acad Sci U S A 101:4865–4870

Tuite AR, Guthrie JL, Alexander DC, Whelan MS, Lee B, Lam K, Ma J, Fisman DN, Jamieson FB (2013) Epidemiological evaluation of spatiotemporal and genotypic clustering of *Mycobacterium tuberculosis* in Ontario, Canada. Int J Tuberc Lung Dis 17:1322–1327

Valway SE, Sanchez MP, Shinnick TF, Orme I, Agerton T, Hoy D, Jones JS, Westmoreland H, Onorato IM (1998) An outbreak involving extensive transmission of a virulent strain of *Mycobacterium tuberculosis*. N Engl J Med 338:633–639

Van Crevel R, Nelwan RHH, De Lenne W, Veeraragu Y, Van Der Zanden AG, Amin Z, Van Der Meer JWM, Van Soolingen D (2001) *Mycobacterium tuberculosis* Beijing genotype strains associated with febrile response to treatment. Emerg Infect Dis 7:880–883

Van Der Spuy GD, Warren RM, Richardson M, Beyers N, Behr MA, Van Helden PD (2003) Use of genetic distance as a measure of ongoing transmission of *Mycobacterium tuberculosis*. J Clin Microbiol 41:5640–5644

Van Der Spuy GD, Kremer K, Ndabambi SL, Beyers N, Dunbar R, Marais BJ, Van Helden PD, Warren RM (2009) Changing *Mycobacterium tuberculosis* population highlights clade-specific pathogenic characteristics. Tuberculosis 89:120–125

Van Laarhoven A, Mandemakers JJ, Kleinnijenhuis J, Enaimi M, Lachmandas E, Joosten LAB, Ottenhoff

THM, Netea MG, Van Soolingen D, Van Crevel R (2013) Low induction of proinflammatory cytokines parallels evolutionary success of modern strains within the *Mycobacterium tuberculosis* Beijing genotype. Infect Immun 81:3750–3756

Van Soolingen D, Hoogenboezem T, De Haas PEW, Hermans PWM, Koedam MA, Teppema KS, Brennan PJ, Besra GS, Portaels F, Top J, Schouls LM, Van Embden JDA (1997) A novel pathogenic taxon of the *Mycobacterium tuberculosis* complex, Canetti: characterization of an exceptional isolate from Africa. Int J Syst Bacteriol 47:1236–1245

Vander Beken S, Al Dulayymi JR, Naessens T, Koza G, Maza-Iglesias M, Rowles R, Theunissen C, De Medts J, Lanckacker E, Baird MS, Grooten J (2011) Molecular structure of the *Mycobacterium tuberculosis* virulence factor, mycolic acid, determines the elicited inflammatory pattern. Eur J Immunol 41:450–460

Velmurugan K, Chen B, Miller JL, Azogue S, Gurses S, Hsu T, Glickman M, Jacobs WR Jr, Porcelli SA, Briken V (2007) *Mycobacterium tuberculosis* nuoG is a virulence gene that inhibits apoptosis of infected host cells. PLoS Pathog 3:e110

Vergne I, Fratti RA, Hill PJ, Chua J, Belisle J, Deretic V. Mycobacterium tuberculosis Phagosome Maturation Arrest: Mycobacterial Phosphatidylinositol Analog Phosphatidylinositol Mannoside Stimulates Early Endosomal Fusion. Molecular Biology of the Cell. 2004;15(2):751–60.

Via LE, Weiner DM, Schimel D, Lin PL, Dayao E, Tankersley SL, Cai Y, Coleman MT, Tomko J, Paripati P, Orandle M, Kastenmayer RJ, Tartakovsky M, Rosenthal A, Portevin D, Eum SY, Lahouar S, Gagneux S, Young DB, Flynn JL, Barry CE (2013) Differential virulence and disease progression following *Mycobacterium tuberculosis* complex infection of the common marmoset (Callithrix jacchus). Infect Immun 81:2909–2919

Voskuil MI, Schnappinger D, Visconti KC, Harrell MI, Dolganov GM, Sherman DR, Schoolnik GK (2003) Inhibition of respiration by nitric oxide induces a *Mycobacterium tuberculosis* dormancy program. J Exp Med 198:705–713

Wampande E, Mupere E, Debanne S, Asiimwe B, Nsereko M, Mayanja H, Eisenach K, Kaplan G, Boom H, Gagneux S, Joloba M (2013) Long-term dominance of *Mycobacterium tuberculosis* Uganda family in peri-urban Kampala-Uganda is not associated with cavitary disease. BMC Infect Dis 13:484

Wang C, Peyron P, Mestre O, Kaplan G, Van Soolingen D, Gao Q, Gicquel B, Neyrolles O (2010) Innate immune response to *Mycobacterium tuberculosis* Beijing and other genotypes. PLoS One 5:e13594

Weiner B, Gomez J, Victor TC, Warren RM, Sloutsky A, Plikaytis BB, Posey JE, Van Helden PD, Gey Van Pittius NC, Koehrsen M, Sisk P, Stolte C, White J, Gagneux S, Birren B, Hung D, Murray M, Galagan J (2012) Independent large scale duplications in multiple M. tuberculosis lineages overlapping the same genomic region. PLoS One 7:e26038

Wickstrum JR, Bokhari SM, Fischer JL, Pinson DM, Yeh HW, Horvat RT, Parmely MJ (2009) Francisella tularensis induces extensive caspase-3 activation and apoptotic cell death in the tissues of infected mice. Infect Immun 77:4827–4836

Wirth T, Hildebrand F, Allix-Beguec C, Wölbeling F, Kubica T, Kremer K, Van Soolingen D, Rüsch-Gerdes S, Locht C, Brisse S, Meyer A, Supply P, Niemann S (2008) Origin, spread and demography of the *Mycobacterium tuberculosis* complex. PLoS Pathog 4:e1000160

Yang Z, Yang D, Kong Y, Zhang L, Marrs CF, Foxman B, Bates JH, Wilson F, Cave MD (2005) Clinical relevance of *Mycobacterium tuberculosis* plcD gene mutations. Am J Respir Crit Care Med 171: 1436–1442

Yindeeyoungyeon W, Likitvivatanavong S, Palittapongarnpim P (2009) Characterization of alpha-isopropylmalate synthases containing different copy numbers of tandem repeats in *Mycobacterium tuberculosis*. BMC Microbiol 9:122

Yu G, Fu X, Jin K, Zhang L, Wu W, Cui Z, Hu Z, Li Y (2011) Integrative analysis of transcriptome and genome indicates two potential genomic islands are associated with pathogenesis of *Mycobacterium tuberculosis*. Gene 489:21–29

Zhang M, Gong J, Yang Z, Samten B, Cave MD, Barnes P-F (1999) Enhanced capacity of a widespread strain of *Mycobacterium tuberculosis* to grow in human macrophages. J Infect Dis 179:1213–1217

Zhou X, Mantis N, Zhang X-R, Potoka DA, Watkins SC, Ford HR (2000) Salmonella typhimurium induces apoptosis in human monocyte-derived macrophages. Microbiol Immunol 44:987–995

Zhu L, Zhong J, Jia X, Liu G, Kang Y, Dong M, et al. (2016) Precision methylome characterization of *Mycobacterium tuberculosis* complex (MTBC) using PacBio single-molecule real-time (SMRT) technology. Nucleic Acids Res. 44(2):730–43.

Zumla A, Raviglione M, Hafner R, Fordham Von Reyn C (2013) Tuberculosis. N Engl J Med 368: 745–755

The Biology and Epidemiology of *Mycobacterium africanum*

6

Dorothy Yeboah-Manu, Bouke C. de Jong, and Florian Gehre

Abstract

West Africa is the only region in the world where six out of seven mycobacterial lineages of human importance are endemic. In particular, two evolutionary ancient lineages, *Mycobacterium africanum* West Africa 1 (MTBC Lineage 5) and *M. africanum* West Africa 2 (MTBC Lineage 6) are of interest as they cause up to 40% of all pulmonary TB cases in some West African countries. Although these *M. africanum* lineages are closely related to *M. tuberculosis* sensu stricto lineages, they differ significantly in respect to biology, epidemiology and in their potential to cause disease in humans. Most importantly the *M. africanum* lineages are exclusive to West Africa. Although the exact mechanisms underlying this geographical restriction are still not understood, it is increasingly suspected that this is due to an adaptation of the bacteria to West African host populations. In this chapter, we summarize the geographical distribution of the *M. africanum* lineages within the region, describe biological and clinical differences and the consequent implications for TB control in West Africa. We also try to shed light on the geographical restriction, based on recently published analyses on whole genomes of *M. africanum* isolates.

Keywords

Mycobacterium africanum • Lineages • Epidemiology • Genomic Diversity • Regions of Differences

D. Yeboah-Manu (✉)
Noguchi Memorial Institute for Medical Research, University of Ghana, Accra, Ghana
e-mail: dyeboah-manu@noguchi.ug.edu.gh

B.C. de Jong
Institute for Tropical Medicine, Antwerp, Belgium
e-mail: bdejong@itg.be

F. Gehre
Institute for Tropical Medicine, Antwerp, Belgium

Medical Research Council (MRC) Unit, The Gambia
Serrekunda, Gambia
e-mail: fgehre@mrc.gm; fgehre@itg.be

S. Gagneux (ed.), *Strain Variation in the* Mycobacterium tuberculosis *Complex: Its Role in Biology, Epidemiology and Control*, Advances in Experimental Medicine and Biology 1019,
DOI 10.1007/978-3-319-64371-7_6

6.1 General Introduction and History

The pathogen *Mycobacterium africanum* (Maf) was first isolated from an expectorate of a Senegalese patient with pulmonary tuberculosis (TB), and described by Castets and colleagues (1968) as a member of the tuberculosis causing mycobacteria with phenotypic characteristics that place it in an intermediate position between *M. tuberculosis* (Mtb) and *M. bovis* (David et al. 1978). Later, Meissner described similar mycobacterial isolates in Ghana in 1969 (Meissner and Schroder 1969) and related organisms were also isolated in Democratic Republic of Congo and Uganda (Sticht-Groh et al. 1994), and subsequently from other African countries (Haas et al. 1997; Frothingham et al. 1999). However, in contrast to Mtb and *M. bovis*, Maf strains isolated from the different countries showed higher degrees of variability in their phenotypic attributes (Haas et al. 1997; Kato-Maeda et al. 2001). This heterogeneity of Maf complicated its unequivocal species identification that could have led to the misclassification of clinical strains. Prat and colleagues concluded and speculated that the biochemical heterogeneity observed among Maf strains represents a continuous spectrum linking the classical human and bovine variants of Mtb (Prat et al. 1974). This made other scientists question the wisdom in the separation of the three pathogens into three species (Pattyn et al. 1970), and suggested its designation as subspecies or biovar status using different classification approaches including Adansonian classification (Tsukamura 1967, 1976; David et al. 1978); DNA homology (Bradley 1972; Baess 1979) and immune-diffusion analysis (Stanford and Grange 1974). Grange in UK in 1979 (Grange 1979) therefore described them all as Mtb with designation as human, bovine or African strains, and Tsukamura (1976) was the first to refer to them as part of the '*M. tuberculosis* complex' (MTBC).

Based on a battery of biochemical assays (Table 6.1), results of which varied with geographical origin, Maf was divided into two major subgroups: Maf subtype I originating from West Africa and Maf subtype II from East Africa (David et al. 1978; Vasconcellos et al. 2010), with Maf subtype I more closely related to *M. bovis*, whereas subtype II more closely resembling Mtb (Table 6.1). However, based on modern molecular analysis, we now know that Maf constitutes two distinct lineages that branched at an early stage in evolution of the MTBC from the "modern" Mtb lineages common to America, Europe and Asia (see Chap. 1) (Smith et al. 2009; Galagan 2014).

Early comparative genomic analysis also indicated that Maf subtype I strains from West African countries were characterized by the presence of low or medium numbers of IS*6110* bands and different spoligotype signatures (absence of spacers 7 through 12 and 37–39) from the subtype II (see Chap. 3) (de Jong et al. 2009; Streicher et al. 2007). Moreover, they showed a characteristic genomic deletion (RD9) and distinct *gyrB* DNA sequences that permitted their clear distinction from Mtb (Sola et al. 2003; Parsons et al. 2002). RD12 and RD4 are further genomic regions, intact in MAF I yet absent in *M. bovis* that allow their differentiation. Maf strains from Uganda (subtype II/ East African) on the contrary showed high IS*6110* copy numbers, spoligotype pattern similar to that of Mtb (absence of spacers 33 through 36), and identical *gyrB* sequence, presence of the RD9 region, and the Mtb specific deletion TbD1 (Brosch et al. 2002; Niemann et al. 2002).

Based on the above studies and others, Maf type II (East African clade) has been reclassified as Mtb genotype "Uganda", a sublineage of Mtb Lineage 4 (also known as the Euro-American lineage) (Sola et al. 2003; Mostowy et al. 2004). The 'true' Maf is therefore Maf subtype 1 (West African clade), which has further been sub-divided into two separate lineages that are genetically diverse; i.e. Maf West African type 1 (Maf1) or Lineage 5, characterized by the deletion of RD711, prevalent around the Eastern part of West Africa (Gulf of Guinea) (Gagneux et al. 2006), and Maf West African type 2 (Maf2) or Lineage 6 characterized by the deletion of RD702, (Gagneux et al. 2006; Mostowy et al. 2004) prevalent in Western West Africa; with countries in the middle harboring both variants.

Table 6.1 Biochemical assays used for distinguishing among *M. tuberculosis*, *M. bovis* and *M. africanum*

Biochemical test	*M. tuberculosis*	*M. africanum* East Africa variant	*M. africanum* West Africa variant I	*M. africanum* West Africa variant II	*M. bovis*	References
Colony morphology	Eugonic	Dysgonic	Dysgonic	Dysgonic	Dysgonic	Wayne and Kubica (1986), Kallenius et al. (1999), Hoffner et al. (1993), and David et al. (1978)
Niacin production	+	+/−	−/+	−/+	−	Wayne and Kubica (1986), Kallenius et al. (1999), Hoffner et al. (1993), and David et al. (1978)
Pyrazinamide[a]	S	S	S	S	R	Kallenius et al. (1999), Niemann et al. (2000), (2002), Hoffner et al. (1993). Pattyn et al. (1970), and Leao et al. (2004)
Nitrate reduction	+	−/+	−	−	−	Wayne and Kubica (1986), Kallenius et al. (1999), Hoffner et al. (1993), and Leao et al. (2004)
Thiophene-2-carboxylic acid hydrazide	R	R	S	S	S	Kallenius et al. (1999), Hoffner et al. (1993), Niemann et al. (2002), and David et al. (1978)
Oxygen preference	Aerophillic	Microaerophilic	Microaerophilic	Microaerophilic	Microaerophilic	Collins et al. (1982), Niemann et al. (2002), Pattyn et al. (1970), and Leao et al. (2004)
Preferred carbon source	Glycerol	Glycerol	Pyruvate	Pyruvate	Pyruvate	David et al. (1978), Chavadi et al. (2009), Keating et al. (2005), and Van Der Groen and Pattyn (1975)

[a]*M. africanum* is usually sensitive but pyrazinamide resistant strains have been identified among multi-resistant strains

Within the subsequent sections of this chapter, 'Maf' will refer only to what was previously known as Maf subtype I including both MTBC Lineage 5 and 6.

6.2 Epidemiology of Maf

6.2.1 Prevalence and Geographical Distribution of Maf in Humans

Though Maf has sporadically been identified in other regions of the world, the pathogen is almost exclusively restricted to West-Africa, where a prevalence of more than 50% among smear-positive TB cases has been indicated in distinct countries and time periods (Huet et al. 1971; Kallenius et al. 1999; de Jong et al. 2010). Maf1 has been isolated as far East as the Democratic Republic of Congo (Hermans-Boveroulle et al. 1965) and used to cause 56% of TB in Cameroon (Huet et al. 1971). Biochemical speciation at that time did not permit distinction between Maf1 and Maf2, although the present distribution suggests that all 56% were Maf1. The distribution of Maf1 and Maf2 overlap in the central part of West Africa or middle of West Africa, particularly Ghana, Benin (Affolabi et al. 2009), Sierra Leone and Ivory Coast (de Jong et al. 2010). Countries outside Western Africa where cases have been occasionally reported include: England (Grange and Yates 1989; Brown et al. 2010), France (Frottier et al. 1990; Grosset et al. 1974), Spain (Perez-de Pedro et al. 2008; Isea-Pena et al. 2012), Germany (Jungbluth et al. 1978; Schroder 1982), Italy (Garzelli et al. 2010), Denmark (Bek et al. 2010), California, USA (Desmond et al. 2004) and Columbia (Hurtado et al. 2016) from both pulmonary and extra-pulmonary sources. While most of the identified TB cases carrying Maf within these reports were African migrants, a few of the studies also identified other nationalities (Schroder 1982; Desmond et al. 2004). For example, the study of Desmond that retrospectively analysed an isolate collection of the California State using spoligotyping and IS*6110* restriction length polymorphism, identified 2/5 Maf isolates from a Vietnamese- and a US born African American patient, neither of whom had any travel history to Africa (Desmond et al. 2004), suggesting that transmission had occurred outside of Africa.

The prevalence of Maf estimated by various studies in West Africa showed that actual species and sublineage prevalence varies from country to country, region within a country, as well as period of sampling with figures ranging between 3% and 56%. During a national survey of primary resistance carried out from May 1995 to May 1996 in Côte d'Ivoire, using randomized cluster sputum sampling, 17 Maf were identified among 321 (5.3%) mycobacterial isolates and Maf was isolated from four regional centers (Dosso et al. 1999). In Nigeria, Ibadan (17%) and Nnewi (14%) both showed medium prevalence, while in Abuja, only 8% of all TB was due to Maf1 infection (Lawson et al. 2012). In a countrywide study in Senegal from the 1970s, the prevalence of Maf ranged from 10% in Casamance to 46% near the Senegal river (Diop et al. 1976). Likewise, in the southern sector of Ghana, a prevalence of 20% have been reported as compared to 26% in the north central sector.

There are some indications that Maf probably is being replaced by variants of Mtb. Based on phenotypic identification a study by Huet et al. conducted over 40 years ago, reported that 56% of cases of TB were due to Maf strains in the West and South regions of Cameroon (Huet et al. 1971). A reduced prevalence of 9% was however observed between 1997 and 1998, (Niobe-Eyangoh et al. 2003) and a recent publication in 2013 suggested that Maf1 has almost disappeared from Cameroon (Koro et al. 2013). Of the total 565 MTBC strains analysed, only 19 were identified as Maf, representing 3.3% of MTBC strains. The authors concluded that their findings showed a real regression of Maf compared to the 9% described in the 1990s ($P = 0.0001$), giving the impression that Maf is gradually being out-competed by the Cameroon sublineage of Lineage 4 Mtb. To confirm that the decline of Maf in Cameroon is not due to identification bias, the study conducted by Niobe-Eyango et al. used the same phenotypic assays as done by Huet et al., as well as molecular methods; both approaches gave almost the same prevalence

(9%) of Maf. Similar trends have been reported in Guinea Bissau; in 2011 Groenheit et al. reported a declining prevalence from 52% in 1994 to 38% in 2008 (Groenheit et al. 2011).

In contrast, two recent studies covering several regions of Nigeria identified a high Maf prevalence, ranging from 14% to 33% (Lawson et al. 2012; Thumamo et al. 2012), and detected active foci of recent Maf transmission in 2009–2010 (Lawson et al. 2012), which is consistent with the prevalence of 37% Maf recorded in Benin (Gehre et al. 2013a). Similarly, various studies conducted in Ghana spanning periods of about two decades using either phenotypic or genotypic assays have recorded a prevalence rate of not less than 20% (Lawn et al. 2001; Addo et al. 2007; Yeboah-Manu et al. 2011; Asante-Poku et al. 2015). Also in the Gambia, two independent studies reported in 2008 and 2013 indicated that the prevalence of Maf2 was consistently around 38%. Among isolates from 552 TB cases the prevalence of Maf 2 (n = 223) was 40%, (Gehre et al. 2013a), while the previous study conducted in the same country by de Jong et al. in 2008 recorded a prevalence of 39%. Other country prevalence values and year of report are indicated in Table 6.2. While some researchers have postulated that Maf may be more prevalent in rural West Africa, within our work in Ghana, where both Maf lineages are prevalent we did not observe this. We however, did find statistical association of Maf with an ethnic group (the Ewe ethnic group, one of the native tribes of coastal West Africa) driven by Lineage 5 (Asante-Poku et al. 2015, 2016).

6.2.2 Isolation of Maf from Animals

The phylogenetic proximity of Maf to the animal strains of the MTBC as well as its restricted geographic distribution, gave the indication that the pathogen may have an animal reservoir in West Africa. Different phylogenetic analyses of MTBC strains have always shown Maf, especially Maf2, is closer to animal adapted strains compared to other lineages of Mtb (de Jong et al. 2010). For example an MTBC strain isolated from a wild chimpanzee in Côte d'Ivoire that was shown by comparative genomic and phylogenomic analyses to belong to a new lineage of MTBC was most closely related to Maf2 (i.e. MTBC Lineage 6) and shares 32 SNPs with this lineage (Coscolla et al. 2013). Two other animal-associated members of the MTBC are known to cluster with lineage 6 rather than with the other classical animal-adapted lineages: *M. mungi* and the dassie bacillus, which infect African mongooses and hyraxes, respectively (Cousins et al. 1994; Alexander et al. 2010). One of the 32 SNPs shared between the chimpanzee bacillus genome and the Maf also occurred in *M. mungi* or dassie bacillus (Coscolla et al. 2013). Similar to Maf, the dassie bacillus was shown to lack the regions RD7, RD8, RD9, and RD10 (Mostowy et al. 2004). Moreover, the association of Maf2 with HIV infection suggests that the ratio of latent/active Maf2 infection is greater than for Mtb infections, maintenance of which could again indicate an animal reservoir sustaining zoonotic infection in humans.

However so far, no animal reservoir has been identified for Maf based on extensive skin testing and abattoir surveys in Senegal, Gambia, Guinea and Guinea Bissau not identifying mycobacterial infection and lesion (Unger et al. 2003). Nevertheless, Maf has been isolated from diverse animal species (Hart and Sutherland 1977; Thorel 1980a; Alfredsen and Saxegaard 1992). A study conducted in North-Central Nigeria, identified 1/5 Maf mycobacterial isolates from 400 unpasteurized milk (Cadmus et al. 2006), and in Ghana, Asante-Poku et al. identified 1/6 mycobacterial isolates obtained from 17 lesions of a cow carcass by spoligotyping as Maf (Asante-Poku et al. 2014). Other studies that have isolated Maf from cattle have been reported in Bangladesh (Rahim et al. 2007) and Bavaria (Weber et al. 1998). Furthermore, Maf has also been isolated from pigs during an outbreak of TB in pigs and cows in 1992, where lesions were found to be similar to those caused by *M. bovis* and *M. avium* (Alfredsen and Saxegaard 1992). Among wild animals, disseminated TB disease

Table 6.2 Prevalence of *M. africanum* from both human and animal sources as reported from different countries at different time periods

Country	Source	Sample size	*M. africanum* (%)	Method	References
Benin	Clinical	194	37.0	Spoligo. miru	Gehre et al. (2013a) and Affolabi et al. (2009)
Burkina Faso	Clinical	229	18.0	Biochemical	Simonet et al. (1989)
Burkina Faso	Clinical	300	18.4		Ledru et al. (1996)
Burkina Faso	Clinical	120	0	Spoligo	Godreuil et al. (2007)
Burkina Faso	Clinical	110	4.2	Biochemical	Sangare et al. (2010)
Burkina Faso	Clinical	72	22.2	Spoligo	Gomgnimbou et al. (2012)
Cameroun	Clinical		56.0	Biochemical	Huet et al. (1971)
Cameroun	Clinical	455	9.0	Spoligo	Niobe-Eyangoh et al. (2003)
Cameroon	Clinical	169	1.2	MIRU; spoligo	Assam et al. (2013)
Cameroun	Clinical	565	3.4	Spoligo	Koro et al. (2013)
Cameroun	Clinical	445	2.3	Biochemical, spoligo	Kuaban et al. (2014)
Congo R	Clinical	34	8.8	spoligo	Ontsira Ngoyi et al. (2014)
Djibouti	Clinical	435	0.5	Spoligo, VNTR	Blouin et al. (2012)
Egypt	Mummies	12	30.0	Spoligo	Zink et al. (2003)
Ethiopia	Clinical	168	9.5	spoligo	Nuru et al. (2015)
Ghana	Clinical	25	8.0	IS6110-RFLP	Lawn et al. (2001)
Ghana	Clinical	64	20.0	Biochemical	Addo et al. (2007)
Ghana	Clinical	1906	29.2	IS6110-RFLP, spoligo	Meyer et al. (2008)
Ghana	Clinical	325	28.9	Spoligo	Homolka et al. (2010)
Ghana	Clinical	162	19.7	Spoligo, LSP	Yeboah-Manu et al. (2011)
Ghana	Cattle	6	16.7	Spoligo, snp	Asante-Poku et al. (2014)
Ghana	Clinical	613	17.1	Spoligo	Asante-Poku et al. (2015)
Ghana	Clinical	2551	18.3	Spoligo	Yeboah-Manu et al. (Manuscript in preparation)
Guinea Bissau	Clinical	414	47.1	LSP; spoligo	Groenheit et al. (2011)
Lebanon	Clinical	60	1.7	Spoligo	Bedrossian et al. (2013)
Japan	Clinical	970	0.2	LSP	Ueyama et al. (2014)
Nigeria	Clinical	30	3.3	LSP, spoligo	Cadmus et al. (2009)
Nigeria	Clinical	375	5.3	Genotype MTBC	Aliyu et al. (2013)
Nigeria	Clinical	105	10.5	LSP	Waziri et al. (2014)
Nigeria	Clinical	180	8.9	Spoligo	Uzoewulu et al. (2016)
Sierra Leone	Clinical	97	23.7	Spoligo, LSP	Homolka et al. (2008)
The Gambia	Clinical	386	38.4	LSP	de Jong et al. (2009)

was observed in an adult female hyrax (*Procavia capensis*) imported from United Arab Emirates and held in captivity at a zoo in Croatia, and confirmed by genotyping to be caused by Maf (Gudan et al. 2008). Thorel et al. also identified isolates in the Veterinary Research Centre in France (Thorel 1980b).

6.3 Genomic Structure of Maf

Until recently, mycobacterial lineages were defined based on a phylogenetic classification system that considered the absence or presence of genomic deletions, namely regions of differences (RDs, see Chap. 1) (Tsolaki et al. 2004). These

original classification systems revealed that the two Maf lineages, together with the animal strains, belong to a group of evolutionary ancient mycobacterial lineages that share a common deletion of RD9. Maf1 has a further deletion in RD711, while Maf2 can be identified by a deletion of RD702 (de Jong et al. 2005). Using spoligotyping, Maf1 is characterised by a loss of spacers 8–12 and 37–39, while in Maf2 spacers 7–9 and spacer 39 are missing (de Jong et al. 2010). With the increased availability of whole genome sequencing data, lineage defining SNPs for either of the Maf lineages were identified and developed into SNP assays for rapid speciation (Stucki et al. 2012).

Although no in-depth phylogenetic studies were conducted to date, spoligotyping in combination with 12-loci MIRU-VNTR genotyping of Maf1 isolates from several West African countries (Sierra Leone, Nigeria, Benin) suggested the presence of at least ten sublienages within the Maf1 lineage, which differ in their capacity to spread within the human population (Gehre et al. 2013a). For Maf2, no comparable data is currently available.

6.4 Genomic Diversity and Implications for Epidemiological Differences Between Maf and Other MTBC Lineages

So far, only few whole genome sequencing studies of Maf isolates have been performed (Bentley et al. 2012; Gehre et al. 2013b; Winglee et al. 2016). Besides the reference genome for Maf2 (Bentley et al. 2012) and a study on a limited number of Maf2 isolates (Gehre et al. 2013b), a third sequencing study (the most comprehensive of all) on 24 Maf2 and 2 Maf1 isolates was recently published (Winglee et al. 2016). No reference genome for Maf1 is available at the moment, although efforts are ongoing and the present preliminary findings (Winglee et al. 2016) need to be confirmed in future large-scale studies. For Maf2, all studies described similar key observations. When compared to Mtb genomes, Maf2 is characterised by increased genome erosion due to a larger number of pseudogenes (Bentley et al. 2012). The above-mentioned decrease of Maf prevalence could be due to a replacement of Maf isolates by rapidly spreading Mtb isolates in West Africa. It was speculated that a loss of functional genes in the Maf genome and therefore a loss of fitness of Maf when compared to Mtb may ultimately lead to an inability of Maf to compete with newly introduced Mtb isolates in the region. If this was the case, one would expect that the genetic diversity of rapidly spreading Mtb lineages would be higher than in the slowly vanishing, decreasing Maf populations (Winglee et al. 2016). However, available sequencing data did not find any differences in genetic diversity between Maf and Mtb isolates, and therefore did not detect a genetic bottleneck indicative of a selective sweep that would have occurred during exerted selective pressure on Maf isolates following introduction of Mtb lineages (Winglee et al. 2016). Similarly, no quantitative difference in mutations of virulence genes between the Maf and Mtb lineages was found that could explain a loss of fitness and competitiveness. Therefore, the exact interaction between the various MTBC lineages and the reason for the geographical restriction of Maf to West Africa is still unknown. One suggested scenario based on sequence data was that while Mtb isolates may be able to establish infection independent of host ethnicity, Maf appears to be particularly well adapted to West African hosts. As Winglee et al. described, Mtb is one of a few bacterial species capable of Vitamin B12 synthesis essential for bacterial survival during Vitamin B12 limiting conditions within the host (Winglee et al. 2016). DNA sequence data however revealed that Maf2 isolates lost the capacity to synthesize Vitamin B12. As West Africans are speculated to have higher Vitamin B12 levels than other ethnicities, the auxotroph Maf will still encounter sufficient Vitamin B12 amounts for bacterial survival within these West African hosts, while the lack of available Vitamin B12 in other host ethnicities outside the region limits Maf growth. This hypothesis is supported by a recent publication investigating Maf infections in the US that found a strong association between

Maf infection and patients born in West Africa or of West African ancestry (Sharma et al. 2016).

6.5 Growth Requirement, Morphology and Biochemical Characteristics of Maf

Mycobacteriological laboratory methods have traditionally utilized a series of phenotypical tests based on growth rate, colony morphology, growth requirements and other biochemical properties to separate the classical members of the MTBC. In the past, Maf strains were generally identified by default, having initally ruled-out both Mtb and *M. bovis* by biochemical tests (Hoffner et al. 1993; Grange and Yates 1989). Maf prefers minimal oxygen (microaerophilic) and produces dysgonic colonies compared to the abundant growth (eugonic colonies) exhibited by Mtb. Similar to *M. bovis* and *M. microti*, Maf is unable to use glycerol as a sole carbon source, thus to improve isolation of Maf, solid medium are normally supplemented with 0.4–1.0% pyruvate for primary culture; this culture medium has shown a better recovery rate (Stonebrink 1958; Wayne and Kubica 1986), which has been attributed to the lack of pyruvate kinase activity. A SNP in *pykA* seems to affect the organization at the active site of the pyruvate kinase enzyme, including association with the co-factor (Mg^{2+}) and substrate [ADP/ATP] and phosphoenolpyruvate binding (Munoz and Ponce 2003).

Most Maf isolates also show catalase activity; in one study it was found that all Maf isolates analysed had catalase activity less than or equal to 10 mm (height of bubbles above agar) at 22 °C, and all were negative for catalase activity at 68 °C (Frothingham et al. 1999). Like *M. bovis*, Maf is nitrate negative, with a variable reaction to accumulation of niacin, between negative and a weak production of niacin. Comparison of the *narGHJI* promoter revealed that at nucleotide −215 prior to the start codon of *narG*, *Maf* has a cytosine residue, compared to Mtb, which carries a thymine residue (Stermann et al. 2003). Furthermore, Maf are susceptible to thiophene-2-carboxylic acid hydrazide (TCH) (5 mg/l) and pyrazinamide (PZA, positive pyrazinamidase activity), although, like Mtb, resistance to PZA can develop, typically reported among multi-drug resistant cases (Frothingham et al. 1999; Homolka et al. 2010). Very importantly, Maf grows more slowly than Mtb. From the first descriptions of Maf, the time to detection on solid medium was known to be longer when compared to Mtb (Grange and Yates 1989; Castets et al. 1968). Gehre et al. compared the growth rates of Mtb and Maf isolates in The Gambia in two liquid culture systems on primary isolation and found Maf to grow significantly more slowly than Mtb (Gehre et al. 2013a). This observation was confirmed in growth experiments under defined laboratory conditions.

6.6 Diagnostic Implications

The two Maf lineages are phylogenetically related yet distinct from the remaining Mtb lineages. Consequently, differences in the bacterial physiology were observed that not only impact on the direct detection of bacteria in the laboratory but also lead to clinical differences that need to be considered during patient diagnosis. As most of culture media and diagnostics were traditionally optimized using Mtb isolates and Mtb infected patients, the sensitivity in detecting Maf is not known for most of the commonly used assays. While clinical researchers in West Africa are slowly progressing in evaluating diagnostic assay performance for Maf2, data on Maf1 is still lacking.

A key first step of laboratory TB diagnostics is microscopy of sputa, with or without Xpert MTB/RIF, and culture on solid or liquid media for patients who need phenotypic drug susceptibility performed. As mentioned above, Maf2 isolates cannot metabolize glycerol and therefore require pyruvate-supplemented media to grow in vitro (Keating et al. 2005). This finding was demonstrated in traditional biochemical growth experiments, and therefore it was recommended to add pyruvate to Loewenstein-Jensen

slopes, which should be commonly done in Maf-endemic West African countries. However, the compositions of other solid media such as Kudoh/Ogawa slopes (Palaci et al. 2013; Jaspe et al. 2009; Kothadia and Sengupta 1982) or the recently described MOD9 plates (Asmar et al. 2015) as well as liquid culture systems such as the widely used BactecMGIT960 were never evaluated for their suitability for Maf culture. Similarly, as previous publications described a deficiency of Maf2 in vitamin B metabolism (Bentley et al. 2012; Winglee et al. 2016), non egg-based media (eggs contain vitamin B) supplemented with vitamin B might result in higher yields of Maf. Also, the impact that the preferable microaerophilic growth of Maf has on automated culture systems is not fully understood, as the majority of culture techniques are either under aerobic growth conditions or, in the case of the Bactec MGIT 960, use oxygen depletion as a proxy for bacterial proliferation. As a microaerophilic organism, Maf2 might consume less oxygen during growth.

Another factor that could influence Maf laboratory diagnosis is the pre-defined length of a 42 day incubation period for TB diagnosis of primary isolation from sputa, or a 12 day incubation period for phenotypic drug-susceptibility testing using the Bactec MGIT 960. As the doubling time of Maf2 was described to be significantly longer (24.12 h) compared to that of MTB (20.16 h) (Gehre et al. 2013b), it is conceivable that incubation periods in Maf endemic countries need to be extended. Therefore, in one of the first publications by Castets et al., an extended incubation period of 90 days for isolates from West Africa was already recommended (Castets et al. 1968), yet never implemented in modern diagnostic laboratory algorithms.

To confirm that growing bacteria in positive cultures are indeed members of the MTBC, lateral flow rapid tests are widely used around the world. These rapid tests detect the secreted MPT64 antigen common to all members of the MTBC (Bekmurzayeva et al. 2013; Jiang et al. 2014; Wang et al. 2007). However, unpublished data from The Gambia demonstrated that Maf2 isolates express the *mpt64* gene at 2.5-fold lower levels than Mtb. Indeed, during the 10 day test window, recommended by most manufacturers, in which positive MGIT tubes should be processed, the sensitivity of the test in correctly identifying Maf2 isolates as members of the MTBC ranged from only 78% to 84% for day 1 to day 10 after MGIT positivity, respectively. Although no evaluation was performed in Maf1 isolates to date, genomic analysis of the *mpt64* gene in Maf1 already demonstrated a lineage-defining nsSNP that could impair the functionality of rapid tests in the Maf1 lineage as well (Gagneux and Small 2007).

To detect whether a patient's immune system was previously exposed to MTBC bacteria, ELISPOT assays are performed. For these assays, PBMCs are stimulated with various antigens, for instance ESAT-6, and the amount of T-cells responding to ESAT-6 stimulation is assessed. However, when comparing Mtb and Maf2, it was found that the sensitivity of ESAT-6 ELISPOT assay in Maf2 infected patients' blood was significantly lower than in patients infected with Mtb (de Jong et al. 2006). However, no difference in responsiveness towards the other two ELISPOT antigens PPD and CFP-10 was observed (de Jong et al. 2006). It was concluded that Maf2 induces a somewhat attenuated anti-ESAT-6 immune response, possibly due to impaired ESAT-6 synthesis or secretion, therefore limiting the use of ESAT-6 ELISPOT tests in Maf endemic areas.

Taken together, various examples show that commonly used diagnostics might have a reduced performance when applied to Maf, due to underlying physiological differences of the bacteria. Therefore, it is strongly recommended that all novel diagnostic assays need to be re-evaluated for their performance in correctly detecting Maf isolates, including Maf1.

6.7 Biological Implications

Besides diagnostic implications, a solid understanding of the differences and similarities between Mtb *sensu stricto* and Maf lineages is essential, as numerous potential drug or vaccine targets are designed to interfere with biological

mechanisms identified in Mtb *sensu stricto* laboratory strains or clinical isolates. However, to assure success of any therapeutic intervention it is imperative that it targets mechanisms equally conserved between Mtb and Maf. Fulfilling this task becomes increasingly more complex, when considering novel research findings on the microbiology and immunology of Maf, which to date have hinted at the distinct niche adaptation by these lineages.

Previous genomic studies already suggested a different lifestyle of Maf from Mtb. For instance, besides the overall gene erosion outlined above, fundamental genetic changes occurred in genes associated with lipid cata– and metabolism in Maf. These include, amongst others, an accumulation of nsSNPs in lipid transporters of the mmpL gene family (Gehre et al. 2013b) as well as frameshift mutations in genes synthesising phenolic glycolipids (Malaga et al. 2008) and genes responsible for the synthesis of sulfolipids, DATs and PATs regulated by a mutated *phoP/R* system (Broset et al. 2015; Gonzalo-Asensio et al. 2014). Confirming this, the phenotypic lipid profile of Maf dramatically differs from Mtb. It was shown that it is essential for Mtb to adjust its lipid catabolism in response to hypoxic, intracellular survival within the human host cell, as lipids are the major intracellular carbon source in the absence of sugars (Niederweis 2008). A key regulon *dosR*/Rv0081 was recently described in Mtb to orchestrate the adaptation to hypoxic conditions experienced during anaerobic growth within the human macrophage or granuloma (Galagan et al. 2013). The fact that Maf might already have acquired fundamental genetic changes in lipid metabolism, ion transport and genes essential for intracellular survival (Gehre et al. 2013b) might indicate alternative survival strategies. Unpublished preliminary mRNA expression data from The Gambia shows that the expression of genes belonging to the *dosR* regulon is significantly downregulated in Maf isolates when compared to Mtb. Although speculation at this point, it is conceivable that Maf does not rely on rapid *dosR*/Rv0081 switching between aerobic and anaerobic growth to the same degree as Mtb. This could be a consequence of increased adaptation to a more persistent, anaerobic lifestyle and explain the fact that patients infected with Maf progress to disease significantly slower when compared to Mtb (de Jong et al. 2008). Mtb outnumbering Maf in several traditionally Maf-endemic West African countries, as mentioned earlier, may be the consequence of this slower progression of Maf. As an accelerated spread of bacteria within the human population, as observed for Mtb, requires the capacity to rapidly switch between aerobic and anaerobic growth due to faster infection-reinfection cycles between hosts, Maf might have lost its capacity for rapid spread as a sacrifice to its preferred anaerobic growth niche.

Another important virulence factor of Mtb is the early secreted antigen 6 kDa (ESAT-6), implicated in the mycobacterial translocation from the phagolysosome to the cytosol (van der Wel et al. 2007) of the macrophage during intracellular survival. The above mentioned reduced sensitivity of ESAT-6 ELISPOTs for the detection of Maf2 (de Jong et al. 2006) was a first indication of a potential lower overall ESAT-6 synthesis or secretion defect of Maf2 isolates. The exact mechanism is still elusive, and, while one publication attributes the overall reduced ESAT-6 immunogenicity of Maf2 to a relative secretion defect of ESAT-6 in Maf2 when compared to Mtb (Bold et al. 2012), other studies suggest a more complex mechanism that leads to reduced synthesis of the ESAT-6 peptide (Gonzalo-Asensio et al. 2014). For instance, it was demonstrated that ESAT-6 expression is directly linked to the expression of the *espACD* operon, which itself is regulated by the *phoP/R* regulatory system (Gonzalo-Asensio et al. 2014). Maf2 isolates possess mutations in the *phoR* gene, which, when transformed into an Mtb background, will lead to an abolishment of ESAT-6 production and a dramatically altered lipid profile. Despite the presence of this deleterious *phoR* mutation, ESAT-6 synthesis is not fully ceased in Maf2 indicating the presence of a Maf2-specific compensatory mechanism that restores ESAT-6 synthesis. Although it does not fully restore ESAT-6 secretions to the levels of wildtype Mtb, the rescue mechanism could be associated with a gain of *espACD* operon function

due to the loss of the RD8 region in Maf2 isolates (Gonzalo-Asensio et al. 2014).

As shown, despite the fact that both Maf and Mtb cause TB, they significantly differ, amongst many others, in two crucial virulence mechanisms: the *dosR*/Rv0081 regulon and *phoP/R* and ESAT-6 regulation, rendering these pathways less suitable drug or vaccine targets. Therefore, identifying common biological mechanisms between Mtb and Maf isolates is essential for drug and vaccine design. Failing to do so risks selecting for Maf isolates in West Africa by the implementation of drugs/vaccines specific only to Mtb targets, yet the outlined microbiological differences might already reveal potential for vaccine escape variants that could emerge in Mtb isolates in the future, similar to their natural evolution in Maf.

6.8 Clinical Differences Between Maf and Mtb

Unlike its close relative *M. bovis*, Maf is sensitive to pyrazinamide. As treatment outcomes on regular Category 1 therapy (Cat 1; 2 months of isoniazid, rifampicin, ethambutol and pyrazinamide, followed by 4 months of isoniazid and rifampicin), do not differ between Maf and Mtb patients (de Jong et al. 2007), there is no clinical urgency in diagnosing TB infection as due to Maf vs Mtb. Several studies found lower rates of drug resistance in Maf than in Mtb isolates, and/or an underrepresentation of Maf among retreatment patients, suggesting that Maf may be *more* effectively treated by Cat 1. Given the complexities of clinical trials for TB, specific regimen shortening studies for Maf infected patients however seem not warranted.

The underrepresentation of Maf among retreatment cases could also be linked to its lower progression to disease, observed in contacts exposed to Maf2 infected index cases. While the skin test conversion rate was the same as in contacts exposed to Mtb, the contacts exposed to Maf2 were significantly less likely to develop TB during the 2-year follow-up period relative to Mtb exposed contacts, suggesting that transmission rates are similar, yet the rate of progression to active disease is lower in Maf2 (de Jong et al. 2008).

Studies from Ghana (unpublished data) and Gambia (de Jong et al. 2005) have identified an association of Maf2 with HIV infection. In Ghana, Maf1 was *not* associated with HIV infection (unpublished data) and an older study from Ghana, in which Maf1 and Maf2 were analysed together, also did not identify such an association (Meyer et al. 2008). In Gambia, Maf2 is also associated with patients with lower BMIs than Mtb. It is however unclear whether Maf2 preferentially causes disease in underweight persons, or causes more immune stimulation of e.g. TNFa, associated with increased weight loss, or causes more indolent disease for which patients present later, after more weight loss has taken place.

While the frequency of extrapulmonary disease has not been systematically studied in Africa, where diagnostic testing facilities for non-pulmonary TB disease are limited, a recent study from the US, in which MAF1 and MAF2 were analysed together, found Maf to be more common in extrapulmonary disease (Sharma et al. 2016). Whether this is related to the microaerophilic adaptation of Maf is unclear. The extrapulmonary predilection of *M. bovis*, ascribed to the additional gastro-intestinal route of transmission, is based on similar mechanisms as Maf, also remains to be investigated.

6.9 Future Research Questions

As outlined above, understanding fundamental biological processes in any of the two Maf lineages will inform research about pathogenicity, virulence and drug/vaccine design of the MTBC. Although research on Maf2 is slowly increasing, major efforts need to be made in investigating Maf1 biology. Knowledge about both lineages will allow optimisation of diagnostic algorithms and targeted intervention strategies.

Identifying the various differences of the molecular interactions between Maf and the human host, might give further insights into overall virulence of MTBC bacteria. Similarly, studying

the regulatory strategy of Maf2 to compensate for a reduced activity of the *dosR*/Rv0081 regulon will reveal novel mycobacterial strategies to adjust to microaerophilic growth, persistence and latency. On an epidemiological level, insights into the causes and consequences of the geographical restriction of Maf isolates to certain regions within West Africa will reveal important clues on sub-regional or even intercontinental TB transmission mechanisms. Moreover, the predilection for a weakened host (especially for Maf2, Lineage 6) suggests a relatively larger reservoir of latent infection relative to that of 'modern' Mtb. As mycobacterial DNA to date has not been isolated from persons with latent TB infection, and antigens specific for Maf2 relative to Mtb proved insufficiently sensitive nor specific (Hill et al. 2005), the only imperfect proxy for latent infection available to date is whether a person was exposed to an index case with Maf vs Mtb. However, novel phylogenetic modelling techniques may be able to estimate the total mycobacterial population size based on the genomes obtained from patients with active disease, and thus test the size of the relative latent reservoirs (see Chap. 15). Taken together, carefully comparing these ancient West African MTBC lineages with modern Mtb lineages will not only result in optimised vaccine and treatment strategies but also in the design of novel transmission blocking interventions.

References

Addo K, Owusu-Darko K, Yeboah-Manu D, Caulley P, Minamikawa M, Bonsu F, Leinhardt C, Akpedonu P, Ofori-Adjei D (2007) Mycobacterial species causing pulmonary tuberculosis at the Korle-Bu Teaching Hospital, Accra, Ghana. Ghana Med J 41(2):52–57

Affolabi D, Anyo G, Faihun F, Sanoussi N, Shamputa IC, Rigouts L, Kestens L, Anagonou S, Portaels F (2009) First molecular epidemiological study of tuberculosis in Benin. Int J Tuberc Lung Dis 13(3): 317–322

Alexander KA, Laver PN, Michel AL, Williams M, van Helden PD, Warren RM, Gey van Pittius NC (2010) Novel Mycobacterium tuberculosis complex pathogen, M. mungi. Emerg Infect Dis 16(8):1296–1299. doi:10.3201/eid1608.100314

Alfredsen S, Saxegaard F (1992) An outbreak of tuberculosis in pigs and cattle caused by Mycobacterium africanum. Vet Rec 131(3):51–53

Aliyu G, El-Kamary SS, Abimiku A, Brown C, Tracy K, Hungerford L, Blattner W (2013) Prevalence of non-tuberculous mycobacterial infections among tuberculosis suspects in Nigeria. PLoS One 8(5):e63170. doi:10.1371/journal.pone.0063170PONE-D-13-06198. [pii]

Asante-Poku A, Aning KG, Boi-Kikimoto B, Yeboah-Manu D (2014) Prevalence of bovine tuberculosis in a dairy cattle farm and a research farm in Ghana. Onderstepoort J Vet Res 81(2):E1–E6

Asante-Poku A, Yeboah-Manu D, Otchere ID, Aboagye SY, Stucki D, Hattendorf J, Borrell S, Feldmann J, Danso E, Gagneux S (2015) Mycobacterium africanum is associated with patient ethnicity in Ghana. PLoS Negl Trop Dis 9(1):e3370. doi:10.1371/journal.pntd.0003370PNTD-D-14-01295. [pii]

Asante-Poku A, Otchere ID, Osei-Wusu S, Sarpong E, Baddoo A, Forson A, Laryea C, Borrell S, Bonsu F, Hattendorf J, Ahorlu C, Koram KA, Gagneux S, Yeboah-Manu D (2016) Molecular epidemiology of Mycobacterium africanum in Ghana. BMC Infect Dis 16(1):385. doi:10.1186/s12879-016-1725-6

Asmar S, Chatellier S, Mirande C, van Belkum A, Canard I, Raoult D, Drancourt M (2015) A novel solid medium for culturing Mycobacterium tuberculosis Isolates from clinical specimens. J Clin Microbiol 53(8):2566–2569. JCM.01149-15 [pii]. doi:10.1128/JCM.01149-15

Assam JPA, Beng Vé P, Cho-Ngwa F, Toukam M, Ngoh AAI, Kitavi M, Nzuki I, Nyonka JN, Tata E, Tedom JC, Skilton RA, Pelle R, Titanji VPK (2013) Mycobacterium tuberculosis is the causative agent of tuberculosis in the southern ecological zones of Cameroon, as shown by genetic analysis. BMC Infect Dis 13:431

Baess I (1979) Deoxyribonucleic acid relatedness among species of slowly-growing mycobacteria. Acta Pathol Microbiol Scand B 87(4):221–226

Bedrossian N, Hamze M, Rahmo AK, Jurjus A, Saliba J, Dabboussi F, Karam W (2013) Mycobacterium tuberculosis spoligotypes circulating in the Lebanese population: a retrospective study. East Mediterr Health J 19(2):119–124

Bek D, Kjeldsen MK, Kamper-Jorgensen Z, Hansen NF, Rasmussen EM (2010) Tuberculosis caused by Mycobacterium africanum. Ugeskr Laeger 172(7):549–550. VP12080608 [pii]

Bekmurzayeva A, Sypabekova M, Kanayeva D (2013) Tuberculosis diagnosis using immunodominant, secreted antigens of Mycobacterium tuberculosis. Tuberculosis (Edinb) 93(4):381–388. S1472-9792(13)00082-6 [pii]. doi:10.1016/j.tube.2013.03.003

Bentley SD, Comas I, Bryant JM, Walker D, Smith NH, Harris SR, Thurston S, Gagneux S, Wood J, Antonio M, Quail MA, Gehre F, Adegbola RA, Parkhill J, de Jong BC (2012) The genome of Mycobacterium africanum West African 2 reveals a lineage-specific locus and genome erosion common to the M. tuberculosis complex.

PLoS Negl Trop Dis 6(2):e1552. doi:10.1371/journal.pntd.0001552PNTD-D-11-01108. [pii]

Blouin Y, Hauck Y, Soler C, Fabre M, Vong R, Dehan C, Cazajous G, Massoure PL, Kraemer P, Jenkins A, Garnotel E, Pourcel C, Vergnaud G (2012) Significance of the identification in the Horn of Africa of an exceptionally deep branching Mycobacterium tuberculosis clade. PLoS One 7(12):e52841. doi:10.1371/journal.pone.0052841PONE-D-12-22563. [pii]

Bold TD, Davis DC, Penberthy KK, Cox LM, Ernst JD, de Jong BC (2012) Impaired fitness of Mycobacterium africanum despite secretion of ESAT-6. J Infect Dis 205(6):984–990. jir883 [pii]. doi:10.1093/infdis/jir883

Bradley SG (1972) Reassociation of deoxyribonucleic acid from selected mycobacteria with that from Mycobacterium bovis and Mycobacterium farcinica. Am Rev Respir Dis 106:122–124

Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, Garnier T, Gutierrez C, Hewinson G, Kremer K, Parsons LM, Pym AS, Samper S, van Soolingen D, Cole ST (2002) A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci U S A 99:3684–3689. doi:10.1073/pnas.052548299

Broset E, Martin C, Gonzalo-Asensio J (2015) Evolutionary landscape of the Mycobacterium tuberculosis complex from the viewpoint of PhoPR: implications for virulence regulation and application to vaccine development. MBio 6(5):e01289–e01215. mBio.01289-15 [pii]. doi:10.1128/mBio.01289-15

Brown T, Nikolayevskyy V, Velji P, Drobniewski F (2010) Associations between Mycobacterium tuberculosis strains and phenotypes. Emerg Infect Dis 16(2):272–280. doi:10.3201/eid1602.091032

Cadmus S, Palmer S, Okker M, Dale J, Gover K, Smith N, Jahans K, Hewinson RG, Gordon SV (2006) Molecular analysis of human and bovine tubercle bacilli from a local setting in Nigeria. J Clin Microbiol 44(1):29–34. 44/1/29 [pii]. doi:10.1128/JCM.44.1.29-34.2006

Cadmus SI, Jenkins AO, Godfroid J, Osinusi K, Adewole IF, Murphy RL, Taiwo BO (2009) Mycobacterium tuberculosis and Mycobacterium africanum in stools from children attending an immunization clinic in Ibadan, Nigeria. Int J Infect Dis 13(6):740–744. S1201-9712(09)00006-X [pii]. doi:10.1016/j.ijid.2008.11.016

Castets M, Boisvert H, Grumbach F, Brunel M, Rist N (1968) Tuberculosis bacilli of the African type: preliminary note. Rev Tuberc Pneumol (Paris) 32(2):179–184

Chavadi S, Wooff E, Coldham NG, Sritharan M, Hewinson RG, Gordon SV, Wheeler PR (2009) Global effects of inactivation of the pyruvate kinase gene in the Mycobacterium tuberculosis complex. J Bacteriol 191(24):7545–7553

Collins CH, Yates MD, Grange JM (1982) Subdivision of Mycobacterium tuberculosis into five variants for epidemiological purposes: methods and nomenclature. J Hyg (Lond) 89(2):235–242

Coscolla M, Lewin A, Metzger S, Maetz-Rennsing K, Calvignac-Spencer S, Nitsche A, Dabrowski PW, Radonic A, Niemann S, Parkhill J, Couacy-Hymann E, Feldman J, Comas I, Boesch C, Gagneux S, Leendertz FH (2013) Novel Mycobacterium tuberculosis complex isolate from a wild chimpanzee. Emerg Infect Dis 19(6):969–976. doi:10.3201/eid1906.121012

Cousins D, Peet R, Gaynor W, Williams S, Gow B (1994) Tuberculosis in imported hyrax (Procavia capensis) caused by an unusual variant belonging to the Mycobacterium tuberculosis complex. Vet Microbiol 42:135. doi:10.1016/0378-1135(94)90013-2

David HL, Jahan M, Jumin A, Grandry J, Lehman EH (1978) Numerical taxonomy analysis of Mycobacterium africanum. Int J Syst Bacteriol 28(4): 467–472

de Jong BC, Hill PC, Brookes RH, Otu JK, Peterson KL, Small PM, Adegbola RA (2005) Mycobacterium africanum: a new opportunistic pathogen in HIV infection? AIDS 19(15):1714–1715

de Jong BC, Hill PC, Brookes RH, Gagneux S, Jeffries DJ, Otu JK, Donkor SA, Fox A, McAdam KP, Small PM, Adegbola RA (2006) Mycobacterium africanum elicits an attenuated T cell response to early secreted antigenic target, 6 kDa, in patients with tuberculosis and their household contacts. J Infect Dis 193(9):1279–1286. JID35711 [pii]. doi:10.1086/502977

de Jong B, Hill P, Aiken A, Jeffries D, Onipede A, Small P, Adegbola R, Corrah T (2007) Clinical presentation and outcome of tuberculosis patients infected by M. africanum versus M. tuberculosis. Int J Tuberc Lung Dis 11:450

de Jong BC, Hill PC, Aiken A, Awine T, Antonio M, Adetifa IM, Jackson-Sillah DJ, Fox A, Deriemer K, Gagneux S, Borgdorff MW, McAdam KP, Corrah T, Small PM, Adegbola RA (2008) Progression to active tuberculosis, but not transmission, varies by Mycobacterium tuberculosis lineage in The Gambia. J Infect Dis 198(7):1037–1043. doi:10.1086/591504 [doi]

de Jong B, Antonio M, Awine T, Ogungbemi K, de Jong Y, Gagneux S, De Riemer K, Zozio T, Rastogi N, Borgdorff M, Hill P, Adegbola R (2009) Use of spoligotyping and large sequence polymorphisms to study the population structure of the Mycobacterium tuberculosis complex in a cohort study of consecutive smear-positive tuberculosis cases in The Gambia. J Clin Microbiol 47:994. doi:10.1128/jcm.01216-08

de Jong BC, Antonio M, Gagneux S (2010) Mycobacterium africanum—review of an important cause of human tuberculosis in West Africa. PLoS Negl Trop Dis 4(9):e744

Desmond E, Ahmed AT, Probert WS, Ely J, Jang Y, Sanders CA, Lin SY, Flood J (2004) Mycobacterium africanum cases, California. Emerg Infect Dis 10(5):921–923. doi:10.3201/eid1005.030016

Diop S, de Medeiros D, de Medeiros G, Baylet R, Sankale M (1976) Incidence and geographic distribution of Mycobacterium africanum in Senegal. Bull Soc Med Afr Noire Lang Fr 21(1):50–56

Dosso M, Bonard D, Msellati P, Bamba A, Doulhourou C, Vincent V, Peyre M, Traore M, Koffi K, Coulibaly IM (1999) Primary resistance to antituberculosis drugs: a national survey conducted in Cote d'Ivoire in 1995–

1996. Ivoirian Study Group on Tuberculosis Resistance. Int J Tuberc Lung Dis 3(9):805–809

Frothingham R, Strickland PL, Bretzel G, Ramaswamy S, Musser JM, Williams DL (1999) Phenotypic and genotypic characterization of Mycobacterium africanum isolates from West Africa. J Clin Microbiol 37(6):1921–1926

Frottier J, Eliaszewicz M, Arlet V, Gaudillat C (1990) Infections caused by Mycobacterium africanum. Bull Acad Natl Med 174(1):29–33. discussion 34–25

Gagneux S, Small PM (2007) Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development. Lancet Infect Dis 7(5):328–337. S1473-3099(07)70108-1 [pii]. doi:10.1016/S1473-3099(07)70108-1

Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, Nicol M, Niemann S, Kremer K, Gutierrez MC, Hilty M, Hopewell PC, Small PM (2006) Variable host-pathogen compatibility in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 103(8):2869–2873. 0511240103 [pii]. doi:10.1073/pnas.0511240103

Galagan JE (2014) Genomic insights into tuberculosis. Nat Rev Genet 15(5):307–320. nrg3664 [pii]. doi:10.1038/nrg3664 [doi]

Galagan JE, Minch K, Peterson M, Lyubetskaya A, Azizi E, Sweet L, Gomes A, Rustad T, Dolganov G, Glotova I, Abeel T, Mahwinney C, Kennedy AD, Allard R, Brabant W, Krueger A, Jaini S, Honda B, Yu WH, Hickey MJ, Zucker J, Garay C, Weiner B, Sisk P, Stolte C, Winkler JK, Van de Peer Y, Iazzetti P, Camacho D, Dreyfuss J, Liu Y, Dorhoi A, Mollenkopf HJ, Drogaris P, Lamontagne J, Zhou Y, Piquenot J, Park ST, Raman S, Kaufmann SH, Mohney RP, Chelsky D, Moody DB, Sherman DR, Schoolnik GK (2013) The Mycobacterium tuberculosis regulatory network and hypoxia. Nature 499(7457):178–183. nature12337 [pii]. doi:10.1038/nature12337

Garzelli C, Lari N, Cuccu B, Tortoli E, Rindi L (2010) Impact of immigration on tuberculosis in a low-incidence area of Italy: a molecular epidemiological approach. Clin Microbiol Infect 16(11):1691–1697. S1198-743X(14)60567-X [pii]. doi:10.1111/j.1469-0691.2009.03149.x

Gehre F, Antonio M, Faihun F, Odoun M, Uwizeye C, de Rijk P, de Jong BC, Affolabi D (2013a) The first phylogeographic population structure and analysis of transmission dynamics of M. africanum West African 1 – combining molecular data from Benin, Nigeria and Sierra Leone. PLoS One 8(10):e77000. doi:10.1371/journal.pone.0077000PONE-D-13-26514. [pii]

Gehre F, Otu J, DeRiemer K, de Sessions PF, Hibberd ML, Mulders W, Corrah T, de Jong BC, Antonio M (2013b) Deciphering the growth behaviour of Mycobacterium africanum. PLoS Negl Trop Dis 7(5):e2220. doi:10.1371/journal.pntd.0002220PNTD-D-12-01634. [pii]

Godreuil S, Torrea G, Terru D, Chevenet F, Diagbouga S, Supply P, Van de Perre P, Carriere C, Banuls AL (2007) First molecular epidemiology study of Mycobacterium tuberculosis in Burkina Faso. J Clin Microbiol 45(3):921–927. JCM.01918-06 [pii]. doi:10.1128/JCM.01918-06

Gomgnimbou MK, Refregier G, Diagbouga SP, Adama S, Kabore A, Ouiminga A, Sola C (2012) Spoligotyping of Mycobacterium africanum, Burkina Faso. Emerg Infect Dis 18(1):117–119. doi:10.3201/eid1801.110275

Gonzalo-Asensio J, Malaga W, Pawlik A, Astarie-Dequeker C, Passemar C, Moreau F, Laval F, Daffe M, Martin C, Brosch R, Guilhot C (2014) Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence regulator. Proc Natl Acad Sci U S A 111(31):11491–11496. 1406693111 [pii]. doi:10.1073/pnas.1406693111

Goyal M, Lawn S, Afful B, Acheampong JW, Griffin G, Shaw R (1999) Spoligotyping in molecular epidemiology of tuberculosis in Ghana. J Infect 38:171–175

Grange JM (1979) Tuberculosis: the changing tubercle. Br J Hosp Med 22(6):540–548

Grange JM, Yates MD (1989) Incidence and nature of human tuberculosis due to Mycobacterium africanum in South-East England: 1977–87. Epidemiol Infect 103(1):127–132

Groenheit R, Ghebremichael S, Svensson J, Rabna P, Colombatti R, Riccardi F, Couvin D, Hill V, Rastogi N, Koivula T, Kallenius G (2011) The Guinea-Bissau family of Mycobacterium tuberculosis complex revisited. PLoS One 6(4):e18601. doi:10.1371/journal.pone.0018601

Grosset J, Sangare S, Rist N, Meyer L (1974) Cultural and biochemical characteristics of tubercle bacilli isolated from 230 cases of tuberculosis in Mali. Bull Int Union Tuberc 49:177–187

Gudan A, Artukovic B, Cvetnic Z, Spicic S, Beck A, Hohsteter M, Naglic T, Bata I, Grabarevic Z (2008) Disseminated tuberculosis in hyrax (Procavia capensis) caused by Mycobacterium africanum. J Zoo Wildl Med 39(3):386–391. doi:10.1638/06-041.1

Haas WH, Bretzel G, Amthor B, Schilke K, Krommes G, Rusch-Gerdes S, Sticht-Groh V, Bremer HJ (1997) Comparison of DNA fingerprint patterns of isolates of Mycobacterium africanum from east and west Africa. J Clin Microbiol 35(3):663–666

Hart PD, Sutherland I (1977) BCG and vole bacillus vaccines in the prevention of tuberculosis in adolescence and early adult life. Br Med J 2(6082):293–295

Hermans-Boveroulle MT, Pattyn SR, Gatti F (1965) Study of human strains of Mycobacterium tuberculosis Isolated in Leopoldville. Ann Soc Belge de Med Trop 45(5):531–540

Hill PC, Jackson-Sillah D, Fox A, Franken KL, Lugos MD, Jeffries DJ, Donkor SA, Hammond AS, Adegbola RA, Ottenhoff TH, Klein MR, Brookes RH (2005) ESAT-6/CFP-10 fusion protein and peptides for optimal diagnosis of mycobacterium tuberculosis infection by ex vivo enzyme-linked immunospot assay in the Gambia. J Clin Microbiol 43(5):2070–2074. 43/5/2070 [pii]. doi:10.1128/JCM.43.5.2070-2074.2005

Hoffner SE, Svenson SB, Norberg R, Dias F, Ghebremichael S, Kallenius G (1993) Biochemical heterogeneity of Mycobacterium tuberculosis complex isolates in Guinea-Bissau. J Clin Microbiol 31(8):2215–2217

Homolka S, Post E, Oberhauser B, George A, Westman L, Dafae F, Rüsch-Gerdes S, Niemann S (2008) High genetic diversity among Mycobacterium tuberculosis complex strains from Sierra Leone. BMC Microbiol 8:103. doi:10.1186/1471-2180-8-103

Homolka S, Meyer CG, Hillemann D, Owusu-Dabo E, Adjei O, Horstmann RD, Browne EN, Chinbuah A, Osei I, Gyapong J, Kubica T, Ruesch-Gerdes S, Niemann S (2010) Unequal distribution of resistance-conferring mutations among Mycobacterium tuberculosis and Mycobacterium africanum strains from Ghana. Int J Med Microbiol 300(7):489–495. S1438-4221(10)00050-0 [pii]. doi:10.1016/j.ijmm.2010.04.019 [doi]

Huet M, Rist N, Boube G, Potier D (1971) Bacteriological study of tuberculosis in Cameroon. Rev Tuberc Pneumol (Paris) 35(4):413–426

Hurtado UA, Solano JS, Rodriguez A, Robledo J, Rouzaud F (2016) Draft genome sequence of a Mycobacterium africanum clinical isolate from Antioquia, Colombia. Genome Announc 4(3):e00486. 4/3/e00486-16 [pii]. doi:10.1128/genomeA.00486-16

Isea-Pena MC, Brezmes-Valdivieso MF, Gonzalez-Velasco MC, Lezcano-Carrera MA, Lopez-Urrutia-Lorente L, Martin-Casabona N, Monforte-Cirac ML, Palacios JJ, Penedo-Pallares A, Ramirez-Rosales A, Sanchez-Silos R, Tortola-Fernandez T, Vinuelas-Bayon J, Vitoria-Agreda A, Esteban J (2012) Mycobacterium africanum, an emerging disease in high-income countries? Int J Tuberc Lung Dis 16(10):1400–1404. doi:10.5588/ijtld.12.0142

Jaspe RC, Rojas YM, Flores LA, Sofia Toro E, Takiff H, de Waard JH (2009) Evaluation of the Kudoh swab method for the culturing of Mycobacterium tuberculosis in rural areas. Tropical Med Int Health 14(4):468–471. TMI2236 [pii]. doi:10.1111/j.1365-3156.2009.02236.x

Jiang Y, Liu H, Wan K (2014) MPT64 polymorphisms of Mycobacterium tuberculosis strains suggest ongoing immune evasion. Tuberculosis (Edinb) 94(6):712–714. S1472-9792(14)20573-7 [pii]. doi:10.1016/j.tube.2014.08.013

Jungbluth H, Fink H, Reusch F (1978) Tuberculous infection caused by Myco. africanum in black africans resident in the German Federal Republic (author's transl). Prax Klin Pneumol 32(5):306–309

Kallenius G, Koivula T, Ghebremichael S, Hoffner SE, Norberg R, Svensson E, Dias F, Marklund BI, Svenson SB (1999) Evolution and clonal traits of Mycobacterium tuberculosis complex in Guinea-Bissau. J Clin Microbiol 37(12):3872–3878

Kato-Maeda M, Bifani PJ, Kreiswirth BN, Small PM (2001) The nature and consequence of genetic variability within Mycobacterium tuberculosis. J Clin Invest 107(5):533–537. doi:10.1172/JCI11426

Keating LA, Wheeler PR, Mansoor H, Inwald JK, Dale J, Hewinson RG, Gordon SV (2005) The pyruvate requirement of some members of the Mycobacterium tuberculosis complex is due to an inactive pyruvate kinase: implications for in vivo growth. Mol Microbiol 56(1):163–174. MMI4524 [pii]. doi:10.1111/j.1365-2958.2005.04524.x

Koro FK, Simo YK, Piam FF, Noeske J, Gutierrez C, Kuaban C, Eyangoh SI (2013) Population dynamics of tuberculous bacilli in Cameroon as assessed by spoligotyping. J Clin Microbiol 51(1):299–302

Kothadia SN, Sengupta SR (1982) Evaluation of swab culture method of Kudoh and Kudoh for cultivation of M. tuberculosis. Indian J Med Sci 36(1):9–11

Kuaban C, Um Boock A, Noeske J, Bekang F, Eyangoh S (2014) Mycobacterium tuberculosis complex strains and drug susceptibility in a cattle-rearing region of Cameroon. Int J Tuberc Lung Dis 18(1):34–38. doi:10.5588/ijtld.13.0333

Lawn SD, Frimpong EH, Al-Ghusein H, Acheampong JW, Uttley AH, Butcher PD, Griffin GE (2001) Pulmonary tuberculosis in Kumasi, Ghana: presentation, drug resistance, molecular epidemiology and outcome of treatment. West Afr J Med 20(2):92–97

Lawson L, Zhang J, Gomgnimbou MK, Abdurrahman ST, Le Moullec S, Mohamed F, Uzoewulu GN, Sogaolu OM, Goh KS, Emenyonu N, Refrégier G, Cuevas LE, Sola C (2012) A molecular epidemiological and genetic diversity study of tuberculosis in Ibadan, Nnewi and Abuja, Nigeria. PLoS One 7(6):e38409

Leao SC, Martin A, Mejia GI, Palomino JC, Robledo J, Telles MAS, Portaels F (2004) Practical handbook for the phenotypic and genotypic identification of mycobacteria. Vanden BROELLE, Brugges

Ledru S, Cauchoix B, Yameogo M, Zoubga A, Lamande-Chiron J, Portaels F, Chiron JP (1996) Impact of short-course therapy on tuberculosis drug resistance in South-West Burkina Faso. Tuber Lung Dis 77(5):429–436. S0962-8479(96)90116-1 [pii]

Malaga W, Constant P, Euphrasie D, Cataldi A, Daffe M, Reyrat JM, Guilhot C (2008) Deciphering the genetic bases of the structural diversity of phenolic glycolipids in strains of the Mycobacterium tuberculosis complex. J Biol Chem 283(22):15177–15184. M710275200 [pii]. doi:10.1074/jbc.M710275200

Meissner G, Schroder K (1969) The so-called African mycobacteria strains from the tropical West Africa. Zentralbl Bakteriol Orig 211(1):69–81

Meyer CG, Scarisbrick G, Niemann S, Browne EN, Chinbuah MA, Gyapong J, Osei I, Owusu-Dabo E, Kubica T, Rusch-Gerdes S, Thye T, Horstmann RD (2008) Pulmonary tuberculosis: virulence of Mycobacterium africanum and relevance in HIV co-infection. Tuberculosis (Edinb) 88(5):482–489. S1472-9792(08)00056-5 [pii]. doi:10.1016/j.tube.2008.05.004

Mostowy S, Onipede A, Gagneux S, Niemann S, Kremer K, Desmond EP, Kato-Maeda M, Behr M (2004) Genomic analysis distinguishes Mycobacterium africanum. J Clin Microbiol 42(8):3594–3599.

doi:10.1128/JCM.42.8.3594-3599.200442/8/3594. [pii]
Munoz ME, Ponce E (2003) Pyruvate kinase: current status of regulatory and functional properties. Comp Biochem Physiol B Biochem Mol Biol 135: 197–218
Niederweis M (2008) Nutrient acquisition by mycobacteria. Microbiology 154(Pt 3):679–692. 154/3/679 [pii]. doi:10.1099/mic.0.2007/012872-0
Niemann S, Richter E, Rusch-Gerdes S (2000) Differentiation among members of the Mycobacterium tuberculosis complex by molecular and biochemical features: evidence for two pyrazinamide-susceptible subtypes of M. bovis. J Clin Microbiol 38(1):152–157
Niemann S, Rüsch-Gerdes S, Joloba ML, Whalen CC, Guwatudde D, Ellner JJ, Eisenach K, Fumokong N, Johnson JL, Aisu T, Mugerwa RD, Okwera A, Schwander SK (2002) Mycobacterium africanum subtype II is associated with two distinct genotypes and is a major cause of human tuberculosis in Kampala, Uganda. J Clin Microbiol 40(9):3398–3405
Niobe-Eyangoh SN, Kuaban C, Sorlin P, Cunin P, Thonnon J, Sola C, Rastogi N, Vincent V, Gutierrez MC (2003) Genetic biodiversity of Mycobacterium tuberculosis complex strains from patients with pulmonary tuberculosis in Cameroon. J Clin Microbiol 41(6):2547–2553
Nuru A, Mamo G, Worku A, Admasu A, Medhin G, Pieper R, Ameni G (2015) Genetic diversity of Mycobacterium tuberculosis complex isolated from tuberculosis patients in Bahir Dar City and its surroundings, Northwest Ethiopia. Biomed Res Int 2015:174732. doi:10.1155/2015/174732
Ontsira Ngoyi EN, Obengui, Taty Taty R, Koumba EL, Ngala P, Ossibi Ibara RB (2014) Mycobacterial species repartition: experience of the Anti-tuberculosis Center in Pointe Noire (Republic of Congo). Bull Soc Pathol Exot 107(5):342–345. doi:10.1007/s13149-014-0395-4 [doi]
Palaci M, Peres RL, Maia R, Cunha EA, Ribeiro MO, Lecco R, de Souza RC, Ferro ESRR, Vinhas SA, Dietze R, Vianna S, de Morais CG (2013) Contribution of the Ogawa-Kudoh swab culture method to the diagnosis of pulmonary tuberculosis in Brazil. Int J Tuberc Lung Dis 17(6):782–786. doi:10.5588/ijtld.12.0500
Parsons L, Brosch R, Cole S, Somoskövi A, Loder A, Bretzel G, Van Soolingen D, Hale Y, Salfinger M (2002) Rapid and simple approach for identification of Mycobacterium tuberculosis complex isolates by PCR-based genomic deletion analysis. J Clin Microbiol 40:2339. doi:10.1128/jcm.40.7.2339-2345.2002
Pattyn SR, Portaels F, Spanoghe L, Magos J (1970) Further studies on African strains of Mycobacterium tuberculosis: comparison with M. bovis and M. microti. Ann Soc Belges Med Trop Parasitol Mycol 50(2):211–227
Perez-de Pedro I, Bermudez P, Artero I, Jimenez MS (2008) Orchiepididymitis due to Mycobacterium africanum. Enferm Infecc Microbiol Clin 26(9):600–602
Prat R, Rist N, Dumitrescu N, Mugabushaka A, Clavel S, Duponchel C (1974) Special characteristics of the cultures of tubercle bacilli isolated in Ruanda. Bull Int Union Tuberc 49:53–62
Rahim Z, Mollers M, te Koppele-Vije A, de Beer J, Zaman K, Matin MA, Kamal M, Raquib R, van Soolingen D, Baqi MA, Heilmann FG, van der Zanden AG (2007) Characterization of Mycobacterium africanum subtype I among cows in a dairy farm in Bangladesh using spoligotyping. Southeast Asian J Trop Med Public Health 38(4):706–713
Sangare L, Diande S, Kouanda S, Dingtoumda BI, Mourfou A, Ouedraogo F, Sawadogo I, Nebie B, Gueye A, Sawadogo LT, Traore AS (2010) Mycobacterium tuberculosis drug-resistance in previously treated patients in Ouagadougou, Burkina Faso. Ann Afr Med 9(1):15–19. AnnAfrMed_2010_9_1_15_62619 [pii]. doi:10.4103/1596-3519.62619
Schroder KH (1982) Occurence of M. africanum in the Federal Republic of Germany (author's transl). Zentralbl Bakteriol Mikrobiol Hyg A 251(3):341–344
Sharma A, Bloss E, Heilig CM, Click ES (2016) Tuberculosis caused by Mycobacterium africanum, United States, 2004–2013. Emerg Infect Dis 22(3):396–403. doi:10.3201/eid2203.151505
Simonet F, Menard M, Le Mao G, Albert JP (1989) Action antituberculeuse au Burkina Faso L'experience du service de Pneumo-phtisiologie de l'Hospital National Sanon Souro de Bobo-Dioulaso. Societe' Sante' Industrie Servoice, Paris
Smith NH, Hewinson RG, Kremer K, Brosch R, Gordon SV (2009) Myths and misconceptions: the origin and evolution of Mycobacterium tuberculosis. Nat Rev Micro 7(7):537–544
Sola C, Rastogi N, Gutierrez MC, Vincent V, Brosch R, Parsons L (2003) Is Mycobacterium africanum subtype II (Uganda I and Uganda II) a genetically well-defined subspecies of the Mycobacterium tuberculosis complex? J Clin Microbiol 41(3):1345–1346. author reply 1346–1348
Stanford JL, Grange JM (1974) The meaning and structure of species as applied to mycobacteria. Tubercle 55:143–152
Stermann M, Bohrssen A, Diephaus C, Maass S, Bange FC (2003) Polymorphic nucleotide within the promoter of nitrate reductase (NarGHJI) is specific for Mycobacterium tuberculosis. J Clin Microbiol 41(7): 3252–3259
Sticht-Groh V, Bretzel G, Rüsch-Gerdes S, Bwire S, Kawuma HJS (1994) M. africanum strains isolated in East-Africa, Uganda. Tubercle and Lung Dis 75:46. doi:10.1016/0962-8479(94)90853-2
Stonebrink B (1958) The use of a pyruvate containing egg medium in the culture of isoniazid resistant strains of Mycobacterium tuberculosis var. hominis. Acta Tuberc Scandia 35:67–80
Streicher EM, Victor TC, van der Spuy G, Sola C, Rastogi N, van Helden PD, Warren RM (2007) Spoligotype signatures in the Mycobacterium tuberculosis complex.

J Clin Microbiol 45(1):237–240. JCM.01429-06 [pii]. doi:10.1128/JCM.01429-06

Stucki D, Malla B, Hostettler S, Huna T, Feldmann J, Yeboah-Manu D, Borrell S, Fenner L, Comas I, Coscolla M, Gagneux S (2012) Two new rapid SNP-typing methods for classifying Mycobacterium tuberculosis complex into the main phylogenetic lineages. PLoS One 7(7):e41253. doi:10.1371/journal.pone.0041253PONE-D-12-11965. [pii]

Thorel MF (1980a) Isolation of Mycobacterium africanum from monkeys. Tubercle 61(2):101–104

Thorel MF (1980b) Mycobacteria identified in a centre for veterinary research between 1973 and 1979 (author's transl). Ann Microbiol (Paris) 131(1):61–69

Thumamo BP, Asuquo AE, Abia-Bassey LN, Lawson L, Hill V, Zozio T, Emenyonu N, Eko FO, Rastogi N (2012) Molecular epidemiology and genetic diversity of Mycobacterium tuberculosis complex in the Cross River State, Nigeria. Infect Genet Evol 12(4):671–677. S1567-1348(11)00290-5 [pii]. doi:10.1016/j.meegid.2011.08.011

Tsolaki AG, Hirsh AE, DeRiemer K, Enciso JA, Wong MZ, Hannan M, Goguet de la Salmoniere YO, Aman K, Kato-Maeda M, Small PM (2004) Functional and evolutionary genomics of Mycobacterium tuberculosis: insights from genomic deletions in 100 strains. Proc Natl Acad Sci U S A 101(14):4865–4870. doi:10.1073/pnas.0305634101 0305634101. [pii]

Tsukamura M (1967) Identification of mycobacteria. Tubercle 48(4):311–338

Tsukamura M (1976) Numerical classification of slowly growing mycobacteria. Int J Syst Bacteriol 26:409–420

Ueyama M, Chikamatsu K, Aono A, Murase Y, Kuse N, Morimoto K, Okumura M, Yoshiyama T, Ogata H, Yoshimori K, Kudoh S, Azuma A, Gemma A, Mitarai S (2014) Sub-speciation of Mycobacterium tuberculosis complex from tuberculosis patients in Japan. Tuberculosis (Edinb) 94(1):15–19. S1472-9792(13)00166-2 [pii]. doi:10.1016/j.tube.2013.09.006

Unger F, Münstermann S, Goumou A, Apia CN, Konte M (2003) Risk associated with Mycobacterium bovis infections detected in selected study herds and slaughter cattle in 4 countries of West Africa. Animal Health Working Paper 1. ITC (International Trypanotolerance Centre), Banjul, The Gambia

Uzoewulu GN, Lawson L, Nnanna IS, Rastogi N, Goyal M (2016) Genetic diversity of Mycobacterium tuberculosis complex strains isolated from patients with pulmonary tuberculosis in Anambra State, Nigeria. Int J Mycobacteriology 5(1):74–79. doi:http://dx.doi.org/10.1016/j.ijmyco.2015.06.008

Van Der Groen G, Pattyn SR (1975) Comparison of Mycobacterium africanum, M. tuberculosis and M. bovis by their utilization of carbon and nitrogen sources. Ann Soc Belg Med Trop 55(6):647–651

van der Wel N, Hava D, Houben D, Fluitsma D, van Zon M, Pierson J, Brenner M, Peters PJ (2007) M. tuberculosis and M. leprae translocate from the phagolysosome to the cytosol in myeloid cells. Cell 129(7):1287–1298. S0092-8674(07)00782-9 [pii]. doi:10.1016/j.cell.2007.05.059

Vasconcellos SEG, Huard RC, Niemann S, Kremer K, Santos AR, Suffys PN, Ho JL (2010) Distinct genotypic profiles of the two major clades of Mycobacterium africanum. BMC Infect Dis 10(1):1–16. doi:10.1186/1471-2334-10-80

Wang Z, Potter BM, Gray AM, Sacksteder KA, Geisbrecht BV, Laity JH (2007) The solution structure of antigen MPT64 from Mycobacterium tuberculosis defines a new family of beta-grasp proteins. J Mol Biol 366(2):375–381. S0022-2836(06)01579-8 [pii]. doi:10.1016/j.jmb.2006.11.039

Wayne LG, Kubica GP (1986) The Mycobacteria. In: Snears PHA, Mair NS, Sharpe ME, Holt JG (eds) Bergey's manual of systematic bacteriology, vol 2, 7th edn. William &Wilkins Co, Baltimore, pp 1436–1457

Waziri NE, Cadmus S, Nguku P, Fawole O, Owolodun OA, Waziri H, Ibrahim L, Biya O, Gidado S, Badung S, Kumbish P, Nsubuga P (2014) Factors associated with tuberculosis among patients attending a treatment centre in Zaria, North-west Nigeria, 2010. Pan Afr Med J 18(Suppl 1):5. doi:10.11694/pamj.supp.2014.18.1.4189PAMJ-SUPP-18-1-05. [pii]

Weber A, Reischl U, Naumann L (1998) Demonstration of Mycobacterium africanum in a bull from North Bavaria. Berl Munch Tierarztl Wochenschr 111(1):6–8

Winglee K, Manson McGuire A, Maiga M, Abeel T, Shea T, Desjardins CA, Diarra B, Baya B, Sanogo M, Diallo S, Earl AM, Bishai WR (2016) Whole genome sequencing of Mycobacterium africanum strains from Mali provides insights into the mechanisms of geographic restriction. PLoS Negl Trop Dis 10 (1):e0004332. 10.1371/journal.pntd.0004332PNTD-D-15-01035 [pii]

Yeboah-Manu D, Asante-Poku A, Bodmer T, Stucki D, Koram K, Bonsu F, Pluschke G, Gagneux S (2011) Genotypic diversity and drug susceptibility patterns among M. tuberculosis complex isolates from South-Western Ghana. PLoS One 6 (7):e21906. 10.1371/journal.pone.0021906PONE-D-10-05029 [pii]

Zink AR, Sola C, Reischl U, Grabner W, Rastogi N, Wolf H, Nerlich AG (2003) Characterization of Mycobacterium tuberculosis complex DNAs from Egyptian mummies by spoligotyping. J Clin Microbiol 41(1):359–367

Mycobacterium tuberculosis Complex Members Adapted to Wild and Domestic Animals

7

Kerri M. Malone and Stephen V. Gordon

Abstract

The *Mycobacterium tuberculosis* complex (MTBC) is composed of several highly genetically related species that can be broadly classified into those that are human-host adapted and those that possess the ability to propagate and transmit in a variety of wild and domesticated animals. Since the initial description of the bovine tubercle bacillus, now known as *Mycobacterium bovis*, by Theobald Smith in the late 1800's, isolates originating from a wide range of animal hosts have been identified and characterized as *M. microti*, *M. pinnipedii*, the Dassie bacillus, *M. mungi*, *M. caprae*, *M. orygis* and *M. suricattae*. This chapter outlines the events resulting in the identification of each of these animal-adapted species, their close genetic relationships, and how genome-based phylogenetic analyses of species-specific variation amongst MTBC members is beginning to unravel the events that resulted in the evolution of the MTBC and the observed host tropism between the human- and animal-adapted member species.

Keywords

Mycobacterium tuberculosis complex • MTBC • *Mycobacterium bovis* • Host adaptation • Animal-adapted species • One Health

K.M. Malone (✉)
UCD School of Veterinary Medicine, University College Dublin, Dublin 4, Ireland
e-mail: kerri.malone@ucd.ie

S.V. Gordon
UCD School of Veterinary Medicine, University College Dublin, Dublin 4, Ireland

UCD School of Medicine, University College Dublin, Dublin 4, Ireland

UCD School of Biomolecular and Biomedical Science, University College Dublin, Dublin 4, Ireland

UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin 4, Ireland
e-mail: stephen.gordon@ucd.ie

S. Gagneux (ed.), *Strain Variation in the* Mycobacterium tuberculosis *Complex: Its Role in Biology, Epidemiology and Control*, Advances in Experimental Medicine and Biology 1019,
DOI 10.1007/978-3-319-64371-7_7

7.1 What's in a Name? Defining the Animal-Adapted Species of the MTBC

The *Mycobacterium tuberculosis* complex (MTBC) is composed of several highly genetically related species (>99% nucleotide identity) that can be broadly classified into those that are human-host adapted, such as *M. tuberculosis* and *Mycobacterium africanum,* and those that possess the ability to propagate and transmit in a wide variety of wild and domesticated animal hosts. Following the identification of *M. tuberculosis* by Robert Koch in the sputum of a tuberculosis patient in 1882, Theobald Smith in 1896 was the first to demonstrate that the causative agent of tuberculosis in cattle, and indeed other animal hosts, was not the same as the human bacillus, a finding that ultimately led to the description of the bovine-adapted species *Mycobacterium bovis* (Smith 1898). Since then, there have been numerous reports of tuberculous isolates originating from a wide range of animal hosts including, but certainly not limited to, possums, voles, deer, shrews, pigs, goats, cats, meercats, mice, dogs, badgers, dassies, seals and rhinoceroses; as these animal-related isolates were identified, they were named after the host from which they were originally or most frequently isolated (Alexander et al. 2010; Boardman et al. 2014; Coscolla et al. 2013; Cousins et al. 1994; De Garine-Wichatitsky et al. 2013; Fitzgerald and Kaneene 2013; Miller et al. 2016; Parra et al. 2003; Parsons et al. 2013; Pesciaroli et al. 2014; Wells 1937, 1946). This has resulted in the nomenclature of *M. bovis* (Karlson and Lessel 1970), *M. microti* (Wells 1937, 1946), *M. pinnipedii* (Cousins et al. 1993), the Dassie bacillus (Cousins et al. 1994), *M. mungi* (Alexander et al. 2010), *M. caprae* (Aranaz et al. 1999), *M. orygis* (van Ingen et al. 2012) and *M. suricattae* (Parsons et al. 2013), while a novel isolate derived from a chimpanzee with pulmonary tuberculosis in Cote d'Ivoire has been recently identified that is tentatively referred to as the chimpanzee bacillus (Coscolla et al. 2013). However, this process and nomenclature can be misleading as it may result in erroneous redundancy within the complex, it does not exclusively define host range and it hinges upon whether the host in question is one to which the associated pathogen is adapted to and can successfully propagate within ('primary' or 'maintenance' host) or one to which the pathogen has been introduced to but is not adapted to for propagation ('secondary' or 'spillover' host). This issue can be seen in the context of the cattle pathogen *M. bovis* and the diverse range of maintenance hosts from which it can be identified (e.g. badgers, white tailed deer, bush-tailed possum) while instances of *M. bovis* infection in humans as a spillover event has also been documented (Cadmus et al. 2006; Chen et al. 2009; De Garine-Wichatitsky et al. 2013; Soto et al. 2004; Waters and Palmer 2015).

A more rational view of the animal-adapted MTBC species is to consider them as 'ecotypes'. In doing so, this concept allows for the definition of MTBC members as evolutionarily related groups, or clades, which adapt to and sustain within distinct host populations and where strains of a given clade are immune to periodic selection events that occur within independent clades (Cohan 2002; Smith et al. 2006). This classification, along with genome-based phylogenetic analyses of species-specific variation amongst MTBC members, is beginning to unravel the evolutionary events that resulted in the formation of the MTBC and the observed host tropism between the human- and animal-adapted member species. With the availability of the *M. tuberculosis* genome sequence in 1998, the genomic sequences of the MTBC species were interrogated to determine if specific molecular barriers associated with sustainability within a given host could be defined (Cole et al. 1998). In doing so, it became apparent that single nucleotide polymorphisms (SNPs) and/or regions of deletion (RDs) could assemble the entire MTBC into a phylogenetic architecture that broadly mirrors the observed host specificity amongst the MTBC members (Brosch et al. 2002; Cole et al. 1998; Huard et al. 2006; Smith et al. 2006, 2012b). The animal adapted species of the MTBC form a clade separate from *M. tuberculosis* and closer to Lineage 6 or *M. africanum* West-African 2

(WA2) that is demarcated by the absence of RD9; this is in stark contrast to traditional ideas whereby animal-adapted strains were thought to be closer to the progenitor of the MTBC (Smith et al. 2006, 2009b). WA2 strains exhibit phenotypic variation and show attenuation in the human host in that those infected with these strains are less likely to progress to active tuberculosis in comparison to other human-adapted strains (de Jong et al. 2008). These findings suggest that early diversification of an *M. africanum*-like strain that was maintained within African human populations or in a yet to be identified animal host subsequently gave rise to the animal-adapted lineages, which coincides with the geographical location of animal-adapted MTBC member hosts such as banded mongooses, oryx, waterbucks, gazelles and meercats.

7.2 The Animal-Adapted MTBC Strains of Small Animals and Rodents

There is a subgroup of animal-adapted MTBC species that infect small animals and rodents which can be differentiated from those that cause tuberculosis in bovine, caprine and pinniped hosts by the presence of intact RD12 and RD13 sequences in their genomes which include *M. microti,* the Dassie bacillus, *M. mungi* and the recently classified *M. suricattae* (Alexander et al. 2010; Cousins et al. 1993; Dippenaar et al. 2015; Frota et al. 2004; Mostowy et al. 2004; Wells 1937, 1946).

7.2.1 *Mycobacterium microti*

Mycobacterium microti, previously known as the vole bacillus, was first identified by Wells in 1937 in Great Britain as the causative agent of tuberculosis in small rodents such as the wood mouse (*Apodemus sylvaticus)*, shrews (*Sorex araneus*) and voles (*Microtus agretis*) with the latter being the most commonly identified host (Cavanagh et al. 2002; Wells 1937, 1946). *M. microti* has also been documented as a zoonosis and has been isolated from immunocompromised patients in the United Kingdom, Germany, Switzerland and the Netherlands (Cavanagh et al. 2002; Niemann et al. 2000; van Soolingen et al. 1998). Besides small rodents, *M. microti* is most frequently isolated from felines in Great Britain and it is speculated that feline hunting practices are responsible for this spillover event (Smith et al. 2009a). After its identification, *M. microti* showed promise for use as a live vaccine against tuberculosis as it was found safe when administered to pigs, cattle and humans and its use did not result in pathology or progressive disease in animal models unless given in high doses; however an *M. microti* vaccine was not developed further after this point as trial results failed to demonstrate increased efficacy in comparison to the BCG vaccine (Hart and Sutherland 1977; Sula and Radkovsky 1976). As *M. microti* infection is protective in this manner, it has been suggested that the lack of recovery of *M. bovis* from cats in the south-western regions of the United Kingdom may be due to high burden levels of *M. microti* in the felines of this area (Smith et al. 2009a).

Prior to the application of genotyping methods for the identification of MTBC members, it was difficult to distinguish *M. microti* from *M. africanum* and *M. bovis* biochemically, with different *M. microti* strains resulting in varying niacin accumulation and nitrate reduction test readings (Levy-Frebault and Portaels 1992). In two studies between 2002 and 2004, the *M. microti* genome was compared to the *M. tuberculosis* H37Rv and *M. bovis* AF2122/97 reference strain genomes (Brodin et al. 2002; Frota et al. 2004). The results revealed that like other animal-adapted species, the *M. microti* genome had RD7, RD9 and RD10 genomic deletions, intact RD12 and RD13 sequences along with a unique *M. microti* deletion known as RD1^{mic}, a 14 kb region that shows partial overlap with the RD1 deletion of *M. bovis* BCG that is responsible for the attenuation of the BCG vaccine strain (Lewis et al. 2003; Pym et al. 2002). Also, there were four other deletions that were found to be specific to *M. microti* and these include RD5^{mic}, MiD1, MiD2 and MiD3. Taken together, these

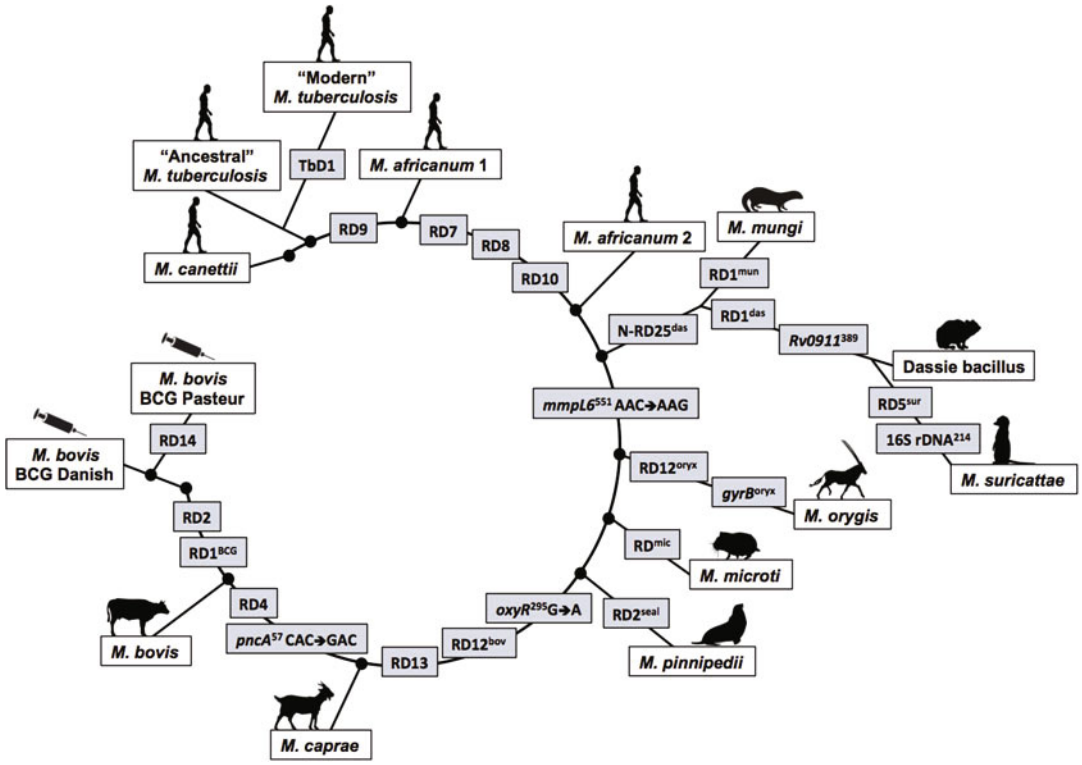

Fig. 7.1 The molecular markers and regions of deletion that distinguish the animal adapted species of the *Mycobacterium tuberculosis* complex. This diagram depicts the regions of deletion (RD) and specific nucleotide polymorphisms (SNPs) that are described in the main text and that can be used to describe the evolution of the MTBC phylogeny, from an *M. canetti*–like ancestor to *M. bovis* BCG Pasteur. Sequential loss of DNA segments act as unique markers with which the broad evolutionary steps of the phylogeny can be elucidated. Animal silhouettes illustrate the primary host from which the associated MTBC species was originally or most frequently isolated from while the syringe denotes a vaccine strain (Alexander et al. 2010; Aranaz et al. 1999; Brosch et al. 1998; Dippenaar et al. 2015; Frota et al. 2004; Garnier et al. 2003; Gordon et al. 1999; Marmiesse et al. 2004; Mostowy et al. 2004; van Ingen et al. 2012)

genomic markers placed *M. microti* between *M. tuberculosis* and *M. bovis* on the evolutionary scale of the MTBC complex (Brodin et al. 2002; Frota et al. 2004) (Fig. 7.1). A complete genome sequence of *M. microti* was published at the start of 2016 (Zhu et al. 2016).

7.2.2 The Dassie bacillus, *Mycobacterium mungi and Mycobacterium suricattae*

The Dassie bacillus was first described in 1960 and was isolated from the South African Cape hydrax (*Procavia capensis*), commonly known as the dassie (Smith 1960). The Dassie bacillus was reported to be attenuated in the vole, mouse, rabbit and guinea pig hosts in experimental infections, with no reports, thus far, of infection in humans (Cousins et al. 1994; Smith 1965).

M. mungi was first identified in a study of the banded mongoose (*Mungos mungo)* in Botswana in 2010 where the population in question was living in close proximity to humans with some of the animals dying from what appeared to be a tuberculosis-like disease (Alexander et al. 2010). The banded mongoose, in contrast to the majority of other mongoose species, lives in a colony with complex social hierarchies; animals scent mark their territory and communicate with others

within the colony through anal gland secretions and urine markings (Alexander et al. 2016). *M. mungi* was found to be environmentally transmitted amongst the studied mongoose population, primarily through abrasions in the nose that most likely occur from foraging activity, and this represents a novel mode of transmission for an MTBC species in contrast to the traditional aerosolized route (Alexander et al. 2016). Furthermore, *M. mungi* was found in the urine, respiratory secretions and anal gland tissue and secretions of infected mongooses, similar to the levels found in tuberculous lung tissue from other host species infected with *M. bovis* (Alexander et al. 2016). Hence, *M. mungi* has effectively usurped the social behaviour of the banded mongoose for dissemination and propagation amongst the population in that scent sniffing effectively transmits the pathogen from contaminated secretions originating from infected mongooses. Also of note is that infection of the mongoose with *M. mungi* is reported to progress rapidly, with infected animals succumbing within 2–3 months of infection; this is in contrast to the chronic disease established with infection with other MTBC species (Alexander et al. 2010). As mongoose troops in Botswana live in close proximity to human settlements, understanding the ecology and the potential for transmission events to occur within this environment will be critical for assessing the potential for the emergence of zoonotic infections.

Currently, the newest member of the MTBC to be described is *M. suricattae* (Dippenaar et al. 2015; Parsons et al. 2013). In 2002, disseminated tuberculosis was described in a troop of free-living meercats (*Suricata suricatta*) in the Kalahari Desert in Botswana and histopathologic examination revealed granulomas in most tissues and organs of infected animals including the lymph nodes, spleen, liver and lungs (Parsons et al. 2013). *M. suricattae* is highly pathogenic for meercats and appears to be more virulent than the Dassie bacillus in its eponymous host (Dippenaar et al. 2015; Parsons et al. 2013).

Comprehensive molecular and whole-genome analyses of the Dassie bacillus, *M. mungi,* and *M. suricattae* have revealed the close genetic relationship amongst these species in that they share genomic deletions RD7, RD8, RD9 and RD10. The Dassie bacillus was first characterized as being most similar to *M. microti* and investigation revealed genomic deletions specific to the Dassie bacillus that include RD5^{das}, RD25^{das} and an RD1-like deletion variant, RD1^{das}, representing the smallest such deletion at the locus encoding the ESX-1 secretion system discovered to date (Mostowy et al. 2004). Further to this, a SNP in *Rv0911* at position 389 was revealed in the Dassie bacillus genome that could distinguish the Dassie bacillus from all other members of the MTBC defined up to then (Huard et al. 2006). With the identification of tuberculosis in the banded mongoose, investigations into the genetic polymorphisms present in the isolated strain sought to identify if a novel primary host for an existing MTBC member or a novel MTBC member had been uncovered; as the latter was shown to be the case, *M. mungi* was defined (Alexander et al. 2010). *M. mungi* was found to share RD25^{das} with the Dassie bacillus but it was the lack of a SNP present in *Rv0911*^{389} and an RD1 variant deletion (RD1^{mun} which is larger than RD1^{das}) that differentiated it from the Dassie bacillus. Furthermore, a unique spoligotyping pattern was identified for *M. mungi* that was found consistent within *M. mungi* strains isolated from varying mongoose troops in diverse geographical locations between 2000 and 2009 (Alexander et al. 2010). Subsequently, following investigations aimed to differentiate the causative agent of tuberculosis in meercats, *M. suricattae,* from human-adapted pathogens *M. tuberculosis* and *M. africanum*, from *M. bovis* and in particular from *M. mungi* and the Dassie bacillus (Parsons et al. 2013). *M. suricattae* was found to share the RD1^{das} and N-RD25^{das} deletions found in the Dassie bacillus and the presence of RD1^{das} and the *Rv0911*^{389} SNP (originally described as specific to the Dassie bacillus) in the *M. suricattae* genome allowed for its differentiation from *M. mungi*. The ability to differentiate *M. suricattae* from the Dassie bacillus relied on two genomic variant events; 21 copies of the IS*6110* insertion element were found in the *M. suricattae* strains analysed in contrast to 15 copies in the

Dassie bacillus, while an RD5^{bovis}-like genomic deletion variant was identified, RD5^{sur}, that is unique to *M. suricattae* and contains an IS*6110* insertion element. Finally, spilogotyping analysis of *M. suricattae* revealed that the entire direct repeat region along with 3500 bp upstream and 1700 bp downstream of the region was deleted, while a SNP was disclosed in the 16S rDNA (T214G) that distinguishes *M. suricattae* from all other members of the MTBC (Parsons et al. 2013). Of course, a natural caveat to the increasing membership of the MTBC is that with the increased resolution of molecular methods, the inclination to designate each new animal isolate as a separate 'species' may complicate discourse around host adaptation.

7.3 The Animal-Adapted MTBC Strains of Bovines, Caprines and Pinnipeds

Another subgroup of animal-adapted MTBC members that can be differentiated from those that cause tuberculosis in small animals and rodents include *M. orygis*, *M. caprae, M. pinnipedii* and the hallmark animal-adapted strain *M. bovis*.

7.3.1 *Mycobacterium orygis*

Mycobacterium orygis, formally known as the oryx bacillus, was first described in 1976 after isolation from two East African oryxes and was elevated to subspecies status in 2012 following the characterization of 22 clinical isolates (Lomme et al. 1976; van Ingen et al. 2012). *M. orygis* has been isolated from many members of the *Bovidae* family and the true extent of its host range is unknown; isolates have been recovered from waterbucks (*Kobus ellipsirpymnus*), different species of deer, free ranging buffalo (*Syncerus caffer*), cattle and rhinoceroses (*Rhinoceros unicornis*) (Gey van Pittius et al. 2012; Thapa et al. 2015; van Soolingen et al. 1994). Furthermore, *M. orygis* has also been identified in primate and human hosts; reports of isolates recovered from rhesus monkeys and patients in Africa and south Asia suggest animal-human transmission with humans representing a spillover host (Thapa et al. 2015). Prior to its elevation as a species, *M. orygis* may have been misidentified as *M. africanum* on account of the *gyrB*^{1450} SNP mutation that they share, and *M. orygis* was also once considered a variant of *M. bovis* (van Ingen et al. 2012). Classification of *M. orygis* discovered species-specific polymorphisms within its genome such as a GGC mutation in *Rv2042, a gyrB*^{113} G-A single nucleotide polymorphism and a subspecies specific RD12 deletion (RD12^{oryx}) that can differentiate *M. orygis* from all other members of the MTBC. Furthermore, unlike the Dassie bacillus, *M. mungi* and *M. microti, M. orygis* displays an intact RD1 locus (van Ingen et al. 2012).

7.3.2 *Mycobacterium caprae*

Mycobacterium caprae, originally known as *M. tuberculosis* subsp. *caprae* or *M. bovis* subsp. *caprae* was first isolated from a goat with disseminated tuberculosis in 1999, and in 2003 it was officially defined as a member of the MTBC (Aranaz et al. 1996, 1999, 2003; Niemann et al. 2002). *M. caprae* has been identified in several European countries including Spain, Austria and Germany and it is not solely restricted to the caprine host; *M. caprae* had been isolated from sheep (*Ovis aries*), red deer (*Cervus elaphus*), cattle, wild boar (*Sus scrofa*), from humans and captive animals such as the Siberian tiger (*Panthera tigris altaica*), camel (*Camelus dromedarius*) and bison (*Bison bison*) (Chiari et al. 2014; Cvetnic et al. 2007; Fink et al. 2015; Hansen et al. 2012; Kubica et al. 2003; Lantos et al. 2003; Munoz Mendoza et al. 2012; Parra et al. 2003; Pate et al. 2006; Prodinger et al. 2014; Rodriguez et al. 2009, 2011). Along with *M. bovis*, *M. caprae* constitutes a sub-cluster within the MTBC delimited by the deletion of RD12 and RD13 along with a SNP in the *oxyR* gene (G295A) (Aranaz et al. 1999; Garnier et al. 2003). *M. caprae* is generally thought of as less diverse than *M. bovis* strains with two main

spoligotype patterns identified, one that is found in the Iberian peninsula and the other which is found in Central and Western Europe (Aranaz et al. 2003; Duarte et al. 2008; Erler et al. 2004; Prodinger et al. 2002; Rodriguez et al. 2011). *M. caprae* can be distinguished from *M. bovis* based on its sensitivity to pyrazinamide, spoligotype pattern, the presence of wild-type *pncA* and RD4 in the *M. caprae* genome, while a SNP in *gyrB* at position T1311G differentiates *M. caprae* from the rest of the complex members (Aranaz et al. 1999; Brosch et al. 2002). The first complete *M. caprae* genomic sequence was published in 2015 and will provide the basis for future in-depth comparative genomic analyses with other members of the MTBC (de la Fuente et al. 2015).

7.3.3 *Mycobacterium pinnipedii*

Once known as the seal bacillus after isolation from its primary host both in the wild and captivity, *M. pinnipedii* was classified as a separate member of the MTBC in 2003 (Cousins et al. 2003). Apart from wild and captive Australian fur seals (*Arctocephalus forsteri*) and sea lions (*Neophoca cinera*), *M. pinnipedii* has been isolated from Brazilian and Malayan tapirs (*Tapirus indicus*) and the camel (*Camelus bactrianus bactrianus*) (Cousins et al. 1993; Forshaw and Phelps 1991; Moser et al. 2008; Woods et al. 1995). Recovery of *M. pinnipedii* from human hosts has also been reported. Seal trainers from both the Netherlands and Australia were found to be infected with *M. pinnipedii* and there are even reports of the recovery of an-*M. pinnipedii* related strain from 1000 year old Peruvian human skeletons; however, whether *M. pinnipedii* possesses the ability to successfully transmit and sustain within the human population is yet unknown (Bos et al. 2014; Kiers et al. 2008; Thompson et al. 1993). The genome of *M. pinnipedii* contains the subspecies specific deletion PiD1 which can distinguish it from the rest of the MTBC species while PiD2/RD2^{seal} is also deleted from the *M. pinnipedii* genome and results in the removal of genes associated with the mammalian cell entry 3 (Mce3) locus, disruption of which is associated with attenuation of *M. tuberculosis* in mice (Bigi et al. 2005; Cousins et al. 2003; Senaratne et al. 2008).

7.3.4 *Mycobacterium bovis*

Mycobacterium bovis is the causative agent of bovine tuberculosis and is the most widely studied animal-adapted MTBC species as bovine tuberculosis exacts a tremendous global economic toll through the loss of productivity and disease control costs, results in a negative impact on animal health and welfare, and is a threat to human health through zoonotic infections (Abernethy et al. 2013; Heath 2013; Mableson et al. 2014; Muller et al. 2013). The primary control of bovine tuberculosis in developed countries with a control policy relies on a 'test and slaughter' policy whereby those cattle that test positive to the tuberculin skin test ('reactor' animals) are subsequently removed from the herd and slaughtered (Barker 2015; More and Good 2006). Failure to eradicate bovine tuberculosis can be attributed to a number of factors including on-farm practices and policies, missed surveillance of disease in abattoirs, a lack of an effective vaccine and specificity/sensitivity issues associated with current diagnostic testing regimens (More and Good 2006; Sheridan 2011). However, one of the many difficulties impeding the control and eradication of bovine tuberculosis is the presence of wildlife reservoirs of *M. bovis* infection. *M. bovis* can be found in a diverse range of maintenance hosts that include (and are not limited to) species such as white-tailed deer (*Odocoileus virginianus*), elk (*Cervus Canadensis*), domestic goats (*Capra aegagrus hircus*), North American buffalo (*Bison bison*), wood bison (*Bison bison athabascae)*, brushtail possums (*Trichosurus vulpecula*), and the European badger (*Meles meles*) (Drewe et al. 2013; Lees et al. 2003; Pesciaroli et al. 2014; Waters and Palmer 2015; Wright et al. 2015).

The badger is considered a highly susceptible host for *M. bovis* infection (Corner et al. 2007). Infected badgers present with a range of disease states of which only a small proportion

develop general pathology; one study determined that 30/57 naturally infected badgers had latent infection with no signs of gross pathology, a difficulty with regards to wildlife disease surveillance (Corner et al. 2012). Epidemiological data has shown that tuberculosis is endemic in Irish badgers and that cross-species transmission may play a significant factor in the failure to totally eradicate tuberculosis from Irish cattle herds. For example, the most widely distributed *M. bovis* genotype found in Ireland could be detected in proximal cattle and badger populations while *M. bovis* strains isolated from the associated badger and cattle populations were similar, indicating cross-species transmission (Furphy et al. 2012; Olea-Popelka et al. 2005). The route of transmission from badger to cattle and vice versa has yet to be fully elucidated although it has been speculated that close contact at water reservoirs, aerosol transmission and to a lesser extent faeces transmission may play a role (Corner et al. 2012; Gormley and Corner 2013). Several studies have shown that the selective removal of badgers led to an improvement in the bovine tuberculosis rates in cattle within a given population and as a result, culling of badgers in both the UK and Ireland is undertaken in an effort to reduce transmission to cattle, particularly in areas of high disease prevalence in cattle (Gormley and Corner 2013; Griffin et al. 2005; Kelly et al. 2008; More and Good 2006, 2015; Olea-Popelka et al. 2009; Smith et al. 2012a). Furthermore, efforts towards developing a vaccine for badgers are currently being pursued as a practical replacement for badger culling activity and to reduce the prevalence of infection and the rate of *M. bovis* transmission to cattle (Chambers et al. 2011; Corner et al. 2008, 2010; Murphy et al. 2014).

Genetic diversity and the geographical distribution of *M. bovis* strains provide insight into the overall success of bovine tuberculosis surveillance programmes. For example, since 2000, France has been officially declared bovine tuberculosis free by the EU. However, a survey of over 4000 *M. bovis* strains isolated from livestock, wildlife and domesticated animals between the years 1978 and 2010 has revealed that the number of cases of *M. bovis* infection are on the increase in France and that three spoligotypes dominate amongst animal populations namely SB0120 (26% of strains tested), SB0134 (11%) and SB0121 (6%) (Hauer et al. 2015). The occurrence of these spoligotypes within livestock, wildlife (red deer, badgers, boar) and domesticated animals (cats, dogs, zoo animals) highlights the inter-species transmission of *M. bovis* and the challenges faced in eradicating infection from animal reservoirs (Hauer et al. 2015).

A study in the United Kingdom explored whether any phenotypic differences existed between the ten major *M. bovis* spoligotypes that occur within the national herd and it was found that strains could be grouped independently of each other based on their global metabolic profiles, suggesting differential metabolic features are associated with different *M. bovis* spoligotypes (Winder et al. 2006). Furthermore, another study highlighted differences with respect to the expression of a subset of genes between the four most common UK *M. bovis* spoligotypes (SB0263, SB0140, SB0129 and SB0134) with some of this diversity linked to the occurrence of SNPs between these highly genetically related strains (Golby et al. 2013). Taken together, these studies suggest that the slight differences in genetic makeup across *M. bovis* lineages affect the phenotypic traits of the associated strains; this diversity may ultimately influence diagnostic test performance and monitoring of disease burden and have consequences for diagnostic test and vaccine development.

M. bovis not only presents as a threat to wild and domesticated animals, but is also a zoonotic threat. In 2006, the WHO declared bovine tuberculosis as one of the world's most neglected zoonotic diseases with little progress being made, particularly in Africa, to combat this threat to human health (Mableson et al. 2014; Muller et al. 2013; WHO 2005). Although *M. bovis* is generally regarded as attenuated within human hosts, there have been several reports of tuberculosis caused by *M. bovis* infection in humans, especially within those in close contact with cattle such as abattoir workers, cattle herders and

those who co-habit with livestock (De Garine-Wichatitsky et al. 2013; Kazoora et al. 2016; Khattak et al. 2016; Malama et al. 2014). Furthermore, genetic diversity within the *M. bovis* lineage also presents as a threat to the human population. For example, a multi-drug resistant strain of *M. bovis* that transmitted between HIV^+ patients was found to possess an IS6110 insertion upstream of the *phoP* gene that may have played a role in potentiating the strain's capacity for human-human transmission (Soto et al. 2004).

Constant vigilance and surveillance of national herds and wildlife reservoirs is paramount for the eventual eradication of bovine tuberculosis. Further research should focus on defining the ecology and interplay of reservoir hosts, cattle and humans and how genetic diversity in *M. bovis* may impact upon these interactions.

7.4 Host Tropism Within the MTBC

As mentioned, the MTBC can be broadly classified into those that are human-host adapted and those that possess the ability to propagate and transmit in animal hosts. Host tropism amongst MTBC species is a balance between host-driven and pathogen-driven processes, where differences in the immune response of the host or differences in the phenotypes of the MTBC species drive infection, persistence, disease and transmission. The focus of studies of the MTBC is largely driven by their threat to human health; *M. tuberculosis* is estimated to latently infect one third of the world's population and tuberculosis was the world's lead cause of death from an infectious agent in 2015 (WHO 2015), while *M. bovis* infection of domesticated animals poses a threat to animal productivity, welfare and zoonotic infection. There are a small number of reports of these two MTBC members sustaining in the opposite host; cattle presenting with *M. tuberculosis* infection have been found in regions of high human tuberculosis prevalence and in African countries such as Ethiopia, Nigeria, and India, while *M. bovis* infection in humans has been documented in those working or living in close proximity with animals, or in the immunocompromised (Ameni et al. 2013; Berg et al. 2009; Cadmus et al. 2006; Chen et al. 2009; Du et al. 2011; Soto et al. 2004; Srivastava et al. 2008). In 2010, Whelan et al. revisited the host specificity between *M. tuberculosis* and *M. bovis* in the bovine host, and the distinct attenuation of *M. tuberculosis* in cattle was demonstrated (Whelan et al. 2010). While experimental infection of Friesian cattle with *M. bovis* AF2122/97 led to pathology characteristic of bovine tuberculosis after 16 weeks, infection of cattle with *M. tuberculosis* H37Rv did not result in gross pathology, despite both groups of animals showing positive skin-test and interferon-gamma responses indicative of successful infection (Whelan et al. 2010). Furthermore, it has also been shown that infection of bovine alveolar and blood-derived macrophages in vitro with *M. bovis* and *M. tuberculosis* resulted in differential production of nitric oxide and select chemokines and cytokines (Magee et al. 2014; Piercy et al. 2007; Widdison et al. 2008). These results therefore demonstrate a difference in the immune response of the host in the face of infection with the two highly related pathogens. From the viewpoint of the pathogen, understanding the evolutionary events that led to the adaptation of MTBC species to specific primary hosts will allow for the molecular definition of host tropism; this will not only aid in understanding the evolution of the MTBC during speciation as a whole but also provide fundamental data to underpin the development of novel control tools. The transition from human-to-animal host within the MTBC is associated with an overall reduction in genome size (Brosch et al. 2002; Garnier et al. 2003). The question as to which specific genome polymorphisms found amongst the members of the MTBC impact upon host tropism, and whether or not a single genomic deletion or multiple deletion events are required, has yet to be resolved; however, there are some genomic events between members of the MTBC that have been found to impact upon pathogenicity. It is clear that loss of the region encompassing RD1, which encodes for components of the ESX-1

secretion system, is responsible for attenuation of *M. bovis* BCG in vivo (Gordon et al. 1999; Pym et al. 2002). That being said, *M. bovis* BCG has been documented to infect immunocompromised hosts receiving immunotherapy while *M. microti* and the Dassie bacillus also contain variants of an RD1 deletion ($RD1^{mic}$ and $RD1^{das}$ respectively) and yet possess the ability to cause tuberculosis in their respective hosts, with additional reports of *M. microti* infecting immunocompromised hosts (Frota et al. 2004; Mostowy et al. 2004; Perez-Jacoiste Asin et al. 2014). Conversely, the Mce3 pathway encoded by RD7 in *M. tuberculosis* is expressed during infection and when disrupted results in the attenuation of *M. tuberculosis* in mice; equally, RD7 is deleted from the *M. bovis* genome and yet its loss clearly does not attenuate *M. bovis* in mice or cattle (Ahmad et al. 2005; Senaratne et al. 2008). As well as the deletion of specific genomic segments, there are also some examples of minor genetic modifications leading to significant phenotypic changes between MTBC members. A SNP present in the promoter of the *narGHJI* locus is responsible for the difference in nitrate reductase activity between *M. bovis* and *M. tuberculosis* and is a trait that is used to distinguish the two species diagnostically while a frameshift event in *pks1-15* results in a loss of phenolic glycolipid production from *M. tuberculosis* H37Rv in contrast to other *M. tuberculosis* strains and *M. bovis* (Pang et al. 2012; Sinsimer et al. 2008; Sohaskey and Modesti 2009; Stermann et al. 2003).

As *M. bovis* is the most studied animal-adapted member of the MTBC at both the genetic and phenotypic level, comparison with the human-adapted *M. tuberculosis* aims to reveal the impact of reductive evolution or genome downsizing on host specificity between the two species. *M. bovis* AF2122/97 was the first bovine mycobacterial strain to be fully sequenced (Garnier et al. 2003). In the original *M. bovis* AF2122/97 genome publication, the authors highlighted the major differences between the *M. bovis* AF2122/97 genome and *M. tuberculosis* H37Rv reference sequence with regards to the potential impact upon the phenotype of both species (Garnier et al. 2003) (Fig. 7.2). These analyses revealed high nucleotide sequence identity (>99%) between the two genomes, no unique genes *per se* found in the *M. bovis* AF2122/97 genome and that a number of genomic deletions led to a reduced genome size for *M. bovis* AF2122/97 e.g. RD4, RD7-10. Also, examples of gene alteration events such as frameshifts (alanine dehydrogenase Ald and lipoprotein LppO) and deletions (Esx-family proteins Rv2346c, Rv2347c, Rv3619c and Rv3620c) were found in the *M. bovis* AF2122/97 genome and a link was suggested between the lack of pyruvate kinase activity in *M. bovis* AF2122/97 and a SNP in the *pykA* gene which was later independently verified (Garnier et al. 2003; Keating et al. 2005). At the time, it was unknown how these nucleotide polymorphisms impacted upon the phenotype of the two species to result in altered host-pathogen interactions, virulence and host tropism. In 2007, two microarray-based publications compared the transcriptome of *M. bovis* to that of *M. tuberculosis* to identify differential global gene expression between the two pathogens. Both studies highlighted that the genes encoding the major antigens MPB83 and MPB70 (MPT83 and MPT70 in *M. tuberculosis*) were expressed at higher levels in *M. bovis* in comparison to *M. tuberculosis* (Golby et al. 2007; Rehren et al. 2007). MPB83 is a lipoprotein that is post-translationally glycosylated and anchored in the cell membrane while MPB70 is a secreted protein (Hewinson et al. 1996; Wiker 2009), with the function of both proteins unknown. The expression of both of these proteins is under the control of the SigK regulator (Charlet et al. 2005; Said-Salim et al. 2006). In *M. tuberculosis* these antigens show low-level expression and selective induction during intracellular growth, while *M. bovis* shows constitutive high-level expression of both genes as a result of the loss of negative regulation due to the mutation of the anti-sigma factor RskA (Charlet et al. 2005; Said-Salim et al. 2006; Schnappinger et al. 2003). Constitutive expression of MPB70 and MPB83 can also be seen in *M. orygis* where an independent mutation in the *rskA* gene can be found; this

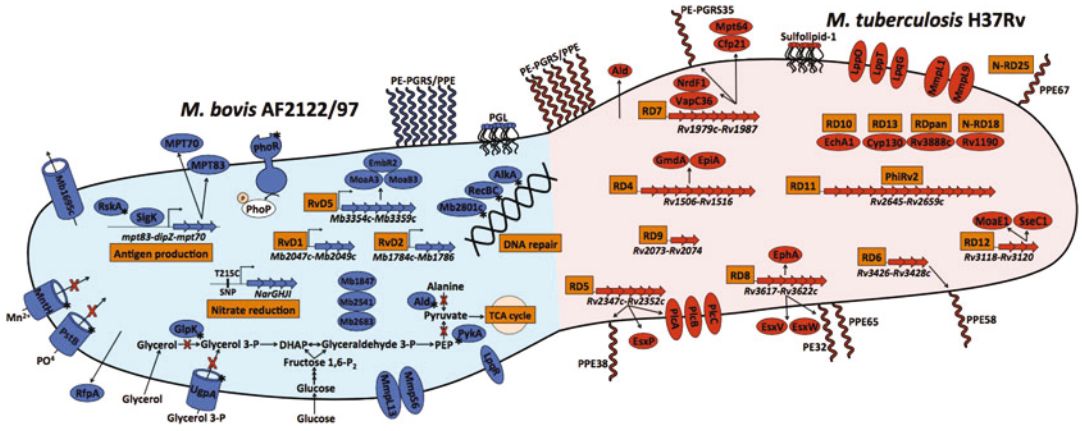

Fig. 7.2 Schematic of the major differences between *M. bovis* AF2122/97 and *M. tuberculosis* H37Rv. Depicted are the genotypic and phenotypic differences predicted and/or experimentally validated to date between *M. bovis* AF2122/97 (*blue shading*) and *M. tuberculosis* H37Rv (*red shading*). Surface-exposed proteins, transport molecules, intracellular and secreted proteins particular to each bacillus are shown. *Red crosses* representing lesions in *M. bovis* AF2122/97 in comparison to *M. tuberculosis* H37Rv while *asterisks* denote mutations that alter the function of the associated protein. Also, the regions of deletion ('RD'/'Rv', *yellow*) that are present in the associated tubercle bacilli and absent in the other, along with their associated genes and gene products are depicted (Brosch et al. 2002; Cole et al. 1998; Garnier et al. 2003; Gordon et al. 1999; Keating et al. 2005; Pang et al. 2012; Pym et al. 2002; Sohaskey and Modesti 2009; Stermann et al. 2003)

suggests that overcoming RskA regulation to drive high expression of MPB70 and MPB83 confers a yet to be elucidated selective advantage for both the bovine and oryx bacillus during infection (Said-Salim et al. 2006). Furthermore, the aforementioned transcriptional profiling studies highlight that certain biosynthetic genes involved in the production of major cell wall proteins sulfolipid-1 (SL-1) and phthiocerol dimycocerosate (PDIM) were differentially expressed between the two species with increases in SL-1 and PDIM related genes expressed to a higher level in *M. tuberculosis* and *M. bovis,* respectively (Golby et al. 2007; Rehren et al. 2007). Comparative global transcriptomic profiling using RNA-sequencing and global proteomic profiling using quantitative SWATH mass spectrometry of *M. tuberculosis* H37Rv and *M. bovis* AF2122/97 grown in vitro also supports the difference in expression of the SL-1 and PDIM biosynthetic genes between the two mycobacterial species (Malone and Gordon, unpublished) (Fig. 7.3). These results are in agreement with previous reports that *M. bovis* lacks trehalose-containing glycolipids in its cell wall, including SL-1, di- and poly-acyl trehalose (DAT, PAT) (Gonzalo Asensio et al. 2006; Goren 1970). The ratio and biosynthesis of PDIM and SL-1 in cell envelope composition of *M. tuberculosis* has been shown to be dependent on the flux of common intermediates such as methyl malonyl CoA within the cell; *M. tuberculosis* mutants in SL-1 synthesis display increased PDIM production (Jain et al. 2007). The consequence of SL-1 loss in the *M. bovis* cell wall with regards to virulence and adaptation to the host cell environment is unclear. SL-1 is one of the most abundant lipids in the mycobacterial outer membrane, it is unique to pathogenic mycobacteria, it is immunogenic and it has been implicated in the alteration of phagolysosome fusion (Brodin et al. 2010; Seeliger et al. 2012; Zhang et al. 1991). That being said, *M. tuberculosis* SL-1^{-}/PDIM^{+} strains do not show an attenuated phenotype in the murine host (Converse et al. 2003). PDIM has recently been reported to mask pathogen-associated molecular patterns (PAMPs) of *M. tuberculosis* during immune evasion in vivo (Arbues et al. 2014; Cambier et al. 2014) and it is

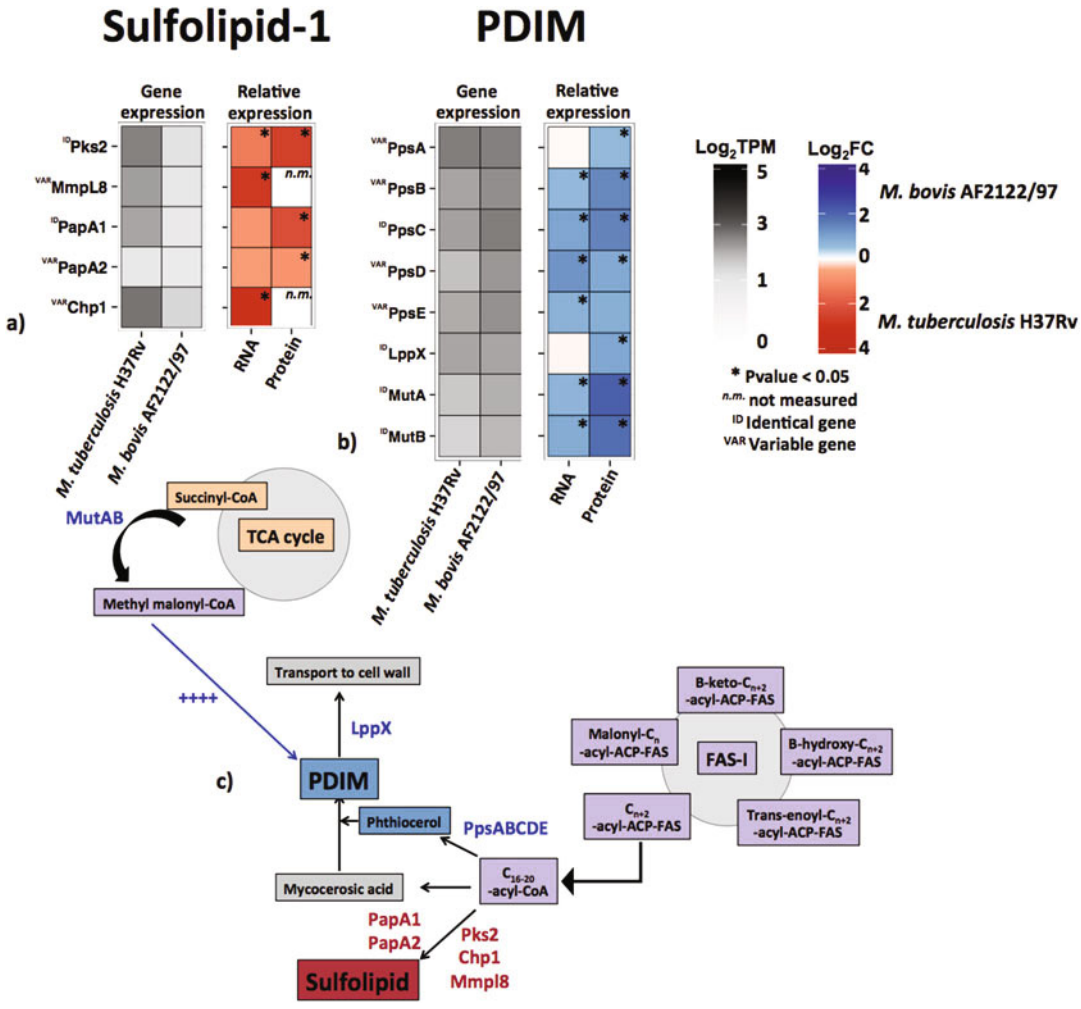

Fig. 7.3 Differential expression of the Sulfolipid-1 and phthiocerol dimycocerosate biosynthesis pathways in *M. tuberculosis* H37Rv and *M. bovis* AF2122/97. Representation of the expression of the (**a**) sulfolipid-1 (*SL-1*) and (**b**) phthiocerol dimycocerosate (*PDIM*) synthesis associated genes at the RNA and protein level in *M. tuberculosis* H37Rv and *M. bovis* AF2122/97. The expression of each gene ("Gene expression") is presented as Log_2TPM (transcripts per million) at the RNA level while the relative abundance of RNA or protein ("Relative expression") between the two species is presented as log_2FC (fold change). Those genes that change significantly at the RNA and protein level (corrected $p < 0.05$) are denoted (*). Genes within either pathway that are of equal length and are conserved 100% at the amino acid level ("ID") or are not 100% conserved at the amino acid level ("VAR") are noted while absence of protein measurements is also noted ("n.m"). (**c**) Diagrammatic overview of the SL-1 and PDIM biosynthesis pathways in *M. tuberculosis*. Those proteins coloured in *blue* represent the genes in (**i**) that are upregulated in *M. bovis* AF2122/97 in contrast to *M. tuberculosis* H37Rv and those in *red* represent the genes (**ii**) that are upregulated in *M. tuberculosis* H37Rv in contrast in *M. bovis* AF2122/97

required for the virulence of *M. tuberculosis* (Day et al. 2014). Whether or not the loss of trehalose containing glycolipids from the cell wall of *M. bovis* results in a seemingly compensatory production of PDIM is unclear and its overall effect on the virulence of the bovine bacilli remains to be demonstrated. Furthermore, the role of either SL-1 or PDIM in host specificity has yet to be determined and whether or not these differential lipid profiles hold under infection-

like conditions between the two species has yet to be elucidated.

The genetic basis for the loss of these glycolipids in *M. bovis* has been linked to the PhoPR two component signalling system. PhoPR signalling is implicated in the regulation of the biosynthesis of complex cell wall lipids and EsxA/ESAT-6 secretion in *M. tuberculosis* (Gonzalo Asensio et al. 2006, 2014). It has been suggested that mutation of PhoR in *M. bovis* in comparison to *M. tuberculosis* may explain the host specificity of these two species, leading to attenuated PhoPR signalling in *M. bovis,* alterations in the secretion of EsxA and SL-1/PDIM ratios in the cell wall (Gonzalo-Asensio et al. 2014); strikingly, the RD8 deletion of *M. bovis* restores EsxA/ESAT-6 secretion in a PhoPR-independent manner (Gonzalo-Asensio et al. 2014). Transfer of the PhoPR system-associated mutations from *M. bovis* into *M. tuberculosis* led to a loss of complex lipids (including SL-1), lower virulence and reduced secretion of EsxA/ESAT-6 (Gonzalo-Asensio et al. 2014). Reduced PhoPR signalling in *M. bovis* and its link to virulence in humans is supported by the report of a hypervirulent and multidrug-resistant strain of *M. bovis* recovered from a human patient that showed constitutive and high expression of PhoP as a result of an insertion sequence event upstream of the *phoP* gene (Soto et al. 2004). However, these mutations do not fully demonstrate that PhoPR is solely responsible for the human/animal host specificity exhibited by these species and thus, the mechanism behind bovine/human host tropism has yet to be fully determined.

Differential cell wall profiles between the human and bovine tubercle bacilli may also be a result of differential responses to extracellular stimuli. External environmental signals can be translated into adaptive responses in the mycobacterial cell by protein kinases, and *M. tuberculosis* encodes 11 serine/threonine protein kinases (STPKs) (*M. bovis* encodes 12) that are involved in the regulation of cell division, pathogenesis and cell wall biosynthesis (Av-Gay and Everett 2000; Chao et al. 2010). PknD is associated with the phosphorylation of the MmpL7 transporter protein of the PDIM biosynthetic pathway; *pknD* is split into two coding sequences in *M. bovis* AF2122/97 in contrast to *M. tuberculosis* H37Rv and therefore PknD-MmpL7 interactions may be perturbed in the bovine bacillus and result in an altered PDIM profile in contrast to *M. tuberculosis* (Perez et al. 2006). PknH has also been associated with PDIM synthesis; an *M. tuberculosis pknH* knockout mutant was attenuated for the production of PDIM, and the PDIM synthesis-associated protein PpsE was identified as a potential target for PknH (Gomez-Velasco et al. 2013). PknH's cognate partner is the transcriptional regulator EmbR, a protein associated with the regulation of the *embCAB* genes that are involved in the production of arabinan and resistance towards ethambutol while a third protein, EmbR2 is found to bind to EmbR and inhibit EmbR phosphorylation by PknH (Brossier et al. 2015; Sharma et al. 2006). EmbR2 is present in *M. bovis* and is absent in *M. tuberculosis* as part of RvD5 and therefore a difference in the regulation of PDIM synthesis between the two species may be a result of differential PknH signalling (Molle et al. 2008); however this remains to be shown.

7.5 Practical Implications

Genomic analyses across the animal-adapted members of the MTBC and comparison to *M. tuberculosis* have exposed the evolutionary steps that generated this unique group of pathogens. A key challenge is to now define how genetic variation across the MTBC translates into functional consequences for host adaptation. By achieving this linkage between genotype and phenotype the precise molecular basis of host preference across the complex can be revealed. This has implications not only for our broad understanding of how pathogens evolve to new hosts, but also conversely reveal what is it is that makes *M. tuberculosis* such a successful human pathogen. This integration of research on human and animal pathogens, in a true 'One Health' framework, holds substantial promise

for the development the next generation of tools and approaches to combat tuberculosis in both humans and animals.

References

Abernethy DA, Upton P, Higgins IM, McGrath G, Goodchild AV, Rolfe SJ, Broughan JM, Downs SH, Clifton-Hadley R, Menzies FD, De La Rua-Domenech R, Blissitt MJ, Duignan A, More SJ (2013) Bovine tuberculosis trends in the UK and the Republic of Ireland, 1995–2010. Vet Rec 172:312

Ahmad S, El-Shazly S, Mustafa AS, Al-Attiyah R (2005) The six mammalian cell entry proteins (Mce3A-F) encoded by the mce3 operon are expressed during in vitro growth of Mycobacterium tuberculosis. Scand J Immunol 62:16–24

Alexander KA, Laver PN, Michel AL, Williams M, Van Helden PD, Warren RM, Gey Van Pittius NC (2010) Novel Mycobacterium tuberculosis complex pathogen, M. mungi. Emerg Infect Dis 16:1296–1299

Alexander KA, Sanderson CE, Larsen MH, Robbe-Austerman S, Williams MC, Palmer MV (2016) Emerging tuberculosis pathogen hijacks social communication behavior in the group-living banded mongoose (*Mungos mungo*). MBio 7

Ameni G, Tadesse K, Hailu E, Deresse Y, Medhin G, Aseffa A, Hewinson G, Vordermeier M, Berg S (2013) Transmission of Mycobacterium tuberculosis between farmers and cattle in central Ethiopia. PLoS One 8:e76891

Aranaz A, Liebana E, Mateos A, Dominguez L, Vidal D, Domingo M, Gonzolez O, Rodriguez-Ferri EF, Bunschoten AE, Van Embden JD, Cousins D (1996) Spacer oligonucleotide typing of Mycobacterium bovis strains from cattle and other animals: a tool for studying epidemiology of tuberculosis. J Clin Microbiol 34:2734–2740

Aranaz A, Liebana E, Gomez-Mampaso E, Galan JC, Cousins D, Ortega A, Blazquez J, Baquero F, Mateos A, Suarez G, Dominguez L (1999) Mycobacterium tuberculosis subsp. caprae subsp. nov.: a taxonomic study of a new member of the Mycobacterium tuberculosis complex isolated from goats in Spain. Int J Syst Bacteriol 49(Pt 3):1263–1273

Aranaz A, Cousins D, Mateos A, Dominguez L (2003) Elevation of Mycobacterium tuberculosis subsp. caprae Aranaz et al. 1999 to species rank as Mycobacterium caprae comb. nov., sp. nov. Int J Syst Evol Microbiol 53:1785–1789

Arbues A, Lugo-Villarino G, Neyrolles O, Guilhot C, Astarie-Dequeker C (2014) Playing hide-and-seek with host macrophages through the use of mycobacterial cell envelope phthiocerol dimycocerosates and phenolic glycolipids. Front Cell Infect Microbiol 4:173

Av-Gay Y, Everett M (2000) The eukaryotic-like Ser/Thr protein kinases of Mycobacterium tuberculosis. Trends Microbiol 8:238–244

Barker C (2015) In: Department for Environment, F. A. R. A (eds) Bovine TB statistics: Great Britain. House of Commons Library, London

Berg S, Firdessa R, Habtamu M, Gadisa E, Mengistu A, Yamuah L, Ameni G, Vordermeier M, Robertson BD, Smith NH, Engers H, Young D, Hewinson RG, Aseffa A, Gordon SV (2009) The burden of mycobacterial disease in ethiopian cattle: implications for public health. PLoS One 4:e5068

Bigi F, Garcia-Pelayo MC, Nunez-Garcia J, Peralta A, Caimi KC, Golby P, Hinds J, Cataldi A, Gordon SV, Romano MI (2005) Identification of genetic markers for Mycobacterium pinnipedii through genome analysis. FEMS Microbiol Lett 248:147–152

Boardman WS, Shephard L, Bastian I, Globan M, Fyfe JA, Cousins DV, Machado A, Woolford L (2014) Mycobacterium pinnipedii tuberculosis in a free-ranging Australian fur seal (Arctocephalus pusillus doriferus) in South Australia. J Zoo Wildl Med 45: 970–972

Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, Forrest SA, Bryant JM, Harris SR, Schuenemann VJ, Campbell TJ, Majander K, Wilbur AK, Guichon RA, Wolfe Steadman DL, Cook DC, Niemann S, Behr MA, Zumarraga M, Bastida R, Huson D, Nieselt K, Young D, Parkhill J, Buikstra JE, Gagneux S, Stone AC, Krause J (2014) Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature 514:494–497

Brodin P, Eiglmeier K, Marmiesse M, Billault A, Garnier T, Niemann S, Cole ST, Brosch R (2002) Bacterial artificial chromosome-based comparative genomic analysis identifies Mycobacterium microti as a natural ESAT-6 deletion mutant. Infect Immun 70:5568–5578

Brodin P, Poquet Y, Levillain F, Peguillet I, Larrouy-Maumus G, Gilleron M, Ewann F, Christophe T, Fenistein D, Jang J, Jang MS, Park SJ, Rauzier J, Carralot JP, Shrimpton R, Genovesio A, Gonzalo-Asensio JA, Puzo G, Martin C, Brosch R, Stewart GR, Gicquel B, Neyrolles O (2010) High content phenotypic cell-based visual screen identifies Mycobacterium tuberculosis acyltrehalose-containing glycolipids involved in phagosome remodeling. PLoS Pathog 6:e1001100

Brosch R, Gordon SV, Billault A, Garnier T, Eiglmeier K, Soravito C, Barrell BG, Cole ST (1998) Use of a Mycobacterium tuberculosis H37Rv bacterial artificial chromosome library for genome mapping, sequencing, and comparative genomics. Infect Immun 66:2221–2229

Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, Garnier T, Gutierrez C, Hewinson G, Kremer K, Parsons LM, Pym AS, Samper S, Van Soolingen D, Cole ST (2002) A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci U S A 99:3684–3689

Brossier F, Sougakoff W, Bernard C, Petrou M, Adeyema K, Pham A, Amy De La Breteque D, Vallet M, Jarlier V, Sola C, Veziris N (2015) Molecular analysis of the embCAB Locus and embR gene involved in ethambutol resistance in clinical isolates of Mycobacterium

tuberculosis in France. Antimicrob Agents Chemother 59:4800–4808

Cadmus S, Palmer S, Okker M, Dale J, Gover K, Smith N, Jahans K, Hewinson RG, Gordon SV (2006) Molecular analysis of human and bovine tubercle bacilli from a local setting in Nigeria. J Clin Microbiol 44:29–34

Cambier CJ, Takaki KK, Larson RP, Hernandez RE, Tobin DM, Urdahl KB, Cosma CL, Ramakrishnan L (2014) Mycobacteria manipulate macrophage recruitment through coordinated use of membrane lipids. Nature 505:218–222

Cavanagh R, Begon M, Bennett M, Ergon T, Graham IM, De Haas PE, Hart CA, Koedam M, Kremer K, Lambin X, Roholl P, Soolingen DV, D. (2002) Mycobacterium microti infection (vole tuberculosis) in wild rodent populations. J Clin Microbiol 40:3281–3285

Chambers MA, Rogers F, Delahay RJ, Lesellier S, Ashford R, Dalley D, Gowtage S, Dave D, Palmer S, Brewer J, Crawshaw T, Clifton-Hadley R, Carter S, Cheeseman C, Hanks C, Murray A, Palphramand K, Pietravalle S, Smith GC, Tomlinson A, Walker NJ, Wilson GJ, Corner LA, Rushton SP, Shirley MD, Gettinby G, McDonald RA, Hewinson RG (2011) Bacillus Calmette-Guerin vaccination reduces the severity and progression of tuberculosis in badgers. Proc Biol Sci 278:1913–1920

Chao J, Wong D, Zheng X, Poirier V, Bach H, Hmama Z, Av-Gay Y (2010) Protein kinase and phosphatase signaling in Mycobacterium tuberculosis physiology and pathogenesis. Biochim Biophys Acta 1804:620–627

Charlet D, Mostowy S, Alexander D, Sit L, Wiker HG, Behr MA (2005) Reduced expression of antigenic proteins MPB70 and MPB83 in Mycobacterium bovis BCG strains due to a start codon mutation in sigK. Mol Microbiol 56:1302–1313

Chen Y, Chao Y, Deng Q, Liu T, Xiang J, Chen J, Zhou J, Zhan Z, Kuang Y, Cai H, Chen H, Guo A (2009) Potential challenges to the stop TB plan for humans in China; cattle maintain M. bovis and M. tuberculosis. Tuberculosis (Edinb) 89:95–100

Chiari M, Zanoni M, Alborali LG, Zanardi G, Avisani D, Tagliabue S, Gaffuri A, Pacciarini ML, Boniotti MB (2014) Isolation of Mycobacterium caprae (Lechtal genotype) from red deer (Cervus elaphus) in Italy. J Wildl Dis 50:330–333

Cohan FM (2002) What are bacterial species? Annu Rev Microbiol 56:457–487

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, Mclean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 393:537–544

Converse SE, Mougous JD, Leavell MD, Leary JA, Bertozzi CR, Cox JS (2003) MmpL8 is required for sulfolipid-1 biosynthesis and Mycobacterium tuberculosis virulence. Proc Natl Acad Sci U S A 100:6121–6126

Corner LA, Costello E, Lesellier S, O'meara D, Sleeman DP, Gormley E (2007) Experimental tuberculosis in the European badger (Meles meles) after endobronchial inoculation of Mycobacterium bovis: I. Pathology and bacteriology. Res Vet Sci 83:53–62

Corner LA, Costello E, Lesellier S, O'meara D, Gormley E (2008) Vaccination of European badgers (Meles meles) with BCG by the subcutaneous and mucosal routes induces protective immunity against endobronchial challenge with Mycobacterium bovis. Tuberculosis (Edinb) 88:601–609

Corner LA, Costello E, O'meara D, Lesellier S, Aldwell FE, Singh M, Hewinson RG, Chambers MA, Gormley E (2010) Oral vaccination of badgers (Meles meles) with BCG and protective immunity against endobronchial challenge with Mycobacterium bovis. Vaccine 28:6265–6272

Corner LA, O'meara D, Costello E, Lesellier S, Gormley E (2012) The distribution of Mycobacterium bovis infection in naturally infected badgers. Vet J 194:166–172

Coscolla M, Lewin A, Metzger S, Maetz-Rennsing K, Calvignac-Spencer S, Nitsche A, Dabrowski PW, Radonic A, Niemann S, Parkhill J, Couacy-Hymann E, Feldman J, Comas I, Boesch C, Gagneux S, Leendertz FH (2013) Novel Mycobacterium tuberculosis complex isolate from a wild chimpanzee. Emerg Infect Dis 19:969–976

Cousins DV, Williams SN, Reuter R, Forshaw D, Chadwick B, Coughran D, Collins P, Gales N (1993) Tuberculosis in wild seals and characterisation of the seal bacillus. Aust Vet J 70:92–97

Cousins DV, Peet RL, Gaynor WT, Williams SN, Gow BL (1994) Tuberculosis in imported hyrax (Procavia capensis) caused by an unusual variant belonging to the Mycobacterium tuberculosis complex. Vet Microbiol 42:135–145

Cousins DV, Bastida R, Cataldi A, Quse V, Redrobe S, Dow S, Duignan P, Murray A, Dupont C, Ahmed N, Collins DM, Butler WR, Dawson D, Rodriguez D, Loureiro J, Romano MI, Alito A, Zumarraga M, Bernardelli A (2003) Tuberculosis in seals caused by a novel member of the Mycobacterium tuberculosis complex: Mycobacterium pinnipedii sp. nov. Int J Syst Evol Microbiol 53:1305–1314

Cvetnic Z, Katalinic-Jankovic V, Sostaric B, Spicic S, Obrovac M, Marjanovic S, Benic M, Kirin BK, Vickovic I (2007) Mycobacterium caprae in cattle and humans in Croatia. Int J Tuberc Lung Dis 11:652–658

Day TA, Mittler JE, Nixon MR, Thompson C, Miner MD, Hickey MJ, Liao RP, Pang JM, Shayakhmetov DM, Sherman DR (2014) Mycobacterium tuberculosis strains lacking surface lipid phthiocerol dimycocerosate are susceptible to killing by an early innate host response. Infect Immun 82:5214–5222

De Garine-Wichatitsky M, Caron A, Kock R, Tschopp R, Munyeme M, Hofmeyr M, Michel A (2013) A review of bovine tuberculosis at the wildlife-livestock-human interface in sub-Saharan Africa. Epidemiol Infect 141:1342–1356

De Jong BC, Hill PC, Aiken A, Awine T, Antonio M, Adetifa IM, Jackson-Sillah DJ, Fox A, Deriemer K, Gagneux S, Borgdorff MW, Mcadam KP, Corrah T, Small PM, Adegbola RA (2008) Progression to active tuberculosis, but not transmission, varies by Mycobacterium tuberculosis lineage in The Gambia. J Infect Dis 198:1037–1043

De La Fuente J, Diez-Delgado I, Contreras M, Vicente J, Cabezas-Cruz A, Manrique M, Tobes R, Lopez V, Romero B, Dominguez L, Garrido JM, Juste R Gortazar C (2015) Complete genome sequences of field isolates of Mycobacterium bovis and Mycobacterium caprae. Genome Announc 3

Dippenaar A, Parsons SD, Sampson SL, Van Der Merwe RG, Drewe JA, Abdallah AM, Siame KK, Gey Van Pittius NC, Van Helden PD, Pain A, Warren RM (2015) Whole genome sequence analysis of Mycobacterium suricattae. Tuberculosis (Edinb) 95:682–688

Drewe JA, O'connor HM, Weber N, McDonald RA, Delahay RJ (2013) Patterns of direct and indirect contact between cattle and badgers naturally infected with tuberculosis. Epidemiol Infect 141:1467–1475

Du Y, Qi Y, Yu L, Lin J, Liu S, Ni H, Pang H, Liu H, Si W, Zhao H, Wang C (2011) Molecular characterization of Mycobacterium tuberculosis complex (MTBC) isolated from cattle in northeast and northwest China. Res Vet Sci 90:385–391

Duarte EL, Domingos M, Amado A, Botelho A (2008) Spoligotype diversity of Mycobacterium bovis and Mycobacterium caprae animal isolates. Vet Microbiol 130:415–421

Erler W, Martin G, Sachse K, Naumann L, Kahlau D, Beer J, Bartos M, Nagy G, Cvetnic Z, Zolnir-Dovc M, Pavlik I (2004) Molecular fingerprinting of Mycobacterium bovis subsp. caprae isolates from central Europe. J Clin Microbiol 42:2234–2238

Fink M, Schleicher C, Gonano M, Prodinger WM, Pacciarini M, Glawischnig W, Ryser-Degiorgis MP, Walzer C, Stalder GL, Lombardo D, Schobesberger H, Winter P, Buttner M (2015) Red deer as maintenance host for bovine tuberculosis, Alpine region. Emerg Infect Dis 21:464–467

Fitzgerald SD, Kaneene JB (2013) Wildlife reservoirs of bovine tuberculosis worldwide: hosts, pathology, surveillance, and control. Vet Pathol 50:488–499

Forshaw D, Phelps GR (1991) Tuberculosis in a captive colony of pinnipeds. J Wildl Dis 27:288–295

Frota CC, Hunt DM, Buxton RS, Rickman L, Hinds J, Kremer K, Van Soolingen D, Colston MJ (2004) Genome structure in the vole bacillus, Mycobacterium microti, a member of the Mycobacterium tuberculosis complex with a low virulence for humans. Microbiology 150:1519–1527

Furphy C, Costello E, Murphy D, Corner LA, Gormley E (2012) DNA typing of Mycobacterium bovis isolates from Badgers (Meles meles) culled from areas in Ireland with different levels of tuberculosis prevalence. Vet Med Int 2012:742478

Garnier T, Eiglmeier K, Camus JC, Medina N, Mansoor H, Pryor M, Duthoy S, Grondin S, Lacroix C, Monsempe C, Simon S, Harris B, Atkin R, Doggett J, Mayes R, Keating L, Wheeler PR, Parkhill J, Barrell BG, Cole ST, Gordon SV, Hewinson RG (2003) The complete genome sequence of Mycobacterium bovis. Proc Natl Acad Sci U S A 100:7877–7882

Gey Van Pittius NC, Perrett KD, Michel AL, Keet DF, Hlokwe T, Streicher EM, Warren RM, Van Helden PD (2012) Infection of African buffalo (Syncerus caffer) by oryx bacillus, a rare member of the antelope clade of the Mycobacterium tuberculosis complex. J Wildl Dis 48:849–857

Golby P, Hatch KA, Bacon J, Cooney R, Riley P, Allnutt J, Hinds J, Nunez J, Marsh PD, Hewinson RG, Gordon SV (2007) Comparative transcriptomics reveals key gene expression differences between the human and bovine pathogens of the Mycobacterium tuberculosis complex. Microbiology 153:3323–3336

Golby P, Nunez J, Witney A, Hinds J, Quail MA, Bentley S, Harris S, Smith N, Hewinson RG, Gordon SV (2013) Genome-level analyses of Mycobacterium bovis lineages reveal the role of SNPs and antisense transcription in differential gene expression. BMC Genomics 14:710

Gomez-Velasco A, Bach H, Rana AK, Cox LR, Bhatt A, Besra GS, Av-Gay Y (2013) Disruption of the serine/threonine protein kinase H affects phthiocerol dimycocerosates synthesis in Mycobacterium tuberculosis. Microbiology 159:726–736

Gonzalo Asensio J, Maia C, Ferrer NL, Barilone N, Laval F, Soto CY, Winter N, Daffe M, Gicquel B, Martin C, Jackson M (2006) The virulence-associated two-component PhoP-PhoR system controls the biosynthesis of polyketide-derived lipids in Mycobacterium tuberculosis. J Biol Chem 281:1313–1316

Gonzalo-Asensio J, Malaga W, Pawlik A, Astarie-Dequeker C, Passemar C, Moreau F, Laval F, Daffe M, Martin C, Brosch R, Guilhot C (2014) Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence regulator. Proc Natl Acad Sci U S A 111:11491–11496

Gordon SV, Brosch R, Billault A, Garnier T, Eiglmeier K, Cole ST (1999) Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. Mol Microbiol 32: 643–655

Goren MB (1970) Sulfolipid I of Mycobacterium tuberculosis, strain H37Rv. I. Purification and properties. Biochim Biophys Acta 210:116–126

Gormley E, Corner LA (2013) Control strategies for wildlife tuberculosis in Ireland. Transbound Emerg Dis 60(Suppl 1):128–135

Griffin JM, Williams DH, Kelly GE, Clegg TA, O'Boyle I, Collins JD, More SJ (2005) The impact of badger removal on the control of tuberculosis in cattle herds in Ireland. Prev Vet Med 67:237–266

Hansen N, Seiler C, Rumpf J, Kraft P, Dlaske H, Abele-Horn M, Muellges W (2012) Human tuberculous meningitis caused by Mycobacterium caprae. Case Rep Neurol 4:54–60

Hart PD, Sutherland I (1977) BCG and vole bacillus vaccines in the prevention of tuberculosis in adolescence and early adult life. Br Med J 2:293–295

Hauer A, De Cruz K, Cochard T, Godreuil S, Karoui C, Henault S, Bulach T, Banuls AL, Biet F, Boschiroli ML (2015) Genetic evolution of Mycobacterium bovis causing tuberculosis in livestock and wildlife in France since 1978. PLoS One 10:e0117103

Heath D (2013) Response to FOI request about various bovine TB costs from 2008 to 2013. In: Department for Environment, F. A. R. A. (ed)

Hewinson RG, Michell SL, Russell WP, Mcadam RA, Jacobs WR Jr (1996) Molecular characterization of MPT83: a seroreactive antigen of Mycobacterium tuberculosis with homology to MPT70. Scand J Immunol 43:490–499

Huard RC, Fabre M, De Haas P, Lazzarini LC, Van Soolingen D, Cousins D, Ho JL (2006) Novel genetic polymorphisms that further delineate the phylogeny of the Mycobacterium tuberculosis complex. J Bacteriol 188:4271–4287

Jain M, Petzold CJ, Schelle MW, Leavell MD, Mougous JD, Bertozzi CR, Leary JA, Cox JS (2007) Lipidomics reveals control of Mycobacterium tuberculosis virulence lipids via metabolic coupling. Proc Natl Acad Sci U S A 104:5133–5138

Karlson AG, Lessel EF (1970) Mycobacterium bovis nom. nov. Int J Syst Evol Microbiol 20:273–282

Kazoora HB, Majalija S, Kiwanuka N, Kaneene JB (2016) Knowledge, attitudes and practices regarding risk to human infection due to Mycobacterium bovis among cattle farming communities in western Uganda. Zoonoses Public Health 63:616–623

Keating LA, Wheeler PR, Mansoor H, Inwald JK, Dale J, Hewinson RG, Gordon SV (2005) The pyruvate requirement of some members of the Mycobacterium tuberculosis complex is due to an inactive pyruvate kinase: implications for in vivo growth. Mol Microbiol 56:163–174

Kelly GE, Condon J, More SJ, Dolan L, Higgins I, Eves J (2008) A long-term observational study of the impact of badger removal on herd restrictions due to bovine TB in the Irish midlands during 1989–2004. Epidemiol Infect 136:1362–1373

Khattak I, Mushtaq MH, Ahmad MU, Khan MS, Haider J (2016) Zoonotic tuberculosis in occupationally exposed groups in Pakistan. Occup Med (Lond) 66:371–376

Kiers A, Klarenbeek A, Mendelts B, Van Soolingen D, Koeter G (2008) Transmission of Mycobacterium pinnipedii to humans in a zoo with marine mammals. Int J Tuberc Lung Dis 12:1469–1473

Kubica T, Rusch-Gerdes S, Niemann S (2003) Mycobacterium bovis subsp. caprae caused one-third of human M. bovis-associated tuberculosis cases reported in Germany between 1999 and 2001. J Clin Microbiol 41:3070–3077

Lantos A, Niemann S, Mezosi L, Sos E, Erdelyi K, David S, Parsons LM, Kubica T, Rusch-Gerdes S, Somoskovi A (2003) Pulmonary tuberculosis due to Mycobacterium bovis subsp. caprae in captive Siberian tiger. Emerg Infect Dis 9:1462–1464

Lees VW, Copeland S, Rousseau P (2003) Bovine tuberculosis in elk (Cervus elaphus manitobensis) near Riding Mountain National Park, Manitoba, from 1992 to 2002. Can Vet J 44:830–831

Levy-Frebault VV, Portaels F (1992) Proposed minimal standards for the genus Mycobacterium and for description of new slowly growing Mycobacterium species. Int J Syst Bacteriol 42:315–323

Lewis KN, Liao R, Guinn KM, Hickey MJ, Smith S, Behr MA, Sherman DR (2003) Deletion of RD1 from Mycobacterium tuberculosis mimics bacille Calmette-Guerin attenuation. J Infect Dis 187:117–123

Lomme JR, Thoen CO, Himes EM, Vinson JW, King RE (1976) Mycobacterium tuberculosis infection in two East African oryxes. J Am Vet Med Assoc 169:912–914

Mableson HE, Okello A, Picozzi K, Welburn SC (2014) Neglected zoonotic diseases-the long and winding road to advocacy. PLoS Negl Trop Dis 8:e2800

Magee DA, Conlon KM, Nalpas NC, Browne JA, Pirson C, Healy C, Mcloughlin KE, Chen J, Vordermeier HM, Gormley E, Machugh DE, Gordon SV (2014) Innate cytokine profiling of bovine alveolar macrophages reveals commonalities and divergence in the response to Mycobacterium bovis and Mycobacterium tuberculosis infection. Tuberculosis (Edinb) 94:441–450

Malama S, Muma J, Munyeme M, Mbulo G, Muwonge A, Shamputa IC, Djonne B, Godfroid J, Johansen TB (2014) Isolation and molecular characterization of Mycobacterium tuberculosis from humans and cattle in Namwala District, Zambia. EcoHealth 11:564–570

Marmiesse M, Brodin P, Buchrieser C, Gutierrez C, Simoes N, Vincent V, Glaser P, Cole ST, Brosch R (2004) Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the Mycobacterium tuberculosis complex. Microbiology 150:483–496

Miller M, Michel A, Van Helden P, Buss P (2016) Tuberculosis in rhinoceros: an underrecognized threat? Transbound Emerg Dis

Molle V, Reynolds RC, Alderwick LJ, Besra GS, Cozzone AJ, Futterer K, Kremer L (2008) EmbR2, a structural homologue of EmbR, inhibits the Mycobacterium tuberculosis kinase/substrate pair PknH/EmbR. Biochem J 410:309–317

More SJ, Good M (2006) The tuberculosis eradication programme in Ireland: a review of scientific and policy advances since 1988. Vet Microbiol 112:239–251

More SJ, Good M (2015) Understanding and managing bTB risk: perspectives from Ireland. Vet Microbiol 176:209–218

Moser I, Prodinger WM, Hotzel H, Greenwald R, Lyashchenko KP, Bakker D, Gomis D, Seidler T, Ellenberger C, Hetzel U, Wuennemann K, Moisson P (2008) Mycobacterium pinnipedii: transmission from South American sea lion (Otaria byronia) to Bactrian camel (Camelus bactrianus bactrianus) and Malayan tapirs (Tapirus indicus). Vet Microbiol 127:399–406

Mostowy S, Cousins D, Behr MA (2004) Genomic interrogation of the dassie bacillus reveals it as a unique RD1 mutant within the Mycobacterium tuberculosis complex. J Bacteriol 186:104–109

Muller B, Durr S, Alonso S, Hattendorf J, Laisse CJ, Parsons SD, Van Helden PD, Zinsstag J (2013) Zoonotic Mycobacterium bovis-induced tuberculosis in humans. Emerg Infect Dis 19:899–908

Munoz Mendoza M, Juan L, Menendez S, Ocampo A, Mourelo J, Saez JL, Dominguez L, Gortazar C, Garcia Marin JF, Balseiro A (2012) Tuberculosis due to Mycobacterium bovis and Mycobacterium caprae in sheep. Vet J 191:267–269

Murphy D, Costello E, Aldwell FE, Lesellier S, Chambers MA, Fitzsimons T, Corner LA, Gormley E (2014) Oral vaccination of badgers (Meles meles) against tuberculosis: comparison of the protection generated by BCG vaccine strains Pasteur and Danish. Vet J 200:362–367

Niemann S, Richter E, Dalugge-Tamm H, Schlesinger H, Graupner D, Konigstein B, Gurath G, Greinert U, Rusch-Gerdes S (2000) Two cases of Mycobacterium microti derived tuberculosis in HIV-negative immunocompetent patients. Emerg Infect Dis 6:539–542

Niemann S, Richter E, Rusch-Gerdes S (2002) Biochemical and genetic evidence for the transfer of Mycobacterium tuberculosis subsp. caprae Aranaz et al. 1999 to the species Mycobacterium bovis Karlson and Lessel 1970 (approved lists 1980) as Mycobacterium bovis subsp. caprae comb. nov. Int J Syst Evol Microbiol 52:433–436

Olea-Popelka FJ, Flynn O, Costello E, Mcgrath G, Collins JD, O'Keeffe J, Kelton DF, Berke O, Martin SW (2005) Spatial relationship between Mycobacterium bovis strains in cattle and badgers in four areas in Ireland. Prev Vet Med 71:57–70

Olea-Popelka FJ, Fitzgerald P, White P, Mcgrath G, Collins JD, O'Keeffe J, Kelton DF, Berke O, More S, Martin SW (2009) Targeted badger removal and the subsequent risk of bovine tuberculosis in cattle herds in county Laois, Ireland. Prev Vet Med 88:178–184

Pang JM, Layre E, Sweet L, Sherrid A, Moody DB, Ojha A, Sherman DR (2012) The polyketide Pks1 contributes to biofilm formation in Mycobacterium tuberculosis. J Bacteriol 194:715–721

Parra A, Fernandez-llario P, Tato A, Larrasa J, Garcia A, Alonso JM, Hermoso de mendoza M, Hermoso de mendoza J (2003) Epidemiology of Mycobacterium bovis infections of pigs and wild boars using a molecular approach. Vet Microbiol 97:123–133

Parsons SD, Drewe JA, Gey Van Pittius NC, Warren RM, Van Helden PD (2013) Novel cause of tuberculosis in meerkats, South Africa. Emerg Infect Dis 19:2004–2007

Pate M, Svara T, Gombac M, Paller T, Zolnir-Dovc M, Emersic I, Prodinger WM, Bartos M, Zdovc I, Krt B, Pavlik I, Cvetnic Z, Pogacnik M, Ocepek M (2006) Outbreak of tuberculosis caused by Mycobacterium caprae in a zoological garden. J Vet Med B Infect Dis Vet Public Health 53:387–392

Perez J, Garcia R, Bach H, De Waard JH, Jacobs WR Jr, Av-Gay Y, Bubis J, Takiff HE (2006) Mycobacterium tuberculosis transporter MmpL7 is a potential substrate for kinase PknD. Biochem Biophys Res Commun 348:6–12

Perez-Jacoiste Asin MA, Fernandez-Ruiz M, Lopez-Medrano F, Lumbreras C, Tejido A, San Juan R, Arrebola-Pajares A, Lizasoain M, Prieto S, Aguado JM (2014) Bacillus Calmette-Guerin (BCG) infection following intravesical BCG administration as adjunctive therapy for bladder cancer: incidence, risk factors, and outcome in a single-institution series and review of the literature. Medicine (Baltimore) 93:236–254

Pesciaroli M, Alvarez J, Boniotti MB, Cagiola M, Di Marco V, Marianelli C, Pacciarini M, Pasquali P (2014) Tuberculosis in domestic animal species. Res Vet Sci 97(Suppl):S78–S85

Piercy J, Werling D, Coffey TJ (2007) Differential responses of bovine macrophages to infection with bovine-specific and non-bovine specific mycobacteria. Tuberculosis (Edinb) 87:415–420

Prodinger WM, Eigentler A, Allerberger F, Schonbauer M, Glawischnig W (2002) Infection of red deer, cattle, and humans with Mycobacterium bovis subsp. caprae in western Austria. J Clin Microbiol 40: 2270–2272

Prodinger WM, Indra A, Koksalan OK, Kilicaslan Z, Richter E (2014) Mycobacterium caprae infection in humans. Expert Rev Anti-Infect Ther 12:1501–1513

Pym AS, Brodin P, Brosch R, Huerre M, Cole ST (2002) Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines Mycobacterium bovis BCG and Mycobacterium microti. Mol Microbiol 46:709–717

Rehren G, Walters S, Fontan P, Smith I, Zarraga AM (2007) Differential gene expression between Mycobacterium bovis and Mycobacterium tuberculosis. Tuberculosis (Edinb) 87:347–359

Rodriguez E, Sanchez LP, Perez S, Herrera L, Jimenez MS, Samper S, Iglesias MJ (2009) Human tuberculosis due to Mycobacterium bovis and M. caprae in Spain, 2004–2007. Int J Tuberc Lung Dis 13: 1536–1541

Rodriguez S, Bezos J, Romero B, De Juan L, Alvarez J, Castellanos E, Moya N, Lozano F, Javed MT, Saez-Llorente JL, Liebana E, Mateos A, Dominguez L, Aranaz A, Spanish Network on, S. & Monitoring of Animal, T (2011) Mycobacterium caprae infection in livestock and wildlife, Spain. Emerg Infect Dis 17:532–535

Said-Salim B, Mostowy S, Kristof AS, Behr MA (2006) Mutations in Mycobacterium tuberculosis Rv0444c, the gene encoding anti-SigK, explain high level expression of MPB70 and MPB83 in Mycobacterium bovis. Mol Microbiol 62:1251–1263

Schnappinger D, Ehrt S, Voskuil MI, Liu Y, Mangan JA, Monahan IM, Dolganov G, Efron B, Butcher PD, Nathan C, Schoolnik GK (2003) Transcriptional Adaptation of Mycobacterium tuberculosis within Macrophages: insights into the phagosomal environment. J Exp Med 198:693–704

Seeliger JC, Holsclaw CM, Schelle MW, Botyanszki Z, Gilmore SA, Tully SE, Niederweis M, Cravatt BF, Leary JA, Bertozzi CR (2012) Elucidation and chemical modulation of sulfolipid-1 biosynthesis in Mycobacterium tuberculosis. J Biol Chem 287:7990–8000

Senaratne RH, Sidders B, Sequeira P, Saunders G, Dunphy K, Marjanovic O, Reader JR, Lima P, Chan S, Kendall S, Mcfadden J, Riley LW (2008) Mycobacterium tuberculosis strains disrupted in mce3 and mce4 operons are attenuated in mice. J Med Microbiol 57:164–170

Sharma K, Gupta M, Pathak M, Gupta N, Koul A, Sarangi S, Baweja R, Singh Y (2006) Transcriptional control of the mycobacterial embCAB operon by PknH through a regulatory protein, EmbR, in vivo. J Bacteriol 188:2936–2944

Sheridan M (2011) Progress in tuberculosis eradication in Ireland. Vet Microbiol 151:160–169

Sinsimer D, Huet G, Manca C, Tsenova L, Koo MS, Kurepina N, Kana B, Mathema B, Marras SA, Kreiswirth BN, Guilhot C, Kaplan G (2008) The phenolic glycolipid of Mycobacterium tuberculosis differentially modulates the early host cytokine response but does not in itself confer hypervirulence. Infect Immun 76: 3027–3036

Smith T (1898) A comparative study of bovine tubercle bacilli and of human bacilli from sputum. J Exp Med 3:451–511

Smith N (1960) The 'Dassie' bacillus. Tubercle 41: 203–212

Smith N (1965) Animal pathogenicity of the 'Dassie Bacillus'. Tubercle 46:58–64

Smith NH, Kremer K, Inwald J, Dale J, Driscoll JR, Gordon SV, Van Soolingen D, Hewinson RG, Smith JM (2006) Ecotypes of the Mycobacterium tuberculosis complex. J Theor Biol 239:220–225

Smith NH, Crawshaw T, Parry J, Birtles RJ (2009a) Mycobacterium microti: more diverse than previously thought. J Clin Microbiol 47:2551–2559

Smith NH, Hewinson RG, Kremer K, Brosch R, Gordon SV (2009b) Myths and misconceptions: the origin and evolution of Mycobacterium tuberculosis. Nat Rev Microbiol 7:537–544

Smith GC, Mcdonald RA, Wilkinson D (2012a) Comparing badger (Meles meles) management strategies for reducing tuberculosis incidence in cattle. PLoS One 7:e39250

Smith SE, Showers-Corneli P, Dardenne CN, Harpending HH, Martin DP, Beiko RG (2012b) Comparative genomic and phylogenetic approaches to characterize the role of genetic recombination in mycobacterial evolution. PLoS One 7:e50070

Sohaskey CD, Modesti L (2009) Differences in nitrate reduction between Mycobacterium tuberculosis and Mycobacterium bovis are due to differential expression of both narGHJI and narK2. FEMS Microbiol Lett 290:129–134

Soto CY, Menendez MC, Perez E, Samper S, Gomez AB, Garcia MJ, Martin C (2004) IS6110 mediates increased transcription of the phoP virulence gene in a multidrug-resistant clinical isolate responsible for tuberculosis outbreaks. J Clin Microbiol 42:212–219

Srivastava K, Chauhan DS, Gupta P, Singh HB, Sharma VD, Yadav VS, Sreekumaran, Thakral SS, Dharamdheeran JS, Nigam P, Prasad HK, Katoch VM (2008) Isolation of Mycobacterium bovis & M. tuberculosis from cattle of some farms in north India – possible relevance in human health. Indian J Med Res 128:26–31

Stermann M, Bohrssen A, Diephaus C, Maass S, Bange FC (2003) Polymorphic nucleotide within the promoter of nitrate reductase (NarGHJI) is specific for Mycobacterium tuberculosis. J Clin Microbiol 41: 3252–3259

Sula L, Radkovsky I (1976) Protective effects of M. microti vaccine against tuberculosis. J Hyg Epidemiol Microbiol Immunol 20:1–6

Thapa J, Nakajima C, Maharjan B, Poudell A, Suzuki Y (2015) Molecular characterization of Mycobacterium orygis isolates from wild animals of Nepal. Jpn J Vet Res 63:151–158

Thompson PJ, Cousins DV, Gow BL, Collins DM, Williamson BH, Dagnia HT (1993) Seals, seal trainers, and mycobacterial infection. Am Rev Respir Dis 147:164–167

Van Ingen J, Rahim Z, Mulder A, Boeree MJ, Simeone R, Brosch R, Van Soolingen D (2012) Characterization of Mycobacterium orygis as M. tuberculosis complex subspecies. Emerg Infect Dis 18:653–655

Van Soolingen D, De Haas PE, Haagsma J, Eger T, Hermans PW, Ritacco V, Alito A, Van Embden JD (1994) Use of various genetic markers in differentiation of Mycobacterium bovis strains from animals and humans and for studying epidemiology of bovine tuberculosis. J Clin Microbiol 32:2425–2433

Van Soolingen D, Van Der Zanden AG, De Haas PE, Noordhoek GT, Kiers A, Foudraine NA, Portaels F, Kolk AH, Kremer K, Van Embden JD (1998) Diagnosis of Mycobacterium microti infections among humans by using novel genetic markers. J Clin Microbiol 36:1840–1845

Waters WR, Palmer MV (2015) Mycobacterium bovis Infection of cattle and white-tailed deer: translational research of relevance to human tuberculosis. ILAR J 56:26–43

Wells AQ (1937) Tuberculosis in wild voles. Lancet 229:1221

Wells AQ (1946) The murine type of tubercle bacilli (The vole acid-fast bacillus). MRC Spec Rep Ser Med Res Counc London

Whelan AO, Coad M, Cockle PJ, Hewinson G, Vordermeier M, Gordon SV (2010) Revisiting host preference in the Mycobacterium tuberculosis complex: experimental infection shows M. tuberculosis H37Rv to be avirulent in cattle. PLoS One 5:e8527

WHO (2005) The control of neglected zoonotic diseases. In: WHO/SDE/FOS/2006 (ed)

WHO (2015) Global tuberculosis report [Online]. Available: http://www.who.int/tb/publications/global_report/en/. Accessed March 2016

Widdison S, Watson M, Piercy J, Howard C, Coffey TJ (2008) Granulocyte chemotactic properties of M. tuberculosis versus M. bovis-infected bovine alveolar macrophages. Mol Immunol 45:740–749

Wiker HG (2009) MPB70 and MPB83 – major antigens of Mycobacterium bovis. Scand J Immunol 69:492–499

Winder CL, Gordon SV, Dale J, Hewinson RG, Goodacre R (2006) Metabolic fingerprints of Mycobacterium bovis cluster with molecular type: implications for genotype-phenotype links. Microbiology 152: 2757–2765

Woods R, Cousins DV, Kirkwood R, Obendorf DL (1995) Tuberculosis in a wild Australian fur seal (Arctocephalus pusillus doriferus) from Tasmania. J Wildl Dis 31:83–86

Wright DM, Reid N, Ian Montgomery W, Allen AR, Skuce RA, Kao RR (2015) Herd-level bovine tuberculosis risk factors: assessing the role of low-level badger population disturbance. Sci Rep 5:13062

Zhang L, English D, Andersen BR (1991) Activation of human neutrophils by Mycobacterium tuberculosis-derived sulfolipid-1. J Immunol 146:2730–2736

Zhu L, Zhong J, Jia X, Liu G, Kang Y, Dong M, Zhang X, Li Q, Yue L, Li C, Fu J, Xiao J, Yan J, Zhang B, Lei M, Chen S, Lv L, Zhu B, Huang H, Chen F (2016) Precision methylome characterization of Mycobacterium tuberculosis complex (MTBC) using PacBio single-molecule real-time (SMRT) technology. Nucleic Acids Res 44:730–743

Evolution and Strain Variation in BCG

8

Abdallah M. Abdallah and Marcel A. Behr

Abstract

BCG vaccines were derived by in vitro passage, during the years 1908–1921, at the Pasteur Institute of Lille. Following the distribution of stocks of BCG to vaccine production laboratories around the world, it was only a few decades before different BCG producers recognized that there were variants of BCG, likely due to different passaging conditions in the different laboratories. This ultimately led to the lyophilization of stable BCG products in the 1950s and 1960s, but not before considerable evolution of the different BCG strains had taken place. The application of contemporary research methodologies has now revealed genomic, transcriptomic and proteomic differences between BCG strains. These molecular differences in part account for phenotypic differences in vitro between BCG strains, such as their variable secretion of antigenic proteins. Yet, the relevance of BCG variability for immunization policy remains elusive. In this chapter we present an overview of what is known about BCG evolution and its resulting strain variability, and provide some speculation as to the potential relevance for a vaccine given to over 100 million newborns each year.

Keywords

Tuberculosis • BCG • Vaccine • Protection • Strain

A.M. Abdallah (✉)
Bioscience Core Laboratory, King Abdullah University of Science and Technology, Thuwal, Jeddah, Kingdom of Saudi Arabia
e-mail: Abdallah.abdallah@kaust.edu.sa

M.A. Behr
Department of Medicine, McGill University Health Centre, Montreal, QC, Canada
e-mail: marcel.behr@mcgill.ca

8.1 Introduction

Humans have been infected with *M. tuberculosis* (Mtb) for millennia. Mtb, the intracellular pathogen that causes tuberculosis (TB), was discovered in 1882 by Robert Koch and is responsible for more human deaths than any other

S. Gagneux (ed.), *Strain Variation in the* Mycobacterium tuberculosis *Complex: Its Role in Biology, Epidemiology and Control*, Advances in Experimental Medicine and Biology 1019,
DOI 10.1007/978-3-319-64371-7_8

single pathogen today (Kaufmann et al. 2010; Kaufmann and Winau 2005; Ottenhoff 2009). *Mycobacterium bovis* Bacillus Calmette-Guérin (BCG), an attenuated strain of *M. bovis*, has been used as a prophylactic measure against TB for nearly a century to immunize over four billion individuals in more than 180 countries or territories (McShane 2011; Ottenhoff and Kaufmann 2012; Zwerling et al. 2011). Over 90% of children worldwide are vaccinated with BCG and more than 120 million doses of BCG are administered annually, making it the world's most widely used vaccine. Given this large denominator, it is possible to infer that there is a remarkable safety record. Unfortunately, despite the large numbers of individuals who have received BCG, in both programmatic settings and in clinical trials, the true efficacy of BCG has been difficult to understand due to many experimental variables (Griffin et al. 2001).

The parent BCG vaccine strain was derived from a virulent strain of *M. bovis,* a primary cause of TB in cattle, which is member of the *M. tuberculosis* complex that mainly affects wild and domesticated mammals (see Chap. 7) (Frothingham et al. 1994; Imaeda et al. 1985). Case-control studies have shown that BCG is associated with protection against childhood disseminated TB including meningitis and miliary TB (Colditz et al. 1995; Trunz et al. 2006). However, its efficacy against pulmonary TB in adults, as measured in randomized controlled trials, has varied from no efficacy at all to as high as 80% protection (Fine 1995). It was noted as far back as 1967 that the protective efficacy of BCG against TB varies substantially between studies, showing an average risk reduction of pulmonary TB of 50%, and of disseminated and meningitic forms of this disease by 70–80% (Brewer 2000; Colditz et al. 1995; Trunz et al. 2006). There are several reasons for this variation in BCG vaccine effectiveness, including differences in host populations, differences in their exposure to environmental mycobacteria and methodologic differences between the studies. In this chapter, we focus on differences between strains of BCG vaccine in use, and consider the possibility that the heterogeneity among these strains has in part contributed to the heterogeneous results from clinical trials (Behr 2002). Nowadays, several BCG strains are currently used worldwide as vaccines. Early clinical trials in indigenous groups in North America, infants in Chicago and school children in the UK demonstrated the efficacy of the vaccine and led to its distribution to several countries for worldwide application (Aronson et al. 1958; Rosenthal et al. 1961). Next to BCG, no other vaccines are available for protecting from TB, and of the many new candidates in the pipeline none is close to market use. Therefore, it is critical to examine the product known as BCG and to ask whether evolution and strain variability in BCG may have a role in determining the capacity to provide protective immunity against TB.

8.2 Historical Aspects of BCG

BCG is named for Albert Calmette and Camille Guérin (Bacillus of Calmette and Guérin) who derived the original BCG vaccine strain from an isolate of *M. bovis* at the Pasteur Institute of Lille, France in the early part of the twentieth century. Since 1900, Calmette and Guérin began their research on the *M. bovis* strain, which had been isolated from the milk of a cow suffering from tuberculous mastitis by Nocard at the Pasteur Institute of Paris, France in 1902. This isolated strain of *M. bovis* was used by Albert Calmette and Camille Guérin to study the pathogenesis of bovine TB (Grange et al. 1983; Oettinger et al. 1999). In early studies, Calmette reported that oral inoculation of *M. bovis* resulted in pulmonary TB, through lymphatic spread from the mesenteric to the mediastinal lymph nodes. This finding presented a public health challenge, as, at the time, the predominant route of acquisition of pulmonary TB was through respiratory aerosols. To produce inocula for these experiments, Calmette and Guérin cultivated these tubercle bacilli on a glycerol-soaked potato medium, but they found that there was difficulty in the production of homogenous suspension of the bacilli and bacteria used to grow in clumps in vitro. In

order to minimize mycobacterial clumping and make the bacteria homogenous for optimized infections, they added ox bile to the potato slices soaked in glycerol. To their surprise, they observed alterations in colony morphology within a few months of growth on this new medium, and when injected into guinea pigs, the bacilli were less virulent than the original Nocard's *M. bovis*. This fortuitous observation became source of long term project of producing a vaccine from this attenuated tubercle bacilli (Calmette 1922; Gheorghiu et al. 1983).

However, appreciating the importance of reduced virulence in terms of vaccine development, Calmette and Guérin continued the serial in vitro sub-culturing of the Nocard's *M. bovis* strain on potato slices soaked in ox bile and glycerol at three weekly intervals for 13 years (1908–1921), leading to total of 231 passages (Corbel et al. 2004). When administered at different doses and by different routes, the lab-adapted Nocard *M. bovis* strain was well tolerated and failed to produce progressive TB in different animal models, including guinea pigs, cows, horses, hamsters, mice, rabbits, dogs, chicken, and non-human primate (Sakula 1983). However, these cultures maintained the clumped morphology, and the same physical properties and exhibited continued immunogenicity in animal models i.e., guinea pigs, cattle, mice and chimpanzees. Furthermore, BCG vaccination protected cows against challenge with virulent *M. bovis*. These results established the safety and efficacy of BCG vaccine in experimental animals. At Guerin's suggestion, they named it Bacille Bilie Calmette-Guerin; later they omitted "Bilie" and so BCG was born (Calmette 1922).

In 1921, the BCG strain was used for the first time as a human vaccine at the request of Benjamin Weill-Halle a French pediatrician and bacteriologist. Weill-Halle wanted to protect an infant born to a mother who had died of TB a few hours after childbirth, and who was now under the care of a grandmother suffering from TB at the Charité Hospital, Paris (Bryder 1999; Sakula 1983). On July 18th, 1921, Benjamin Weill-Halle assisted by Raymond Turpin administered orally the culture of the lab-adapted Nocard's *M. bovis* strain in three doses of 2 mg each (6 mg total; $\sim 2.4 \times 10^8$ bacilli). On follow-up, there were no serious side effects, and the child did not develop any sign of TB. Based on this first anecdotal success, over the next year, additional newborns were vaccinated and no ill effects were reported. By 1924, they were able to report a series of more than 660 oral BCG vaccinations of infants (Calmette et al. 1924). For the first time, a safe and apparently effective vaccine was available for protection against human TB. The Pasteur Institute of Lille started mass-production of the BCG vaccine for medical applications.

As early as 1924, the original culture of BCG strain was sub-cultured and distributed to several laboratories throughout the world (Oettinger et al. 1999). These cultures were further propagated on non-synthetic culture media that varied around the world, and used for local vaccine production. This propagation of BCG on different culture media, at times following different passaging schedules, led to its diversification into a number of genetically distinct BCG sub-strains (daughter strains) (Liu et al. 2009). The first documented distribution of a daughter strain was BCG-Russia obtained in 1924 (Dubos and Pierce 1956). In the absence of lyophilisation or freezing and the production of seed-lots in 1961, BCG-Pasteur continually underwent serial passage of in vitro evolution, with daughter strains obtained directly or indirectly from the Pasteur Institute of Lille. After BCG Russia, records of BCG transfers between laboratories have shown that BCG Moreau and Japan (Tokyo-172), were obtained in 1925 (Behr and Small 1999; Obayashi 1955), Sweden in 1926 and Birkhaug in 1927 (Lind 1983; Wallgren 1928). Over the years, more than 14 sub-strains of BCG have evolved and have been used as BCG vaccine strains in different parts of the world. The main strains or seed-lots that are currently in use are: BCG Pasteur 1173 P2 (lyophilized strain in 1961 after 1173 serial passages (Gheorghiu et al. 1983)), BCG Danish 1331, BCG Glaxo 1077 (derived from the Danish strain), and BCG Moreau RDJ strains. The

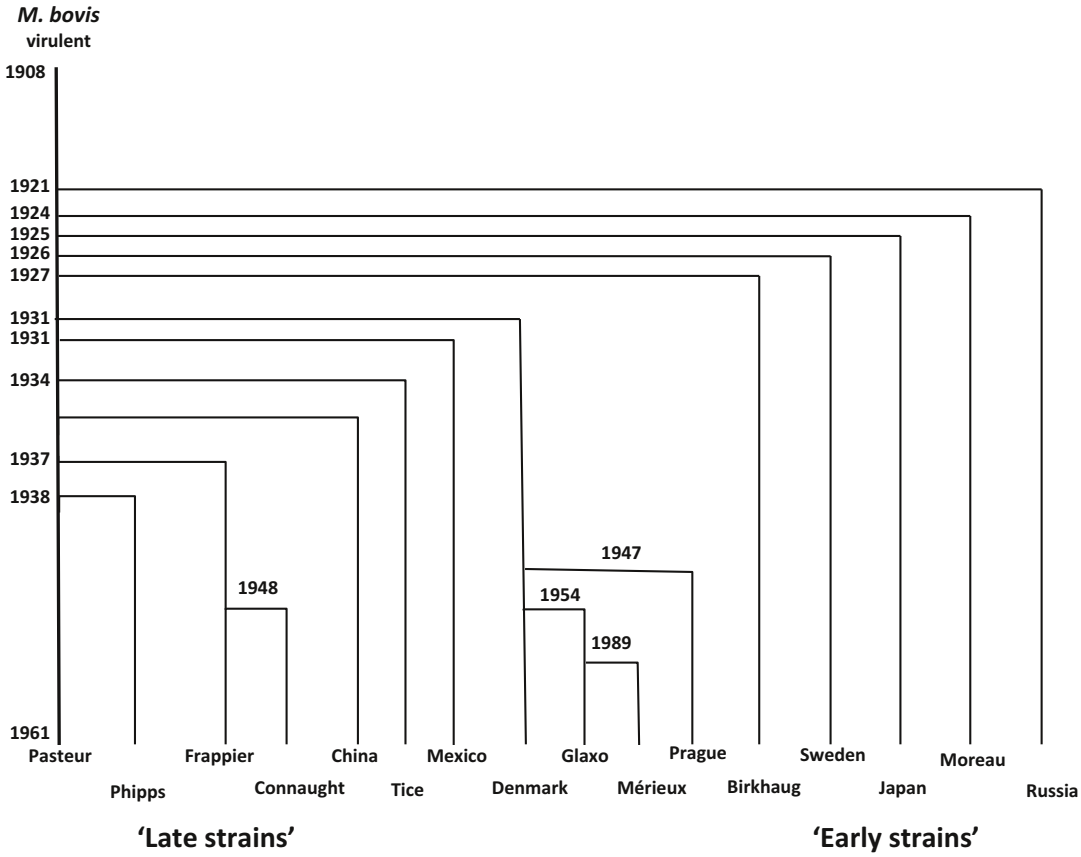

Fig. 8.1 Revised historical genealogy of BCG strain dissemination. The *vertical axis* represents time. The *horizontal axis* denotes different geographic locations of BCG propagation. Strains obtained before 1927 are labelled as "early strains"; strains obtained in 1931 or later are indicated as "late strains" (Adapted from a figure by Behr et al. 1999)

historical records of BCG dissemination separated "early strains" that were obtained in the early 1920s (BCG Russia, BCG Tokyo, BCG Moreau, BCG Sweden, BCG Birkhaug) from "late strains" that were obtained from the Pasteur Institute after 1927 (Fig. 8.1). The distinction between early strains and late strains coincides with reports of ongoing attenuation of BCG in the late 1920s (Dreyer and Vollum 1931), and is correlated with severely reduced production of the antigenic proteins MPB70, MPB83 and MPB64 in late strains (Milstien and Gibson 1990; Wiker et al. 1996). However, whether this historical distinction is critical for BCG phenotypes remains uncertain, as certain properties such as production of virulence lipids, have been independently lost in both early and late strains (Chen et al. 2007). Furthermore, the relevance of an early versus late dichotomy in term of protective efficacy in humans cannot be ascertained, as randomized clinical trials of BCG vaccination only employed late strains.

8.3 Phylogeny of BCG

It is fundamental to recognize that the currently available BCG strains have undergone two phases of genomic modification. The initial phase (1908–1921) comprises the 231 in vitro passages conducted by Calmette and Guérin to produce the lab-adapted Nocard's *M. bovis*

strain (the original vaccine). The second phase starts circa 1924 with widespread use and distribution of the culture of the lab-adapted Nocard's *M. bovis* strain. It ends several decades, and hundreds of passages, later (1961 for BCG-Pasteur 1173 P2, but different years for different daughter strains) with the establishment of frozen seed lots. Due to the initial phase, it is expected that BCG daughter strains should share the acquired particular genomic modification (BCG versus virulent *M. bovis*), whereas the second phase should give rise to additional genomic modification, specific to individual BCG strains and lineages. During the past two decades, several studies on BCG daughter strains have demonstrated changes at the genome level, using a variety of comparative genomic techniques, including subtractive hybridization, BAC libraries, spotted oligonucleotide arrays, microarray based resequencing, and whole genome sequencing. These studies have not only documented differences between BCG and *M. bovis*, but also documented additional genomic modifications that apparently were not present in the original culture obtained by Calmette and Guerin (Abdallah et al. 2015; Behr et al. 1999; Brosch et al. 2000, 2007; Gordon et al. 1999; Mahairas et al. 1996; Mostowy et al. 2003; Salamon et al. 2000). This extensive genotypic diversity in BCG daughter strains includes uncovered regions of difference (RD), single nucleotide polymorphisms (SNPs), insertion sequences (IS*6110*), deletions and tandem duplications.

Notably, due to the loss of the original strain of *M. bovis* (Nocard strain) used to derive BCG during the First World War, genomic studies have compared BCG vaccine to a variety of different *M. bovis* strains until the sequence of the *M. bovis* AF2122/97 strain was completed in 2003 (Garnier et al. 2003). As a consequence, certain differences uncovered between a particular BCG strain and the chosen *M. bovis* strain could represent *M. bovis*-*M. bovis* variants rather than BCG-specific differences. For instance, most circulating strains of *M. bovis*, including the sequenced strain AF2122/97, have only one copy of the IS*6110* element. A straightforward comparison with most BCG strains might suggest that this was the ancestral state of BCG. However, a second copy of IS*6110* element has been shown to be present in certain BCG daughter strains, and these were limited to BCG strains obtained before 1925 (BCGs Russia, Moreau, and Japan). While a potential explanation would be that a second IS*6110* element was introduced into these different strains, the finding that they each have IS*6110* element at base pair 851,592 of the *M. tuberculosis* H37Rv genome suggest the possibility that the ancestral *M. bovis* which gave rise to BCG had two copies of IS*6110*, with one IS*6110* element deleted in 1925–1926 (Mostowy et al. 2003).

Based on the accessible data, it is now well defined which genetic characteristics are shared across all BCG daughter strains relative to *M. bovis* that may be directly involved in the attenuation of BCG, and which genetic variations are specific for only certain BCG daughter strains that may account for variation in protective efficacy and over attenuation of certain BCG daughter strains. RD1 is the example for the genomic particularities that apply to all BCG daughter strains and shown to be implicated in virulence. Deletion of RD3 and Del_Mb2377c most likely occurred during the first attenuation period (1908–1921) as these regions are absent in all BCG daughter strains (Liu et al. 2009). The variations that were observed for only certain daughter strains of BCG consist of deletions, duplications, and point mutations, and probably occurred during the following period of divergence (1921–1966). Of note, RD2 was deleted only from daughter strains derived from the BCG Pasteur strain after 1927 while nRD18 is only deleted in strains obtained after 1933.

Based on the RD1 and RD2 deletions among BCG daughter strains, two main groups of BCG have been suggested. The first group includes BCG Russia, Moreau, Japan, Sweden and Birkhaug, which were distributed from the Pasteur Institute between 1921 and 1926 and

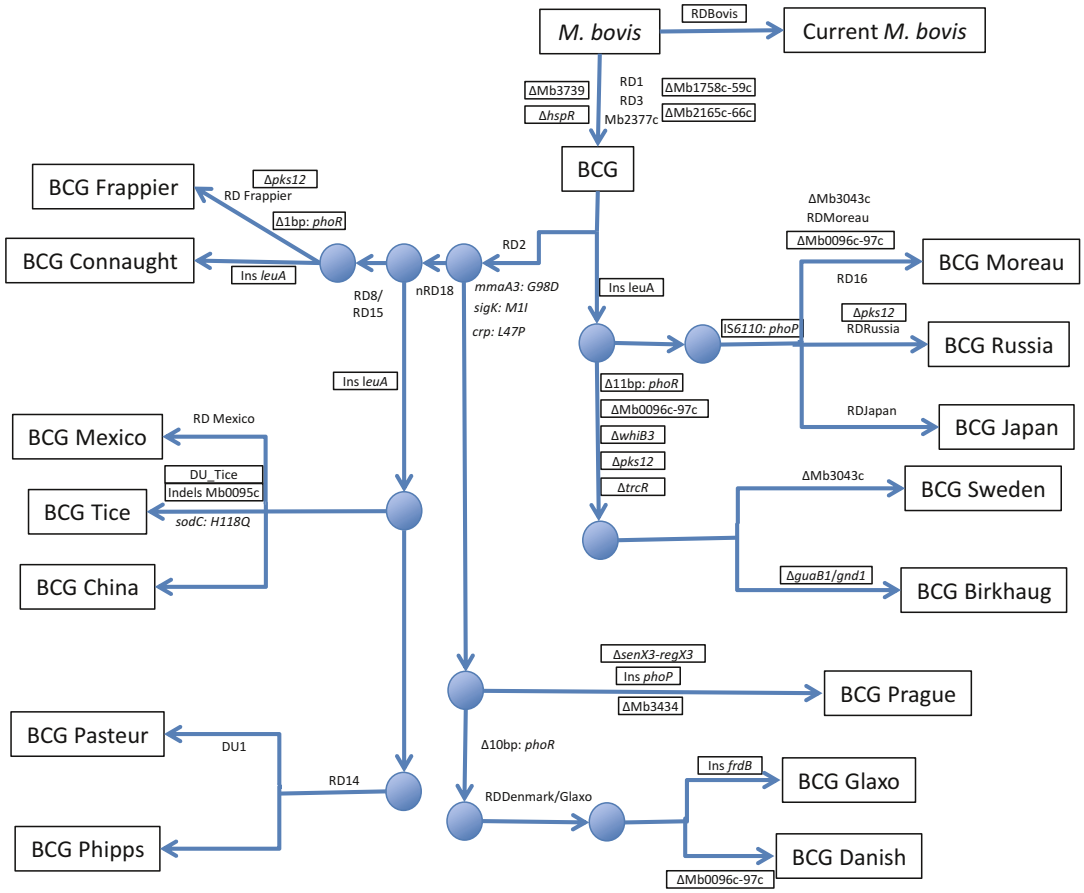

Fig. 8.2 Revised genealogy of BCG vaccines. Evolutionary scheme of BCG vaccine strains (From reference Abdallah et al. 2015), displaying the original virulent *M. bovis* ancestor strain and the subsequent series of genomic alteration including deletions of regions of difference (*RD*), SNPs and some strain-specific insertions (‘*Ins*’) and deletions (‘Δ’)

have only the RD1, RD3 and Del_Mb2377c deletions. The second group, distributed after 1927, includes BCG Prague, Glaxo, Merieux, Danish, Frappier, Connaught, Tice, Mexico, China, Phipps and Pasteur, and has the first group of deletions as well as the RD2 deletion. Based on tandem duplication marker, each of the above two groups have been further divided into two DU types. BCG Russia, Moreau and Japan as DU2 group I; BCG Sweden and Birkhaug as DU2 group II; BCG Prague, Glaxo, Merieux and Danish as DU2 group III and Frappier, Connaught, Tice, Mexico, China, Phipps and Pasteur as DU2 group IV (Brosch et al. 2007). Therefore, a molecular phylogeny based on this available data has been established and is generally consistent with the historical records of BCG dissemination (Fig. 8.2).

8.4 Molecular Evolution of BCG Daughter Strains Between 1908 and 1921

During continuous in vitro passage between 1908 and 1921, BCG lost 38 open reading frames, which were identified using subtractive hybridization (Mahairas et al. 1996), BAC libraries (Gordon et al. 1999), and spotted oligonucleotide arrays (Behr et al. 1999). These deleted genes include the RD1 region. The suggestion that the loss of RD1 contributes

to BCG attenuation has been confirmed by subsequent studies. RD1 encodes the ESX-1 protein secretion system, which is one of the five type-VII secretion systems found in the *M. tuberculosis* genome (Abdallah et al. 2007; Behr and Sherman 2007). RD1 is 9.5 kilobases (kb) in length and comprises nine genes, including the genes that encode the secreted proteins ESAT-6 (early secreted antigenic target of 6 kDa) and CFP-10 (culture filtrate protein of 10 kDa). Both of these proteins are important T-cell antigenic targets and are essential for the virulence of *M. tuberculosis*. The first experimental evidence for such contribution was obtained when the BCG Pasteur strain was complemented with the RD1 locus. The recombinant BCG strain was more virulent in severely immunodeficient mice than the BCG Pasteur strain (Pym et al. 2003). In other studies, deletion of RD1 from virulent *M. bovis* and *M. tuberculosis* strains resulted in ΔRD1 mutants that were significantly attenuated for virulence in both immunocompromised and immunocompetent mice (Hsu et al. 2003; Lewis et al. 2003). In a separate approach, individual genes in the RD1 locus were identified as virulence factors for *M. tuberculosis* (Guinn et al. 2004; Hsu et al. 2003; Stanley et al. 2003). The disruption of these individual genes also resulted in the attenuation of virulence in mice. Collectively, these studies provide convincing evidence that the RD1 deletion is a key mechanism of BCG attenuation and plays a major role in virulence; although these studies did not exclude additional genetic lesions as also contributing to the loss of virulence and attenuation of BCG described by Calmette and Guerin between 1908 and 1921. Since it is common to all BCG strains, the loss of RD1 likely occurred in the initial stage of BCG attenuation.

A recent comparative genome analysis of multiple BCG daughter strains using whole genome sequencing has uncovered a 103 bp deletion present across all BCG daughter strains, and which eliminates the distal end of *hspR* (Abdallah et al. 2015). This gene is involved in transcriptional regulation (repression) of heat shock proteins and is known to impact virulence. The *hspR* locus activates a subset of the heat-shock general stress response upon macrophage invasion (Stewart et al. 2002), and is necessary in the persistent phase since strains with an hspR deletion (ΔhspR) exhibit attenuated growth in chronic infection (Stewart et al. 2001). Furthermore, the same study showed that a duplication of 2900 bp segments spanning the region between 1,276,501 and 1,279,400 base pairs (*M. bovis* coordinates) in all BCG daughter strains (Abdallah et al. 2015).

Although the most attributed reason for the primary attenuation of BCG compared to *M. bovis* is the loss of the RD1 locus, complementation of BCG with this region does not fully restore virulence to wild-type levels (Pym et al. 2003), and the RD1 deletion mutant of *M. tuberculosis* is still more virulent than BCG in long-term murine infection experiments (Sherman et al. 2004). This leads to speculation that additional genetic lesions that have occurred in BCG also contribute to its attenuation. Whole genome sequence comparison revealed 736 single nucleotide polymorphisms (SNPs) between BCG-Pasteur 1173P2 and *M. bovis* AF2122/97 (Brosch et al. 2007). Recent comprehensive studies of SNPs between *M. bovis* and BCG daughter strains revealed two types of SNPs that could also play a role in the early evolution of BCG (Abdallah et al. 2015, Pelayo et al. 2009). Most are common to the entire BCG lineage, but some SNPs are specific to individual BCG strains. This suggests that loss of virulence during the initial 231 in vitro passages of BCG involved both the loss of RD1 and other mutations (the aforementioned deletion and duplication, plus SNPs), which together likely contributed to the attenuation of virulence during the derivation of BCG.

Of the SNPs that are common to all BCG daughter strains, some have been explored in greater depth, but further study is required. For example, SNPs in genes predicted to be involved in glycerol catabolism were identified in two genes, *glpK* (codon 191) and *pykA* (codon 220). *glpK* encodes glycerol kinase, the enzyme which catalyzes the MgATP-dependent phosphorylation of glycerol to yield sn-glycerol 3-phosphate, the rate limiting step in glycerol

utilization in *E. coli* (Zwaig et al. 1970). Remarkably, this SNP is present in *M. bovis* AF2122/97 and causes a frameshift in the *glpK* gene at codon 191, leading to a truncated coding sequence (Garnier et al. 2003). Noteworthy, the *glpK* frameshift is not evident in the BCG daughter strains or in the *M. bovis* AN5 strain, strains that can grow on glycerol as the sole carbon source (Keating et al. 2005). The second SNP is a nucleotide substitution in *pykA*, the gene that encodes pyruvate kinase, an enzyme that catalyzes the final step in glycolysis, the conversion of phosphoenolpyruvate to pyruvate. The *pykA* SNP results in the substitution of glutamic acid 220 by aspartic acid and renders pyruvate kinase nonfunctional. This highly conserved glutamic acid residue is predicted to play an important role in the active site of pyruvate kinase and has been associated with cofactor (Mg++) and substrate (ADP/ATP, PEP) binding (Munoz and Ponce 2003). As with the *glpK* SNP, the *pykA* mutation does not occur in the BCG daughter strains or *M. bovis* AN5 but is present in *M. bovis* AF2122/97 and was suggested to account for inability of *M. bovis* to grow on glycerol as a sole carbon source (Keating et al. 2005). This suggests that progenitors of BCG daughter strains (Nocard's *M. bovis*) and *M. bovis* AN5 strain had an in-frame *glpK* and *pykA* coding sequence.

Another study implicated one of the first BCG polymorphism described in Mb3700, a gene encoding a transcriptional regulator of the cyclic AMP (cAMP) receptor protein (CRP)- fumarate and nitrate reduction regulator (FNR) family that could affect the DNA binding activity of this putative global transcriptional regulator and therefore, could contribute to the attenuation of BCG strains. Further studies demonstrated that this point mutation resulting in a base substitution of glutamic acid to lysine (E178K) present in all BCG daughter strains examined altered DNA binding of CRP to target sites as well as global gene expression, without playing any role in the attenuation of BCG (Bai et al. 2007; Hunt et al. 2008). Although these lesions have not been directly associated with the original attenuation of BCG, it remains possible that the accumulation of multiple lesions, including SNPs, between 1908 and 1921 may have resulted in a complex effect that is yet to be fully resolved.

8.5 Molecular Evolution of BCG Daughter Strains After 1924

Dissemination of BCG from Pasteur Institute to various parts of the world began in 1924. Subsequent to the original derivation of BCG, strains were passaged in a variety of non-synthetic media, based on variable growth factors such as the local potatoes. As these laboratory conditions were not uniform, it was perhaps not surprising that vaccine producers recognized the emergence of BCG daughter strains with distinct morphological, biochemical and immunological phenotypic properties by the 1940s and 1950s. Thanks to genomic studies, it is now known that different BCG strains differ both from the original BCG of 1921 and from each other, due to deletions, SNP and duplications. Thus there is no "ancestral" BCG in existence. For instance, BCG Pasteur may be considered the reference strain of BCG vaccines, but the sequenced BCG Pasteur 1173 is separated from the BCG of 1921 by a number of genetics events (e.g., deletion of IS*6110*, RD2, nRD18, and RD14; SNPs in *mmA3*, *sigK*, and *crp_L47P*; duplication of DU1) (Fig. 8.2). Evidence to suggest a closer relationship between BCG Russia and the original BCG progenitor strain emerged from the discovery that BCG Russia is a natural *recA* mutant (Keller et al. 2008). Because a RecA mutation might result in fewer mutation, it was speculated that this mutation might have kept BCG Russia in a state 'closer' to the 1921 ancestor of all BCG strains. Indeed, upon determining the complete genome sequence of a panel of BCG strains, it was shown that BCG Russia contains fewer SNPs and deletions compared to other BCG strains (Abdallah et al. 2015).

A critical divergence in the evolution of BCG was the loss of the 10.8 kb region of RD2 during the ongoing propagation of BCG between 1927 and 1931, a time that coincides with reports of the ongoing attenuation of the vaccine. This

led to "early" (RD2 present) and "late" (RD2 absent) BCG strains. The role of RD2-associated virulence was evaluated by a targeted knockout in the *M. tuberculosis* reference strain H37Rv, which showed that the deletion mutant was attenuated in murine models of infection (Kozak et al. 2011). Moreover, some of the genes encoded by RD2 stand out as candidates for the ongoing attenuation of BCG daughter strains in the laboratory. The deletion of RD2, which contains the gene *mpb64* and encodes the antigenic protein MPB64, accounts for the lack of MPB64 in the late strains. Complementation of the BCG daughter strain Pasteur with *mpb64* gene improved the immunogenicity of the vaccine strain but did not improve protection against pulmonary TB (Kozak and Behr 2011).

In addition to RD2, several other genomic polymorphisms that are common to all strains obtained from the Pasteur Institute after 1927 have been investigated, including point mutations that may have contributed to the later evolution of BCG. For instance, protein antigens MPB70 and MPB83 are found to be at high levels in early strains prior to 1927 (acquired before 1927) but are present only in trace quantities in late BCG strains (Milstien and Gibson 1990). Follow-up studies using immunoblot and quantitative reverse transcription polymerase chain reaction (RT-PCR) separated BCG into high- and low-producing strains, and determined that transcription of the antigen-encoding genes, *mpb70* and *mpb83*, follows the same strain pattern with mRNA levels reduced over 50-fold in low-producing strains (Charlet et al. 2005). Using transcriptomic analysis, Charlet et al. identified two regions of the genome that had dysregulated gene expression between high-producers and low-producers: the set of genes including *mpb70* and *mpb83*, and a distant set of genes including the gene coding for Sigma Factor K. DNA sequence analysis showed a point mutation in the *SigK* gene, leading to gene complementation studies that formally implicated this mutation in the decreased expression of MPB70 and MPB83 in the later strains (Charlet et al. 2005).

Strains obtained from the Pasteur Institute prior to 1927 produced methoxymycolates, a subclass of cell wall mycolic acids in vitro but those obtained later could not synthesize methoxymycolates. This phenotype has been attributed to a point mutation in the *mmaA3* gene, which encodes methoxy mycolic acid synthase 3 and is responsible for O-methylation of hydroxymycolate precursors to form methoxymycolic acids. This base substitution at position 293 in *mmaA3* results in an amino acid change from glycine to aspartic acid and inhibits methoxymycolates production (Behr et al. 2000). Also, it has been postulated that this SNP leads to low-level isoniazid (INH) resistance (Abdallah et al. 2015), as INH is known to inhibit the synthesis of α-mycolate, methoxymycolate and ß-mycolate (Takayama et al. 1972). Intriguingly, the loss of methoxymycolate appears to have no impact on the virulence of late BCG strains (Belley et al. 2004). Furthermore, two predicted regulators have been disrupted during the serial passage of BCG strains. The *sigI* gene (possible alternate RNA polymerase sigma factor SigI) encoded in nRD18 is missing from BCG strains obtained after 1933. Additionally, RD14, which encodes Mb1802 and has been annotated as a probable transcriptional regulatory protein, is deleted from BCG-Pasteur (Behr et al. 1999) and some strains of BCG-Phipps (Abdallah et al. 2015); this predicted role as a transcriptional regulator was subsequently confirmed with promoter fusion assays (Alexander and Behr 2007). Finally, Mb3439c, a gene encoded in RD16 and also annotated as a possible transcriptional regulator has been disrupted by independent deletions in BCG-Moreau and BCG-Japan. The loss of regulatory genes from a number of different BCG strains argues that mutation of regulatory genes can be tolerated during conditions of laboratory growth.

BCG-Japan, BCG-Moreau and BCG-Glaxo do not produce the lipid virulence factors phthiocerol dimycocerosates (PDIMs) and phenolic glycolipids (PGLs), and were found to be naturally deficient in PDIMs and PGLs, whereas the other BCG strains do produce these lipids (Chen et al. 2007). Because some strains of BCG have been associated with higher rates of adverse effects, such as disseminated BCGosis, investigators have queried whether the presence/absence

of these virulence factors might in part explain the variable rates of these events. Indeed, the loss of these lipids has been shown to correlate with the superior safety records of these strains in clinical studies (Chen et al. 2007; Lotte et al. 1984). Moreover, deletion of PIDMs/PGLs from BCG Pasteur reduces its virulence and protective efficacy (Tran et al. 2016). Intriguingly, variation in PDIM and PGL production does not coincide with the genealogy of BCG strains (Abdallah et al. 2015; Brosch et al. 2007), suggesting that this particular phenotype has emerged multiple times and by multiple mechanisms. Indeed, in BCG-Moreau, the PDIM and PGL defect is likely due to a 975-bp deletion that affects *fadD26* and *ppsA* (Leung et al. 2008), which are members of the PDIMs and PGLs biosynthetic locus (Azad et al. 1997, Leung et al. 2008). Similarly, a point mutation in *ppsA* is responsible for the lack of PDIMs/PGLs in BCG-Japan (Naka et al. 2011). However, this region is intact in BCG-Glaxo (Leung et al. 2008), indicating that the PDIM/PGL defect in BCG-Glaxo is caused by other, currently unknown genetic lesions. Together, this suggests that these three BCG strains independently acquired mutations in PDIM/PGL biosynthesis that contributed to their further attenuation.

Considering its prominent role in virulence, it may not be surprising that BCG strains exhibit a number of genetic polymorphism in the *phoP-phoR* locus (Leung et al. 2008), a two-component system known to regulate the expression of multiple genes, including some well-established T-cell antigens (Walters et al. 2006). Of note, a frame-shift mutation within the *phoP* gene of BCG-Prague eliminates the majority of C-terminal DNA binding domain, and makes this strain a natural *phoP* mutant (Gupta et al. 2006; Sinha et al. 2008; Wang et al. 2007), and it is possible that this could account for this strain's reported low immunogenicity (Ladefoged et al. 1976; Vallishayee et al. 1974). Furthermore, three early strains as mentioned earlier, BCG-Russia, BCG-Japan and BCG-Moreau have an *IS6110* insertion in the promoter region of *phoP* (Leung et al. 2008), which may eliminate the auto-repression regulatory mechanism of this two component system (Gupta et al. 2006). BCG-Sweden and BCG-Birkhaug contain a deletion that truncates the C-terminal of *PhoR* (Leung et al. 2008). In BCG-Danish, BCG-Glaxo and BCG-Frappier, frame-shift mutations in the *phoR* gene abolish the PhoR protein (Leung et al. 2008).

Unlike other BCG strains, two closely related BCG strains, BCG-Sweden and BCG-Birkhaug, contain deletions in *whiB3*, a transcriptional regulator implicated in virulence of *M. bovis* (Steyn et al. 2002) and *M. tuberculosis* (Singh et al. 2009), in *pks12,* a polyketide synthase that necessary for CD1c-mediated T-cell response, which may affect the immunogenicity of the vaccine (Matsunaga et al. 2004), and in *trcR*, the response regulator of the TrcR-TrcS two-component system that controls expression of the *trcRS* operon, including *mmpS5/mmpL5* transporter and *bfrB* bacteroferrin (Wernisch et al. 2003). Intriguingly, deletion of *trcS* from *M. tuberculosis* produced a hypervirulent phenotype in SCID mice (Parish et al. 2003), suggesting a role for this gene in the attenuation of BCG. However, these deletions are not present in other BCG strains and therefor distinguish the BCG-Sweden and BCG-Birkhaug lineage from other early strains (e.g., BCG-Russia, BCG-Japan and BCG-Moreau).

Comparative genomics has also uncovered two large tandem duplications, DU1 and DU2 of 29-kb and 36-kb, respectively, in BCG-Pasteur (Brosch et al. 2000, 2007). These seem to have arisen independently, as their presence and/or their size varies between the different BCG daughter strains. While DU1 appears to be restricted to BCG-Pasteur, DU2 has detected in all BCG daughter strains tested so far but at variable size (Abdallah et al. 2015; Brosch et al. 2007). Interestingly, DU1 contains the *oriC* locus, the site of chromosomal origin of replication, indicating that BCG-Pasteur is diploid for *oriC*, and several key genes involved in replication initiation and cell division machinery (Brosch et al. 2000). For DU2, the tandem duplication resulted in diploidy for 30 genes, for which probable functions are known. These include *aroA*, a key enzyme in aromatic amino acid biosynthesis, and the coding

sequences for variety of regulatory proteins that could exert pleiotropic effects, including a histidine kinase *asnC* and *tetR* homologues, *whiB1* and *sigH*, a sigma factor implicated in the heat shock response (Fernandes et al. 1999). Chromosomal duplications are a common evolutionary response in bacteria exposed to different selection pressures in the laboratory and presumably in nature, as they provide a means for increasing gene dosage and for generating novel functions from potential gene fusion events at duplication endpoints. They also represent a source of redundant DNA for divergence. As such, the presence of DU1 and DU2 suggests that the process of tandem duplications in BCG is ongoing and remains a potent source of genome dynamics. However, the potential role of these duplication events on the immunogenicity of BCG strains remains to be explored. Other duplications specific to certain BCG strains have also been uncovered, such as a 22-kb duplication present only in BCG Tice (DU-Tice) (Leung et al. 2008). DU-Tice contains the entire ESX-5 secretion system, which is present only in pathogenic mycobacteria (Abdallah et al. 2007), and has been shown to be responsible for the transport of cell envelope proteins that are required for nutrient uptake (Ates et al. 2015), and directly or indirectly modulate the human macrophages response (Abdallah et al. 2008). Interestingly, sequencing of BCG-China repositioned this strain from the DU2-III group into the DU2-IV group that includes BCG-Pasteur, BCG-Phipps, BCG-Tice, BCG-Mexico, BCG-Connaught and BCG-Frappier. This analysis revealed that BCG-China was not originally derived from BCG-Danish, as previously thought, and that this inconsistency was likely attributed to multiple circulating strains of "BCG China" (Abdallah et al. 2015).

8.6 Implications for BCG Immunization

BCG vaccines are given to over a hundred millions of newborns each year, with the goal of safely preventing TB disease. The ideal profile of such a vaccine should be: (1) prevents disease, (2) rarely causes progressive vaccine-associated infection, and (3) if vaccine infection occurs, this should be readily treatable. At present, it is unclear, despite all known from genomic studies, whether a particular BCG is better, or less desirable, at addressing the aforementioned criteria.

For protective efficacy, a lingering question is whether early strains, which produce antigens such as MPB64, MPB70 and MPB83, are more likely to protect against TB. However, there have been no randomized trials that have competed an early strain against a late strain of BCG; in fact, there have been no randomized controlled trials whatsoever of the early BCG strains. For safety, there have been reports that the early strains have been associated with a greater risk of vaccine-associated infection (also known as BCG-osis). However, these reports were generally confined to before-and-after studies in countries with a change of vaccine strain for political reasons. Formal demonstration that early BCG strains are more virulent, and therefore, a greater risk in immunization schedules, is lacking. Finally, for treatment of rare cases of BCG-osis, it is worth noting that all BCG strains are resistant to pyrazinamide, as this is a genetic feature of *M. bovis* (Huard et al. 2006), and that some strains are apparently resistant to isoniazid (Abdallah et al. 2015; Kolibab et al. 2011). The degree of isoniazid resistance is considered low-level, such that some consider this not to be a clinical issue (Arend and van Soolingen 2011). In addition, they are also resistant to cycloserine, partly due to the G122S mutation in *cycA* (Chen et al. 2012). Interestingly, expression of a functional copy of *cycA* from *M. tuberculosis* or *M. bovis* increased the susceptibility of BCG Pasteur to d-cycloserine, albeit not to the levels of *M. bovis* or *M. tuberculosis* (Chen et al. 2012), suggesting that other genetic lesions may contribute to BCG resistance to d-cycloserine. Nonetheless, it is not intuitive that we should be giving a live attenuated vaccine to millions of newborns, knowing that some of these vaccines are inherently resistant to two of the four antibiotics used to treat TB. Further study should aim to formally identify the mutation(s) responsible for isoniazid resistance,

as there could be policy implications in knowing which strains might not be ideal for vaccination purposes.

Finally, BCG is used not only to prevent TB, but also as an immunotherapeutic for bladder cancer. The lessons of BCG evolution and strain variability have yet to affect the choice of BCG strains for this indication, yet, many parallels can be imagined. How does BCG work in bladder cancer? What are the adverse effects? How is BCG-osis treated in an elderly, debilitated patient suffering from a malignancy? Further study of the BCG strain family, particularly with the goal of linking genotype to clinically-meaningful phenotype, is required, to find the best use of these bacteria given to us by the work of Calmette and Guerin.

8.7 Conclusion

More than 100 years ago, Albert Calmette and Camille Guérin began their research for an anti-TB vaccine, which lasted 13 years and led to one of the most widely used vaccines in human history. Despite a century of investigation, the BCG vaccine continues to be controversial and remains the only vaccine for the prevention of TB. Shaped by human history, BCG has also evolved to the various daughter strains recognized today. These evolved strains differ from each other and from the original BCG first used in 1921, both genetically and phenotypically, such that these changes may translate into variable vaccine properties, including protective efficacy, tuberculin reactivity and propensity for adverse effects. Remarkably, the current collection of BCG comprises various natural mutants of established virulence factors. Continuing studies on the molecular factors that impact properties of BCG vaccine strains would shed light on the specific differences among BCG strains and further our understanding of the mechanisms of attenuation specific to each lineage. This would be useful in further delineating BCG strains into phenotypic and genetic categories that could mediate their protective efficacy.

References

Abdallah AM, Gey Van pittius NC, Champion PA, Cox J, Luirink J, Vandenbroucke-Grauls CM, Appelmelk BJ, Bitter W (2007) Type VII secretion – mycobacteria show the way. Nat Rev Microbiol 5:883–891

Abdallah AM, Savage ND, Van zon M, Wilson L, Vandenbroucke-Grauls CM, Van Der Wel NN, Ottenhoff TH, Bitter W (2008) The ESX-5 secretion system of Mycobacterium marinum modulates the macrophage response. J Immunol 181:7166–7175

Abdallah AM, Hill-Cawthorne GA, Otto TD, Coll F, Guerra-Assuncao JA, Gao G, Naeem R, Ansari H, Malas TB, ADROUB SA, Verboom T, Ummels R, Zhang H, Panigrahi AK, Mcnerney R, Brosch R, Clark TG, Behr MA, Bitter W, Pain A (2015) Genomic expression catalogue of a global collection of BCG vaccine strains show evidence for highly diverged metabolic and cell-wall adaptations. Sci Rep 5: 15443

Alexander DC, Behr MA (2007) Rv1773 is a transcriptional repressor deleted from BCG-Pasteur. Tuberculosis (Edinb) 87:421–425

Arend SM, Van Soolingen D (2011) Editor's choice: editorial commentary: low level INH-resistant BCG: a sheep in wolf's clothing? Clin Infect Dis 52:89–93

Aronson JD, Aronson CF, Taylor HC (1958) A twenty-year appraisal of BCG vaccination in the control of tuberculosis. AMA Arch Intern Med 101:881–893

Ates LS, Ummels R, Commandeur S, Van de Weerd R, Sparrius M, Weerdenburg E, Alber M, Kalscheuer R, Piersma SR, Abdallah AM, Abd el Ghany M, Abdel-Haleem AM, Pain A, Jimenez CR, Bitter W, Houben EN (2015) Essential role of the ESX-5 secretion system in outer membrane permeability of pathogenic mycobacteria. PLoS Genet 11:e1005190

Azad AK, Sirakova TD, Fernandes ND, Kolattukudy PE (1997) Gene knockout reveals a novel gene cluster for the synthesis of a class of cell wall lipids unique to pathogenic mycobacteria. J Biol Chem 272:16741–16745

Bai G, Gazdik MA, Schaak DD, Mcdonough KA (2007) The Mycobacterium bovis BCG cyclic AMP receptor-like protein is a functional DNA binding protein in vitro and in vivo, but its activity differs from that of its M. tuberculosis ortholog, Rv3676. Infect Immun 75:5509–5517

Behr MA (2002) BCG – different strains, different vaccines? Lancet Infect Dis 2:86–92

Behr MA, Sherman DR (2007) Mycobacterial virulence and specialized secretion: same story, different ending. Nat Med 13:286–287

Behr MA, Small PM (1999) A historical and molecular phylogeny of BCG strains. Vaccine 17:915–922

Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, Small PM (1999) Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science 284:1520–1523

Behr MA, Schroeder BG, Brinkman JN, Slayden RA, Barry CE 3rd (2000) A point mutation in the mma3 gene is responsible for impaired methoxymycolic acid production in Mycobacterium bovis BCG strains obtained after 1927. J Bacteriol 182:3394–3399

Belley A, Alexander D, Di Pietrantonio T, Girard M, Jones J, Schurr E, Liu J, Sherman DR, Behr MA (2004) Impact of methoxymycolic acid production by Mycobacterium bovis BCG vaccines. Infect Immun 72:2803–2809

Brewer TF (2000) Preventing tuberculosis with Bacillus Calmette-Guerin vaccine: a meta-analysis of the literature. Clin Infect Dis 31(Suppl 3):S64–S67

Brosch R, Gordon SV, Buchrieser C, Pym AS, Garnier T, Cole ST (2000) Comparative genomics uncovers large tandem chromosomal duplications in Mycobacterium bovis BCG Pasteur. Yeast 17:111–123

Brosch R, Gordon SV, Garnier T, Eiglmeier K, Frigui W, VALENTI P, Dos Santos S, Duthoy S, Lacroix C, Garcia-Pelayo C, Inwald JK, Golby P, Garcia JN, Hewinson RG, Behr MA, Quail MA, Churcher C, Barrell BG, Parkhill J, Cole ST (2007) Genome plasticity of BCG and impact on vaccine efficacy. Proc Natl Acad Sci U S A 104:5596–5601

Bryder L (1999) 'We shall not find salvation in inoculation': BCG vaccination in Scandinavia, Britain and the USA, 1921–1960. Soc Sci Med 49:1157–1167

Calmette A (1922) L'ninfection bacillaire et la tuberculose chez l'homme et chez les animaux

Calmette A, Guerin C, Weill-Halle B (1924) Essai d'immunisation contre l'infection tuberculeuse. Bull Acad Med Paris 91:787–796

Charlet D, Mostowy S, Alexander D, Sit L, Wiker HG, Behr MA (2005) Reduced expression of antigenic proteins MPB70 and MPB83 in Mycobacterium bovis BCG strains due to a start codon mutation in sigK. Mol Microbiol 56:1302–1313

Chen JM, Islam ST, Ren H, Liu J (2007) Differential productions of lipid virulence factors among BCG vaccine strains and implications on BCG safety. Vaccine 25:8114–8122

Chen JM, Uplekar S, Gordon SV, Cole ST (2012) A point mutation in cycA partially contributes to the D-cycloserine resistance trait of Mycobacterium bovis BCG vaccine strains. PLoS One 7:e43467

Colditz GA, Berkey CS, Mosteller F, Brewer TF, Wilson ME, Burdick E, Fineberg HV (1995) The efficacy of bacillus Calmette-Guerin vaccination of newborns and infants in the prevention of tuberculosis: meta-analyses of the published literature. Pediatrics 96:29–35

Corbel MJ, Fruth U, Griffiths E, Knezevic I (2004) Report on a WHO consultation on the characterisation of BCG strains, Imperial College, London 15–16 December 2003. Vaccine 22:2675–2680

Dreyer G, Vollum RL (1931) Mutation and pathogenicity experiments with BC G. Lancet 1:9–15

Dubos RJ, Pierce CH (1956) Differential characteristics in vitro and in vivo of several substrains of BCG. IV. Immunizing effectiveness. Am Rev Tuberc 74:699–717

Fernandes ND, Wu QL, Kong D, Puyang X, Garg S, Husson RN (1999) A mycobacterial extracytoplasmic sigma factor involved in survival following heat shock and oxidative stress. J Bacteriol 181:4266–4274

Fine PE (1995) Variation in protection by BCG: implications of and for heterologous immunity. Lancet 346:1339–1345

Frothingham R, Hills HG, Wilson KH (1994) Extensive DNA sequence conservation throughout the Mycobacterium tuberculosis complex. J Clin Microbiol 32:1639–1643

Garnier T, Eiglmeier K, Camus JC, Medina N, Mansoor H, Pryor M, Duthoy S, Grondin S, Lacroix C, Monsempe C, Simon S, Harris B, Atkin R, Doggett J, Mayes R, Keating L, Wheeler PR, Parkhill J, Barrell BG, Cole ST, Gordon SV, Hewinson RG (2003) The complete genome sequence of Mycobacterium bovis. Proc Natl Acad Sci U S A 100:7877–7882

Gheorghiu M, Augier J, Lagrange PH (1983) Maintenance and control of the French Bcg strain 1173-P2 (primary and secondary seed-lots). Bull Inst Pasteur 81: 281–288

Gordon SV, Brosch R, Billault A, Garnier T, Eiglmeier K, Cole ST (1999) Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. Mol Microbiol 32:643–655

Grange JM, Gibson J, Osborn TW, Collins CH, Yates MD (1983) What is BCG? Tubercle 64:129–139

Griffin JF, Chinn DN, Rodgers CR, Mackintosh CG (2001) Optimal models to evaluate the protective efficacy of tuberculosis vaccines. Tuberculosis (Edinb) 81:133–139

Guinn KM, Hickey MJ, Mathur SK, Zakel KL, Grotzke JE, Lewinsohn DM, Smith S, Sherman DR (2004) Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of Mycobacterium tuberculosis. Mol Microbiol 51:359–370

Gupta S, Sinha A, Sarkar D (2006) Transcriptional autoregulation by Mycobacterium tuberculosis PhoP involves recognition of novel direct repeat sequences in the regulatory region of the promoter. FEBS Lett 580:5328–5338

Hsu T, Hingley-Wilson SM, Chen B, Chen M, Dai AZ, Morin PM, Marks CB, Padiyar J, Goulding C, Gingery M, Eisenberg D, Russell RG, Derrick SC, Collins FM, Morris SL, King CH, Jacobs WR Jr (2003) The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. Proc Natl Acad Sci U S A 100:12420–12425

Huard RC, Fabre M, De Haas P, Lazzarini LCO, Van Soolingen D, Cousins D, Ho JL (2006) Novel genetic polymorphisms that further delineate the phylogeny of the Mycobacterium tuberculosis complex. J Bacteriol 188:4271–4287

Hunt DM, Saldanha JW, Brennan JF, Benjamin P, Strom M, Cole JA, Spreadbury CL, Buxton RS (2008) Single nucleotide polymorphisms that cause structural changes in the cyclic AMP receptor protein

transcriptional regulator of the tuberculosis vaccine strain Mycobacterium bovis BCG alter global gene expression without attenuating growth. Infect Immun 76:2227–2234

Imaeda T, Coppola KM, Furness G (1985) Deoxyribonucleic acids of Corynebacterium genitalium and Corynebacterium pseudogenitalium: their genome molecular weights, base ratios, and DNA relatedness with other corynebacteria involved in urinary tract infections. Can J Microbiol 31:1068–1070

Kaufmann SHE, Winau F (2005) From bacteriology to immunology: the dualism of specificity. Nat Immunol 6:1063–1066

Kaufmann SHE, Hussey G, Lambert PH (2010) New vaccines for tuberculosis. Lancet 375:2110–2119

Keating LA, Wheeler PR, Mansoor H, Inwald JK, Dale J, Hewinson RG, Gordon SV (2005) The pyruvate requirement of some members of the Mycobacterium tuberculosis complex is due to an inactive pyruvate kinase: implications for in vivo growth. Mol Microbiol 56:163–174

Keller PM, Bottger EC, Sander P (2008) Tuberculosis vaccine strain Mycobacterium bovis BCG Russia is a natural recA mutant. BMC Microbiol 8:120

Kolibab K, Derrick SC, Morris SL (2011) Sensitivity to isoniazid of Mycobacterium bovis BCG strains and BCG disseminated disease isolates. J Clin Microbiol 49:2380–2381

Kozak R, Behr MA (2011) Divergence of immunologic and protective responses of different BCG strains in a murine model. Vaccine 29:1519–1526

Kozak RA, Alexander DC, Liao R, Sherman DR, Behr MA (2011) Region of difference 2 contributes to virulence of Mycobacterium tuberculosis. Infect Immun 79:59–66

Ladefoged A, Bunch-Christensen K, Guld J (1976) Tuberculin sensitivity in guinea-pigs after vaccination with varying doses of BCG of 12 different strains. Bull World Health Organ 53:435–443

Leung AS, Tran V, Wu Z, Yu X, Alexander DC, Gao GF, Zhu B, Liu J (2008) Novel genome polymorphisms in BCG vaccine strains and impact on efficacy. BMC Genomics 9:413

Lewis KN, Liao R, Guinn KM, Hickey MJ, Smith S, Behr MA, Sherman DR (2003) Deletion of RD1 from Mycobacterium tuberculosis mimics bacille Calmette-Guerin attenuation. J Infect Dis 187:117–123

Lind A (1983) The Swedish strain of Bcg. Tubercle 64:233–234

Liu J, Tran V, Leung AS, Alexander DC, Zhu B (2009) BCG vaccines: their mechanisms of attenuation and impact on safety and protective efficacy. Hum Vaccin 5:70–78

Lotte A, Wasz-Hockert O, Poisson N, Dumitrescu N, Verron M, Couvet E (1984) BCG complications. Estimates of the risks among vaccinated subjects and statistical analysis of their main characteristics. Adv Tuberc Res 21:107–193

Mahairas GG, Sabo PJ, Hickey MJ, Singh DC, Stover CK (1996) Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M-bovis. J Bacteriol 178:1274–1282

Matsunaga I, Bhatt A, Young DC, Cheng TY, Eyles SJ, Besra GS, Briken V, Porcelli SA, Costello CE, Jacobs WR Jr, Moody DB (2004) Mycobacterium tuberculosis pks12 produces a novel polyketide presented by CD1c to T cells. J Exp Med 200:1559–1569

Mcshane H (2011) Tuberculosis vaccines: beyond bacille Calmette-Guerin. Philos Trans R Soc Lond Ser B Biol Sci 366:2782–2789

Milstien JB, Gibson JJ (1990) Quality control of BCG vaccine by WHO: a review of factors that may influence vaccine effectiveness and safety. Bull World Health Organ 68:93–108

Mostowy S, Tsolaki AG, Small PM, Behr MA (2003) The in vitro evolution of BCG vaccines. Vaccine 21:4270–4274

Munoz ME, Ponce E (2003) Pyruvate kinase: current status of regulatory and functional properties. Comp Biochem Physiol B Biochem Mol Biol 135:197–218

Naka T, Maeda S, Niki M, Ohara N, Yamamoto S, Yano I, Maeyama J, Ogura H, Kobayashi K, Fujiwara N (2011) Lipid phenotype of two distinct subpopulations of Mycobacterium bovis Bacillus Calmette-Guerin Tokyo 172 substrain. J Biol Chem 286:44153–44161

Obayashi Y (1955) Dried BCG vaccine. Monogr Ser World Health Organ, 1–220

Oettinger T, Jorgensen M, Ladefoged A, Haslov K, Andersen P (1999) Development of the Mycobacterium bovis BCG vaccine: review of the historical and biochemical evidence for a genealogical tree. Tuber Lung Dis 79:243–250

Ottenhoff TH (2009) Overcoming the global crisis: "yes, we can", but also for TB ... ? Eur J Immunol 39:2014–2020

Ottenhoff THM, Kaufmann SHE (2012) Vaccines against tuberculosis: where are we and where do we need to go? PLoS Pathog 8

Parish T, Smith DA, Kendall S, Casali N, Bancroft GJ, Stoker NG (2003) Deletion of two-component regulatory systems increases the virulence of Mycobacterium tuberculosis. Infect Immun 71:1134–1140

Pelayo MCG, Uplekar S, Keniry A, Lopez PM, Garnier T, Garcia JN, Boschiroli L, Zhou XM, Parkhill J, Smith N, Hewinson RG, Cole ST, Gordon SV (2009) A comprehensive survey of single nucleotide polymorphisms (SNPs) across Mycobacterium bovis strains and M. bovis BCG vaccine strains refines the genealogy and defines a minimal set of SNPs that separate virulent M. bovis strains and M. bovis BCG strains. Infect Immun 77:2230–2238

Pym AS, Brodin P, Majlessi L, Brosch R, Demangel C, Williams A, Griffiths KE, Marchal G, Leclerc C, Cole ST (2003) Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. Nat Med 9:533–539

Rosenthal SR, Loewinsohne, Graham ML, Liveright D, Thorne G, Johnson V (1961) BCG vaccination against tuberculosis in Chicago. A twenty-year study statistically analyzed. Pediatrics 28:622–641

Sakula A (1983) BCG: who were Calmette and Guerin? Thorax 38:806–812

Salamon H, Kato-Maeda M, Small PM, Drenkow J, Gingeras TR (2000) Detection of deleted genomic DNA using a semiautomated computational analysis of GeneChip data. Genome Res 10:2044–2054

Sherman DR, Guinn KM, Hickey MJ, Mathur SK, Zakel KL, Smith S (2004) Mycobacterium tuberculosis H37Rv: Delta RD1 is more virulent than M. bovis bacille Calmette-Guerin in long-term murine infection. J Infect Dis 190:123–126

Singh A, Crossman DK, Mai D, Guidry L, Voskuil MI, Renfrow MB, Steyn AJ (2009) Mycobacterium tuberculosis WhiB3 maintains redox homeostasis by regulating virulence lipid anabolism to modulate macrophage response. PLoS Pathog 5:e1000545

Sinha A, Gupta S, Bhutani S, Pathak A, Sarkar D (2008) PhoP-PhoP interaction at adjacent PhoP binding sites is influenced by protein phosphorylation. J Bacteriol 190:1317–1328

Stanley SA, Raghavan S, Hwang WW, Cox JS (2003) Acute infection and macrophage subversion by Mycobacterium tuberculosis require a specialized secretion system. Proc Natl Acad Sci U S A 100:13001–13006

Stewart GR, Snewin VA, Walzl G, Hussell T, Tormay P, O'Gaora P, Goyal M, Betts J, Brown IN, Young DB (2001) Overexpression of heat-shock proteins reduces survival of Mycobacterium tuberculosis in the chronic phase of infection. Nat Med 7:732–737

Stewart GR, Wernisch L, Stabler R, Mangan JA, Hinds J, Laing KG, Young DB, Butcher PD (2002) Dissection of the heat-shock response in Mycobacterium tuberculosis using mutants and microarrays. Microbiology 148:3129–3138

Steyn AJ, Collins DM, Hondalus MK, Jacobs WR Jr, Kawakami RP, Bloom BR (2002) Mycobacterium tuberculosis WhiB3 interacts with RpoV to affect host survival but is dispensable for in vivo growth. Proc Natl Acad Sci U S A 99:3147–3152

Takayama K, Wang L, David HL (1972) Effect of isoniazid on the in vivo mycolic acid synthesis, cell growth, and viability of Mycobacterium tuberculosis. Antimicrob Agents Chemother 2:29–35

Tran V, Ahn SK, Ng M, Li M, Liu J (2016) Loss of lipid virulence factors reduces the efficacy of the BCG vaccine. Sci Rep 6:29076

Trunz BB, Fine P, Dye C (2006) Effect of BCG vaccination on childhood tuberculous meningitis and miliary tuberculosis worldwide: a meta-analysis and assessment of cost-effectiveness. Lancet 367:1173–1180

Vallishayee RS, Shashidhara AN, Bunch-Christensen K, Guld J (1974) Tuberculin sensitivity and skin lesions in children after vaccination with 11 different BCG strains. Bull World Health Organ 51:489–494

Wallgren A (1928) Intradermal vaccinations with BCG virus – preliminary note. J Am Med Assoc 91:1876–1881

Walters SB, Dubnau E, Kolesnikova I, Laval F, Daffe M, Smith I (2006) The Mycobacterium tuberculosis PhoPR two-component system regulates genes essential for virulence and complex lipid biosynthesis. Mol Microbiol 60:312–330

Wang S, Engohang-Ndong J, Smith I (2007) Structure of the DNA-binding domain of the response regulator PhoP from Mycobacterium tuberculosis. Biochemistry 46:14751–14761

Wernisch L, Kendall SL, Soneji S, Wietzorrek A, Parish T, Hinds J, Butcher PD, Stoker NG (2003) Analysis of whole-genome microarray replicates using mixed models. Bioinformatics 19:53–61

Wiker HG, Nagai S, Hewinson RG, Russell WP, Harboe M (1996) Heterogenous expression of the related MPB70 and MPB83 proteins distinguish various substrains of Mycobacterium bovis BCG and Mycobacterium tuberculosis H37Rv. Scand J Immunol 43:374–380

Zwaig N, Kistler WS, Lin EC (1970) Glycerol kinase, the pacemaker for the dissimilation of glycerol in Escherichia coli. J Bacteriol 102:753–759

Zwerling A, Behr MA, Verma A, Brewer TF, Menzies D, Pai M (2011) The BCG world atlas: a database of global BCG vaccination policies and practices. PLoS Med 8:e1001012

Antigenic Variation and Immune Escape in the MTBC

9

Joel D. Ernst

Abstract

Microbes that infect other organisms encounter host immune responses, and must overcome or evade innate and adaptive immune responses to successfully establish infection. Highly successful microbial pathogens, including *M. tuberculosis*, are able to evade adaptive immune responses (mediated by antibodies and/or T lymphocytes) and thereby establish long-term chronic infection. One mechanism that diverse pathogens use to evade adaptive immunity is antigenic variation, in which structural variants emerge that alter recognition by established immune responses and allow those pathogens to persist and/or to infect previously-immune hosts. Despite the wide use of antigenic variation by diverse pathogens, this mechanism appears to be infrequent in *M. tuberculosis*, as indicated by findings that known and predicted human T cell epitopes in this organism are highly conserved, although there are exceptions. These findings have implications for diagnostic tests that are based on measuring host immune responses, and for vaccine design and development.

Keywords

Antigenic variation • Immune Escape • T lymphocytes • Epitopes • Antibodies • Tuberculosis

9.1 Immune Evasion by Pathogens

All host organisms possess mechanisms to defend themselves from foreign microbial invaders, especially those that have detrimental effects on the hosts. Bacteria and archaea possess nucleic

J.D. Ernst (✉)
Division of Infectious Diseases and Immunology, Departments of Medicine, Microbiology, and Pathology, New York University School of Medicine, Smilow Building, 9th floor, Rooms 901-907, 522 First Avenue, New York, NY 10016, USA
e-mail: joel.ernst@med.nyu.edu

S. Gagneux (ed.), *Strain Variation in the* Mycobacterium tuberculosis *Complex: Its Role in Biology, Epidemiology and Control*, Advances in Experimental Medicine and Biology 1019,
DOI 10.1007/978-3-319-64371-7_9

acid restriction-modification systems (Makarova et al. 2013) and CRISPR systems (Abudayyeh et al. 2016; Jinek et al. 2012) to defend against invaders such as bacteriophages, while higher organisms have evolved progressively more complex systems of immunity to balance their interactions with beneficial and pathogenic microbes in their environments, to remain fit and capable of reproduction.

In metazoans, immune responses are categorized as innate or adaptive, although the boundaries between these sometimes blur (Cerwenka and Lanier 2016; Cheng et al. 2014; Netea et al. 2016; Saeed et al. 2014). Innate immune responses do not require genetic rearrangements in host somatic cells to respond to a pathogen, and as such are rapidly available and provide broad defense against diverse pathogens (Murphy and Weaver 2016). Innate immune responses are initiated by host molecular recognition of foreign molecular targets, including structural components of bacteria and fungi, certain microbial signalling molecules, and viral nucleic acids. In contrast, adaptive immune responses, represented by antibodies and by T lymphocytes, require genetic rearrangements of immunoglobulin or T cell antigen receptor genes in individual host cells, followed by clonal expansion of antigen-specific B and T lymphocytes and production of antibodies or T cell effector molecules, in response to the individual host's encounter with a microbe bearing a specific repertoire of antigens. Because of the requirement for clonal expansion and differentiation of lymphocytes after specific antigen exposure, initial adaptive immune responses are less rapid than innate immune responses, but a specialized property of adaptive immunity is that of immune memory, which allows for more rapid and higher quality responses when a host encounters a given antigen for a second (or third, or more) time (Murphy and Weaver 2016).

If innate and/or adaptive immune responses operated successfully all of the time, pathological infections would not happen, so the high frequency and diverse nature of infections indicates that host immune responses either fail, or can be evaded by specific mechanisms.

The range of mechanisms of immune evasion by pathogens is vast, and discovery of those mechanisms has yielded considerable understanding of the pathogenesis of specific infections and the roles of specific arms of immunity in control of certain pathogens. Representative mechanisms of immune evasion by pathogens include: (1) masking or modification of pathogen molecules that are typically recognized by innate pattern recognition receptors (Cambier et al. 2014; Vladimer et al. 2012), (2) manipulation (enhancing or inhibiting, depending on the pathogen) of host cell entry, to occupy optimal compartments for replication (Fu and Galan 1999; Hardt et al. 1998); (3) manipulation of intracellular signalling (Xu et al. 2014) or vesicle trafficking (Martinez et al. 2016; Newton et al. 2014) (Mordue et al. 1999; Mordue and Sibley 1997); (4) resistance to or evasion of host microbicidal mechanisms (LaRock et al. 2015; Potter et al. 2012); (5) molecular mimicry, where pathogen molecules resemble host molecules and are not recognized as foreign (Elde and Malik 2009; Hill et al. 2010); (6) antigenic variation, where structural variants of pathogen molecules develop to escape recognition by pre-existing antibodies or T lymphocytes (Deitsch et al. 2009).

9.2 Antigenic Variation in Human Pathogens

Antigenic variation is a common mechanism of immune evasion in diverse pathogens. In a typical example, a pathogen with a given antigenic composition infects a population of hosts that respond by developing adaptive immune responses (antibodies and/or T cells). In many cases, those adaptive immune responses provide protective immunity, which prevents members of that population from subsequent infection. However, since antibodies and T cells have restricted specificities, pathogen variants can develop that are not recognized by the antibodies or T cells in immune hosts, and those pathogen variants can then cause infection in hosts that are immune to the original pathogen (Fig. 9.1). Sequential

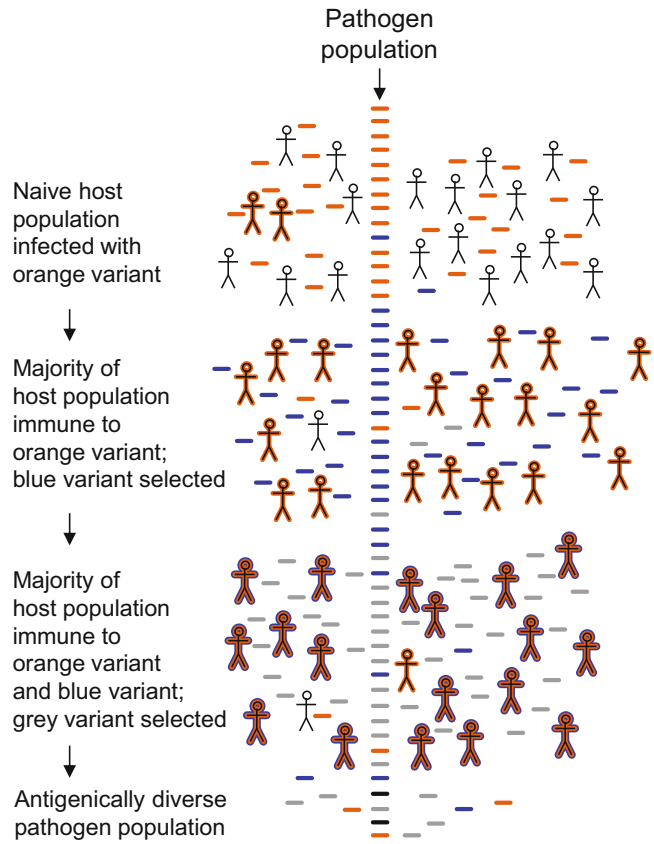

Fig. 9.1 Antigenic variation allows a pathogen to infect hosts immune to original pathogen variant. In this example, a pathogen (*orange*) is able to infect immunologically naive hosts, but not hosts immune to the orange pathogen variant. Blue pathogen variant arises, and can infect host population largely immune to the orange pathogen variant. Infection with blue variant causes additional immunity to blue variant, yet allows infection with grey variant. In each population of hosts, a minor subset of hosts may remain nonimmune, and these can sustain minor variants in the pathogen population (indicated by the mix of colored variants after three rounds of infection and antigenic variation)

rounds of immune escape can generate populations of antigenically diverse pathogens. The consequences of this process of successive generation of pathogen antigenic variants are illustrated by the examples described below.

9.2.1 Influenza

Infection with influenza virus generates T cell and antibody responses, both of which contribute to protective immunity, although it is thought that antibody responses make the more important contribution to preventing reinfection after immunization or infection with a given influenza viral strain. The major target of protective antibody responses to influenza virus is the viral hemagglutinin (HA) protein (Ohmit et al. 2011; Virelizier 1975), one of the two viral proteins exposed on the viral surface. The HA protein is responsible for virus binding to host cells, a requirement for viral entry and replication, and the domain of HA that directly mediates viral binding to host cells is termed the globular head (Tsibane et al. 2012).

The HA globular head is also the predominant target of antibodies generated in response to influenza vaccination or infection (Wrammert et al. 2008). Since the globular head can tolerate a large number and broad range of mutations while retaining its cell-binding function, the virus mutates to escape recognition by anti-HA antibodies that act by blocking virus-host cell binding (Smith et al. 2004; Thyagarajan and Bloom 2014). In addition to mutations in HA that allow escape from antibody recognition, recent evidence indicates that influenza virus can also mutate to escape CD8 T cell recognition of matrix and nucleoproteins (Valkenburg et al. 2013, 2016). Antigenic variation in influenza virus is the result of point mutations in the viral genome, which happen with high frequencies, due to the

lack of proofreading and error correction by the viral RNA polymerase. Antigenic variation in influenza virus can also result from development of reassortant viruses, in which a novel HA antigenic type is acquired from a virus that has not previously circulated in the human population.

The effect of escape mutations and antigenic variation in influenza viruses is significant for human and public health, as it contributes to the need to generate and administer new influenza vaccines nearly every year (Grohskopf et al. 2015), and it allows for global pandemics, when new antigenic variants are generated that are not recognized by antibodies generated by prior vaccination or infection (Bautista et al. 2010).

9.2.2 Human Immunodeficiency Virus (HIV)

Like other pathogens, HIV induces adaptive immune responses that include CD4 and CD8 T cells, as well as antibodies. However, unlike influenza, HIV establishes a chronic infection, because adaptive immune responses are unable to clear the virus and cure the infection. Antigenic variation, which happens at an exceptionally high frequency in HIV, is a major reason that antibodies and T cells are unable to clear the virus.

Although antibodies and CD8 T cell responses are both directed against multiple HIV proteins, these two mechanisms of immunity have distinct effects, depending on the antigenic target: antibody responses drive antigenic variation of the HIV surface envelope glycoprotein (Hraber et al. 2015), while CD8 T cell responses drive antigenic variation of the HIV internal gag protein (Crawford et al. 2007, 2009; Goepfert et al. 2008).

Neutralizing antibodies to HIV develop early after infection, but most of these are specific for the initial infecting virus, and do not cross-react with other viral strains (Li et al. 2009). Antibodies to the HIV envelope glycoprotein are directed at surface-exposed loops of the protein (Pantophlet and Burton 2006), and specific domains within these loops represent hotspots of variation during evolution of HIV (Korber and Gnanakaran 2009). Together, these data indicate that HIV develops escape mutants that allow it to evade the predominant neutralizing antibodies that develop after infection, thus generating considerable sequence and antigenic diversity in the viral envelope.

CD8 T cell responses in HIV-infected individuals bearing HLA-B27, HLA-B57, or HLA-B58 alleles, are especially efficacious in limiting viral replication and progression to AIDS (Kaslow et al. 1996; Lazaryan et al. 2010). Intensive studies have identified several dominant epitope targets of CD8 T cells that are presented by these HLA alleles, and revealed that recognition of those epitopes in the HIV gag protein by CD8 T cells contributes to limiting the HIV viral load in plasma. The same studies have revealed a high frequency of escape mutations in those epitopes, with concomitant decrease in viral fitness, which also contributes to limiting the level of the HIV viral load (Crawford et al. 2007, 2009; Goepfert et al. 2008). That the observed mutations represent escape from CD8 T cell recognition is supported by a high rate of reversion to the ancestral sequence in the involved epitopes when HIV is transmitted to an individual lacking a protective HLA allele.

Although antigenic diversity in HIV can develop as the consequence of viral recombination (Vuilleumier and Bonhoeffer 2015), the extremely high mutation rate of HIV (3×10^{-5} per base per round of replication from polymerase errors (Roberts et al. 1988) and approximately 100-fold higher rates due to host cytidine deaminases (Cuevas et al. 2015) indicates that most antigenic variation in HIV is attributable to point mutations.

9.2.3 *Streptococcus pneumoniae*

Streptococcus pneumoniae is a gram-positive bacterial pathogen that inhabits respiratory mucosal surfaces and can cause potentially-fatal invasive infections including pneumonia,

bacteremia and sepsis, and meningitis. Immunity to *S. pneumoniae* depends on specific antibodies, antibody-mediated phagocytosis (Hosea et al. 1981) or netosis (Beiter et al. 2006), and intracellular or extracellular killing of the bacteria, respectively.

Antigenic variation in *S. pneumoniae* has been known for more than 100 years (Dochez and Gillespie 1913), and its significance in protective immunity has been known for nearly as long (Cooper et al. 1932). As a predominantly extracellular pathogen, the most important mechanism of adaptive immunity to *S. pneumoniae* is mediated by antibodies: individuals deficient in anti-pneumococcal IgG are especially susceptible to invasive infection, and administration of anti-pneumococcal IgG reduces the risk (Chapel et al. 1994). The dominant target of protective antibodies against *S. pneumoniae* is the bacterial surface polysaccharide capsule (Avery and Heidelberger 1925), which exists in ~90 antigenically distinct forms.

Studies performed to evaluate the efficacy of vaccination with *S. pneumoniae* capsular polysaccharide antigens have provided strong evidence for antigenic variation, driven by population immunity. Introduction of a pneumococcal polysaccharide conjugate vaccine containing seven antigenic types markedly reduced the frequency of invasive infections due to the antigenic variants included in the vaccine in several distinct populations (Feikin et al. 2013; Hsu et al. 2009; Pichon et al. 2013; Richter et al. 2013). This indicates that the vaccine was highly efficacious in producing protection against infection, and thus exerted strong immunological selection pressure on the bacterial population. However, each of these studies also revealed increases in the frequency of infections due to *S. pneumoniae* expressing polysaccharide antigens that were not included in the vaccine, indicated that antigenic variation allows escape from vaccine-induced immunity (Feikin et al. 2013; Hsu et al. 2009; Pichon et al. 2013; Richter et al. 2013).

Unlike in influenza virus or HIV, antigenic variation in *S. pneumoniae* is not predominantly due to immune selection of point mutants, but is the result of horizontal gene transfer, in which a given virulent bacterial strain acquires a genomic locus for biosynthesis of an antigenically distinct capsular polysaccharide from another strain (Wyres et al. 2013). Although capsular antigen switching by horizontal gene transfer is characteristic of *S. pneumoniae*, selection pressure exerted by vaccine-induced immunity amplifies the effect and impact of this mechanism of antigenic variation (Croucher et al. 2011).

9.2.4 Antigenic Variation in Other Major Global Pathogens

Although the mechanisms and significance of antigenic variation are best characterized for the examples described above, antigenic variation is a common strategy in multiple other pathogens (Table 9.1). The importance of antigenic variation to pathogen survival, replication, and transmission is demonstrated by its employment by a broad range of pathogens, including viruses, intracellular and extracellular bacteria, and intracellular and extracellular eukaryotic parasites. Moreover, multiple molecular mechanisms underlie antigenic variation, including nucleotide substitution (point mutation), gene conversion, recombination to place different copies of a multigene family into a transcriptionally-active site, alternative transcription without recombination, and transcriptional selection by phase variation. Finally, antigenic variation involves targets of antibody, CD4, and CD8 T cell recognition. It is also likely that additional examples and mechanisms for this biologically important pathogen strategy remain to be discovered.

9.3 *M. tuberculosis* Immune Evasion

As a highly successful pathogen, *M. tuberculosis* possesses numerous mechanisms for manipulating and modulating immune responses to opti-

Table 9.1 Antigenic variation in other human pathogens

Pathogen	Mechanism of protective immunity	Antigenic target(s)	Mechanism of antigenic variation	References
Hepatitis C virus	CD8 T cells	Multiple structural and nonstructural proteins	Point mutations	Erickson et al. (2001), Kasprowicz et al. (2010), and Nivarthi et al. (2014)
Neisseria gonorrhoeae	Antibodies	Lipooligosaccharide (LOS)	Phase variation	Yang and Gotschlich (1996)
		Opacity (Opa) proteins	Multiple loci with phase variation	Bhat et al. (1991)
		Pilin	Recombination	Segal et al. (1986)
Streptococcus pyogenes	Antibodies	M protein	Point mutations	Lannergard et al. 2011; Persson et al. (2006)
		SMEZ protein		Hoe et al. 1999; Proft et al. (2000)
Borrelia relapsing fever agents (e.g., *B. hermsii*)	Antibodies	Vlp and Vsp lipoproteins	Gene conversion; alternative transcription of multicopy genes	Barbour et al. (2006), Plasterk et al. (1985), and Stoenner et al. (1982)
Borrelia Lyme disease agents (e.g., *B. burgdorferi*)	Antibodies	VlsE lipoprotein	Gene conversion	Lawrenz et al. (2004) and Liang et al. (2004)
Anaplasma phagocytophilum/marginale	Antibodies, CD4 T cells	MSP2	Segmental gene conversion	Brown et al. (2003)
Trypanosoma brucei	Antibodies	Surface glycoprotein	Recombination; transcription switching	Morrison et al. (2009)
Plasmodium falciparum	Antibodies	Pfemp1	Multicopy genes with transcriptional switching	Claessens et al. (2014), Peters et al. (2002), and Ralph et al. (2005)

mize its survival, replication, and transmission. Unlike viruses, bacteria such as *M. tuberculosis* can respond to distinct environmental conditions and signals to optimize their gene expression and deploy mechanisms that optimally suit the bacteria in a given context. For example, during certain parts of the infection cycle, the bacteria may benefit from going undetected by host mechanisms, while during other parts of the infection cycle, *M. tuberculosis* may gain most by inducing vigorous inflammatory and immune responses (Ernst 2012). While a comprehensive review of mechanisms of immune evasion is beyond the scope of this chapter, selected examples are given below, to provide context for the main points of the chapter.

9.3.1 *M. tuberculosis* Manipulation of Innate Immunity

Innate immune responses provide rapidly-available responses to the presence of diverse pathogens, including *M. tuberculosis*. In particular, innate immune responses to bacteria, including *M. tuberculosis*, include production of proinflammatory and regulatory cytokines, as well as chemotactic cytokines (also known as chemokines). Chemokines and their receptors govern the types and numbers of cells recruited to a site of infection or inflammation, and contribute to early responses to mycobacterial infection. As one example, C-C Chemokine Receptor-2 (CCR2), which is expressed by

mononuclear cells and is important for their trafficking to sites of inflammation, plays a complex role in innate immune responses to *M. tuberculosis*, as it is essential for control of infection but also paradoxically contributes to pathogenesis. In the former case, CCR2-dependent mononuclear cell recruitment to the lungs after *M. tuberculosis* infection is crucial for accumulation of mononuclear cell-derived dendritic cells in the lungs (Peters et al. 2001, 2004), which in turn become infected and are required for activation of antigen-specific CD4 T cells that ultimately control infection (Wolf et al. 2007, 2008). In the latter case, *M. tuberculosis* infection uses CCR2-dependent monocyte recruitment to generate a population of monocyte-derived cells that actively supports intracellular bacterial growth and spread (Antonelli et al. 2010). The factors that determine whether the host-beneficial or the pathogen-beneficial effects of CCR2-dependent cell trafficking predominate remain to be determined. Together, these data indicate that innate immune responses induced by *M. tuberculosis* that involve CCR2-dependent cell recruitment are critical determinants of the course of infection. It is not surprising, then, that pathogenic mycobacteria possess mechanisms to manipulate cell recruitment to their own advantage.

Mycobacterial manipulation of monocyte/macrophage recruitment is mediated by masking of bacterial Toll-like receptor agonist molecules by the lipoglycan, pthiocerol dimycoceroserate (PDIM), thus reducing recruitment of macrophages with mycobactericidal potential (Cambier et al. 2014). While using PDIM to reduce recruitment of mycobactericidal macrophages, mycobacteria use surface phenolic glycolipid (PGL) for CCR2-dependent recruitment of monocyte/macrophages that support intracellular growth of the bacteria and thereby promote infection (Cambier et al. 2014).

In an additional mechanism to manipulate innate immune responses for its own benefit, *M. tuberculosis* uses a cytoplasmic signalling pathway involving the DNA sensor, cyclic GMP-AMP synthase (cGAS) and its downstream signalling molecule, stimulator of interferon genes (STING) to induce expression of type I interferon (Watson et al. 2015; Wiens and Ernst 2016). Type I interferons act as regulatory cytokines, and are implicated in promoting progressive infection with *M. tuberculosis* at least in part by suppressing expression of the proinflammatory cytokine, interleukin-1 (Mayer-Barber et al. 2011, 2014). Type I interferons regulate a large number of genes, and an interferon-responsive transcriptional signature is found in the context of active tuberculosis in humans (Berry et al. 2010; Bloom et al. 2012; Maertzdorf et al. 2011; Ottenhoff et al. 2012).

9.3.2 *M. tuberculosis* Manipulates the Intersection of Innate and Adaptive Immunity

In experimental infections with *M. tuberculosis*, the bacterial population expands in the lungs until the onset of adaptive (e.g., T cell) responses, which are able to arrest progressive growth of the bacteria (Wolf et al. 2008). A remarkable characteristic of tuberculosis in humans and in experimental animals is that the time required for development of T cell responses is much longer than for T cell responses to other infections. In humans with close, but limited ($\leq$24 h), exposure to a person with active tuberculosis, the average time to develop a detectable cellular immune response is 6 weeks (Poulsen 1950); this contrasts with an average of 7 days for humans to generate T cell responses after experimental exposure to a live virus (the vaccine strain of yellow fever virus) (Miller et al. 2008). The onset of CD4 and CD8 T cell responses after aerosol infection with *M. tuberculosis* is similarly delayed, and averages approximately 17 days (Chackerian et al. 2002; Wolf et al. 2008), compared with 3–5 days after experimental influenza infection

(Lawrence and Braciale 2004; Lawrence et al. 2005). The rate-limiting step in priming naive antigen-specific CD4 T cells after aerosol infection of mice with *M. tuberculosis* is the acquisition of bacteria by dendritic cells in the lungs, which happens 8–10 days after aerosol infection (Blomgran et al. 2012; Blomgran and Ernst 2011; Wolf et al. 2007). This is followed by dendritic cell transport of the bacteria from the lungs to the local lymph node (Wolf et al. 2008) and transfer of bacterial antigens (but not the bacteria) (Samstein et al. 2013; Srivastava and Ernst 2014) to resident dendritic cells, which process and present the mycobacterial antigens to antigen-specific CD4 T cells. The latter step (transfer of bacterial antigens from migratory to resident dendritic cells) compensates for *M. tuberculosis* inhibition of MHC class II antigen presentation by the infected cells themselves, but the compensation is incomplete (Srivastava et al. 2016). In mice, the long delay before appearance of antigen-specific T cells in the lungs allows ~40,000-fold expansion of the bacterial population, indicating that *M. tuberculosis* benefits from delaying the onset of T cell responses.

The mechanisms of immune evasion and manipulation by *M. tuberculosis* given above are only a partial sampling of the numerous mechanisms employed by this persistent pathogen. Additional mechanisms, and a comprehensive review of mechanisms of immunity to *M. tuberculosis* in humans and experimental animals are found in (O'Garra et al. 2013).

9.4 Antigenic Variation in *M. tuberculosis*

The well-documented ability of *M. tuberculosis* to survive in immunocompetent hosts and reactivate to cause active disease decades later (Lillebaek et al. 2002), together with evidence that infection with *M. tuberculosis* does not reliably confer protection from reinfection (Caminero et al. 2001; du Plessis et al. 2001; Interrante et al. 2015), indicates that this pathogen has a very effective armamentarium for evading the human immune responses that successfully eradicate other pathogens. Considering the numerous precedents provided by pathogens that evade immunity through antigenic variation, it is important to understand the role of antigenic variation in the success of *M. tuberculosis* as a pathogen.

9.4.1 Initial Studies of Antigen Sequence Variation in *M. tuberculosis*

Early evidence that antigenic diversity may be low in *M. tuberculosis* was derived from sequence analysis of 24 antigen-coding genes in 16 strains of *M. tuberculosis* (Musser et al. 2000). That study found that 19 of the 24 genes studied lacked any sequence variation; this included 5 of the 8 PE or PPE genes selected for analysis. Although the study was limited in its scope and it did not distinguish antigens recognized by antibodies or by T cells, or focus on specific epitopes in the proteins analyzed, it provided valuable information regarding the sequence diversity of the *M. tuberculosis* antigens that were known at that time.

Among the unique insights gained from the genome sequence of *M. tuberculosis* H37Rv was the discovery of the genes encoding proteins of the PE/PPE family, including the PE_PGRS family (Cole et al. 1998). Based on precedents in other pathogens, the presence of this large family of proteins suggested that one of their functions was in antigenic variation. This hypothesis prompted several analyses of sequence variation in individual members of the *pe_pgrs* family. Analysis of *pe_pgrs33* in 123 strains revealed a higher frequency of variants than in other *M. tuberculosis* genes studied. Using the H37Rv reference sequence, variants of *pe_pgrs33* were found in 84 of the 123 strains, and represented 25 distinct variants (Talarico et al. 2005). Of the 25 variants, 13 involved insertions or deletions (collectively termed indels), and 12 were single nucleotide polymorphisms (SNPs). All of the insertions or deletions were found

in the PRGS domain, as were 9 of the 12 SNPs; synonymous and nonsynonymous SNPs were equally frequent. Similar observations were obtained from sequencing the *pe_pgrs16* and *pe_pgrs26* genes in 200 *M. tuberculosis* isolates (Talarico et al. 2008). While these genes were similar to each other and to *pe_pgrs33* in their frequency of sequence variants (indels and SNPs), they differed in the effect of their indels: in 40% of the strains, the indels in *pe_pgrs16* caused frameshifts, while the indels in *pe_pgrs26* caused frameshifts in only 2% of the strains. Together, these studies provided evidence for a high frequency of sequence variants in members of the *pe_pgrs* gene family, although they did not reveal whether the sequence variants altered immune recognition of the proteins by either antibodies or T cells. A subsequent analysis of sequence variants of the *ppe18* gene, combined with independently-reported T cell epitope mapping data, predicted that some, but not all, of the nSNPs were contained in epitope domains of PPE18 (Hebert et al. 2007).

9.4.2 Genomewide Evaluation of Antigenic Variation in *M. tuberculosis*

Two major advances have facilitated comprehensive, high-resolution analysis of antigenic variation in *M. tuberculosis*, at least for human T cell recognition. The first advance is the wide availability of economical genome sequencing and bioinformatic analyses, which is dealt with in depth in other chapters in this volume. The other advance is the development and readily-accessed Immune Epitope Database (IEDB; www.iedb.org), funded by the U.S. National Institutes of Health. The IEDB identifies, stores, and provides analytical tools for T cell and B cell epitopes from a variety of sources of antigens, including infectious agents such as *M. tuberculosis*. As of August, 2016, the IEDB contains information on 2148 human T cell epitopes of *M. tuberculosis*, derived from 440 antigens. The sources of epitope data in the IEDB include published papers, meeting abstracts, and the results of several separately-funded epitope discovery projects. The availability of information on *M. tuberculosis* T cell epitopes in the IEDB has markedly facilitated studies revealing the frequency and nature of antigenic variation in this organism. The paucity of information on B cell antigens and epitopes of *M. tuberculosis* limits the availability of this knowledge in the IEDB, so the preponderance of analyses have focused on T cell epitopes. T cell epitopes are short (8–20 amino acids) peptide fragments of proteins (in the present case, from *M. tuberculosis*) generated by intracellular (host) proteolysis, bound to specific MHC (HLA in human) class I or class II alleles, for recognition by CD8 or CD4 T cells, respectively.

The first genomewide analysis of human T cell epitope sequence diversity examined 491 epitopes (from 78 antigens) available in the IEDB in 21 strains representative of the six major global lineages of the *M. tuberculosis* complex (MTBC) known at the time (Comas et al. 2010). Generation and comparison of the genome sequences of these strains (plus that of *M. canettii*) revealed an expected result: that essential genes (as defined in Sassetti et al. 2003; Sassetti and Rubin 2003) are more conserved than are nonessential genes. The unexpected result was that the great majority (468 of 491; 95%) of the experimentally-verified human T cell epitopes analyzed are completely conserved, with no amino acid sequence variants in these representative strains of the MTBC. Additional analyses revealed that the ratio of the rates of nonsynonymous SNPs (nSNPs) to synonymous SNPs (sSNPs) (dN/dS) is lower for the identified T cell epitopes than for nonessential genes, and is lower for the epitope than for the nonepitope domains in their source protein. These findings indicate that the known human T cell epitopes of *M. tuberculosis* are not under diversifying selection, implying that little, if any, selection pressure is exerted on these epitopes by human T cell recognition. This unique finding, that antigenic variation is rare in *M. tuberculosis,* provides evidence that the nature of the interaction of this pathogen with its human hosts is

distinct from that of other pathogens, and that principles established with other pathogens may not reliably apply to *M. tuberculosis*.

9.4.3 PE_PGRS Proteins as Variable Antigens

A limitation of the aforementioned study is that the sequencing technology used generates short reads, and this prohibited analysis of epitopes located in protein members of large families with long stretches of sequence identity, such as the PE/PPE genes. As a consequence, it remained possible that the PE/PPE family is a source of antigenic (and epitope) diversity that could not be identified with the genome sequencing approach that was used. This possibility was especially plausible, in light of the evidence discussed above that certain members of the PE_PGRS family exhibit a high frequency of sequence variation.

To determine whether PE_PGRS proteins are a source of antigenic variation in *M. tuberculosis*, we used Sanger dideoxynucleotide chain termination sequencing to analyse 27 *pe_pgrs* genes in 94 phylogenetically diverse strains of *M. tuberculosis* (Copin et al. 2014). These analyses revealed that the nucleotide diversity in the *pg_pgrs* gene family is higher than the genome average in the MTBC, consistent with the results of the studies of *pe_pgrs16*, *pe_pgrs26*, and *pe_pgrs33* (Talarico et al. 2005, 2008). However, the comparison of 27 *pe_pgrs* genes revealed previously-unsuspected properties of this gene family. First, the individual members of the *pe_pgrs* family differ greatly in their nucleotide diversity, with some members (e.g., *pe_pgrs38*) much more diverse ($\pi > 0.0015$) than the genomewide average for the MTBC ($\pi \sim 0.0003$), while other members are even more conserved than the genome as a whole (e.g., *pe_pgrs23*; $\pi < 0.00006$). By this analysis, which does not take indels into account, the previously-characterized *pe_pgrs26* gene is among the group with the highest nucleotide diversity ($\pi \sim 0.00089$), while *pe_pgrs33* is close to the genome average. The vast range of nucleotide diversity extended to a vast range of ratios of substitution rates at nonsynonymous and synonymous sites (dN/dS): as extreme examples, *pe_pgrs59* exhibited evidence of exceptionally strong purifying selection (dN/dS $\sim$0.12), while *pe_pgrs23* is under potent diversifying selection (dN/dS ∞; no sSNPs identified). In a similar manner, individual members of the *pe_pgrs* family exhibited a wide range of frequencies of indels; there was little concordance between nucleotide diversity and indel diversity in these genes. Notably, nSNPs and indels were clustered in the PGRS domains of the proteins, while the N-terminal ($\sim$110 amino acid) PE domains were significantly more conserved.

To directly determine whether the PE_PGRS proteins are a source of antigenic variation in *M. tuberculosis*, we used bioinformatic analyses to predict human T cell epitopes, using HLA alleles found at the highest frequencies in six major human population groups. This revealed that the great majority of the predicted epitopes are found in the PE domain, with very few in the PGRS domain of any of the proteins (Fig. 9.2). The validity of the epitope predictions was confirmed using synthetic peptides designed from the sequences of the predicted epitopes in the PE domains, which activated T cells from blood of patients with active pulmonary tuberculosis (Copin et al. 2014). Together the sequencing and epitope prediction efforts revealed that, while structural variants of PE_PGRS proteins are common, they are clustered in the PGRS domains, while the human T cell epitopes are clustered in the PE domains.

These findings indicated that the PE_PGRS proteins are not a major source of variation of human T cell epitopes, in accord with the findings that human T cell epitopes in *M. tuberculosis* are highly conserved. However, the findings are consistent with the possibility that the PE_PGRS proteins are a source of antigenic variation for antibody recognition, as other studies have determined that the PGRS domains are targets of antibodies in humans infected with *M. tuberculosis* (Cohen et al. 2014; Delogu and Brennan 2001; Koh et al. 2009).

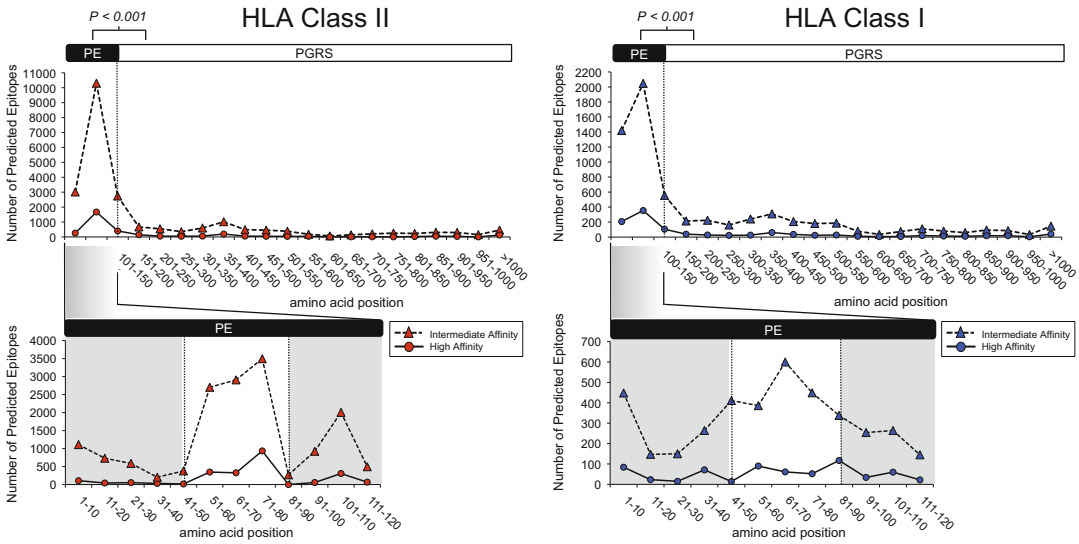

Fig. 9.2 Human T cell epitopes are concentrated in the PE domain of PE_PGRS proteins. The amino acid sequences of all 64 annotated PE_PGRS were used for in silico epitope prediction. The graph represents the number of high binding affinity (IC50 <50 nM; *solid line*) and intermediate binding affinity (IC50 <500 nM; *dashed line*) epitopes for the selected CD4 (MHC class II and CD8 (MHC class I alleles, by amino acid position of the proteins). The lower part of each panel shows an expanded view of the PE domain containing a high density of predicted epitopes (Adapted from MBio 5 (1):e00960–00913)

9.4.4 Esx Proteins as Variable Antigens

The short-read genome sequencing technology used in the original study of epitope diversity also necessitated exclusion of certain *Esx* genes, which encode closely-related immunogenic low molecular weight secreted proteins (Comas et al. 2010). An independent study of these proteins in 108 phylogenetically-diverse strains used Sanger dideoxynucleotide chain termination sequencing to identify sequence variants in 23 genes of the *Esx* family (Uplekar et al. 2011). This revealed an overall SNP frequency similar to the *M. tuberculosis* genomewide average, and identified a total of 109 unique SNPs (50 sSNPs and 59 nSNPs) in the 23 genes. *Esx* genes in the ESX-1 to ESX-4 secretion system loci contained few SNPs, while genes that belong to the Mtb9.9 and QILSS subfamilies, including those in the ESX-5 secretion system locus, contained the majority of the variants. Coincidentally, these latter subfamilies exhibit an especially high degree of amino acid sequence similarity (93–98%) in paralogous protein members. Comparison of the identified nSNPs with previously-identified human T cell epitopes in Esx proteins revealed one amino acid substitution in an epitope in EsxB (CFP-10); otherwise there were no sequence variants found in the 80 identified epitopes in EsxA (ESAT-6) or EsxB. In contrast, of the 13 epitopes identified in the other members of the Esx protein family, 9 contained one or more amino acid substitutions, including 1 in EsxH, which was also identified in the initial genomewide epitope analysis (Comas et al. 2010). This study provided valuable insight into sequence variation in the Esx proteins that was not revealed by the initial genomewide epitope analysis. It confirmed that the high level of epitope conservation extends to epitopes in the immunodominant EsxA and EsxB proteins, and revealed the presence of potentially-important sequence variants in epitopes of other members of the Esx protein family.

9.4.5 Population genomics and computational and experimental immunology to identify and quantitate human T cell epitope sequence variation in the *M. tuberculosis* complex

An additional limitation to the first genomewide analysis of human T cell epitope variation is that it used a database of experimentally-verified peptide epitopes, and was thus subject to factors that might skew epitope discovery toward conserved epitopes. For example, studies of *M. tuberculosis* antigen or epitope discovery typically use the H37Rv strain (from Lineage 4) as the source of antigen or sequences for preparation of synthetic peptides or recombinant antigens. If studies of T cell recognition or epitope mapping are performed with cells from individuals infected with bacteria from a different lineage (e.g., Lineage 6), then antigens or epitopes that are conserved in Lineages 4 and 6 are most likely to be discovered.

To circumvent the potential for epitope discovery favoring conserved epitopes, we took an alternative approach, using comparative genomics to identify the most diverse regions of the *M. tuberculosis* genome, reasoning that antigens or epitopes undergoing diversifying selection by T cell recognition would exhibit nucleotide diversities and dN/dS ratios higher than the genomewide average.

To identify and enumerate potential epitopes undergoing diversifying selection by human T cell recognition, we first analyzed 3774 coding regions in newly-generated whole genome sequences of 216 human-adapted strains representative of the seven main lineages of the *M. tuberculosis* complex (Coscolla et al. 2015), and identified loci with the highest nucleotide diversity in the 216 genomes. We then focused our analyses on coding regions exhibiting high rates of nSNPs. For this, we selected the most variable 5% of the genes (N = 189) and determined that the dN/dS in these genes ranged from 0 to 4.21; 88 of the 189 genes showed a dN/dS > 1, indicating evidence of diversifying selection. From these 88 genes, we selected those with at least one nSNP present in an entire MTBC lineage; this narrowed the analysis to seven genes, which are associated with various functional categories, though none of the protein products are known or predicted to be secreted.

To explore whether sequence diversity in the seven genes of interest could be related to human T cell recognition, we first computationally predicted human CD4 and CD8 T cell epitopes in the protein products of these genes using HLA class I and class II alleles that are prevalent in diverse human populations. This predicted a mean of 207 high-affinity epitopes per protein for HLA class I and 150 epitopes per protein for HLA class II; comparison of the locations of the predicted epitopes and the nSNPs in the seven genes of interest revealed that 51 of the 56 (91%) nSNPs identified coincided with predicted CD4 and/or CD8 T cell epitopes. We then examined the impact of the identified nSNPs (and corresponding amino acid substitutions) on the capacity of the predicted epitope peptides to bind to selected HLA alleles. On average, a given naturally-occurring amino acid substitution decreased the binding of 25% and 18% of the predicted peptide:HLA class I and peptide:HLA class II interactions, respectively. The naturally occurring sequence variation could also lead to an increase in the number of HLA alleles capable of binding the corresponding epitope peptide, indicating that the naturally-occurring sequence variants in predicted epitopes in the protein products of the seven genes of interest could result in either loss or gain of recognition by human T cells.

Since computational prediction of epitopes has greater sensitivity than specificity, it was essential to determine whether the new predicted T cell epitopes we identified with naturally-occurring sequence variants are genuinely recognized by T cells of humans with tuberculosis. Of the putative epitopes we identified, we selected those predicted to bind to HLA class I and class II alleles prevalent in the Gambia, and studied their ability to stimulate interferon gamma production in a modified diluted whole blood assay, performed in the Gambia using cells from 82 individuals

with active pulmonary tuberculosis. For this experimental verification, we used 30 peptides corresponding to 14 ancestral and 16 variant sequences from the seven proteins of interest. Fifty-two (63%) of the subjects responded to at least one candidate epitope peptide; cells from individual subjects responded to an average of 3 of the 30 candidate epitope peptides (ancestral and/or variant form), although cells of some subjects responded to as many as 22 of the peptides. An especially important finding is that an average of 72% of the responding subjects for a given candidate epitope exhibited differential responses to the ancestral compared with the variant sequences of each of the 14 candidate epitopes, indicating that the naturally-occurring amino acid substitutions in these epitopes have a functional impact on human T cell recognition and activation.

The most important conclusions from this study are the confirmation that the majority of human T cell epitopes in *M. tuberculosis* are conserved, and that a small number of antigens and epitopes in this pathogen exhibit evidence of diversifying selection and antigenic variation. Therefore, unlike other pathogens that employ antigenic variation as a mechanism to evade human adaptive immune responses, antigenic variation appears to be exceptional in *M. tuberculosis*.

9.5 Practical Implications

The findings that antigenic variation is unusual, and has little if any effect on the widely-studied immunodominant protein antigens of *M. tuberculosis*, has practical implications that differ in specific contexts.

In the context of immunodiagnostic assays, whether they are designed to detect the antigens or responses to the antigens, antigen and epitope conservation implies that specific diagnostic reagents can be considered universal. That is, antigenic variation has little, if any, effect on the sensitivity of assays that employ a single antigenic sequence. While this can be a confident conclusion for assays involving T cell recognition and/or antigens identified specifically as T cell antigens, studies to date are insufficient to conclude that the same rule would apply for assays that detect antibody responses.

The identification of rare antigens and epitopes that do exhibit naturally-occurring sequence variation may provide opportunities to enhance studies of TB epidemiology and transmission, especially in regions where strains from several bacterial lineages coexist. For example, it is currently only possible to detect reinfection with *M. tuberculosis* if a person progresses to active disease and the disease-causing isolate can be identified and found to be distinct from the isolate that caused an initial episode of disease. It is possible that use of the newly-identified antigens and epitopes with sequence variants can be used in T cell assays to detect reinfection with a new strain based on T cell responses to distinct epitope variants, even in the absence of progression to active disease. Development and validation of such assays will require close collaborations between individuals with expertise in optimizing conditions and readouts in T cell assays and experts in TB epidemiology.

The practical implications of antigen and epitope conservation for TB vaccine development are complex. On the surface, that there is little sequence variation in the immunodominant antigens of *M. tuberculosis* that are included in current candidate TB vaccines could be regarded as a favorable finding. That is, the results imply that it may not be necessary to develop distinct TB vaccines for different populations, and that a polyvalent TB vaccine is unnecessary, to account for bacterial strain-dependent antigenic variation. However, the finding that most currently-favored TB vaccine antigens are not under diversifying selection pressure from human T cell recognition may be interpreted as evidence that *M. tuberculosis* has effective mechanisms to disarm T cell responses to these antigens, potentially limiting the efficacy of vaccine-induced T cells. Whether this should be taken into account and given significant weight in TB vaccine development deserves in-depth consideration and experimental study. One consideration is that one or more of the seven newly-discovered antigens that exhibit naturally-

occurring sequence diversity might actually be superior TB vaccine antigens, in light of the evidence that their recognition is sufficient to select escape mutants. In that vein, the finding that none of the identified epitopes in those antigens has more than three sequence variants implies that the bacteria have limited options for immune escape by further mutation and antigenic variation.

The suggestion that antigen and epitope conservation prevails in *M. tuberculosis* because human T cell recognition of these antigens provides an evolutionary advantage to the bacteria also deserves consideration and in-depth study, to avoid pursuing TB vaccine strategies that induce T cell responses that benefit the bacteria (for example, by promoting transmission).

9.6 Remaining Questions and Future Directions

Although studies of antigenic variation and immune escape in the genomic era of tuberculosis research have revealed considerable valuable and unexpected information, many important matters are only partially understood.

First, it is of utmost importance to determine whether the low frequency of antigenic variation in the majority of *M. tuberculosis* antigens is truly due to little diversifying selection pressure exerted by human T cell recognition. If this is so, then it will be especially important to identify the most important mechanisms used by *M. tuberculosis* to evade effector T cell responses and to develop ways to overcome or circumvent them.

Second, it is important to determine whether human T cell responses to *M. tuberculosis* antigens that do exhibit evidence of escape mutation are more advantageous to the host than are responses to conserved antigens. This may be studied in at least two ways. One is through prospective studies of individuals with latent TB, to determine whether T cell responses to the variable antigens are more advantageous than T cell responses to conserved antigens, as manifest by differential association with subsequent progression to active TB disease. Another is through experimental vaccine studies in relevant model systems, in which protection conferred by responses to vaccines containing different antigens can be assessed.

Third, it is currently difficult to reconcile *M. tuberculosis* T cell epitope conservation with the vast diversity of HLA alleles and differential HLA allele frequencies in distinct populations. On the surface, it would seem that a given epitope might be conserved in a population with a given prevalent HLA allele, but that same epitope would be subject to genetic drift in a population where that HLA allele is rare or absent. One potential explanation for this is that conservation of epitope sequences is attributable to mechanisms other than T cell recognition, that is, that *M. tuberculosis* epitopes have low inherent mutational tolerance (Thyagarajan and Bloom 2014). While there is currently no direct evidence that allows exclusion of this as a predominant mechanism for epitope sequence conservation, several observations weigh against it: *M. tuberculosis* T cell epitopes are not concentrated in essential genes, nor are they disproportionately located in or near functional protein domains such as enzyme active sites (Comas et al. 2010); and *M. tuberculosis* T cell epitopes are found in diverse structural domains of proteins, including beta sheets, alpha helices, and loops, implying that their ability to tolerate sequence variants is not constrained by protein structural requirements (J. Ernst, unpublished observation). One alternative possibility is that *M. tuberculosis* epitopes are more promiscuous for binding diverse HLA alleles than are epitopes in other pathogens (McKinney et al. 2013; Paul et al. 2015), and that HLA allele diversity has little influence on responses to *M. tuberculosis* T cell epitopes. This latter possibility is consistent with the observation that no single HLA class I or class II locus or allele is reproducibly associated with susceptibility or resistance to tuberculosis.

Fourth, rigorous studies of the importance of antibody responses in the immune control of tuberculosis are needed, to determine whether the principles recently established for *M. tuberculosis* T cell epitopes also apply to B cell/antibody

epitopes, and whether TB vaccines that induce antibody responses should be a high priority for development. The finding that the PGRS domains of certain PE_PGRS proteins are under potent diversifying selection by a process other than T cell recognition, coupled with the evidence that PGRS domains are targets for antibody recognition, suggests that certain of these proteins are worthy of greater investigation as potential TB vaccine antigens.

In conclusion, *M. tuberculosis* differs from other successful pathogens in the rarity of its employment of antigenic variation. Further investigation is needed to understand the basis of antigen and epitope conservation in *M. tuberculosis*, and to exploit the resulting knowledge to impact the global epidemic of human TB.

References

Abudayyeh OO, Gootenberg JS, Konermann S, Joung J, Slaymaker IM, Cox DB, Shmakov S, Makarova KS, Semenova E, Minakhin L, Severinov K, Regev A, Lander ES, Koonin EV, Zhang F (2016) C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science (New York, NY). doi:10.1126/science.aaf5573

Antonelli LR, Gigliotti Rothfuchs A, Goncalves R, Roffe E, Cheever AW, Bafica A, Salazar AM, Feng CG, Sher A (2010) Intranasal poly-IC treatment exacerbates tuberculosis in mice through the pulmonary recruitment of a pathogen-permissive monocyte/macrophage population. J Clin Invest 120(5):1674–1682. doi:10.1172/jci40817

Avery OT, Heidelberger M (1925) Immunological relationships of cell constituents of pneumococcus: second paper. J Exp Med 42(3):367–376

Barbour AG, Dai Q, Restrepo BI, Stoenner HG, Frank SA (2006) Pathogen escape from host immunity by a genome program for antigenic variation. Proc Natl Acad Sci U S A 103(48):18290–18295. doi:10.1073/pnas.0605302103

Bautista E, Chotpitayasunondh T, Gao Z, Harper SA, Shaw M, Uyeki TM, Zaki SR, Hayden FG, Hui DS, Kettner JD, Kumar A, Lim M, Shindo N, Penn C, Nicholson KG (2010) Clinical aspects of pandemic 2009 influenza A (H1N1) virus infection. N Engl J Med 362(18):1708–1719. doi:10.1056/NEJMra1000449

Beiter K, Wartha F, Albiger B, Normark S, Zychlinsky A, Henriques-Normark B (2006) An endonuclease allows Streptococcus pneumoniae to escape from neutrophil extracellular traps. Curr Biol: CB 16(4):401–407. doi:10.1016/j.cub.2006.01.056

Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, Wilkinson KA, Banchereau R, Skinner J, Wilkinson RJ, Quinn C, Blankenship D, Dhawan R, Cush JJ, Mejias A, Ramilo O, Kon OM, Pascual V, Banchereau J, Chaussabel D, O'Garra A (2010) An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. Nature 466(7309):973–977. doi:10.1038/nature09247

Bhat KS, Gibbs CP, Barrera O, Morrison SG, Jahnig F, Stern A, Kupsch EM, Meyer TF, Swanson J (1991) The opacity proteins of Neisseria gonorrhoeae strain MS11 are encoded by a family of 11 complete genes. Mol Microbiol 5(8):1889–1901

Blomgran R, Ernst JD (2011) Lung neutrophils facilitate activation of naive antigen-specific CD4+ T cells during Mycobacterium tuberculosis infection. J Immunol 186(12):7110–7119. doi:10.4049/jimmunol.1100001

Blomgran R, Desvignes L, Briken V, Ernst JD (2012) Mycobacterium tuberculosis inhibits neutrophil apoptosis, leading to delayed activation of naive CD4 T cells. Cell Host Microbe 11(1):81–90. doi:10.1016/j.chom.2011.11.012

Bloom CI, Graham CM, Berry MP, Wilkinson KA, Oni T, Rozakeas F, Xu Z, Rossello-Urgell J, Chaussabel D, Banchereau J, Pascual V, Lipman M, Wilkinson RJ, O'Garra A (2012) Detectable changes in the blood transcriptome are present after two weeks of antituberculosis therapy. PLoS One 7(10):e46191. doi:10.1371/journal.pone.0046191

Brown WC, Brayton KA, Styer CM, Palmer GH (2003) The hypervariable region of Anaplasma marginale major surface protein 2 (MSP2) contains multiple immunodominant CD4+ T lymphocyte epitopes that elicit variant-specific proliferative and IFN-gamma responses in MSP2 vaccinates. J Immunol 170(7):3790–3798

Cambier CJ, Takaki KK, Larson RP, Hernandez RE, Tobin DM, Urdahl KB, Cosma CL, Ramakrishnan L (2014) Mycobacteria manipulate macrophage recruitment through coordinated use of membrane lipids. Nature 505(7482):218–222. doi:10.1038/nature12799

Caminero JA, Pena MJ, Campos-Herrero MI, Rodriguez JC, Afonso O, Martin C, Pavon JM, Torres MJ, Burgos M, Cabrera P, Small PM, Enarson DA (2001) Exogenous reinfection with tuberculosis on a European island with a moderate incidence of disease. Am J Respir Crit Care Med 163(3 Pt 1):717–720. doi:10.1164/ajrccm.163.3.2003070

Cerwenka A, Lanier LL (2016) Natural killer cell memory in infection, inflammation and cancer. Nat Rev Immunol 16(2):112–123. doi:10.1038/nri.2015.9

Chackerian AA, Alt JM, Perera TV, Dascher CC, Behar SM (2002) Dissemination of Mycobacterium tuberculosis is influenced by host factors and precedes the initiation of T-cell immunity. Infect Immun 70(8):4501–4509

Chapel HM, Lee M, Hargreaves R, Pamphilon DH, Prentice AG (1994) Randomised trial of intravenous immunoglobulin as prophylaxis against infection in

plateau-phase multiple myeloma. The UK Group for Immunoglobulin Replacement Therapy in Multiple Myeloma. Lancet 343(8905):1059–1063

Cheng SC, Quintin J, Cramer RA, Shepardson KM, Saeed S, Kumar V, Giamarellos-Bourboulis EJ, Martens JH, Rao NA, Aghajanirefah A, Manjeri GR, Li Y, Ifrim DC, Arts RJ, van der Veer BM, Deen PM, Logie C, O'Neill LA, Willems P, van de Veerdonk FL, van der Meer JW, Ng A, Joosten LA, Wijmenga C, Stunnenberg HG, Xavier RJ, Netea MG (2014) mTOR- and HIF-1alpha-mediated aerobic glycolysis as metabolic basis for trained immunity. Science (New York, NY) 345(6204):1250684. doi:10.1126/science.1250684

Claessens A, Hamilton WL, Kekre M, Otto TD, Faizullabhoy A, Rayner JC, Kwiatkowski D (2014) Generation of antigenic diversity in Plasmodium falciparum by structured rearrangement of Var genes during mitosis. PLoS Genet 10(12):e1004812. doi:10.1371/journal.pgen.1004812

Cohen I, Parada C, Acosta-Gio E, Espitia C (2014) The PGRS domain from PE_PGRS33 of Mycobacterium tuberculosis is target of humoral immune response in mice and humans. Front Immunol 5:236. doi:10.3389/fimmu.2014.00236

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 393(6685):537–544. doi:10.1038/31159

Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, Kremer K, Ernst JD, Gagneux S (2010) Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved. Nat Genet 42(6):498–503. doi:10.1038/ng.590

Cooper G, Rosenstein C, Walter A, Peizer L (1932) The further separation of types among the pneumococci hitherto included in group iv and the development of therapeutic antisera for these types. J Exp Med 55(4):531–554

Copin R, Coscolla M, Seiffert SN, Bothamley G, Sutherland J, Mbayo G, Gagneux S, Ernst JD (2014) Sequence diversity in the pe_pgrs genes of Mycobacterium tuberculosis is independent of human T cell recognition. MBio 5(1):e00960–e00913. doi:10.1128/mBio.00960-13

Coscolla M, Copin R, Sutherland J, Gehre F, de Jong B, Owolabi O, Mbayo G, Giardina F, Ernst JD, Gagneux S (2015) M. tuberculosis T cell epitope analysis reveals paucity of antigenic variation and identifies rare variable TB antigens. Cell Host Microbe 18(5):538–548. doi:10.1016/j.chom.2015.10.008

Crawford H, Prado JG, Leslie A, Hue S, Honeyborne I, Reddy S, van der Stok M, Mncube Z, Brander C, Rousseau C, Mullins JI, Kaslow R, Goepfert P, Allen S, Hunter E, Mulenga J, Kiepiela P, Walker BD, Goulder PJ (2007) Compensatory mutation partially restores fitness and delays reversion of escape mutation within the immunodominant HLA-B*5703-restricted Gag epitope in chronic human immunodeficiency virus type 1 infection. J Virol 81(15):8346–8351. doi:10.1128/jvi.00465-07

Crawford H, Lumm W, Leslie A, Schaefer M, Boeras D, Prado JG, Tang J, Farmer P, Ndung'u T, Lakhi S, Gilmour J, Goepfert P, Walker BD, Kaslow R, Mulenga J, Allen S, Goulder PJ, Hunter E (2009) Evolution of HLA-B*5703 HIV-1 escape mutations in HLA-B*5703-positive individuals and their transmission recipients. J Exp Med 206(4):909–921. doi:jem.20081984 [pii] 231084/jem.20081984

Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, McGee L, von Gottberg A, Song JH, Ko KS, Pichon B, Baker S, Parry CM, Lambertsen LM, Shahinas D, Pillai DR, Mitchell TJ, Dougan G, Tomasz A, Klugman KP, Parkhill J, Hanage WP, Bentley SD (2011) Rapid pneumococcal evolution in response to clinical interventions. Science (New York, NY) 331(6016):430–434. doi:10.1126/science.1198545

Cuevas JM, Geller R, Garijo R, Lopez-Aldeguer J, Sanjuan R (2015) Extremely high mutation rate of HIV-1 in vivo. PLoS Biol 13(9):e1002251. doi:10.1371/journal.pbio.1002251

Deitsch KW, Lukehart SA, Stringer JR (2009) Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. Nat Rev Microbiol 7(7):493–503. doi:nrmicro2145 [pii] 281038/nrmicro2145

Delogu G, Brennan MJ (2001) Comparative immune response to PE and PE_PGRS antigens of Mycobacterium tuberculosis. Infect Immun 69(9):5606–5611

Dochez AR, Gillespie LJ (1913) A biologic classification of pneumococci by means of immunity reactions. JAMA 61:727–730

Elde NC, Malik HS (2009) The evolutionary conundrum of pathogen mimicry. Nat Rev Microbiol 7(11):787–797. doi:10.1038/nrmicro2222

Erickson AL, Kimura Y, Igarashi S, Eichelberger J, Houghton M, Sidney J, McKinney D, Sette A, Hughes AL, Walker CM (2001) The outcome of hepatitis C virus infection is predicted by escape mutations in epitopes targeted by cytotoxic T lymphocytes. Immunity 15(6):883–895

Ernst JD (2012) The immunological life cycle of tuberculosis. Nat Rev Immunol 12(8):581–591. doi:10.1038/nri3259

Feikin DR, Kagucia EW, Loo JD, Link-Gelles R, Puhan MA, Cherian T, Levine OS, Whitney CG, O'Brien KL, Moore MR (2013) Serotype-specific changes in invasive pneumococcal disease after pneumococcal conjugate vaccine introduction: a pooled analysis of

multiple surveillance sites. PLoS Med 10(9):e1001517. doi:10.1371/journal.pmed.1001517

Fu Y, Galan JE (1999) A salmonella protein antagonizes Rac-1 and Cdc42 to mediate host-cell recovery after bacterial invasion. Nature 401(6750):293–297. doi:10.1038/45829

Goepfert PA, Lumm W, Farmer P, Matthews P, Prendergast A, Carlson JM, Derdeyn CA, Tang J, Kaslow RA, Bansal A, Yusim K, Heckerman D, Mulenga J, Allen S, Goulder PJ, Hunter E (2008) Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in linked recipients. J Exp Med 205(5):1009–1017. doi:10.1084/jem.20072457

Grohskopf LA, Sokolow LZ, Olsen SJ, Bresee JS, Broder KR, Karron RA (2015) Prevention and control of influenza with vaccines: recommendations of the advisory committee on immunization practices, United States, 2015–16 influenza season. MMWR Morb Mortal Wkly Rep 64(30):818–825

Hardt WD, Chen LM, Schuebel KE, Bustelo XR, Galan JE (1998) S. typhimurium encodes an activator of Rho GTPases that induces membrane ruffling and nuclear responses in host cells. Cell 93(5):815–826

Hebert AM, Talarico S, Yang D, Durmaz R, Marrs CF, Zhang L, Foxman B, Yang Z (2007) DNA polymorphisms in the pepA and PPE18 genes among clinical strains of Mycobacterium tuberculosis: implications for vaccine efficacy. Infect Immun 75(12):5798–5805. doi:IAI.00335-07 [pii] 401128/IAI.00335-07

Hill DJ, Griffiths NJ, Borodina E, Virji M (2010) Cellular and molecular biology of Neisseria meningitidis colonization and invasive disease. Clin Sci (Lond) 118(9):547–564. doi:10.1042/cs20090513

Hoe NP, Nakashima K, Lukomski S, Grigsby D, Liu M, Kordari P, Dou SJ, Pan X, Vuopio-Varkila J, Salmelinna S, McGeer A, Low DE, Schwartz B, Schuchat A, Naidich S, De Lorenzo D, Fu YX, Musser JM (1999) Rapid selection of complement-inhibiting protein variants in group A Streptococcus epidemic waves. Nat Med 5(8):924–929. doi:10.1038/11369

Hosea SW, Brown EJ, Hamburger MI, Frank MM (1981) Opsonic requirements for intravascular clearance after splenectomy. N Engl J Med 304(5):245–250. doi:10.1056/nejm198101293040501

Hraber P, Korber B, Wagh K, Giorgi EE, Bhattacharya T, Gnanakaran S, Lapedes AS, Learn GH, Kreider EF, Li Y, Shaw GM, Hahn BH, Montefiori DC, Alam SM, Bonsignori M, Moody MA, Liao HX, Gao F, Haynes BF (2015) Longitudinal antigenic sequences and sites from intra-host evolution (LASSIE) identifies immune-selected HIV variants. Virus 7(10):5443–5475. doi:10.3390/v7102881

Hsu HE, Shutt KA, Moore MR, Beall BW, Bennett NM, Craig AS, Farley MM, Jorgensen JH, Lexau CA, Petit S, Reingold A, Schaffner W, Thomas A, Whitney CG, Harrison LH (2009) Effect of pneumococcal conjugate vaccine on pneumococcal meningitis. N Engl J Med 360(3):244–256. doi:10.1056/NEJMoa0800836

Interrante JD, Haddad MB, Kim L, Gandhi NR (2015) Exogenous reinfection as a cause of late recurrent tuberculosis in the United States. Ann Am Thorac Soc 12(11):1619–1626. doi:10.1513/AnnalsATS.201507-429OC

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science (New York, NY) 337(6096):816–821. doi:10.1126/science.1225829

Kaslow RA, Carrington M, Apple R, Park L, Munoz A, Saah AJ, Goedert JJ, Winkler C, O'Brien SJ, Rinaldo C, Detels R, Blattner W, Phair J, Erlich H, Mann DL (1996) Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. Nat Med 2(4):405–411

Kasprowicz V, Kang YH, Lucas M, Schulze zur Wiesch J, Kuntzen T, Fleming V, Nolan BE, Longworth S, Berical A, Bengsch B, Thimme R, Lewis-Ximenez L, Allen TM, Kim AY, Klenerman P, Lauer GM (2010) Hepatitis C virus (HCV) sequence variation induces an HCV-specific T-cell phenotype analogous to spontaneous resolution. J Virol 84(3):1656–1663. doi:10.1128/jvi.01499-09

Koh KW, Soh SE, Seah GT (2009) Strong antibody responses to Mycobacterium tuberculosis PE-PGRS62 protein are associated with latent and active tuberculosis. Infect Immun 77(8):3337–3343. doi:10.1128/iai.01175-08

Korber B, Gnanakaran S (2009) The implications of patterns in HIV diversity for neutralizing antibody induction and susceptibility. Curr Opin HIV AIDS 4(5):408–417. doi:10.1097/COH.0b013e32832f129e

Lannergard J, Gustafsson MC, Waldemarsson J, Norrby-Teglund A, Stalhammar-Carlemalm M, Lindahl G (2011) The Hypervariable region of Streptococcus pyogenes M protein escapes antibody attack by antigenic variation and weak immunogenicity. Cell Host Microbe 10(2):147–157. doi:10.1016/j.chom.2011.06.011

LaRock CN, Dohrmann S, Todd J, Corriden R, Olson J, Johannssen T, Lepenies B, Gallo RL, Ghosh P, Nizet V (2015) Group a streptococcal M1 protein sequesters cathelicidin to evade innate immune killing. Cell Host Microbe 18(4):471–477. doi:10.1016/j.chom.2015.09.004

Lawrence CW, Braciale TJ (2004) Activation, differentiation, and migration of naive virus-specific CD8+ T cells during pulmonary influenza virus infection. J Immunol 173(2):1209–1218

Lawrence CW, Ream RM, Braciale TJ (2005) Frequency, specificity, and sites of expansion of CD8+ T cells during primary pulmonary influenza virus infection. J Immunol 174(9):5332–5340

Lawrenz MB, Wooten RM, Norris SJ (2004) Effects of vlsE complementation on the infectivity of Borrelia burgdorferi lacking the linear plasmid lp28-1. Infect Immun 72(11):6577–6585. doi:10.1128/iai.72.11.6577-6585.2004

Lazaryan A, Song W, Lobashevsky E, Tang J, Shrestha S, Zhang K, Gardner LI, McNicholl JM, Wilson CM, Klein RS, Rompalo A, Mayer K, Sobel J, Kaslow RA (2010) Human leukocyte antigen class I supertypes and HIV-1 control in African Americans. J Virol 84(5):2610–2617. doi:10.1128/jvi.01962-09

Li Y, Svehla K, Louder MK, Wycuff D, Phogat S, Tang M, Migueles SA, Wu X, Phogat A, Shaw GM, Connors M, Hoxie J, Mascola JR, Wyatt R (2009) Analysis of neutralization specificities in polyclonal sera derived from human immunodeficiency virus type 1-infected individuals. J Virol 83(2):1045–1059. doi:10.1128/jvi.01992-08

Liang FT, Yan J, Mbow ML, Sviat SL, Gilmore RD, Mamula M, Fikrig E (2004) Borrelia burgdorferi changes its surface antigenic expression in response to host immune responses. Infect Immun 72(10):5759–5767. doi:10.1128/iai.72.10.5759-5767.2004

Lillebaek T, Dirksen A, Baess I, Strunge B, Thomsen VO, Andersen AB (2002) Molecular evidence of endogenous reactivation of Mycobacterium tuberculosis after 33 years of latent infection. J Infect Dis 185(3):401–404. doi:JID010932 [pii] 561086/338342

Maertzdorf J, Ota M, Repsilber D, Mollenkopf HJ, Weiner J, Hill PC, Kaufmann SH (2011) Functional correlations of pathogenesis-driven gene expression signatures in tuberculosis. PLoS One 6(10):e26938. doi:10.1371/journal.pone.0026938

Makarova KS, Wolf YI, Koonin EV (2013) Comparative genomics of defense systems in archaea and bacteria. Nucleic Acids Res 41(8):4360–4377. doi:10.1093/nar/gkt157

Martinez E, Allombert J, Cantet F, Lakhani A, Yandrapalli N, Neyret A, Norville IH, Favard C, Muriaux D, Bonazzi M (2016) Coxiella burnetii effector CvpB modulates phosphoinositide metabolism for optimal vacuole development. Proc Natl Acad Sci U S A 113(23):E3260–E3269. doi:10.1073/pnas.1522811113

Mayer-Barber KD, Andrade BB, Barber DL, Hieny S, Feng CG, Caspar P, Oland S, Gordon S, Sher A (2011) Innate and adaptive interferons suppress IL-1alpha and IL-1beta production by distinct pulmonary myeloid subsets during Mycobacterium tuberculosis infection. Immunity 35(6):1023–1034. doi:10.1016/j.immuni.2011.12.002

Mayer-Barber KD, Andrade BB, Oland SD, Amaral EP, Barber DL, Gonzales J, Derrick SC, Shi R, Kumar NP, Wei W, Yuan X, Zhang G, Cai Y, Babu S, Catalfamo M, Salazar AM, Via LE, Barry CE 3rd, Sher A (2014) Host-directed therapy of tuberculosis based on interleukin-1 and type I interferon crosstalk. Nature 511(7507):99–103. doi:10.1038/nature13489

McKinney DM, Southwood S, Hinz D, Oseroff C, Arlehamn CS, Schulten V, Taplitz R, Broide D, Hanekom WA, Scriba TJ, Wood R, Alam R, Peters B, Sidney J, Sette A (2013) A strategy to determine HLA class II restriction broadly covering the DR, DP, and DQ allelic variants most commonly expressed in the general population. Immunogenetics 65(5):357–370. doi:10.1007/s00251-013-0684-y

Miller JD, van der Most RG, Akondy RS, Glidewell JT, Albott S, Masopust D, Murali-Krishna K, Mahar PL, Edupuganti S, Lalor S, Germon S, Del Rio C, Mulligan MJ, Staprans SI, Altman JD, Feinberg MB, Ahmed R (2008) Human effector and memory CD8+ T cell responses to smallpox and yellow fever vaccines. Immunity 28(5):710–722. doi:10.1016/j.immuni.2008.02.020

Mordue DG, Sibley LD (1997) Intracellular fate of vacuoles containing Toxoplasma gondii is determined at the time of formation and depends on the mechanism of entry. J Immunol 159(9):4452–4459

Mordue DG, Hakansson S, Niesman I, Sibley LD (1999) Toxoplasma gondii resides in a vacuole that avoids fusion with host cell endocytic and exocytic vesicular trafficking pathways. Exp Parasitol 92(2):87–99

Morrison LJ, Marcello L, McCulloch R (2009) Antigenic variation in the African trypanosome: molecular mechanisms and phenotypic complexity. Cell Microbiol 11(12):1724–1734. doi:10.1111/j.1462-5822.2009.01383.x

Murphy KP, Weaver C (2016) Janeway's immunobiology. Garland Science, New York

Musser JM, Amin A, Ramaswamy S (2000) Negligible genetic diversity of mycobacterium tuberculosis host immune system protein targets: evidence of limited selective pressure. Genetics 155(1):7–16

Netea MG, Joosten LA, Latz E, Mills KH, Natoli G, Stunnenberg HG, O'Neill LA, Xavier RJ (2016) Trained immunity: a program of innate immune memory in health and disease. Science (New York, NY) 352(6284):aaf1098. doi:10.1126/science.aaf1098

Newton HJ, Kohler LJ, McDonough JA, Temoche-Diaz M, Crabill E, Hartland EL, Roy CR (2014) A screen of Coxiella burnetii mutants reveals important roles for Dot/Icm effectors and host autophagy in vacuole biogenesis. PLoS Pathog 10(7):e1004286. doi:10.1371/journal.ppat.1004286

Nivarthi UK, Gras S, Kjer-Nielsen L, Berry R, Lucet IS, Miles JJ, Tracy SL, Purcell AW, Bowden DS, Hellard M, Rossjohn J, McCluskey J, Bharadwaj M (2014) An extensive antigenic footprint underpins immunodominant TCR adaptability against a hypervariable viral determinant. J Immunol 193(11):5402–5413. doi:10.4049/jimmunol.1401357

O'Garra A, Redford PS, McNab FW, Bloom CI, Wilkinson RJ, Berry MP (2013) The immune response in tuberculosis. Annu Rev Immunol 31:475–527. doi:10.1146/annurev-immunol-032712-095939

Ohmit SE, Petrie JG, Cross RT, Johnson E, Monto AS (2011) Influenza hemagglutination-inhibition antibody titer as a correlate of vaccine-induced protection. J Infect Dis 204(12):1879–1885. doi:10.1093/infdis/jir661

Ottenhoff TH, Dass RH, Yang N, Zhang MM, Wong HE, Sahiratmadja E, Khor CC, Alisjahbana B, van Crevel R, Marzuki S, Seielstad M, van de Vosse E, Hibberd ML (2012) Genome-wide expression profiling identifies type 1 interferon response pathways in active tuberculosis. PLoS One 7(9):e45839. doi:10.1371/journal.pone.0045839

Pantophlet R, Burton DR (2006) GP120: target for neutralizing HIV-1 antibodies. Annu Rev Immunol 24:739–769. doi:10.1146/annurev.immunol.24.021605.090557

Paul S, Dillon MB, Lindestam Arlehamn CS, Huang H, Davis MM, McKinney DM, Scriba TJ, Sidney J, Peters B, Sette A (2015) A population response analysis approach to assign class II HLA-epitope restrictions. J Immunol 194(12):6164–6176. doi:10.4049/jimmunol.1403074

Persson J, Beall B, Linse S, Lindahl G (2006) Extreme sequence divergence but conserved ligand-binding specificity in Streptococcus pyogenes M protein. PLoS Pathog 2(5):e47. doi:10.1371/journal.ppat.0020047

Peters W, Scott HM, Chambers HF, Flynn JL, Charo IF, Ernst JD (2001) Chemokine receptor 2 serves an early and essential role in resistance to Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 98(14):7958–7963. doi:10.1073/pnas.131207398

Peters J, Fowler E, Gatton M, Chen N, Saul A, Cheng Q (2002) High diversity and rapid changeover of expressed var genes during the acute phase of Plasmodium falciparum infections in human volunteers. Proc Natl Acad Sci U S A 99(16):10689–10694. doi:10.1073/pnas.162349899

Peters W, Cyster JG, Mack M, Schlondorff D, Wolf AJ, Ernst JD, Charo IF (2004) CCR2-dependent trafficking of F4/80dim macrophages and CD11cdim/intermediate dendritic cells is crucial for T cell recruitment to lungs infected with Mycobacterium tuberculosis. J Immunol 172(12):7647–7653

Pichon B, Ladhani SN, Slack MP, Segonds-Pichon A, Andrews NJ, Waight PA, Miller E, George R (2013) Changes in molecular epidemiology of streptococcus pneumoniae causing meningitis following introduction of pneumococcal conjugate vaccination in England and Wales. J Clin Microbiol 51(3):820–827. doi:10.1128/jcm.01917-12

Plasterk RH, Simon MI, Barbour AG (1985) Transposition of structural genes to an expression sequence on a linear plasmid causes antigenic variation in the bacterium Borrelia hermsii. Nature 318(6043):257–263

du Plessis DG, Warren R, Richardson M, Joubert JJ, van Helden PD (2001) Demonstration of reinfection and reactivation in HIV-negative autopsied cases of secondary tuberculosis: multilesional genotyping of Mycobacterium tuberculosis utilizing IS 6110 and other repetitive element-based DNA fingerprinting. Tuberculosis (Edinb) 81(3):211–220. doi:10.1054/tube.2000.0278

Potter AJ, Trappetti C, Paton JC (2012) Streptococcus pneumoniae uses glutathione to defend against oxidative stress and metal ion toxicity. J Bacteriol 194(22):6248–6254. doi:10.1128/jb.01393-12

Poulsen A (1950) Some clinical features of tuberculosis. 1. Incubation period. Acta Tuberc Scand 24(3–4):311–346

Proft T, Moffatt SL, Weller KD, Paterson A, Martin D, Fraser JD (2000) The streptococcal superantigen SMEZ exhibits wide allelic variation, mosaic structure, and significant antigenic variation. J Exp Med 191(10):1765–1776

Ralph SA, Scheidig-Benatar C, Scherf A (2005) Antigenic variation in Plasmodium falciparum is associated with movement of var loci between subnuclear locations. Proc Natl Acad Sci U S A 102(15):5414–5419. doi:10.1073/pnas.0408883102

Richter SS, Heilmann KP, Dohrn CL, Riahi F, Diekema DJ, Doern GV (2013) Pneumococcal serotypes before and after introduction of conjugate vaccines, United States, 1999–2011. Emerg Infect Dis 19(7):1074–1083. doi:10.3201/eid1907.121830

Roberts JD, Bebenek K, Kunkel TA (1988) The accuracy of reverse transcriptase from HIV-1. Science (New York, NY) 242(4882):1171–1173

Saeed S, Quintin J, Kerstens HH, Rao NA, Aghajanirefah A, Matarese F, Cheng SC, Ratter J, Berentsen K, van der Ent MA, Sharifi N, Janssen-Megens EM, Ter Huurne M, Mandoli A, van Schaik T, Ng A, Burden F, Downes K, Frontini M, Kumar V, Giamarellos-Bourboulis EJ, Ouwehand WH, van der Meer JW, Joosten LA, Wijmenga C, Martens JH, Xavier RJ, Logie C, Netea MG, Stunnenberg HG (2014) Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. Science (New York, NY) 345(6204):1251086. doi:10.1126/science.1251086

Samstein M, Schreiber HA, Leiner IM, Susac B, Glickman MS, Pamer EG (2013) Essential yet limited role for CCR2(+) inflammatory monocytes during Mycobacterium tuberculosis-specific T cell priming. elife 2:e01086. doi:10.7554/eLife.01086

Sassetti CM, Rubin EJ (2003) Genetic requirements for mycobacterial survival during infection. Proc Natl Acad Sci U S A 100(22):12989–12994. doi:10.1073/pnas.2134250100

Sassetti CM, Boyd DH, Rubin EJ (2003) Genes required for mycobacterial growth defined by high density mutagenesis. Mol Microbiol 48(1):77–84

Segal E, Hagblom P, Seifert HS, So M (1986) Antigenic variation of gonococcal pilus involves assembly of separated silent gene segments. Proc Natl Acad Sci U S A 83(7):2177–2181

Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, Fouchier RA (2004) Mapping the antigenic and genetic evolution of influenza virus. Science (New York, NY) 305(5682):371–376. doi:10.1126/science.1097211

Srivastava S, Ernst JD (2014) Cell-to-cell transfer of M. tuberculosis antigens optimizes CD4 T cell priming. Cell Host Microbe 15(6):741–752. doi:10.1016/j.chom.2014.05.007

Srivastava S, Grace PS, Ernst JD (2016) Antigen export reduces antigen presentation and limits T cell control of M. tuberculosis. Cell Host Microbe 19(1):44–54. doi:10.1016/j.chom.2015.12.003

Stoenner HG, Dodd T, Larsen C (1982) Antigenic variation of Borrelia hermsii. J Exp Med 156(5):1297–1311

Talarico S, Cave MD, Marrs CF, Foxman B, Zhang L, Yang Z (2005) Variation of the Mycobacterium tuberculosis PE_PGRS 33 gene among clinical isolates. J Clin Microbiol 43(10):4954–4960. doi:43/10/4954 [pii]10.1128/JCM.43.904954-4960.2005

Talarico S, Zhang L, Marrs CF, Foxman B, Cave MD, Brennan MJ, Yang Z (2008) Mycobacterium tuberculosis PE_PGRS16 and PE_PGRS26 genetic polymorphism among clinical isolates. Tuberculosis (Edinb) 88(4):283–294. doi:S1472-9792(08)00002-4 [pii] 921016/j.tube.2008.01.001

Thyagarajan B, Bloom JD (2014) The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. elife 3. doi:10.7554/eLife.03300

Tsibane T, Ekiert DC, Krause JC, Martinez O, Crowe JE Jr, Wilson IA, Basler CF (2012) Influenza human monoclonal antibody 1F1 interacts with three major antigenic sites and residues mediating human receptor specificity in H1N1 viruses. PLoS Pathog 8(12):e1003067. doi:10.1371/journal.ppat.1003067

Uplekar S, Heym B, Friocourt V, Rougemont J, Cole ST (2011) Comparative genomics of Esx genes from clinical isolates of Mycobacterium tuberculosis provides evidence for gene conversion and epitope variation. Infect Immun 79(10):4042–4049. doi:10.1128/iai.05344-11

Valkenburg SA, Quiñones-Parra S, Gras S, Komadina N, McVernon J, Wang Z, Halim H, Iannello P, Cole C, Laurie K, Kelso A, Rossjohn J, Doherty PC, Turner SJ, Kedzierska K (2013) Acute emergence and reversion of influenza A virus quasispecies within CD8+ T cell antigenic peptides. Nat Commun 4. doi:10.1038/ncomms3663

Valkenburg SA, Josephs TM, Clemens EB, Grant EJ, Nguyen TH, Wang GC, Price DA, Miller A, Tong SY, Thomas PG, Doherty PC, Rossjohn J, Gras S, Kedzierska K (2016) Molecular basis for universal HLA-A*0201-restricted CD8+ T-cell immunity against influenza viruses. Proc Natl Acad Sci U S A 113(16):4440–4445. doi:10.1073/pnas.1603106113

Virelizier JL (1975) Host defenses against influenza virus: the role of anti-hemagglutinin antibody. J Immunol 115(2):434–439

Vladimer GI, Weng D, Paquette SW, Vanaja SK, Rathinam VA, Aune MH, Conlon JE, Burbage JJ, Proulx MK, Liu Q, Reed G, Mecsas JC, Iwakura Y, Bertin J, Goguen JD, Fitzgerald KA, Lien E (2012) The NLRP12 inflammasome recognizes Yersinia pestis. Immunity 37(1):96–107. doi:10.1016/j.immuni.2012.07.006

Vuilleumier S, Bonhoeffer S (2015) Contribution of recombination to the evolutionary history of HIV. Curr Opin HIV AIDS 10(2):84–89. doi:10.1097/coh.0000000000000137

Watson RO, Bell SL, MacDuff DA, Kimmey JM, Diner EJ, Olivas J, Vance RE, Stallings CL, Virgin HW, Cox JS (2015) The cytosolic sensor cGAS detects Mycobacterium tuberculosis DNA to induce type I interferons and activate autophagy. Cell Host Microbe 17(6):811–819. doi:10.1016/j.chom.2015.05.004

Wiens KE, Ernst JD (2016) The mechanism for type I interferon induction by Mycobacterium tuberculosis is bacterial strain-dependent. PLoS Pathog 12(8):e1005809. doi:10.1371/journal.ppat.1005809

Wolf AJ, Linas B, Trevejo-Nunez GJ, Kincaid E, Tamura T, Takatsu K, Ernst JD (2007) Mycobacterium tuberculosis infects dendritic cells with high frequency and impairs their function in vivo. J Immunol 179(4):2509–2519

Wolf AJ, Desvignes L, Linas B, Banaiee N, Tamura T, Takatsu K, Ernst JD (2008) Initiation of the adaptive immune response to Mycobacterium tuberculosis depends on antigen production in the local lymph node, not the lungs. J Exp Med 205(1):105–115. doi:10.1084/jem.20071367

Wrammert J, Smith K, Miller J, Langley WA, Kokko K, Larsen C, Zheng N-Y, Mays I, Garman L, Helms C, James J, Air GM, Capra JD, Ahmed R, Wilson PC (2008) Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. Nature 453(7195):667–671. http://www.nature.com/nature/journal/v453/n7195/suppinfo/nature06890_S1.html

Wyres KL, Lambertsen LM, Croucher NJ, McGee L, von Gottberg A, Linares J, Jacobs MR, Kristinsson KG, Beall BW, Klugman KP, Parkhill J, Hakenbeck R, Bentley SD, Brueggemann AB (2013) Pneumococcal capsular switching: a historical perspective. J Infect Dis 207(3):439–449. doi:10.1093/infdis/jis703

Xu W, Edwards MR, Borek DM, Feagins AR, Mittal A, Alinger JB, Berry KN, Yen B, Hamilton J, Brett TJ, Pappu RV, Leung DW, Basler CF, Amarasinghe GK (2014) Ebola virus VP24 targets a unique NLS binding site on karyopherin alpha 5 to selectively compete with nuclear import of phosphorylated STAT1. Cell Host Microbe 16(2):187–200. doi:10.1016/j.chom.2014.07.008

Yang QL, Gotschlich EC (1996) Variation of gonococcal lipooligosaccharide structure is due to alterations in poly-G tracts in lgt genes encoding glycosyl transferases. J Exp Med 183(1):323–327

PE and PPE Genes: A Tale of Conservation and Diversity

10

Giovanni Delogu, Michael J. Brennan, and Riccardo Manganelli

Abstract

PE and PPE are two large families of proteins typical of mycobacteria whose structural genes in the *Mycobacterium tuberculosis* complex (MTBC) occupy about 7% of the total genome. The most ancestral PE and PPE proteins are expressed by genes that belong to the same operon and in most cases are found inserted in the *esx* clusters, encoding a type VII secretion system. Duplication and expansion of *pe* and *ppe* genes, coupled with intragenomic and intergenomic recombination events, led to the emergence of the polymorphic *pe_pgrs* and *ppe_mptr* genes in the MTBC genome. The role and function of these proteins, and particularly of the polymorphic subfamilies, remains elusive, although it is widely accepted that PE and PPE proteins may represent a specialized collection used by MTBC to interact with the complex host immune system of mammals. In this chapter, we summarize what has been discovered since the identification of these genes in 1998, focusing on *M. tuberculosis* genetic variability, host-pathogen interaction and TB pathogenesis.

Keywords

PE • PPE • PE_PGRS • PPE_MPTR • Genetic variability • Gene family • Polymorphism

G. Delogu (✉)
Institute of Microbiology, Università Cattolica del Sacro Cuore, Largo A. Gemelli, 8, 00168, Rome, Italy
e-mail: giovanni.delogu@unicatt.it

M.J. Brennan
Aeras, 1405 Research Blvd., Rockville, MD 20850, USA
e-mail: mbrennan@aeras.org

R. Manganelli
Department of Molecular Medicine, University of Padua, Via A. Gabelli, 63, 35121, Padua, Italy
e-mail: riccardo.manganelli@unipd.it

S. Gagneux (ed.), *Strain Variation in the* Mycobacterium tuberculosis *Complex: Its Role in Biology, Epidemiology and Control*, Advances in Experimental Medicine and Biology 1019,
DOI 10.1007/978-3-319-64371-7_10

10.1 General Features

Sequences encoding PE_PGRS proteins in *Mycobacterium tuberculosis* (*Mtb*) are the Polymorphic-GC-Rich Sequences (PGRS) that were first used in the molecular typing of *Mtb* (Poulet and Cole 1995). This use occurred prior to the publication of the first *Mtb* genome sequence of the laboratory strain H37Rv in 1998 (Cole et al. 1998). The sequencing was slowed by the homology and redundancy of *pe_pgrs*, as well as *pe* and *ppe*. Since then, we have learned much about the evolution and cell localization of certain PE and PPE proteins, but we have much more to learn about their function and immunogenicity.

In 1998, the first *Mtb* genome sequence highlighted for the first time the presence of genes grouped into two large families that were shown to comprise approximately 10% of the genome size and to account for 7–8% of the *Mtb* coding potential (Cole et al. 1998; Brennan and Delogu 2002). Based on the presence of conserved Pro-Glu (PE) and Pro-Pro-Glu (PPE) motifs at their N-terminus, the proteins encoded by the members of these two gene families were named PE and PPE, respectively. The *Mtb* H37Rv genome comprises 100 *pe* and 69 *ppe* genes encoding the corresponding proteins, which are further classified into subfamilies depending on the amino acid sequence at the C-terminus. The 100 proteins of the PE family share a highly conserved N-terminal domain of about 90–100 amino acids in length and are further subdivided into the PE and PE_PGRS subfamilies (Gey van Pittius et al. 2006). PE_PGRS proteins are characterized by the presence of a polymorphic domain, rich in Gly-Gly-Ala/Gly-Gly-Asn amino acid repeats, which can vary in sequence and size. *pe_pgrs* genes are found scattered in the genome and are mostly not co-transcribed with other genes. Conversely, many of the *pe* genes are adjacent to *ppe* genes and a number of studies have demonstrated that these *pe/ppe* couplets are co-expressed, and at least some of the corresponding proteins are heterodimers that are secreted (Adindla and Guruprasad 2003; Ates et al. 2016). A number of PE proteins, characterized by a unique C-terminal domain, cannot be included in the PE_PGRS and PE/PPE couplet subfamilies and can be classified as PE unique (Delogu et al. 2008). The 69 PPE proteins are characterized by the presence of a highly conserved N-terminal domain which is approximately 180 amino acids in length and that, similar to the PE domain, seems to play a key role in driving protein localization or secretion (Daleke et al. 2011; Dona et al. 2013). Many PPE proteins appear to be co-expressed with the PE partner (as mentioned) and belong to the PE/PPE pair subfamily; others are encoded by genes found scattered in the chromosome. Twenty of them contain, downstream of the conserved N-terminal PPE domain, the Major Polymorphic Tandem Repeat (MPTR) region characterized by multiple C-terminal repeats of the amino acid sequence motif Asn-X-Gly-X-Gly-Asn-X-Gly (Gey van Pittius et al. 2006).

The discovery of these two protein families in 1998 (Cole et al. 1998) prompted many to speculate on their relevance in *Mtb* biology and, given their abundance in *M. tuberculosis* complex (MTBC), also about their role in tuberculosis (TB) pathogenesis. Since then, a number of studies have started to provide experimental data that specifies a function that, however, remains mostly vague. In this chapter, we provide a general overview of these families and their impact on TB pathogenesis, with a special focus on the biological significance of the genetic conservation as well as diversity of these apparently redundant proteins.

10.2 Origin and Expansion

Pe and *ppe* genes are found in the Mycobacterium genus with the fast-growing, non-pathogenic, saprophytic species such as *Mycobacterium smegmatis* or *Mycobacterium abscessus*, which are the closest to the last common ancestor of the genus, having only a few *pe* and *ppe* genes. In contrast, the slow-growing pathogenic mycobacterial species contain many *pe* and *ppe* genes (Gey van Pittius et al. 2006; Sassi and Drancourt 2014). Formative studies by Gey van Pittius and colleagues (2006) aimed at investigating the evolutionary history of *pe* and *ppe* genes, provided a beginning for defining their genetic

relationship in the MTBC complex. Based on in silico analysis of the available genome sequences, DNA hybridization assays and comparative genomics, these authors (Gey van Pittius et al. 2006) reconstructed the phylogenetic relationship of these genes and subdivided both the *pe* and *ppe* genes into five subfamilies. The progenitors of these families belong to subfamily I and are associated with the *esx-1* gene cluster. Non-pathogenic fast-growing species tend to have few *pe*/*ppe* genes, while slow-growing pathogenic species, such as MTBC, *Mycobacterium marinum* and *Mycobacterium ulcerans*, possess a high number of *pe*/*ppe* genes, and an accumulation of the most polymorphic genes (subfamilies V), including *pe_pgrs* and *ppe_mptr* (Gey van Pittius et al. 2006; Delogu et al. 2008). Expansion of the *pe* and *ppe* genes is thought to have been initiated with the duplication of the *esx-1* gene cluster, followed by multiple duplication events (Gey van Pittius et al. 2006). The association of *esx-1*, *esx-3* and *esx-5* with virulence of *Mtb* (Pym et al. 2002; Simeone et al. 2009; Serafini et al. 2013; Tufariello et al. 2016; Bottai et al. 2012), the discovery that duplication and expansion of *pe_pgrs* and *ppe_mptr* genes followed the emergence of *esx-5* (Gey van Pittius et al. 2006), and the elevated polymorphisms observed in these genes, were seen as an indication that PE and PPE proteins are key for the pathogenesis of TB (Brennan and Delogu 2002; Gey van Pittius et al. 2006).

Comparative genomic studies in smooth tubercle bacilli (STB) strains (see Chap. 2), have also provided useful insights into the evolution of these genes and their contribution in shaping *Mtb* as a human pathogen. Namouchi and colleagues (2013) identified strong signals of intragenic and intergenic recombination events in *pe_pgrs* genes that, together with mutation events, were important for generating genetic diversity. This was highlighted by the positive selection demonstrated in STB strains for a few genes (Namouchi et al. 2013). An interesting example of this is the evolution of *pe_pgrs17* and *pe_pgrs18* that emerged as a duplication event in an ancestral strain. This was followed by an accumulation of genetic changes that resulted in a polymorphic region (named 12/40) in *pe_pgrs17*. This polymorphic region, which was conserved during the expansion of the *Mycobacterium africanum* – *Mycobacterium bovis* lineage, was relocated to the other paralog *pe_pgrs18* only in modern *Mtb* strains by a process of gene conversion (Karboul et al. 2006). Hence, some modern MTBC strains have the 12/40 polymorphism and others do not. It has been proposed that the acquisition of the 12/40 polymorphism coincided with the emergence of the most successful MTBC strains (Karboul et al. 2006). This implies that these two PE_PGRS proteins are important for human-*Mtb* interactions. In a broader perspective, it has been suggested that homologous recombination between genetic sequences emerged by duplication events, and may have shaped the evolution of *pe* and *ppe* genes (and particularly of *pe_pgrs* and *ppe_mptr* genes) in STB. Expansion of these gene families may have provided the raw material for functional innovation as well as adaptation to the human host (Fig. 10.1) (McEvoy et al. 2009; Namouchi et al. 2013; Karboul et al. 2008; Supply et al. 2013).

10.3 Regulation of Gene Expression

The first study characterizing the level of expression of *pe* and *ppe* genes, in a systematic manner, is that of Voskuil and colleagues (2004). Using DNA microarrays, they found that, of 169 *pe* and *ppe* genes, 128 were differentially expressed, in 15 different environmental conditions. No clear conserved pattern of expression was found for *pe*, *ppe* or even any of their subgroups, with some genes specifically induced in a single condition (e.g. *pe3* was only induced in activated macrophages), and others differentially expressed in several different conditions (e.g. *pe11* was differentially regulated in 10 out of 15 different conditions). Expression of different subsets of PE and PPE proteins, in response to diverse environmental conditions may result in a wide antigen variation, allowing *Mtb* to evade the host immune response. This also suggests that, despite their structural similarity, small differences in

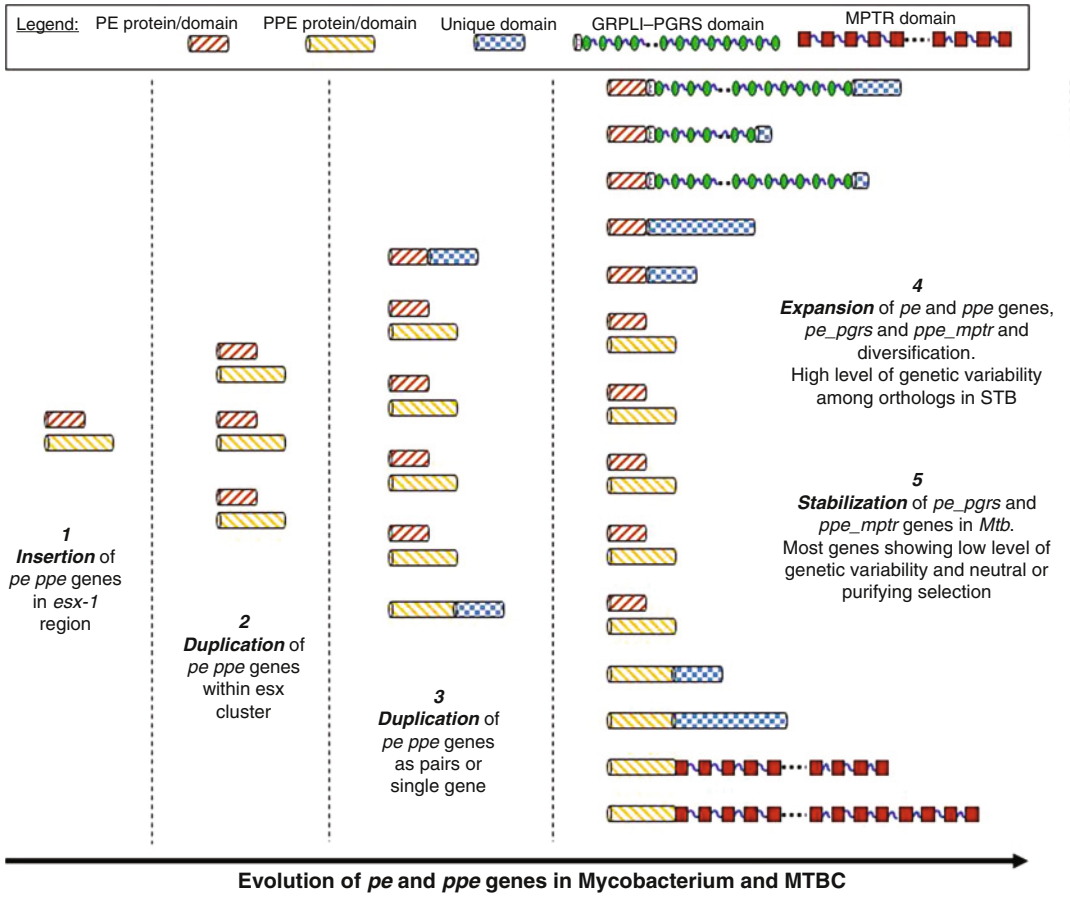

Fig. 10.1 Schematic of the PE and PPE proteins showing the expansion of these two families occuring during the evolution that led to MTBC

their sequence may result in diverse functions, so that different sets of proteins are expressed depending on the environmental conditions encountered by the bacteria (Delogu et al. 2006).

Differential expression of PE and PPE proteins has also been studied in vivo. Rachman and colleagues (2006) identified two *PE,* four *PE_PGRS* and three *PPE_MPTR* induced in lung granulomas from human patients, while Ramakrishnan and colleagues (2000) identified two PE_PGRS-coding genes of *M. marinum* preferentially expressed in macrophages. In addition, studying the expression of *ppe_mptr* genes throughout the infection of mice, it was demonstrated that while some of these genes are expressed constitutively, others are repressed or alternatively induced during infection (Soldini et al. 2011).

In vivo expression of PE and PPE proteins was also demonstrated by proteomics in a study which analyzed the bacterial proteome during infection of guinea pig lungs. Interestingly, members of the PE and PPE protein family (mostly PE_PGRS and PPE_MPTR) were found to represent the third most abundant category of proteins expressed during infection (Kruh et al. 2010).

Beyond showing a different pattern of regulation in response to changing environmental conditions, *pe* and *ppe* genes were also shown to be expressed at different levels in different members of the MTBC (Rehren et al. 2007; Golby et al. 2007) and among different *Mtb* clinical isolates (Gao et al. 2005). Specifically, *Mtb* was found to have 18 *pe* and *ppe* genes expressed at a superior level compared to its

close relative *M. bovis* (Rehren et al. 2007). Conversely, 20 *pe* and *ppe* genes were found to be differentially expressed among 10 clinical isolates of *Mtb* (Gao et al. 2005).

Several transcriptional factors have been reported to be involved in the modulation of *pe* and *ppe* genes expression (Ahmed et al. 2015; Mohareer et al. 2011). In particular, (*i*) the Iron DEpendent Regulator IdeR regulates around 11 *pe/ppe* genes (Gold et al. 2001; Rodriguez et al. 2002); (*ii*) PhoPR, a two component system involved in the regulation of genes in the synthesis of cell wall components such as the diacyltrehaloses, polyacyltrehaloses and sulfolipids (Solans et al. 2014), also regulates the expression of several *pe* and *ppe* genes; (*iii*) Rv0485, which was identified in a transposon mutagenesis approach as a regulator able to modify the expression of the *pe13-ppe18* operon (Goldstone et al. 2009); (*iv*) finally, some alternative sigma factors are involved in the differential expression of PE and PPE proteins (Mohareer et al. 2011). Overexpression of SigB causes induction of 7 *pe_pgrs* (Lee et al. 2008), and absence of SigD causes induction of 14 *pe_pgrs* (Raman et al. 2004).

10.4 PE/PPE Couplets and PE and PPE Unique Proteins

The most ancient *pe* and *ppe* genes are organized in operons and found within the *esx-1* gene cluster. A number of other *pe/ppe* couplets are found within the *esx-2*, *-3* and *-5* gene clusters or scattered throughout the genome (Gey van Pittius et al. 2006). The presence of these *pe/ppe* gene pairs in operons was immediately indicative of a functional link. Tundup and colleagues (2006) demonstrated that the *pe25/ppe41* genes were co-transcribed and their gene products were shown to interact. To investigate the structure of these PE and PPE proteins, Strong and colleagues (2006) co-expressed PE25 and PPE41 in *Escherichia coli*, demonstrating that co-expression is required for protein stability. These authors were able to obtain the crystal structure of the PE25/PPE41 complex and established that the PE protein interacts with the PPE protein through conserved apolar residues that warrant strong hydrophobic interactions (Strong et al. 2006). The structure of PE25/PPE41 resembles that of the ESXA/B heterodimer, providing further support for the observation that PE/PPE heterodimers are substrates for the Type 7 Secretion System (T7SS) in mycobacteria (Ates et al. 2016). Secretion of PE/PPE pairs has been demonstrated in mycobacteria and recent studies showed that the $EspG_5$ protein serves as a chaperone for the PE25/PPE41 heterodimer, that is then driven for secretion through the ESX-5 T7SS (Ekiert and Cox 2014; Korotkova et al. 2014). PE/PPE couplet proteins are therefore recognized by the EspG chaperone of the respective T7SS and then translocated through the mycobacterial membranes to be secreted or to remain associated with the capsular layer (Ates et al. 2016).

Inactivation of the *ppe25/pe19* genes, inserted in the *esx-5* gene cluster in *Mtb*, caused impaired secretion and attenuated virulence in mice, implicating these PE/PPE proteins in TB pathogenesis (Bottai et al. 2012). The role of these PE/PPE proteins in the context of the T7SS has been thoroughly reviewed elsewhere (Ates et al. 2016), but the functional role of these proteins in *Mtb* biology and their interaction with the other T7SS substrates remains to be established.

Co-regulated genes encoding PE/PPE couplets are very homologous and are major targets of the host T cell responses during infection. However, only few of the PE/PPE couplets have actually been analyzed (Akhter et al. 2012; Gey van Pittius et al. 2006; Mukhopadhyay and Balaji 2011; Fishbein et al. 2015). PPE18, which is a part of a putative PE/PPE couplet, has been implicated in the pathogenesis of TB via its ability to specifically bind to toll-like receptor (TLR) 2 and trigger secretion of the anti-inflammatory cytokine IL-10. These are features of fully virulent *Mtb* (Nair et al. 2009; Bhat et al. 2012). Interestingly, the immunomodulatory properties of PPE18 are observed when the protein is either expressed alone or as a recombinant protein in *E. coli* or *M. smegmatis*. This raises the question of whether heterodimerization with PE13 or PE31 is required for proper folding and function as predicted by bioinformatics studies (Nair et al. 2009; Riley et al. 2008). PPE17 is available on the mycobacterial surface and may deliver a heterologous antigen to the surface, when it is expressed in the absence of the predicted PE part-

ner (PE11) (Dona et al. 2013). Together with the observation that a *M. smegmatis* overexpressing PE11 alone shows enhanced virulence compared with the parental strain (Singh et al. 2016), these findings suggest that heterodimerization with the paired PE may not be a requirement for all PPEs (and vice versa).

A significant level of genetic variability in the *ppe18* gene was observed among MTBC strains, particularly in the region encoding the N-terminal domain of the protein. This region is involved in the interaction with TLR2 and is usually highly conserved in the PPE proteins (Hebert et al. 2007; Homolka et al. 2016). With this background, it would be of interest to investigate the impact of this genetic variability on PPE18-dependent *Mtb* inflammatory properties. Since PPE18 is one of the most promising vaccine candidates and it is included in the GSK Mtb72F subunit polyprotein (see below in Sect. 10.7.1), there has been concern that the genetic variability observed in *Mtb* strains could affect vaccine efficacy. Recent data also indicates that the insertions/deletions and mutations observed in *pe* and *ppe* do not impact the predicted putative T cell epitopes that cover a wide range of HLAI and II (Mortier et al. 2015).

Another *ppe* gene which has been analyzed because of *Mtb* genetic variability is *ppe38*, which has been shown to be hypervariable in the MTBC due to IS*6110*-recombinations, gene conversion and recombination with the *ppe71* paralog (McEvoy et al. 2009). In *M. marinum*, inactivation by transposon mutagenesis of the gene encoding the surface exposed PPE38 protein, results in an attenuated strain defective in phagocytosis uptake and in the ability to trigger inflammatory responses in adult zebrafish (Dong et al. 2012). The lack of a clear understanding of the role of PPE38 in *Mtb* biology prevents formulating any hypothesis on the impact of the observed polymorphisms. Moreover, it would be important to assess whether the predicted loss of function in *ppe38* is somehow compensated by other highly homologous paralogs (McEvoy et al. 2009).

Other PE and PPE proteins demonstrate immunomodulatory properties. PPE37, a 473 amino acid unique protein, was shown to interfere with the inflammatory responses triggered by infected macrophages (Daim et al. 2011). In addition, PE3, a 468 amino acid unique protein, was shown to be overexpressed during the chronic steps of *Mtb* infection, and immunization with recombinant PE3 elicited significant levels of protective immune responses (Singh et al. 2013). Also, in a recent report, PE9 and PE10, two proteins encoded by adjacent co-expressed genes, were shown to physically interact and to promote macrophage apoptosis through interaction with TLR4 (Tiwari et al. 2015), demonstrating the possibility of heterodimerization between two PE proteins.

PE and PPE domains were shown to serve a similar functional role in different mycobacterial species by mediating secretion of the enzymatic domain with lipase activity (LipY). For example, in *Mtb* the N-terminal domain of LipY is a PE domain, in *M. marinum* LipY is a PPE domain and in fast-growing mycobacteria it is a signal sequence (Mishra et al. 2008; Daleke et al. 2011).

10.5 PPE_MPTR

PPE_MPTR proteins are characterized by a N-terminal PPE domain followed by a variable-size C-terminus. The C-terminal domain (300–4000 amino acids) is composed of multiple Asn-X-Gly-X-Gly-Asn-X-Gly repeats and arose from a recent evolutionary event (Gey van Pittius et al. 2006). In *Mtb* H37Rv, there are 20 *ppe_mptr* genes, which are commonly scattered throughout the chromosome, with few exceptions. For example, see the locus containing the genes encoding PPE_MPTR56-PPE_MPTR55-PE_PGRS50-PE_PGRS49-PPE_MPTR54 (Cole et al. 1998).

10.5.1 Genetic Variability

The repeat structure encoding the MPTR domain of these PPEs is clearly predicted to go through several variations due to either intramolecular recombination or polymerase slippage. Indeed, McEvoy and colleagues (2012) demonstrated that several *ppe_mptr* genes were subjected to extremely high levels of variation. However, macromolecular recombinatorial

events were shown to play a minor part in the variation of these genes. The authors focused, in particular, on six of the most variable *ppe_mptr* genes (*ppe_mptr*5/6 7/8, 24, 34, 54 and 55) and found that most of the mutations were due to non-synonymous single nucleotide polymorphisms (nSNPs), frameshifts or in-frame indels. Although these genes show an extremely high variability among strains belonging to different phylogenetic lineages, their sequence was found to be identical in closely related isolates suggesting a slow evolution. Another example of a *ppe_mptr* gene showing an elevated level of variation is *ppe_mptr42*, whose PCR-RFLP profile was shown to differ in 16% of isolates (Chakhaiyar et al. 2004).

10.5.2 Potential Role in Pathogenesis

The role of PPE_MPTR proteins in *Mtb* physiology or virulence is still unknown. Three mutants missing the genes encoding PPE_MPTR10, PPE_MPTR16 and PPE_MPTR21 were selected in a high-throughput screening of an *M. bovis* BCG transposon mutant library enriched in acidified phagosomes (Stewart et al. 2005).

The highly repetitive structure of PPE_MPTR suggests that they might act as dominant antigens. Chakhaiyar and colleagues (2004) for example, demonstrated that PPE_MPTR42 was able to elicit a significant humoral response in patients with relapsed TB, compared to a generally low T cell response in other TB patients. Finally, PPE_MPTR34 was shown to trigger functional maturation of human dendritic cells associated with the secretion of IL-10. This facilitates a shift toward a Th2-driven response, which is often ineffective against intracellular pathogens, and is suggestive of an immune evasion mechanism (Bansal et al. 2010b).

10.6 PE_PGRS

The PE_PGRS proteins share a molecular architecture characterized by the presence at the C-terminus of a polymorphic PGRS domain that varies in sequence and size, the presence of a conserved N-terminal PE domain of 90–100 amino acids in length and the existence of a highly conserved domain that links the PE and PGRS domains. The linker region consistently includes a sequence with the GRPLI motif which in four proteins is also found as an extra copy in the PGRS domain. A number of PE proteins, lacking the GRPLI and PGRS motifs were previously included in the PE_PGRS family, although the lack of these specific amino acid sequences (PGRS and GRPLI motif), suggest that these should rather be included in the PE unique group. Hence, in *Mtb* H37Rv, there are 51 *pe_pgrs* genes that are found scattered throughout the genome and that appear to be mostly expressed as a monocistronic operon (Delogu et al. 2008).

10.6.1 Protein Structure and Function

The paucity of experimental data on PE_PGRS proteins prevents a satisfactory understanding of their localization in the *Mtb* cell. However, PE_PGRS33 was found associated with the mycobacterial cell wall, with the PGRS domain available on the surface of *M. bovis* BCG and *Mtb*, and accessible for interactions with host components (Brennan et al. 2001; Delogu et al. 2004; Cascioferro et al. 2011; Zumbo et al. 2013; Palucci et al. 2016). Functional dissection of the PE_PGRS33 protein identified the first 140 amino acids of the protein as necessary and sufficient to mediate translocation and localization of the protein to the mycobacterial cell wall (Cascioferro et al. 2007). Indeed, these first 140 amino acids of PE_PGRS33, that contain the PE domain and the GRPLI motif, were used as a delivery system to express a heterologous protein on the mycobacterial surface (Sali et al. 2010). Proteomic analysis of lipid phase fractions enriched following TritonX-114 extraction, provided further evidence for the presence of a number of PE_PGRS proteins in the mycobacterial cell wall (De Souza and Wiker 2011). Green-fluorescent protein (GFP) fusions with PE_PGRS33 and PE_PGRS30 have been used to monitor protein localization and polarization in mycobacteria. The data indicates that while the first 140 amino acids are sufficient to drive GFP localization to the mycobacterial

surface, polarization can be affected by other sequences found in the PGRS domain or also in the region found between the PE domain and the GRPLI motif (Chatrath et al. 2011, 2014; De Maio et al. 2014).

A number of PE_PGRS proteins were shown to have immunomodulatory properties, at least in some cases, probably due to their ability to interact with TLR2 (Balaji et al. 2007; Bansal et al. 2010a; Chen et al. 2013; Singh et al. 2008; Zumbo et al. 2013; Palucci et al. 2016). Findings obtained in *M. marinum* indicate that PE_PGRS proteins are secreted into the media via an ESX-5 dependent mechanism (Abdallah et al. 2009). This suggests that these proteins, similar to other PE and PPE proteins, can translocate through the mycobacterial membranes. However, when experiments were performed in *Mtb*, secretion of PE_PGRS proteins by ESX-5 was not observed (Bottai et al. 2012). This implies that there might be significant differences between *Mtb* and *M. marinum* species in the translocation of PE_PGRSs.

The available experimental data and the significant homology observed between the amino acid sequences of PE_PGRSs, at least in the first 90 amino acids of the N-terminal domain and the GRPLI motif, point toward a shared mechanism of protein translocation among PE_PGRSs. Although it remains to be determined whether the GRPLI motif is also required for the anchoring of the protein on the mycomembrane, or if it has another yet undefined role in translocation. Another issue concerns the structure of the PE domain of PE_PGRS and whether, similarly to what has been shown for the PE/PPE couplets (Strong et al. 2006), interaction with another PPE protein to form heterodimers is required to warrant protein stability as predicted by computational studies (Riley et al. 2008). Characterization of the structure of the PE domain of PE_PGRS proteins may also provide a better understanding of how PE_PGRSs are localized to the mycobacterial cell wall and the role of the PE domain in this process.

The polymorphic PGRS domain shows the utmost level of diversity among the paralogs, with significant differences in size and sequence among the PE_PGRS proteins (see Chap. 9). However, the lack of any significant homology with other protein sequences in existing databases is a major obstacle toward understanding the functional role of this domain. The partial homology with the domain found in the EBNA protein of Epstein-Barr Virus, prompted some speculation on their role in the evasion of the cytotoxic T cell responses (Cole et al. 1998; Brennan and Delogu 2002), though this hypothesis has not been supported by experimental evidence. A number of PE_PGRS proteins have been implicated in the pathogenesis of TB, primarily with evidence showing that, following expression of PE_PGRS in saprophytic *M. smegmatis*, enhanced cytotoxity and inflammation were observed in in vitro and in vivo models. This has been described for PE_PGRS33 (Brennan et al. 2001; Dheenadhayalan et al. 2006; Balaji et al. 2007; Basu et al. 2007; Zumbo et al. 2013), for PE_PGRS16, for PE_PGRS26 (Singh et al. 2008), for PE_PGRS30 (Iantomasi et al. 2012; Chatrath et al. 2014), for PE_PGRS11 (Bansal et al. 2010a) and for PE_PGRS17 (Bansal et al. 2010a; Chen et al. 2013). In all these instances, the PGRS domain has been identified as a ligand, although it has not been possible to equate a specific region or sequence within the PGRS domain accountable for the observed activity.

Cadieux and colleagues (2011), established that PE_PGRS33 co-localizes to the mitochondria of transfected cells, a phenomenon dependent on the linker GRPLI region and the PGRS domain, but not the PE domain. Furthermore, although all constructs localizing to the mitochondria did induce apoptosis, only the wild-type PE_PGRS33 plus its own PE domain induced primary necrosis as well as apoptosis, signifying a potentially important role for the PE domain in this phenomenon. Considering the importance of primary necrosis in *Mtb* dissemination during natural infection, PE_PGRS33 may play a crucial role in TB pathogenesis. Interestingly, other PE_PGRS proteins tested (PE-PGRS1, -18 and -24) did not co-localize to the mitochondria of transfected cells, suggesting this is a phenomenon that may be unique to PE_PGRS33.

Functional and immunological characterization of PE_PGRS11 provided interesting

insights into the multifunctional role of these proteins during *Mtb* infections. PE_PGRS11 is a 584 amino acid protein, which has a highly conserved unique domain of 314 amino acids that encode a functional phosphoglycerate mutase (Chaturvedi et al. 2010) downstream of a relatively small PGRS domain (≈100 aa). The enzyme was properly active when the full length PE_PGRS11 was expressed in *M. smegmatis* and provided protection from oxidative stress. Interestingly, PE_PGRS11 was also capable of binding TLR2 inducing anti-apoptotic signals. It has been proposed that this surface exposed protein is engaged in the rescue of pathogenic *Mtb*-infected lung epithelial cells from oxidative stresses (Chaturvedi et al. 2010). The short PGRS domain could be responsible for the interaction with TLR2, and recent data, obtained with PE_PGRS33, indicates that even a small PGRS region, containing few repeats, can activate the TLR2-dependent entry into macrophages (Palucci et al. 2016). Hence, PE_PGRS11 is an example of the complexity of the functional potential of PE_PGRSs that can be seen as proteins with moonlighting properties. It would be of interest to explore whether this moonlighting property is a common trait for other PE_PGRSs, and it is expected that investigation of similar PE_PGRS proteins showing a large unique C-terminal domain (>100 aa in length) downstream of the PGRS (such as PE_PGRS11, PE_PGRS16, -17, -18, -30, -35, and -39) may help address this issue (Delogu et al. 2008).

The unique C-terminal domain found in PE_PGRS30 is highly homologous to the protein encoded by Rv3812 (which is still named PE_PGRS62 although it is missing two key features of PE_PGRSs, that is the GRPLI motif and the GGA-GGN repeats). Interestingly, the unique ≈ 300 amino acid long C-terminal domain in PE_PGRS30, but not the PGRS domain, has been shown to be dispensable for the full virulence of *Mtb* (Iantomasi et al. 2012).

The functional characterization of the PGRS domain remains elusive. The identification of the sequences within PGRS that interact with TLR2 or mediate other functions are unknown, although they likely include one or more of the GGA-GGN repeats.

Expression of a number of PE_PGRS proteins in *M. smegmatis*, or overexpression of PE_PGRSs in *Mtb*, results in changes in colony morphology and cell ultrastructure, suggesting significant alteration of the mycobacterial cell wall. However, it is not clear whether parts of the PGRS domain are embedded in the mycomembrane, or whether the observed changes in ultrastructure are simply a result of protein hindrance. The finding that under physiological conditions, where the PE_PGRS is expressed under the control of its own promoter in *Mtb*, the PGRS domain mediates protein polarization (De Maio et al. 2014; Chatrath et al. 2014) suggesting that this domain is directly interacting with other mycobacterial components found in the mycomembrane or in the periplasmic space.

It is expected that, despite the significant homology, a specific and unique function for all or most of the PE_PGRSs will be established. Differences in function between PE_PGRSs will likely be due to specific amino acids sequences found in the PGRS domain or regions found at the C-terminus of the protein. Based on the information available at this time, the protein specific intercalating sequences, of different sizes and patterns found between the GGA-GGN repeats, are likely to provide a specific function to each PE_PGRS.

10.6.2 Genetic Variability and Impact on Virulence and Pathogenesis

Demonstration that the polymorphic PGRS domain is available on the mycobacterial surface (Brennan et al. 2001; Banu et al. 2002; Delogu et al. 2004) and that it can be the target of the host humoral response (Singh et al. 2001; Delogu and Brennan 2001; Cohen et al. 2014) led further support to the hypothesis that PE_PGRSs were involved in antigenic variation. Sequencing of the *pe_pgrs33* gene in hundreds of different *Mtb* clinical isolates demonstrated a number of polymorphisms, mostly occurring in the PGRS region (Talarico et al. 2005, 2007) and characterized by the presence of single nucleotide polymor-

phisms (SNP)s, insertions or deletions. Interestingly, deletion of large sequences in the PGRS region of *pe_pgrs33* resulted in an attenuation of the ability of PE_PGRS33 to elicit TNFα (Basu et al. 2007). Also, alleles with large deletions in the *pgrs* region were found in *Mtb* strains isolated from patients with absence of cavitation in the lung (Talarico et al. 2007). Moreover, *Mtb* strains showing *pe_pgrs33* alleles significantly different from the dominant one, were associated with TB meningitis and negative PPD skin tests in children (Wang et al. 2011). These studies support the notion that variation in PE_PGRS33 could have consequences for the clinical features of TB. A remarkable level of genetic variation was also observed in PE_PGRS16 and PE_PGRS26, though the lack of an understanding of the function of these proteins in TB pathogenesis prevents a full assessment of the implications of these polymorphisms (Talarico et al. 2008).

Whole-genome sequences for many *Mtb* strains was not used to investigate sequence variation in *pe_pgrs* (and *ppe_mptr*) genes, since the GC-rich sequences are not reliably resolved by Next-Generation- Sequencing and often contain many sequence errors (McEvoy et al. 2012). A more systematic approach was taken by Copin and colleagues (2014), who Sanger sequenced 27 different *pe_pgrs* genes in a collection of 94 *Mtb* clinical isolates which were representatives of five phylogeographic lineages, and provided a broad overview on the genetic variability in this gene family (see Chap. 9). While *pe_pgrs* genes as a group were found to show more diversity than the rest of the MTBC genome, great differences in sequence diversity for different *pe_pgrs* genes were observed, with some genes under diversifying selection and others under purifying selection. This prompted the authors to conclude that distinct selection pressures act on individual genes. Genetic diversity was mostly observed in the *pgrs* sequences, with SNPs and indels, that when resulting in frameshift mutations were apparently mostly removed by purifying selection. The data obtained support the hypothesis that members of the *pe_pgrs* family have a distinct and non-redundant function and that *pe_pgrs* genes are mostly conserved in *Mtb* (Copin et al. 2014).

T cell immune responses play a key role in protection against TB. It has been assumed that genetic variation in *Mtb* serves to evade or manipulate host immune responses to avoid killing, promote disease and assure transmission (Comas et al. 2010). PE_PGRS proteins are enriched in T cell epitopes that appear to be restricted to the PE domain, which displays a low level of polymorphisms. In contrast, although the PGRS domain is polymorphic, it harbors only a few predicted T cell epitopes, prompting the authors to exclude that T cell recognition is the driving force of sequence diversity in *pe_pgrs* genes (see Chap. 9) (Copin et al. 2014). Hence, it remains to be determined what biological forces are driving the genetic diversity observed in the few *pe_pgrs* genes showing positive selection. Moreover, of the genetic variations observed in those genes that appear under purifying selection, what are the consequences of *Mtb*-host interaction and TB pathogenesis? Since most of the PE_PGRS proteins for which we have experimental data have been shown to interact with TLRs, it can be postulated that *pe_pgrs* genetic variation may have an impact on the immunomodulatory properties of the *Mtb* strains.

10.7 Practical Implications

10.7.1 Vaccines

Although one-third of the human population world-wide is infected with *Mtb*, we still do not understand what antigens are needed or what immune responses should be elicited to make a TB vaccine effective. A number of studies showed that PE and PPE proteins are the main target of T cell responses in humans and animals infected with *Mtb*, in different stages of infection and in different disease states (Dillon et al. 1999; Delogu and Brennan 2001; Vordermeier et al. 2012). Detailed analysis at the peptide level indicates that T cell epitopes are mainly found in the conserved N-terminal domains of these proteins (Lindestam Arlehamn et al. 2013; Copin et al. 2014) and that also cross-reactivity exists among the different PE and PPE antigens (Vordermeier et al. 2012). Moreover, T

cell responses directed against some of the PE and PPE proteins were shown to elicit protection in animal models (Skeiky et al. 2000) and PPE18 emerged as one of the most promising vaccine candidates and is currently included in the GSK vaccine Mtb72F (Skeiky et al. 2004). Mtb72f is a polyprotein subunit fusion vaccine that when adjuvanted with AS01B was able to elicit protection in several animal models, when either administered alone, or in a prime-boost strategy in previously BCG immunized animals (Tsenova et al. 2006; Reed et al. 2009). Results of phase I/II clinical trials showed that *Mtb*72f is safe and immunogenic in humans (Spertini et al. 2013). Another polyprotein subunit vaccine which exhibited good protection in animal models is ID93, which is composed of four *Mtb* antigens, one of which is PPE_MPTR42 (Bertholet et al. 2010).

The results obtained with these two vaccines indicates that PPE18 and PPE_MPTR42 are immuno-dominant, likely because they cross-react with many PPEs expressed by *Mtb* (Vordermeier et al. 2012). Indeed, mice immunized with a *Mtb* attenuated mutant lacking *ppe25* and *pe19*, were able to mount significant T cell responses when stimulated with peptides corresponding to shared epitopes of both PE and PPE proteins, demonstrating cross-reactivity of these antigens (Sayes et al. 2012). These data, together with other experimental evidence, which showed that protective activity in animal models is elicited by a number of other PE and PPE proteins (Romano et al. 2008; Parra et al. 2006; Vipond et al. 2006; Singh et al. 2013), highlights the vaccine potential of PE and PPE proteins.

As suggested by Comas and colleagues (2010), the immune responses to major antigens of *Mtb* are evolutionarily conserved to the benefit of the pathogen. Although the PPE and PE families were not included in that study, subsequent findings that many T cell epitopes in PE and PPE proteins are detected in the conserved regions (Copin et al. 2014), raises the same concerns that have been highlighted for other *Mtb* immune-dominant antigens (Comas et al. 2010).

Several PE and PPE proteins, and primarily PE_PGRS, PPE_MPTR, PE-unique and PPE-unique, are surface exposed antigens that directly interact with host components to modulate innate immune responses or exert other specific functions. Since the genes encoding these proteins showed a level of conservation greater than expected, and some of these proteins play a key role in TB pathogenesis, proteins of these families may represent a suitable target for a humoral response. In principle, antibodies specifically directed against the surface available domains may neutralize the activity of these proteins and provide some level of protection. Antibodies directed against the PGRS domain of PE_PGRS33 were observed in mice and humans with active TB (Singh et al. 2001; Delogu and Brennan 2001; Cohen et al. 2014), although immunization with a DNA vaccine that elicits antibodies against the PGRS domain did not provide a meaningful level of protection in mice (Delogu and Brennan 2001). It remains to be seen whether the antibodies developed following infection, or those induced by the PE_PGRS33 vaccine, are capable of properly binding the critical epitopes involved in the interaction with host components. A better functional characterization of these proteins is required before ruling out a potential activity of antibodies against these surface molecules. PGRS-containing vaccines could represent a promising target of a protective humoral response against *Mtb* (Delogu et al. 2014).

10.7.2 Drug Targets

The lack of functional information on the role of PE and PPE proteins have so far impeded studies aimed at targeting these proteins with drugs. Meszaros and colleagues (2011) suggested that PE and PPE should be considered ideal candidates as drug targets because they are specific for mycobacteria. Also, many of these proteins are found only in *Mtb*, and they may be less hostile to normal microbiota, which are often affected by the long lasting traditional anti-TB treatment (Wu et al. 2013). Some of the PE and PPE proteins, and primarily PE_PGRS,

were shown to be overrepresented in the lung tissue of infected guinea pigs (Kruh et al. 2010), implying that they may impact TB disease. In addition, these proteins are available on the surface and in the mycobacterial cell wall and are more accessible to drugs compared to mycobacterial cytoplasmic targets, that need to pass the highly impermeable mycobacterial cell wall.

Computational studies have proposed that a drug, capable of inhibiting or displacing the hydrophobic interaction between PE and PPE domains, may block important processes required for *Mtb* virulence (Meszaros et al. 2011). They also indicate that even the disordered domain found at the C-terminus of many PE and PPE proteins may be a valid drug target. As recently proposed (Ahmed et al. 2015), biological and molecular characterization of these PE and PPE may be instrumental for identifying potential drug targets. These drugs could then block the interaction between surface available PE and PPE and host receptors such as TLR2, and could be used for treating TB.

10.7.3 Diagnostics

Several studies have commenced to characterize the cellular and humoral immune responses to PE and PPE proteins found in *Mtb* infected subjects. A main goal in TB diagnosis is the identification of immunological biomarkers that can help distinguish between *Mtb* infected subjects (LTBI), and patients with active TB as well as normal subjects. In a recent study, stimulation of human whole-blood with PE35 and PPE68, using procedures routinely used for Interferon-γ release assays, but where IL-2 was assessed, often discriminated between active and LTBI patients (Pourakbari et al. 2015). Robust CD8 T cell responses, directed against PE_PGRS33, the PE-unique Rv3812 and PPE46, were also detected in LTBI subjects. A better immunological characterization of these proteins, particularly those included in the RD regions which are missing in the vaccine strain, BCG (see Chap. 8), may lead to the development of improved Interferon-γ release assays with improved prognostic value. Moreover, serological diagnosis of TB may potentially be improved by the use of a number of PE and PPE proteins for which a robust antibody response has been observed, such as for PE_PGRS33, PPE41 (Choudhary et al. 2003), PPE55 (Singh et al. 2005), PPE57 (Zhang et al. 2007), and PPE17 (Khan et al. 2008).

10.8 Conclusion

Despite a growing interest in PE and PPE proteins since their discovery, their role in *Mtb* physiology and pathogenicity is still elusive. In summary, in the last few years we have learned that: (i) PE/PPE couplets are substrates of the ESX T7SSs whereby they are secreted or surface exposed; (ii) several PE and PPE proteins have the ability to modulate the host immune system by directly engaging with TLRs; (iii) PE and PPE proteins are also the main targets of the adaptive host T cell response; (iv) PE and PPE domains can serve as shuttles to deliver PGRS and MPTR domains to the cell surface, or other unique domains, some of which possess enzymatic activity as LipY or PE_PGRS11. While the role of the PE and PPE domains has been associated with protein localization, the biological significance of the structural redundancy observed in PE_PGRSs and PPE_PMTRs is still a matter of debate. Likewise, the consequences of the differential regulation of these abundant genes in *Mtb* during the infectious process are unclear. A major puzzle is that, despite the abundance of these genes and the presence of a degree of genetic polymorphisms higher than the average *Mtb* genome, most *pe* and *ppe* genes, including the polymorphic *pe_pgrs* and *ppe_mptr*, are under neutral or purifying selection. This suggests that the corresponding proteins have, despite the presence of repeat motifs, structural and/or functional constraints and that they are critical for TB pathogenesis. A better understanding of the role and function of the PE and PPE proteins will require clarification of the impact of the contradictory forces of both conservation and diversity acting on *pe* and *ppe* genes.

References

Abdallah AM, Verboom T, Weerdenburg EM, Gey van Pittius NC, Mahasha PW, Jimenez C et al (2009) PPE and PE_PGRS proteins of Mycobacterium marinum are transported via the type VII secretion system ESX-5. Mol Microbiol 73:329–340

Adindla S, Guruprasad L (2003) Sequence analysis corresponding to the PPE and PE proteins in Mycobacterium tuberculosis and other genomes. J Biosci 28:169–179

Ahmed A, Das A, Mukhopadhyay S (2015) Immunoregulatory functions and expression patterns of PE/PPE family members: roles in pathogenicity and impact on anti-tuberculosis vaccine and drug design. IUBMB Life 67:414–427

Akhter Y, Ehebauer MT, Mukhopadhyay S, Hasnain SE (2012) The PE/PPE multigene family codes for virulence factors and is a possible source of mycobacterial antigenic variation: perhaps more? Biochimie 94:110–116

Ates LS, Houben EN, Bitter W (2016) Type VII secretion: a highly versatile secretion system. Microbiol Spectr 4

Balaji KN, Goyal G, Narayana Y, Srinivas M, Chaturvedi R, Mohammad S (2007) Apoptosis triggered by Rv1818c, a PE family gene from Mycobacterium tuberculosis is regulated by mitochondrial intermediates in T cells. Microbes Infect 9:271–281

Bansal K, Elluru SR, Narayana Y, Chaturvedi R, Patil SA, Kaveri SV et al (2010a) PE_PGRS antigens of Mycobacterium tuberculosis induce maturation and activation of human dendritic cells. J Immunol 184:3495–3504

Bansal K, Sinha AY, Ghorpade DS, Togarsimalemath SK, Patil SA, Kaveri SV et al (2010b) Src homology 3-interacting domain of Rv1917c of Mycobacterium tuberculosis induces selective maturation of human dendritic cells by regulating PI3K-MAPK-NF-kappaB signaling and drives Th2 immune responses. J Biol Chem 285:36511–36522

Banu S, Honore N, Saint-Joanis B, Philpott D, Prevost MC, Cole ST (2002) Are the PE-PGRS proteins of Mycobacterium tuberculosis variable surface antigens? Mol Microbiol 44:9–19

Basu S, Pathak SK, Banerjee A, Pathak S, Bhattacharyya A, Yang Z et al (2007) Execution of macrophage apoptosis by PE_PGRS33 of Mycobacterium tuberculosis is mediated by Toll-like receptor 2-dependent release of tumor necrosis factor-alpha. J Biol Chem 282:1039–1050

Bertholet S, Ireton GC, Ordway DJ, Windish HP, Pine SO, Kahn M et al (2010) A defined tuberculosis vaccine candidate boosts BCG and protects against multidrug-resistant Mycobacterium tuberculosis. Sci Transl Med 2:53ra74

Bhat KH, Ahmed A, Kumar S, Sharma P, Mukhopadhyay S (2012) Role of PPE18 protein in intracellular survival and pathogenicity of Mycobacterium tuberculosis in mice. PLoS One 7:e52601

Bottai D, Di LM, Majlessi L, Frigui W, Simeone R, Sayes F et al (2012) Disruption of the ESX-5 system of Mycobacterium tuberculosis causes loss of PPE protein secretion, reduction of cell wall integrity and strong attenuation. Mol Microbiol 83:1209

Brennan MJ, Delogu G (2002) The PE multigene family: a 'molecular mantra' for mycobacteria. Trends Microbiol 10:246–249

Brennan MJ, Delogu G, Chen Y, Bardarov S, Kriakov J, Alavi M, Jacobs WR Jr (2001) Evidence that mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells. Infect Immun 69:7326–7333

Cadieux N, Parra M, Cohen H, Maric D, Morris SL, Brennan MJ (2011) Induction of cell death after localization to the host cell mitochondria by the Mycobacterium tuberculosis PE_PGRS33 protein. Microbiology 157:793–804

Cascioferro A, Delogu G, Colone M, Sali M, Stringaro A, Arancia G et al (2007) PE is a functional domain responsible for protein translocation and localization on mycobacterial cell wall. Mol Microbiol 66:1536–1547

Cascioferro A, Daleke MH, Ventura M, Dona V, Delogu G, Palu G et al (2011) Functional dissection of the PE domain responsible for translocation of PE_PGRS33 across the mycobacterial cell wall. PLoS One 6:e27713

Chakhaiyar P, Nagalakshmi Y, Aruna B, Murthy KJ, Katoch VM, Hasnain SE (2004) Regions of high antigenicity within the hypothetical PPE major polymorphic tandem repeat open-reading frame, Rv2608, show a differential humoral response and a low T cell response in various categories of patients with tuberculosis. J Infect Dis 190:1237–1244

Chatrath S, Gupta VK, Dixit A, Garg LC (2011) The Rv1651c-encoded PE-PGRS30 protein expressed in Mycobacterium smegmatis exhibits polar localization and modulates its growth profile. FEMS Microbiol Lett 322:194–199

Chatrath S, Gupta VK, Garg LC (2014) The PGRS domain is responsible for translocation of PE_PGRS30 to cell poles while the PE and the C-terminal domains localize it to the cell wall. FEBS Lett 588:990–994

Chaturvedi R, Bansal K, Narayana Y, Kapoor N, Sukumar N, Togarsimalemath SK et al (2010) The multifunctional PE_PGRS11 protein from Mycobacterium tuberculosis plays a role in regulating resistance to oxidative stress. J Biol Chem 285:30389–30403

Chen T, Zhao Q, Li W, Xie J (2013) Mycobacterium tuberculosis PE_PGRS17 promotes the death of host cell and cytokines secretion via Erk kinase accompanying with enhanced survival of recombinant Mycobacterium smegmatis. J Interf Cytokine Res 33:452–458

Choudhary RK, Mukhopadhyay S, Chakhaiyar P, Sharma N, Murthy KJ, Katoch VM, Hasnain SE (2003) PPE antigen Rv2430c of Mycobacterium tuberculosis induces a strong B-cell response. Infect Immun 71:6338–6343

Cohen I, Parada C, Acosta-Gio E, Espitia C (2014) The PGRS Domain from PE_PGRS33 of Mycobacterium tuberculosis is Target of Humoral Immune Response in Mice and Humans. Front Immunol 5:236

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D et al (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence [see comments] [published erratum appears in Nature 1998 Nov 12;396(6707):190]. Nature 393:537–544

Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, Kremer K et al (2010) Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved. Nat Genet 42:498–503

Copin R, Coscolla M, Seiffert SN, Bothamley G, Sutherland J, Mbayo G et al (2014) Sequence diversity in the pe_pgrs genes of Mycobacterium tuberculosis is independent of human T cell recognition. MBio 5:e00960–e00913

Daim S, Kawamura I, Tsuchiya K, Hara H, Kurenuma T, Shen Y et al (2011) Expression of the Mycobacterium tuberculosis PPE37 protein in Mycobacterium smegmatis induces low tumour necrosis factor alpha and interleukin 6 production in murine macrophages. J Med Microbiol 60:582–591

Daleke MH, Cascioferro A, de Punder K, Ummels R, Abdallah AM, van der Wel N et al (2011) Conserved PE and PPE protein domains target LipY lipases of pathogenic mycobacteria to the cell surface via ESX-5. J Biol Chem

De Maio F, Maulucci G, Minerva M, Anoosheh S, Palucci I, Iantomasi R et al (2014) Impact of protein domains on PE_PGRS30 polar localization in Mycobacteria. PLoS One 9:e112482

De Souza GA, Wiker HG (2011) A proteomic view of mycobacteria. Proteomics 11:3118–3127

Delogu, G, Brennan, MJ (2001) Comparative Immune Response to PE and PE_PGRS Antigens of Mycobacterium tuberculosis. Infect Immun 69: 5606–5611

Delogu G, Pusceddu C, Bua A, Fadda G, Brennan MJ, Zanetti S (2004) Rv1818c-encoded PE_PGRS protein of Mycobacterium tuberculosis is surface exposed and influences bacterial cell structure. Mol Microbiol 52:725–733

Delogu G, Sanguinetti M, Pusceddu C, Bua A, Brennan MJ, Zanetti S, Fadda G (2006) PE_PGRS proteins are differentially expressed by Mycobacterium tuberculosis in host tissues. Microbes Infect 8:2061–2067

Delogu G, Cole ST, Brosch R (2008) The PE and PPE protein families of mycobacterium tuberculosis. In: Kaufmann SH, Rubin E (eds) Handbook of tuberculosis. Wiley-VCH Verlag GmbH%Co. KGaA, Weinheim, pp 131–150

Delogu, G, Manganelli, R, Brennan, MJ (2014) Critical research concepts in tuberculosis vaccine development. Clin Microbiol Infect 20 Suppl 5: 59–65

Dheenadhayalan V, Delogu G, Brennan MJ (2006) Expression of the PE_PGRS 33 protein in Mycobacterium smegmatis triggers necrosis in macrophages and enhanced mycobacterial survival. Microbes Infect 8:262–272

Dillon DC, Alderson MR, Day CH, Lewinsohn DM, Coler R, Bement T et al (1999) Molecular characterization and human T-cell responses to a member of a novel Mycobacterium tuberculosis mtb39 gene family. Infect Immun 67:2941–2950

Dona V, Ventura M, Sali M, Cascioferro A, Provvedi R, Palu G et al (2013) The PPE domain of PPE17 is responsible for its surface localization and can be used to express heterologous proteins on the mycobacterial surface. PLoS One 8:e57517

Dong D, Wang D, Li M, Wang H, Yu J, Wang C et al (2012) PPE38 modulates the innate immune response and is required for Mycobacterium marinum virulence. Infect Immun 80:43–54

Ekiert DC, Cox JS (2014) Structure of a PE-PPE-EspG complex from Mycobacterium tuberculosis reveals molecular specificity of ESX protein secretion. Proc Natl Acad Sci U S A 111:14758–14763

Fishbein S, van Wyk N, Warren RM, Sampson SL (2015) Phylogeny to function: PE/PPE protein evolution and impact on Mycobacterium tuberculosis pathogenicity. Mol Microbiol 96:901–916

Gao Q, Kripke KE, Saldanha AJ, Yan W, Holmes S, Small PM (2005) Gene expression diversity among *Mycobacterium tuberculosis* clinical isolates. Microbiology 151:5–14

Gey van Pittius NC, Sampson SL, Lee H, Kim Y, van Helden PD, Warren RM (2006) Evolution and expansion of the Mycobacterium tuberculosis PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. BMC Evol Biol 6:95

Golby P, Hatch KA, Bacon J, Cooney R, Riley P, Allnutt J et al (2007) Comparative transcriptomics reveals key gene expression differences between the human and bovine pathogens of the Mycobacterium tuberculosis complex. Microbiology 153:3323–3336

Gold B, Rodriguez GM, Marras SAE, Pentecost M, Smith I (2001) The *Mycobacterium tuberculosis* IdeR is a dual functional regulator that controls transcription of genes involved in iron acquisition, iron storage and survival in macrophages. Mol Microbiol 42:851–865

Goldstone RM, Goonesekera SD, Bloom BR, Sampson SL (2009) The transcriptional regulator Rv0485 modulates the expression of a pe and ppe gene pair and is required for Mycobacterium tuberculosis virulence. Infect Immun 77:4654–4667

Hebert AM, Talarico S, Yang D, Durmaz R, Marrs CF, Zhang L et al (2007) DNA polymorphisms in the pepA and PPE18 genes among clinical strains of Mycobacterium tuberculosis: implications for vaccine efficacy. Infect Immun 75:5798–5805

Homolka S, Ubben T, Niemann S (2016) High sequence variability of the ppE18 gene of clinical Mycobacterium tuberculosis complex strains potentially impacts effectivity of vaccine candidate M72/AS01E. PLoS One 11:e0152200

Iantomasi R, Sali M, Cascioferro A, Palucci I, Zumbo A, Soldini S et al (2012) PE_PGRS30 is required for the full virulence of Mycobacterium tuberculosis. Cell Microbiol 14:356–367

Karboul A, Gey van Pittius NC, Namouchi A, Vincent V, Sola C, Rastogi N et al (2006) Insights into the evolutionary history of tubercle bacilli as disclosed by genetic rearrangements within a PE_PGRS duplicated gene pair. BMC Evol Biol 6:107

Karboul A, Mazza A, Gey van Pittius NC, Ho JL, Brousseau R, Mardassi H (2008) Frequent homologous recombination events in Mycobacterium tuberculosis PE/PPE multigene families: potential role in antigenic variability. J Bacteriol 190:7838–7846

Khan N, Alam K, Nair S, Valluri VL, Murthy KJ, Mukhopadhyay S (2008) Association of strong immune responses to PPE protein Rv1168c with active tuberculosis. Clin Vaccine Immunol 15:974–980

Korotkova N, Freire D, Phan TH, Ummels R, Creekmore CC, Evans TJ et al (2014) Structure of the Mycobacterium tuberculosis type VII secretion system chaperone EspG5 in complex with PE25-PPE41 dimer. Mol Microbiol 94:367–382

Kruh NA, Troudt J, Izzo A, Prenni J, Dobos KM (2010) Portrait of a pathogen: the Mycobacterium tuberculosis proteome in vivo. PLoS One 5:e13938

Lee JH, Karakousis PC, Bishai WR (2008) Roles of SigB and SigF in the Mycobacterium tuberculosis sigma factor network. J Bacteriol 190:699–707

Lindestam Arlehamn CS, Gerasimova A, Mele F, Henderson R, Swann J, Greenbaum JA et al (2013) Memory T cells in latent Mycobacterium tuberculosis infection are directed against three antigenic islands and largely contained in a CXCR3+CCR6+ Th1 subset. PLoS Pathog 9:e1003130

McEvoy CR, van Helden PD, Warren RM, Gey van Pittius NC (2009) Evidence for a rapid rate of molecular evolution at the hypervariable and immunogenic Mycobacterium tuberculosis PPE38 gene region. BMC Evol Biol 9:237

McEvoy CR, Cloete R, Muller B, Schurch AC, van Helden PD, Gagneux S et al (2012) Comparative analysis of Mycobacterium tuberculosis pe and ppe genes reveals high sequence variation and an apparent absence of selective constraints. PLoS One 7:e30593

Meszaros B, Toth J, Vertessy BG, Dosztanyi Z, Simon I (2011) Proteins with complex architecture as potential targets for drug design: a case study of Mycobacterium tuberculosis. PLoS Comput Biol 7:e1002118

Mishra KC, de Chastellier C, Narayana Y, Bifani P, Brown AK, Besra GS et al (2008) Functional role of the PE domain and immunogenicity of the Mycobacterium tuberculosis triacylglycerol hydrolase LipY. Infect Immun 76:127–140

Mohareer K, Tundup S, Hasnain SE (2011) Transcriptional regulation of Mycobacterium tuberculosis PE/PPE genes: a molecular switch to virulence? J Mol Microbiol Biotechnol 21:97–109

Mortier MC, Jongert E, Mettens P, Ruelle JL (2015) Sequence conservation analysis and in silico human leukocyte antigen-peptide binding predictions for the Mtb72F and M72 tuberculosis candidate vaccine antigens. BMC Immunol 16:63

Mukhopadhyay, S, Balaji, KN (2011) The PE and PPE proteins of Mycobacterium tuberculosis. Tuberculosis (Edinb) 91: 441–447

Nair S, Ramaswamy PA, Ghosh S, Joshi DC, Pathak N, Siddiqui I et al (2009) The PPE18 of Mycobacterium tuberculosis interacts with TLR2 and activates IL-10 induction in macrophage. J Immunol 183:6269–6281

Namouchi A, Karboul A, Fabre M, Gutierrez MC, Mardassi H (2013) Evolution of smooth tubercle Bacilli PE and PE_PGRS genes: evidence for a prominent role of recombination and imprint of positive selection. PLoS One 8:e64718

Palucci I, Camassa S, Cascioferro A, Sali M, Anoosheh S, Zumbo A et al (2016) PE_PGRS33 Contributes to Mycobacterium tuberculosis Entry in Macrophages through Interaction with TLR2. PLoS One 11:e0150800

Parra M, Cadieux N, Pickett T, Dheenadhayalan V, Brennan MJ (2006) A PE protein expressed by Mycobacterium avium is an effective T-cell immunogen. Infect Immun 74:786–789

Poulet S, Cole ST (1995) Characterization of the highly abundant polymorphic GC-rich-repetitive sequence (PGRS) present in Mycobacterium tuberculosis. Arch Microbiol 163:87–95

Pourakbari B, Mamishi S, Marjani M, Rasulinejad M, Mariotti S, Mahmoudi S (2015) Novel T-cell assays for the discrimination of active and latent tuberculosis infection: the diagnostic value of PPE family. Mol Diagn Ther 19:309–316

Pym AS, Brodin P, Brosch R, Huerre M, Cole ST (2002) Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines Mycobacterium bovis BCG and Mycobacterium microti. Mol Microbiol 46:709–717

Rachman H, Strong M, Ulrichs T, Grode L, Schuchhardt J, Mollenkopf H et al (2006) Unique transcriptome signature of Mycobacterium tuberculosis in pulmonary tuberculosis. Infect Immun 74:1233–1242

Ramakrishnan L, Federspiel NA, Falkow S (2000) Granuloma-Specific Expression of Mycobacterium Virulence Proteins from the Glycine-Rich PE-PGRS Family. Science 288:1436–1439

Raman S, Hazra R, Dascher CC, Husson RN (2004) Transcription regulation by the Mycobacterium tuberculosis alternative sigma factor SigD and its role in virulence. J Bacteriol 186:6605–6616

Reed SG, Coler RN, Dalemans W, Tan EV, DeLa Cruz EC, Basaraba RJ et al (2009) Defined tuberculosis vaccine, Mtb72F/AS02A, evidence of protection in cynomolgus monkeys. Proc Natl Acad Sci U S A 106: 2301–2306

Rehren G, Walters S, Fontan P, Smith I, Zarraga AM (2007) Differential gene expression between Mycobacterium bovis and Mycobacterium tuberculosis. Tuberculosis (Edinb) 87:347–359

Riley R, Pellegrini M, Eisenberg D (2008) Identifying cognate binding pairs among a large set of paralogs: the case of PE/PPE proteins of Mycobacterium tuberculosis. PLoS Comput Biol 4:e1000174

Rodriguez GM, Voskuil MI, Gold B, Schoolnik GK, Smith I (2002) ideR, An essential gene in mycobacterium tuberculosis: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. Infect Immun 70:3371–3381

Romano M, Rindi L, Korf H, Bonanni D, Adnet PY, Jurion F et al (2008) Immunogenicity and protective efficacy of tuberculosis subunit vaccines expressing PPE44 (Rv2770c). Vaccine 26:6053–6063

Sali M, Di SG, Cascioferro A, Zumbo A, Nicolo C, Dona V et al (2010) Surface expression of MPT64 as a fusion with the PE domain of PE_PGRS33 enhances Mycobacterium bovis BCG protective activity against Mycobacterium tuberculosis in mice. Infect Immun 78:5202–5213

Sassi M, Drancourt M (2014) Genome analysis reveals three genomospecies in Mycobacterium abscessus. BMC Genomics 15:359

Sayes F, Sun L, Di LM, Simeone R, Degaiffier N, Fiette L et al (2012) Strong immunogenicity and cross-reactivity of Mycobacterium tuberculosis ESX-5 type VII secretion: encoded PE-PPE proteins predicts vaccine potential. Cell Host Microbe 11:352–363

Serafini A, Pisu D, Palu G, Rodriguez GM, Manganelli R (2013) The ESX-3 secretion system is necessary for iron and zinc homeostasis in Mycobacterium tuberculosis. PLoS One 8:e78351

Simeone R, Bottai D, Brosch R (2009) ESX/type VII secretion systems and their role in host-pathogen interaction. Curr Opin Microbiol 12:4–10

Singh KK, Zhang X, Patibandla AS, Chien P Jr, Laal S (2001) Antigens of Mycobacterium tuberculosis expressed during preclinical tuberculosis: serological immunodominance of proteins with repetitive amino acid sequences. Infect Immun 69:4185–4191

Singh KK, Dong Y, Patibandla SA, McMurray DN, Arora VK, Laal S (2005) Immunogenicity of the Mycobacterium tuberculosis PPE55 (Rv3347c) protein during incipient and clinical tuberculosis. Infect Immun 73:5004–5014

Singh, PP, Parra, M, Cadieux, N, Brennan, MJ (2008) A comparative study of host response to three Mycobacterium tuberculosis PE_PGRS proteins. Microbiology 154: 3469–3479

Singh SK, Tripathi DK, Singh PK, Sharma S, Srivastava KK (2013) Protective and survival efficacies of Rv0160c protein in murine model of Mycobacterium tuberculosis. Appl Microbiol Biotechnol 97:5825–5837

Singh, P, Rao, RN, Reddy, JR, Prasad, R, Kotturu, SK, Ghosh, S, Mukhopadhyay, S (2016) PE11, a PE/PPE family protein of Mycobacterium tuberculosis is involved in cell wall remodeling and virulence. Sci Rep 6: 21624

Skeiky YA, Ovendale PJ, Jen S, Alderson MR, Dillon DC, Smith S et al (2000) T cell expression cloning of a Mycobacterium tuberculosis gene encoding a protective antigen associated with the early control of infection. J Immunol 165:7140–7149

Skeiky YA, Alderson MR, Ovendale PJ, Guderian JA, Brandt L, Dillon DC et al (2004) Differential immune responses and protective efficacy induced by components of a tuberculosis polyprotein vaccine, Mtb72F, delivered as naked DNA or recombinant protein. J Immunol 172:7618–7628

Solans L, Gonzalo-Asensio J, Sala C, Benjak A, Uplekar S, Rougemont J et al (2014) The PhoP-dependent ncRNA Mcr7 modulates the TAT secretion system in Mycobacterium tuberculosis. PLoS Pathog 10:e1004183

Soldini S, Palucci I, Zumbo A, Sali M, Ria F, Manganelli R et al (2011) PPE_MPTR genes are differentially expressed by Mycobacterium tuberculosis in vivo. Tuberculosis (Edinb)

Spertini F, Audran R, Lurati F, Ofori-Anyinam O, Zysset F, Vandepapeliere P et al (2013) The candidate tuberculosis vaccine Mtb72F/AS02 in PPD positive adults: a randomized controlled phase I/II study. Tuberculosis (Edinb) 93:179–188

Stewart GR, Patel J, Robertson BD, Rae A, Young DB (2005) Mycobacterial mutants with defective control of phagosomal acidification. PLoS Pathog 1:269–278

Strong M, Sawaya MR, Wang S, Phillips M, Cascio D, Eisenberg D (2006) Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 103:8060–8065

Supply P, Marceau M, Mangenot S, Roche D, Rouanet C, Khanna V et al (2013) Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of Mycobacterium tuberculosis. Nat Genet 45:172–179

Talarico S, Cave MD, Marrs CF, Foxman B, Zhang L, Yang Z (2005) Variation of the Mycobacterium tuberculosis PE_PGRS 33 gene among clinical isolates. J Clin Microbiol 43:4954–4960

Talarico S, Cave MD, Foxman B, Marrs CF, Zhang L, Bates JH, Yang Z (2007) Association of Mycobacterium tuberculosis PE PGRS33 polymorphism with clinical and epidemiological characteristics. Tuberculosis (Edinb) 87:338–346

Talarico S, Zhang L, Marrs CF, Foxman B, Cave MD, Brennan MJ, Yang Z (2008) Mycobacterium tuberculosis PE_PGRS16 and PE_PGRS26 genetic polymorphism among clinical isolates. Tuberculosis (Edinb) 88:283–294

Tiwari B, Ramakrishnan UM, Raghunand TR (2015) The Mycobacterium tuberculosis protein pair PE9 (Rv1088)-PE10 (Rv1089) forms heterodimers and induces macrophage apoptosis through Toll-like receptor 4. Cell Microbiol 17:1653–1669

Tsenova L, Harbacheuski R, Moreira AL, Ellison E, Dalemans W, Alderson MR et al (2006) Evaluation of the Mtb72F polyprotein vaccine in a rabbit model of tuberculous meningitis. Infect Immun 74:2392–2401

Tufariello JM, Chapman JR, Kerantzas CA, Wong KW, Vilcheze C, Jones CM et al (2016) Separable roles for Mycobacterium tuberculosis ESX-3 effectors in iron

acquisition and virulence. Proc Natl Acad Sci U S A 113:E348–E357

Tundup S, Akhter Y, Thiagarajan D, Hasnain SE (2006) Clusters of PE and PPE genes of Mycobacterium tuberculosis are organized in operons: evidence that PE Rv2431c is co-transcribed with PPE Rv2430c and their gene products interact with each other. FEBS Lett 580:1285–1293

Vipond J, Vipond R, Allen-Vercoe E, Clark SO, Hatch GJ, Gooch KE et al (2006) Selection of novel TB vaccine candidates and their evaluation as DNA vaccines against aerosol challenge. Vaccine 24:6340–6350

Vordermeier HM, Hewinson RG, Wilkinson RJ, Wilkinson KA, Gideon HP, Young DB, Sampson SL (2012) Conserved immune recognition hierarchy of mycobacterial PE/PPE proteins during infection in natural hosts. PLoS One 7:e40890

Voskuil MI, Schnappinger D, Rutherford R, Liu Y, Schoolnik GK (2004) Regulation of the *Mycobacterium tuberculosis* PE/PPE genes. Tuberculosis (Edinb) 84:256–262

Wang J, Huang Y, Zhang A, Zhu C, Yang Z, Xu H (2011) DNA polymorphism of Mycobacterium tuberculosis PE_PGRS33 gene among clinical isolates of pediatric TB patients and its associations with clinical presentation. Tuberculosis (Edinb) 91:287–292

Wu J, Liu W, He L, Huang F, Chen J, Cui P et al (2013) Sputum microbiota associated with new, recurrent and treatment failure tuberculosis. PLoS One 8:e83445

Zhang H, Wang J, Lei J, Zhang M, Yang Y, Chen Y, Wang H (2007) PPE protein (Rv3425) from DNA segment RD11 of Mycobacterium tuberculosis: a potential B-cell antigen used for serological diagnosis to distinguish vaccinated controls from tuberculosis patients. Clin Microbiol Infect 13:139–145

Zumbo A, Palucci I, Cascioferro A, Sali M, Ventura M, D'Alfonso P et al (2013) Functional dissection of protein domains involved in the immunomodulatory properties of PE_PGRS33 of Mycobacterium tuberculosis. Pathog Dis 69: 232–239

Epidemiology of Drug-Resistant Tuberculosis

11

Anna S. Dean, Helen Cox, and Matteo Zignol

Abstract

As we move into the era of the Sustainable Development Goals (SDGs), the World Health Organization (WHO) has developed the End TB strategy 2016–2035 with a goal to end the global epidemic of tuberculosis (TB) by 2035. Achieving the targets laid out in the Strategy will require strengthening of the whole TB diagnosis and treatment cascade, including improved case detection, the establishment of universal drug susceptibility testing and rapid treatment initiation. An estimated 3.9% of new TB cases and 21% of previously treated cases had rifampicin-resistant (RR) or multidrug-resistant (MDR) TB in 2015. These levels have remained stable over time, although limited data are available from major high burden settings. In addition to the emergence of drug resistance due to inadequate treatment, there is growing evidence that direct transmission is a large contributor to the RR/MDR-TB epidemic. Only 340,000 of the estimated 580,000 incident cases of RR/MDR-TB were notified to WHO in 2015. Among these, only 125,000 were initiated on second-line treatment. RR/MDR-TB epidemics are likely to be driven by direct transmission. The most important risk factor for MDR-TB is a history of previous treatment. Other risk factors vary according to setting but can include hospitalisation, incarceration and HIV infection. Children have the same risk of MDR-TB as adults and represent a diagnostic and treatment challenge. Rapid molecular technologies have revolutionized the diagnosis of drug-resistant TB. Until capacity can be established to

A.S. Dean (✉) • M. Zignol
Global TB Programme, World Health Organization, Geneva, Switzerland
e-mail: deanan@who.int zignolm@who.int

H. Cox
Division of Medical Microbiology and Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa
e-mail: helen.cox@uct.ac.za

S. Gagneux (ed.), *Strain Variation in the* Mycobacterium tuberculosis *Complex: Its Role in Biology, Epidemiology and Control*, Advances in Experimental Medicine and Biology 1019,
DOI 10.1007/978-3-319-64371-7_11

test every TB patient for rifampicin resistance, countries should focus on gradually expanding their coverage of testing. DNA sequencing technologies are being increasingly incorporated into patient management and drug resistance surveillance. They offer additional benefits over conventional culture-based phenotypic testing, including a faster turn-around time for results, assessment of resistance patterns to a range of drugs, and investigation of strain clustering and transmission.

Keywords
Epidemiology • Surveillance • Antimicrobial Resistance • Transmission • Diagnostics

11.1 Emergence and Spread of Drug Resistance in the World

Drug-resistant *Mycobacterium tuberculosis* was first documented in the late 1940s, shortly after the introduction of the first anti-tuberculosis (TB) treatment, streptomycin (Crofton and Mitchison 1948). However, drug-resistant TB did not receive global attention until the 1990s when large outbreaks of multidrug-resistant tuberculosis (MDR-TB) were identified in the United States and Europe (Monno et al. 1991; Centers for Disease Control (CDC) 1991; Frieden et al. 1993). The urgent need for global monitoring of the emergence and spread of resistance to anti-TB drugs became clear. In 1994, the Global Tuberculosis Programme of the World Health Organization (WHO) and the International Union Against Tuberculosis and Lung Disease (The Union) established the Global Project on Anti-Tuberculosis Drug Resistance Surveillance (Pablos-Méndez et al. 1998). This remains the oldest and largest initiative for antimicrobial resistance surveillance in the world (Zignol et al. 2016).

In 2014, the World Health Assembly called upon WHO member states and the international community to take stronger measures to prevent and control the spread of antimicrobial resistance. As we move into the era of the Sustainable Development Goals (SDGs), WHO has developed the End TB Strategy 2016–2035 with a goal to end the global TB epidemic by 2035 (Uplekar et al. 2015; World Health Organization). The Strategy calls for universal drug susceptibility testing of all TB patients to ensure more rapid diagnosis of drug-resistant TB and initiation of appropriate treatment (World Health Organization 2016).

11.2 Coverage of Surveillance Data

High quality data on the burden and distribution of drug resistance are essential to understanding the global epidemic of drug-resistant TB. In 2015, only 30% of the 3.4 million new bacteriologically-confirmed and previously treated TB cases notified to WHO were tested for susceptibility to rifampicin. Major gains in coverage will be needed in order to achieve universal drug susceptibility testing. By 2015, reliable data on anti-TB drug resistance were available from 155 countries, representing more than 95% of the world's population and TB burden. This is a more than fourfold increase compared to the early years of the Global Project when data were available for 35 countries only (World Health Organization 2016).

Epidemiological data on the burden of drug resistance are derived from two principal sources: continuous surveillance systems based on routine drug susceptibility testing of TB patients and ad hoc surveys of a representative group of TB patients. Testing results for at least rifampicin and isoniazid are collected for both

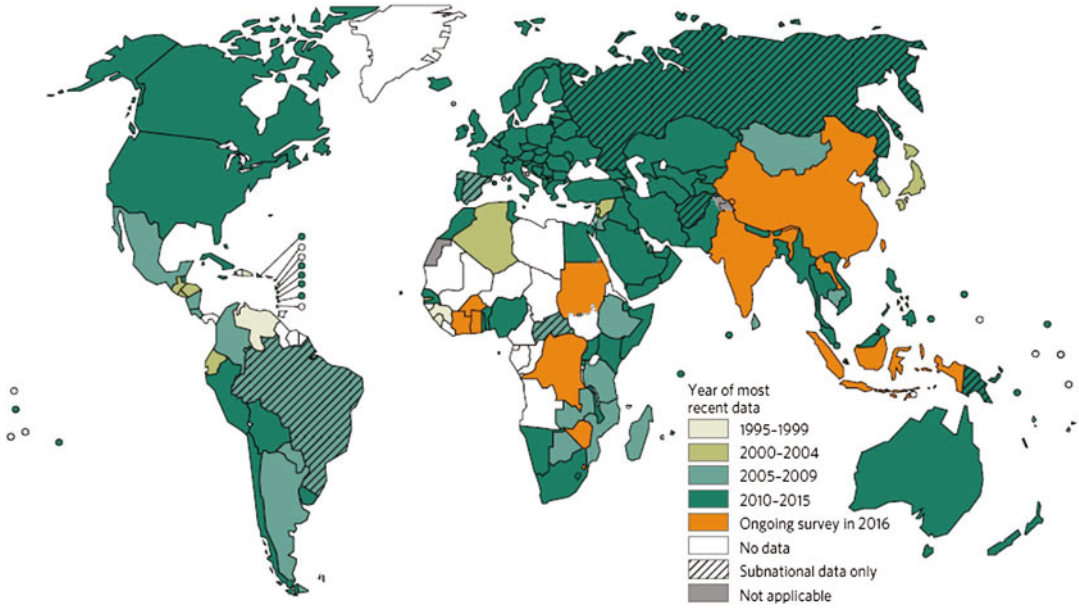

Fig. 11.1 Global coverage of surveillance data on drug resistance, 1994–2016 (Reproduced from World Health Organization 2016)

new and previously treated patients. Among patients with multidrug-resistant TB (MDR-TB), testing results for fluoroquinolones and second-line injectable agents are also collected. In 2015, representative data were available from 83 countries with continuous surveillance systems, most of which were high-income countries. Data were available for a remaining 72 countries from surveys. From this total of 155 countries, data were considered to be nationally representative for 151 and only subnationally representative for the remaining countries. Of the 30 high MDR-TB burden countries defined by the WHO for the period 2016–2020, only 5 countries have continuous national surveillance systems (Belarus, Kazakhstan, Peru, Republic of Moldova and Tajikistan) (World Health Organization 2016). Central and Francophone Africa remain the parts of the world where drug resistance surveillance data are most lacking, largely as a result of weak laboratory infrastructure. Progress towards achieving global coverage of drug resistance surveillance data is shown in Fig. 11.1.

11.3 Burden of Drug-resistant TB

Globally, 580,000 (range: 520,000–640,000) new cases of rifampicin-resistant TB (RR-TB), including MDR-TB are estimated to have occurred in 2015, with 250,000 (range: 160,000–340,000) deaths.

11.3.1 Resistance to First-Line Anti-TB Drugs

An estimated 3.9% (95%CI: 2.7–5.1%) of new cases and 21% (95%CI: 15–28%) of previously treated cases had MDR/RR-TB in 2015. Levels of drug resistance among new cases remain low in many parts of the world, including in almost all countries in the Region of the Americas, most African countries where drug resistance surveys have been conducted, most of the South-East Asia Region, most of western Europe, and several countries in the Western Pacific Region. However, a low proportion of TB cases with drug resistance can still mean a substantial absolute

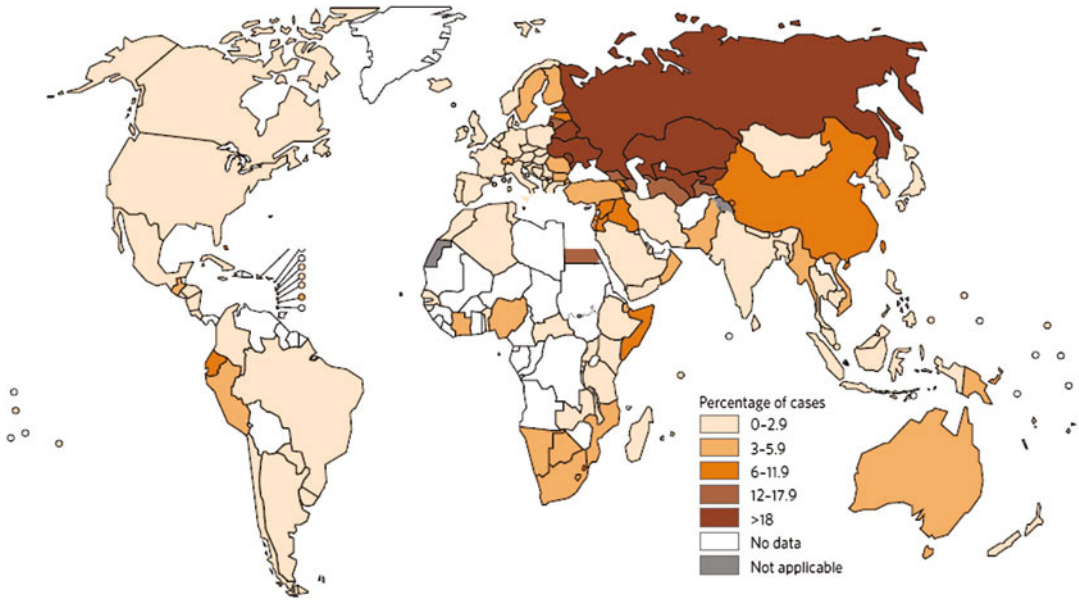

Fig. 11.2 Percentage of new TB cases with MDR-TB in 2015 (Reproduced from World Health Organization 2016)

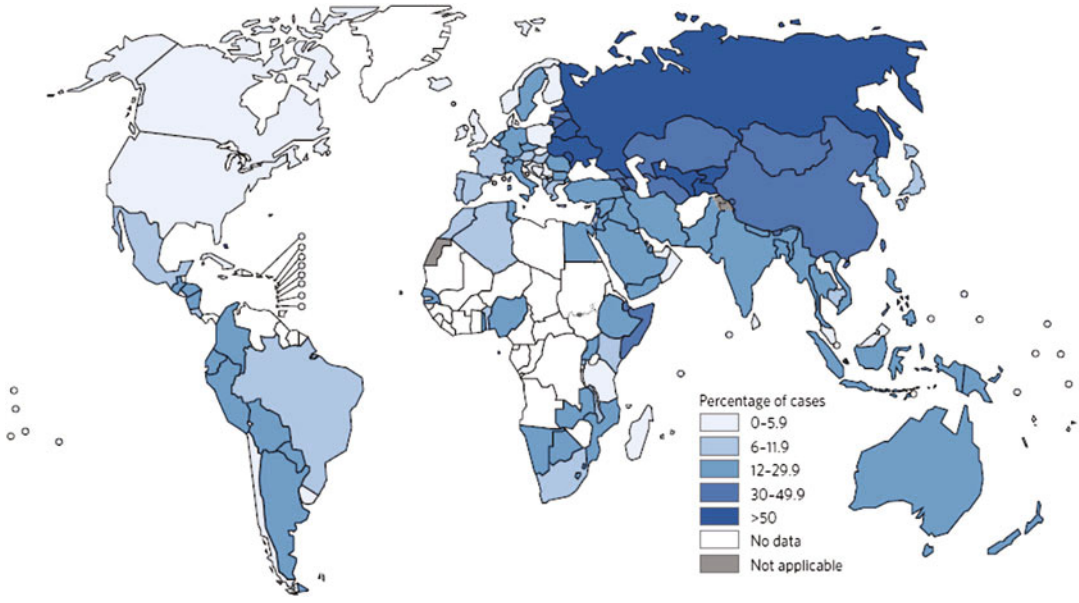

Fig. 11.3 Percentage of previously treated TB cases with MDR-TB in 2015 (Reproduced from World Health Organization 2016)

number of drug-resistant cases in countries with a high TB incidence. The WHO European Region has the highest proportions of MDR/RR-TB among TB cases, with MDR/RR-TB occurring in 16% (95%CI: 11–20%) of new cases and 48% (95%CI: 42–53%) of previously treated cases. The proportions of new and previously treated TB cases with MDR/RR-TB at the country level are shown in Figs. 11.2 and 11.3 (World Health Organization 2016).

Among new cases, the proportions with RR/MDR-TB were highest in Belarus (37% in 2015), Kazakhstan (26.3% in 2013), Kyrgyzstan (32% in 2011), the Republic of Moldova (32.0% in 2015), the Russian Federation (average: 22.0% in 2013), Ukraine (25.0% in 2014) and Uzbekistan (24.0% in 2011). Among previously treated TB cases, the proportions with RR/MDR-TB were highest in Belarus (69.1% in 2015), Estonia (54.0% in 2015), Kyrgyzstan (56% in 2013), the Republic of Moldova (69.0% in 2015), Tajikistan (77.0% in 2014), Ukraine (58.0% in 2014) and Uzbekistan (63.0% in 2011). In the Russian Federation, even though the average proportion of previously treated cases with MDR-TB does not exceed 50%, the proportion is well above 50% in several oblasts (World Health Organization 2016).

While the European Region has the highest proportion of RR/MDR-TB, high numbers of RR/MDR-TB cases are estimated to occur in countries in other regions. The burden of RR/MDR-TB among notified cases is particularly high in countries where populations are large and significant overall TB burdens still exist, such as India and China, as well as in countries with very high TB incidence, such as those in Southern Africa (Table 11.1). Assessment of population level RR/MDR-TB burdens are important for programmatic planning and as they indicate transmission risk in the community.

11.3.2 Resistance to Second-Line Anti-TB Drugs

Extensively drug-resistant TB (XDR-TB), defined as MDR-TB plus resistance to at least one fluoroquinolone and a second-line injectable drug, had been reported by 117 countries by the end of 2015. As representative data on the proportions of MDR-TB cases with second-line drug resistance were only available for 93 countries/territories, settings with high levels of second-line drug resistance may be missed. The proportion of MDR-TB cases with XDR-TB was estimated as 9.5% (95%CI: 7.0–12.1%) in 2015. However, this proportion is much higher in Eastern Europe, at 29% in Belarus, 15% in Georgia, 19% in Latvia and 25% in Lithuania. Globally, the proportion of MDR-TB cases with resistance to any fluoroquinolone for which testing was done, including ofloxacin, levofloxacin and moxifloxacin, was 21.0% (95% CI: 8.8–33.3%) (World Health Organization 2016).

11.3.3 Trends in Anti-TB Drug Resistance

Of the 40 countries with a high TB and/or MDR-TB burden, only 20 have repeated a survey to evaluate trends in drug resistance over time. An analysis of trends during the period 2008–2013 indicated that the proportion of new cases with MDR-TB has remained relatively low and stable over time, ranging from 2 to 4% (World Health Organization 2014a). While serious epidemics do exist in some countries, particularly in Eastern Europe and central Asia, less than 10% of the population living in high MDR-TB burden countries reside in areas with sufficient data to assess trends over time. Countries in the Baltic states, with low overall numbers of cases, have demonstrated a considerable reduction in the burden of MDR-TB over time, while increases have been detected in some oblasts of Russian Federation and in Botswana (the sole African setting for which trends could be assessed) (Cohen et al. 2014).

11.4 Transmission of Drug-Resistant TB

MDR-TB can occur via two mechanisms: through the selection of resistant *M. tuberculosis* bacilli during poorly managed anti-TB treatment or via direct transmission from an infectious patient to another. Given that the risk of drug resistance is higher among TB patients who have previously received TB treatment, there has been a widespread belief that MDR-TB is due primarily to acquisition of resistance during inadequate treatment (Caminero 2008). This

Table 11.1 Diagnosis and treatment gaps across the 30 high MDR-TB burden countries in 2015

Country	RR/MDR-TB cases estimated to occur among notified pulmonary TB patients	Notified RR/MDR-TB cases	% Notified among estimated RR/MDR-TB cases	Cases enrolled on RR/MDR-TB treatment	% Enrolled on treatment among notified RR/MDR-TB cases	% Enrolled on treatment among estimated RR/MDR-TB cases
Angola	2700	227	8.4	227	100.0	8.4
Azerbaijan	1400	1171	83.6	827	70.6	59.1
Bangladesh	5100	954	18.7	880	92.2	17.3
Belarus	1800	1340	74.4	1949	145.4	108.3
China	57000	9662	17.0	5691	58.9	10.0
Democratic Republic of the Congo	4000	499	12.5	413	82.8	10.3
Ethiopia	3300	597	18.1	597	100.0	18.1
Indonesia	10000	2135	21.4	1519	71.1	15.2
India	79000	28876	36.6	26966	93.4	34.1
Kazakhstan	4000	6497	162.4	6948	106.9	173.7
Kenya	1400	368	26.3	368	100.0	26.3
Kyrgyzstan	2400	1116	46.5	1158	103.8	48.3
Republic of Moldova	1700	1042	61.3	1020	97.9	60.0
Myanmar	9000	2793	31.0	2207	79.0	24.5
Mozambique	2800	646	23.1	646	100.0	23.1
Nigeria	4700	1241	26.4	656	52.9	14.0
Pakistan	14000	3059	21.9	2553	83.5	18.2
Peru	2000	1534	76.7	1738	113.3	86.9
Philippines	15000	3788	25.3	4142	109.3	27.6
Papua New Guinea	1100	254	23.1	225	88.6	20.5
Democratic People's Republic of Korea	4600	209	4.5	125	59.8	2.7
Russian Federation	42000	17132	40.8	26756	156.2	63.7
Somalia	1200	75	6.3	74	98.7	6.2
Thailand	2500	466	18.6	506	108.6	20.2
Tajikistan	1300	675	51.9	636	94.2	48.9
Ukraine	12000	9397	78.3	9787	104.2	81.6
Uzbekistan	5800	2149	37.1	2149	100.0	37.1
Viet Nam	5200	2602	50.0	2131	81.9	41.0
South Africa	10000	19613	196.1	12527	63.9	125.3
Zimbabwe	1100	468	42.5	433	92.5	39.4

Adapted from World Health Organization (2016)

view was supported by early observations of lower fitness among *M. tuberculosis* strains with resistance-conferring mutations (Billington et al. 1999; Gagneux et al. 2006). However, growing evidence for the significant burden of MDR-TB among new cases, along with genotyping data from studies across different settings, suggests that direct transmission is a large contributor to the epidemic, particularly in high burden settings (Almeida et al. 2005; Cox et al. 2010; Nodieva et al. 2010; Zhao et al. 2012).

The relative contribution of transmission versus acquisition is important for guiding effective control strategies in different settings. Early modelling studies suggested that effective treatment of drug-susceptible TB would prevent the emergence of resistance (Dye and Williams 2000). Subsequent studies suggested that failure to diagnose and treat MDR-TB would result in "hot spots" of disease (Blower and Chou 2004), with MDR-TB eventually even outcompeting drug-susceptible TB in some settings (Cohen and Murray 2004). While estimates of the prevalence of MDR-TB among notified TB cases are often used for programmatic planning and evaluation within countries, they do not provide information on the relative importance of direct transmission.

A more recent study aimed to estimate the relative importance of transmission versus acquisition using a dynamic transmission model incorporating data from six countries differing by geographical location, MDR-TB burden and ratio of MDR-TB among new compared to previously treated TB patients. It was estimated that 96% of all incident MDR-TB was due to transmission. Importantly, it was also estimated that 61% of incident MDR-TB among previously treated cases was due to direct transmission. The six settings differed in overall TB incidence, proportions of MDR-TB among TB cases, and the ratio of MDR-TB among new and previously treated TB cases. The lowest transmission estimate of 48% was found for Bangladesh, which has a high TB incidence and a much greater proportion of MDR-TB among previously treated cases than among new cases. The estimate of 99% transmission in Uzbekistan reflects a scenario of relatively low TB incidence and similarly high MDR-TB among both previously treated and new TB cases. Clearly, factors such as the temporal emergence of MDR-TB and the extent of diagnosis and appropriate MDR-TB treatment are likely to impact on the contribution of direct transmission to the current burden (Kendall et al. 2015).

Direct transmission has also been identified as a significant contributor to the burden of XDR-TB in several settings (Perdigao et al. 2010; Murase et al. 2010; Gandhi et al. 2013; Klopper et al. 2013), although acquisition of second-line resistance due to inadequate second-line treatment likely remains significant (Mlambo et al. 2008). High rates of transmission of both MDR-TB and XDR-TB, particularly in high MDR-TB burden settings, emphasises the need for improved case detection and treatment for MDR-TB and XDR-TB (Nardell and Dharmadhikari 2010).

11.5 Factors Associated with Drug-Resistant TB

The most consistent factor associated with a diagnosis of MDR-TB in a patient is a previous history of anti-TB treatment, over a period of 1 month or more. While this association reflects the potential for acquisition of drug resistance during prior inadequate treatment, it may also represent confounding with other shared risk factors. For example, patients who have previously received TB treatment are likely to have had significantly greater contact with health services, with a greater subsequent risk of nosocomial transmission of MDR-TB. Indeed, previous hospitalization has been found to be a significant risk factor for MDR-TB and XDR-TB in several studies (Gandhi et al. 2006; Gelmanova et al. 2007; Zhao et al. 2012; Rifat et al. 2015). An increased risk of nosocomial transmission of MDR-TB likely arises from delayed or non-existent MDR-TB diagnosis and subsequent inadequate treatment with first-line anti-TB therapy. This results in failure of treatment among source cases who then transmit the organism when seeking care in a health care facility setting. Given inadequate infection control, the risk of nosocomial trans-

mission to other patients as well as health care workers is high (Shenoi et al. 2010).

HIV infection may also contribute to the risk of MDR-TB through a range of potential mechanisms, including greater risk of nosocomial transmission (Wells et al. 2007). Institutional MDR-TB outbreaks in well-resourced settings have primarily involved HIV-infected individuals (Edlin et al. 1992; Coronado et al. 1993; Frieden et al. 1996). Several studies have shown a positive association between HIV infection and MDR-TB in more generalised settings (Jenkins et al. 2013). However, the association between MDR-TB and HIV infection at the population level varies between countries. A systematic review of published studies found a moderate association with HIV, particularly for primary, transmitted MDR-TB (Mesfin et al. 2014). In contrast, in countries in sub-Saharan Africa where HIV prevalence is high, lower proportions of MDR-TB were found among countries with the highest reported HIV prevalences (Lukoye et al. 2015). This could be due to higher early mortality among HIV-infected MDR-TB patients compared to HIV-infected TB patients. Dual epidemics of HIV and MDR-TB have been shown to converge in socially vulnerable groups in some settings, such as Kazakhstan (van den Hof et al. 2015). Globally, there is no evidence of a clear association between the levels of HIV infection and MDR-TB in a population, although reliable national-level surveillance data are predominantly from high- and middle-income settings (Dean et al. 2014).

While national level notifications and estimates are often used for reporting and programmatic planning, there is consistent evidence of heterogeneity in MDR-TB burden within countries (Paramasivan and Venkataraman 2004; Ulmasova et al. 2013; Jenkins et al. 2014a; Aia et al. 2016; Heidebrecht et al. 2016; Zelner et al. 2016). In Moldova, analysis has shown that geographical areas with higher proportions of previously incarcerated individuals also have higher burdens of MDR-TB (Jenkins et al. 2013). In Peru, variation in per capita MDR-TB rates was associated with certain health centre catchment areas (Zelner et al. 2016). In other settings, variable access to MDR-TB diagnosis and care, lower socio-economic status and greater reliance on inadequate MDR-TB treatment in the private sector are likely contributors to variations in MDR-TB prevalence.

As suggested by the study from Moldova, prisons have been shown to harbour particularly high burdens of MDR-TB (Kimerling et al. 1999). Annual TB notifications have reached as high as 7% of the prison population in some settings (Coninx et al. 2000), with high proportions of MDR-TB (Kimerling et al. 1999; Habeenzu et al. 2007). Indeed, transmission of TB and MDR-TB in the community has been shown to be driven by high burdens of TB and MDR-TB in prisons, along with a high level of incarceration at the population level (Stuckler et al. 2008). Overcrowding, malnutrition and poor access to diagnosis and treatment all contribute to high TB transmission rates within prisons.

Around 30,000 children are estimated to develop MDR-TB every year. As children are more likely than adults to have paucibacillary disease and often cannot expectorate sputum, drug susceptibility testing is particularly challenging (Jenkins et al. 2014b). However, a review of available data from multiple countries suggested that children with TB do not have a lower risk of MDR-TB than adult patients, with similar proportions of MDR-TB detected in both groups (Zignol et al. 2013).

11.6 The Diagnosis and Treatment Cascade for Drug-Resistant TB

Given the significant contribution of direct transmission to the MDR-TB epidemic in high MDR-TB burden settings, prevention will require better case detection, rapid treatment initiation of appropriate treatment and improved treatment outcomes. In 2014, only 123,000 of the estimated 480,000 incident MDR-TB cases were notified to WHO, and among these only 111,000 were reported to have been initiated on second-line treatment. Globally, treatment success for MDR-TB is reported to be 50%. Diagnosis and treatment gaps vary substantially across the 30

high MDR-TB burden countries defined for the period 2016–2020. The numbers of MDR-TB cases estimated to have occurred among notified TB cases in 2014 are given in Table 11.1. Among these estimated MDR-TB cases, the proportions that were actually notified range from 5% in DPR Korea to over 100% in Kazakhstan, South Africa, Tajikistan. Notably, China, Myanmar and Somalia report large treatment gaps with less than 50% of notified cases enrolled on treatment (World Health Organization 2016). While, there are inherent uncertainties in these estimates (Cohen et al. 2008, 2010), this diagnosis and treatment cascade results in cure of only 12% of the estimated burden of incident cases, or even lower if one considers the MDR-TB burden in terms of prevalence (Borgdorff 2004).

Globally, the largest gap in the cascade is diagnosis. Poor access to drug susceptibility testing, even among TB cases with previous TB treatment, is common. In 2014, only 12% of new, bacteriologically confirmed TB cases were tested for drug susceptibility (World Health Organization 2016). However, given that a large proportion of TB cases are not bacteriologically confirmed in many settings, a much higher proportion of all TB cases do not receive drug susceptibility testing. Many high TB burden countries lack sputum transport networks and equipped laboratories to safely and reliably perform culture-based drug susceptibility testing. Due to a lack of capacity, drug susceptibility testing is usually only performed for prioritised patient groups, such as those who have failed treatment. From a clinical perspective, the turn-around time for receiving testing results from culture-based methods can be weeks to months, delaying the initiation of appropriate therapy.

Lack of drug susceptibility testing capacity is particularly evident for second-line drug resistance. Only 24% of notified MDR-TB cases were tested in 2014. Given the relatively high proportions of XDR-TB among MDR-TB cases found in surveys, the 4044 XDR-TB cases enrolled on treatment in 2014 suggests a large undiagnosed burden. In many settings, these undiagnosed cases are likely to receive inappropriate treatment, suffer high mortality and contribute to transmission. Overall, among the 2685 XDR-TB cases treated in 2012, only 26% were reported to be successfully treated (World Health Organization 2016).

The ongoing expansion of rapid molecular technologies have revolutionised the diagnosis of drug-resistant TB. For example, South Africa now notifies substantially more rifampicin-resistant (RR) TB cases than the number of estimated incident cases following scale up of the Xpert MTB/RIF® assay for all presumptive TB cases (see Chap. 12), although this may be partially due to duplication in notified cases. Indeed in South Africa, while the diagnostic gap is likely to be relatively small, the gap between the number of notified cases and those reported to have started a second-line treatment regimen is large; in 2014, only 62% of the notified, bacteriologically-confirmed RR-TB cases were reported to have started treatment (Schnippel et al. 2013; World Health Organization 2016). High early mortality among HIV infected RR-TB cases may contribute to this gap, although a significant shortfall in health system capacity to follow up and treat all diagnosed cases is also a likely contributor. Globally, gaps between the numbers of cases diagnosed with resistance and those beginning second-line therapy represent major public health and ethical problems (World Health Organization 2016).

11.7 Future Direction

The establishment of universal drug susceptibility testing via continuous surveillance systems for at least rifampicin resistance should be a priority for the following reasons: (i) rifampicin is the most powerful first-line anti-TB drug, (ii) reliable and easily implemented rapid technologies exist for accurate diagnosis. In addition, rifampicin resistance can be a marker of resistance to additional drugs. However, it is important to note that levels of isoniazid resistance without concurrent rifampicin resistance are not insignificant, at 8.1% (95%CI: 6.5–9.7%) of new cases and 14.0% (95%CI: 11.6–16.3%) of previously treated cases. Rifampicin-susceptible

patients with poor treatment outcomes should therefore always be investigated for isoniazid resistance (World Health Organization 2014b). Additionally, all patients detected with rifampicin resistance should be tested to exclude further resistance to fluoroquinolones and/or second-line injectable agents. Resistance to these classes of drugs seriously compromises the treatment options that remain available.

Until capacity can be established to test every TB patient for rifampicin resistance, countries could focus on gradually expanding their coverage of testing with rapid molecular technologies, such as Xpert MTB/RIF (World Health Organization 2014c). Xpert MTB/RIF provides a result within 90 min and is easy to use at lower level health facilities. Ad hoc surveys of TB patients will continue to be important for generating nationally representative estimates of the proportion of new and previously treated patients with drug-resistant TB, which are essential for resource planning and allocation. Surveys can also allow the investigation of other drug resistance patterns beyond rifampicin. This includes drugs from existing first-line and second-line regimens as well as new or repurposed drugs. For the first time in over 40 years, two new anti-TB drugs are available on the market, bedaquiline and delamanid (World Health Organization 2013, 2014d). These drugs will be key to improving treatment outcomes and quality of life among MDR-TB patients, and close monitoring of their continued effectiveness is essential.

DNA sequencing technologies have primarily been used for research purposes, but are being increasingly incorporated into patient management and drug resistance surveillance (see Chap. 12). They offer additional benefits over conventional culture-based phenotypic testing, including a faster turn-around time for results, assessment of resistance patterns to a range of drugs, and investigation of strain clustering and transmission (see Chap. 4) (Roetzer et al. 2013; Guerra-Assunção et al. 2015). However, whole genome sequencing still requires specialised bioinformatics skills, limiting its broader application in global TB control. Targeted gene sequencing performed directly on sputum delivers rapid results for a range of drugs and bypasses the need for culture (Rigouts et al. 2007). This may be a more practical approach to expanding the coverage of testing to other drugs.

References

Aia P, Kal M, Lavu E et al (2016) The burden of drug-resistant tuberculosis in Papua New Guinea: results of a large population-based survey. PLoS One 11:e0149806. doi:10.1371/journal.pone.0149806

Almeida D, Rodrigues C, Ashavaid TF et al (2005) High incidence of the Beijing genotype among multidrug-resistant isolates of Mycobacterium tuberculosis in a tertiary care center in Mumbai, India. Clin Infect Dis 40:881–886

Billington OJ, McHugh TD, Gillespie SH (1999) Physiological cost of rifampin resistance induced in vitro in Mycobacterium tuberculosis. Antimicrob Agents Chemother 43:1866–1869

Blower SM, Chou T (2004) Modeling the emergence of the "hot zones": tuberculosis and the amplification dynamics of drug resistance. Nat Med 10:1111–1116

Borgdorff MW (2004) New measurable indicator for tuberculosis case detection. Emerg Infect Dis 10:1523–1528. doi:10.3201/eid1009.040349

Caminero JA (2008) Likelihood of generating MDR-TB and XDR-TB under adequate National Tuberculosis Control Programme implementation. Int J Tuberc Lung Dis 12:869–877

Centers for Disease Control (CDC) (1991) Nosocomial transmission of multidrug-resistant tuberculosis among HIV-infected persons – Florida and New York, 1988–1991. MMWR Morb Mortal Wkly Rep 40:585–591

Cohen T, Murray M (2004) Modeling epidemics of multidrug-resistant M. tuberculosis of heterogeneous fitness. Nat Med 10:1117–1121

Cohen T, Colijn C, Finklea B et al (2008) Are survey-based estimates of the burden of drug resistant TB too low? Insight from a simulation study. PLoS One 3:e2363. doi:10.1371/journal.pone.0002363

Cohen T, Hedt BL, Pagano M (2010) Estimating the magnitude and direction of bias in tuberculosis drug resistance surveys conducted only in the public sector: a simulation study. BMC Public Health 10:355. doi:10.1186/1471-2458-10-355

Cohen T, Jenkins HE, Lu C et al (2014) On the spread and control of MDR-TB epidemics: an examination of trends in anti-tuberculosis drug resistance surveillance data. Drug Resist Updat 17:105–123. doi:10.1016/j.drup.2014.10.001

Coninx R, Maher D, Reyes H, Grzemska M (2000) Tuberculosis in prisons in countries with high prevalence. BMJ 320:440–442

Coronado VG, Beck-Sague CM, Hutton MD et al (1993) Transmission of multidrug-resistant Mycobacterium tuberculosis among persons with human immunodefi-

ciency virus infection in an urban hospital: epidemiologic and restriction fragment length polymorphism analysis. J Infect Dis 168:1052–1055

Cox HS, McDermid C, Azevedo V et al (2010) Epidemic levels of drug resistant tuberculosis (MDR and XDR-TB) in a high HIV prevalence setting in Khayelitsha, South Africa. PLoS One 5:e13901. doi:10.1371/journal.pone.0013901

Crofton J, Mitchison DA (1948) Streptomycin resistance in pulmonary tuberculosis. BMJ 2:1009–1015

Dean AS, Zignol M, Falzon D et al (2014) HIV and multidrug-resistant tuberculosis: overlapping epidemics. Eur Respir J 44:251–254. doi:10.1183/09031936.00205413

Dye C, Williams BG (2000) Criteria for the control of drug-resistant tuberculosis. Proc Natl Acad Sci U S A 97:8180–8185

Edlin BR, Tokars JI, Grieco MH et al (1992) An outbreak of multidrug-resistant tuberculosis among hospitalized patients with the acquired immunodeficiency syndrome. N Engl J Med 326:1514–1521. doi:10.1056/NEJM199206043262302

Frieden TR, Sterling T, Pablos-Mendez A et al (1993) The emergence of drug-resistant tuberculosis in New York City. N Engl J Med 328:521–526. doi:10.1056/NEJM199302253280801

Frieden TR, Sherman LF, Maw KL et al (1996) A multi-institutional outbreak of highly drug-resistant tuberculosis: epidemiology and clinical outcomes. JAMA 276:1229–1235

Gagneux S, Burgos MV, DeRiemer K et al (2006) Impact of bacterial genetics on the transmission of isoniazid-resistant Mycobacterium tuberculosis. PLoS Pathog 2:e61. doi:10.1371/journal.ppat.0020061

Gandhi NR, Moll A, Sturm AW et al (2006) Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. Lancet 368:1575–1580

Gandhi NR, Weissman D, Moodley P et al (2013) Nosocomial transmission of extensively drug-resistant tuberculosis in a rural hospital in South Africa. J Infect Dis 207:9–17. doi:10.1093/infdis/jis631

Gelmanova IY, Keshavjee S, Golubchikova VT et al (2007) Barriers to successful tuberculosis treatment in Tomsk, Russian Federation: non-adherence, default and the acquisition of multidrug resistance. Bull World Health Organ 85:703–711

Guerra-Assunção JA, Crampin AC, RMGJ H et al (2015) Large-scale whole genome sequencing of M. tuberculosis provides insights into transmission in a high prevalence area. Elife. doi:10.7554/eLife.05166

Habeenzu C, Mitarai S, Lubasi D et al (2007) Tuberculosis and multidrug resistance in Zambian prisons, 2000–2001. Int J Tuberc Lung Dis 11:1216–1220

Heidebrecht CL, Podewils LJ, Pym A et al (2016) Assessing loval risk of rifampicin-resistant tuberculosis in KwaZulu-Natal, South Africa using lot quality assurance sampling. PLoS One 11:e0153143

Jenkins HE, Plesca V, Ciobanu A et al (2013) Assessing spatial heterogeneity of multidrug-resistant tuberculosis in a high-burden country. Eur Respir J 42:1291–1301. doi:10.1183/09031936.00111812

Jenkins HE, Gegia M, Furin J et al (2014a) Geographical heterogeneity of multidrug-resistant tuberculosis in Georgia, January 2009 to June 2011. Euro Surveill 19:20743

Jenkins HE, Tolman AW, Yuen CM et al (2014b) Incidence of multidrug-resistant tuberculosis disease in children: systematic review and global estimates. Lancet 383:1572–1579. doi:10.1016/S0140-6736(14)60195-1

Kendall EA, Fofana MO, Dowdy DW (2015) Burden of transmitted multidrug resistance in epidemics of tuberculosis: a transmission modelling analysis. Lancet Respir Med 3:963–972. doi:10.1016/S2213-2600(15)00458-0

Kimerling ME, Kluge H, Vezhnina N et al (1999) Inadequacy of the current WHO re-treatment regimen in a central Siberian prison: treatment failure and MDR-TB. Int J Tuberc Lung Dis 3:451–453

Klopper M, Warren RM, Hayes C et al (2013) Emergence and spread of extensively and totally drug-resistant tuberculosis, South Africa. Emerg Infect Dis 19:449–455. doi:10.3201//EID1903.120246

Lukoye D, Ssengooba W, Musisi K et al (2015) Variation and risk factors of drug resistant tuberculosis in Sub-Saharan Africa: a systematic review and meta-analysis. BMC Public Health 15:291. doi:10.1186/s12889-015-1614-8

Mesfin YM, Hailemariam D, Biadglign S, Kibret KT (2014) Association between HIV/AIDS and multidrug resistance tuberculosis: a systematic review and meta-analysis. PLoS One 9:e82235. doi:10.1371/journal.pone.0082235

Mlambo CK, Warren RM, Poswa X et al (2008) Genotypic diversity of extensively drug-resistant tuberculosis (XDR-TB) in South Africa. Int J Tuberc Lung Dis 12:99–104

Monno L, Angarano G, Carbonara S et al (1991) Emergence of drug-resistant Mycobacterium tuberculosis in HIV-infected patients. Lancet 337:852

Murase Y, Maeda S, Yamada H et al (2010) Clonal expansion of multidrug-resistant and extensively drug-resistant tuberculosis, Japan. Emerg Infect Dis 16:948–954. doi:10.3201/eid1606.091844

Nardell E, Dharmadhikari A (2010) Turning off the spigot: reducing drug-resistant tuberculosis transmission in resource-limited settings. Int J Tuberc Lung Dis 14:1233–1243

Nodieva A, Jansone I, Broka L et al (2010) Recent nosocomial transmission and genotypes of multidrug-resistant Mycobacterium tuberculosis. Int J Tuberc Lung Dis 14:427–433

Pablos-Méndez A, Raviglione MC, Laszlo A et al (1998) Global surveillance for antituberculosis-drug resistance, 1994–1997. World Health Organization-International Union against Tuberculosis and Lung Disease Working Group on Anti-Tuberculosis Drug Resistance Surveillance. N Engl J Med 338:1641–1649. doi:10.1056/NEJM199806043382301

Paramasivan CN, Venkataraman P (2004) Drug resistance in tuberculosis in India. Indian J Med Res 120:377–386

Perdigao J, Macedo R, Malaquias A et al (2010) Genetic analysis of extensively drug-resistant Mycobacterium tuberculosis strains in Lisbon, Portugal. J Antimicrob Chemother 65:224–227. doi:10.1093/jac/dkp452

Rifat M, Hall J, Oldmeadow C et al (2015) Factors related to previous tuberculosis treatment of patients with multidrug-resistant tuberculosis in Bangladesh. BMJ Open 5:e008273. doi:10.1136/bmjopen-2015-008273

Rigouts L, Nolasco O, de Rijk P et al (2007) Newly developed primers for comprehensive amplification of the rpoB gene and detection of rifampin resistance in Mycobacterium tuberculosis. J Clin Microbiol 45:252–254. doi:10.1128/JCM.01489-06

Roetzer A, Diel R, Kohl TA et al (2013) Whole genome sequencing versus traditional genotyping for investigation of a Mycobacterium tuberculosis outbreak: a longitudinal molecular epidemiological study. PLoS Med 10:e1001387. doi:10.1371/journal.pmed.1001387

Schnippel K, Meyer-Rath G, Long L et al (2013) Diagnosing Xpert MTB/RIF negative TB: impact and cost of alternative algorithms for South Africa. S Afr Med J 103:101–106. doi:10.7196/samj.6182

Shenoi SV, Escombe AR, Friedland G (2010) Transmission of drug-susceptible and drug-resistant tuberculosis and the critical importance of airborne infection control in the era of HIV infection and highly active antiretroviral therapy rollouts. Clin Infect Dis 50(Suppl 3):S231–S237. doi:10.1086/651496

Stuckler D, Basu S, Mckee M, King L (2008) Mass incarceration can explain population increases in TB and multidrug-resistant TB in European and central Asian countries. P Natl Acad Sci USA 105:13280–13285. doi:10.1073/pnas.0801200105

Ulmasova D, Uzakova G, Tillyashayhov M et al (2013) Multidrug-resistant tuberculosis in Uzbekistan: results of a nationwide survey, 2010 to 2011. Euro Surveill. doi:20609 [pii]

Uplekar M, Weil D, Lonnroth K et al (2015) WHO's new end TB strategy. Lancet 385:1799–1801. doi:10.1016/S0140-6736(15)60570-0

van den Hof S, Tursynbayeva A, Abildaev T et al (2015) HIV and multidrug-resistant tuberculosis: overlapping risk factors. Eur Respir J 45:567–569. doi:10.1183/09031936.00131014

Wells CD, Cegielski JP, Nelson LJ et al (2007) HIV infection and multidrug-resistant tuberculosis: the perfect storm. J Infect Dis 196(Suppl 1):S86–107. doi:10.1086/518665

World Health Organization (2013) The use of bedaquiline in the treatment of multidrug-resistant tuberculosis. Interim policy guidance. WHO/HTM/TB/2013.6. Geneva

World Health Organization (2014a) Drug-resistant TB – surveillance and response. Supplement to global tuberculosis report 2014. WHO/HQ/TB/2014.12. Geneva

World Health Organization (2014b) Global tuberculosis report 2014. WHO/HTM/TB/2014.08. Geneva

World Health Organization (2014c) Xpert MTB/RIF implementation manual. Technical and operational "how-to": practical considerations

World Health Organization (2014d) The use of delamanid in the treatment of multidrug-resistant tuberculosis. WHO/HTM/TB2014.23. Geneva

World Health Organization (2016) Global tuberculosis report 2015. Geneva

World Health Organization WHO End TB Strategy. http://www.who.int/tb/post2015_strategy/en/

Zelner JL, Murray MB, Becerra MC et al (2016) Identifying hotspots of multidrug-resistant tuberculosis transmission using spatial and molecular genetic data. J Infect Dis 213:287–294. doi:10.1093/infdis/jiv387

Zhao Y, Xu S, Wang L et al (2012) National survey of drug-resistant tuberculosis in China. N Engl J Med 366:2161–2170. doi:10.1056/NEJMoa1108789

Zignol M, Sismanidis C, Falzon D et al (2013) Multidrug-resistant tuberculosis in children: evidence from global surveillance. Eur Respir J 42:701–707. doi:10.1183/09031936.00175812

Zignol M, Dean A, Falzon D, et al (2016) Twenty years of global surveillance of anti-tuberculosis drug resistance: key achievements and findings, remaining challenges, future direction

Evolution of Phenotypic and Molecular Drug Susceptibility Testing

12

Daniela M. Cirillo, Paolo Miotto, and Enrico Tortoli

Abstract

Drug Resistant Tuberculosis (DRTB) is an emerging problem world-wide. In order to control the disease and decrease the number of cases overtime a prompt diagnosis followed by an appropriate treatment should be provided to patients. Phenotypic DST based on liquid automated culture has greatly reduced the time needed to generate reliable data but has the drawback to be expensive and prone to contamination in the absence of appropriate infrastructures. In the past 10 years molecular biology tools have been developed. Those tools target the main mutations responsible for DRTB and are now globally accessible in term of cost and infrastructures needed for the implementation. The dissemination of the Xpert MTB/rif has radically increased the capacity to perform the detection of rifampicin resistant TB cases. One of the main challenges for the large scale implementation of molecular based tests is the emergence of conflicting results between phenotypic and genotypic tests. This mines the confidence of clinicians in the molecular tests and delays the initiation of an appropriate treatment. A new technique is revolutionizing the genotypic approach to DST: the WGS by Next-Generation Sequencing technologies. This methodology promises to become the solution for a rapid access to universal DST, able indeed to overcome the limitations of the current phenotypic and genotypic assays. Today the use of the generated information is still challenging in decentralized facilities due to the lack of automation for sample processing and standardization in the analysis.

D.M. Cirillo (✉) • P. Miotto • E. Tortoli
IRCCS San Raffaele Scientific Institute, Milan, Italy
e-mail: cirillo.daniela@hsr.it; miotto.paolo@hsr.it; tortoli.enrico@hsr.it

S. Gagneux (ed.), *Strain Variation in the* Mycobacterium tuberculosis *Complex: Its Role in Biology, Epidemiology and Control*, Advances in Experimental Medicine and Biology 1019,
DOI 10.1007/978-3-319-64371-7_12

The growing knowledge of the molecular mechanisms at the basis of drug resistance and the introduction of high-performing user-friendly tools at peripheral level should allow the very much needed accurate diagnosis of DRTB in the near future.

Keywords
Molecular drug susceptibility test • Phenotypic drug susceptibility test

12.1 Introduction

Proper managing of tuberculosis requires starting an effective antitubercular therapy as soon as possible to prevent spreading of the disease and to increase the cure rate of the affected individuals (Uys et al. 2009; Dowdy et al. 2008). Appropriate therapy can only be provided if the drug susceptibility pattern of the infecting strain is known. Drug susceptibility testing (DST) for *M. tuberculosis* can be performed by conventional phenotypic methods or by molecular detection of genetic determinants associated with drug resistance. The two approaches have advantages and disadvantages and in the most difficult cases a combination of the two may be required. Conventional DST based on mycobacterial growth on both solid and liquid media is time-consuming, and challenged by technical difficulties and biosafety issues (Kim 2005; WHO 2012a, b; Jiang et al. 2013; Somoskovi and Salfinger 2015). The development of molecular technologies has led to the emergence of rapid diagnostic assays suitable for the detection of drug-resistant tuberculosis. Despite these advancements in technology and the large amount of data that are going to be collected by Whole Genome Sequencing (WGS) of drug resistant and drug sensitive *M. tuberculosis* strains, we cannot abandon completely phenotypic DST at this time. Achieving a more comprehensive understanding of the genotype-phenotype-clinical outcome associations could lead to a future when molecular DST will become the routine and phenotypic will be restricted as a referral test for few cases.

We can predict that deeper knowledge will be available in the near future allowing designing a full molecular DST for routine testing. Phenotypic tests and Minimal Inhibitory Concentration (MIC) will be reserved for the most challenging cases.

12.2 Phenotypic DST

Streptomycin, the first antituberculosis drug was first experimented in 1944 (Jones et al. 1944; Emmart 1945; Smith and Waskman 1947). Shortly after its introduction in clinical practice, the first cases of resistance were reported (Youmans et al. 1946). Not substantially different was the history of all other antimycobacterial agents; resistance was rapidly emerging in particular when drugs were used inappropriately or in monotherapy (Guernsey and Alexander 1978; Smith et al. 2013). It became clear that a combination of several effective drugs was essential to achieve cured in patients with tuberculosis. The need to develop a laboratory test able to predict antibacterial sensitivity to a specific drug soon emerged and in the 1960s the pioneering experiments of Canetti, at the Pasteur Institute of Paris, lead to the development of the "Proportion method" (Canetti et al. 1963). It is based on the empiric observation that when the proportion of resistant mutants within the *M. tuberculosis* population infecting the patient is approximately $\geq 1\%$, the probability of treatment failure is very high. The proportion method uses a set of media, each containing the "critical" concentration of a single drug, to test the growth of the strain in comparison with that obtained on a drug-free medium (the growth control). The susceptibility to single antimicrobials is inferred by determining the percentage between the counts of colonies grown on

the medium with the drug and on the control; whenever this proportion is ≥1% the strain is classified as resistant. Several years earlier the same researcher had developed the "Absolute Concentration" method (Canetti et al. 1969), a kind of MIC determination with multiple drug concentrations; in this case each laboratory was requested to define its critical concentration. The Resistance Ratio method relies, for each determination, on a parallel testing of the susceptible reference strain *M. tuberculosis* H37Rv (Kent and Kubica 1985). The results are interpreted on the basis of the ratio between the MICs of reference and test strains. The proportion method, thanks to its easy implementation and interpretation, rapidly prevailed and its principle is still at the basis of modern phenotypic susceptibility testing for *M. tuberculosis.* A feature shared by all methods on solid media is the long incubation time which, added to the time for culture, make results available for clinical use in not less than 2–3 months.

In the last 40 years the liquid media, suitable to shorten the incubation time, have progressively replaced the classical solid media for the culture of *M. tuberculosis*.

At present the commercial Mycobacteria Growth Indicator Tube (MGIT) 960 system (Becton Dickinson) has monopolized the market worldwide. It is an automated system that infers the bacterial growth rate from the oxygen consumption. When used for the determination of antimicrobial susceptibility a set of tubes of liquid medium, added with critical concentrations of drugs, is inoculated with a standardized suspension of the strain to be tested; while the control tube is inoculated with the standardized suspension above, diluted 1/100. The tubes, with and without drug, are monitored in parallel and the software, upon quality validation, reports susceptibility or resistance on the basis of the comparison of respective growth curves (Rüsch-Gerdes et al. 1999; Tortoli et al. 2002).

Although the MGIT is widespread, it is not the only method available. The classical proportion methods on egg-based solid media, along with its variant on agar media (Middlebrook 7H10 or Middlebrook 7H11), are still considered the reference for several antibiotics.

Micro-dilution methods in liquid medium have also been developed which, combining inexpensiveness and ease reading, are especially suited for low-income countries. In the Microscopic Observation Drugs Susceptibility (MODS) assay the drug resistance is detected by the low magnification observation of the growth in liquid medium dispensed in microwells and added with drugs. The reading is made easy by the characteristic corded morphology of *M. tuberculosis* colonies (Moore et al. 2004). In another microtiter assay, the addition of a redox indicator is used to detect the bacterial growth (TeMA, MABA, ReMA, according the indicator used: tetrazolium, alamar blue, resazurin). The color change due to indicator reduction is consistent with bacterial growth and indicates resistance to the antimicrobial present in the micro-well (Collins and Franzblau 1997).

Differently from MODS and other microtitre assays, which are home-made, a microdilution method based on microtitre plates containing twofold concentrations of freeze-dried drugs has been recently commercialized. This method, still under validation, combines a number of potential benefits: the inclusion in a single test of both first- and second-line drugs, the possibility of MIC determination, it is user-friendly and relatively inexpensive (Hall et al. 2012).

First-line drugs (isoniazid, rifampicin, pyrazinamide and rifampicin) are normally tested with the phenotypic approach; in case of simultaneous isoniazid- and rifampicin-resistance the test must be widened to second line molecules, at least to fluoroquinolones and injectables (amikacin, kanamycin and capreomycin).

In general, the phenotypic susceptibility testing produces reliable results, in particular for the two major antitubercular drugs, rifampicin and isoniazid. For ethambutol, a bacteriostatic drug, the results are less reliable (Madison et al. 2002) and DST for this drug is not considered a priority by the World Health Organization (WHO). Pyrazinamide testing has been reported as highly challenging by several laboratories, and results may not be fully reliable (Piersimoni et al. 2013).

Pyrazinamide is a prodrug and needs a low pH for activation, a condition that is difficult to control in an "in vitro" test (Table 12.1).

Critical Concentrations (CC) for the major antitubercular drugs were proposed by WHO in 2008, revised in 2012 and under revision in 2017. Table 12.2 shows the critical concentrations endorsed by WHO in 2012. Drugs used for treatment of rifampicin resistant tuberculosis and MDR tuberculosis are listed according to the new classification published in 2016 in the last WHO manual for drug resistant tuberculosis (http://www.who.int/tb/areas-of-work/drug-resistant tb/MDRTBguidelines2016.pdf). It must be noted that some of the CCs will be revised very soon as reported in the table legend.

CCs should be established at the epidemiological cut-off value (ECOFF) or one dilution higher. For the majority of the drugs, the ECOFF separates wild-type strains expected to be sensitive from those expected to be resistant to a selected drug. If drugs can be used at higher doses without high risk of toxicity, concentration higher than the CC can be tested to predict sensitivity to treatment when high doses of the drug can be used to achieve higher plasmatic concentration. In this case a "clinical breakpoint" can be established; for example, a clinical breakpoint has been recently established for moxifloxacin. For some drugs, CCs cannot be established due to the lack of data. For drugs such cycloserin, imipenem, amoxi/clavulanate *in vitro* testing is still not recommended due to the absence of reliable protocols.

As a general rule, it is advisable to test the drug in use for treatment and to perform susceptibility tests under quality assurance conditions and strictly adhering to the recommended protocols.

Recently, two new drugs, delamanid and bedaquiline, received conditional approval for treatment of MDR-TB cases. Interim CCs were recently discussed and will be reported officially by WHO at the end of 2017. Protocols for susceptibility testing on liquid and solid media have been published (Schena et al. 2016; Torrea et al. 2015).

Interpretation of discrepant results obtained in different high level laboratories, from phenotypic tests performed on different media or between genotype and phenotype has underlined the need of MIC determination for a correct management of difficult cases.

Plates for MIC determination are commercially available but have been only evaluated on a small scale so far (Hall et al. 2012). These microtitre plates allow the MIC determination for several first and second line antimycobacterial agents. Yu et al. (2016) have recently reported an agreement between plates and LJ of 99.2% for rifampicin, ofloxacin, amikacin, kanamycin and cycloserin, 98.4% for isoniazid and PAS and lower than 90% for ethambutol. The use of microtitre plates highly reduces the cost of DST compared to other liquid media but poses several questions on the feasibility in terms of implementation in laboratories located in high burden countries for the level of biosafety requested to handle the plates and the risk of cross-contamination. Automation in plate-reading and a "sealed" layout could improve the use of MIC plates in the future.

Although all the methods in liquid medium have drastically reduced the turnaround time, the cost of maintaining an adequate level of bio containment and the high rates of contaminated culture remain a major limit to implement phenotypic DST in TB high burden settings.

12.3 Molecular DST

12.3.1 Molecular Basis of Drug Resistance in *M. tuberculosis*

In *M. tuberculosis*, drug resistance is mainly caused by chromosomal mutations and evidences exclude horizontal transfer of genetic material as a source of resistance (Gillespie 2002; Marttila and Soini 2003). In a limited number of cases, mobile genetic elements (such as insertion sequences, *e.g.* IS*6110* see Chap. 3) can contribute to the insurgence of phenotypic drug resistance (Lemaitre et al. 1999).

Table 12.1 Grouping of antitubercular drug and proposed CC according to WHO 2012

Drug Groups[a]	Drug	DST method available	DST critical concentration (μg/ml)			
			Löwenstein-Jensen[b]	Middlebrook 7H10[b]	Middlebrook 7H11[b]	MGIT960
First-line oral anti-TB agents	Ethambutol[c]	Solid, liquid	2.0	5	7.5	5.0
	Isoniazid	Solid, liquid	0.2	0.2	0.2	0.1
	Pyrazinamide	Liquid	–	–	–	100.0
	Rifampicin[d]	Solid, liquid	40.0	1	1	1.0
Injectable anti-TB agents (B)	Amikacin[e]	Solid, liquid	30.0	**4.0**	–	1.0
	Capreomycin	Solid, liquid	40.0	4.0	–	2.5
	Kanamycin	Solid, liquid	30.0	4.0	–	2.5
	Streptomycin	Solid, liquid	4.0	2.0	2.0	1.0
Fluoroquinolones (A)	Gatifloxacin[f]	Solid	**–**	**1.0**	**–**	**–**
	Levofloxacin	Solid, liquid	–	1.0	**–**	**1.5**
	Moxifloxacin[g]	Solid, liquid	**–**	**0.5/2.0**	**–**	**0.5/2.0**
	Ofloxacin[h]	Solid, liquid	**4.0**	**2.0**	**2.0**	**1.0**
Additional Core second line agents (C)	Clofazimine[i]	Liquid	**–**	**–**	**–**	**–**
	Cycloserine/Terizidone[j]	Solid	**30.0**	–	–	–
	Ethionamide	Solid, liquid	40.0	5.0	10.0	5.0
	Linezolid[k]	Liquid	–	**–**	**–**	1.0
	Prothionamide	Solid, liquid	40.0	–	–	2.5

(continued)

Table 12.1 (continued)

Drug Groups[a]	Drug		DST method available	DST critical concentration (μg/ml)			
				Löwenstein-Jensen[b]	Middlebrook 7H10[b]	Middlebrook 7H11[b]	MGIT960
Add-on agents (D) (not part of the core MDR-TB regimen)	D1	First-line oral anti-TB agents (Pyrazinamide, Ethambutol, Isoniazid HD)					
	D2	Bedaquiline[l]	Solid, liquid	–	–	**–**	**–**
		Delamanid[m]	Solid, liquid	–	–	**–**	**–**
	D3	P-aminosalicylic acid	Solid, liquid	1.0	2.0	8.0	4.0
		Imepenen-cilastatin	None	–	–	–	–
		Meropenem	None	–	–	–	–
		Amoxicillin/clavulanate	None	–	–	–	–
		(Thioacetazone)	None	–	–	–	–

In bold the CC that will be updated in 2017

[a]Modified according to WHO Guidelines for the programmatic management of drug-resistant tuberculosis

[b]Indirect proportion method recommended. Other solid media methods (resistance ratio) have not been adequately validated for second-line drugs. Concentrations for the absolute concentration method were not evaluated

[c]Ethambutol 5 ug/ml in MGIT is not equivalent to other methods. Ethambutol testing in 7H11 not equivalent to 7H10

[d]Rifampicin borderline resistance can missed by MGIT and can cause discrepant phenotypic/genotypic results

[e]Amikacin. Critical concentration on 7H10 it is expected to be decreased

[f]Gatifloxacin the drug is not widely available. Few data are available to establish the CC

[g]Moxifloxacin. Two concentrations were proposed in 2012. Testing for moxifloxacin at 2 mg/L underestimate resistance and truncates the distribution of mutant strains. It is anticipated that new recommendations will decrease the CC for moxifloxacin and levofloxacin in order to reflect the distribution of wild type and mutated strains. A clinical breakpoint for moxifloxacin is under discussion

[h]Ofloxacin is not used for therapy . it is suggested to use the fluoroquinolone included in regimens

[i]Clofazimine provisional CC for MGIT will be proposed in 2017

[j]Cycloserine. Critical concentration in LJ will be withdrawn due to concerns of unreliable testing and reproducibility

[k]Linezolid temptative CC, CC on solid media will be proposed in 2017

[l]Bedaquiline CC will be proposed in 2017

[m]Delamanid CC will be proposed in 2017

Table 12.2 Main genomic regions associated with drug resistance in M. tuberculosis

Drug	Gene(s) containing drug resistance-conferring mutations
Amikacin	*rrs*
Bedaquiline	*mmp*L5, *mmp*S5, *Rv0678*, *atp*E, *pep*Q
Capreomycin	*rrs*, *tly*A
Clofazimine	*Rv0678*, *pep*Q
Delamanid	*ddn*, *fgd*1, *fbi*A, *fbi*B and *fbi*C
Ethambutol	*emb*A, *emb*B, *emb*C, *ubi*A
Ethionamide	*inh*A, *ndh*, *eth*A, *eth*R
Fluoroquinolones	*gyr*A, *gyr*B
Isoniazid	*kat*G, *inh*A, *ndh*, *fur*A, *kas*A
Kanamycin	*rrs*, *gid*B, *eis*, *tap*, *whi*B7
Linezolid	*rpl*C, *rrl*
Pyrazinamide	*pnc*A, *rps*A, *pan*D
Rifampicin	*rpo*B
Streptomycin	*rps*L, *rrs*, *gid*B, *tap*, *whi*B7

Drug resistance can be defined according to different criteria. Clinical resistance is based on breakpoints determining the likelihood of therapeutic failure during treatment. Even though this definition is useful for clinical practice, it does not consider low-level resistance mechanisms that increase the MIC without reaching the breakpoint, thus representing an hallmark of a possible evolutionary trend towards high-level (clinically relevant) resistance (Baquero 2001; Martínez et al. 2015). Alternatively, resistance could be referred according to an epidemiological definition: the European Committee on Antimicrobial Susceptibility Testing (EUCAST) established common ECOFFs defined as the MIC value corresponding to the upper limit of the wild-type population of a specific bacterial species (Kronvall 2010). In this case, low-level resistance can be also determined. Table 12.2 summarizes main genomic regions involved in the emergence of drug resistance to preeminent anti-tubercular drugs (see Chap. 14).

Rifampicin is an excellent example for molecular evaluation of phenotypic resistance: more than 95% of the mutations causing drug resistance are mapping in a short (81 nucleotides in length) "hot-spot" region in the *rpo*B gene (Ramaswamy and Musser 1998). Despite the variability of mutations found, four mutations (D516V, H526Y, H526D, and S531L – refer to Andre 2016 for corresponding *M. tuberculosis* codon numbering system (Andre et al. 2016)) account for more than 70% of resistant clinical isolates. Different mutations can also be associated to different fitness (Comas et al. 2011; Brandis et al. 2012; de Vos et al. 2013). Due to the lower rate of spontaneous mutations, rifampicin resistance alone is rarely observed, and the majority of rifampicin-resistant isolates have already accumulated mutations conferring resistance to isoniazid (Coker 2004). This observation initially leads policy-makers to consider rifampicin-resistance as a marker for MDR tuberculosis. However, further data showed that rifampicin resistance is not a good surrogate marker at the global level for MDR-TB and its reliability is strictly dependent on the rate of primary resistance to rifampicin in the setting considered (WHO 2004).

Isoniazid resistance is caused by mutations affecting different genomic regions: approximately 60% and 20% of resistant cases are caused respectively by mutations affecting the codon 315 in the *kat*G gene and the position −15 in the *inh*A gene promoter region (Seifert et al. 2015). Different mutations in these regions slightly increase the percentage of isolates with at least one reported mutation. The mutation frequency was

also found to be different according to the geographical distribution (Seifert et al. 2015). Other genomic regions involved in isoniazid resistance are in the *ndh*, *fur*A and *kas*A genes, however their contribution is usually below 1% (Banerjee et al. 1994; Kelley et al. 1997; Sreevatsan et al. 1997; Banerjee et al. 1998; Slayden and Barry 2000; Lee et al. 2001; van Doorn et al. 2003; Vilchèze et al. 2005). Recently, another potential target of the drug has been proposed: the MymA protein, a flavin-containing monooxygenase was found to be inhibited by isoniazid in modeling and biochemical analyses (Saraav et al. 2017). Further studies will help in understanding any role of this putative target in the development of isoniazid resistance. Other studies focusing on bacterial persistence underlined the complexity of the interaction between antibiotics and bacteria. In particular, studies on isoniazid showed that single-cell dynamics of the isoniazid-activating enzyme catalase-peroxidase (KatG) are driving the success of the drug; thus, persistence to isoniazid was likely due to reversible phenotypic tolerance rather than stable genetic mutations (Wakamoto et al. 2013). Several mechanisms of bacterial persistence have been described, and many were also found in *M. tuberculosis* (Nathan 2012; Dartois et al. 2016; Harms et al. 2016).

Ethambutol resistance is caused by mutations affecting the *emb*CAB operon, and the *emb*R gene encoding for its regulator (Telenti et al. 1997). Most frequent mutations are observed at codon 306 of the *emb*B gene (approximately 70% of resistant cases) (Zhang and Yew 2015). Recently, mutations affecting the *ubi*A gene have also been associated with ethambutol resistance (Safi et al. 2013; He et al. 2015). Researchers started to characterize the frequency of mutations at this locus in clinical isolates. However, they seems associated to particular geographical regions (Xu et al. 2015; Lingaraju et al. 2016).

Pyrazinamide resistance is caused by mutations affecting the *pnc*A gene encoding the pyrazinamidase enzyme required to convert pyrazinamide to its active form pyrazinoic acid (Zhang and Mitchison 2003). Resistance-conferring mutations are spread along the entire gene, including the promoter region, and a clear hot-spot region cannot be identified. In addition, more than 600 different mutations (including indels) have been described in the literature, making the detection of individual mutations or a restricted subset of them useless for diagnostic purposes (Miotto et al. 2014; Ramirez-Busby and Valafar 2015; Whitfield et al. 2015). In some cases, insertion of the IS*6110* mobile genetic element in the gene has been also reported (Gillespie 2002). According to the most recent systematic review available, mutations in the *pnc*A locus are responsible of 83% of pyrazinamide resistant cases (Ramirez-Busby and Valafar 2015). Mutations affecting two novel genetic loci have been associated with the emergence of pyrazinamide resistance: evidences suggest that *rps*A and *pan*D genes are involved in the mechanism of action of the drug (Yang et al. 2015; Pandey et al. 2016), however only a relatively small number of clinical isolates have been characterized for these genomic regions, and the clinical relevance of these targets remains questionable. Pyrazinamide monoresistance is an extremely rare phenotype (with the exception of *M. bovis* and *M. canettii*) and any phenotypic resistance in the absence of *pncA* mutations should be interpreted with extreme caution.

Streptomycin resistance is caused by mutations in the *rps*L gene, with codons 43 and 88 as the most affected ones. Mutations in this genetic locus account for approximately 50% of streptomycin-resistant clinical isolates. An additional 10% of resistant strains harbor mutations in regions 530 and 912 of the *rrs* gene (Sreevatsan et al. 1996). Other genes described as involved in streptomycin resistance are *gid*B, *tap* and *whi*B7 genes (Okamoto et al. 2007; Wong et al. 2011; Reeves et al. 2013). Mutations in the *rrs* gene but at different region (region 1400) are also associated with resistance to amikacin, kanamycin and capreomycin (Alangaden et al. 1998; Maus et al. 2005). In addition, mutations in the promoter region of *eis* gene and in the *tly*A gene are responsible of kanamycin and capreomycin resistance, respectively (Johansen et al. 2006; Zaunbrecher et al. 2009). According to Georghiou et al., the *rrs* 1401 mutation alone is found in 70–80% of amikacin and capreomycin

resistant isolates, and in 60% of strains resistant to kanamycin (Georghiou et al. 2012). The contribution of mutations in the promoter region of the *eis* gene to kanamycin resistance is variable across different geographical regions, however in some settings it can rise up to more than 60% of resistant cases (Hoshide et al. 2014). The contribution of mutations in *tly*A gene in determining capreomycin resistance remains low to moderate (5–10%) (Campbell et al. 2011; Georghiou et al. 2012).

Similarly to the other anti-tubercular drugs, resistance to fluoroquinolones is promoted by sub-optimal bacterial-drug exposure (Miotto et al. 2015). Main drug resistance-associated mutations are found in the quinolone resistance determining regions (QRDRs) of the *gyr*A and *gyr*B genes. In particular, 80% of resistant cases harbor mutations in the QRDR of the *gyr*A gene, whereas mutations in the *gyr*B gene contribute for less than 1% of resistant cases (Avalos et al. 2015). Several studies showed that the levels of resistance to the different fluoroquinolones (namely levofloxacin, ciprofloxacin, ofloxacin, moxifloxacin, gatifloxacin) are associated with the type of mutation present in the *gyr*A and *gyr*B genes, however MIC values for isolates with the same mutation vary widely (Kam et al. 2006; Nosova et al. 2013; Kambli et al. 2015; Chien et al. 2016; Farhat et al. 2016a, b). Concordance in resistance testing among fluoroquinolones was found to be low (Farhat et al. 2015; Coeck et al. 2016). These observations lead to consider the possibility to use later-generation fluoroquinolones (moxifloxacin, gatifloxacin) to treat cases resistant to earlier generation (levofloxacin, ciprofloxacin, ofloxacin) fluoroquinolones (resistance detected by phenotypic tests using CC published in 2012). According to Farhat et al. (2015), nearly 30% of strains had moxifloxacin MICs in the intermediate range, likely below the peak serum concentrations of the drug, and thus clinically treatable with standard doses of moxifloxacin (despite distribution of drugs in the granulomas could not reflect the serum level of the antibiotic, as suggested by Dartois and Barry 2013). Similarly, a previous study described ofloxacin-resistant cases showing improved treatment outcomes when treated with moxifloxacin (Jo et al. 2014). Van Deun et al. suggested that specific *gyr*A gene mutations can be used to predict poor treatment outcome in MDR tuberculosis, and in particular mutations other than Ala substitution at codon 94 are associated with high-level resistance to gatifloxacin and moxifloxacin (Rigouts et al. 2016).

Data for drugs recently introduced in the treatment of MDR and XDR tuberculosis such as linezolid, delamanid, bedaquiline and clofazimine are scarce, both in terms of genetic loci involved and prevalence (Bloemberg 2015; Xu 2017). Further work is needed to fill present knowledge gaps. Resistance to linezolid is mainly caused by mutations in *rpl*C and *rrl* genes (Hillemann et al. 2008; Beckert et al. 2012; Makafe et al. 2016), whereas resistance to delamanid is mediated by mutations affecting *ddn*, *fgd*1, *fbi*A, *fbi*B and *fbi*C genes (Stover et al. 2000; Choi et al. 2001; Matsumoto et al. 2006; Manjunatha et al. 2006; Feuerriegel et al. 2011). Resistance to delamanid has been observed in strains never exposed to the drug and presenting mutations in *ddn* (Schena et al. 2016). Mutations affecting *mmp*L5, *mmp*S5, *pep*Q, *Rv0678*, *atp*E genes cause resistant to bedaquiline (Andries et al. 2005; Huitric et al. 2010; Andries et al. 2014; Hartkoorn et al. 2014; Almeida et al. 2016; Segala 2012). Mutations in *Rv0678* and *pep*Q were found to confer cross-resistance between bedaquiline and clofazimine, an antileprosy drug recently gaining attention for treating MDR tuberculosis (Hartkoorn et al. 2014; Almeida et al. 2016). In particular, mutations in Tv0678 were linked to the upregulation of *mmp*S5 and *mmp*L5 genes. Resistance mediated by *pep*Q mutation seems to be associated with increased drug efflux, but this is not due to upregulation of *mmp*L5 and *mmp*S5 expression as in the case of resistance mediated by mutations in the *Rv0678* gene.

The frequency of mutations in these genomic regions remain to be determined in clinical isolates and only limited data on small number of samples are available.

The mycobacterial cell wall permeability barrier and active multidrug efflux pumps represent

a relevant role in the development of phenotypic drug resistance in *M. tuberculosis* (De Rossi et al. 2006; Escribano et al. 2007; Louw et al. 2009; da Silva et al. 2011). Studying these specific aspects of the biology of *M. tuberculosis* requires more sophisticated experiments and the identification of mutations related to these mechanisms is more complex; therefore, there are very limited data available on clinical isolates covering these mechanisms of resistance (see Chap. 14).

For drugs sharing the same molecular target cross-resistance is commonly observed. This phenomenon has been reported for fluoroquinolones, second-line injectable drugs, but also for isoniazid and ethionamide. However, it should be noted that certain mutations confer different level and pattern of resistance among the different drugs. Table 12.3 summarizes some of the main information available on cross-resistance and its molecular bases, together with main references.

12.3.2 Molecular Identification of the Drug Susceptibility Profile of *M. tuberculosis*

The bottleneck in the molecular detection of drug resistance in *M. tuberculosis* is the limited knowledge of the relevant mutations responsible of the resistant phenotype (Zhang and Yew 2015). There are several genotypic approaches to detect known mutations causing drug resistance; the use of the polymerase chain reaction (PCR) in the diagnosis of tuberculosis was introduced in 1990 (Patel et al. 1990) and few years later, starting on 1993 (Telenti et al. 1993), PCR-based assays have been developed for the detection of drug resistance in *M. tuberculosis* (for reviews on early developed methods please refer to Caws 2001; García de Viedma 2003). Together with undeniable advantages in terms of both ease of use and reduced time-to-results, the introduction of molecular assays for direct drug susceptibility profiling of *M. tuberculosis* in specimens also reduced biosafety risks associated with the manipulation of live pathogens, especially in resource-limited settings (WHO 2008; Parsons et al. 2011; WHO 2013; Somoskovi and Salfinger 2015).

Although molecular techniques can detect a single bacillus in a specimen (at least in theory), sensitivity can be hampered by the presence of inhibitors in clinical specimens and loss of nucleic acids during specimen processing. In addition, the need to detect mutations affecting multiple genes (*e.g.* those required to identify delamanid or bedaquiline resistances) and/or the multiplicity of mutations on a single target (*e.g.* mutations in *pnc*A gene associated with pyrazinamide resis-

Table 12.3 Genes involved in cross resistance to anti-TB drugs

Cross-resistances	Genomic region	Example mutations	References
Amikacin, kanamycin, capreomycin	*rrs*	a1401g, a1484t	Blumberg et al. (2003), Maus et al. (2005), and Georghiou et al. (2012)
Bedaquiline, clofazimine	*Rv0678*	S63R, R134Stop	Hartkoorn et al. 2014
	*pep*Q	A14ins c, R271del c, L44P	Almeida et al. (2016)
Ciprofloxacin, levofloxacin, ofloxacin, moxifloxacin	*gyr*A	G88C, A90V, S91P, D94N, D94G, D94Y N538D, E540V,	Malik et al. (2012), Nosova et al. (2013), Imperiale et al. (2014), and Willby et al. (2015)
	*gyr*B	R485C + T539N	
Levofloxacin, ofloxacin	*gyr*B	D500H, D500N	Malik et al. (2012)
Isoniazid, ethionamide	*inh*A	c-15t	Imperiale et al. (2014) and Rueda et al. (2015)
Rifampicin, rifabutin	*rpo*B	Q513E, Q513K, Q513L, Q513P, S531L, S531F, S531W, H526D, H526Y, H526R, D516A + R529Q	Bodmer et al. (1995), Sintchenko et al. (1999), Goldstein (2014), Jamieson et al. (2014), Imperiale et al. (2014), and Berrada et al. (2016)

tance) considerably limit the choice to the techniques with a sufficient capacity for multiplexing.

Starting in 2008, molecular tools detecting rifampicin resistance-associated mutations have been formally endorsed by the World Health Organization, as they represent cost-effective rapid diagnostics for fast detection of resistant cases. In particular, based on evidence and expert opinion, the WHO endorsed the use of molecular line probe assays (LiPAs), and the Xpert MTB/RIF assay (Cepheid) for the rapid detection of MDR tuberculosis cases (WHO 2008, 2013; Gilpin et al. 2016).

The commercially available LiPAs are based on targeted amplification of specific regions of the *M. tuberculosis* genome followed by hybridization of the amplicons to oligo probes immobilized on nitrocellulose strips. The INNO-LiPA Rif.TB (Innogenetics) has been designed to detect rifampicin resistance alone by targeting the hot-spot region of the *rpo*B gene with five wild-type probes and four probes specific for most frequent rifampicin resistance-associated mutations (D516V, H526Y, H526D, and S531L). Reported pooled sensitivity and specificity are 97% (95% CI 95.0–98.0) and 99% (95% CI 98.0–100.0), respectively (Morgan et al. 2005). The Genotype MTBDR*plus* assay by Hain Lifescience detects both rifampicin and isoniazid resistance. The test targets the hot-spot region of the *rpo*B gene with eight wild-type probes and four probes specific for most common rifampicin resistance-associated mutations (D516V, H526Y, H526D, and S531L). The detection of isoniazid resistance is enabled by one wild-type probe plus two mutated probes for the *kat*G gene (S315T, nucleotidic substitutions agc/acc and agc/aca) and two wild-type probes plus four mutated probes for the promoter region of the *inh*A gene (c-15t, a-16g, t-8c, and t-8a). The pooled sensitivity for the detection of rifampicin resistance was reported 98.1% (95% CI 95.9–99.1), with a specificity of 98.7% (95% CI 97.3–99.4). Results for isoniazid showed lower sensitivity (84.3%, 95% CI 76.6–89.8) but high specificity (99.5%, 95% CI 97.5–99.9) (Ling et al. 2008). More recently a non-inferiority study of the new version of the GenoType MTBDR*plus* assay (version 2) and a newly developed LiPA assay named Nipro NTM + MTBDRTB assay (Nipro Corporation) was published. Both tests have been designed to detect rifampicin and isoniazid resistance. The GenoType MTBDR*plus* assay ver. 2 targets the regions already described for the MTBDR*plus* ver.1 assay. Similarly, the Nipro assay targets the hot-spot region of the *rpo*B gene for rifampicin resistance (5 wild-type probes plus four mutated probes targeting D516V, H526Y, H526D, and S531L substitutions), whereas for isoniazid resistance the assay targets four wild-type probes plus two mutated probes for the *kat*G gene (S315T, and S315N) and one wild-type probe plus four mutated probes for the promoter region of the *inh*A gene (c-15t, a-16g, t-8c, and t-8a). Non-inferiority of MTBDR*plus* ver.2 and Nipro assays to MTBDR*plus* ver.1 was demonstrated for rifampicin and isoniazid resistance detection (Nathavitharana et al. 2016).

The Genotype MTBDR*sl* (Hain Lifescience) is the only molecular test designed for the detection of resistance to second-line drugs. The assay targets the QRDR of the *gyr*A gene (three wild-type probes plus six mutated probes targeting mutations A90V, S91P, D94A, D94N/Y, D94G, and D94H) and the region 1400 of the *rrs* gene (two wild-type probes plus two mutated probes targeting a1401g, and g1484t). In addition, the assay targets the codon 306 of the *emb*B gene for ethambutol resistance. The pooled sensitivity for detecting fluoroquinolone resistance on isolates was 83.1% (95% C.I. 78.7–86.7) and the pooled specificity was 97.7% (95% C.I. 94.3–99.1), respectively. Similar performances were found for direct testing in clinical specimens. Performance on second-line injectable drugs resistance (amikacin, kanamycin, and capreomycin) showed a pooled sensitivity of 76.9% (95% C.I. 61.1–87.6) and a pooled specificity of 99.5% (95% C.I. 97.1–99.9), respectively (Theron et al. 2014). The new version of the MTBDR*sl* assay (version 2) in addition to *gyr*A and *rrs* genes targets also the QRDR of the *gyr*B gene for fluoroquinolone resistance (one wild-type probe plus two mutated probes targeting N538D, and E540V), and the promoter region of the *eis* gene for kanamycin resistance (three wild-type probes

plus one mutated probe targeting c-14t); the target region for ethambutol has been removed. The new version of the assay showed improved performances (Tagliani et al. 2015; Brossier et al. 2016), and very recently, the WHO also provided recommendations for the use of the MTBDR*sl* for the detection of resistance to second-line antituberculosis drugs (WHO 2016).

Despite the good performances of the LiPAs on clinical isolates and smear-positive clinical specimens, it should be noted that the WHO does not recommend the use of these assays for smear-negative samples. MTBDR*sl* assay (version 2) is recommended to triage patients for the short MDR regimen. Although not perfect, this assay can identify patients that are resistant to all second line injectable drugs (presenting the a1401g mutation in *rrs*) from those that may still respond to amikacin (position 1402, *rrs* gene). Patients presenting mutations in the promoter region of *eis* could also still be treatable with capreomycin.

At present guidelines for interpreting the genotype to predict drug responses are under preparation and more data are collected to support the interpretation of mutation pattern for clinical management of DR tuberculosis.

The Cepheid Gene Xpert MTB/RIF system is a fully automated real time PCR-based assay for the detection of *M. tuberculosis* DNA and mutations associated with rifampicin resistance, directly in clinical specimens (Boehme et al. 2010). The assay uses semi-nested PCR to amplify the hot-spot region of the *rpo*B gene (RRDR) resistance is detected by five molecular beacons targeting both wild-type sequences and the most common mutated codons. For the detection of rifampicin resistance, the pooled sensitivity and specificity were 95% and 98%, respectively (Steingart et al. 2014). Due to the high sensitivity of the assay, the Xpert MTB/RIF is also recommended for smear-negative specimens, and extra-pulmonary samples (WHO 2013; Denkinger et al. 2014). Despite these recommendations the sensitivity on paucibacillary samples remains suboptimal compared to liquid culture. Additional limitations for this test are the suboptimal negative predictive value (Theron et al. 2014), the capacity to detect heteroresistance to rifampicin (Zetola 2014), low capacity to detect the C533G mutation (Rufai et al. 2014) and occasional rifampicin resistant false positive cases due to delays in the signal generated by the probes D and E (Williamson 2012) or detection of silent mutations such as F514F. A new generation of Xpert assay, named Ultra, was recently developed and evaluated in a non inferiority study including the previous version (https://www.finddx.org/wp-content/uploads/2017/03/Ultra-WHO-report_24MAR2017_FINAL.pdf). The Ultra is an improved version of the previous test (G4) working on the same platform after upgrade of software. The main differences between G4 and Ultra are: the increased volume of the PCR chamber, the target genes for detection of MTB (two multicopy genes IS6110 and IS 1081) and faster reaction kinetics. Rifampicin resistance is detected using the melting temperature curves of RRDR-specific probes. Samples are defined positive according to five semiquantitative categories. The first four (high, medium, low and very low) correspond to the G4 categories while the last one, "trace", is new.

The Ultra cartridges have been endorsed by WHO in April 2017. In the first prospective multicenter study comparing G4 and Ultra, the Ultra showed an increased sensitivity (+17%) in smear negative respiratory samples and HIV coinfected subjects (+ 14%). Specificities of Ultra and G4 for case detection were 95.6% and 98.3%, overall, and 93.5% and 98.4% among patients with a history of tuberculosis. The decreased specificity in patients with history of tuberculosis is mainly in the category of results labeled as "trace". Samples resulting positive as "trace" will not be further analyzed for rifampicin resistance. Large amount of data will be needed to define the diagnostic value of the "trace" category. "Trace" positive subjects could be managed differently based on local TB epidemiology, patient's history and immunological status (Chakravorty et al. 2017).

As already mentioned, the clinical value of molecular assays relies upon our knowledge of the mutations involved in the emergence of drug resistance in *M. tuberculosis*, evidences

suggest that future drug resistance diagnostics will need to be able to detect high numbers of mutations to impact on the management of patients with drug-resistance tuberculosis (Farhat et al. 2016b). However, additional considerations should be taken into account in the evaluation of such tools. First, whereas in settings with a high prevalence of rifampicin resistance and MDR-TB, these tests may be a valuable component of an MDR-TB management strategy, molecular tests for rifampicin resistance alone cannot accurately predict resistance in areas with a low prevalence of rifampicin resistance (Arentz et al. 2013; WHO 2013; Drobniewski et al. 2015). Thus, careful evaluation of the setting should performed prior introduction of molecular assays. Second, although some molecular tools are often easier to be performed compared to phenotypic DST, interpretation challenges may arise. Whereas rare or novel mutations usually do not account for the majority of resistance determination based on the absence of wild-type probe hybridization, continuous evaluation of geographical mutation frequencies might be needed for maximizing the impact of molecular diagnostics (Seifert et al. 2016; Sanchez-Padilla et al. 2015). Similarly, false-positive rifampicin resistance detection or detection of *M. tuberculosis* DNA by Xpert MTB/RIF assay in culture-negative patients can be confusing and detrimental for patient management (Huh et al. 2014; Lippincott et al. 2015). Third, clear data on the relationship between genotype, phenotype and response to treatment are limited. Phylogenetic polymorphisms, mutations associated with hyper-susceptibility and/or different level of resistance, and differences related to the phenotypic testing method used as reference have been described. Distinctive geographical distributions of drug resistance-associated mutations further complicate the clinical interpretation of genetic polymorphisms (Aubry et al. 2006; Rigouts et al. 2013; Feuerriegel et al. 2014; Hoshide et al. 2014; Van Deun et al. 2015; Kambli et al. 2015; Singh et al. 2015; Coeck et al. 2016; Berrada et al. 2016; Kambli et al. 2016).

Target product profiles in terms of minimal requirements, performances and controls for developing new molecular diagnostic assays for drug-resistant tuberculosis and guidelines for their successful evaluation have been developed to guide the development of new assays (Wells et al. 2013; Kik et al. 2014; WHO 2015; Denkinger et al. 2015). Currently, large efforts are devoted to fill the gaps in our understanding of the genotype-phenotype relationships. We can now take advantage from the onset of next generation sequencing (NGS). NGS is making whole genome sequencing (WGS) affordable in the broader field of microbiology (Punina et al. 2015; Gilchrist et al. 2015). Several automated or semi-automated tools for interpreting *M. tuberculosis* drug resistance in WGS data are already available (Steiner et al. 2014; Flandrois et al. 2014; Bradley et al. 2015; Feuerriegel et al. 2015; Coll et al. 2015). Recent studies highlight the need for standardized databases for interpreting genotype-phenotype correlation in clinical contexts (Witney et al. 2015). At this end, a large collaborative project involving academic institutions, public health agencies, and nongovernmental organizations has been established to develop a tuberculosis relational sequencing data platform (ReSeqTB) for improving understanding of the relationships between genotype, phenotype and clinical outcomes (Starks et al. 2015; Schito and Dolinger 2015; https://platform.resqtb.org/). Another consortium named "CRyPTIC" (Comprehensive Resistance Prediction for Tuberculosis, www.crypticproject.org/) aims at developing a sufficient number of sequences to unveil all possible variants leading to drug resistance.

12.4 Discrepancies Between Phenotypic and Genotypic Tests

Since the introduction of molecular tests for the diagnosis of drug resistant tuberculosis, several reports showing conflicting results were published. Discrepancies between molecular methods to detect drug resistance (including

Table 12.4 Main reasons for genotype/phenotype discrepancies

Mutation is out side the region targeted by the molecular assay
Mutation confers low level resistance and CC is set to high
Presence of unknown mechanisms conferring DR
Trivial errors in the performance of DST
Presence of heteroresistance not detected by molecular methods
Molecular assay detect silent mutations
Errors due to probe interaction/binding in LPA or other molecular assays

rifampicin resistance) and traditional phenotypic methods have caused confusion and in many settings have decreased the confidence in molecular tests. This has resulted in delay in starting treatment or inappropriate treatment when priority was given to phenotypic data. These discrepancies could be due to real "false positive" or "false negative" results of the tests used for the determination (both genotypic or phenotypic) or can be linked to more complex reasons. Table 12.4 lists the most frequent reasons for the discrepancies. As already mentioned, phenotypic tests performed on different media may yield conflicting results when the MIC of the strain for the drug tested is close to the critical concentration (Coeck et al. 2016). Technical issues related to the methodology used for DST, quality of media and drugs, experience of the staff can strongly affect the reliability of the test. Testing *M. tuberculosis* sensitivity by phenotypic methods should only be performed in laboratories maintaining a high standard in performance and with a consistent workload.

Additional causes for discrepancies are linked to the use of CCs established at value that are too high and don't represent the true distribution of the wild type and mutant bacterial population. When drugs can be used at higher doses without causing serious side effects, a clinical breakpoint can be established, and in some case it is recommended to test the drug at two concentrations. Phenotypic sensitivity at the higher concentration should not be interpreted as a "true discrepancy" but as useful information for the clinical management of the patient. Moxifloxacin is a candidate drug for testing at two concentrations.

Our knowledge on the mechanisms conferring drug resistance is still limited, and in some case determinants that are causing resistance are not properly investigated. The implementation of whole genome sequencing has highly improved our knowledge of genomic variants causing drug resistance. For some minority variants the association to drug resistance will need to be confirmed by reverse genetic experiments. Some molecular assays are not able to discriminate "silent" from "non-silent" mutations: F514F is the most common silent mutation in rpoB, if detected by molecular assay the sample may be misinterpretated as rifampicin-resistant.

Table 12.5 summarizes some of the most common mutations causing DR and expected phenotype.

12.5 Whole Genome Sequencing as Novel Approach to Susceptibility Testing

The whole genome sequencing (WGS) approach offers a powerful alternative for diagnosis of drug-resistant tuberculosis, promising a rapid and accurate determination of all the clinically-relevant mutations (Drobniewski et al. 2015; Witney et al. 2015; Walker et al. 2015). Indeed, using this methodology, clinicians could promptly obtain relevant information on the best therapy to adopt, receiving information on sensitivity to the first-line drugs, as well as to second-line and new agents. In addition, it is now accepted that the emergence of drug resistance is not always caused by point mutations affecting only single genes, but the presence of other mechanisms, such as compensatory mutations,

Table 12.5 Most common mutations causing DR and the expected phenotype

Antimicrobial agent	Gene	Common mutations	Less common mutations	Expected phenotype
Isoniazid	*katG*	S315T	S315T, S315N, S315I, S315L, additional mutations in *katG* coding region	S315T is the most common mutation conferring resistance to INH, is associated to medium to high level of resistance, is the most represented mutation in MDR strains
	inhA promoter	−15C/T	−8T/C, −8T/A, −8T/G, −9G/T, −16A/G, −17G/T	Often associated with low-level INH resistance, confers ethionamide resistance.
				Some strains may tests sensitive by phenotypic DST
	inhA coding region fabG1	Different codons may be affected		Mutations in the coding regions are non detected by LPA commercial assays and/or assays targeting the promoter. Presence of mutations in the coding region in association with mutations in the promoter increases the level of resistance.
	ahpC promoter	−48C/T	−52G/A, −54G/A, −51C/T, −52G/T, −49A/G, −57G/A	Associated with INH resistance. Few data available
Rifampin	*rpoB*	S531L	L511P, L533P, D516Y, H526N,	S531 L is the most frequent mutation identified in MDR TB, associated with resistance to all rifamicins
		H526Y	S522L, H526L, H526A, H526C, D516F, D516V	F514F is the most common silent mutation. It can cause misinterpretation of resistance to rifampicin if not recognised.
		H526D	Q513A, Q513E, H526Y, H526D, H526R, S531W, S531F, S531L, V176F.	L511P, L533P, D516Y, H526N "disputed" mutations. Are associated to rifampicin resistance if tested on solid media, may test sensitive when tested in MGIT. Associated to poor clinical outcome should be interpreted as conferring resistance and rifampicin should not be counted as a fully active drug in the therapeutic regimen.
				S522 L, H526L, H526A, H526C, D516F, D516V: those mutations have been associated to rifampicin resistance and rifabutin sensitivity. This is still a disputed issue and more data on mic distribution for rifabutin are needed
Ethambutol	*embB*	M306V	M306L, M306I D354A, G406D, etc	M306V is the most frequent mutation associated with R to EMB.
				Not all mutations in *embB* are associated with EMB-R. Discrepant phenotypic/genotypic results are expected due to the CC used for in vitro testing

(continued)

Table 12.5 (continued)

Antimicrobial agent	Gene	Common mutations	Less common mutations	Expected phenotype
Pyrazinamide	*pncA*	No predominant mutations		Pyrazinamide monoresistance is associated to *M.bovis* and *M.bovis*/BCG due to the presence of a characteristic mutation in position 57. For all other cases phenotypic resistance to pyrazinamide in the absence of mutations should be interpreted as false resistance. *pncA* mutations are widely distributed throughout the gene and its promoter.
				Some mutations such as E37V, D110G, V163A, A170V and V180I, are not associated with PZA-R
	panD		I49V, I115T	Few data to support the role of *panD* as main determinant for PZA resistance. It is causing resistance in *M. canetti*
Quinolones	*gyrA*	D94G	D94Y, D94H, D94A, D94N, S91P; G88A, mutations outside hot spot	D94G causes resistance to all fluoroquinolones (including moxifloxacin and gatifloxacin) in the presence of this mutation fluoroquinolone treatment is not recommended or if performed should be condered potentially not effective despite results of "in vitro" phenotypic testing
		A90V		A90V is associated resistance to levofloxacin and to lower level resistance to moxifloxacin and gatifloxacin
Amikacin	*rrs*	1401A/G	1484G/T	1401A/G is the most frequent mutation; confers resistance to AMK and all second line injectables
				1484G/T may be associated with R to AMK.
	eis		−14C/T	may confer very low level resistance
Capreomycin	*rrs*	1401A/G	1402C/T, 1484G/T	1401A/G is the most common mutation; Usually associated with R to CAP.
				1402C/T and 1484G/T mutations may also be associated with R to CAP.
	tlyA	No predominant mutations		Mutations are widely distributed throughout the gene.
				Not all mutations are associated with R to CAP.
				role of the different mutations is still disputed
Kanamycin	*rrs*	1401A/G,	1402C/T, 1484G/T	1401A/G is the most common mutation and associated with high level R to KAN.
				1402C/T may be associated with low-level R to KAN.
	eis promoter	−10G/A		Confer KAN-resistance, short MDR regimen with kanamycin cannot be used for treatment. May induces increased mic to Amikacin. Capreomycin could still be effective
		−12C/T		
		−14C/T		
		−37G/T		
				−12C/T may confer low level R to KAN.

(continued)

Table 12.5 (continued)

Antimicrobial agent	Gene	Common mutations	Less common mutations	Expected phenotype
Bedaquiline	*atpE*	D28N, A63V		Mutations in C ring of the ATP synthase may be associated with BDQ resistance.
	mmpR	No predominant mutations		A new publication indicated mutations in mmpR may be associated with BDQ-R.
Delamanid	*fbiA (Rv3261), fbiB (Rv3262), fbiC (Rv1173), fgd1 (Rv0407)*			Mutations in genes involved in coenzyme F420 biosynthesis and metabolism has been proposed as possible mechanisms of resistance to DLM (Choi KP *et al*., *J. Bacteriol*. 2002) *several mutations have been observed but data correlating to DR are not yet available*
	ddn			*Stop codons in the ddn have been associated to high level resistance*

could explain the discrepancies observed between phenotypic and genotypic results. With the increasing number of genomic loci identified by WGS as linked to resistance, the value of this approach will increase in particular for use in laboratory routine (Drobniewski et al. 2015; Pankhurst et al. 2016). Recent studies underlined that over 100 genetic regions are involved in the drug resistance pathways and that mutations found within these regions could play relevant roles. WGS therefore appears the most suitable approach for a comprehensive analysis, given an appropriate validation of all the mutations by MIC and allelic exchange experiments, and considering the correlation with clinical outcomes (Zhang and Yew 2015). At the moment, WGS can be used to rapidly identify the known conferring-resistance mutations and, consequently, to guide individualized treatment decisions, even supporting for some drugs the phenotypic DST results, due to the reliability issues of the latter (Koser 2013). Among the advantages of WGS over the molecular tools currently recommended by the WHO, there is the possibility to provide information on the specific nucleotide substitutions that confer different levels of phenotypic resistance (e.g. mutations affecting codons 90 and 94 in gyrA) and the analysis of large genomic regions not limited to hotspot fragments (e.g. *pncA* complete coding and promoter sequencing; mutations outside the *rpoB* RRDR and *gyrA*-*gyrB* QRDR). WGS can also provide information to support conventional contact tracing for epidemiological studies, given its high discriminatory power in determining phylogenetic lineages (see Chap. 4), and in tracking the circulating strains and their relatedness (Drobniewski et al. 2015; WHO/UNITAID 2015; Witney et al. 2015). Thus, it may be possible to diagnose drug resistance and monitor transmission events at the same moment, with considerable impact on public health strategies (Arinaminpathy 2015). WGS platforms have been already adopted in many TB supranational and national reference laboratories, as well as in research laboratories: several groups are working to reduce the complexity of such technologies, from the hardware to the analysis part, with the final aim to make this technology accessible to all (Chap. 3). Already, several Countries are moving towards a centralized genomic approach for detection of sensitivity at least to first line antitubercular drugs. In addition WGS provides detailed information on the prevalence of strains and drug resistance patterns in the different settings, thus helping the strategies adopted by TB control programs at local and national levels (WHO/UNITAID 2015; Zignol et al. 2016). The cost of WGS varies depending on the technologies and numbers of sample analysed, and it has now probably reached the price range of the other tests performed in the hospital laboratories. The cost benefit depends also on the time needed to provide results, with a reduction of around 4 weeks compared to phenotypic

DST, avoiding also the use of ineffective and expensive drugs and hospital isolation sectors for long period of time (Drobniewski et al. 2015; Witney et al. 2015). Despite the great opportunity to provide a comprehensive analysis of MTB primary cultures including species identification, simultaneous determination of resistance to all the anti-TB drugs through the interrogation of the known molecular targets, and genotyping and phylogenetic investigation to track the transmission events, the use of the generated information is challenging in decentralized facilities due to the computational capacity and bioinformatics skills required, and to the lack of standardized reference, analysis pipelines, and interpretation tools (Schito and Dolinger 2015).

Moving from culture-based WGS to direct analysis from clinical samples with fully automated platforms could be the next step to make this approach suitable for high burden settings. Commercial tests based on NGS of specific targets are under development and will be available in a near future.

12.6 Clinical Considerations

Moving into the era of "personalized medicine" requires an appropriate and accurate classification of the bacterial strains causing TB for both the sensitivity patterns and the genotype. Treatment of TB and of drug-resistant TB in particular is still very long and associated with toxicity and irreversible side effects. Treatment initiation in the absence of data on the susceptibility of the strain to the drugs selected should be avoided whenever possible. Each patient deserves a reliable drug sensitivity test done under the best conditions in a quality assured laboratory.

The introduction of additional therapeutic options, ranging from the adoption of the short MDR regimen to the introduction of new or repurposed drugs, requires a "triaging" of the patient with MDR-TB in the shortest possible time, it is clear that only molecular tools can respond to this need.

In the past few years DST for *M. tuberculosis* has evolved from a mostly "home made" test performed in few laboratories with doubtful results with turnaround time of months to a much needed high-tech test. The promise of WGS is now the "all in one" approach, with a prediction of the resistance pattern associated to epidemiological and genotypic information from clinical samples.

Although we recognize that rapid molecular tests are still unable to predict sensitivity or resistance in 100% of cases, they are still able to guide therapy in the high majority of cases allowing not starting or early discontinuation of potentially toxic therapy in cases in which resistance can be predicted.

It is becoming clear that the concept of "one gold standard method" for testing *M.tuberculosis* susceptibility to antibiotics is challenged by the fact that the different tests are providing results that at first may appear conflicting. This is causing confusion among clinicians and reluctance in modifying therapy. We need to accept that each drug may have a different testing standard and that for some drug the genotypic results will overrule the phenotype.

In some cases the use of MIC will provide substantial information to decide on the discontinuation of a therapy.

In the future, the same investment should be made in training clinician in the interpretation of molecular tests and MIC-based test, that we have devoted to train microbiologists in the use of molecular tests in order to translate into clinical action the information that the technology will allow to collect.

References

Alangaden GJ, Kreiswirth BN, Aouad A, Khetarpal M, Igno FR, Moghazeh SL, Manavathu EK, Lerner SA (1998) Mechanism of resistance to amikacin and kanamycin in Mycobacterium tuberculosis. Antimicrob Agents Chemother 42(5):1295–1297

Almeida D, Ioerger T, Tyagi S, Li SY, Mdluli K, Andries K, Grosset J, Sacchettini J, Nuermberger E (2016) Mutations in pepQ confer low-level resistance to bedaquiline and clofazimine in Mycobacterium tuberculosis. Antimicrob Agents Chemother 60(8): 4590–4599

Andre E, Goeminne L, Cabibbe A, Beckert P, Kabamba Mukadi B, Mathys V, Gagneux S, Niemann S, Van Ingen J, Cambau E (2016) Consensus numbering sys-

tem for the rifampicin resistance-associated rpoB gene mutations in pathogenic mycobacteria. Clin Microbiol Infect 23:167–172. pii: S1198-743X(16)30393–7

Andries K, Verhasselt P, Guillemont J, Göhlmann HW, Neefs JM, Winkler H, Van Gestel J, Timmerman P, Zhu M, Lee E, Williams P, de Chaffoy D, Huitric E, Hoffner S, Cambau E, Truffot-Pernot C, Lounis N, Jarlier V (2005) A diarylquinoline drug active on the ATP synthase of Mycobacterium tuberculosis. Science 307(5707):223–227

Andries K, Villellas C, Coeck N, Thys K, Gevers T, Vranckx L, Lounis N, de Jong BC, Koul A (2014) Acquired resistance of Mycobacterium tuberculosis to bedaquiline. PLoS One 9(7):e102135

Arentz M, Sorensen B, Horne DJ, Walson JL (2013) Systematic review of the performance of rapid rifampicin resistance testing for drug-resistant tuberculosis. PLoS One 8(10):e76533

Arinaminpathy N, Dowdy D (2015) Understanding the incremental value of novel diagnostic tests for tuberculosis. Nature 528(7580):S60–S67. https://doi.org/10.1038/nature16045

Aubry A, Veziris N, Cambau E, Truffot-Pernot C, Jarlier V, Fisher LM (2006) Novel gyrase mutations in quinolone-resistant and -hypersusceptible clinical isolates of Mycobacterium tuberculosis*:* functional analysis of mutant enzymes. Antimicrob Agents Chemother 50(1):104–112

Avalos E, Catanzaro D, Catanzaro A, Ganiats T, Brodine S, Alcaraz J, Rodwell T (2015) Frequency and geographic distribution of gyrA and gyrB mutations associated with fluoroquinolone resistance in clinical Mycobacterium tuberculosis isolates: a systematic review. PLoS One 10(3):e0120470

Banerjee A, Dubnau E, Quemard A, Balasubramanian V, Um KS, Wilson T, Collins D, de Lisle G, Jacobs WR Jr (1994) inhA, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis. Science 263(5144):227–230

Banerjee A, Sugantino M, Sacchettini JC, Jacobs WR Jr (1998) The mabA gene from the inhA operon of Mycobacterium tuberculosis encodes a 3-ketoacyl reductase that fails to confer isoniazid resistance. Microbiology 144(Pt 10):2697–2704

Baquero F (2001) Low-level antibacterial resistance: a gateway to clinical resistance. Drug Resist Updat 4(2):93–105

Beckert P, Hillemann D, Kohl TA, Kalinowski J, Richter E, Niemann S, Feuerriegel S (2012) rplC T460C identified as a dominant mutation in linezolid-resistant Mycobacterium tuberculosis strains. Antimicrob Agents Chemother 56(5):2743–2745

Berrada ZL, Lin SG, Rodwell TC, Nguyen D, Schecter GF, Pham L, Janda JM, Elmaraachli W, Catanzaro A, Desmond E (2016) Rifabutin and rifampin resistance levels and associated rpoB mutations in clinical isolates of Mycobacterium tuberculosis complex. Diagn Microbiol Infect Dis 85:177–181. pii: S0732–8893(16)30001–3

Bloemberg GV, Keller PM, Stucki D, Trauner A, Borrell S, Latshang T, Coscolla M, Rothe T, Hömke R, Ritter C, Feldmann J, Schulthess B, Gagneux S, Böttger EC (2015) Acquired resistance to Bedaquiline and Delamanid in therapy for tuberculosis. N Engl J Med 373(20):1986–1988

Blumberg HM, Burman WJ, Chaisson RE, Daley CL, Etkind SC, Friedman LN, Fujiwara P, Grzemska M, Hopewell PC, Iseman MD, Jasmer RM, Koppaka V, Menzies RI, O'Brien RJ, Reves RR, Reichman LB, Simone PM, Starke JR, Vernon AA, American Thoracic Society, Centers for Disease Control and Prevention and the Infectious Diseases Society (2003) American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America: treatment of tuberculosis. Am J Respir Crit Care Med 167(4):603–662

Bodmer T, Zürcher G, Imboden P, Telenti A (1995) Mutation position and type of substitution in the beta-subunit of the RNA polymerase influence in-vitro activity of rifamycins in rifampicin-resistant Mycobacterium tuberculosis. J Antimicrob Chemother 35(2):345–348

Boehme CC, Nabeta P, Hillemann D, Nicol MP, Shenai S, Krapp F, Allen J, Tahirli R, Blakemore R, Rustomjee R, Milovic A, Jones M, O'Brien SM, Persing DH, Ruesch-Gerdes S, Gotuzzo E, Rodrigues C, Alland D, Perkins MD (2010) Rapid molecular detection of tuberculosis and rifampin resistance. N Engl J Med 363(11):1005–1015

Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, Earle S, Pankhurst LJ, Anson L, de Cesare M, Piazza P, Votintseva AA, Golubchik T, Wilson DJ, Wyllie DH, Diel R, Niemann S, Feuerriegel S, Kohl TA, Ismail N, Omar SV, Smith EG, Buck D, McVean G, Walker AS, Peto TE, Crook DW, Iqbal Z (2015) Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis. Nat Commun 6:10063

Brandis G, Wrande M, Liljas L, Hughes D (2012) Fitness-compensatory mutations in rifampicin-resistant RNA polymerase. Mol Microbiol 85(1):142–151

Brossier F, Guindo D, Pham A, Reibel F, Sougakoff W, Veziris N, Aubry A, French National Reference Center for Mycobacteria (Members of the French National Reference Center for Mycobacteria: Christine Bernard, Emmanuelle Cambau, Vincent Jarlier, Faiza Mougari, Laurent Raskine, Jérôme Robert) (2016) Performance of the New V2.0 of the GenoType MTBDRsl test for the detection of resistance to second-line drugs in multidrug-resistant Mycobacterium tuberculosis complex strains. J Clin Microbiol 54:1573–1580. pii: JCM.00051–16. [Epub ahead of print]

Campbell PJ, Morlock GP, Sikes RD, Dalton TL, Metchock B, Starks AM, Hooks DP, Cowan LS, Plikaytis BB, Posey JE (2011) Molecular detection of mutations associated with first- and second-line drug resistance compared with conventional drug susceptibility testing of mycobacterium tuberculosis. Antimicrob Agents Chemother 55(5):2032–2041

Canetti G, Froman S, Grosset J, Haudoroy P, Langerová M, Mahler HT, Meissner G, Mitchison DA, Sula L (1963) Mycobacteria: laboratory methods for testing drug susceptibility and resistance. Bull WHO 29:565–578

Canetti G, Fox W, Khomenco A, Mahler HT, Menon NK, Mitchison DA, Rist N, Šmelev NA (1969) Advances in techniques of testing mycobacterial drug sensitivity, and the use of sensitivity tests in tuberculosis control programmes. Bull WHO 41:21–43

Caws M, Drobniewski FA (2001) Molecular techniques in the diagnosis of Mycobacterium tuberculosis and the detection of drug resistance. Ann N Y Acad Sci 953:138–145

Chakravorty S, Simmonds AM, Parmar H, Cao Y et al (2017) The New Xpert MTB/RIF Ultra: Improving Detection of *Mycobacterium tuberculosis* and Resistance to Rifampin in an Assay Suitable for Point-of-Care Testing. mBio 8(4):e00812–e00817

Chien JY, Chiu WY, Chien ST, Chiang CJ, CJ Y, Hsueh PR (2016) Mutations in gyrA and gyrB among fluoroquinolone- and multidrug-resistant mycobacterium tuberculosis isolates. Antimicrob Agents Chemother 60(4):2090–2096

Choi KP, Bair TB, Bae YM, Daniels L (2001) Use of transposon Tn5367 mutagenesis and a nitroimidazopyran-based selection system to demonstrate a requirement for fbiA and fbiB in coenzyme F(420) biosynthesis by Mycobacterium Bovis BCG. J Bacteriol 183(24):7058–7066

Coeck N, de Jong BC, Diels M, de Rijk P, Ardizzoni E, Van Deun A, Rigouts L (2016) Correlation of different phenotypic drug susceptibility testing methods for four fluoroquinolones in mycobacterium tuberculosis. J Antimicrob Chemother 71(5):1233–1240

Coker RJ (2004) Review: multidrug-resistant tuberculosis: public health challenges. Tropical Med Int Health 9(1):25–40

Coll F, McNerney R, Preston MD, Guerra-Assunção JA, Warry A, Hill-Cawthorne G, Mallard K (2015) Nair M4, Miranda a, Alves a, Perdigão J, Viveiros M, Portugal I, Hasan Z, Hasan R, Glynn JR, Martin N, pain a, Clark TG. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. Genome Med 7(1):51

Collins LA, Franzblau SG (1997) Microplate Alamar blue assay versus BACTEC 460 system for high-throughput screening of compounds against *Mycobacterium tuberculosis* and *Mycobacterium avium*. Antimicrob Agents Chemother 41:1004–1009

Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, Galagan J, Niemann S, Gagneux S (2011) Whole-genome sequencing of rifampicin-resistant mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. Nat Genet 44(1):106–110

da Silva PE, Von Groll A, Martin A, Palomino JC (2011) Efflux as a mechanism for drug resistance in mycobacterium tuberculosis. FEMS Immunol Med Microbiol 63(1):1–9

Dartois V, Barry CE 3rd (2013) A medicinal chemists' guide to the unique difficulties of lead optimization for tuberculosis. Bioorg Med Chem Lett 23(17):4741–4750

Dartois V, Saito K, Warrier T, Nathan C (2016) New evidence for the complexity of the population structure of mycobacterium tuberculosis increases the diagnostic and biologic challenges. Am J Respir Crit Care Med 194(12):1448–1451

De Rossi E, Aínsa JA, Riccardi G (2006) Role of mycobacterial efflux transporters in drug resistance: an unresolved question. FEMS Microbiol Rev 30(1):36–52

de Vos M, Müller B, Borrell S, Black PA, van Helden PD, Warren RM, Gagneux S, Victor TC (2013) Putative compensatory mutations in the rpoC gene of rifampin-resistant mycobacterium tuberculosis are associated with ongoing transmission. Antimicrob Agents Chemother 57(2):827–832

Denkinger CM, Schumacher SG, Boehme CC, N D, Pai M, Steingart KR (2014) Xpert MTB/RIF assay for the diagnosis of extrapulmonary tuberculosis: a systematic review and meta-analysis. Eur Respir J 44(2):435–446

Denkinger CM, Dolinger D, Schito M, Wells W, Cobelens F, Pai M, Zignol M, Cirillo DM, Alland D, Casenghi M, Gallarda J, Boehme CC, Perkins MD (2015) Target product profile of a molecular drug-susceptibility test for use in microscopy centers. J Infect Dis 211(Suppl 2):S39–S49

Dowdy DW, Chaisson RE, Maartens G, Corbett EL, Dorman SE (2008) Impact of enhanced tuberculosis diagnosis in South Africa: a mathematical model of expanded culture and drug susceptibility testing. Proc Natl Acad Sci U S A 105(32):11293–11298

Drobniewski F, Cooke M, Jordan J, Casali N, Mugwagwa T, Broda A, Townsend C, Sivaramakrishnan A, Green N, Jit M, Lipman M, Lord J, White PJ, Abubakar I (2015) Systematic review, meta-analysis and economic modelling of molecular diagnostic tests for antibiotic resistance in tuberculosis. Health Technol Assess 19(34):1–188. vii-viii

Emmart EW (1945) The tuberculostatic action of streptothricin and streptomycin with special reference to the action of streptomycin on the chorioallantoic membrane of the chick embryo. Public Health Rep 60:1415–1421

Escribano I, Rodríguez JC, Llorca B, García-Pachon E, Ruiz M, Royo G (2007) Importance of the efflux pump systems in the resistance of Mycobacterium tuberculosis to fluoroquinolones and linezolid. Chemotherapy 53(6):397–401

Farhat MR, Mitnick CD, Franke MF, Kaur D, Sloutsky A, Murray M, Jacobson KR (2015) Concordance of Mycobacterium tuberculosis fluoroquinolone resistance testing: implications for treatment. Int J Tuberc Lung Dis 19(3):339–341

Farhat MR, Jacobson KR, Franke MF, Kaur D, Sloutsky A, Mitnick CD, Murray M (2016a) Gyrase mutations are associated with variable levels of fluoroquinolone

resistance in Mycobacterium tuberculosis. J Clin Microbiol 54(3):727–733

Farhat MR, Sultana R, Iartchouk O, Bozeman S, Galagan J, Sisk P, Stolte C, Nebenzahl-Guimaraes H, Jacobson K, Sloutsky A, Kaur D, Posey J, Kreiswirth BN, Kurepina N, Rigouts L, Streicher EM, Victor TC, Warren RM, van Soolingen D, Murray M (2016b) Genetic determinants of drug resistance in mycobacterium tuberculosis and their diagnostic value. Am J Respir Crit Care Med 194(5):621–630

Feuerriegel S, Köser CU, Baù D, Rüsch-Gerdes S, Summers DK, Archer JA, Marti-Renom MA, Niemann S (2011) Impact of Fgd1 and ddn diversity in Mycobacterium tuberculosis complex on in vitro susceptibility to PA-824. Antimicrob Agents Chemother 55(12):5718–5722

Feuerriegel S, Köser CU, Niemann S (2014) Phylogenetic polymorphisms in antibiotic resistance genes of the Mycobacterium tuberculosis complex. J Antimicrob Chemother 69(5):1205–1210

Feuerriegel S, Schleusener V, Beckert P, Kohl TA, Miotto P, Cirillo DM, Cabibbe AM, Niemann S, Fellenberg K (2015) PhyResSE: a web tool delineating Mycobacterium tuberculosis antibiotic resistance and lineage from whole-genome sequencing data. J Clin Microbiol 53(6):1908–1914

Flandrois JP, Lina G, Dumitrescu O (2014) MUBII-TB-DB: a database of mutations associated with antibiotic resistance in Mycobacterium tuberculosis. BMC Bioinformatics 15:107

García de Viedma D (2003) Rapid detection of resistance in Mycobacterium tuberculosis: a review discussing molecular approaches. Clin Microbiol Infect 9(5):349–359

Georghiou SB, Magana M, Garfein RS, Catanzaro DG, Catanzaro A, Rodwell TC (2012) Evaluation of genetic mutations associated with Mycobacterium tuberculosis resistance to amikacin, kanamycin and capreomycin: a systematic review. PLoS One 7(3):e33275

Gilchrist CA, Turner SD, Riley MF, Petri WA Jr, Hewlett EL (2015) Whole-genome sequencing in outbreak analysis. Clin Microbiol Rev 28(3):541–563

Gillespie SH (2002) Evolution of drug resistance in Mycobacterium tuberculosis: clinical and molecular perspective. Antimicrob Agents Chemother 46(2):267–274

Gilpin C, Korobitsyn A, Weyer K (2016) Current tools available for the diagnosis of drug-resistant tuberculosis. Ther Adv Infectious Dis 3:1–7

Goldstein BP (2014) Resistance to rifampicin: a review. J Antibiot (Tokyo) 67(9):625–630

Guernsey BG, Alexander MR (1978) Tuberculosis: review of treatment failure, relapse and drug resistance. Am J Hosp Pharm 35(6):690–698

Hall L, Jude KP, Clark SL, Dionne K, Merson R, Boyer A, Parrish NM, Wengenack NL (2012) Evaluation of the Sensititre MycoTB plate for susceptibility testing of the Mycobacterium tuberculosis complex against first- and second-line agents. J Clin Microbiol 50(11):3732–3734

Harms A, Maisonneuve E, Gerdes K (2016) Mechanisms of bacterial persistence during stress and antibiotic exposure. Science 354(6318):aaf4268

Hartkoorn RC, Uplekar S, Cole ST (2014) Cross-resistance between clofazimine and bedaquiline through upregulation of MmpL5 in Mycobacterium tuberculosis. Antimicrob Agents Chemother 58(5):2979–2981

He L, Wang X, Cui P, Jin J, Chen J, Zhang W, Zhang Y (2015) ubiA (Rv3806c) encoding DPPR synthase involved in cell wall synthesis is associated with ethambutol resistance in Mycobacterium tuberculosis. Tuberculosis (Edinb) 95(2):149–154

Hillemann D, Rüsch-Gerdes S, Richter E (2008) In vitro-selected linezolid-resistant Mycobacterium tuberculosis mutants. Antimicrob Agents Chemother 52(2):800–801

Hoshide M, Qian L, Rodrigues C, Warren R, Victor T, Evasco HB 2nd, Tupasi T, Crudu V, Douglas JT (2014) Geographical differences associated with single-nucleotide polymorphisms (SNPs) in nine gene targets among resistant clinical isolates of Mycobacterium tuberculosis. J Clin Microbiol 52(5):1322–1329

Huh HJ, Jeong BH, Jeon K, Koh WJ, Ki CS, Lee NY (2014) Performance evaluation of the Xpert MTB/RIF assay according to its clinical application. BMC Infect Dis 14:589

Huitric E, Verhasselt P, Koul A, Andries K, Hoffner S, Andersson DI (2010) Rates and mechanisms of resistance development in Mycobacterium tuberculosis to a novel diarylquinoline ATP synthase inhibitor. Antimicrob Agents Chemother 54(3):1022–1028

Imperiale BR, Di Giulio ÁB, Adrián Cataldi A, Morcillo NS (2014) Evaluation of Mycobacterium tuberculosis cross-resistance to isoniazid, rifampicin and levofloxacin with their respective structural analogs. J Antibiot (Tokyo) 67(11):749–754

Jamieson FB, Guthrie JL, Neemuchwala A, Lastovetska O, Melano RG, Mehaffy C (2014) Profiling of rpoB mutations and MICs for rifampin and rifabutin in Mycobacterium tuberculosis. J Clin Microbiol 52(6):2157–2162

Jiang GL, Chen X, Song Y, Zhao Y, Huang H, Kam KM (2013) First proficiency testing of second-line anti-tuberculosis drug susceptibility testing in 12 provinces of China. Int J Tuberc Lung Dis 17(11):1491–1494

Jo KW, Lee SD, Kim WS, Kim DS, Shim TS (2014) Treatment outcomes and moxifloxacin susceptibility in ofloxacin-resistant multidrug-resistant tuberculosis. Int J Tuberc Lung Dis 18(1):39–43

Johansen SK, Maus CE, Plikaytis BB, Douthwaite S (2006) Capreomycin binds across the ribosomal subunit interface using tlyA-encoded 2′-O-methylations in 16S and23S rRNAs. Mol Cell 23(2):173–182

Jones D, Metzger HJ, Schatz A, Waksman SA (1944) Control of gram-negative bacteria in experimental animals by streptomycin. Science 100(2588):103–105

Kam KM, Yip CW, Cheung TL, Tang HS, Leung OC, Chan MY (2006) Stepwise decrease in moxifloxacin

susceptibility amongst clinical isolates of multidrug-resistant Mycobacterium tuberculosis: correlation with ofloxacin susceptibility. Microb Drug Resist 12(1): 7–11

Kambli P, Ajbani K, Sadani M, Nikam C, Shetty A, Udwadia Z, Rodwell TC, Catanzaro A, Rodrigues C (2015) Correlating minimum inhibitory concentrations of ofloxacin and moxifloxacin with gyrA mutations using the genotype MTBDRsl assay. Tuberculosis (Edinb) 95(2):137–141

Kambli P, Ajbani K, Nikam C, Sadani M, Shetty A, Udwadia Z, Georghiou SB, Rodwell TC, Catanzaro A, Rodrigues C (2016) Correlating rrs and eis promoter mutations in clinical isolates of Mycobacterium tuberculosis with phenotypic susceptibility levels to the second-line injectables. Int J Mycobacteriol 5(1):1–6

Kelley CL, Rouse DA, Morris SL (1997) Analysis of ahpC gene mutations in isoniazid-resistant clinical isolates of Mycobacterium tuberculosis. Antimicrob Agents Chemother 41(9):2057–2058

Kent PT, Kubica GP (1985) Antituberculosis chemotherapy and drug susceptibility testing. In: Kent PT, Kubica GP (eds) Public health mycobacteriology. A guide for the level III laboratory. U.S. Department of Health and Human Services, Atlanta, pp 159–184

Kik SV, Denkinger CM, Casenghi M, Vadnais C, Pai M (2014) Tuberculosis diagnostics: which target product profiles should be prioritised? Eur Respir J 44(2):537–540

Kim SJ (2005) Drug-susceptibility testing in tuberculosis: methods and reliability of results. Eur Respir J 25(3):564–569

Köser CU, Bryant JM, Becq J, Török ME, Ellington MJ, Marti-Renom MA, Carmichael AJ, Parkhill J, Smith GP, Peacock SJ (2013) Whole-genome sequencing for rapid susceptibility testing of M. tuberculosis. N Engl J Med 369(3):290–292. https://doi.org/10.1056/NEJMc1215305. No abstract available

Kronvall G (2010) Normalized resistance interpretation as a tool for establishing epidemiological MIC susceptibility breakpoints. J Clin Microbiol 48(12):4445–4452

Lee AS, Teo AS, Wong SY (2001) Novel mutations in ndh in isoniazid-resistant Mycobacterium tuberculosis isolates. Antimicrob Agents Chemother 45(7):2157–2159

Lemaitre N, Sougakoff W, Truffot-Pernot C, Jarlier V (1999) Characterization of new mutations in pyrazinamide-resistant strains of Mycobacterium tuberculosis and identification of conserved regions important for the catalytic activity of the pyrazinamidase PncA. Antimicrob Agents Chemother 43(7):1761–1763

Ling DI, Zwerling AA, Pai M (2008) GenoType MTBDR assays for the diagnosis of multidrug-resistant tuberculosis: a meta-analysis. Eur Respir J 32(5):1165–1174

Lingaraju S, Rigouts L, Gupta A, Lee J, Umubyeyi AN, Davidow AL, German S, Cho E, Lee JI, Cho SN, Kim CT, Alland D, Safi H (2016) Geographic differences in the contribution of ubiA mutations to high-level Ethambutol Resistance in Mycobacterium tuberculosis. Antimicrob Agents Chemother. pii: AAC.03002-15. [Epub ahead of print]

Lippincott CK, Miller MB, Van Rie A, Weber DJ, Sena AC, Stout JE (2015) The complexities of Xpert® MTB/RIF interpretation. Int J Tuberc Lung Dis 19(3):273–275

Louw GE, Warren RM, Gey van Pittius NC, McEvoy CR, Van Helden PD, Victor TC (2009) A balancing act: efflux/influx in mycobacterial drug resistance. Antimicrob Agents Chemother 53(8):3181–3189

Madison B, Robinson-Dunn B, George I et al (2002) Multicenter evaluation of ethambutol susceptibility testing of Mycobacterium tuberculosis by agar proportion and radiometric methods. J Clin Microbiol 40:3976–3979

Makafe GG, Cao Y, Tan Y, Julius M, Liu Z, Wang C, Njire MM, Cai X, Liu T, Wang B, Pang W, Tan S, Zhang B, Yew WW, Lamichhane G, Guo J, Zhang T (2016) Role of the Cys154Arg substitution in ribosomal protein L3 in Oxazolidinone resistance in Mycobacterium tuberculosis. Antimicrob Agents Chemother 60(5):3202–3206

Malik S, Willby M, Sikes D, Tsodikov OV, Posey JE (2012) New insights into fluoroquinolone resistance in Mycobacterium tuberculosis: functional genetic analysis of gyrA and gyrB mutations. PLoS One 7(6): e39754

Manjunatha UH, Boshoff H, Dowd CS, Zhang L, Albert TJ, Norton JE, Daniels L, Dick T, Pang SS, Barry CE 3rd (2006) Identification of a nitroimidazo-oxazine-specific protein involved in PA-824 resistance in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 103(2):431–436

Martínez JL, Coque TM, Baquero F (2015) What is a resistance gene? Ranking risk in resistomes. Nat Rev Microbiol 13(2):116–123

Marttila HJ, Soini H (2003) Molecular detection of resistance to antituberculous therapy. Clin Lab Med 23(4):823–841. v-vi

Matsumoto M, Hashizume H, Tomishige T, Kawasaki M, Tsubouchi H, Sasaki H, Shimokawa Y, Komatsu M (2006) OPC-67683, a nitro-dihydro-imidazooxazole derivative with promising action against tuberculosis in vitro and in mice. PLoS Med 3(11):e466

Maus CE, Plikaytis BB, Shinnick TM (2005) Molecular analysis of cross-resistance to capreomycin, kanamycin, amikacin, and viomycin in Mycobacterium tuberculosis. Antimicrob Agents Chemother 49(8):3192–3197

Miotto P, Cabibbe AM, Feuerriegel S, Casali N, Drobniewski F, Rodionova Y, Bakonyte D, Stakenas P, Pimkina E, Augustynowicz-Kopeć E, Degano M, Ambrosi A, Hoffner S, Mansjö M, Werngren J, Rüsch-Gerdes S, Niemann S, Cirillo DM (2014) Mycobacterium tuberculosis pyrazinamide resistance determinants: a multicenter study. MBio 5(5):e01819-14

Miotto P, Cirillo DM, Migliori GB (2015) Drug resistance in Mycobacterium tuberculosis: molecular mechanisms challenging fluoroquinolones and pyrazinamide effectiveness. Chest 147(4):1135–1143

Moore DA, Mendoza D, Gilman RH, Evans CA, Hollm Delgado MG, Guerra J, Caviedes L, Vargas D, Ticona E, Ortiz J, Soto G, Serpa J (2004) Microscopic observation drug susceptibility assay, a rapid, reliable diagnostic test for multidrug-resistant tuberculosis suitable for use in resource-poor settings. J.Clin.Microbiol. 42:4432–4437

Morgan M, Kalantri S, Flores L, Pai M (2005) A commercial line probe assay for the rapid detection of rifampicin resistance in Mycobacterium tuberculosis: a systematic review and meta-analysis. BMC Infect Dis 5:62

Nathan C (2012) Fresh approaches to anti-infective therapies. Sci Transl Med 4(140):140sr2

Nathavitharana RR, Hillemann D, Schumacher SG, Schlueter B, Ismail N, Vally Omar S, Sikhondze W, Havumaki J, Valli E, Boehme C, Denkinger CM (2016) A Multi-Center Non-inferiority Evaluation of Hain GenoType MTBDRplus Version 2 and Nipro NTM+MDRTB lineprobe assays for the diagnosis of Rifampin and Isoniazid Resistance. J Clin Microbiol. pii: JCM.00251-16. [Epub ahead of print]

Nosova EY, Bukatina AA, Isaeva YD, Makarova MV, Galkina KY, Moroz AM (2013) Analysis of mutations in the gyrA and gyrB genes and their association with the resistance of Mycobacteriumtuberculosis to levofloxacin, moxifloxacin and gatifloxacin. J Med Microbiol 62(Pt 1):108–113

Okamoto S, Tamaru A, Nakajima C, Nishimura K, Tanaka Y, Tokuyama S, Suzuki Y, Ochi K (2007) Loss of a conserved 7-methylguanosine modification in 16S rRNA confers low-level streptomycin resistance in bacteria. Mol Microbiol 63(4):1096–1106

Pandey B, Grover S, Tyagi C, Goyal S, Jamal S, Singh A, Kaur J, Grover A (2016) Molecular principles behind pyrazinamide resistance due to mutations in panD gene in Mycobacterium tuberculosis. Gene 581(1):31–42

Pankhurst LJ, del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies J, Fermont JM, Gascoyne-Binzi DM, Kohl T, Kong C, Lemaitre N, Niemann S, Paul J, Rogers TR, Roycroft E, Smith G, Supply P, Tang P, Wilcox MH, Wordsworth S, Wyllie D, Xu L, Crook W, Rapid DW (2016) Comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. Lancet Respir Med 4(1):49–58. https://doi.org/10.1016/S2213-2600(15)00466-X

Parsons LM, Somoskövi A, Gutierrez C, Lee E, Paramasivan CN, Abimiku A, Spector S, Roscigno G, Nkengasong J (2011) Laboratory diagnosis of tuberculosis in resource-poor countries: challenges and opportunities. Clin Microbiol Rev 24(2):314–350

Patel RJ, Fries JW, Piessens WF, Wirth DF (1990) Sequence analysis and amplification by polymerase chain reaction of a cloned DNA fragment for identification of Mycobacterium tuberculosis. J Clin Microbiol 28(3):513–518

Piersimoni C, Mustazzolu A, Giannoni F, Bornigia S, Gherardi G, Fattorini L (2013) Prevention of false resistance results obtained in testing the susceptibility of Mycobacterium tuberculosis to pyrazinamide with the Bactec MGIT 960 system using a reduced inoculum. J Clin Microbiol 51(1):291–294

Punina NV, Makridakis NM, Remnev MA, Topunov AF (2015) Whole-genome sequencing targets drug-resistant bacterial infections. Hum Genomics 9:19

Ramaswamy S, Musser JM (1998) Molecular genetic basis of antimicrobial agent resistance in Mycobacterium tuberculosis: 1998 update. Tuber Lung Dis 79(1): 3–29

Ramirez-Busby SM, Valafar F (2015) Systematic review of mutations in pyrazinamidase associated with pyrazinamide resistance in Mycobacterium tuberculosis clinical isolates. Antimicrob Agents Chemother 59(9):5267–5277

Reeves AZ, Campbell PJ, Sultana R, Malik S, Murray M, Plikaytis BB, Shinnick TM, Posey JE (2013) Aminoglycoside cross-resistance in Mycobacterium tuberculosis due to mutations in the 5′ untranslated region of whiB7. Antimicrob Agents Chemother 57(4):1857–1865

Rigouts L, Gumusboga M, de Rijk WB, Nduwamahoro E, Uwizeye C, de Jong B, Van Deun A (2013) Rifampin resistance missed in automated liquid culture system for Mycobacterium tuberculosis isolates with specific rpoB mutations. J Clin Microbiol 51(8):2641–2645

Rigouts L, Coeck N, Gumusboga M, de Rijk WB, Aung KJ, Hossain MA, Fissette K, Rieder HL, Meehan CJ, de Jong BC, Van Deun A (2016) Specific gyrA gene mutations predict poor treatment outcome in MDR-TB. J Antimicrob Chemother 71(2):314–323

Rueda J, Realpe T, Mejia GI, Zapata E, Rozo JC, Ferro BE, Robledo J (2015) Genotypic analysis of genes associated with independent resistance and cross-resistance toIsoniazid and Ethionamide in Mycobacterium tuberculosis clinical isolates. Antimicrob Agents Chemother 59(12):7805–7810

Rufai SB, Kumar P, Singh A, Prajapati S, Balooni V, Singh S (2014) Comparison of Xpert MTB/RIF with line probe assay for detection of rifampin-monoresistant Mycobacterium tuberculosis. J Clin Microbiol 52(6):1846–1852. https://doi.org/10.1128/JCM.03005-13. Epub 2014 Mar 19

Rüsch-Gerdes S, Domehl C, Nardi G, Gismondo MR, Welscher HM, Pfyffer GE (1999) Multicenter evaluation of the mycobacteria growth indicator tube for testing susceptibility of Mycobacterium tuberculosis to first-line drugs. J Clin Microbiol 37(1):45–48

Safi H, Lingaraju S, Amin A, Kim S, Jones M, Holmes M, McNeil M, Peterson SN, Chatterjee D, Fleischmann R, Alland D (2013) Evolution of high-level ethambutol-resistant tuberculosis through interacting mutations in decaprenylphosphoryl-β-D-arabinose biosynthetic and utilization pathway genes. Nat Genet 45(10):1190–1197

Sanchez-Padilla E, Merker M, Beckert P, Jochims F, Dlamini T, Kahn P, Bonnet M, Niemann S (2015) Detection of drug-resistant tuberculosis by Xpert MTB/RIF in Swaziland. N Engl J Med 372(12):1181–1182

Saraav I, Pandey K, Misra R, Singh S, Sharma M, Sharma S (2017) Characterization of MymA protein as a flavin-containing monooxygenase and as a target of isoniazid. Chem Biol Drug Des 89(1):152–160

Schena E, Nedialkova L, Borroni E, Battaglia S, Cabibbe AM, Niemann S, Utpatel C, Merker M, Trovato A, Hofmann-Thiel S, Hoffmann H, Cirillo DM (2016) Delamanid susceptibility testing of Mycobacterium tuberculosis using the resazurin microtitre assay and the BACTEC™ MGIT™ 960 system. J Antimicrob Chemother 71(6):1532–1539. https://doi.org/10.1093/jac/dkw044

Schito M, Dolinger DL (2015) A collaborative approach for "ReSeq-ing" Mycobacterium tuberculosis drug resistance: convergence for drug and diagnostic developers. EBioMedicine 2(10):1262–1265

Segala E, Sougakoff W, Nevejans-Chauffour A, Jarlier V, Petrella S (2012) New mutations in the mycobacterial ATP synthase: new insights into the binding of the diarylquinoline TMC207 to the ATP synthase C-ring structure. Antimicrob Agents Chemother 56(5):2326–2334

Seifert M, Catanzaro D, Catanzaro A, Rodwell TC (2015) Genetic mutations associated with isoniazid resistance in Mycobacterium tuberculosis: a systematic review. PLoS One 10(3):e0119628

Seifert M, Georghiou SB, Catanzaro D, Rodrigues C, Crudu V, Victor TC, Garfein RS, Catanzaro A, Rodwell TC (2016) MTBDRplus and MTBDRsl assays: absence of wild-type probe hybridization and implications for detection of drug-resistant tuberculosis. J Clin Microbiol 54(4):912–918

Singh P, Jain A, Dixit P, Prakash S, Jaiswal I, Venkatesh V, Singh M (2015) Prevalence of gyrA and B gene mutations in fluoroquinolone-resistant and -sensitive clinical isolates of Mycobacterium tuberculosis and their relationship with MIC of ofloxacin. J Antibiot (Tokyo) 68(1):63–66

Sintchenko V, Chew WK, Jelfs PJ, Gilbert GL (1999) Mutations in rpoB gene and rifabutin susceptibility of multidrug-resistant Mycobacterium tuberculosis strains isolated in Australia. Pathology 31(3):257–260

Slayden RA, Barry CE 3rd (2000) The genetics and biochemistry of isoniazid resistance in mycobacterium tuberculosis. Microbes Infect 2(6):659–669

Smith DG, Waskman SA (1947) Tuberculostatic and tuberculocidal action of streptomycin. J Bacteriol 54(1):67

Smith T, Wolff KA, Nguyen L (2013) Molecular biology of drug resistance in *Mycobacterium tuberculosis*. Curr Top Microbiol Immunol 374:53–80

Somoskovi A, Salfinger M (2015) The race is on to shorten the turnaround time for diagnosis of multidrug-resistant tuberculosis. J Clin Microbiol 53(12):3715–3718

Sreevatsan S, Pan X, Stockbauer KE, Williams DL, Kreiswirth BN, Musser JM (1996) Characterization of rpsL and rrs mutations in streptomycin-resistant Mycobacterium tuberculosis isolates from diverse geographic localities. Antimicrob Agents Chemother 40(4):1024–1026

Sreevatsan S, Pan X, Zhang Y, Deretic V, Musser JM (1997) Analysis of the oxyR-ahpC region in isoniazid-resistant and -susceptible Mycobacterium tuberculosis complex organisms recovered from diseased humans and animals in diverse localities. Antimicrob Agents Chemother 41(3):600–606

Starks AM, Avilés E, Cirillo DM, Denkinger CM, Dolinger DL, Emerson C, Gallarda J, Hanna D, Kim PS, Liwski R, Miotto P, Schito M, Zignol M (2015) Collaborative effort for a centralized worldwide tuberculosis relational sequencing data platform. Clin Infect Dis 61(Suppl 3):S141–S146

Steiner A, Stucki D, Coscolla M, Borrell S, Gagneux S (2014) KvarQ: targeted and direct variant calling from fastq reads of bacterial genomes. BMC Genomics 15:881

Steingart KR, Schiller I, Horne DJ, Pai M, Boehme CC, Dendukuri N (2014) Xpert® MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. Cochrane Database Syst Rev 1:CD009593

Stover CK, Warrener P, VanDevanter DR, Sherman DR, Arain TM, Langhorne MH, Anderson SW, Towell JA, Yuan Y, McMurray DN, Kreiswirth BN, Barry CE, Baker WRA (2000) Small-molecule nitroimidazopyran drug candidate for the treatment of tuberculosis. Nature 405(6789):962–966

Tagliani E, Cabibbe AM, Miotto P, Borroni E, Toro JC, Mansjö M, Hoffner S, Hillemann D, Zalutskaya A, Skrahina A, Cirillo DM (2015) Diagnostic performance of the new version (v2.0) of GenoType MTBDRsl assay for detection of resistance to fluoroquinolones and second-line injectable drugs: a multicenter study. J Clin Microbiol 53(9):2961–2969

Telenti A, Imboden P, Marchesi F, Lowrie D, Cole S, Colston MJ, Matter L, Schopfer K, Bodmer T (1993) Detection of rifampicin-resistance mutations in Mycobacterium tuberculosis. Lancet 341(8846):647–650

Telenti A, Philipp WJ, Sreevatsan S, Bernasconi C, Stockbauer KE, Wieles B, Musser JM, Jacobs WR Jr (1997) The emb operon, a gene cluster of Mycobacterium tuberculosis involved in resistance to ethambutol. Nat Med 3(5):567–570

Theron G, Peter J, Richardson M, Barnard M, Donegan S, Warren R, Steingart KR, Dheda K (2014) The diagnostic accuracy of the GenoType® MTBDRsl assay for the detection of resistance to second-line anti-tuberculosis drugs. Cochrane Database Syst Rev 10:CD010705

Torrea G, Coeck N, Desmaretz C, Van De Parre T, Van Poucke T, Lounis N, de Jong BC, Rigouts L (2015) Bedaquiline susceptibility testing of Mycobacterium tuberculosis in an automated liquid culture system. J Antimicrob Chemother 70(8):2300–2305

Tortoli E, Benedetti M, Fontanelli A, Simonetti MT (2002) Evaluation of automated BACTEC MGIT 960 system for testing susceptibility of *Mycobacterium tuberculosis* to four major antituberculous drugs: comparison with the radiometric BACTEC 460TB method

and the agar plate method of proportion. JClinMicrobiol 40:607–610

Uys PW, Warren R, van Helden PD, Murray M, Victor TC (2009) Potential of rapid diagnosis for controlling drug-susceptible and drug-resistant tuberculosis in communities where Mycobacterium tuberculosis infections are highly prevalent. J Clin Microbiol 47(5):1484–1490

Van Deun A, Aung KJ, Hossain A, de Rijk P, Gumusboga M, Rigouts L, de Jong BC (2015) Disputed rpoB mutations can frequently cause important rifampicin resistance among newtuberculosis patients. Int J Tuberc Lung Dis 19(2):185–190

van Doorn HR, Claas EC, Templeton KE, van der Zanden AG, te Koppele Vije A, de Jong MD, Dankert J, Kuijper EJ (2003) Detection of a point mutation associated with high-level isoniazid resistance in Mycobacterium tuberculosis by using real-time PCR technology with 3′-minor groove binder-DNA probes. J Clin Microbiol 41(10):4630–4635

Vilchèze C, Weisbrod TR, Chen B, Kremer L, Hazbón MH, Wang F, Alland D, Sacchettini JC, Jacobs WR Jr (2005) Altered NADH/NAD+ ratio mediates coresistance to isoniazid and ethionamide in mycobacteria. Antimicrob Agents Chemother 49(2):708–720

Wakamoto Y, Dhar N, Chait R, Schneider K, Signorino-Gelo F, Leibler S, McKinney JD (2013) Dynamic persistence of antibiotic-stressed mycobacteria. Science 339(6115):91–95

Walker T, Kohl T, Omar S, Hedge J, Del Ojo E, Bradley P, Iqbal Z, Feuerriegel S, Niehaus K, Wilson D, Clifton D, Kapatai G, Ip C, Bowden C, Drobniewski F, Allix-Béguec C, Gaudin C, Parkhill J, Diel R, Supply P, Crook D, Smith G, Walker S, Ismail N, Niemann S, Peto T (2015) Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. The Lancet Infectious Diseases 15:1193. https://doi.org/10.1016/S1473-3099(15)00062-6

Wells WA, Boehme CC, Cobelens FG, Daniels C, Dowdy D, Gardiner E, Gheuens J, Kim P, Kimerling ME, Kreiswirth B, Lienhardt C, Mdluli K, Pai M, Perkins MD, Peter T, Zignol M, Zumla A, Schito M (2013) Alignment of new tuberculosis drug regimens and drug susceptibility testing: a framework for action. Lancet Infect Dis 13(5):449–458

Whitfield MG, Soeters HM, Warren RM, York T, Sampson SL, Streicher EM, van Helden PD, van Rie A (2015) A global perspective on pyrazinamide resistance: systematic review and meta-analysis. PLoS One 10(7):e0133869

Willby M, Sikes RD, Malik S, Metchock B, Posey JE (2015) Correlation between GyrA substitutions and ofloxacin, levofloxacin, and moxifloxacin cross-resistance in Mycobacterium tuberculosis. Antimicrob Agents Chemother 59(9):5427–5434

Williamson DA, Basu I, Bower J, Freeman JT, Henderson G, Roberts SA (2012) An evaluation of the Xpert MTB/RIF assay and detection of false-positive rifampicin resistance in Mycobacterium tuberculosis. Diagn Microbiol Infect Dis 74(2):207–209. https://doi.org/10.1016/j.diagmicrobio.2012.06.013. Epub 2012 Jul 20

Witney AA, Gould KA, Arnold A, Coleman D, Delgado R, Dhillon J, Pond MJ, Pope CF, Planche TD, Stoker NG, Cosgrove CA, Butcher PD, Harrison TS, Hinds J (2015) Clinical application of whole-genome sequencing to inform treatment for multidrug-resistant tuberculosis cases. J Clin Microbiol 53(5):1473–1483

Wong SY, Lee JS, Kwak HK, Via LE, Boshoff HI, Barry CE 3rd (2011) Mutations in gidB confer low-level streptomycin resistance in Mycobacterium tuberculosis. Antimicrob Agents Chemother 55(6):2515–2522

World Health Organization (2004) Anti-tuberculosis drug resistance in the world, Third global report. WHO, Geneva. ISBN 92 4 156285 4. WHO/HTM/TB/2004.343

World Health Organization (2008) Molecular line probe assays for rapid screening of patients at risk of multidrug resistant tuberculosis (MDR-TB), Expert group report. WHO, Geneva

World Health Organization (2012a) Tuberculosis laboratory biosafety manual. WHO, Geneva. ISBN 978 92 4 150463 8

World Health Organization (2012b) Summary of outcomes from WHO expert group meeting on drug susceptibility testing. 4th annual GLI meeting 17 April 2012

World Health Organization (2013) Automated real-time nucleic acid amplification technology for rapid and simultaneous detection of tuberculosis and rifampicin resistance: Xpert MTB/RIF assay for the: diagnosis of pulmonary and extrapulmonary TB in adults and children, Policy update. WHO, Geneva. ISBN: 978 92 4 150633 5. WHO/HTM/TB/2013.16

World Health Organization, UNITAID (2015) Tuberculosis. Diagnostics technology and market land-scape, 4th ed. Geneva

World Health Organization (2016) The use of molecular line probe assays for the detection of resistance to second-line antituberculosis drugs, Policy guidance. WHO, Geneva. ISBN 978 92 4 150963 3

World Health Organization, UNITAID (2015) Tuberculosis diagnostics technology and market landscape, 4th edn. WHO, Geneva

Xu Y, Jia H, Huang H, Sun Z, Zhang Z (2015) Mutations found in embCAB, embR, and ubiA genes of ethambutol-sensitive and -resistant Mycobacterium tuberculosis clinical isolates from China. Biomed Res Int 2015:951706

Xu J, Wang B, Hu M, Huo F, Guo S, Jing W, Nuermberger E, Lu Y (2017) Primary clofazimine and bedaquiline resistance among isolates from patients with multidrug-resistant tuberculosis. Antimicrob Agents Chemother 61(6):pii: e00239-17. doi:https://doi.org/10.1128/AAC.00239-17. Print 2017 Jun

Yang J, Liu Y, Bi J, Cai Q, Liao X, Li W, Guo C, Zhang Q, Lin T, Zhao Y, Wang H, Liu J, Zhang X, Lin D (2015) Structural basis for targeting the ribosomal protein S1

of Mycobacterium tuberculosis by pyrazinamide. Mol Microbiol 95(5):791–803

Youmans GP, Williston EH et al (1946) Increase in resistance of tubercle bacilli to streptomycin; a preliminary report. Proc Staff Meet Mayo Clin 21:126

Yu X, Ma YF, Jiang GL, Chen ST, Wang GR, Huang HR (2016) Sensititre® MYCOTB MIC plate for drug susceptibility testing of Mycobacterium tuberculosis complex isolates. Int J Tuberc Lung Dis 20(3):329–334

Zaunbrecher MA, Sikes RD Jr, Metchock B, Shinnick TM, Posey JE (2009) Overexpression of the chromosomally encoded aminoglycoside acetyltransferase eis confers kanamycinresistance in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 106(47):20004–20009

Zetola NM, Shin SS, Tumedi KA, Moeti K, Ncube R, Nicol M, Collman RG, Klausner JD, Modongo C (2014) Mixed Mycobacterium tuberculosis complex infections and false-negative results for rifampin resistance by GeneXpert MTB/RIF are associated with poor clinical outcomes. J Clin Microbiol 52(7):2422–2429. https://doi.org/10.1128/JCM.02489-13. Epub 2014 Apr 30

Zhang Y, Mitchison D (2003) The curious characteristics of pyrazinamide: a review. Int J Tuberc Lung Dis 7(1):6–21

Zhang Y, Yew WW (2015) Mechanisms of drug resistance in Mycobacterium tuberculosis: update 2015. Int J Tuberc Lung Dis 19(11):1276–1289

Zignol M, Dean AS, Alikhanova N, Andres S, Cabibbe AM, Cirillo DM et al (2016) Population-based resistance of Mycobacterium tuberculosis isolates to pyrazinamide and fluoroquinolones: results from a multicountry surveillance project. Lancet Infect Dis 16(10):1185–1192. https://doi.org/10.1016/S1473-3099(16)30190-6

DNA Replication Fidelity in the *Mycobacterium tuberculosis* Complex

13

Digby F. Warner, Jeremy M. Rock, Sarah M. Fortune, and Valerie Mizrahi

Abstract

Mycobacterium tuberculosis is genetically isolated, with no evidence for horizontal gene transfer or the acquisition of episomal genetic information in the modern evolution of strains of the *Mycobacterium tuberculosis* complex. When considered in the context of the specific features of the disease *M. tuberculosis* causes (e.g., transmission via cough aerosol, replication within professional phagocytes, subclinical persistence, and stimulation of a destructive immune pathology), this implies that to understand the mechanisms ensuring preservation of genomic integrity in infecting mycobacterial populations is to understand the source of genetic variation, including the emergence of microdiverse sub-populations that may be linked to the acquisition of drug resistance. In this chapter, we focus on mechanisms involved in maintaining DNA replication fidelity in *M. tuberculosis*, and consider the potential to target components of the

D.F. Warner • V. Mizrahi (✉)
MRC/NHLS/UCT Molecular Mycobacteriology Research Unit and DST/NRF Centre of Excellence for Biomedical TB Research, Department of Pathology and Institute of Infectious Disease and Molecular Medicine, University of Cape Town, P/Bag X3, Rondebosch, 7700, South Africa
e-mail: digby.warner@uct.ac.za; Valerie.mizrahi@uct.ac.za

J.M. Rock
Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA, USA
e-mail: rock@hsph.harvard.edu

S.M. Fortune
Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA, USA

Broad Institute of MIT and Harvard, Cambridge, MA, USA

Ragon Institute of Massachusetts General Hospital, Harvard and Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: sfortune@hsph.harvard.edu

S. Gagneux (ed.), *Strain Variation in the* Mycobacterium tuberculosis *Complex: Its Role in Biology, Epidemiology and Control*, Advances in Experimental Medicine and Biology 1019,
DOI 10.1007/978-3-319-64371-7_13

DNA replication machinery as part of novel therapeutic regimens designed to curb the emerging threat of drug-resistance.

Keywords
Mutation rate • DNA polymerase • PHP domain • Drug resistance

13.1 Introduction

Horizontal gene transfer (HGT) was essential to the emergence of *M. tuberculosis* as a human-adapted pathogen (Boritsch et al. 2014). In contrast, and as noted elsewhere in this book (see Chaps. 3 and 5), all available evidence suggests a very limited role (if any) for HGT in the modern evolution of members of the *Mycobacterium tuberculosis* complex (Galagan 2014). Instead, comparative genomics indicate that the *M. tuberculosis* genome is shaped almost exclusively by chromosomal rearrangements and mutations, a conclusion consistent with the ecological isolation of the tubercle bacillus, as well as the regular bottlenecks encountered during the transmission events which are essential to the ability of the organism to persist within the human population (Hershberg et al. 2008). As an intracellular pathogen that preferentially resides within cells of the host immune system, *M. tuberculosis* is exposed to genotoxic environments (Darwin and Nathan 2005; Dutta et al. 2010; Sassetti and Rubin 2003), requiring the activity of specialist DNA repair systems throughout infection (Gorna et al. 2010). Moreover, and as noted previously (Warner et al. 2014), the ability of *M. tuberculosis* bacilli to persist within the human host while retaining the ability to reactivate decades later (Lillebaek et al. 2002), suggests a critical role for DNA metabolic pathways in the maintenance of bacterial viability throughout extended infection. In combination, these factors – genetic isolation, exposure to genotoxic stress, and long-term persistence – identify the processes of DNA replication, repair, and recombination (the so-called "3Rs") as integral components of the metabolic repertoire which defines mycobacterial pathogenesis and evolution. The comparatively high levels of polymorphisms in 3R genes (Dos Vultos et al. 2008) has implicated these genes in the evolution of *M. tuberculosis* by balancing the need to maintain genome stability with the need to adapt to the fluctuating and hostile environments encountered during host infection. It may seem surprising, therefore, that very little is known about the specific contributions of each of the 3R systems, and the constituent components thereof, to the maintenance of genomic integrity on the one hand, and the fixation of mutations which are essential to microdiversity (Warner et al. 2015) and adaptive evolution, including the emergence of drug resistance (McGrath et al. 2014), on the other.

In this chapter, we will concentrate specifically on replicative fidelity, highlighting the central role of the intrinsic proofreading activity associated with the Polymerase and Histidinol Phosphatase (PHP) domain of the major replicative polymerase DnaE1, and the roles of other DNA polymerases in genome replication and maintenance in *M. tuberculosis*; the roles of DNA repair and recombination in mycobacterial genome dynamics, pathogenesis, and evolution have been extensively reviewed elsewhere (McGrath et al. 2014; Warner et al. 2014), and will not be discussed here. Rather, we will provide a brief overview of recent results pertaining to potential contributions of the expanded complement of DNA polymerases in the evolution of *M. tuberculosis* (Fig. 13.1). In addition, we review the potential strategies to target specific 3R pathways as part of novel therapeutic interventions against tuberculosis.

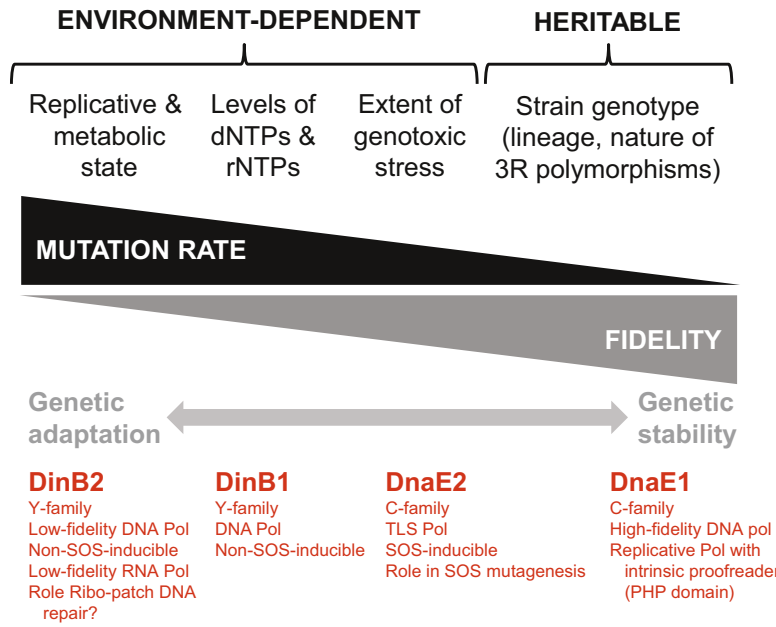

Fig. 13.1 Factors impacting upon replication fidelity in *M. tuberculosis*, including the DNA polymerases reviewed in this chapter

13.2 The DNA Polymerase Complement in Mycobacteria

The *M. tuberculosis* genome encodes eight DNA polymerases: the C-family DNA polymerases, DnaE1 (Rv1547) (Rock et al. 2015; Gu et al. 2016) and DnaE2 (Rv3370c) (Boshoff et al. 2003; Warner et al. 2010); the A-family polymerase, PolI (PolA; Rv1629) (Gordhan et al. 1996; Mizrahi and Huberts 1996); three AEP family polymerases, LigD-POL (Rv0938) (Zhu et al. 2006), PolD1 (Rv3730c) and PolD2 (Rv0269c) (Brissett et al. 2007; Zhu et al. 2012); and two Y-family polymerases, DinB1 (Rv1537) and DinB2 (Rv3056) (Kana et al. 2010; Ghosh et al. 2015; Sharma and Nair 2012). In addition to these *bona fide* DNA polymerases, *M. tuberculosis* also possesses a cryptic Y-family DNA polymerase, ImuB (Rv3394c), which lacks the active site acidic residues necessary for catalysis of polymerase activity. ImuB is one of the components of the mycobacterial "mutasome", which also comprises DnaE2 (Boshoff et al. 2003) and ImuA' (Rv3395c) and is involved in DNA damage tolerance and damage-induced mutagenesis in *M. tuberculosis* and *M. smegmatis* (Warner et al. 2010).

The DNA polymerases of *M. tuberculosis* and/or their counterparts from *M. smegmatis* have been the subject of intense investigation in several laboratories. Recombinant forms of some of these proteins have been produced and enzymatically characterised in terms of DNA polymerase activity, replication fidelity, and also, in certain cases, 3′–5′ exonuclease and/or RNA polymerase activity. For some of the polymerases, biochemical studies have been coupled with genetic approaches which have sought to elucidate the role of individual polymerases or polymerase families in mycobacterial DNA metabolism through phenotypic characterization of mutant strains of *M. tuberculosis* and/or *M. smegmatis* under various experimental conditions.

13.3 The Mycobacterial Replication Machinery

In bacteria, replication of chromosomal DNA is mediated by a large, multi-protein complex, which interacts with a number of additional accessory factors (Miggiano et al. 2013). As with many aspects of bacterial DNA replication

and repair, the composition and operation of the replication machinery has been most thoroughly characterized in *E. coli* (Farhat et al. 2013) and *B. subtilis* (Colangeli et al. 2014). These bacterial systems have informed the model "replisome" comprising two major sub-complexes: firstly, a primosome that consists of a replicative helicase, DnaB, that unwinds the double-stranded DNA, and a primase which generates short RNA primers during discontinuous synthesis of the lagging-strand (Miggiano et al. 2013); and, secondly, a DNA polymerase III (Pol III) holoenzyme, which includes a Pol III core (itself made up of the α catalytic subunit, ε proofreading subunit and its stabilizer, θ), the β_2-sliding clamp, and the clamp-loader complex (Miggiano et al. 2013). Like many other bacteria, *M. tuberculosis* lacks some components of the *E. coli* system (Dos Vultos et al. 2008), and is predicted to function instead with a reduced replication module consisting of a *dnaA*-encoded replication initiator protein (Rv0001), DnaB helicase (Rv0058), DnaG primase (Rv2343c), Pol IIIα subunit (DnaE1), β_2 sliding clamp (DnaN, Rv0002), ε proofreading subunit (DnaQ, Rv3711c), τ (DnaZX, Rv3721c), δ (HolA, Rv2413c) and δ' (HolB, Rv3644c), single-strand DNA (ssDNA)-binding protein (Ssb, Rv0054), DNA ligase (LigA, Rv3014c), and Pol I (PolA, Rv1629). To date, no one has fully reconstituted the mycobacterial replication machinery in vitro; however, as described below, two recent studies (Gu et al. 2016; Rock et al. 2015) of the composition and function of specific replisome components provide compelling evidence of the potential to add *M. tuberculosis* to the list of model organisms, for which detailed structural and biochemical information about the replisome are known.

13.4 Fidelity of the Major Replicative DNA Polymerase

According to the well-established *E. coli* model, DNA replication fidelity in bacteria is ensured by three principal components: (correct) base selection by the replicative polymerase subunit, proofreading by a separate 3′–5′ exonuclease subunit, and post-replicative mismatch repair (MMR). Combined, these processes ensure an overall mutation rate of $\sim 10^{-10}$ mutations per base pair per generation, which is significantly increased in mutator strains deficient in proofreading or MMR (Fijalkowska et al. 2012). Fluctuation analyses utilizing *rpoB* as the target for rifampicin resistance estimated the mutation rates of *M. tuberculosis* clinical and laboratory isolates at comparable values with the basal mutation rates falling within tenfold of $\sim 2 \times 10^{-10}$ per base pair per round of replication (Ford et al. 2011, 2013). Importantly, however, lineage-specific effects were discerned, with lineage 2 strains being associated with an elevated mutation rate. While the underlying mechanisms are unknown, it has been postulated that polymorphisms in 3R genes found in lineage 2 isolates may lead to relaxed fidelity in DNA repair (Dos Vultos et al. 2008). The observed increase in mutation rate in lineage 2 strains predict a substantially higher probability that patients infected with drug susceptible *M. tuberculosis* will harbor multi-drug resistant organisms at the time of diagnosis (Ford et al. 2013). Moreover, such an increase in mutation rate may also lead to the emergence of immunologically advantageous variants potentially including antigenic variants, which might improve the bacterium's selective advantage in vivo.

Like all actinomycetes, *M. tuberculosis* does not possess a canonical MMR system (Mizrahi and Andersen 1998); therefore, the absence of a severe mutator phenotype in turn implies a disproportionate role for base selection and proofreading in maintaining the mycobacterial mutation rate at a level comparable with *E. coli*. This inference may be challenged by the recent identification of an archaeal mismatch-specific endonuclease (Ishino et al. 2016); however, the role, if any, of the mycobacterial homolog in MMR has yet to be established. The suggestion that the low basal mutation rate in *M. tuberculosis* is a function primarily of intrinsic polymerase fidelity was reinforced recently in a key study which demonstrated the presence of

intrinsic PHP domain exonuclease function in the essential *dnaE1*-encoded replicative polymerase subunit in *M. tuberculosis* (Rock et al. 2015). The *M. tuberculosis* genome contains two putative homologs of the *dnaQ*-encoded ε subunit which, in *E. coli*, provides sole 3′–5′ exonuclease proofreading function (Barros et al. 2013). Deletion of the closest homolog, *Rv3711c*, had no effect on the *M. tuberculosis* mutation rate in vitro; moreover, the mutation rate was similarly unaffected in corresponding *M. smegmatis* mutants lacking one or both DnaQ homologs. This was an unexpected observation, and it suggested that, unlike the established *E. coli* model, proofreading function in mycobacteria was not dependent on extrinsic 3′–5′ exonuclease activity. Consistent with this result, it was further shown that, while purified DnaQ possesses de facto exonuclease activity in vitro (an observation that was confirmed in a separate study (Gu et al. 2016), discussed below), the putative proofreading subunit did not associate with DnaE1 in pull-down assays. This suggests that the inability of the respective replicative polymerase and exonuclease subunits to interact physically renders the mycobacterial DnaQ non-essential for proofreading.

A panel of mutant *dnaE1* alleles containing targeted mutations of PHP domain residues was then utilized to demonstrate that mycobacteria rely exclusively on PHP-domain-intrinsic 3′–5′ exonuclease activity for proofreading. Whereas purified recombinant DnaE1 was shown to possess both DNA polymerase and exonuclease functions, mutant proteins containing site-directed substitutions of residues predicted to be essential for metal co-factor coordination retained polymerase function, but lost 3′–5′ exonuclease capacity. Critically, in microbiological assays, *M. smegmatis* mutants lacking PHP exonuclease function were associated with severe growth defects and mutation rates 2,300–3,500-fold greater than wildtype. Notably, whole-genome sequence (WGS) analyses of the mutants revealed an accumulation of insertion and deletion events, confirming the requirement for functional PHP exonuclease activity in maintaining genomic integrity.

Subsequent work by Lijun Bi and colleagues (Gu et al. 2016) has extended these observations, providing detailed biochemical insight into the potential composition and activity of the core replicase complex in *M. tuberculosis.* At the same time, their latest results highlight the potential disconnect between the phenotypes observed in whole-cell (microbiological) assays, versus the enzymatic properties characterized in biochemical systems in vitro. In a *tour de force* of protein expression and purification, the authors were able to reconstitute functional *M. tuberculosis* DNA polymerase III holoenzyme – comprising the α (DnaE1), ε (DnaQ), β (DnaN), τ (DnaZX), δ (HolA), δ′ (HolB), and SSB (Ssb) subunits – in a leading-strand replication assay. Notably, in yet another departure from the *E. coli* model, biochemical analyses suggested that the core mycobacterial replicase consists of $\alpha\beta_2\varepsilon$, with the β_2 processivity factor serving as bridging protein between α and ε, which do not interact directly. Moreover, by utilizing a catalytically dead ε subunit and by titrating the concentrations of each of the three replicase subunits (α, β, and ε) or preventing α-mediated extension by dNTP starvation, the authors derived a model in which the *M. tuberculosis* DNA polymerase III replicase transitions between polymerization and proofreading modes according to the presence of the β_2 clamp (which maintains the $\alpha\beta_2\varepsilon$ replicase in polymerization mode) and dNTP availability. While these results are largely consistent with those reported previously (Gu et al. 2016; Rock et al. 2015), the absence of a mutator phenotype in any of the mycobacterial *dnaQ* deletion mutants (Rock et al. 2015) highlights the need for further research into the factors which determine relative contribution of PHP versus DnaQ exonuclease function in *M. tuberculosis* under conditions relevant to host infection. This need is further underscored by the fact that the *dnaQ* gene is highly polymorphic (Dos Vultos et al. 2008) and was also identified as a "target of independent mutation" associated with drug resistance in *M. tuberculosis* (Farhat et al. 2013).

13.5 The Function and Fidelity of Specialist DNA Polymerases in *M. tuberculosis*

Some genetic studies of mycobacterial DNA polymerases have been highly informative by virtue of the fact that functionally inactivating mutations in their encoding genes conferred phenotypes that were readily discerned under conditions predicted to be phenotypically revealing (Gordhan et al. 1996; Boshoff et al. 2003; Warner et al. 2010). For example, functional inactivation of *dnaE2* was shown to confer hypersensitivity of *M. tuberculosis* and *M. smegmatis* to DNA damage caused by exposure to mitomycin C or UV irradiation and completely abrogated UV-induced mutagenesis to rifampicin resistance (Boshoff et al. 2003; Warner et al. 2010). These phenotypes were recapitulated by transferring point mutations into the *dnaE2* gene in the *M. smegmatis* chromosome which replaced two aspartic acid residues that are highly conserved in the active site of C-family polymerases with alanine residues, thus strongly suggesting that the DNA damage tolerance and damage-induced mutagenesis phenotypes observed for *dnaE2* mutant strains were attributable to loss of the polymerase activity of DnaE2. In other cases, however, the genetic studies have been hampered by the lack of a discernible phenotypic consequence in mycobacteria of functionally inactivating mutations, either individually, or in combination with others. This is reminiscent of the failure to detect a phenotype consequent on *dnaQ* inactivation (discussed above), and is likely due to functional redundancy within, and potentially across, DNA polymerase families and/or the inadequacy of the microbiological assay systems employed (Ordonez and Shuman 2014), which have focused on mycobacterial survival and/or mutagenesis in response to genotoxic stress in vitro, ex vivo and in vivo (Kana et al. 2010; Zhu et al. 2012).

The mycobacterial Y-family polymerases provide a compelling illustration of this conundrum. Showing a profound departure from the *E. coli* DinB homologue, PolIV, the mycobacterial polymerases DinB1 and DinB2 are not induced as part of the SOS response (Boshoff et al. 2003; Kana et al. 2010); instead, DnaE2 exclusively provides SOS-inducible DNA polymerase activity in mycobacteria. Like *dnaQ*, the *dinB1* and *dinB2* genes of *M. tuberculosis* are also highly polymorphic (Dos Vultos et al. 2008). While the functional significance of the non-synonymous SNPs observed in these genes remains unclear, it is worth noting that loss of *dinB1* and/or *dinB2* function had no discernible impact on growth or survival of *M. tuberculosis* in vitro or in the lungs of infected mice; on the susceptibility of *M. tuberculosis* to the cytotoxic effects of a range of genotoxic agents; or on the rate of spontaneous mutation to rifampicin resistance (Kana et al. 2010). Recently, however, key biochemical insights have been obtained which might help to guide more informative studies on mycobacterial strains carrying mutations in *dinB1*and/or *dinB2*. In the first of two important studies, Shuman and colleagues showed that, while *M. smegmatis* DinB1 and DinB2 both function as DNA-dependent DNA polymerases, DinB2 is also able to catalyse the incorporation ribonucleotides – a feature attributable to the presence of a leucine residue in place of the canonical aromatic steric gate which blocks ribonucleotide incorporation by other DNA polymerases, including DinB1 (Ordonez et al. 2014). This intriguing feature makes DinB2 one of four mycobacterial polymerases with the ability to incorporate ribonucleotides; however, unlike the other three – the AEP family polymerases – DinB2 is capable of synthesising long stretches of RNA when copying a DNA template (Ordonez et al. 2014). In addition, *M. smegmatis* DinB2 was subsequently shown to have low-fidelity DNA polymerase activity, displaying a characteristic signature for misincorporation in vitro, and particularly prone to misincorporate dGTP and dTTP as opposed to dCTP during DNA template-directed replication (Ordonez and Shuman 2014). The fidelity of template-directed polymerisation by DinB2 is divalent metal-ion-dependent, with the mutational spectrum noticeably broader in the presence of Mn^{2+} *vs*. Mg^{2+} cofactor. Importantly, DinB2

was also shown capable of promiscuous incorporation of 8-oxo-dGMP opposite any template nucleobase, and was particularly adept at 8-oxo-dG:dT mispairing, and similarly indiscriminate when copying an oxo-dG lesion in the template strand. The low-fidelity character of DinB2 extends to its RNA polymerase activity which is promiscuous when copying an undamaged template, particularly in the presence of Mn^{2+}, and is also evidenced by the utilisation of oxo-rGTP as a substrate for replication of an undamaged template, and misincorporation of rNTPs across an 8-oxo-dG lesion in the DNA template strand. Together, these biochemical characteristics of DinB2 suggest a potential role for this specialist polymerase in "ribo patch" DNA repair under conditions in which dNTP pools are limiting and/or the organism is exposed to oxidative stress (Ordonez and Shuman 2014).

The ribonucleotide reductase (RNR) and thymidylate synthase (TS) enzymes which are responsible for the provision of dATP/dCTP/dGTP (RNR) and dTTP (TS), respectively, in mycobacteria have been extensively investigated biochemically (Liu et al. 1998, 2000; Elleingand et al. 1998; Hammerstad et al. 2014; Georgieva et al. 2008; Uppsten et al. 2004; Basta et al. 2012; Hunter et al. 2008; Yang et al. 1994) and genetically (Mowa et al. 2009; Singh et al. 2015; Fivian-Hughes et al. 2012; Dawes et al. 2003; Yang et al. 1997). Although the cellular concentrations of dNTPs are likely to be a critical determinant of the rate and fidelity of DNA replication in *M. tuberculosis*, there have been no reports to date on the measurement of dNTP pool sizes in this organism (Warner et al. 2014). Importantly, however, rapid advances in mycobacterial metabolomics (Nandakumar et al. 2015), together with the availability of a suite of conditional knockdown and knockout mutants in genes involved in dNTP biosynthesis, offer the prospect of addressing the question of dNTP pool size variation, and the physiological implications thereof comprehensively in the near future. Such studies are currently underway (V. Singh, D.F.W., V.M. & L. P. de Carvalho unpublished). Evidence from other systems suggests that rNTP pool sizes are likely to be significantly greater than those of dNTPs (Schroeder et al. 2015). The extent of the imbalance would have a profound impact on the degree of rNTP incorporation into the genome, especially under conditions in which DinB2 is likely to function. In this regard, an intriguing relationship between RNR and TS function, dTTP pool size, and *dinB2* gene dosage in *M. smegmatis* was established in a recent study by Ghosh et al. (2015). These authors made the surprising discovery that expression of the class II RNR-encoding gene from mycobacteriophage D29, gene 50, selectively conferred growth inhibitory and replication defects on the strain, ΔDRKIN, a derivative of *M. smegmatis* mc^2155 which lacks a 56-kb region of duplication in the genome (Warner et al. 2006). Comparative quantification of dNTP pool sizes implicated the induction of dTTP deficiency – and hence, thymineless death – in the growth and replication phenotype induced by expression of the RNR-encoding gene from D29 in the ΔDRKIN strain. This conclusion was supported by rescue of the phenotype by co-expression of the D29-encoded TS gene, gene 48, and consequent restoration of dTTP levels. Included among the genes present in the 56-kb duplication are those encoding the essential class Ib RNR (*nrdE* and *nrdF*) as well as *dinB2*. Importantly, ΔDRKIN was rescued from the toxic effects of gene 50 expression by ectopic expression of *dinB2*. Together, these studies have implicated the low-fidelity polymerase, DinB2, in scavenge of dNTPs and the possible misincorporation thereof, as well as in rNTP incorporation into genomic DNA when dNTP pools are limiting, i.e., when *M. tuberculosis* is in a low-growth state (Fig. 13.1).

13.6 DNA Replication as a TB Drug Target

The emergence and spread of drug-resistant strains of *M. tuberculosis* necessitates the development of new antimicrobials with novel mechanisms of action. In comparison to other essential macromolecular processes such as transcription, translation and cell wall synthesis,

relatively few antimicrobials are known that target bacterial DNA replication (Robinson et al. 2012). Does this mean that the DNA replication machinery is a poor target for drug development? Growing evidence of natural products that target bacterial DNA replication suggests the contrary, and instead identifies DNA replication as a clinically underexploited drug target.

There are currently three families of natural products that directly or indirectly interfere with DNA replication: DNA gyrase inhibitors, sliding clamp inhibitors, and DNA polymerase III inhibitors. In addition, antifolate drugs such as sulfamethoxazole, trimethoprim, and *para*-aminosalicylic acid, which are biomimetic but not natural products, deplete dNTP pools, thereby preventing DNA replication (Minato et al. 2015). However, the polypharmacologic mechanism of action of these drugs, which also involves inhibition of RNA and protein synthesis (Minato et al. 2015), excludes them from further discussion here.

DNA gyrase is a type II topoisomerase that functions during DNA replication to relieve torsional strain by introducing transient double-strand DNA (dsDNA) breaks which generate negative supercoils in the bacterial chromosome. The aminocoumarins (e.g., novobiocin) were the first of many natural products found to act as gyrase inhibitors (Barreiro and Ullan 2016). However, the relatively poor penetration of aminocoumarins across Gram-negative cell membranes, their limited solubility, and the development of the synthetic fluoroquinolones have limited the clinical utility of this compound class (Barreiro and Ullan 2016). In contrast, fluoroquinolones (e.g., moxifloxacin) which are potent gyrase poisons that stabilize the gyrase-DNA intermediate, thereby resulting in dsDNA breaks, constitute the backbone of second-line therapies for multi-drug resistant (MDR) TB.

In exciting recent developments, two classes of natural products have been identified that target the bacterial replisome. Griselimycins are cyclic peptides originally isolated from *Streptomyces griseus* by Sanofi in the 1960s. Despite potent antimycobacterial activity, these compounds were abandoned owing to poor pharmacokinetic properties (Hoagland et al. 2016). Sanofi recently revived this program, generating a series of fully synthetic griselimycin analogs with improved therapeutic potential (Kling et al. 2015). The griselimycin scaffold binds to the *dnaN*-encoded β sliding clamp with picomolar affinity to abrogate the interaction of β with DnaE1, thereby disrupting DNA replication. The preclinical lead compound, CGM, inhibits *M. tuberculosis* growth both *in vitro* and *in vivo* in a mouse model of TB and, importantly, appears to enhance *M. tuberculosis* clearance when co-administered with first-line anti-TB drugs. Moreover, spontaneous resistance to CGM occurs infrequently, and the resistance mechanism (amplification of the *dnaN* gene) is associated with dramatically reduced fitness relative to wild-type bacteria (Kling et al. 2015).

The second natural product found to target the bacterial replisome is nargenicin, a macrolide isolated from *Nocardia sp.* ACC18 with potent but narrow spectrum bactericidal antibiotic against gram-positive bacteria (Painter et al. 2015). Isolation of spontaneous nargenicin-resistant mutants of *Staphylococcus aureus* identified the DnaE replicative polymerase as the molecular target of nargenicin (Painter et al. 2015). Nargenicin binds to *Staphylococcus aureus* DnaE in the presence of DNA and, both in vitro and in vivo, inhibits DNA replication. However, given that the sole SNP identified in *dnaE* was located far from the DnaE active site, the precise mechanism by which nargenicin inhibits DNA replication remains to be identified. Importantly, counter screening of nargenicin against human DNA polymerases showed no inhibition. The spontaneous rate of resistance to nargenicin was low in *S. aureus* (1×10^{-9}), and resistant mutants displayed only low-level nargenicin resistance (MIC increased fourfold). Moreover, nargenicin showed efficacy in two independent mouse models of *S. aureus* infection (Painter et al. 2015). Somewhat unexpectedly, nargenicin compounds have recently been shown to be bactericidal against *M. tuberculosis* (International Patent Number WO2016/061772A1). Therefore, it will be important to elucidate the mechanism of action

of nargenicin in *M. tuberculosis* and to define the associated resistance mechanisms.

In addition to these natural products, there are active medicinal chemistry efforts to target *M. tuberculosis* DNA ligase and DNA primase. There are also efforts to target the bacterial helicase, DnaB (Shadrick et al. 2013); however, this work is at an early stage and has not been extended to *M. tuberculosis* DnaB, and so will not be discussed further.

DNA ligases seal nicks left after DNA replication and DNA repair. Bacterial DNA ligase (LigA) has long been thought to be an ideal drug target as it is essential and displays unique structural features and cofactor dependencies relative to its human counterpart (Brotz-Oesterhelt et al. 2003). Bayer was the first to identify a series of pyridochromanones as potent in vitro inhibitors of *E. coli* and *Streptococcus pneumoniae* LigA, with bactericidal activity in vivo (Brotz-Oesterhelt et al. 2003). The same compound series was subsequently shown to inhibit *M. tuberculosis* LigA in vitro (Gong et al. 2004), while additional work by AstraZeneca and other academic groups has identified a number of novel compound classes that selectively inhibit bacterial LigA while sparing human ligase (Mills et al. 2011; Srivastava et al. 2005a, b). Unfortunately, follow-up studies on the lead pyridochromanones demonstrated very high rates of spontaneous, high-level drug resistance in *S. aureus* (Podos et al. 2012). Moreover, drug resistance appeared to incur no decrease in fitness (Podos et al. 2012). The high rate of resistance with no associated fitness defect was attributed to a large excess of DNA ligase capacity in wild-type *S. aureus* cells, thereby allowing numerous possible ligase hypomorphic escape paths (Podos et al. 2012). Comparable genetic evidence suggests an excess ligase capacity in both *E. coli* (Lehman 1974) and *M. smegmatis* (Korycka-Machala et al. 2007), and so undermines the attractiveness of bacterial DNA ligase as a potential drug target.

DNA primase (DnaG) is an RNA polymerase that synthesizes short RNA fragments that function as primers for the replicative DNA polymerase during DNA replication. High-throughput screens have identified potent inhibitors of *M. tuberculosis* primase (Biswas et al. 2013). Unfortunately, the most efficacious molecules are doxorubicin, a non-specific DNA intercalator, and suramin, which has also been shown to inhibit human DNA polymerase α and human primase; moreover, neither compound possesses whole-cell activity against *M. smegmatis* (Biswas et al. 2013; Gajadeera et al. 2015; Kuron et al. 2014). *M. smegmatis* is able to tolerate substantial reductions in DnaG levels without obvious growth defects, again raising the question of the therapeutic vulnerability of this target (Kuron et al. 2014). Therefore, future efforts will require the identification of potent primase inhibitors that are selective for the (myco)bacterial enzyme.

13.7 Targeting DNA Metabolism to Limit the Evolution of Drug Resistance

There is yet to be a clinically relevant antibiotic for which bacteria have not developed resistance (CDC 2013). Is resistance inevitable? While most microbiologists would likely answer in the affirmative, an interesting, but relatively unexplored, space for antimicrobials is to directly target the molecular mechanisms that generate the genetic diversity that is the source of de novo antibiotic resistance.

Drug resistance in *M. tuberculosis* is caused by chromosomal mutations (Sandgren et al. 2009). Therefore, to prevent drug resistance one must prevent mutation. *M. tuberculosis* mutates at a low rate in vitro (Ford et al. 2011); the situation in vivo is less clear. Within the infected host, *M. tuberculosis* is confronted by a variety of exogenous and endogenous DNA-damaging assaults including antimicrobial reactive oxygen species (ROS) and nitrogen intermediates. Moreover, there is evidence that sub-lethal exposure of bacteria to antibiotics can generate ROS and elevate the mutation rate (Cohen et al. 2013). Despite this, WGS of *M. tuberculosis* isolates from nonhuman primate lung lesions indicates that the *M.*

tuberculosis mutation rate during infection (in the absence of treatment) is equivalent to that observed in vitro (Ford et al. 2011). On the other hand, deep sequencing of serial sputum samples has revealed significant genetic diversity in *M. tuberculosis* isolates in patients failing antibiotic therapy, raising the possibility that the mycobacterial mutation rate may be elevated in the host, perhaps specifically in the context of antibiotic therapy (Sun et al. 2012). Moreover, there is substantial evidence of selection for elevated mutation rates in clinical isolates of *E. coli*, *Salmonella enterica*, and *Pseudomonas aeruginosa* (LeClerc et al. 1996; Oliver et al. 2000). As discussed above, there is evidence of a modestly increased mutation rate in some clinical *M. tuberculosis* isolates from lineage 2 (Ford et al. 2013; McGrath et al. 2014; Werngren and Hoffner 2003). These increases in mutation rate are smaller than found in "mutator" strains of *E. coli*, *Salmonella enterica*, and *Pseudomonas aeruginosa*. However, mathematical modelling suggests that they could substantially contribute to the evolution of drug resistance because they are distributed across large bacterial populations.

As noted above, mutation can be spontaneous, as in the case of errors in DNA replication, or it can be induced, as in the case of error-prone DNA repair. Can these processes be targeted to lower the mutation rate and prevent the evolution of drug resistance? While in theory it may be possible to increase the fidelity of DNA replication by increasing the fidelity of base insertion or increasing the activity of the DNA replication proofreading, neither approach seems straightforward. A more plausible scenario would be to increase the fidelity of DNA damage repair.

In response to DNA damage, *M. tuberculosis* induces a suite of >100 genes in what is known as the SOS response (Rand et al. 2003). In *E. coli*, DNA damage is sensed by the polymerization of the DNA damage sensor RecA on single-stranded DNA (ssDNA) (Simmons et al. 2008). The RecA/ssDNA nucleoprotein filament then acts as a co-protease to cleave the LexA transcriptional repressor, resulting in activation of the SOS response. Interestingly, the genetic circuitry governing the SOS response is distinct in *M. tuberculosis*: the organism possesses both the canonical RecA-LexA regulon (Smollett et al. 2012) as well as a poorly understood RecA-LexA independent signal that regulates the majority of DNA damage inducible genes (Gamulin et al. 2004; Rand et al. 2003; Wang et al. 2011). As described above, *dnaE2* is included in the mycobacterial RecA/LexA-dependent SOS response. In an important study, Boshoff and colleagues demonstrated that *dnaE2* is induced during a mouse model of *M. tuberculosis* infection and that deletion of *dnaE2* reduced the emergence of drug resistance relative to a wild-type control (Boshoff et al. 2003). Moreover, mice infected with the *dnaE2* knockout mutant survived nearly twice as long as those infected with wild-type *M. tuberculosis*. Therefore, this work provided the first in vivo evidence that induced mutation may contribute to the evolution of drug resistance.

Much work remains to be done to demonstrate unambiguously the clinical significance of induced mutation for the evolution of drug resistance in *M. tuberculosis*. Furthermore, we still know very little about the molecular mechanisms by which induced mutation occurs during *M. tuberculosis* infection in vivo. However, available data suggest potential targets for compounds to prevent induced mutation, so called "anti-evolution" compounds. At present, the most promising targets are the regulators of the SOS response RecA and LexA and the error-prone polymerase, DnaE2. In principle, RecA could be inhibited by preventing DNA binding, ATP binding (necessary for filament formation), or oligomerization. Interestingly, small molecule inhibitors of *E. coli* RecA DNA binding have been identified that partially inhibit SOS induction (Alam et al. 2016). It is important to recognize, however, that there are up to seven RecA homologues in humans (the Rad51 family), which could render efforts to ensure target specificity extremely challenging (Culyba et al. 2015).

LexA too could be inhibited by small molecules that inhibit protease activity, either

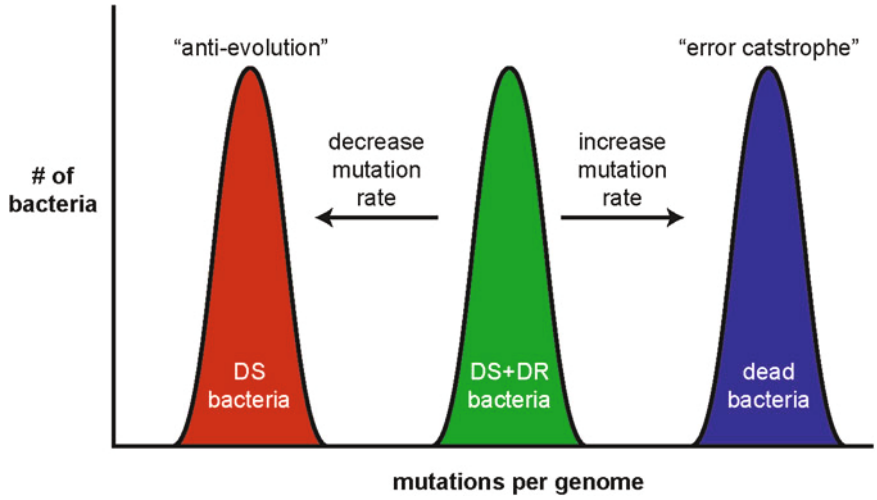

Fig. 13.2 Targeting mechanisms that drive the evolution of drug resistance in *M. tuberculosis*

by direct active site inhibition or allosteric modulators that prevent access of the LexA cleavage loop to the active site. While active site protease inhibitors are extremely well studied, any such LexA inhibitor will have to overcome the very high local concentration of the *cis*-tethered LexA cleavage loop (Culyba et al. 2015). Similarly, DnaE2 could be inhibited by polymerase active site inhibitors or compounds that prevent the interaction of DnaE2 with ImuB, or of ImuB with β sliding clamp (Warner et al. 2010). Although there are no structural data available for DnaE2, sequence alignments suggest that the active site of DnaE2 may differ significantly from the replicative polymerase DnaE1 (Timinskas et al. 2014), which would be consistent with its role as a translesion polymerase. These unique structural features could potentially be exploited to develop DnaE2-specific inhibitors. Small molecule inhibitors of DnaE2 protein interactions could also be efficacious. In this regard, it is tempting to speculate that the natural product, griselimycin, may inhibit both DNA replication and error-prone translesion synthesis by interrupting the various interactions (β-DnaE1 and β-ImuB-DnaE2) that are essential to these processes (Kling et al. 2015).

The possibility of targeting the mechanisms that drive the evolution drug resistance in *M. tuberculosis* represents an exciting opportunity and significant challenge for future research. Were such "anti-evolution" compounds to be developed (Fig. 13.2), they would be co-administered with conventional antibiotics during *M. tuberculosis* chemotherapy, thereby decreasing treatment failure rates due to de novo drug resistance and preserving the utility of our increasingly dwindling pool of effective antibiotics.

13.8 Targeting DNA Metabolism to Induce Error Catastrophe

A much more controversial idea is to identify mechanisms to increase the mutation rate beyond the viability threshold, a concept referred to as "error catastrophe" (Fig. 13.2). Such an approach has been experimentally validated for small RNA viruses, whose naturally elevated mutation rates ensure that they are perched on the edge of viability (Crotty et al. 2001; Loeb et al. 1999). For example, treatment of poliovirus with the mutagenic nucleoside analog, ribavirin, increases the mutation rate by only ~10-fold, but results in a 99.3% reduction in infectivity (Crotty et al. 2001). Bacteria have generally not been considered susceptible to error catastrophe since their mutation rates are three to four

orders of magnitude lower than RNA viruses (Drake 1999), and it has been difficult to identify mechanisms to elevate the mutation to lethal levels. Moreover, elevated mutation rates have been identified at high frequency in clinically-isolated bacterial pathogens, indicating that an increased mutation rate can provide a selective advantage (LeClerc et al. 1996; Oliver et al. 2000). However, it is important to distinguish between the magnitude of mutation rate increase identified in clinical isolates and that necessary to achieve lethal mutagenesis. Clinically observed "mutator" strains are almost universally MMR defective (LeClerc et al. 1996; Oliver et al. 2000) which, in *E. coli*, results in a mutation rate of ~0.15 mutations per genome per generation (Lee et al. 2012). Mathematical models have predicted that a mutation rate of ~3–6 mutations per genome per generation is necessary to reach the lethal threshold in bacteria (Bull and Wilke 2008). Consistent with these predictions, a mutation rate of ~50 mutations per genome per generation is lethal in haploid *Saccharomyces cerevisiae*, a budding yeast which displays a basal mutation rate that is very similar to that of *M. tuberculosis* (Morrison et al. 1993).

Given the caveats discussed above, how might one push *M. tuberculosis* over the error threshold? The most plausible mechanism would be to disable DNA replication proofreading with a small molecule that inhibits the PHP domain of the replicative polymerase DnaE1 (Rock et al. 2015). As discussed previously, inactivation of replicative proofreading increases the mutation rate to approximately 6–11 mutations per genome per generation in *M. smegmatis*, which is near the predicted lethal threshold for bacteria (Bull and Wilke 2008; Rock et al. 2015). Consistent with this notion, PHP domain inhibition results in a dramatic fitness defect, although it remains to be shown that this defect is a direct result of the mutation rate increase (Rock et al. 2015). Similar to the case of ribavirin and poliovirus (Crotty et al. 2001), the simultaneous inactivation of the DNA replication proofreading coupled with the use of mutagenic nucleoside analogs could be used to further increase the mutation rate to induce error catastrophe in *M. tuberculosis*.

13.9 Conclusions

The problem appears simple: given the fact that *M. tuberculosis* relies solely on chromosomal mutations for evolution, elucidating the mechanisms underlying their generation and fixation will enable a clear understanding of the countermeasures required to interrupt drug resistance and virulence. Experimental mycobacteriology, including the use of genetic tools for targeted disruption of specific functions and pathways, has provided important insights into the operation of different DNA replication processes and components; however, the extensive redundancies and hierarchies of operation which likely determine their relative utilization in vivo, render extremely difficult any definitive assignment of causality. Instead, there is likely a significant disconnect between the functions and capacities inferred in vitro using a limited set of growth conditions and stimuli, none of which can adequately recapitulate the conditions which enable the generation of microdiverse subpopulations in a single infected host. For this reason, a number of key research questions remain unresolved: What is the source of mutations in *M. tuberculosis* during host infection? Do they arise predominantly at standard (in vitro) rates during normal replication? Are they a consequence of error-prone DNA damage repair induced in a genotoxic environment? If so, what is the predominant error-prone repair mechanism, and what is the source of the genotoxic stress? And, finally, what is the size of the infecting bacillary population that is capable of generating sufficient genomic diversity through mutation to enable selection and fixation of alternative alleles – including those conferring drug resistance – in individual patients undergoing combination therapy? Addressing each of these is essential to developing novel intervention strategies to disrupt the adaptation of *M. tuberculosis* to its human host and, equally, to curbing the emergence of multidrug resistant strains which

nevertheless retain the ability to transmit, infect, and cause disease.

Acknowledgements This work was supported by a subcontract from the US National Institute of Child Health and Human Development (NICHD) U01HD085531-02 (to D.F.W.); the South African Medical Research Council (to V.M.); the National Research Foundation of South Africa (to D.F.W. and V.M.); a Senior International Research Scholars grant from the Howard Hughes Medical Institute (to V.M.); a Helen Hay Whitney fellowship (to J.M.R.); an NIH Director's New Innovator Award 1DP20D001378 (to S.F.), and a subcontract from NIAID U19 AI107774-0 (to S.F.).

References

Alam MK, Alhhazmi A, DeCoteau JF, Luo Y, Geyer CR (2016) RecA inhibitors potentiate antibiotic activity and block evolution of antibiotic resistance. Cell Chem Biol 23(3):381–391

Barreiro C, Ullan RV (2016) DNA-synthesizing enzymes as antibacterial targets. In: Villa TG, Vinas M (eds) New weapons to control bacterial growth. Springer International Publishing, Switzerland

Barros T, Guenther J, Kelch B, Anaya J, Prabhakar A, O'Donnell M, Kuriyan J, Lamers MH (2013) A structural role for the PHP domain in *E. coli* DNA polymerase III. BMC Struct Biol 13:8

Basta T, Boum Y, Briffotaux J, Becker HF, Lamarre-Jouenne I, Lambry JC, Skouloubris S, Liebl U, Graille M, van Tilbeurgh H, Myllykallio H (2012) Mechanistic and structural basis for inhibition of thymidylate synthase ThyX. Open Biol 2:120120

Biswas T, Resto-Roldan E, Sawyer SK, Artsimovitch I, Tsodikov OV (2013) A novel non-radioactive primase-pyrophosphatase activity assay and its application to the discovery of inhibitors of *Mycobacterium tuberculosis* primase DnaG. Nucleic Acids Res 41(4):e56

Boritsch EC, Supply P, Honore N, Seeman T, Stinear TP, Brosch R (2014) A glimpse into the past and predictions for the future: the molecular evolution of the tuberculosis agent. Mol Microbiol 93:835–852. doi:10.1111/mmi.12720

Boshoff HI, Reed MB, Barry CE 3rd, Mizrahi V (2003) DnaE2 polymerase contributes to in vivo survival and the emergence of drug resistance in *Mycobacterium tuberculosis*. Cell 113(2):183–193

Brissett NC, Pitcher RS, Juarez R, Picher AJ, Green AJ, Dafforn TR, Fox GC, Blanco L, Doherty AJ (2007) Structure of a NHEJ polymerase-mediated DNA synaptic complex. Science 318(5849):456–459

Brotz-Oesterhelt H, Knezevic I, Bartel S, Lampe T, Warnecke-Eberz U, Ziegelbauer K, Habich D, Labischinski H (2003) Specific and potent inhibition of NAD+-dependent DNA ligase by pyridochromanones. J Biol Chem 278(41):39435–39442

Bull JJ, Wilke CO (2008) Lethal mutagenesis of bacteria. Genetics 180(2):1061–1070

CDC (2013) Antiobiotic resistance threats in the United States. Centres for Disease Conotrl and Prevention

Cohen NR, Lobritz MA, Collins JJ (2013) Microbial persistence and the road to drug resistance. Cell Host Microbe 13(6):632–642

Colangeli R, Arcus VL, Cursons RT, Ruthe A, Karalus N, Coley K, Manning SD, Kim S, Marchiano E, Alland D (2014) Whole genome sequencing of *Mycobacterium tuberculosis* reveals slow growth and low mutation rates during latent infections in humans. PLoS One 9(3):e91024

Crotty S, Cameron CE, Andino R (2001) RNA virus error catastrophe: direct molecular test by using ribavirin. Proc Natl Acad Sci U S A 98(12):6895–6900

Culyba MJ, Mo CY, Kohli RM (2015) Targets for combating the evolution of acquired antibiotic resistance. Biochemistry 54(23):3573–3582

Darwin KH, Nathan CF (2005) Role for nucleotide excision repair in virulence of *Mycobacterium tuberculosis*. Infect Immun 73(8):4581–4587

Dawes SS, Warner DF, Tsenova L, Timm J, McKinney JD, Kaplan G, Rubin H, Mizrahi V (2003) Ribonucleotide reduction in *Mycobacterium tuberculosis*: function and expression of genes encoding class Ib and class II ribonucleotide reductases. Infect Immun 71(11):6124–6131

Dos Vultos T, Mestre O, Rauzier J, Golec M, Rastogi N, Rasolofo V, Tonjum T, Sola C, Matic I, Gicquel B (2008) Evolution and diversity of clonal bacteria: the paradigm of *Mycobacterium tuberculosis*. PLoS One 3(2):e1538

Drake JW (1999) The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes. Ann N Y Acad Sci 870:100–107

Dutta NK, Mehra S, Didier PJ, Roy CJ, Doyle LA, Alvarez X, Ratterree M, Be NA, Lamichhane G, Jain SK, Lacey MR, Lackner AA, Kaushal D (2010) Genetic requirements for the survival of tubercle bacilli in primates. J Infect Dis 201(11):1743–1752

Elleingand E, Gerez C, Un S, Knupling M, Lu G, Salem J, Rubin H, Sauge-Merle S, Laulhere JP, Fontecave M (1998) Reactivity studies of the tyrosyl radical in ribonucleotide reductase from *Mycobacterium tuberculosis* and *Arabidopsis thaliana* – comparison with *Escherichia coli* and mouse. Eur J Biochem 258(2):485–490

Farhat MR, Shapiro BJ, Kieser KJ, Sultana R, Jacobson KR, Victor TC, Warren RM, Streicher EM, Calver A, Sloutsky A, Kaur D, Posey JE, Plikaytis B, Oggioni MR, Gardy JL, Johnston JC, Rodrigues M, Tang PK, Kato-Maeda M, Borowsky ML, Muddukrishna B, Kreiswirth BN, Kurepina N, Galagan J, Gagneux S, Birren B, Rubin EJ, Lander ES, Sabeti PC, Murray M (2013) Genomic analysis identifies targets of convergent positive selection in drug-resistant

Mycobacterium tuberculosis. Nat Genet 45(10): 1183–1189

Fijalkowska IJ, Schaaper RM, Jonczyk P (2012) DNA replication fidelity in *Escherichia coli*: a multi-DNA polymerase affair. FEMS Microbiol Rev 36(6):1105–1121

Fivian-Hughes AS, Houghton J, Davis EO (2012) *Mycobacterium tuberculosis* thymidylate synthase gene *thyX* is essential and potentially bifunctional, while *thyA* deletion confers resistance to p-aminosalicylic acid. Microbiol-Sgm 158:1388–1388

Ford CB, Lin PL, Chase MR, Shah RR, Iartchouk O, Galagan J, Mohaideen N, Ioerger TR, Sacchettini JC, Lipsitch M, Flynn JL, Fortune SM (2011) Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. Nat Genet 43(5):482–486

Ford CB, Shah RR, Maeda MK, Gagneux S, Murray MB, Cohen T, Johnston JC, Gardy J, Lipsitch M, Fortune SM (2013) *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. Nat Genet 45(7):784–790

Gajadeera C, Willby MJ, Green KD, Shaul P, Fridman M, Garneau-Tsodikova S, Posey JE, Tsodikov OV (2015) Antimycobacterial activity of DNA intercalator inhibitors of *Mycobacterium tuberculosis* primase DnaG. J Antibiot (Tokyo) 68(3):153–157

Galagan JE (2014) Genomic insights into tuberculosis. Nat Rev Genet 15(5):307–320

Gamulin V, Cetkovic H, Ahel I (2004) Identification of a promoter motif regulating the major DNA damage response mechanism of *Mycobacterium tuberculosis*. FEMS Microbiol Lett 238(1):57–63

Georgieva ER, Narvaez AJ, Hedin N, Graslund A (2008) Secondary structure conversions of *Mycobacterium tuberculosis* ribonucleotide reductase protein R2 under varying pH and temperature conditions. Biophys Chem 137(1):43–48

Ghosh S, Samaddar S, Kirtania P, Das Gupta SK (2015) A DinB ortholog enables mycobacterial growth under dTTP-limiting conditions induced by the expression of a mycobacteriophage-derived ribonucleotide reductase gene. J Bacteriol 198(2):352–362

Gong C, Martins A, Bongiorno P, Glickman M, Shuman S (2004) Biochemical and genetic analysis of the four DNA ligases of mycobacteria. J Biol Chem 279(20):20594–20606

Gordhan BG, Andersen SJ, De Meyer AR, Mizrahi V (1996) Construction by homologous recombination and phenotypic characterization of a DNA polymerase domain *polA* mutant of *Mycobacterium smegmatis*. Gene 178(1–2):125–130

Gorna AE, Bowater RP, Dziadek J (2010) DNA repair systems and the pathogenesis of *Mycobacterium tuberculosis*: varying activities at different stages of infection. Clin Sci (Lond) 119(5):187–202

Gu S, Li W, Zhang H, Fleming J, Yang W, Wang S, Wei W, Zhou J, Zhu G, Deng J, Hou J, Zhou Y, Lin S, Zhang XE, Bi L (2016) The b_2 clamp in the *Mycobacterium tuberculosis* DNA polymerase III ab_2e replicase promotes polymerization and reduces exonuclease activity. Sci Rep 6:18418

Hammerstad M, Rohr AK, Andersen NH, Graslund A, Hogbom M, Andersson KK (2014) The class Ib ribonucleotide reductase from *Mycobacterium tuberculosis* has two active R2F subunits. J Biol Inorg Chem 19(6):893–902

Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, Roach JC, Kremer K, Petrov DA, Feldman MW, Gagneux S (2008) High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. PLoS Biol 6(12):e311

Hoagland DT, Liu J, Lee RB, Lee RE (2016) New agents for the treatment of drug-resistant *Mycobacterium tuberculosis*. Adv Drug Deliv Rev 102:55

Hunter JH, Pang CKT, Gujjar R, Rathod PK (2008) Kinetics and ligand-binding preferences of *Mycobacterium tuberculosis* thymidylate synthases, ThyA and ThyX. PLoS One 3(5):e2237

Ishino S, Nishi Y, Oda S, Uemori T, Sagara T, Takatsu N, Yamagami T, Shirai T, Ishino Y (2016) Identification of a mismatch-specific endonuclease in hyperthermophilic Archaea. Nucleic Acids Res 44(7):2977–2986

Kana BD, Abrahams GL, Sung N, Warner DF, Gordhan BG, Machowski EE, Tsenova L, Sacchettini JC, Stoker NG, Kaplan G, Mizrahi V (2010) Role of the DinB homologs Rv1537 and Rv3056 in *Mycobacterium tuberculosis*. J Bacteriol 192(8):2220–2227

Kling A, Lukat P, Almeida DV, Bauer A, Fontaine E, Sordello S, Zaburannyi N, Herrmann J, Wenzel SC, Konig C, Ammerman NC, Barrio MB, Borchers K, Bordon-Pallier F, Bronstrup M, Courtemanche G, Gerlitz M, Geslin M, Hammann P, Heinz DW, Hoffmann H, Klieber S, Kohlmann M, Kurz M, Lair C, Matter H, Nuermberger E, Tyagi S, Fraisse L, Grosset JH, Lagrange S, Muller R (2015) Targeting DnaN for tuberculosis therapy using novel griselimycins. Science 348(6239):1106–1112

Korycka-Machala M, Rychta E, Brzostek A, Sayer HR, Rumijowska-Galewicz A, Bowater RP, Dziadek J (2007) Evaluation of NAD(+) -dependent DNA ligase of mycobacteria as a potential target for antibiotics. Antimicrob Agents Chemother 51(8):2888–2897

Kuron A, Korycka-Machala M, Brzostek A, Nowosielski M, Doherty A, Dziadek B, Dziadek J (2014) Evaluation of DNA primase DnaG as a potential target for antibiotics. Antimicrob Agents Chemother 58(3):1699–1706

LeClerc JE, Li B, Payne WL, Cebula TA (1996) High mutation frequencies among *Escherichia coli a*nd *Salmonella* pathogens. Science 274(5290): 1208–1211

Lee H, Popodi E, Tang H, Foster PL (2012) Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. Proc Natl Acad Sci U S A 109(41):E2774–E2783

Lehman IR (1974) DNA ligase: structure, mechanism, and function. Science 186(4166):790–797

Lillebaek T, Dirksen A, Baess I, Strunge B, Thomsen VO, Andersen AB (2002) Molecular evidence of endogenous reactivation of *Mycobacterium tuberculosis* after 33 years of latent infection. J Infect Dis 185(3):401–404

Liu A, Potsch S, Davydov A, Barra AL, Rubin H, Graslund A (1998) The tyrosyl free radical of recombinant ribonucleotide reductase from *Mycobacterium tuberculosis* is located in a rigid hydrophobic pocket. Biochemistry 37(46):16369–16377

Liu AM, Barra AL, Rubin H, Lu GZ, Graslund A (2000) Heterogeneity of the local electrostatic environment of the tyrosyl radical in *Mycobacterium tuberculosis* ribonucleotide reductase observed by high-field electron paramagnetic resonance. J Am Chem Soc 122(9):1974–1978

Loeb LA, Essigmann JM, Kazazi F, Zhang J, Rose KD, Mullins JI (1999) Lethal mutagenesis of HIV with mutagenic nucleoside analogs. Proc Natl Acad Sci U S A 96(4):1492–1497

McGrath M, Gey van Pittius NC, van Helden PD, Warren RM, Warner DF (2014) Mutation rate and the emergence of drug resistance in *Mycobacterium tuberculosis*. J Antimicrob Chemother 69(2):292–302

Miggiano R, Casazza V, Garavaglia S, Ciaramella M, Perugino G, Rizzi M, Rossi F (2013) Biochemical and structural studies of the *Mycobacterium tuberculosis* O^6-methylguanine methyltransferase and mutated variants. J Bacteriol 195(12):2728–2736

Mills SD, Eakin AE, Buurman ET, Newman JV, Gao N, Huynh H, Johnson KD, Lahiri S, Shapiro AB, Walkup GK, Yang W, Stokes SS (2011) Novel bacterial NAD+-dependent DNA ligase inhibitors with broad-spectrum activity and antibacterial efficacy in vivo. Antimicrob Agents Chemother 55(3):1088–1096

Minato Y, Thiede JM, Kordus SL, McKlveen EJ, Turman BJ, Baughn AD (2015) *Mycobacterium tuberculosis* folate metabolism and the mechanistic basis for para-aminosalicylic acid susceptibility and resistance. Antimicrob Agents Chemother 59(9):5097–5106

Mizrahi V, Andersen SJ (1998) DNA repair in *Mycobacterium tuberculosis*. What have we learnt from the genome sequence? Mol Microbiol 29(6): 1331–1339

Mizrahi V, Huberts P (1996) Deoxy- and dideoxynucleotide discrimination and identification of critical 5′ nuclease domain residues of the DNA polymerase I from *Mycobacterium tuberculosis*. Nucleic Acids Res 24(24):4845–4852

Morrison A, Johnson AL, Johnston LH, Sugino A (1993) Pathway correcting DNA replication errors in *Saccharomyces cerevisiae*. EMBO J 12(4):1467–1473

Mowa MB, Warner DF, Kaplan G, Kana BD, Mizrahi V (2009) Function and regulation of class I ribonucleotide reductase-encoding genes in mycobacteria. J Bacteriol 191(3):985–995

Nandakumar M, Prosser GA, de Carvalho LP, Rhee K (2015) Metabolomics of *Mycobacterium tuberculosis*. Methods Mol Biol 1285:105–115

Oliver A, Canton R, Campo P, Baquero F, Blazquez J (2000) High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. Science 288(5469):1251–1254

Ordonez H, Shuman S (2014) *Mycobacterium smegmatis* DinB2 misincorporates deoxyribonucleotides and ribonucleotides during templated synthesis and lesion bypass. Nucleic Acids Res 42(20):12722–12734

Ordonez H, Uson ML, Shuman S (2014) Characterization of three mycobacterial DinB (DNA polymerase IV) paralogs highlights DinB2 as naturally adept at ribonucleotide incorporation. Nucleic Acids Res 42(17):11056–11070

Painter RE, Adam GC, Arocho M, DiNunzio E, Donald RG, Dorso K, Genilloud O, Gill C, Goetz M, Hairston NN, Murgolo N, Nare B, Olsen DB, Powles M, Racine F, Su J, Vicente F, Wisniewski D, Xiao L, Hammond M, Young K (2015) Elucidation of DnaE as the antibacterial target of the natural product, Nargenicin. Chem Biol 22(10):1362–1373

Podos SD, Thanassi JA, Pucci MJ (2012) Mechanistic assessment of DNA ligase as an antibacterial target in *Staphylococcus aureus*. Antimicrob Agents Chemother 56(8):4095–4102

Rand L, Hinds J, Springer B, Sander P, Buxton RS, Davis EO (2003) The majority of inducible DNA repair genes in *Mycobacterium tuberculosis* are induced independently of RecA. Mol Microbiol 50(3):1031–1042

Robinson A, Causer RJ, Dixon NE (2012) Architecture and conservation of the bacterial DNA replication machinery, an underexploited drug target. Curr Drug Targets 13(3):352–372

Rock JM, Lang UF, Chase MR, Ford CB, Gerrick ER, Gawande R, Coscolla M, Gagneux S, Fortune SM, Lamers MH (2015) DNA replication fidelity in *Mycobacterium tuberculosis* is mediated by an ancestral prokaryotic proofreader. Nat Genet 47(6):677–681

Sandgren A, Strong M, Muthukrishnan P, Weiner BK, Church GM, Murray MB (2009) Tuberculosis drug resistance mutation database. PLoS Med 6(2):e2

Sassetti CM, Rubin EJ (2003) Genetic requirements for mycobacterial survival during infection. Proc Natl Acad Sci U S A 100(22):12989–12994

Schroeder JW, Randall JR, Matthews LA, Simmons LA (2015) Ribonucleotides in bacterial DNA. Crit Rev Biochem Mol Biol 50(3):181–193

Shadrick WR, Ndjomou J, Kolli R, Mukherjee S, Hanson AM, Frick DN (2013) Discovering new medicines targeting helicases: challenges and recent progress. J Biomol Screen 18(7):761–781

Sharma A, Nair DT (2012) MsDpo4-a DinB Homolog from *Mycobacterium smegmatis*-is an error-prone DNA polymerase that can promote G:T and T:G mismatches. J Nucleic Acids 2012:285481

Simmons LA, Foti JJ, Cohen SE, Walker GC (2008) The SOS regulatory network. EcoSal Plus 3(1)

Singh V, Brecik M, Mukherjee R, Evans JC, Svetlikova Z, Blasko J, Surade S, Blackburn J, Warner DF, Mikusova K, Mizrahi V (2015) The complex mechanism of

antimycobacterial action of 5-fluorouracil. Chem Biol 22(1):63–75

Smollett KL, Smith KM, Kahramanoglou C, Arnvig KB, Buxton RS, Davis EO (2012) Global analysis of the regulon of the transcriptional repressor LexA, a key component of SOS response in Mycobacterium tuberculosis. J Biol Chem 287(26):22004–22014

Srivastava SK, Dube D, Tewari N, Dwivedi N, Tripathi RP, Ramachandran R (2005a) *Mycobacterium tuberculosis* NAD+-dependent DNA ligase is selectively inhibited by glycosylamines compared with human DNA ligase I. Nucleic Acids Res 33(22):7090–7101

Srivastava SK, Tripathi RP, Ramachandran R (2005b) NAD+-dependent DNA ligase (Rv3014c) from *Mycobacterium tuberculosis*. Crystal structure of the adenylation domain and identification of novel inhibitors. J Biol Chem 280(34):30273–30281

Sun G, Luo T, Yang C, Dong X, Li J, Zhu Y, Zheng H, Tian W, Wang S, Barry CE 3rd, Mei J, Gao Q (2012) Dynamic population changes in *Mycobacterium tuberculosis* during acquisition and fixation of drug resistance in patients. J Infect Dis 206(11): 1724–1733

Timinskas K, Balvociute M, Timinskas A, Venclovas C (2014) Comprehensive analysis of DNA polymerase III alpha subunits and their homologs in bacterial genomes. Nucleic Acids Res 42(3):1393–1413

Uppsten M, Davis J, Rubin H, Uhlin U (2004) Crystal structure of the biologically active form of class 1b ribonucleotide reductase small subunit from *Mycobacterium tuberculosis*. FEBS Lett 569(1–3):117–122

Wang Y, Huang Y, Xue C, He Y, He ZG (2011) ClpR protein-like regulator specifically recognizes RecA protein-independent promoter motif and broadly regulates expression of DNA damage-inducible genes in mycobacteria. J Biol Chem 286(36): 31159–31167

Warner DF, Etienne G, Wang XM, Matsoso LG, Dawes SS, Soetaert K, Stoker NG, Content J, Mizrahi V (2006) A derivative of *Mycobacterium smegmatis* mc^2155 that lacks the duplicated chromosomal region. Tuberculosis 86(6):438–444

Warner DF, Ndwandwe DE, Abrahams GL, Kana BD, Machowski EE, Venclovas C, Mizrahi V (2010) Essential roles for *imuA′*- and *imuB*-encoded accessory factors in DnaE2-dependent mutagenesis in *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A 107(29):13093–13098

Warner DF, Tonjum T, Mizrahi V (2014) DNA metabolism in mycobacterial pathogenesis. Curr Top Microbiol 374:27–51

Warner DF, Koch A, Mizrahi V (2015) Diversity and disease pathogenesis in *Mycobacterium tuberculosis*. Trends Microbiol 23(1):14–21

Werngren J, Hoffner SE (2003) Drug-susceptible *Mycobacterium tuberculosis* Beijing genotype does not develop mutation-conferred resistance to rifampin at an elevated rate. J Clin Microbiol 41(4): 1520–1524

Yang F, Lu G, Rubin H (1994) Isolation of ribonucleotide reductase from *Mycobacterium tuberculosis* and cloning, expression, and purification of the large subunit. J Bacteriol 176(21):6738–6743

Yang F, Curran SC, Li LS, Avarbock D, Graf JD, Chua MM, Lu G, Salem J, Rubin H (1997) Characterization of two genes encoding the *Mycobacterium tuberculosis* ribonucleotide reductase small subunit. J Bacteriol 179(20):6408–6415

Zhu H, Nandakumar J, Aniukwu J, Wang LK, Glickman MS, Lima CD, Shuman S (2006) Atomic structure and nonhomologous end-joining function of the polymerase component of bacterial DNA ligase D. Proc Natl Acad Sci U S A 103(6):1711–1716

Zhu H, Bhattarai H, Yan HG, Shuman S, Glickman MS (2012) Characterization of *Mycobacterium smegmatis* PolD2 and PolD1 as RNA/DNA polymerases homologous to the POL domain of bacterial DNA ligase D. Biochemistry 51(51):10147–10158

Strain Diversity and the Evolution of Antibiotic Resistance

14

Sonia Borrell and Andrej Trauner

Abstract

Drug resistance is best thought of as an ongoing biological process. Resistant bacteria must emerge, become established and ultimately transmit in order to be relevant to human health. In this context, genetic diversity can influence the rate and likelihood of resistance emerging; it can also modulate the net physiological impact of resistance and the propensity of an organism to improve any defects that arise from it. Combined, these effects can have an impact on a larger scale, with highly transmissible drug-resistant bacterial strains posing a formidable threat to global health. These considerations are pertinent to the future of tuberculosis control as well. In this chapter, we review our current understanding of the impact of genetic diversity in the broadest sense on the evolution of drug-resistant members of the *Mycobacterium tuberculosis* complex.

Keywords

Drug resistance • Evolution • Epistasis • Fitness Landscape

14.1 Drug Resistance as a Process (Not an Endpoint)

The emergence of drug resistance is a process that spans biological and ecological scales. For instance, the frequency and mechanics of mutation are largely defined by biophysical properties of DNA and the enzyme that replicates it: DNA polymerase. Once they occur, most mutations are not made equal. In the specific case of drug resistance, the impact of a mutation on the structure of a drug target or prodrug activator will determine the extent to which a mutation confers resistance. Next, the biochemistry of mutant proteins can influence the ability of a cell to grow and the interactions between cells carrying different mutations determine the ability of each mutant to survive in competition with the others. These steps highlight the fact that resistance mutations must first emerge, and

S. Borrell (✉) • A. Trauner
Swiss Tropical and Public Health Institute, Socinstrasse 57, 4002, Basel, Switzerland

University of Basel, Basel, Switzerland
e-mail: sonia.borrell@swisstph.ch; andrej.trauner@swisstph.ch

S. Gagneux (ed.), *Strain Variation in the* Mycobacterium tuberculosis *Complex: Its Role in Biology, Epidemiology and Control*, Advances in Experimental Medicine and Biology 1019,
DOI 10.1007/978-3-319-64371-7_14

then they need to be selected for. For obligate pathogens, all of these processes need to occur while retaining the ability to survive, evade or counter host defenses. But more importantly, in order to establish itself within a population of hosts, the drug-resistant pathogen needs to be able to spread, adding successful transmission to the set of evolutionary constrains. The success of this process is modulated by a number of factors, including the genetic interactions between mutations, the intensity of selection, the size of the pathogen population and immunological determinants of the host. In this chapter, we will focus specifically on the role and consequences of genetic diversity on the evolution of drug resistance in the *Mycobacterium tuberculosis* complex (MTBC).

14.2 Treatment and Drug Resistance of the MTBC

MTBC is normally treated by a combination of drugs via directly observed therapy (DOTS). For drug-susceptible TB, DOTS relies on the supervised administration of four drugs: isoniazid (INH), rifampicin (RIF), ethambutol (EMB) and pyrazinamide (PZA) over a period of 6 months. Resistance to various components of the combination has disparate effects, however from a perspective of disease control, the combined resistance to INH and RIF, defined as MultiDrug-Resistant (MDR) TB, poses the greatest threat. Treatment regimens used for patients with MDR-TB currently depend on fluoroquinolones (FQ) and injectable aminoglycosides (AG) for favourable outcomes. Acquisition of resistance to these classes of antibiotics defines eXtensively Drug-Resistant (XDR) TB. MDR- and XDR-TB emerged on several epidemiologically independent occasions, showing the evolutionary potential of MTBC (see Chap. 11 for a more detailed review).

The use of combination therapy directs the evolutionary process in the MTBC towards the eventual emergence of multidrug resistance. In many bacterial pathogens, multidrug resistance occurs through the acquisition of plasmids encoding several resistance determinants. But there is little evidence for recent horizontal gene transfer in MTBC (see Chap. 13); instead the resistance to each agent is mediated by chromosomal mutations. The MTBC relies heavily on two types of resistance mechanisms: target modification and loss of prodrug activation. Target mutation results in loss of binding affinity and occurs across many drug classes. For example, RIF inhibits RNA polymerase (RNAP) and mutations in the RIF resistance-determining region (RRDR) of the β subunit of RNAP (RpoB) decrease the affinity of its interaction with the antibiotic (Campbell et al. 2001). Similarly, mutations in the quinolone resistance-determining region (QRDR) of the α subunit of DNA gyrase (GyrA) have the same effect on FQ-binding affinity (Piton et al. 2010) and mutations in the 16S ribosomal RNA disrupt the binding of many AGs (Magnet and Blanchard 2005). Many MTBC-specific antibiotics are prodrugs and become active only through interaction with intracellular enzymes: prodrug activators. Resistance to this type of antibiotics, which includes INH, PZA, ethionamide (ETH) and the new nitroimidazole class of antibiotics, often occurs by the modification or loss of the activating enzyme, rather than mutation of the target (Haver et al. 2015). Drug efflux and transcriptional modulation also occur as resistance mechanisms (for bedaquiline and INH/ETH, respectively) but their relative frequency is much lower, possibly because their contribution to resistance in MTBC tends to be less pronounced (Trauner et al. 2014). Nonetheless, transcriptional modulation of MmpL4/S4 through mutations in MmpR leads to the efflux of bedaquiline and cross-resistance with clofazimine (Hartkoorn et al. 2014). The existence of such mechanisms may be coincidental; however it raises the possibility that efflux pumps may play an important role as an unspecific resistance mechanism in future drug regimens.

14.3 Geographical Hotspots of MDR-TB

The most recent survey conducted by the WHO in 2015 indicates that 4.6% of the 10.4 million new cases of TB were caused by MDR-TB. This relatively low rate conceals significant variation between and within countries and regions (see Chap. 11) ((WHO) 2016). MTBC resistance rates in high-income countries are not representative of the frequency of MDR/XDR-TB strains in other parts of the world. Regions with ongoing MDR-TB epidemics are defined as hotspots – areas where the prevalence of MDR-TB cases is >5% (Blower and Chou 2004). Localized high incidence rates of MDR-TB have been found only in particular regions, mostly in the Russian federation, India and China. Despite a relatively low prevalence, the fact that 36% of the world's total population resides in China and India means that as much as 50% of the burden of MDR-TB falls on these two countries. Results from first-line diagnostics reported annually in South Africa found 6.6% of all TB cases in South Africa to be RIF-resistant in 2013/2014. A study conducted in Belarus in 2009–2010 revealing that almost half of all TB cases were MDR ((WHO) 2015). Among MDR-TB isolates, 12% were XDR (Skrahina et al. 2013). This latter is a common occurrence – a significant proportion of MDR isolates are resistant to additional drugs; for example one study found 10.6% of all MDR-TB isolates to qualify for XDR-TB status (Dalal et al. 2015).

The reason why the prevalence of MDR-TB varies significantly across the globe is a result of many contributing factors. These include variably effective control programs (Zignol et al. 2012; Cohen et al. 2014), environmental factors such as permanence in prisons or mines (Dalton et al. 2012; Zignol et al. 2016), the presence of co-morbidities, which may influence the emergence and spread of MDR-TB (e.g. HIV co-infection) (Wells et al. 2007; Escombe et al. 2008; Dean et al. 2014), and numerous patient-related factors. These range from societal and psychosomatic to immunological and genetic factors, as well as patient noncompliance with chemotherapy, pharmacokinetic variability and even circulation of counterfeit drugs (Perlman et al. 2005; Reynolds and Heysell 2014; Zignol et al. 2016).

Adding to the heterogeneity of environmental and host factors is the genetic diversity of MTBC which shows a discrete phylogeographical distribution (Coscolla and Gagneux 2014, Chap. 1). It is therefore possible that regional differences in the burden of resistance could be due to phenotypic differences between MTBC lineages. If the dominant MTBC lineage in a region was more prone to acquire resistance when compared to another, this could exacerbate the resistance burden in the region. Supporting this possibility is the fact that predominant lineages in MDR hotspots are often similar. Russia and China show a predominance of Lineage 2, which includes the Beijing family of strains (Merker et al. 2015). Several reports have associated Beijing strains with an increased probability of MDR-TB (Munsiff et al. 2003). It is therefore possible that the global distribution of MDR hotspots is in fact driven by the local predominance of Lineage 2/Beijing. Next we will review the evidence pertaining to this hypothesis.

14.4 MTBC Strain Diversity and Resistance: The Case of Lineage 2/Beijing Family

Lineage 2 is one of the most successful MTBC lineages with increasing prevalence in the global population. Lineage 2, which includes the Beijing family, is the predominant lineage in large parts of Asia (East and South-East) and Eastern Europe. Several studies have revealed a strong link between Beijing and drug resistance (Munsiff et al. 2003; Borrell and Gagneux 2009), however the exact biological basis for this association remains poorly understood.

One hypothesis linking the two is based on the observation that Beijing strains have a higher

basal mutation rate. Ford et al. (2013) used Luria-Delbrück fluctuation assays to measure the frequency of resistance to different antibiotics in isolates belonging to Lineage 2 and Lineage 4. They found that Lineage 2 strains acquired resistance to RIF, INH and ethambutol at much higher rates than Lineage 4 strains. Using their measurements to parameterize a stochastic MDR-TB evolution model, they propose that individuals infected with Lineage 2 strains have an increased risk (∼22-fold higher) of harboring MDR-TB than individuals infected with a Lineage 4 strain. The importance of mutation rates in the emergence of resistance is obvious: more mutations mean a higher probability of resistance mutations. Therefore, strains from the Beijing family with a higher basal mutation rate will have an increased chance of acquiring resistance, but also mutations that facilitate the acquisition of resistance (for example, those affecting efflux pumps), including compensatory mutations that may enhance their fitness in vivo mitigating the cost of resistance. The net results are that Beijing strains may be more likely to become resistant without loss of fitness. The association between resistance and higher mutation rates has been observed in other bacteria (Denamur and Matic 2006). However, as we alluded to earlier, most mutations are deleterious and therefore a mutation rate that is too high can have a net negative effect, particularly when considering the constrained evolutionary space, within which drug resistant strains operate. Therefore, a high in vitro frequency of resistance does not necessarily translate into frequent emergence of resistance in the clinic, especially if the associated fitness cost is sufficient to prevent the mutants to establish in the population (Trauner et al. 2014). Furthermore, in vitro models do not necessarily recapitulate in vivo environments; as a result there might be factors specific to the in vitro or in vivo environment that could drive mutagenesis – see Chap. 13 (Denamur and Matic 2006; Rock et al. 2015). An example of this disparity can be found in the fact that some resistance mutations appear to be lineage-specific (Sandgren et al. 2009; Comas et al. 2010; Spies et al. 2011) or more likely to be present in certain MTBC lineages; several reports associate Beijing to specific INH resistant mutation with no fitness cost (Gagneux et al. 2006; Fenner et al. 2012; Feuerriegel et al. 2014).

Another explanation for the association between Beijing and resistance comes from epidemiological studies that suggest that Lineage 2/Beijing strains have a higher ability to transmit. In fact, the genetic background of strains has been observed to affect transmissibility in several studies (see Chap. 5, Baker et al. 2004; Gagneux et al. 2006; Van Doorn et al. 2006; Fenner et al. 2012). For example, molecular epidemiological studies in China found that Beijing family strains were more likely to be in a genotype cluster than non-Beijing strains (Yang et al. 2015; Hu et al. 2016). In their multi-center population-based study, which included a large collection of isolates, Hu et al. also found that the Beijing family represented over 90.7% of MDR isolates; this percentage increased to around 95% among pre-XDR and XDR clinical isolates. The observation that Beijing isolates are associated with accelerated progression to active TB (De Jong et al. 2008), and the fact that they often seem to cause drug-resistant outbreaks could be part of the explanation for the relatively recent and ongoing global expansion of the lineage. On the other hand, the expansion of may be coincidental and linked to human demographic factors (Merker et al. 2015; Luo et al. 2015).

According to all these studies, the association of the Beijing genotype and resistance is likely to be the conflation of different factors: higher basal mutation rate, accelerated progression to disease, increased transmission, favorable host demographics. However, it is important to consider that all these studies were performed in different settings, and we may therefore be combining disparate information leading to over-generalizations. A study from Japan and Taiwan illustrates this point because it found that specific Beijing sub-lineages differed in their association with resistance (Yang et al. 2012). Similarly, two studies associated the higher transmission rate with a specific sub-lineage of the Beijing family (Hu et al. 2016; Yuan et al. 2016).

14.5 In Vivo Fitness of MDR-TB

Because resistance in MTBC occurs through chromosomal mutations, MDR-TB and XDR-TB must emerge sequentially, placing each acquired resistance mutation (beyond the first one) in the context of pre-existing resistance. This poses a challenge to MTBC because gain of DR mutations is normally accompanied by a decrease in the rate of bacterial growth (Trauner et al. 2014). In other words, DR mutations are said to have a "fitness cost" in the absence of antibiotic. Therefore sequentially adding costly mutations makes the problem of adaptation progressively harder. Furthermore, resistance in vitro is more easily attained than resistance in patients. Darwinian fitness in the context of communicable diseases is not defined only as rate of growth but rather as an infectious agent's ability to survive, reproduce and transmit.

The cost of drug resistance in MTBC was established already by studies from the early 1950s, which showed a reduced pathogenicity of isoniazid-resistant tubercle bacilli compared to drug-sensitive bacteria when inoculated into guinea pigs (Da 1954). This, and similar later studies (As 2002), lay the foundations for the hypothesis that drug-resistant MTBC strains show a reduced "in vivo fitness". As a result DR-TB should be less transmissible and therefore unlikely to spread successfully in immunocompetent human populations. This hypothesis, however, proved misleading. It is now clear that resistance mutations inflict different levels of fitness-cost in absence of the drug, with some showing mild or no detrimental effects. In MTBC, such "low-cost mutations" also account for the majority of the resistance detected in clinical isolates. Thus, strains harboring low-cost mutations are positively selected for among DR strains either within the bacterial population infecting a host and/or as a result of variable transmission efficiency, therefore between hosts. In addition, compensatory evolution seems to be a key driver ensuring that fitness is restored in strains with multiple costly DR mutations (Comas et al. 2012; De Vos et al. 2013). So, while there is a clear cost associated with the initial emergence of resistance, the true impact on MTBC can be successfully mitigated by selective processes acting on the bacteria.

For a long time, it was not well understood which bacterial factors, beyond resistance mutations, facilitate the emergence of MDR-TB. Molecular epidemiological studies carried out in different settings, covering the global diversity of MTBC and focused on the evolution of MDR-TB identified the relative fitness of drug-resistant strains as one of the most important determinants of MDR-TB spread and epidemic size. Importantly, the use of whole genome sequencing (WGS) to characterize community outbreaks of MDR-TB lead to the conclusion that MDR-TB is more frequently caused by transmission over de novo acquisition. For example, Casali et al. propose a plausible evolution of a clade of Beijing TB strains in the Russian Federation from mono-resistant to XDR-TB with no apparent loss of transmissibility (Casali et al. 2014). Similarly, a genomic study investigating strains that gave rise to the first known outbreak of XDR-TB, points to transmission as a driving force behind the emergence of DR in South Africa: data from the Kwazulu-Natal province appear to support person to person spread of MDR- and XDR-TB, showing that resistant strains are sufficiently fit to transmit and cause morbidity in both immunocompetent and immunosuppressed persons (Cohen et al. 2015). An analogous study in Argentina, reconstructed the past demography and timeline of acquisition of DR-TB mutations by generating WGS for 252 clinical isolates collected from a single large outbreak in Buenos Aires. The study shows efficient transmission of MDR-TB strains while there is a DR expansion to become XDR-TB, evidencing direct transmission as a significant contributor to the burden of XDR-TB (Eldholm et al. 2015). In addition to molecular epidemiology, modeling studies supported by the WHO and aimed at estimating the relative contribution of transmission and de novo acquisition of resistance reach similar conclusions. Authors estimate that 96% of all incident MDR-TB cases were due to transmission (Zignol et al. 2016).

In summary, MDR-TB has emerged in several settings globally. While the "in vivo fitness" of resistant clones varies, we have witnessed the emergence of a number of successful strains. As a result, transmission poses a much greater risk than de novo emergence, pointing to the fact that clinically evolved drug-resistant strains can successfully overcome the genetic and metabolic burden of resistance. More population-based studies are needed to identify bacterial factors that lead to the generation of successful clones. One important consequence of such studies would be the ability to identify successful MDR-TB genotypes, which could pose a significant problem to the future eradication of the DR-TB epidemic.

14.6 Successful Clones of MDR-TB: The Role of Epistasis

The amount of WGS information that continues to be made available is setting the foundations for the definition of successful MDR genotypes. It is apparent that there seems to be a clear preference for which combination of DR mutations shows a higher "*in vivo* fitness" and what order these DR mutations were acquired. This genetic dimension influencing the path towards MDR/XDR-TB can be formalized in the concept of epistasis. Epistasis is defined as the effect of the interaction between two or more mutations on an organism's phenotype. Or said differently, epistasis describes the phenomenon where the same mutation has different effects depending on the presence of other mutations. For example, the cost of a given drug resistance mutation could differ depending on whether or not a compensatory mutation is present in the genome. Similarly, different DR-conferring mutations can interact epistatically; this happens when strains resistant to two drugs had a higher fitness than at least one of the corresponding single DR mutants (Borrell et al. 2013).

Epistasis represents one way in which genetic diversity modulates the evolution of DR. Epistatic interactions occurring between DR mutations and pre-existing genomic differences, can lead the same resistance mutation to have a different cost depending on the strain it is in (Vogwill et al. 2016; Gagneux et al. 2006, Brandis and Hughes 2013). As mentioned above, the "Beijing family" of MTBC has often been associated with DR. The association between specific MTBC lineages and DR could result from an increased ability to tolerate a larger range of resistance mutations, either due to a higher baseline fitness, or more favorable epistatic interactions between the strain genetic background and DR mutations. This idea is supported by the presence of various lineage-specific non-synonymous nucleotide substitutions in genes associated with DR (Feuerriegel et al. 2014). Thus, it is possible that the pre-existence of non-resistance mutations in DR-associated genes could modulate the fitness effects of subsequently acquired resistance-conferring mutations in these genes.

Epistatic interactions may also determine the order in which DR-conferring mutations and associated compensatory mutations are acquired over time (Muller et al. 2013a). The need to harbor multiple resistance mutations places a selective pressure on MTBC to minimize the impact of each acquired mutation. To some extent, this selection limits the number of available evolutionary trajectories resulting in recurring patterns of resistance in the clinic (Muller et al. 2013b; Eldholm et al. 2015). On the molecular level, this means an enrichment of a set of low-cost resistance mutations followed by the acquisition of secondary, compensatory mutations that decrease the fitness impact of resistance determinants. Epistatic interactions may determine the optimal mutational pathway; for resistance to meaningfully arise, it needs to be accessible within a small number of mutational steps – preferably a single point mutation. This is indeed the case for rifampicin resistance in MTBC where the majority of clinically relevant resistance mutations affect a single base resulting in a non-synonymous amino acid substitution (Campbell et al. 2001; Zhang and Yew 2015). In the short and medium runs, this means that MTBC will accumulate the best mutations it can access, which may not be the absolutely optimal combination of resistance

determinants. Over time, this may change, as MTBC is able to explore its mutational space further and generate less costly combinations of mutations.

An analogy that is often used to illustrate the concept of evolutionary trajectories is that of a topologically diverse landscape where fitness is represented by elevation and movement through it mediated by mutations. The accumulation of costly mutations leads to a descent into fitness valleys while combinations of no-cost or beneficial mutations can be thought of as a climb towards accessible peaks. The optimal solution to the problem of fitness cost minimization will depend on the starting point within this landscape, and therefore, on the pre-existing genetic context. By extending the geographical analogy, we can say that genetic diversity influences, together with the environment, the overall shape of the fitness landscape, defines the position of an individual within it, and the possible trajectories an individual can pursue. Evolutionary trajectories are expected to differ considerably between drug-sensitive TB and MDR-TB strains.

14.7 Overcoming the Cost of Drug Resistance: The Case of Rifampicin (RIF)

So far, we have seen how the genetic background of MTBC influences the emergence of DR in the clinic: it can influence the emergence of resistance, it can impose a preference for specific resistance mutations and possibly define the overall *in clinico* fitness of drug-resistant strains. Next, we will use the specific example of RIF resistance to build on the importance of relative fitness as a driving force of strain evolution. As described earlier, RIF inhibits transcription by binding to RNAP and interfering with the early steps of transcript elongation (Campbell et al. 2001). The core RNAP enzyme is a protein complex composed of five subunits: $\alpha_2\beta\beta'\omega$, which are encoded by *rpoA, rpoB, rpoC* and *rpoZ,* respectively. Mutations conferring resistance to RIF occur away from the active site and mostly in the part of the β subunit of RNAP that is involved directly in drug binding (Campbell et al. 2001). This section of RpoB is defined as the RIF-Resistance Determining Region (RRDR) and in MTBC covers residues 426–452, which are sometimes referred to, using *Escherichia coli* residue numbering (507–533). Approximately 95% of all clinically observed RIF-resistance in MTBC can be attributed to mutations in the RRDR. As a result, the majority of molecular diagnostic tools focus on detecting mutations in this locus to identify RIF-resistant MTBC in the clinic. While many mutations in the RRDR can confer drug resistance, their relative frequency in the clinic can vary widely. A likely reason for these is the disparity in the fitness of drug-resistant strains in the absence of drug (Mariam et al. 2004; Gagneux et al. 2006; Borrell and Gagneux 2009; Song et al. 2014), a concept that can be explained using fitness landscapes.

14.7.1 The Dynamic Fitness Landscape

Recall the geographical analogy of bacterial fitness: we said that sequence context defines the topology of the terrain and the elevation of a bacterium corresponds to its ability to grow and spread. However, unlike a real landscape, the fitness landscape is dynamic and constantly redrawn by the environment the bacterium finds itself in: what is advantageous in one setting may be deleterious in another. Selective forces, such as antibiotic treatment or the challenges of the host environment, can be thought of as rivers that wash away all individuals that are below the water line in any given landscape. Given the dynamic nature of the contours, this can mean that the same genotype can thrive in one setting and be flushed away once the landscape changes with the environmental conditions (see Fig. 14.1). If we place a whole bacterial population in a fitness landscape of infection (see Fig. 14.1a, b), the majority of individuals will be grouped close to the starting position (most probably an accessible optimal fitness – a local maximum), with a minority scattered further afield due to the presence of mutations. If we now treat this population with RIF, the landscape will change to favour drug resistance, and suddenly almost all of the

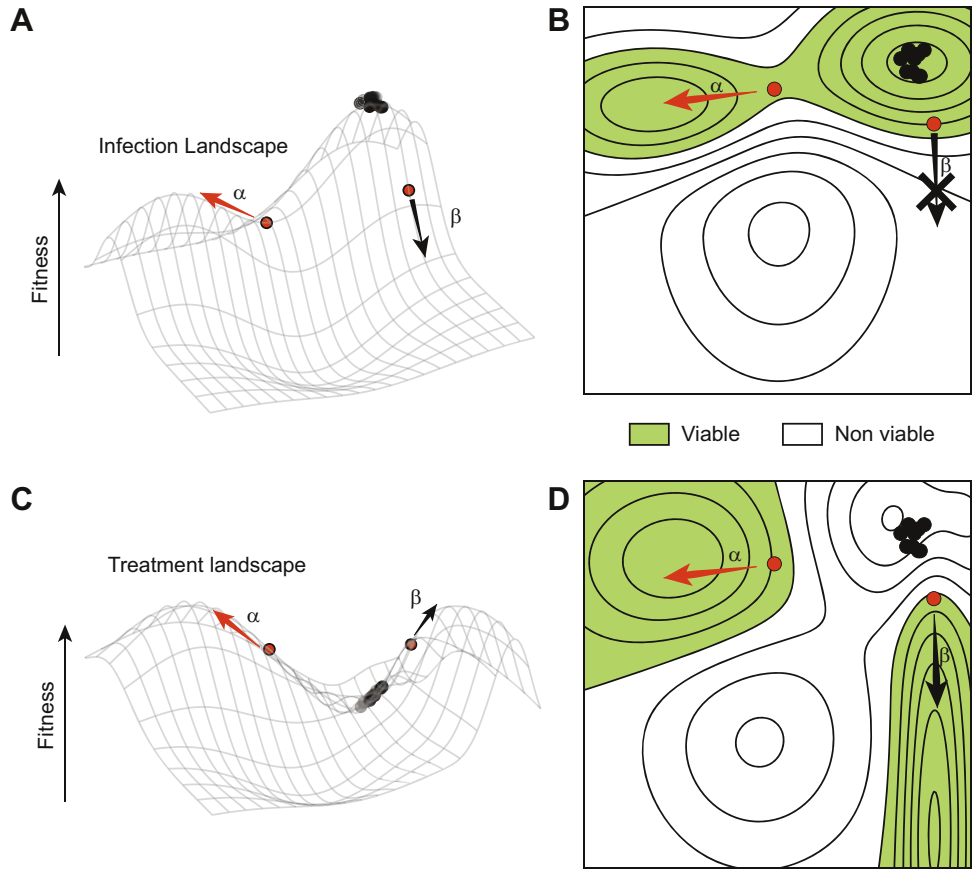

Fig. 14.1 Geographical representation of bacterial fitness. Bacterial fitness in a given environment can be illustrated using a geographical analogy. Imagine a set of bacterial cells scattered across a landscape. The grid-position of each bacterium is defined by its genotype, while the elevation reflects its ability to thrive in a given environment – its fitness. Varying the environment of a bacterium might change the overall landscape but not the position of a bacterium within it, while the acquisition of mutations will move the bacterium through a given landscape. Panels **a**, **b** provide a 3D projection and a contour map respectively of a hypothetical landscape pertinent to the infection context of MTBC. The *green* shaded area of the contour map indicates regions of sufficient overall fitness to support growth. Consider a set of drug susceptible strains (*black circles*) and a pair of dissimilar resistant strains (*red circles*). Drug susceptible strains are clustered around the highest point of the infection landscape, showing that they are well adapted to the human host with little room left for improvement. On the other hand, strains harboring resistance mutations show a marked decrease in fitness. These bacteria are at a disadvantage compared to their fitter cousins and are therefore unlikely to be successful in the long run. However, upon the introduction of an antibiotic (panels **c**, **d**) the landscape changes dramatically and susceptible strains are no longer able to survive, leaving resistant strains free to proliferate. Proliferation leads to the acquisition of mutations and therefore movement through the landscape (paths α and β). Because bacteria need to balance the pressures imposed by both: the host environment and antibiotics, the only successful evolutionary strategy is one that leads to higher elevation in both contexts (path α). The bacterium on path β cannot improve its positions in both landscapes simultaneously, and is therefore unlikely to prevail

individuals will not be high enough to survive (see Fig. 14.1c, d). For the handful that do, they will almost invariably find themselves lower than the population starting position once the drug is removed and the landscape returns to its previous state. In order, therefore, to be able to not only survive treatment, but to continue to proliferate, bacteria must move through the landscape further by acquiring additional, compensatory, mutations (see Fig. 14.1, arrows). The march of drug-resistant strains across the landscape is constrained by the fact that RIF treatment will

continue to occur in the future. Therefore, drug-resistant strains must reach higher grounds both in the context of the "infection landscape" and in the "resistance landscape". In the figure, we illustrate these using two resistant strains (red dots). One can follow the trajectory which leads to an improved fitness in both landscapes (α, red arrows), while the other can follow only the trajectory that leads to improved fitness in the presence of RIF at the cost of fitness in the absence of drug (β, black arrows). It follows that α will give rise to transmissible drug-resistant bacteria, while β will not. In reality, all of this is complicated by the fact that MTBC will be faced with multiple drugs at the same time. The aim of combination therapy is precisely that of minimizing the probability of resistance emerging (Bonhoeffer et al. 1997), therefore stemming all possible evolutionary trajectories by removing the bacteria that may follow them. But even combination therapy is not perfect, and issues such as imperfect drug penetration can provide a sufficient foothold for drug-resistant clones to emerge and transmit (Moreno-Gamez et al. 2015), so that the general principles described so far remain valid.

How does this abstraction play out in reality? The de facto shape of a landscape is hard to describe in a meaningful way using experimental approaches, because we can only ever recapitulate known parts of the infection environment. Therefore, predicting all available trajectories may not be possible. However, by combining experimental data with molecular epidemiological findings, we should be able to recapitulate important aspects of the landscape, such as the availability of α-type trajectories and key mutations that define them. It is not surprising that modeling studies identify the distribution of fitness costs of drug resistance-conferring mutations as a key determinant for the future burden of MDR-TB (Knight et al. 2015). As we know, drug treatment is very effective at clearing infection; however it is also the driver of the selection for resistance. As discussed earlier, drug-resistant MTBC can be less transmissible than fully susceptible strains (Grandjean et al. 2015), however in some cases, drug-resistant strains can grow just as well, if not even faster, than their susceptible ancestors (Gagneux et al. 2006; Comas et al. 2012), suggesting that resistant strains can evolve to be just as transmissible as susceptible strains. Clinically, successful drug-resistant strains can therefore be thought to follow a trajectory similar to α in Fig. 14.1: they are able to both transmit and withstand RIF treatment. Does it then follow that other RIF-resistance mutations are confined to β-type trajectories? What do we know about the internal perturbation imposed by each mutation that may explain the differences in clinical success?

14.7.2 Biological Basis of the Cost of RIF-Resistance and Its Compensation

The exact mechanism by which mutations in the RRDR impart a fitness cost on bacteria is unclear. However, given the central role of RNAP in gene expression its mutation is likely to have complex consequences (see (Koch et al. 2014) for a more detailed discussion). In the case of *E. coli, Pseudomonas aeruginosa and Mycobacterium smegmatis*, there is a clear correlation between fitness cost and transcription rate (Qi et al. 2014; Reynolds 2000; Song et al. 2014). This relationship reflects the need for optimal synthesis of stable RNA molecules such as ribosomal RNA (rRNA) and transfer RNA (tRNA) to support maximal growth; a factor that becomes rate limiting in RIF-resistant mutants. There is also clear evidence of gene expression differences between drug-susceptible strains and RNAP mutants (Bisson et al. 2012; De Knegt et al. 2013). These impact biosynthetic processes in the bacterium resulting in a perturbation of the metabolic steady state that can affect a range of processes from cell wall biosynthesis to iron acquisition (Du Preez and Loots 2012; Lahiri et al. 2016). In some organisms, *rpoB* mutations can lead to the expression of otherwise silent (cryptic) genes (Ochi et al. 2014) that result in the production of metabolites that were previously absent from the cell. Overall, such metabolic changes could impose a biosynthetic cost and shift steady state concentrations of metabolic intermediates into a sub-optimal equilibrium. Both such effects could explain a net cost of RIF-resistance. It is likely

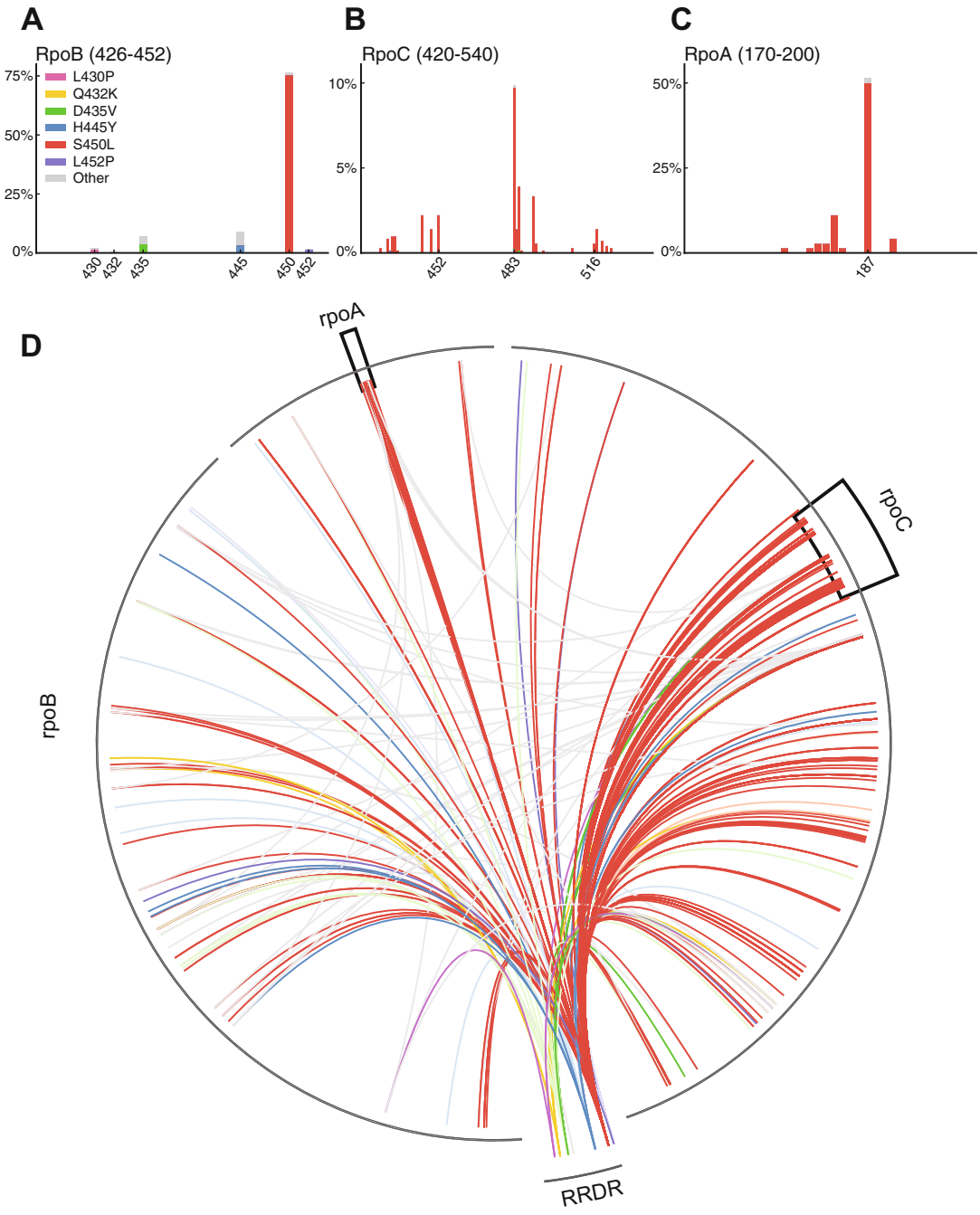

Fig. 14.2 Genetic diversity of rifampicin resistance. Collating data regarding the co-occurrence of RNAP mutations across several studies that include globally well distributed populations (Comas et al. 2012; De Vos et al. 2013; Farhat et al. 2013; Lanzas et al. 2013; Merker et al. 2013; Casali et al. 2014; Song et al. 2014; Bloemberg et al. 2015; Eldholm et al. 2015) reveals a clear pattern in the diversity of rifampicin-resistant strains. In panel **a** we show the frequency distribution of mutations in the rifampicin resistance determining region (RRDR) which confer rifampicin resistance in the clinic. The most frequent substitution at each position, within our dataset, was attributed a colour that is then used throughout the figure. By far the most frequent resistance mutation present in almost 75% of all strains is S450L. Panels **b**, **c** show a histogram of putative compensatory mutations that occur

that the true cost of resistance is a composite of several factors. To some extent, studies focused on the evolution of drug-resistant strains have begun to address the question of fitness cost by looking for signatures of positive selection (Farhat et al. 2013; Zhang et al. 2013). The picture to emerge from these studies shows a complex situation with the likely involvement of a multitude of players. Perhaps the clearest genetic signal for compensation points to the cell wall and associated macromolecules (Farhat et al. 2013) as well as RNAP as hotspots for adaptation (Comas et al. 2012; Farhat et al. 2013; Zhang et al. 2013; Song et al. 2014). Parsing out the actual contributions to compensation will require further work, nonetheless, several studies identified a link between the presence of additional mutations in RNAP and improved transmission of MDR-TB (De Vos et al. 2013; Casali et al. 2014; Song et al. 2014; Cohen et al. 2015). This suggests that RIF-resistant MTBC improves its position in the "infection landscape" by following a trajectory that includes acquiring additional, compensatory RNAP mutations.

As with the cost of RIF-resistance, the mechanism of compensation mediated by secondary RNAP mutations in MTBC remains unknown. In *Mycobacterium smegmatis* and several pseudomonads, compensatory mutations restore RNAP processivity (Qi et al. 2014; Song et al. 2014; Vogwill et al. 2016). In *P. aeruginosa*, they also modify the expression profile of resistant strains to one that more closely resembles their susceptible ancestor (Qi et al. 2014). Studies from *Salmonella typhymurium* suggest that this is mediated by influencing the stability of RNAP, restoring the disrupted structural properties of the RRDR and modulating the way RNAP interacts with both RNA and ribonucleotide triphosphates (Brandis et al. 2012). Knowledge of these processes in MTBC is limited to genetic studies. Early reports suggested that many of the compensatory mutations can be found at the interaction interface between α, β and β′ subunits potentially modulating the stability of the protein complex (Comas et al. 2012). Song et al. performed perhaps the most detailed analysis of compensatory mutations in MTBC to date (Song et al. 2014). They draw several important conclusions: firstly they single out the "Double psi-beta barrel" DPBB domain of the β′ subunit, which contains the active site and includes the RNA exit tunnel as an important region for the accumulation of compensatory mutations pertinent to mitigating the fitness cost of S450L RpoB mutations. They propose that S450L may interfere with the translocation of RNA during transcription, and that the compensatory substitutions modify RNAP to facilitate the extrusion of nascent RNA. They also point out that compensatory mutations exhibit some degree of allele specificity, as they did not observe an accumulation of mutations in the DPBB domain for other RRDR mutations. We complemented their analysis by collating data on RNAP mutations from several recent studies addressing the genetic diversity of drug-resistant MTBC in multiple global hotspots of MDR evolution (Comas et al. 2012; De Vos et al. 2013; Muller et al. 2013b; Farhat et al. 2013; Lanzas et al. 2013; Casali et al. 2014; Song et al. 2014; Bloemberg et al. 2015; Eldholm et al. 2015; Merker et al. 2015) (see Fig. 14.2). The first striking feature is that, as reported previously (Song et al. 2014; Trauner et al. 2014; Cohen et al. 2015; Eldholm et al. 2015), the greatest number and diversity of compensatory

Fig. 14.2 (continued) in defined regions of RpoC (residues 420–540) and RpoA (residues 170–200) respectively. The histograms are colour-coded to reflect the identity rifampicin-resistance mutations and show that these regions are overwhelmingly mutated in strains carrying the S450L resistance mutation. The presence of multiple RNAP mutations within any given strain is depicted (Krzywinski et al. 2009) in Panel (**d**). We drew a link between residues of RNAP subunits that occur within the same strain. The links are colour-coded to represent different RRDR mutations, as delineated in panel (**a**). For the sake of clarity, we show all the RRDR mutations separately and not as part of RpoB. The rectangles highlight regions containing an exceptionally high number of mutations and correspond to histograms in panels (**b**, **c**). In total we analysed 1488 rifampicin-resistant strains, of these 73 contained additional mutations in RpoA and 729 in RpoC. Adapted from Gygli et al. 2017

mutations can be observed in strains with RpoB S450L. Nonetheless, by compiling data from several studies we can see that compensatory mutations occur in several different RIF-resistant backgrounds. As expected, we see clear evidence for the clustering of compensatory mutations in specific parts of RNAP. Specifically, we see a wealth of mutations in the DPBB domain of β′ (residues 423–565) and the proximal region of α (residues 170–200) (Mechold et al. 2013). We plotted the frequency of mutations in these regions across our samples (Fig. 14.2a–c): 265 of a total of 729 *rpoC* mutations fell into the DPBB domain and 56 of the 73 *rpoA* mutations were within the DPBB-proximal region of α. Consistent with the assertions of Song et al. we see that there is an overwhelming over-representation of putative compensatory mutations linked to RpoB S450L in these regions. It seems that residues at positions 483 and 485 of β' and the residue at position 187 of α are particularly frequent in the clinic. Furthermore, no other obvious patterns emerged; implying that other resistance mutations have more diverse ways of achieving compensation or perhaps that further studies are needed to strengthen the genetic signal for less prevalent RIF resistance-conferring mutations. One aspect of RNAP mutations that has not been extensively addressed yet is the presence of a number of mutations in β that co-occur with resistance mutations. Unlike mutations in the DPBB and α, these are not dominated by RpoB S450L and may provide options for compensation of other RIF-resistance mutations. Another open question that requires attention is whether or not the presence of secondary mutations in RNAP impacts the minimum inhibitory concentration (MIC) of RIF.

14.7.3 The Impact of Genetic Diversity on RIF-Resistance

As we already saw, RIF-resistance mutations differ in their net fitness effect. Furthermore, as more and more data become available it is becoming apparent that their path to compensation may also differ. For example, in some studies there was no discernable correlation between the acquisition of additional RNAP mutations and improved transmission of resistant strains (Lanzas et al. 2013; Eldholm et al. 2015). This disparity of roles can result from observing populations at different evolutionary times; for example long standing transmission of drug-resistant strains may be more likely to associate with compensation than newer occurrences, but it may also speak to the fact that not all strains can (or have a need to) compensate the cost of RIF resistance. One ramification of this observation is that the same mutation could reasonably have a different fitness cost depending on the genetic context it enters – a phenomenon that has been described for RIF-resistant pseudomonas (Vogwill et al. 2016). Given the role of additional RNAP mutations in the context of compensation, it is possible that RNAP mutations that were accumulated during the evolution of MTBC and do not confer resistance may influence the cost of resistance. We were able to identify RNAP mutations carried by several strains that did not have any mutations in the RRDR; by interrogating a previously published collection of highly diverse MTBC strains (Bos et al. 2014) (Table 14.1). We recorded both non-synonymous and synonymous mutations since both can potentially have an effect on RNAP. The impact of the former is straightforward – modifying an amino acid can have any number of biochemical consequences, however the latter may also influence RNAP by modifying the type of codon used resulting in a measurable phenotype (Agashe et al. 2013). Lineage 6 strains have three non-synonymous substitutions in β with a subset of strains having also non-synonymous mutations in α. Lineage 6 strains are known to be less likely to evolve drug resistance in the clinic (see Chap. 6 and Homolka et al. 2008; Traore et al. 2012; Winglee et al. 2016). Linking these two observations, it is possible to hypothesize that pre-existing mutations in RNAP of Lineage 6 strains increases the fitness cost of RIF-resistance therefore decreasing the probability of its emergence. Similarly, some Lineage 4 strains carry the same non-synonymous substitution in β′ (RpoC G594E) that can be found in one of the

Table 14.1 Natural variation in RNA polymerase sequence in MTBC

	Mutation	Background
α (RpoA, Rv3457c)	*Synonymous*	
	Q5Q	Lineage 5 (subset)
	G156G	Lineage 7
	Nonsynonymous	
	T271I	Lineage 6 (subset)
	E319K	Lineage 4 (subset)
β (RpoB, Rv0667)	*Synonymous*	
	G876G	Lineage 3
	A1075A	Lineage 4
	I1158I	Lineage 5 (subset)
	Nonsynonymous	
	A10G	Lineage 3 (subset)
	T350I	Lineage 6 (subset)
	S388L	Lineage 6
	E639D	Lineage 6
β' (RpoC, Rv0668)	*Synonymous*	
	P54P	Lineage 1 (subset)
	R173R	Lineage 1
	R480R	Lineage 1 (subset)
	A542A	Lineage 4 (subset)
	G584G	Lineage 2 (subset)
	A954A	Lineage 6 (subset)
	Nonsynonymous	
	A172V	Lineage 1
	E215A	Lineage 7
	D271G	Lineage 3 (subset)
	G594E	Lineage 4 (subset)
	P601L	Lineage 1 (subset)
	A621T	Lineage 1 (subset)
	E1196K	Lineage 7

settings where the presence of additional mutations in RNAP does not correlate with increased transmission (Eldholm et al. 2015), raising the possibility that RpoC G594E mitigates the cost of RIF-resistance. We did not observe any RNAP mutations in Lineage 2 strains included in the collection.

Genetic background has been shown to bias the emergence of resistance mutations in other bacteria (Vogwill et al. 2014), but these hypotheses need to be substantiated by further experimentation in MTBC. Nonetheless, the true impact of genetic background may be limited. We have already discussed that successfully transmitted drug-resistant strains of MTBC are relatively homogeneous from the perspective of resistance mutations. Convergent genotypes emerged in geographically distinct locations – for example in the data collated for this chapter, RpoB S450L accounts for more than 75% of all RIF-resistance (Fig. 14.2a). Intuitively, we would expect that the path to compensation is driven by the genetic context and is likely to be idiosyncratic to a particular genetic starting point, or said differently, if you know the background mutation you should be able to predict the path to compensation. To some extent this is true, there may be measurable epistasis between resistance mutations, genetic background and compensatory mutations (Vogwill et al. 2016; Song et al. 2014), but unfortunately, the full picture is more complex. Studies in budding yeast (Kryazhimskiy et al. 2014), show that while the overall consequence of compensation is predictable, the genetic trajectory leading up to it may not be. This fact may complicate the analysis and limit the knowledge gleaned from genomic studies, but more importantly, it increases the difficulty of trying to infer the potential for a strain to pose a significant public health threat based solely on its genotype (Trauner et al. 2014).

14.8 Concluding Remarks

In recent years, we have seen next generation sequencing become a workhorse for two fields that are fundamentally intertwined: molecular epidemiology and evolution. Combining genomic insights with decades of research into drug resistance, we are now able to trace the trajectory of drug-resistant MTBC better than ever before. Cross-referencing experimental and clinical studies has led to the recognition of the fact that: while drug resistance can be costly to begin with, MTBC can evolve to compensate for this effect, giving rise to highly transmissible drug-resistant strains. This conclusion is only one manifestation of the consequences of treating drug resistance as a process rather than an end-point – princi-

ples of biochemistry, molecular and evolutionary biology underpin the dynamics of bacterial populations in the most relevant of scenarios: human health. Nevertheless, there are still important gaps in our knowledge. For example, what lies at the core of the correlation between virulence and competitive fitness? It is clear that genetic determinants are at the heart of it, but to what extent is this driven by lineage-specific, phylogenetic mutations? Can we predict what constitutes a dangerous MTBC genotype? Can this information be leveraged in a meaningful way to inform health policy or community-wide interventions? We therefore need population-based studies to build on our understanding, and determine what the real impact of a specific genetic background is for the success of an MTBC clone in the clinical setting.

References

Agashe D, Martinez-Gomez NC, Drummond DA, Marx CJ (2013) Good codons, bad transcript: large reductions in gene expression and fitness arising from synonymous mutations in a key enzyme. Mol Biol Evol 30:549–560

As P (2002) Effect of Katg mutations on the virulence of mycobacterium tuberculosis and the implication for transmission in humans. Infect Immun 70:4955–4960

Baker L, Brown T, Maiden MC, Drobniewski F (2004) Silent nucleotide polymorphisms and a phylogeny for mycobacterium tuberculosis. Emerg Infect Dis 10:1568–1577

Bisson GP, Mehaffy C, Broeckling C, Prenni J, Rifat D, Lun DS, Burgos M, Weissman D, Karakousis PC, Dobos K (2012) Upregulation of the phthiocerol dimycocerosate biosynthetic pathway by rifampin-resistant, rpoB mutant Mycobacterium tuberculosis. J Bacteriol 194:6441–6452

Bloemberg GV, Keller PM, Stucki D, Trauner A, Borrell S, Latshang T, Coscollà M, Rothe T, Hömke R, Ritter C et al (2015) Acquired resistance to bedaquiline and delamanid in therapy for tuberculosis. N Engl J Med 373:1986–1988

Blower SM, Chou T (2004) Modeling the emergence of the 'hot zones': tuberculosis and the amplification dynamics of drug resistance. Nat Med 10: 1111–1116

Bonhoeffer S, Lipsitch M, Levin BR (1997) Evaluating treatment protocols to prevent antibiotic resistance. Proc Natl Acad Sci U S A 94:12106–12111

Borrell S, Gagneux S (2009) Infectiousness, reproductive fitness and evolution of drug-resistant Mycobacterium tuberculosis [State of the art]. Int J Tuberc Lung Dis 13:1456–1466

Borrell S, Teo Y, Giardina F, Streicher EM, Klopper M, Feldmann J, Muller B, Victor TC, Gagneux S (2013) Epistasis between antibiotic resistance mutations drives the evolution of extensively drug-resistant tuberculosis. Evol Med Public Health 2013:65–74

Bos KI, Harkins KM, Herbig A, Coscollà M, Weber N, Comas I, Forrest SA, Bryant JM, Harris SR, Schuenemann VJ et al (2014) Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature 514:494–497

Brandis G, Hughes D (2013) Genetic characterization of compensatory evolution in strains carrying rpoB Ser531Leu, the rifampicin resistance mutation most frequently found in clinical isolates. J Antimicrob Chemother 68:2493–2497

Brandis G, Wrande M, Liljas L, Hughes D (2012) Fitness-compensatory mutations in rifampicin-resistant RNA polymerase. Mol Microbiol 85:142–151

Campbell EA, Korzheva N, Mustaev A, Murakami K, Nair S, Goldfarb A, Darst SA (2001) Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. Cell 104:901–912

Casali N, Nikolayevskyy V, Balabanova Y, Harris SR, Ignatyeva O, Kontsevaya I, Corander J, Bryant J, Parkhill J, Nejentsev S et al (2014) Evolution and transmission of drug resistant tuberculosis in a Russian population. Nat Genet 46:279

Cohen T, Jenkins HE, Lu C, Mclaughlin M, Floyd K, Zignol M (2014) On the spread and control of MDR-TB epidemics: an examination of trends in anti-tuberculosis drug resistance surveillance data. Drug Resist Updat 17:105–123

Cohen KA, Abeel T, Mcguire AM, Desjardins CA, Munsamy V, Shea TP, Walker BJ, Bantubani N, Almeida DV, Alvarado L et al (2015) Evolution of extensively drug-resistant tuberculosis over four decades: whole genome sequencing and dating analysis of Mycobacterium tuberculosis isolates from KwaZulu-Natal. PLoS Med 12:e1001880

Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, Kremer K, Ernst JD, Gagneux S (2010) Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved. Nat Genet 42: 498–503

Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, Galagan J, Niemann S, Gagneux S (2012) Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. Nat Genet 44:106–110

Coscolla M, Gagneux S (2014) Consequences of genomic diversity in Mycobacterium tuberculosis. Semin Immunol 26:431–444

Da M (1954) Tubercle bacilli resistant to isoniazid: virulence and response to treatment with isoniazid in guinea-pigs. Br Med J 1:128–130

Dalal A, Pawaskar A, Das M, Desai R, Prabhudesai P, Chhajed P, Rajan S, Reddy D, Babu S, Jayalakshmi TK,

Saranchuk P, Rodrigues C, Isaakidis P (2015) Resistance patterns among multidrug-resistant tuberculosis patients in greater metropolitan Mumbai: trends over time. PLoS One 10:e0116798

Dalton T, Cegielski P, Akksilp S, Asencios L, Campos Caoili J, Cho SN, Erokhin VV, Ershova J, Gler MT, Kazennyy BY, Kim HJ, Kliiman K, Kurbatova E, Kvasnovsky C, Leimane V, Van Der Walt M, Via LE, Volchenkov GV, Yagui MA, Kang H, Akksilp R, Sitti W, Wattanaamornkiet W, Andreevskaya SN, Chernousova LN, Demikhova OV, Larionova EE, Smirnova TG, Vasilieva IA, Vorobyeva AV, Barry CE 3rd, Cai Y, Shamputa IC, Bayona J, Contreras C, Bonilla C, Jave O, Brand J, Lancaster J, Odendaal R, Chen MP, Diem L, Metchock B, Tan K, Taylor A, Wolfgang M, Cho E, Eum SY, Kwak HK, Lee J, Lee J, Min S, Degtyareva I, Nemtsova ES, Khorosheva T, Kyryanova EV, Egos G, Perez MT, Tupasi T, Hwang SH, Kim CK, Kim SY, Lee HJ, Kuksa L, Norvaisha I, Skenders G, Sture I, Kummik T, Kuznetsova T, Somova T, Levina K, Pariona G, Yale G, Suarez C, Valencia E, Viiklepp P (2012) Prevalence of and risk factors for resistance to second-line drugs in people with multidrug-resistant tuberculosis in eight countries: a prospective cohort study. Lancet 380:1406–1417

De Jong BC, Hill PC, Aiken A, Awine T, Antonio M, Adetifa IM, Jackson-Sillah DJ, Fox A, Deriemer K, Gagneux S, Borgdorff MW, Mcadam KP, Corrah T, Small PM, Adegbola RA (2008) Progression to active tuberculosis, but not transmission, varies by Mycobacterium tuberculosis lineage in The Gambia. J Infect Dis 198:1037–1043

De Knegt GJ, Bruning O, Marian T, De Jong M, Van Belkum A, Endtz HP, Breit TM, Bakker-Woudenberg IAJM, De Steenwinkel JEM (2013) Rifampicin-induced transcriptome response in rifampicin-resistant Mycobacterium tuberculosis. Tuberculosis 93:96–101

De Vos M, Muller B, Borrell S, Black PA, Van Helden PD, Warren RM, Gagneux S, Victor TC (2013) Putative compensatory mutations in the rpoC gene of rifampin-resistant Mycobacterium tuberculosis are associated with ongoing transmission. Antimicrob Agents Chemother 57:827–832

Dean AS, Zignol M, Falzon D, Getahun H, Floyd K (2014) HIV and multidrug-resistant tuberculosis: overlapping epidemics. Eur Respir J 44:251–254

Denamur E, Matic I (2006) Evolution of mutation rates in bacteria. Mol Microbiol 60:820–827

Du Preez I, Loots DT (2012) Altered fatty acid metabolism due to rifampicin-resistance conferring mutations in the rpoB gene of Mycobacterium tuberculosis: mapping the potential of pharmaco-metabolomics for global health and personalized medicine. Omics: J Integr Biol 16:596–603

Eldholm V, Monteserin J, Rieux A, Lopez B, Sobkowiak B, Ritacco V, Balloux F (2015) Four decades of transmission of a multidrug-resistant Mycobacterium tuberculosis outbreak strain. Nat Commun 6

Escombe AR, Moore DA, Gilman RH, Pan W, Navincopa M, Ticona E, Martinez C, Caviedes L, Sheen P, Gonzalez A, Noakes CJ, Friedland JS, Evans CA (2008) The infectiousness of tuberculosis patients coinfected with HIV. PLoS Med 5:e188

Farhat MR, Shapiro BJ, Kieser KJ, Sultana R, Jacobson KR, Victor TC, Warren RM, Streicher EM, Calver A, Sloutsky A et al (2013) Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis. Nat Genet 45:1183–1189

Fenner L, Egger M, Bodmer T, Altpeter E, Zwahlen M, Jaton K, Pfyffer GE, Borrell S, Dubuis O, Bruderer T, Siegrist HH, Furrer H, Calmy A, Fehr J, Stalder JM, Ninet B, Bottger EC, Gagneux S (2012) Effect of mutation and genetic background on drug resistance in Mycobacterium tuberculosis. Antimicrob Agents Chemother 56:3047–3053

Feuerriegel S, Koser CU, Niemann S (2014) Phylogenetic polymorphisms in antibiotic resistance genes of the Mycobacterium tuberculosis complex. J Antimicrob Chemother 69:1205–1210

Ford CB, Shah RR, Maeda MK, Gagneux S, Murray MB, Cohen T, Johnston JC, Gardy J, Lipsitch M, Fortune SM (2013) Mycobacterium tuberculosis mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. Nat Genet 45:784–790

Gagneux S, Long CD, Small PM, Van T, Schoolnik GK, Bohannan BJM (2006) The competitive cost of antibiotic resistance in Mycobacterium tuberculosis. Science 312:1944–1946

Grandjean L, Gilman RH, Martin L, Soto E, Castro B, Lopez S, Coronel J, Castillo E, Alarcon V, Lopez V et al (2015) Transmission of multidrug-resistant and drug-susceptible tuberculosis within households: a prospective cohort study. PLoS Med 12:e1001843

Gygli S, Borrell S, Trauner A, Gagneux S (2017) FEMS Microbiol Rev 41(3):354–373 doi: 10.1093/femsre/fux011

Hartkoorn RC, Uplekar S, COLE ST (2014) Cross-resistance between clofazimine and bedaquiline through upregulation of MmpL5 in Mycobacterium tuberculosis. Antimicrob Agents Chemother 58:2979–2981

Haver HL, Chua A, Ghode P, Lakshminarayana SB, Singhal A, Mathema B, Wintjens R, Bifani P (2015) Mutations in genes for the F420 biosynthetic pathway and a nitroreductase enzyme are the primary resistance determinants in spontaneous in vitro-selected PA-824-resistant mutants of Mycobacterium tuberculosis. Antimicrob Agents Chemother 59:5316–5323

Homolka S, Post E, Oberhauser B, George AG, Westman L, Dafae F, Rüsch-Gerdes S, Niemann S (2008) High genetic diversity among Mycobacterium tuberculosis complex strains from Sierra Leone. BMC Microbiol 8:1

Hu Y, Mathema B, Zhao Q, Rzheng X, Li D, Jiang W, Wang W, Xu B (2016) Comparison of the socio-demographic and clinical features of pulmonary TB patients infected with sub-lineages within the W-Beijing and non-Beijing Mycobacterium tuberculosis. Tuberc (Edinb) 97:18–25

Knight GM, Colijn C, Shrestha S, Fofana M, Cobelens F, White RG, Dowdy DW, Cohen T (2015) The distribution of fitness costs of resistance-conferring mutations is a key determinant for the future burden of drug-resistant tuberculosis: a model-based analysis. Clin Infect Dis 61(Suppl 3):S147–S154

Koch A, Mizrahi V, Warner DF (2014) The impact of drug resistance on Mycobacterium tuberculosis physiology: what can we learn from rifampicin? Emerg Microbes Infect 3:e17

Kryazhimskiy S, Rice DP, Jerison ER, Desai MM (2014) Global epistasis makes adaptation predictable despite sequence-level stochasticity. Science 344:1519–1522

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. Genome Res 19:1639–1645

Lahiri N, Shah RR, Layre E, Young D, Ford C, Murray MB, Fortune SM, Moody DB (2016) Rifampin resistance mutations are associated with broad chemical remodeling of mycobacterium tuberculosis. J Biol Chem 291:14248–14256. jbc-M116

Lanzas F, Karakousis PC, Sacchettini JC, Ioerger TR (2013) Multidrug-resistant tuberculosis in panama is driven by clonal expansion of a multidrug-resistant Mycobacterium tuberculosis strain related to the KZN extensively drug-resistant M. tuberculosis strain from South Africa. J Clin Microbiol 51:3277–3285

Luo T, Comas I, Luo D, Lu B, Wu J, Wei L, Yang C, Liu Q, Gan M, Sun G, Shen X, Liu F, Gagneux S, Mei J, Lan R, Wan K, Gao Q (2015) Southern East Asian origin and coexpansion of Mycobacterium tuberculosis Beijing family with Han Chinese. Proc Natl Acad Sci U S A 112:8136–8141

Magnet S, Blanchard JS (2005) Molecular insights into aminoglycoside action and resistance. Chem Rev 105:477–498

Mariam DH, Mengistu Y, Hoffner SE, Andersson DI (2004) Effect of rpoB mutations conferring rifampin resistance on fitness of Mycobacterium tuberculosis. Antimicrob Agents Chemother 48:1289–1294

Mechold U, Potrykus K, Murphy H, Murakami KS, Cashel M (2013) Differential regulation by ppGpp versus pppGpp in Escherichia coli. Nucleic Acids Res 41:6175–6189

Merker M, Kohl TA, Roetzer A, Truebe L, Richter E, Rüsch-Gerdes S, Fattorini L, Oggioni MR, Cox H, Varaine F et al (2013) Whole genome sequencing reveals complex evolution patterns of multidrug-resistant Mycobacterium tuberculosis Beijing strains in patients. PLoS One 8:e82551

Merker M, Blin C, Mona S, Duforet-Frebourg N, Lecher S, Willery E, Blum MG, Rusch-Gerdes S, Mokrousov I, Aleksic E, Allix-Beguec C, Antierens A, Augustynowicz-Kopec E, Ballif M, Barletta F, Beck HP, Barry CE 3rd, Bonnet M, Borroni E, Campos-Herrero I, Cirillo D, Cox H, Crowe S, Crudu V, Diel R, Drobniewski F, Fauville-Dufaux M, Gagneux S, Ghebremichael S, Hanekom M, Hoffner S, Jiao WW, Kalon S, Kohl TA, Kontsevaya I, Lillebaek T, Maeda S, Nikolayevskyy V, Rasmussen M, Rastogi N, Samper S, Sanchez-Padilla E, Savic B, Shamputa IC, Shen A, Sng LH, Stakenas P, Toit K, Varaine F, Vukovic D, Wahl C, Warren R, Supply P, Niemann S, Wirth T (2015) Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage. Nat Genet 47:242–249

Moreno-Gamez S, Hill AL, Rosenbloom DI, Petrov DA, Nowak MA, Pennings PS (2015) Imperfect drug penetration leads to spatial monotherapy and rapid evolution of multidrug resistance. Proc Natl Acad Sci U S A 112:E2874–E2883

Muller B, Borrell S, Rose G, Gagneux S (2013a) The heterogeneous evolution of multidrug-resistant Mycobacterium tuberculosis. Trends Genet 29:160–169

Muller B, Chihota VN, Pillay M, Klopper M, Streicher EM, Coetzee G, Trollip A, Hayes C, Bosman ME, Gey Van Pittius NC, Victor TC, Gagneux S, Van Helden PD, Warren RM (2013b) Programmatically selected multidrug-resistant strains drive the emergence of extensively drug-resistant tuberculosis in South Africa. PLoS One 8:e70919

Munsiff SS, Nivin B, Sacajiu G, Mathema B, Bifani P, Kreiswirth BN (2003) Persistence of a highly resistant strain of tuberculosis in New York City during 1990-1999. J Infect Dis 188:356–363

Ochi k, Tanaka Y, Tojo S (2014) Activating the expression of bacterial cryptic genes by rpoB mutations in RNA polymerase or by rare earth elements. J Ind Microbiol Biotechnol 41:403–414

Perlman DC, Segal Y, Rosenkranz S, Rainey PM, Remmel RP, Salomon N, Hafner R, Peloquin CA (2005) The clinical pharmacokinetics of rifampin and ethambutol in HIV-infected persons with tuberculosis. Clin Infect Dis 41:1638–1647

Piton J, Petrella S, Delarue M, André-Leroux G, Jarlier V, Aubry A, Mayer C (2010) Structural insights into the quinolone resistance mechanism of Mycobacterium tuberculosis DNA gyrase. PLoS One 5:e12245

Qi Q, Preston GM, Maclean RC (2014) Linking system-wide impacts of RNA polymerase mutations to the fitness cost of rifampin resistance in Pseudomonas aeruginosa. MBio 5:e01562–e01514

Reynolds MG (2000) Compensatory evolution in rifampin-resistant Escherichia coli. Genetics 156:1471–1481

Reynolds J, Heysell SK (2014) Understanding pharmacokinetics to improve tuberculosis treatment outcome. Expert Opin Drug Metab Toxicol 10:813–823

Rock JM, Lang UF, Chase MR, Ford CB, Gerrick ER, Gawande R, Coscolla M, Gagneux S, Fortune SM, Lamers MH (2015) DNA replication fidelity in Mycobacterium tuberculosis is mediated by an ancestral prokaryotic proofreader. Nat Genet 47:677–681

Sandgren A, Strong M, Muthukrishnan P, Weiner BK, Church GM, Murray MB (2009) Tuberculosis drug resistance mutation database. PLoS Med 6:e2

Skrahina A, Hurevich H, Zalutskaya A, Sahalchyk E, Astrauko A, Hoffner S, Rusovich V, Dadu A, De Colombani P, Dara M, Van Gemert W, Zignol M (2013)

Multidrug-resistant tuberculosis in Belarus: the size of the problem and associated risk factors. Bull World Health Organ 91:36–45

Song T, Park Y, Shamputa IC, Seo S, Lee SY, Jeon HS, Choi H, Lee M, Glynne RJ, Barnes SW, Walker JR, Batalov S, Yusim K, Feng S, Tung CS, Theiler J, Via LE, Boshoff HI, Murakami KS, Korber B, Barry CE 3rd, Cho SN (2014) Fitness costs of rifampicin resistance in Mycobacterium tuberculosis are amplified under conditions of nutrient starvation and compensated by mutation in the beta' subunit of RNA polymerase. Mol Microbiol 91:1106–1119

Spies FS, Ribeiro AW, Ramos DF, Ribeiro MO, Martin A, Palomino JC, Rossetti ML, Da Silva PE, Zaha A (2011) Streptomycin resistance and lineage-specific polymorphisms in Mycobacterium tuberculosis gidB gene. J Clin Microbiol 49:2625–2630

Traore B, Diarra B, Dembele BPP, Somboro AM, Hammond AS, Siddiqui S, Maiga M, Kone B, Sarro YS, Washington J et al (2012) Molecular strain typing of Mycobacterium tuberculosis complex in Bamako, Mali. Int J Tuberc Lung Dis 16:911–916

Trauner A, Borrell S, Reither K, Gagneux S (2014) Evolution of drug resistance in tuberculosis: recent progress and implications for diagnosis and therapy. Drugs 74:1063–1072

Van Doorn HR, De Haas PE, Kremer K, Vandenbroucke-Grauls CM, Borgdorff MW, Van Soolingen D (2006) Public health impact of isoniazid-resistant Mycobacterium tuberculosis strains with a mutation at amino-acid position 315 of katG: a decade of experience in The Netherlands. Clin Microbiol Infect 12:769–775

Vogwill T, Kojadinovic M, Furió V, Maclean RC (2014) Testing the role of genetic background in parallel evolution using the comparative experimental evolution of antibiotic resistance. Mol Biol Evol. msu262 31(12):3314–3323 doi:10.1093/molbev/msu262

Vogwill T, Kojadinovic M, Maclean RC (2016) Epistasis between antibiotic resistance mutations and genetic background shape the fitness effect of resistance across species of Pseudomonas. Proc R Soc B 283(1830):20160151

(WHO), W. H. O (2016) Global tuberculosis report 2016. WHO, Geneva

Wells CD, Cegielski JP, Nelson LJ, Laserson KF, Holtz TH, Finlay A, Castro KG, Weyer K (2007) HIV infection and multidrug-resistanttuberculosis: the perfect storm. J Infect Dis 196(Suppl 1): S86–107

Winglee K, Mcguire AM, Maiga M, Abeel T, Shea T, Desjardins CA, Diarra B, Baya B, Sanogo M, Diallo S et al (2016) Whole genome sequencing of mycobacterium africanum strains from Mali Provides insights into the mechanisms of geographic restriction. PLoS Negl Trop Dis 10:e0004332

Yang C, Luo T, Sun G, Qiao K, Sun G, Deriemer K, Mei J, Gao Q (2012) Mycobacterium tuberculosis Beijing strains favor transmission but not drug resistance in China. Clin Infect Dis 55:1179–1187

Yang C, Shen X, Peng Y, Lan R, Zhao Y, Long B, Luo T, Sun G, Li X, Qiao K, Gui X, Wu J, Xu J, Li F, Li D, Liu F, Shen M, Hong J, Mei J, Deriemer K, Gao Q (2015) Transmission of Mycobacterium tuberculosis in China: a population-based molecular epidemiologic study. Clin Infect Dis 61:219–227

Yuan L, Mi L, Li Y, Zhang H, Zheng F, Li Z (2016) Genotypic characteristics of Mycobacterium tuberculosis circulating in Xinjiang, China. Infect Dis (Lond) 48:108–115

Zhang Y, Yew WW (2015) Mechanisms of drug resistance in Mycobacterium tuberculosis: update 2015. Int J Tuberc Lung Dis 19:1276–1289

Zhang H, Li D, Zhao L, Fleming J, Lin N, Wang T, Liu Z, Li C, Galwey N, Deng J et al (2013) Genome sequencing of 161 Mycobacterium tuberculosis isolates from China identifies genes and intergenic regions associated with drug resistance. Nat Genet 45:1255–1260

Zignol M, Van Gemert W, Falzon D, Sismanidis C, Glaziou P, Floyd K, Raviglione M (2012) Surveillance of anti-tuberculosis drug resistance in the world: an updated analysis, 2007–2010. Bull World Health Organ 90:111–119d

Zignol M, Dean AS, Alikhanova N, Andres S, Cabibbe AM, Cirillo DM, Dadu A, Dreyer A, Driesen M, Gilpin C, Hasan R, Hasan Z, Hoffner S, Husain A, Hussain A, Ismail N, Kamal M, Mansjo M, Mvusi L, Niemann S, Omar SV, Qadeer E, Rigouts L, Ruesch-Gerdes S, Schito M, Seyfaddinova M, Skrahina A, Tahseen S, Wells WA, Mukadi YD, Kimerling M, Floyd K, Weyer K, Raviglione MC (2016) Population-based resistance of Mycobacterium tuberculosis isolates to pyrazinamide and fluoroquinolones: results from a multicountry surveillance project. Lancet Infect Dis

Mathematical Models for the Epidemiology and Evolution of *Mycobacterium tuberculosis*

15

Jūlija Pečerska, James Wood, Mark M. Tanaka, and Tanja Stadler

Abstract

This chapter reviews the use of mathematical and computational models to facilitate understanding of the epidemiology and evolution of *Mycobacterium tuberculosis*. First, we introduce general epidemiological models, and describe their use with respect to epidemiological dynamics of a single strain and of multiple strains of *M. tuberculosis*. In particular, we discuss multi-strain models that include drug sensitivity and drug resistance. Second, we describe models for the evolution of *M. tuberculosis* within and between hosts, and how the resulting diversity of strains can be assessed by considering the evolutionary relationships among different strains. Third, we discuss developments in integrating evolutionary and epidemiological models to analyse *M. tuberculosis* genetic sequencing data. We conclude the chapter with a discussion of the practical implications of modelling – particularly modelling strain diversity – for controlling the spread of tuberculosis, and future directions for research in this area.

Keywords

Strain variation • Heterogeneity • Population biology • Phylogeny • Phylodynamics • Molecular epidemiology • Compartmental model

J. Pečerska • T. Stadler
Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland
e-mail: julija.pecerska@bsse.ethz.ch; tanja.stadler@bsse.ethz.ch

J. Wood
School of Public Health and Community Medicine, UNSW Sydney, Sydney, NSW, Australia
e-mail: james.wood@unsw.edu.au

M.M. Tanaka (✉)
School of Biotechnology & Biomolecular Sciences, and Evolution & Ecology Research Centre, UNSW Sydney, Sydney, NSW, Australia
e-mail: m.tanaka@unsw.edu.au

S. Gagneux (ed.), *Strain Variation in the* Mycobacterium tuberculosis *Complex: Its Role in Biology, Epidemiology and Control*, Advances in Experimental Medicine and Biology 1019,
DOI 10.1007/978-3-319-64371-7_15

15.1 Introduction

The causative agent of tuberculosis, *Mycobacterium tuberculosis*, emerged as a human pathogen around 70,000 years ago (Comas et al. 2013; Gagneux 2012), although conflicting estimates point to much later dates of around 5,000 years ago (Bos et al. 2014). Forms of tuberculosis such as *M. bovis* that infect non-human animals evolved from human tuberculosis, indicating that the disease first appeared in humans before adapting to other animals. Mounting genetic evidence indicates that strain-to-strain variation in *M. tuberculosis* is more extensive than previously thought (Borrell and Gagneux 2011; Gagneux and Small 2007). Seven major lineages of modern-day *M. tuberculosis* have been identified (Galagan 2014) and specific strains are highly associated with geographic location (see Chap. 1) (Gagneux and Small 2007; Hershberg et al. 2008). Molecular methods have helped identify finer scale variation within lineages, which we discuss in more detail in Sect. 15.3.

The increased availability of data both at the epidemiological and molecular level allows us to start raising complex questions about data interpretation and analysis. For instance, how do we understand and predict tuberculosis epidemics on the population level? How do we best use molecular data to shed light on the transmission dynamics of different *M. tuberculosis* lineages? These questions typically require collated data analysis under specific assumptions on the properties of *M. tuberculosis*, such as, the mechanisms of mutational change. Mathematical, or computational, modelling is a methodology that enables the precise description of assumptions in order to investigate model behaviour, qualitatively or quantitatively. Defined models can be combined with data and thus provide answers to scientific questions concerning a given dataset. This approach has been instrumental in the understanding of physical sciences, and it has become more widely used in biology as biological data have become increasingly refined and quantitative in nature.

Mathematical models that are applied in biology range from being extremely simple and generic to being complex and specific. Simple models often enable a qualitative understanding of complex phenomena, while complex models have the advantage of being more realistic and detailed and thus may offer detailed quantitative insight. In the words of statistician George Box, however, "all models are wrong but some are useful". The aim of modelling is to shed light on a phenomenon rather than to create a maximally realistic description of it.

In the study of infectious diseases models can extend our understanding of an epidemic by allowing us to predict population dynamics from basic knowledge of the natural history of a disease. Models can help evaluate the effects of any potential or actual interventions at the population level. By providing precise quantitative predictions mathematical models also play a role in drawing inferences from observational data, for example, by producing estimates of parameters relating to disease transmission.

In this chapter we consider how mathematical and computational models can be used to understand the variation in *M. tuberculosis* that has been revealed using molecular techniques. Two different modelling traditions are pertinent to this topic. First, epidemiological models address the dynamics of infectious diseases at the population level and enable researchers to consider possible outcomes including the effects of intervention strategies. Second, models of molecular evolution and population genetics concern the processes by which genomes undergo change. These models are generic in that they have not been developed for any particular species, and can be applied to *M. tuberculosis* to understand its variation and to reconstruct its evolutionary history. We will describe both of these approaches and their applications to *M. tuberculosis*. We will further discuss progress made in combining epidemiological and evolutionary elements within the same framework to analyse the diversity of *M. tuberculosis*.

15.2 Epidemiological Modelling and Analysis

In this section we will focus on epidemiological models and their application to *M. tuberculosis*. We will primarily consider models that assume that the host population is homogeneous, ignoring possible effects of heterogeneity in host behaviour on the dynamics of the epidemic. We will begin with generic models that describe epidemiological dynamics of a single disease variant and then describe models of TB epidemiology with heterogeneity in the pathogen population, e.g. due to the occurrence of drug-resistant strains.

15.2.1 Epidemiological Modelling of *M. tuberculosis*

Epidemiological models traditionally separate host populations into distinct compartments according to their infection status. In the simplest scenario, an individual is either susceptible to infection, infectious, or recovered and therefore immune to reinfection. The numbers of individuals in each compartment are tracked by $S(t)$, $I(t)$ and $R(t)$ respectively, where t stands for time. In what follows we will drop the "(t)" from the notation except where it is clearer to retain it. Typically, a susceptible host is assumed to transition to the infected state at a rate proportional to the number of infected individuals I (say βI), and an infected individual transitions to the recovered state at a constant rate (say γ). Without host birth or host death events, the total number of individuals in the three compartments is constant $N = S+I+R$ where N is the total population size. The structure of this model is illustrated in the left panel of Fig. 15.1. This model, which is known as the Susceptible-Infectious-Recovered (SIR) model, was initially studied in depth by Kermack and McKendrick

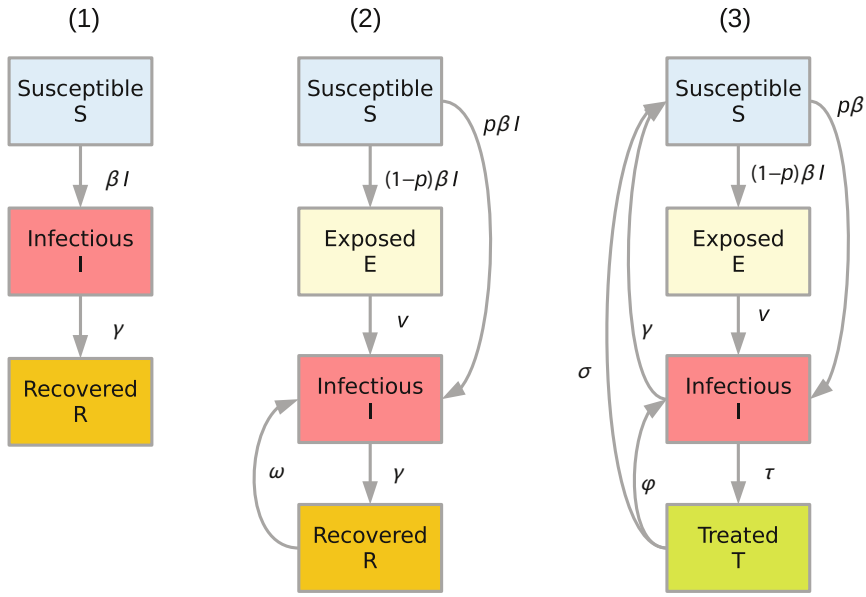

Fig. 15.1 Examples of compartmental models of infectious disease dynamics. (*1*) Classic SIR model with transmission rate β and recovery rate γ. (*2*) A more complex model with an exposed class E. In the case of tuberculosis, a proportion $(1-p)$ of new cases enter a state of latent infection modelled with E while the remainder (p) progress to active disease I. Latent infections reactivate at rate v, active cases recover at rate γ and recovered individuals regress to the disease state at rate ω. (*3*) When modelling antibiotic use and drug resistance it is useful to modify the model to include a state for infected treated individuals T. In this model, active cases are detected and treated at rate τ, stop treatment at rate ϕ and treated individuals return to the uninfected class S at rate σ. Not shown are death from each compartment or birth into S. We note that published models of TB dynamics are varied and while they are in general similar to the structures shown in (*2*) and (*3*) there are differences that reflect differences in the questions being addressed

(1927) and has since been elaborated upon in many ways (Diekmann et al. 1990; Keeling and Rohani 2008).

The SIR dynamics can be modelled deterministically or stochastically. In the deterministic version, the change in compartment sizes follows ordinary differential equations

$$\frac{d}{dt}S(t) = -\beta I(t)S(t) \tag{15.1}$$

$$\frac{d}{dt}I(t) = \beta I(t)S(t) - \gamma I(t) \tag{15.2}$$

and $R(t) = N - S(t) - I(t)$. If the epidemic starts with a single introduction of the infection into the population, the initial conditions are set to $S(0) = N - 1$ and $I(0) = 1$.

In the stochastic formulation of the SIR model, S, I, and R are integer-valued rather than real-valued, and when an infection occurs I increases by one and S decreases by one. Given a very small time interval Δt, the probability for infection to happen is assumed to be $\beta SI\Delta t + o(\Delta t)$. The term $\beta SI\Delta t$ is the probability for precisely one infection event to happen in time interval Δt. The term $o(\Delta t)$ summarises the probability for more than one infection event to happen, with the term $o(\Delta t)/\Delta t$ approaching zero as Δt approaches zero. This means that the waiting time until an infection event where I increases by one and S decreases by one is exponentially distributed with parameter βSI. Similarly, upon recovery I decreases by one and R increases by one, and this event occurs with probability $\gamma I\Delta t + o(\Delta t)$.

The dynamics of the SIR model are well understood and are described in multiple sources such as the text by Keeling and Rohani (2008). Unfortunately, the assumptions of the SIR model are clearly too narrow to be directly applicable to *M. tuberculosis*. In particular, *M. tuberculosis* infection is characterised by a long and highly variable incubation period known as latent infection. Furthermore, hosts generally do not have strong immune protection against further infection and asymptomatic hosts often relapse to disease years after an acute infection. However, the basic methodology of SIR modelling can be modified to reflect the natural history of *M. tuberculosis*. Adaptation of the SIR model to *M. tuberculosis* began with the work of Waaler et al. (1962), with the inclusion of long-term latency in the form of non-symptomatic cases as a key feature. The model divides individuals into noninfected, noncase, actual disease case, and recovered compartments. The noncase individuals are the latently infected individuals that do not show symptoms immediately upon infection but can potentially progress (with some rate ν). The actual disease case individuals show symptoms and are thus infectious. Individuals move into the recovered compartment after an active *M. tuberculosis* infection.

Figure 15.1 (middle) shows a simple version of a model structure that captures key features, in particular latency and relapse, of *M. tuberculosis* natural history. Note that the R compartment here represents "recovery" from active tuberculosis, but individuals in this compartment have not necessarily cleared the *M. tuberculosis* pathogen which is why they may relapse with some rate ω.

Different approaches have been used to incorporate latency due to the lack of detailed quantitative information about the long-term dynamics of infection and the immune response within humans. Blower et al. (1995) popularised the use of a dichotomous short-term/long-term characterisation of latency based around the rule of thumb that around 5% of infections progress quickly to active disease and another 5% progress slowly over the remainder of a person's lifespan. This would be captured by $p = 0.05$ in Fig. 15.1 (middle). In these models, a fraction of infected individuals progress immediately to active disease while the remaining fraction enter a latent state and progress to disease at a low rate. More recently, a modified form of this dichotomous transition has been introduced that more accurately captures the timing of active disease in relation to infection through stratification of latency into 2-stages (see for instance Dowdy et al. 2013).

Models of tuberculosis epidemiology have been used to characterise the decline of *M. tuberculosis* epidemics in the US and UK (Blower et al. 1995) and to determine the

contribution of endogenous reactivation and exogenous reinfection to the overall risk of disease (Vynnycky and Fine 1997). Blower et al. (1995) used a model that allowed for infection, reactivation, and relapse, and showed that the apparent decline might be explained as a temporary effect following a large epidemic. In the work by Vynnycky and Fine (1997), a more data-driven approach to *M. tuberculosis* infection risk estimation was taken, without modelling the underlying transmission process directly. The model was designed to evaluate the impact of new infections compared to reinfection and reactivation of the disease. The results suggest that in the UK reinfection made a strong epidemiological impact during the first half of the twentieth century, but had negligible effects by 1980, by which time the incidence reached its lowest point. This approach is less relevant to more recent epidemiological history in countries such as the UK, where cases have been increasingly driven by migration from high-incidence settings.

The observed variation in the effect of the BCG vaccine has also been investigated, for instance through the work by Gomes et al. (2004), drawing on the earlier model-based analysis by Fine and Vynnycky (1998). The latter work was aimed at explaining the differential success of the BCG vaccine in different settings, ranging from high efficacy observed in the UK trials as opposed to no efficacy in the large Chingleput trial in India. This paper noted how exposure to *M. tuberculosis* and/or related mycobacteria could have a confounding effect on the estimates of vaccine efficacy. Animal studies have shown that the BCG vaccine did not provide a further benefit over the immunity derived from tuberculosis infection. The study by Fine and Vynnycky showed how variation in this background level of immunity affected estimates of vaccine efficacy, with these estimates becoming negligibly small when natural infection rates were very high and provided similar levels of protection as vaccination. Gomes et al. (2004) explored these issues in a population dynamic model and linked the analysis with the concept of reinfection threshold, which is defined as the reproduction number (see Sect. 15.2.2) becoming greater than 1 in a population of previously exposed individuals.

Further developments in tuberculosis modelling have focused on potential effects of variation in treatment strategies. Treatment strongly affects incidence, prevalence and mortality from TB and reduces the average duration of infectiousness and thus reduces the possibility of transmission. Treatment is typically incorporated in models through a modification of the rate of cure (as in Fig. 15.1, right), although in models seeking to capture treatment programs more precisely, this may be described through multiple model states and transitions. Such models have been used to examine the traditional World Health Organisation DOTS (directly observed treatment, short-course) approach (Dye et al. 1998) and various means of improvement, for example, through active case finding, changes in diagnosis and treatment regimens and wider application of preventive treatment of latent tuberculosis infection (Dye and Williams 2010). Of particular interest within the context of this book are the models that have looked at interactions between treatment programs and the development of resistance, which we cover in more detail in Sect. 15.5.

Modelling studies have also aimed to understand the origins of tuberculosis (Chisholm et al. 2016) and the longer term evolution of *M. tuberculosis* and its characteristics, such as latency (Chisholm and Tanaka 2016; Zheng et al. 2014) and virulence (Basu and Galvani 2009).

15.2.2 Transmission Parameters and Their Estimation

Epidemiological modelling has identified a critical quantity in infectious disease dynamics known as the basic reproduction number or R_0 (Anderson and May 1979; Diekmann et al. 1990). This quantity is defined as the average number of new infectious cases arising from a typical case in a completely susceptible population. One of the principal reasons for interest in R_0 is that it constitutes a threshold quantity for a large class of infectious disease

models; namely, $R_0 > 1$ implies that the disease can persist whereas $R_0 < 1$ leads to disease elimination. Despite this simplicity, the mathematical definition of R_0 is a function of model structure and reflects details of disease epidemiology.

Under the SIR model, $R_0 = \beta S(0)/\gamma$. However, even quite simple models of TB such as that defined in Blower et al. (1995) lead to more complex expressions, with R_0 defined as the sum

$$R_0 = R_0^{\text{FAST}} + R_0^{\text{SLOW}} + R_0^{\text{RELAPSE}}.$$

Each component here is a function of multiple parameters. The SIR model formulation of R_0 resembles only the R_0^{FAST} component whereas the transmission potential for tuberculosis is complicated by the processes of reactivation (R_0^{SLOW}) and relapse (R_0^{RELAPSE}). Other studies have additionally included the process of reinfection (Gomes et al. 2004). For a general derivation and discussion of the basic reproduction number we refer readers to the text by Diekmann and Heesterbeek (2000).

In practice, populations are not usually fully susceptible and so the production of new cases is slower than indicated by R_0. New cases are better understood through the related quantity known as the *effective* reproduction number, usually labelled R_e. This quantity is defined as the average number of new infectious cases produced by a typical case regardless of the susceptible proportion. At the start of a local epidemic $R_e = R_0$, and over time it decreases to a value < 1 if the epidemic ends and is unable to persist, or remains close to 1 if the disease persists endemically. Knowledge of reproduction numbers is important for informing strategies for controlling infectious diseases but estimating these quantities poses some challenges. In particular, direct observation of the average number of secondary infections produced by infectious individuals is not feasible with conventional epidemiological methods. Instead reproduction numbersare typically estimated implicitly through comparisons of model outcomes with infection history data or incidence and prevalence time-series. Earlier approaches are summarised in Dietz (1993), while examples of approaches for stochastic models and structured communities are provided in the text by Becker (2015).

Empirical data can be used to estimate parameters which in turn allow investigation of the long term characteristics of epidemics using epidemiological models. The dynamics can also be studied through computer simulations of stochastic formulations of models, which can in turn be compared with data that directly reflect transmission such as contact tracing studies or analysis of *M. tuberculosis* infection within households. Brooks-Pollock et al. (2011), for example, used incidence data across multiple-person households to estimate the relative contributions of community and households to *M. tuberculosis* transmission.

For tuberculosis, the differing components of the reproduction number as described above make this more challenging. Estimation of the basic reproduction number is also complicated by the absence of good diagnostic markers for immunity. As such, in contrast to infections such as influenza and measles, practical epidemiological models of tuberculosis have focused more on projections of disease incidence than analysis of the reproduction number and the potential for elimination (Dye and Williams 2010). However, the growing availability of molecular data provides opportunities to overcome some of the issues outlined above. In particular, molecular data provide information that can potentially allow models to separate between recent infection, reactivation and relapse (Small et al. 1994). Stochastic formulations of the epidemiological models in combination with molecular data, rather than deterministic formulations in combination with the traditional epidemiological data discussed here, will be the focus of Sect. 15.4.

Fig. 15.2 Examples of compartmental models with pathogen variation and host variation. (*1*) When pathogen variation is taken into account the infected classes are partitioned into multiple states corresponding to the multiple pathogen types. The variation can be with respect to resistance or other genetic characters. Here we do not show transitions between types, which depend on the details of the model. (*2*) When host variation is taken into account the hosts are partitioned into susceptibility classes. For simplicity the figure does not show births and deaths for either model

15.2.3 Modelling Heterogeneous Epidemics

Thus far we have discussed models in which infections are homogenous with all prevalent infectious cases represented by the single I compartment. However, pathogen populations may be variable, for example, in levels of drug resistance. Such variation could cause hosts infected with different strains to transmit with different rates. One can model a simple case by subdividing the I class into I_S and I_R classes: the hosts infected with a drug sensitive strain, and the hosts infected with a drug resistant strain, respectively. Correspondingly, one needs to divide the E class into E_S, E_R. Hosts may move from I_S to I_R through resistance evolution, and potentially move from I_R to I_S through loss of resistance.

Similarly, there may be variation in the host population both in terms of transmission risks (e.g. geographic variation in incidence) and in risks of developing active TB for example through HIV co-infection. Again, we can divide the compartments according to this variation, for example into I_1 and I_2. As done above, we subdivide the S class as well as the E class as the difference in risk behaviour does not depend on the TB infection status. Figure 15.2 illustrates an example of how variation in pathogens (left) and variation in hosts (right) could be modelled, extending the homogeneous epidemic model displayed on the right of Fig. 15.1.

Such models have been used to represent different types of variation in viral epidemics, such as geographical variation, drug resistance levels and super spreader behaviour. For *M. tuberculosis*, outside of settings with high HIV prevalence, the main source of variation that has been considered to date is in relation to drug-resistance and its interaction with treatment programs, which we now discuss in more detail.

As with other infectious diseases, the rise of antimicrobial drug resistance is a problem for the control of *M. tuberculosis*. While multi-drug resistant (MDR) or extremely-drug resistant (XDR) tuberculosis is likely to be the result of treatment

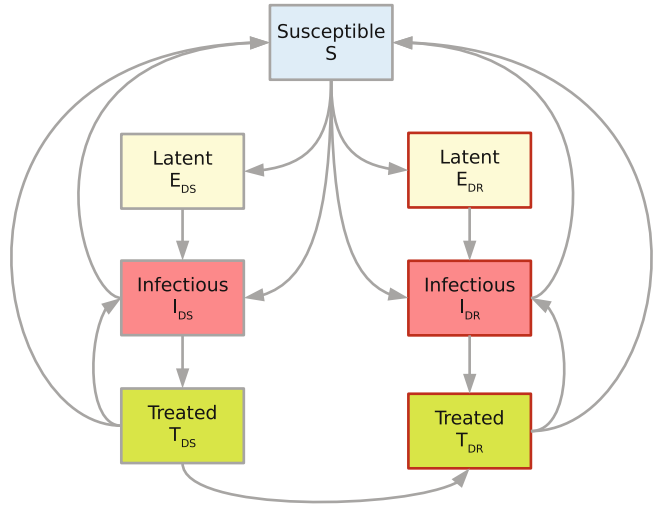

Fig. 15.3 An example of a compartmental model of tuberculosis with drug resistance evolution and transmission. This model is an extension of the treated TB model shown in Fig. 15.1 (right) in which drug sensitive *M. tuberculosis* can evolve resistance de novo and subsequently transmit to new hosts. The subscripts indicate drug sensitive (DS) and drug resistant (DR) infection states. For simplicity, rates are omitted from this diagram

failure, they also occur with lower frequency in new cases. These issues are discussed in more detail in Chaps. 11 and 14 but here we briefly describe key features of epidemiological models of resistance.

In the context of population dynamics, drug resistant strains have two critical properties. First, they can persist longer in patients because standard treatment is less effective or not effective at all. Second, they may come with a fitness cost that lowers their rate of transmission. The cost may be due to a trade-off between the original function of the gene and the resistance phenotype (Chap. 11). The fitness cost has been measured in vitro (Gagneux et al. 2006) and it has been observed to be variable across *M. tuberculosis* strains. Such in vitro measurements assess the replicative capacity of a strain which differs from the ability of the strain to infect new hosts, i.e. the transmission fitness, but the two fitness concepts are assumed to be linked. In vitro, strains have shown a replication fitness disadvantage when compared to their rifampicin-susceptible ancestors, while some of the clinically-derived strains show no fitness costs. One reason for this variability is the possibility that further mutations occur that lower or compensate the cost of resistance. Thus, it is unlikely that drug resistant *M. tuberculosis* will easily revert to sensitivity (Cohen and Murray 2004).

The dynamics of drug resistance in *M. tuberculosis* have been studied using compartmental models as introduced above and shown in Fig. 15.2 (left). The first such models were introduced by Blower and colleagues (Blower and Gerberding 1998; Blower et al. 1996) modelling two types of infection: drug sensitive and drug resistant. Models have since been extended and refined, maintaining a structure similar to that shown in Fig. 15.3 (see also Ozcaglar et al. 2012). This figure is a special case of models with pathogen heterogeneity as shown on the left in Fig. 15.2, with the addition of transition between sensitive and resistant treatment classes.

Such models of the evolution and spread of drug resistance have been studied to address a number of epidemiological problems, in particular to clarify the importance and variability of the reproductive and transmission fitness of drug resistant strains. Such insight allows us to quantify the future burden of drug resistant strains on an epidemiological scale. The replicative fitness of resistant strains is variable (Dye and Williams 2009; Gagneux et al. 2006), but there is some evidence for lower rates of transmission of resistant strains (Dye and Williams 2009; Luciani et al. 2009). On the epidemiological scale this cost is balanced by the advantage resistant strains have in patients under treatment (Luciani et al. 2009). Furthermore, since the cost of resistance can be

lowered by compensatory mutations, the resulting variation in *M. tuberculosis* fitness means that without adequate control strategies, resistant strains will likely dominate in the long run (Blower and Gerberding 1998; Cohen and Murray 2004; Luciani et al. 2009; Shrestha et al. 2014; Trauer et al. 2014). Models can be used to explore control scenarios – e.g. rates of detection and cure success – that could lead to the control of drug-resistant tuberculosis (Dye and Espinal 2001; Dye and Williams 2000). Modelling also allows quantification of the relative importance of the two factors contributing to the spread of drug resistant strains, namely de novo evolution of resistance versus transmitted resistance. Modelling studies strongly suggest that in most settings a large majority of drug resistant cases are due to the transmission of resistant strains rather than the de novo acquisition of resistance (Cohen and Murray 2004; Kendall et al. 2015; Luciani et al. 2009; Trauer et al. 2014).

15.3 Molecular Evolution of *M. tuberculosis*

The fields of population genetics and molecular evolution are dedicated to analysing and understanding genetic variation within and among species. Many models and methods have been developed for these purposes, most of which are general and not specific to pathogens. In this section we discuss applications of this theory to *M. tuberculosis*. We start with standard models and methods in molecular evolution and phylogenetics, and then discuss some of the specific issues that arise when applying these methods to genetic data obtained from *M. tuberculosis* isolates.

All organisms may accumulate mutations through replication, including *M. tuberculosis*. Many of these changes are never observed as the resulting mutants may suffer a large fitness disadvantage so that they are eliminated from the population. Other mutations, however, rise in frequency and may become fixed in a population. A mutation which is fixed in a population is called a *substitution*. To define this term more precisely, it is necessary to clarify what population is being considered. The evolution of *M. tuberculosis* occurs on at least two different population scales. First, the global population of *M. tuberculosis* may undergo evolutionary substitution. This process of substitution is important to consider when comparing *M. tuberculosis* to other species. Second, evolutionary substitution occurs at a local level within hosts. In this case, mutants arise within each host and may reach fixation in the host. In this section we are primarily concerned with substitution at the local host scale rather than the global population scale. Genetic changes accumulate within hosts so that bacterial populations in different hosts generally diverge. This genetic divergence provides information about the evolutionary history of the bacterium, which we can represent with a phylogenetic tree. A similar but distinct tree concept used in population genetics studies and implemented in some software is the coalescent, which describes genealogical history viewing time going backwards (for further details see Wakeley 2009). Coalescents are also known as gene trees.

Phylogenies are appropriate when there is no horizontal gene transfer or recombination and no convergent evolution. In contrast to many other bacteria, *M. tuberculosis* exhibits remarkably little recombination (Liu et al. 2006) and there is currently no evidence it carries plasmids (Zainuddin and Dale 1990). This makes the analysis of genomes relatively straightforward as only substitutions on a phylogenetic tree need to be modelled. In particular, phylogenies of *M. tuberculosis* based on different genetic loci have tree-like structures rather than being reticulate or networks. On the other hand, some genes in the *M. tuberculosis* genome may be under strong natural selection and can therefore exhibit convergent evolution. In particular, sites conferring drug resistance are known to undergo convergence and are therefore typically excluded from any phylogenetic analysis.

In the next subsections we will first discuss evolutionary models for markers of *M. tuberculosis* and then evolutionary models for all nucleotides within the *M. tuberculosis* genome.

We will then illustrate how these models are used in population genetics to assess the within- and between-host variation of *M. tuberculosis* strains.

15.3.1 Evolutionary Models for Markers of *M. tuberculosis*

In an effort to characterise and understand the transmission of tuberculosis, isolates of *M. tuberculosis* have been genotyped for many years using a variety of molecular methods (Chap. 3). Early methods included typing based on the mobile gene IS6110 (Alland et al. 1994; Cave et al. 1991; Small et al. 1994), spacer-oligonucleotide typing or spoligotyping (Kamerbeek et al. 1997) and MIRU-VNTR typing (Supply et al. 2006). In recent years, the declining cost of DNA sequencing has enabled the use of whole-genome sequencing (WGS) as a strategy for studying *M. tuberculosis* transmission (Gardy et al. 2011; Walker et al. 2013). WGS allows a fine-scale genetic characterisation of *M. tuberculosis* strains within a population, and has the advantage over previous technologies of minimising the impact of parallel evolution (Chap. 3).

The genetic variation observed at marker loci ultimately arises from mutation processes. To exploit this variation for studying *M. tuberculosis* epidemiology it is useful to know how and at what rate the underlying genetic loci mutate and undergo within-host substitution (Tanaka and Francis 2005). The simplest model to apply is a Poisson process in which mutations appear randomly at a constant rate per unit time and reach fixation instantly. This model and elaborations of it have been used to estimate mutation rates. An alternative approach is to compare the extent of variation using different markers to estimate rates against known mutation rates (Reyes and Tanaka 2010).

Genomic variation in *M. tuberculosis* is partly due to the movement of the mobile gene IS6110. The rate of movement of IS6110 has been estimated to be around 0.1 to 0.3 changes per year using serial isolates of *M. tuberculosis* (de Boer et al. 1999; Rosenberg et al. 2003; Warren et al. 2002). Spoligotype variation is due to deletion of repeats at a CRISPR locus (Aranaz et al. 2004). The evolutionary process at this locus, estimated to be around 0.02 to 0.09 per year (Reyes and Tanaka 2010), is slower than changes due to IS6110 movement. VNTR loci mutate by expanding or contracting in the number of repeats. The stepwise mutation models developed for microsatellites in humans (Di Rienzo et al. 1994) are likely to apply well to bacteria (Vogler et al. 2006). Under the simplest version of such models, repeats increase or decrease by a single step and with equal probability. Estimates of the mutation rate of VNTR loci have varied considerably with low rates around 10^{-5} per locus per year (Grant et al. 2008) and $10^{-3.9}$ (Wirth et al. 2008) and high rates of 10^{-3} to 10^{-2} (Aandahl et al. 2012; Ragheb et al. 2013; Reyes and Tanaka 2010). The variation in these estimates may reflect the diversity of models, methods and data used to obtain them.

Single nucleotide polymorphisms (SNPs) occur through point mutation which can appear throughout the entire genome. Whole genome sequences analysed with phylogenetic and similar methods have yielded conflicting estimates of mutation rates varying from 3×10^{-9} (Comas et al. 2013) to 10^{-7} (Bos et al. 2014; Walker et al. 2013). The higher rates are supported by in vitro studies of mutation rates (Ford et al. 2013) but it should be noted that the "long-term" rate of evolution is likely to be lower because estimates based on recent variation includes polymorphisms that ultimately will not be fixed in the population (e.g. deleterious mutations) (Ho et al. 2005). Mutation rates also vary substantially depending on genetic background (Ford et al. 2013). Whether mutation rates during latent infection are equal to (Ford et al. 2011; Lillebaek et al. 2016) or lower than (Colangeli et al. 2014) rates during active infection is not settled. We suggest that the uncertainty in mutation rates can be further addressed in the future through modelling and careful consideration of assumptions underlying models and data.

One of the challenges in using genetic markers for phylogenetic analysis is the occurrence of parallel evolution (or homoplasy), by which iden-

tical states are reached by mutation in independent infections. Such events can potentially undermine the phylogenetic analysis of *M. tuberculosis* (Comas et al. 2009), although a simulation study has shown that the impact of homoplasy is not necessarily large on the epidemiological scale (Reyes et al. 2012). The arrival of low-cost genome sequencing has removed the obstacle of homoplasy since it does not strongly affect genome-wide SNPs (after removing SNPs implicated in drug resistance).

15.3.2 Evolutionary Models for Whole Genomes of *M. tuberculosis*

A wide array of molecular evolution models have been developed for nucleotide changes along a genome, covering a range of complexities. These models all assume that each nucleotide evolves independently of the other nucleotides. The simplest is the Jukes-Cantor model (JC69) which assumes that all substitutions among nucleotide bases occur at an equal rate and that base frequencies are all equal (Jukes and Cantor 1969). Kimura's two parameter model (K80) allows different transition and transversion rates, while keeping equal base frequencies (Kimura 1980), whereas Felsenstein's model (F81) keeps equal rates while allowing varying nucleotide base frequencies (Felsenstein 1981). More complex models include the HKY85 model which does not assume equal base frequencies and accounts for the difference between transitions and transversions (Hasegawa et al. 1985), the TN93 model which distinguishes not only transitions/transversions, but also differentiates between purine and pyrimidine transitions (Tamura and Nei 1993), and the generalised time-reversible (GTR) model, which is the least restrictive time-reversible model possible (Tavaré 1986).

Generally, the rate of substitution has been inferred to vary across sites. In order to account for such variation, it is common to assume either a continuous gamma distribution or a discrete approximation with a suitable number of rate classes (Yang 1994a, 1996) for variation across sites in the substitution rate. These substitution models can then be used in combination with genetic data to compute likelihoods or genetic distances for phylogenetic reconstruction, as we next discuss.

15.3.3 Phylogenetic Reconstruction

A number of approaches have been developed for reconstruction of phylogenetic trees from genetic data. Distance-based methods use a measure of genetic distance – an estimate of the degree of similarity two sequences share – to reconstruct evolutionary history. Other methods optimise a criterion measuring how well a tree explains the genetic sequence alignment over the space of possible phylogenetic trees.

In distance-based methods, similar sets of taxa are grouped together whereas more distant ones are placed further apart on a tree according to entries in a pairwise distance matrix. The branch lengths of the inferred phylogeny are a close, but in general imperfect, representation of the inter-sequence distance matrix. An example of a distance-based method is the neighbour-joining method which is a bottom-up clustering algorithm that joins nodes of the tree according to the shortest distance between two existing nodes (Saitou and Nei 1987). It is statistically consistent,[1] but does not maximise a criterion for measuring how well a tree explains the data.

Alternative methods of tree reconstruction from genetic data include maximum parsimony, maximum likelihood and Bayesian tree reconstruction, which all search tree space while optimising a criterion. The maximum parsimony method minimises the number of substitutions required to explain the inferred phylogeny. This method is quick and easy but it has been shown to be statistically inconsistent (Felsenstein 1978).

[1] Statistical consistency of a phylogenetic method means that when given infinitely long genetic sequences, the method –employing the model under which the sequences evolved– will recover the true underlying phylogeny.

The maximum likelihood method searches the tree space to maximise the probability of the data given a particular tree structure (the likelihood of the tree given the data). Bayesian methods also search the tree space, yielding a posterior distribution of trees. As such, Bayesian methods produce multiple trees; they also accommodate prior distributions on trees and parameters. In the maximum likelihood and Bayesian frameworks, the substitution model is used to evaluate the likelihood of trees and model parameters when considering the data. Thus, the methods allow multiple substitutions, parallel substitution, convergent evolution and back substitution along branches, and the complexity of the substitution model can be adjusted to improve the fit to the data. Both frameworks are statistically consistent (Steel 2013; Yang 1994b), and are currently the most widely used phylogenetic reconstruction methods.

Phylogenetic reconstruction methods produce trees with branch lengths measured in numbers of substitutions. Exceptions are the Bayesian methods assuming a clock model (Drummond et al. 2006; Zuckerkandl and Pauling 1965) which leads to branch lengths in units of calendar time. Branch lengths in calendar time are important for quantifying the timing of epidemic spread. Recently, a method to infer branch lengths in calendar time based on a tree with branch lengths on number of substitutions was introduced (LSD (least-squares dating) software (To et al. 2016)). For more details on phylogenetic methods, we refer readers to the texts by Yang (2014) or Felsenstein (2004).

15.3.4 Classification of TB Using Genetic Data

The application of molecular technologies and phylogenetic methods also enables the classification of pathogen isolates into broad classes. In relation to *M. tuberculosis* extensive genetic research has identified the structure of the complex of closely related *M. tuberculosis* species and the seven major extant lineages (see Chap. 1) (Galagan 2014). The classification of *M. tuberculosis* can be further refined by considering relationships within lineages using fast-evolving molecular markers. Attempts have been made to date the first introduction of *M. tuberculosis* into the human population using phylogenetic methods and to describe the patterns of world-wide *M. tuberculosis* distribution (Gagneux 2012). The suitability of alternative genetic approaches can be evaluated by comparing phylogenies reconstructed from different types of markers to a "gold standard" phylogeny in order to identify flaws in commonly used methods and to provide a means of quickly typing isolates (Filliol et al. 2006).

When the isolates in a data set are closely related to each other – such as isolates from a single outbreak – an alternative approach to phylogenies is to show the direct relationships among the genotypes in graphs such as minimal spanning trees. The underlying assumption is that all substitutions that occurred are observed, so that complex substitution models to account for hidden ancestral substitutions are not needed. This approach to visualisation and classification is aided by specifically modelling the processes of substitution underlying the genetic markers. Models of spoligotype evolution have been used to show relationships among isolates (Brudey et al. 2006; Ozcaglar et al. 2011; Reyes et al. 2008; Shabbeer et al. 2012a,b). Relationships among isolates based on MIRU-VNTR can also be visualised using graphs (Weniger et al. 2012). Large international databases based on multiple genotyping methods including MIRU-VNTR and spoligotypes aid in the classification of isolates (see Chap. 3) (Demay et al. 2012; Weniger et al. 2012).

SNPs obtained from whole genome sequencing can also be visualised through graphs within outbreaks (Walker et al. 2013) or through phylogenies when analysing highly divergent isolates with ancient origins (Bos et al. 2014; Comas et al. 2013; Wirth et al. 2008). It is anticipated that future epidemiological studies will increasingly make use of whole genome sequencing.

15.3.5 Population Genetics of TB

So far we have considered models of substitution (i.e. fixation of a mutation in a population), and how the variation between isolates can be represented in a tree. Population genetics "zooms into" the process of substitution by modelling the origins and dynamics of genetic variation including the process of fixation. A natural null model in population genetics is the neutral model in which all genetic variants are selectively equivalent. In this model the process of mutation generates new variants (alleles) while randomness – genetic drift – leads to loss of variation or to fixation (i.e. substitution). The dynamic balance between these two processes – mutation and drift – has been characterised along with properties of samples from a population in such balance (Ewens 1972). The theory of this balance between mutation and drift is generally applicable to bacteria including *M. tuberculosis*.

Because most viable mutations are expected to have zero or negligible effect on bacterial phenotype, variation at marker loci can be considered selectively neutral, as a first approximation. An important exception is antibiotic resistance genes which are removed for the purposes of phylogenetic analysis because they are known to be under selection (Comas et al. 2013). Exceptions to strict neutrality can also occur at marker loci: for instance, movement of IS6110 can lead to deleterious or even advantageous effects. Moreover, in the absence of recombination – as is the case for *M. tuberculosis* – any mutation under positive selection will be linked to neutral variation throughout the genome which may hitchhike to fixation in a selective sweep (Smith et al. 2009). Nevertheless, a first step to analysing most molecular variation is to treat it as selectively neutral to understand broad patterns in data using theory from population genetics (Ewens 2004).

Neutral models have been found to often adequately describe the distribution of cluster sizes under IS6110-RFLP typing and spoligotyping, and the fit can be improved compared to the standard Wright-Fisher population genetic model by allowing the infected population to expand according to a birth-death process (Luciani et al. 2008). These simple population models focus on genetic variation without considering details of the population dynamics or the natural history of disease. For example, at the epidemiological level the population of interest is the set of infected hosts, and the replication process is the transmission process. However, standard population genetic models do not explicitly account for the process of transmission nor other processes such as host recovery, death or latent infection. Thus there is a need to integrate population genetic models with epidemiological dynamics such as those described in Sect. 15.2. Section 15.4 will describe progress towards this goal.

15.3.6 Models of Within-Host Variation and Mixed Infections

At epidemiological scales, it is convenient to assume that each infected case corresponds to a single strain and that mutation leads to the instantaneous replacement of the ancestor, but in reality more than one strain can exist within an infection (Sergeev et al. 2011; Warren et al. 2004). Such infections are called *mixed* or *complex infections*. This bacterial variation may be due to mutation within the host or reinfection of the host by another strain. In order to understand the source and consequences of variation, models of bacterial dynamics at the within-host level are needed. Such modelling has led, for example, to methods to detect mixed infection (Plazzotta et al. 2015) and to classify whether the variation is due to reinfection or mutation (Chindelevitch et al. 2016) using genetic data. In cases where within-host variation is due to mutation, serial isolates of *M. tuberculosis* from an infection can be used to estimate mutation rates (Ragheb et al. 2013; Rosenberg et al. 2003; Tanaka et al. 2004). A benefit of these estimates is that unlike "snapshot" data they make use of temporal information. Within-host data can also be used to study population genetic statistics to quantify the action of natural selection (O'Neill et al. 2015).

We note that the dynamics of *M. tuberculosis* within patients are highly complex and involve a large number of interactions between the pathogen and the host. The roles of the immune system, inter-cellular signals and spatial effects have been modelled (Gammack et al. 2004; Wigginton and Kirschner 2001). In those models, the variation in disease dynamics is due to the complex interactions between *M. tuberculosis* and the host response. To be able to draw conclusions for population level dynamics, models must suppress some of the complexities of the intra-host dynamics while focusing on competition between strains. For example, different *M. tuberculosis* strains can be modelled by imposing structure on the pathogen population (Fig. 15.2, left), and different immune responses can be modelled by imposing structure on host population (Fig. 15.2, right).

Just as drug resistance is an important source of variation on the epidemiological scale, it is also important to consider on the within-host scale. Models that combine the population dynamics of *M. tuberculosis* with the pharmacokinetics of drugs provide a quantitative description of the emergence of resistance within a patient, which can then be used to optimise treatment regimens to minimise drug resistance (Lipsitch and Levin 1997). The effects of nonadherence and drug synergies can be considered under such models (Lipsitch and Levin 1997). Within-host modelling of multidrug resistant *M. tuberculosis* has led to predictions that such strains can arise at an unexpectedly high rate from apparently pansensitive within-host populations because of standing genetic variation in those populations (Colijn et al. 2011; Lipsitch and Levin 1997).

15.4 Integrating Epidemiological and Evolutionary Models

In Sect. 15.2.1 we described models of the epidemiology of disease and of *M. tuberculosis* in particular. In Sect. 15.3 we described how models from the fields of molecular evolution and population genetics can be applied to genetic data from *M. tuberculosis* isolates. We suggest that to fully understand genetic data, it is necessary to combine both kinds of approaches. We will describe developments in this area, including the integration of epidemiological and evolutionary models that involve the analysis of phylogenies – a young field dubbed *phylodynamics* (Grenfell et al. 2004).

15.4.1 Between-Host Variation, Clustering and Transmission

A central goal of molecular epidemiology is to draw inferences about disease transmission using genetic information (Glynn et al. 1999b; Mathema et al. 2006; Murray 2002a; Tanaka and Francis 2005). Above we introduced phylogenetic trees to investigate the diversity of *M. tuberculosis* strains, namely to infer the past history (or state) of the epidemic. In this section, we go a step further and in fact assess the transmission dynamics in addition to the state.

The degree to which isolates cluster into identical genotypes carries information about the extent of recent transmission. The underlying assumption is that genotypes evolve on the same timescale as the process of disease transmission so that each cluster of isolates represents a set of cases that arose recently through transmission, but isolates that are not connected via recent transmission are different through accumulated mutations. Simple clustering statistics have therefore been used to quantify recent transmission of tuberculosis (Alland et al. 1994; Small et al. 1994). Widely used are the "n" and "n-1" statistics ((Glynn et al. 1999b), also known as RTI_n and RTI_{n-1} respectively (Tanaka and Francis 2005)). These are defined as:

$$RTI_n = n_c/n \tag{15.3}$$

$$RTI_{n-1} = (n_c - c)/n \tag{15.4}$$

where n_c is the number of isolates in clusters (of 2 or more isolates), c is the number of clusters and n is the total number of isolates. These can be alternatively written as

$$RTI_n = 1 - u/n \quad (15.5)$$

$$RTI_{n-1} = 1 - g/n \quad (15.6)$$

where g is the number of distinct genotypes in the sample and u is the number of unique genotypes in the sample (also called non-clustered isolates or singletons).

Mathematical modelling has contributed to the goal of improving inferences from these data. First, modelling and simulation studies have shown that incomplete sampling leads to an underestimation of the extent of recent transmission which also interferes with assessing risk factors for recent transmission (Borgdorff et al. 2011; Glynn et al. 1999a,b; Murray 2002b; Murray and Alland 2002). Second, the degree of clustering cannot easily be compared across different types of markers because different markers mutate at different rates, a fact not accounted for in these simple clustering statistics (Tanaka and Francis 2005). Thus, in order to interpret patterns of clustering in terms of disease transmission it is important to know the mutation rates of different markers. Furthermore, separate clusters may not be completely different strains – a single mutation event could split a cluster into two clusters, which should be treated as a single epidemiological cluster of cases. Ideally then, methods of inference from genetic data would both incorporate the speed of evolution and account for sampling.

An approach to analysing genetic data is to use mathematical models that account for disease transmission, marker mutation and sampling. The extent of transmission can be quantified by estimating the effective reproduction number of the pathogen. Because models can be complex and difficult to work with directly, computational methods that approximate the likelihood have been applied to analyse data using models (Aandahl et al. 2014; Luciani et al. 2009; Tanaka et al. 2006). These models do not consider the phylogenetic history of the genetic data. An alternative approach is to augment the data by using trees which permits exact calculation of likelihoods conditional on trees (Stadler 2011). This approach requires a clear definition of trees that represent the evolutionary and epidemiological history of a sample of bacterial isolates. The next section introduces the concept of the transmission tree as a step towards integrating epidemiology and phylogeny.

15.4.2 Transmission Trees

In order to characterise the connection between epidemiological and evolutionary relationships it is necessary to introduce the concept of transmission trees to be studied in light of phylogenetic trees. If all of the infected individuals and the times and sources of every new infection were known, we could represent the spread of an infection as a bifurcating tree. In such a tree, which we will refer to as the transmission tree, each branch represents an infectious case, each bifurcation event represents a secondary case, and the root branch represents the initial introduction of the infection to the population. One of the very few cases of perfect transmission tree reconstruction via complete contact tracing was done in a 1980 study on a quarantined US naval vessel described by Houk (1980).

Unfortunately, in most epidemics it is difficult to achieve complete sampling and to record the precise timing of events; nor can we perform contact tracing, as the infection event might have taken place years prior to the onset of symptoms. Instead, genetic sequencing data are used to estimate the transmission tree relying on the following observation. Upon each infection a subset of the genotypes occurring in the donor is transferred to the recipient, producing a new case. From the time of that transmission event, the pathogen populations in the two distinct infections evolve independently of each other within the two hosts. Thus patients close in the transmission chain have *M. tuberculosis* genomes closer to each other compared to patients distant in the transmission chain. The phylogenetic tree of the pathogen sequences puts similar sequences close together and serves as a proxy for the transmission chain. Note that upon a bifurcation,

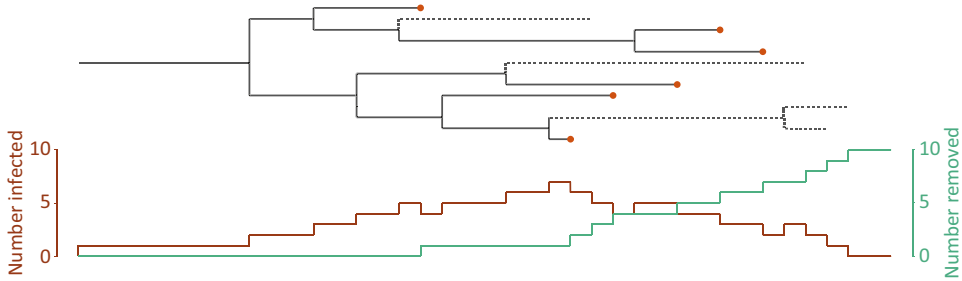

Fig. 15.4 An example of a transmission tree depicting SIR dynamics. The tree shows sampled branches with solid lines and unsampled ones with dotted lines, with samples shown as orange circles. The curves underneath the tree show the changes in the number of infected individuals and the recovered individuals

we do not know which patient is the donor or the recipient in this reconstructed tree.

Genetic sequencing data represent only a subset of active tuberculosis infection because of incomplete sampling due to financial or other constraints. Reconstructed transmission chains are therefore also incomplete. Figure 15.4 shows an example of a transmission tree in an epidemic with SIR dynamics, with incomplete sampling. By observing the reconstructed phylogenetic trees and interpreting them as a proxy for the transmission chain, one can make conclusions about multiple characteristics of the epidemic, such as possible hotspots of infection (e.g. households where multiple family members were infected) and identify whether a specific patient could have been a source of infection in a cluster. Methods to infer the exact transmission tree of *M. tuberculosis* from a phylogenetic tree, including the direction of infection have been recently proposed in Didelot et al. (2014, 2017). Essentially, these methods assign a corresponding sampled or unsampled infected host to lineages.

15.4.3 Phylodynamic Methods

As sequencing of genomes becomes more cost-effective, fast and reliable, increasing amounts of sequence data are sampled from ongoing epidemics, and phylogenetic trees as well as transmission trees are thus being reconstructed. This increase in sequence data has stimulated the development of phylodynamic methods that combine evolutionary and epidemiological analyses to quantify the parameters of the epidemiological models. In fact, the structure and branch length of the reconstructed phylogenetic tree contains information on the different compartments and rates of movement (i.e. dynamics) between the compartments in the underlying epidemiological model. For example, in the case of an SIR model with all patients being sampled upon recovery, the waiting times for a new branching event will be exponentially distributed with mean rate $\lambda = \beta I$ and the branches will terminate (the individual will stop being infectious) with mean rate γ. Incomplete sampling requires the development of sophisticated statistical tools integrating over all non-sampled patients (Stadler et al. 2013; Volz et al. 2009). The dependency will be more complex for a model including a latent or other non-infectious class, or allowing for heterogeneous pathogen or host population as in Fig. 15.2 (Kühnert et al. 2016; Stadler and Bonhoeffer 2013; Volz 2012). In such a case the reconstructed phylogenetic tree is extended by labelling each tip in the tree with the compartment the corresponding host belongs to: if tips from the same compartment cluster in the tree we conclude that there is transmission within the compartment, while if tips from the same compartment are spread over the tree we conclude independent migration into that compartment. The phylodynamic methods allow us to quantify the rates of transmission and migration in the epidemiological models based on the phylogenetic trees with the tip labels.

There are two main approaches to infer epidemiological parameters from pathogen sequencing data, which we call the two-step and the one-step approaches.

In the two-step approach, one first produces a phylogenetic tree with branch lengths in calendar units, using a tree reconstruction method as discussed in Sect. 15.3.3. The reconstructed tree is used as fixed input in the second analysis step to infer epidemiological parameters (see e.g. Kühnert et al. 2014, 2016; Stadler and Bonhoeffer 2013; Stadler et al. 2013; Volz 2012; Volz et al. 2009). Most of these methods are available within the software package BEAST2 (Bouckaert et al. 2014) and stand-alone implementations are mentioned in the individual papers. This parameter inference based on a fixed tree can be done using maximum likelihood (ML) or Bayesian inference. ML inference focuses on finding the combination of parameters that was the most likely to have produced the phylogenetic tree that is being studied under the given epidemiological model. The ML framework does not make use of prior knowledge of the parameters of the underlying models. In contrast, Bayesian methods can incorporate prior distributions of parameters and yield posterior distributions over the parameter space. Posterior distributions are a natural way to interpret the uncertainty in the resulting estimates. The incorporation of prior distributions allows for the better use of all the information available but requires great care in prior specification, as inappropriate priors can significantly and incorrectly influence the results of the analysis.

One-step approaches simultaneously estimate trees and parameters from the genetic sequences, typically in a Bayesian framework. In the one-step Bayesian approach the uncertainty in the phylogenetic trees is naturally integrated out. In other words, the posterior distributions of our epidemiological parameters take into account tree uncertainty. This will become particularly useful for *M. tuberculosis* analyses as the low diversity in whole genome *M. tuberculosis* strains leads to high uncertainty in trees. For this one-step approach we jointly assume an epidemiological model such as the SIR model, which gives rise to a probability distribution over the tree space, and an evolutionary model such as GTR, which defines the probability distribution of the alignment of sequences. The output is a posterior distribution of trees, and the epidemiological and evolutionary parameters. Software packages BEAST2 (Bouckaert et al. 2014) and BEAST1 (Drummond et al. 2012) can simultaneously infer the evolutionary history and the epidemiological parameters under some simple epidemiological models.[2]

This one-step approach had been used for viral datasets such as HCV (Pybus et al. 2003), HIV (Stadler et al. 2012), or influenza (Kühnert et al. 2016). Phylodynamics on viruses can be done based on single genes as the substitution rates are high enough to see differences in single genes of the virus in donor compared to recipient. The slower-evolving *M. tuberculosis* requires whole genomes to reconstruct the phylogenetic trees and the epidemiological parameters. With next-generation sequencing technologies, such whole genomes become increasingly available, opening the door for phylodynamic analysis.

Such analyses have been done for assessing the rate of transmitted drug resistance. In Casali et al. (2014) the tips in the reconstructed phylogenetic tree were labelled according to the drug resistance status. Short inter-sequence distance was used to infer transmission links and to assess the transmission fitness costs in drug-resistant strains.

15.5 Practical Implications

Mathematical modelling helps us explain and predict the dynamics of tuberculosis, including the origins and future of strain diversity. Models aid in estimating the rates of transmission and reactivation, which in turn can influence the design of population interventions, and therefore models that incorporate at least the conclusions

[2]BEAST2 started out as a re-design of BEAST1, but over the course of time the two platforms continued to evolve independently with new features being implemented in both.

from strain diversity studies are of importance in targeting interventions to achieve WHO goals of eliminating *M. tuberculosis*.

15.5.1 Classification and Outbreaks

In a practical sense, model-based analysis of genetic diversity data for *M. tuberculosis* can be useful both for reactive purposes such as outbreak and contact investigations and longer-term policy definitions for addressing problems such as drug-resistance.

In relation to *M. tuberculosis* cases and contact interventions, modelling has most potential to be useful in high-resource environments where relevant data collection on cases and contacts provide nearly complete strain information including genotyping and sequencing of samples. A recent demonstration study suggested that use of WGS-based testing and identification of resistance was a more cost-effective solution for resistance testing and case follow-up than existing methods (Pankhurst et al. 2016).

Application of modelling tools such as BEAST2 to analyse genotype and whole genome sequencing data (Gardy et al. 2011; Outhred et al. 2016) for *M. tuberculosis* isolates can inform outbreak investigations by more precisely identifying links, which in turn has implications for epidemiological analyses of risk-factors for transmission and disease. In addition, modelling can be used to better understand measurement processes and biases in collection of samples, for instance the simulations conducted by Plazzotta et al. (2015) show how the prevalence of mixed-infection can be corrected for based on modelled properties of the infection and sample collection process.

A related question that can be informed by diversity data is the estimation of the proportions of recent transmission, relapse and reactivation. Determining this balance is important as it can help decide which aspects of treatment and prevention programs require attention. For instance a high rate of recent transmission would suggest prioritising case finding and treatment success rates, while high rates of reactivation might suggest the need for increased use of preventive therapy. Studies using WGS to investigate transmission were recently reviewed by Hatherell et al. (2016) who suggest that while these approaches are very helpful, improvements are still needed not only to data fidelity but also to the models of transmission trees and to the development of model-based thresholds for genetic distance to distinguish linked and unlinked cases.

15.5.2 Correlation Between Pathogen Genetics and Host Outcomes

A separate question that can be addressed with genetic data is whether differences in infection and disease outcomes are due to pathogen diversity or due to factors unrelated to the pathogen characteristics (see Chap. 5). For example, differences in progression to disease from infections with *M. africanum* as opposed to *M. tuberculosis* have been observed in a large epidemiological cohort study in the Gambia (de Jong et al. 2008).

While such differences in disease natural history between related species might be expected, it does raise the question of whether epidemiological differences in infection, progression or disease outcomes differ between strain groupings and whether this might need attention in terms of disease control. Evidence for this variation, discussed in Chap. 5 and in the comprehensive review by Coscolla and Gagneux (2014), demonstrates a range of differences between *M. tuberculosis* lineages at molecular, in vivo and in vitro levels. However, epidemiological evidence for special properties of, for example, the Beijing strain, is less conclusive and has not yet been a major focus in modelling studies. We note that Comas and Gagneux (2011) have argued for a "systems epidemiology" approach using computational models to address such questions and expect the increased use of the mathematical techniques such as those summarised here.

15.5.3 Dynamics of Resistance

A major practical focus for modelling strain diversity is the phenomena of multi-drug and extremely drug resistant tuberculosis. While poor individual outcomes and the high cost of such treatment have been influential in altering WHO policy for detecting and treating MDR-TB, models have played a key part in showing the potential for expansion of resistant strains (Blower and Gerberding 1998; Cohen and Murray 2004; Dye and Espinal 2001; Dye and Williams 2000) and identifying the need for drug-sensitivity testing as part of approaches to mitigate these effects. Implementation of treatment programs for MDR-TB in lower income settings has been facilitated by the development of molecular genetic tests such as *geneXpert* (Boehme et al. 2010) and more detailed modelling studies that directly assess and compare the cost-effectiveness of relevant population treatment strategies.

These issues have been explored primarily through deterministic epidemiological models, with strain heterogeneity simplified to sensitive and resistant classes, including the potential for latent infection with both resistant and sensitive TB (Colijn et al. 2009). These approaches are valuable for assessing the impacts of relative fitness under strong treatment related selection, either for active TB (Cohen and Murray 2004) or preventive therapy (Cohen et al. 2006) in latently infected populations.

Extensions of such models have also been applied to describe trends in MDR+ resistance through an expanded classification of resistance properties from mono-resistance through to MDR and XDR-TB (Menzies et al. 2012). Models with this additional detail on resistance allow prediction of the effects of molecular testing for resistance and appropriate on epidemic trajectories and the prevalence of MDR+ resistance. Modelling studies have in general found that tests such as *geneXpert* offer acceptable value for money even in lower resource settings and these studies have helped facilitate a rapid scale-up in testing and treatment for MDR-TB since 2013 (WHO 2016). There has been a recent proliferation of studies assessing the effectiveness and cost-effectiveness of such strategies, as summarised by Zwerling et al. (2016) who note that molecular tests have generally had positive cost-effectiveness findings but that future models need to feature increased use of setting-specific parameters in relation to treatment and diagnostic programs.

In 2015 WHO established its end-TB strategy which sets ambitious goals for reducing new cases by 90% by 2035. This target does not seem achievable without reducing the burden of reactivated TB cases through treatment of latent tuberculosis (Rangaka et al. 2015) and suggests that the prevalence of resistance in latent infections will have substantial impact on the success of preventative treatment strategies. It will be important to detect and know the extent of mixed infection involving sensitive and resistant bacteria (Cohen et al. 2012; Mills et al. 2013), through both data collection and use of models to correct biases in observations.

Models have also considered that increases in the prevalence of resistance relate to the variance of reproductive fitness among resistant strains. In accordance with expectations from evolutionary theory, variance in fitness enhances the success of resistance even if the mean fitness is relatively low (Knight et al. 2015). In particular these results indicate that while resistance that emerges under treatment will in general be poorly transmissible, transmission will nevertheless become dominated by resistant strains with the highest fitness (Knight et al. 2015) in the absence of rapid identification and appropriate treatment for MDR-TB. Such models have the advantage of illustrating how high-fitness resistant strains can gain an increase in prevalence in poorly controlled epidemics.

Dynamic models that consider strain diversity more directly have been less common and have been concerned with scientific questions such as estimation of fitness costs of resistance. For instance stochastic epidemiological models with genotype evolution have been used to estimate the relative fitness of drug resistant strains, and to estimate the relative importance of transmission of resistant strains versus acquired resistance (Luciani et al. 2009). The practical impli-

cation of such studies is to emphasise the value in reducing the average period of infectiousness for individuals with resistant *M. tuberculosis* infections. Models can then be used as decision tools to help guide us to the most effective means to achieve this goal, for example through increasing overall case-finding, reducing the time between identification of a case and tests for resistance or improving treatment compliance and outcomes. These findings can have both policy and research implications, for instance in terms of suggesting the characteristics of potential diagnostics (Dowdy et al. 2014) or treatment regimens that would be most beneficial (Zwerling et al. 2016).

15.5.4 Future Directions

Tuberculosis modelling is a rapidly growing field, with a number of key directions in which modelling research is progressing. In relation to models of *M. tuberculosis* variation, it is particularly important to refine models for whole genome sequencing (WGS) data analysis because much future data will be of this kind. We discussed some of the recent developments in this broad area in Sect. 15.4 but further work can be done. For example, more realistic models could be developed to link dynamics at within- and between-host levels.

Models of the within-host *M. tuberculosis* infection should ideally feature a more fine-grained characterisation of the natural history of disease, including interactions between the immune system and the pathogen population and understanding features of TB infection such as granulomas. Research in this area has moved from more theoretical explorations (Chang et al. 2005) to studies of potential biomarkers (Marino et al. 2016) and enhancements to therapy (Linderman and Kirschner 2015). One other open topic of research is the quantification of substitution rates for *M. tuberculosis* in the latent and in the acute stage of the disease.

Another aspect of tuberculosis epidemiology that has been relatively underexplored is the variation in host susceptibility to infection and disease. We briefly described approaches to describing host variation through models in Sect. 15.2.3. Future work may extend such models to consider coevolution between the host immune system and *M. tuberculosis*. We note that variation in hosts can be due to genetic or non-genetic factors and that although genetic susceptibility to tuberculosis infection has been studied (Bellamy et al. 2000) the important sources of host variation are arguably non-genetic factors such as age, HIV status and other immunosuppression, diabetes, BCG vaccination and living conditions including nutrition, crowding and smoking behaviour. While host factors such as these are commonly included in risk-prediction models for non-communicable diseases (e.g. coronary artery disease), their adoption in transmission models for *M. tuberculosis* has been slow. However, there is increasing recognition that social determinants of transmission and disease are important in characterizing *M. tuberculosis* epidemics and in setting priorities for control (Andrews et al. 2015). As host factors are often closely linked to key characteristics of treatment programs (e.g. compliance with treatment), there can be flow on effects to pathogen variation. As such, it is likely that there is value in integrated approaches that take both host and pathogen variation into account. We see these approaches as valuable not only in terms of explaining existing epidemiology but in more local prediction of epidemic outcomes under changes to control measures.

References

Aandahl RZ, Reyes JF, Sisson SA, Tanaka MM (2012) A model-based Bayesian estimation of the rate of evolution of VNTR loci in *Mycobacterium tuberculosis*. PLoS Comput Biol 8(6):e1002573. doi:10.1371/journal.pcbi.1002573, http://dx.doi.org/10.1371/journal.pcbi.1002573

Aandahl RZ, Stadler T, Sisson SA, Tanaka MM (2014) Exact vs. approximate computation: reconciling different estimates of *Mycobacterium tuberculosis* epidemiological parameters. Genetics 196(4):1227–1230. doi:10.1534/genetics.113.158808

Alland D, Kalkut G, Moss A, McAdam R, Hahn J, Bosworth W, Drucker E, Bloom B (1994) Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. N Engl J Med 330(24):1710–1716

Anderson RM, May RM (1979) Population biology of infectious diseases: Part I. Nature 280:361–367

Andrews JR, Basu S, Dowdy DW, Murray MB (2015) The epidemiological advantage of preferential targeting of tuberculosis control at the poor. Int J Tuberc Lung Dis 19(4):375–380

Aranaz A, Romero B, Montero N, Alvarez J, Bezos J, de Juan L, Mateos A, Domínguez L (2004) Spoligotyping profile change caused by deletion of a direct variable repeat in a *Mycobacterium tuberculosis* isogenic laboratory strain. J Clin Microbiol 42(11):5388–5391. doi:10.1128/JCM.42.11.5388-5391.2004, http://dx.doi.org/10.1128/JCM.42.11.5388-5391.2004

Basu S, Galvani AP (2009) The evolution of tuberculosis virulence. Bull Math Biol 71(5):1073–1088. doi:10.1007/s11538-009-9394-x, http://dx.doi.org/10.1007/s11538-009-9394-x

Becker N (2015) Modeling to inform infectious disease control. Chapman & Hall/CRC biostatistics series. CRC Press/Taylor & Francis, Boca Raton. https://books.google.com.au/books?id=F_MQrgEACAAJ

Bellamy R, Beyers N, McAdam KP, Ruwende C, Gie R, Samaai P, Bester D, Meyer M, Corrah T, Collin M, Camidge DR, Wilkinson D, Hoal-Van Helden E, Whittle HC, Amos W, van Helden P, Hill AV (2000) Genetic susceptibility to tuberculosis in Africans: a genome-wide scan. Proc Natl Acad Sci U S A 97(14):8005–8009. doi:10.1073/pnas.140201897, http://dx.doi.org/10.1073/pnas.140201897

Blower SM, Gerberding JL (1998) Understanding, predicting and controlling the emergence of drug-resistant tuberculosis: a theoretical framework. J Mol Med (Berl) 76(9):624–636

Blower SM, McLean AR, Porco TC, Small PM, Hopewell PC, Sanchez MA, Moss AR (1995) The intrinsic transmission dynamics of tuberculosis epidemics. Nature Med 1:815–821

Blower SM, Small PM, Hopewell PC (1996) Control strategies for tuberculosis epidemics: new models for old problems. Science 273:497–500

Boehme CC, Nabeta P, Hillemann D, Nicol MP, Shenai S, Krapp F, Allen J, Tahirli R, Blakemore R, Rustomjee R, Milovic A, Jones M, O'Brien SM, Persing DH, Ruesch-Gerdes S, Gotuzzo E, Rodrigues C, Alland D, Perkins MD (2010) Rapid molecular detection of tuberculosis and rifampin resistance. N Engl J Med 363(11):1005–1015

Borgdorff MW, van den Hof S, Kalisvaart N, Kremer K, van Soolingen D (2011) Influence of sampling on clustering and associations with risk factors in the molecular epidemiology of tuberculosis. Am J Epidemiol 174(2):243–251. doi:10.1093/aje/kwr061, http://dx.doi.org/10.1093/aje/kwr061

Borrell S, Gagneux S (2011) Strain diversity, epistasis and the evolution of drug resistance in *Mycobacterium tuberculosis*. Clin Microbiol Infect 17(6):815–820. doi:10.1111/j.1469-0691.2011.03556.x, http://www.ncbi.nlm.nih.gov/pubmed/21682802

Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, Forrest SA, Bryant JM, Harris SR, Schuenemann VJ et al (2014) Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature 514(7523):494–497

Bouckaert R, Heled J, Kuhnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol 10(4):e1003537. doi:10.1371/journal.pcbi.1003537, http://www.ncbi.nlm.nih.gov/pubmed/24722319

Brooks-Pollock E, Becerra MC, Goldstein E, Cohen T, Murray MB (2011) Epidemiologic inference from the distribution of tuberculosis cases in households in Lima, Peru. J Infect Dis 203(11):1582–1589. doi:10.1093/infdis/jir162, http://www.ncbi.nlm.nih.gov/pubmed/21592987

Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajoj SA, Allix C, Aristimuño L, Arora J, Baumanis V et al (2006) *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol 6(1):23

Casali N, Nikolayevskyy V, Balabanova Y, Harris SR, Ignatyeva O, Kontsevaya I, Corander J, Bryant J, Parkhill J, Nejentsev S, Horstmann RD, Brown T, Drobniewski F (2014) Evolution and transmission of drug-resistant tuberculosis in a Russian population. Nat Genet 46(3):279–286. doi:10.1038/ng.2878, http://www.ncbi.nlm.nih.gov/pubmed/24464101

Cave MD, Eisenach KD, McDermott PF, Bates JH, Crawford JT (1991) IS *6110*: conservation of sequence in the *Mycobacterium tuberculosis* complex and its utilization in DNA fingerprinting. Mol Cell Probes 5(1):73–80

Chang ST, Linderman JJ, Kirschner DE (2005) Multiple mechanisms allow *Mycobacterium tuberculosis* to continuously inhibit MHC class II-mediated antigen presentation by macrophages. Proc Natl Acad Sci U S A 102(12):4530–4535

Chindelevitch L, Colijn C, Moodley P, Wilson D, Cohen T (2016) ClassTR: classifying within-host heterogeneity based on tandem repeats with application to *Mycobacterium tuberculosis* infections. PLoS Comput Biol 12(2):e1004475. doi:10.1371/journal.pcbi.1004475, http://dx.doi.org/10.1371/journal.pcbi.1004475

Chisholm RH, Tanaka MM (2016) The emergence of latent infection in the early evolution of *Mycobacterium tuberculosis*. Proc Biol Sci 283(1831). doi:10.1098/rspb.2016.0499, http://dx.doi.org/10.1098/rspb.2016.0499

Chisholm RH, Trauer JM, Curnoe D, Tanaka MM (2016) Controlled fire use in early humans might have triggered the evolutionary emergence of tuberculosis. Proc Natl Acad Sci U S A 113(32):9051–9056. doi:10.1073/pnas.1603224113

Cohen T, Murray M (2004) Modeling epidemics of multidrug-resistant M. tuberculosis of heterogeneous fitness. Nat Med 10(10):1117–1121. doi:10.1038/nm1110, http://dx.doi.org/10.1038/nm1110

Cohen T, Lipsitch M, Walensky RP, Murray M (2006) Beneficial and perverse effects of isoniazid preventive therapy for latent tuberculosis infection in HIV-tuberculosis coinfected populations. Proc Natl Acad Sci U S A 103(18):7042–7047. doi:10.1073/pnas.0600349103, https://www.scopus.com/inward/record.uri?eid=2-s2.0-33646493800&partnerID=40&md5=affd24e21a2915b31375d88cb57caf88

Cohen T, van Helden PD, Wilson D, Colijn C, McLaughlin MM, Abubakar I, Warren RM (2012) Mixed-strain *Mycobacterium tuberculosis* infections and the implications for tuberculosis treatment and control. Clin Microbiol Rev 25(4):708–719. doi:10.1128/CMR.00021-12, https://www.scopus.com/inward/record.uri?eid=2-s2.0-84867198547&partnerID=40&md5=bb6952b52ec058f2821621f8f8b3c52d

Colangeli R, Arcus VL, Cursons RT, Ruthe A, Karalus N, Coley K, Manning SD, Kim S, Marchiano E, Alland D (2014) Whole genome sequencing of *Mycobacterium tuberculosis* reveals slow growth and low mutation rates during latent infections in humans. PLoS One 9(3):e91024. doi:10.1371/journal.pone.0091024, http://dx.doi.org/10.1371/journal.pone.0091024

Colijn C, Cohen T, Murray M (2009) Latent coinfection and the maintenance of strain diversity. Bull Math Biol 71(1):247–263. doi:10.1007/s11538-008-9361-y, http://dx.doi.org/10.1007/s11538-008-9361-y

Colijn C, Cohen T, Ganesh A, Murray M (2011) Spontaneous emergence of multiple drug resistance in tuberculosis before and during therapy. PLoS One 6(3):e18327. doi:10.1371/journal.pone.0018327, http://dx.doi.org/10.1371/journal.pone.0018327

Comas I, Gagneux S (2011) A role for systems epidemiology in tuberculosis research. Trends Microbiol 19(10):492–500

Comas I, Homolka S, Niemann S, Gagneux S (2009) Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. PLoS One 4(11):e7815. doi:10.1371/journal.pone.0007815, http://dx.doi.org/10.1371/journal.pone.0007815

Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, Parkhill J, Malla B, Berg S, Thwaites G et al (2013) Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. Nat Genet 45(10):1176–1182

Coscolla M, Gagneux S (2014) Consequences of genomic diversity in *Mycobacterium tuberculosis*. Semin Immunol 26(6):431–444. doi:10.1016/j.smim.2014.09.012, https://www.ncbi.nlm.nih.gov/pubmed/25453224

de Boer AS, Borgdorff MW, de Haas PE, Nagelkerke NJ, van Embden JD, van Soolingen D (1999) Analysis of rate of change of IS6110 RFLP patterns of *Mycobacterium tuberculosis* based on serial patient isolates. J Infect Dis 180(4):1238–1244. doi:10.1086/314979, http://dx.doi.org/10.1086/314979

Demay C, Liens B, Burguière T, Hill V, Couvin D, Millet J, Mokrousov I, Sola C, Zozio T, Rastogi N (2012) SITVITWEB – a publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. Infect Genet Evol 12(4):755–766

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci U S A 91(8):3166–3170

Didelot X, Gardy J, Colijn C (2014) Bayesian inference of infectious disease transmission from whole-genome sequence data. Mol Biol Evol 31(7):1869–1879. doi:10.1093/molbev/msu121, http://dx.doi.org/10.1093/molbev/msu121

Didelot X, Fraser C, Gardy J, Colijn C (2017) Genomic infectious disease epidemiology in partially sampled and ongoing outbreaks. Mol Biol Evol 34(4):997–1007.

Diekmann O, Heesterbeek JAP (2000) Mathematical epidemiology of infectious diseases: model building, analysis and interpretation. Wiley, Chichester

Diekmann O, Heesterbeek JAP, Metz JAJ (1990) On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. J Math Biol 28(4). doi:10.1007/bf00178324

Dietz K (1993) The estimation of the basic reproduction number for infectious diseases. Stat Methods Med Res 2(1):23–41

Dowdy DW, Dye C, Cohen T (2013) Data needs for evidence-based decisions: a tuberculosis modeler's 'wish list'. Int J Tuberc Lung Dis 17(7):866–877. doi:10.5588/ijtld.12.0573, http://www.ncbi.nlm.nih.gov/pubmed/23743307

Dowdy DW, Houben R, Cohen T, Pai M, Cobelens F, Vassall A, Menzies NA, Gomez GB, Langley I, Squire SB, White R (2014) Impact and cost-effectiveness of current and future tuberculosis diagnostics: the contribution of modelling. Int J Tuberc Lung Dis 18(9):1012–1018

Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. PLoS Biol 4(5):e88. doi:10.1371/journal.pbio.0040088, https://www.ncbi.nlm.nih.gov/pubmed/16683862

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution 29(8):1969–1973. doi:10.1093/molbev/mss075

Dye C, Espinal MA (2001) Will tuberculosis become resistant to all antibiotics? Proc R Soc Lond B Biol Sci 268(1462):45–52

Dye C, Williams BG (2000) Criteria for the control of drug-resistant tuberculosis. Proc Natl Acad Sci U S A 97(14):8180–8185

Dye C, Williams BG (2009) Slow elimination of multidrug-resistant tuberculosis. Sci Transl Med 1(3):3ra8. doi:10.1126/scitranslmed.3000346, http://dx.doi.org/10.1126/scitranslmed.3000346

Dye C, Williams BG (2010) The population dynamics and control of tuberculosis. Science 328(5980):856–861

Dye C, Garnett GP, Sleeman K, Williams BG (1998) Prospects for worldwide tuberculosis control under the WHO DOTS strategy. The Lancet 352(9144):1886–1891. http://dx.doi.org/10.1016/S0140-6736(98)03199-7, http://www.sciencedirect.com/science/article/pii/S0140673698031997

Ewens WJ (1972) The sampling theory of selectively neutral alleles. Theor Popul Biol 3(1):87–112

Ewens WJ (2004) Mathematical population genetics 1: theoretical introduction, vol 27, 2nd edn. Springer, New York

Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool 27(4):401–410. doi:10.2307/2412923, <GotoISI>://WOS:A1978GH36300002

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17(6):368–376. doi:10.1007/bf01734359, http://www.ncbi.nlm.nih.gov/pubmed/7288891

Felsenstein J (2004) Inferring phylogenies, vol 2. Sinauer Associates Sunderland

Filliol I, Motiwala AS, Cavatore M, Qi W, Hazbon MH, Bobadilla del Valle M, Fyfe J, Garcia-Garcia L, Rastogi N, Sola C, Zozio T, Guerrero MI, Leon CI, Crabtree J, Angiuoli S, Eisenach KD, Durmaz R, Joloba ML, Rendon A, Sifuentes-Osornio J, Ponce de Leon A, Cave MD, Fleischmann R, Whittam TS, Alland D (2006) Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. J Bacteriol 188(2):759–772. doi:10.1128/JB.188.2.759-772.2006, http://www.ncbi.nlm.nih.gov/pubmed/16385065

Fine PE, Vynnycky E (1998) The effect of heterologous immunity upon the apparent efficacy of (e.g. bcg) vaccines. Vaccine 16(20):1923–1928. http://dx.doi.org/10.1016/S0264-410X(98)00124-8, http://www.sciencedirect.com/science/article/pii/S0264410X98001248

Ford CB, Lin PL, Chase MR, Shah RR, Iartchouk O, Galagan J, Mohaideen N, Ioerger TR, Sacchettini JC, Lipsitch M, Flynn JL, Fortune SM (2011) Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. Nat Genet 43(5):482–486. doi:10.1038/ng.811, http://dx.doi.org/10.1038/ng.811

Ford CB, Shah RR, Maeda MK, Gagneux S, Murray MB, Cohen T, Johnston JC, Gardy J, Lipsitch M, Fortune SM (2013) *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. Nat Genet 45(7):784–790

Gagneux S (2012) Host-pathogen coevolution in human tuberculosis. Philos Trans R Soc B 367(1590):850–859

Gagneux S, Small PM (2007) Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. Lancet Infect Dis 7(5):328–337. doi:10.1016/S1473-3099(07)70108-1, http://dx.doi.org/10.1016/S1473-3099(07)70108-1

Gagneux S, Long CD, Small PM, Van T, Schoolnik GK, Bohannan BJM (2006) The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. Science 312(5782):1944–1946. doi:10.1126/science.1124410, http://dx.doi.org/10.1126/science.1124410

Galagan JE (2014) Genomic insights into tuberculosis. Nat Rev Genet 15(5):307–320. doi:10.1038/nrg3664, http://dx.doi.org/10.1038/nrg3664

Gammack D, Doering CR, Kirschner DE (2004) Macrophage response to *Mycobacterium tuberculosis* infection. J Math Biol 48(2):218–242

Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodkin E, Rempel S, Moore R, Zhao Y, Holt R, Varhol R, Birol I, Lem M, Sharma MK, Elwood K, Jones SJM, Brinkman FSL, Brunham RC, Tang P (2011) Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N Engl J Med 364(8):730–739. doi:10.1056/NEJMoa1003176, http://dx.doi.org/10.1056/NEJMoa1003176

Glynn JR, Bauer J, de Boer AS, Borgdorff MW, Fine PE, Godfrey-Faussett P, Vynnycky E (1999a) Interpreting DNA fingerprint clusters of *Mycobacterium tuberculosis*. European Concerted Action on Molecular Epidemiology and Control of Tuberculosis. Int J Tuberc Lung Dis 3(12):1055–1060

Glynn JR, Vynnycky E, Fine PE (1999b) Influence of sampling on estimates of clustering and recent transmission of *Mycobacterium tuberculosis* derived from DNA fingerprinting techniques. Am J Epidemiol 149(4):366–371

Gomes MGM, Franco AO, Gomes MC, Medley GF (2004) The reinfection threshold promotes variability in tuberculosis epidemiology and vaccine efficacy. Proc Biol Sci 271(1539):617–623. doi:10.1098/rspb.2003.2606, http://dx.doi.org/10.1098/rspb.2003.2606

Grant A, Arnold C, Thorne N, Gharbia S, Underwood A (2008) Mathematical modelling of *Mycobacterium tuberculosis* VNTR loci estimates a very slow mutation rate for the repeats. J Mol Evol 66(6):565–574. doi:10.1007/s00239-008-9104-6, http://dx.doi.org/10.1007/s00239-008-9104-6

Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, Mumford JA, Holmes EC (2004) Unifying the epidemiological and evolutionary dynamics of pathogens. Science 303(5656):327–332. doi:10.1126/science.1090727, http://dx.doi.org/10.1126/science.1090727

Hasegawa M, Kishino H, Yano Ta (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22(2): 160–174

Hatherell HA, Colijn C, Stagg HR, Jackson C, Winter JR, Abubakar I (2016) Interpreting whole genome sequencing for investigating tuberculosis transmission: a systematic review. BMC Med 14:21. doi:10.1186/s12916-016-0566-x, https://www.ncbi.nlm.nih.gov/pubmed/27005433

Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, Roach JC, Kremer K, Petrov DA, Feldman MW, Gagneux S (2008) High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. PLoS Biol 6(12):e311. doi:10.1371/journal.pbio.0060311, http://dx.doi.org/10.1371/journal.pbio.0060311

Ho SYW, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol Biol Evol 22(7):1561–1568. doi:10.1093/molbev/msi145, http://dx.doi.org/10.1093/molbev/msi145

Houk VN (1980) Spread of tuberculosis via recirculated air in a naval vessel: the Byrd study. Ann N Y Acad Sci 353:10–24. http://www.ncbi.nlm.nih.gov/pubmed/6939378

de Jong BC, Hill PC, Aiken A, Awine T, Antonio M, Adetifa IM, Jackson-Sillah DJ, Fox A, Deriemer K, Gagneux S, Borgdorff MW, McAdam KP, Corrah T, Small PM, Adegbola RA (2008) Progression to active tuberculosis, but not transmission, varies by *Mycobacterium tuberculosis* lineage in The Gambia. J Infect Dis 198(7):1037–1043. doi:10.1086/591504, https://www.ncbi.nlm.nih.gov/pubmed/18702608

Jukes T, Cantor C (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism, vol 3. Academic, New York, pp 21–132

Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, van Embden J (1997) Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol 35(4):907–914

Keeling MJ, Rohani P (2008) Modeling infectious diseases in humans and animals. Princeton University Press, Princeton

Kendall EA, Fofana MO, Dowdy DW (2015) Burden of transmitted multidrug resistance in epidemics of tuberculosis: a transmission modelling analysis. Lancet Respir Med 3(12):963–972. doi:10.1016/s2213-2600(15)00458-0

Kermack WO, McKendrick AG (1927) Contributions to the mathematical theory of epidemics. Proc R Soc Ser 115:700–721

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16(2):111–120

Knight GM, Colijn C, Shrestha S, Fofana M, Cobelens F, White RG, Dowdy DW, Cohen T (2015) The distribution of fitness costs of resistance-conferring mutations is a key determinant for the future burden of drug-resistant tuberculosis: a model-based analysis. Clin Infect Dis 61(Suppl 3):S147–S154. doi:10.1093/cid/civ579, http://dx.doi.org/10.1093/cid/civ579

Kühnert D, Stadler T, Vaughan TG, Drummond AJ (2014) Simultaneous reconstruction of evolutionary history and epidemiological dynamics from viral sequences with the birth-death SIR model. J R Soc Interface 11(94):20131106. doi:10.1098/rsif.2013.1106, http://dx.doi.org/10.1098/rsif.2013.1106

Kühnert D, Stadler T, Vaughan TG, Drummond AJ (2016) Phylodynamics with migration: a computational framework to quantify population structure from genomic data. Mol Biol Evol. doi:10.1093/molbev/msw064, http://dx.doi.org/10.1093/molbev/msw064

Lillebaek T, Norman A, Rasmussen EM, Marvig RL, Folkvardsen DB, Andersen ÅB, Jelsbak L (2016) Substantial molecular evolution and mutation rates in prolonged latent *Mycobacterium tuberculosis* infection in humans. Int J Med Microbiol. doi:10.1016/j.ijmm.2016.05.017, http://dx.doi.org/10.1016/j.ijmm.2016.05.017

Linderman JJ, Kirschner DE (2015) In silico models of m. tuberculosis infection provide a route to new therapies. Drug Discov Today Dis Models 15:37–41. http://dx.doi.org/10.1016/j.ddmod.2014.02.006, http://www.sciencedirect.com/science/article/pii/S1740675714000115. Computational models of lung diseases

Lipsitch M, Levin BR (1997) The within-host population dynamics of antibacterial chemotherapy: conditions for the evolution of resistance. Ciba Found Symp 207:112–127; discussion 127–130

Liu X, Gutacker MM, Musser JM, Fu YX (2006) Evidence for recombination in *Mycobacterium tuberculosis*. J Bacteriol 188(23):8169–8177. doi:10.1128/JB.01062-06, http://www.ncbi.nlm.nih.gov/pubmed/16997954

Luciani F, Francis AR, Tanaka MM (2008) Interpreting genotype cluster sizes of *Mycobacterium tuberculosis* isolates typed with IS6110 and spoligotyping. Infect Genet Evol 8(2):182–190. doi:10.1016/j.meegid.2007.12.004, http://dx.doi.org/10.1016/j.meegid.2007.12.004

Luciani F, Sisson SA, Jiang H, Francis AR, Tanaka MM (2009) The epidemiological fitness cost of drug resistance in *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A 106(34):14711–14715. doi:10.1073/pnas.0902437106, http://dx.doi.org/10.1073/pnas.0902437106

Marino S, Gideon HP, Gong C, Mankad S, McCrone JT, Lin PL, Linderman JJ, Flynn JL, Kirschner DE (2016) Computational and empirical studies predict *Mycobacterium tuberculosis*-specific T cells as a biomarker for infection outcome. PLoS Comput Biol 12(4):e1004804

Mathema B, Kurepina NE, Bifani PJ, Kreiswirth BN (2006) Molecular epidemiology of tuberculosis: current insights. Clin Microbiol Rev 19(4):658–685. doi:10.1128/CMR.00061-05, http://dx.doi.org/10.1128/CMR.00061-05

Menzies NA, Cohen T, Lin HH, Murray M, Salomon JA (2012) Population health impact and cost-effectiveness of tuberculosis diagnosis with Xpert MTB/RIF: a dynamic simulation and economic evaluation. PLoS Med. 9(11). doi:10.1371/journal.pmed.1001347, https://www.scopus.com/inward/record.uri?eid=2-s2.0-84870266092&partnerID=40&md5=557a31f6d3f2341c03f44cb526513bdf

Mills HL, Cohen T, Colijn C (2013) Community-wide isoniazid preventive therapy drives drug-resistant tuberculosis: a model-based analysis. Sci Transl Med 5(180). doi:10.1126/scitranslmed.3005260, https://www.scopus.com/inward/record.uri?eid=2-s2.0-84877765955&partnerID=40&md5=feec00c0015ae17a30d3181030db7f0c

Murray M (2002a) Determinants of cluster distribution in the molecular epidemiology of tuberculosis. Proc Natl Acad Sci U S A 99(3):1538–1543

Murray M (2002b) Sampling bias in the molecular epidemiology of tuberculosis. Emerg Infect Dis 8(4):363–369

Murray M, Alland D (2002) Methodological problems in the molecular epidemiology of tuberculosis. Am J Epidemiol 155(6):565–571

O'Neill MB, Mortimer TD, Pepperell CS (2015) Diversity of *Mycobacterium tuberculosis* across evolutionary scales. PLoS Pathog 11(11):e1005257. doi:10.1371/journal.ppat.1005257, http://dx.doi.org/10.1371/journal.ppat.1005257

Outhred AC, Holmes N, Sadsad R, Martinez E, Jelfs P, Hill-Cawthorne GA, Gilbert GL, Marais BJ, Sintchenko V (2016) Identifying likely transmission pathways within a 10-Year community outbreak of tuberculosis by high-depth whole genome sequencing. PLoS One 11(3):e0150550. doi:10.1371/journal.pone.0150550, https://www.ncbi.nlm.nih.gov/pubmed/26938641

Ozcaglar C, Shabbeer A, Vandenberg S, Yener B, Bennett KP (2011) Sublineage structure analysis of *Mycobacterium tuberculosis* complex strains using multiple-biomarker tensors. BMC Genomics 12(Suppl 2):S1. doi:10.1186/1471-2164-12-S2-S1, http://dx.doi.org/10.1186/1471-2164-12-S2-S1

Ozcaglar C, Shabbeer A, Vandenberg SL, Yener B, Bennett KP (2012) Epidemiological models of *Mycobacterium tuberculosis* complex infections. Math Biosci 236(2):77–96. doi:10.1016/j.mbs.2012.02.003, http://dx.doi.org/10.1016/j.mbs.2012.02.003

Pankhurst LJ, Del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies J, Fermont JM, Gascoyne-Binzi DM, Kohl TA, Kong C, Lemaitre N, Niemann S, Paul J, Rogers TR, Roycroft E, Smith EG, Supply P, Tang P, Wilcox MH, Wordsworth S, Wyllie D, Xu L, Crook DW, Group CTS (2016) Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. Lancet Respir Med 4(1):49–58. doi:10.1016/S2213-2600(15)00466-X, https://www.ncbi.nlm.nih.gov/pubmed/26669893

Plazzotta G, Cohen T, Colijn C (2015) Magnitude and sources of bias in the detection of mixed strain *M. tuberculosis* infection. J Theor Biol 368:67–73

Pybus OG, Drummond AJ, Nakano T, Robertson BH, Rambaut A (2003) The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: a Bayesian coalescent approach. Mol Biol Evol 20(3):381–387. doi:10.1093/molbev/msg043, https://www.ncbi.nlm.nih.gov/pubmed/12644558

Ragheb MN, Ford CB, Chase MR, Lin PL, Flynn JL, Fortune SM (2013) The mutation rate of mycobacterial repetitive unit loci in strains of M. tuberculosis from cynomolgus macaque infection. BMC Genomics 14:145. doi:10.1186/1471-2164-14-145, http://dx.doi.org/10.1186/1471-2164-14-145

Rangaka MX, Cavalcante SC, Marais BJ, Thim S, Martinson NA, Swaminathan S, Chaisson RE (2015) Controlling the seedbeds of tuberculosis: diagnosis and treatment of tuberculosis infection. Lancet 386(10010):2344–2353. doi:10.1016/S0140-6736(15)00323-2, https://www.ncbi.nlm.nih.gov/pubmed/26515679

Reyes JF, Tanaka MM (2010) Mutation rates of spoligotypes and variable numbers of tandem repeat loci in *Mycobacterium tuberculosis*. Infect Genet Evol 10(7):1046–1051. doi:10.1016/j.meegid.2010.06.016, http://dx.doi.org/10.1016/j.meegid.2010.06.016

Reyes JF, Francis AR, Tanaka MM (2008) Models of deletion for visualizing bacterial variation: an application to tuberculosis spoligotypes. BMC Bioinformatics 9:496. doi:10.1186/1471-2105-9-496, http://dx.doi.org/10.1186/1471-2105-9-496

Reyes JF, Chan CHS, Tanaka MM (2012) Impact of homoplasy on variable numbers of tandem repeats and spoligotypes in *Mycobacterium tuberculosis*. Infect Genet Evol 12(4):811–818. doi:10.1016/j.meegid.2011.05.018, http://dx.doi.org/10.1016/j.meegid.2011.05.018

Rosenberg NA, Tsolaki AG, Tanaka MM (2003) Estimating change rates of genetic markers using serial samples: applications to the transposon IS6110 in *Mycobacterium tuberculosis*. Theor Popul Biol 63(4):347–363

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4):406–425. http://www.ncbi.nlm.nih.gov/pubmed/3447015

Sergeev R, Colijn C, Cohen T (2011) Models to understand the population-level impact of mixed strain M. tuberculosis infections. J Theor Biol 280(1):88–100. doi:10.1016/j.jtbi.2011.04.011, http://dx.doi.org/10.1016/j.jtbi.2011.04.011

Shabbeer A, Cowan LS, Ozcaglar C, Rastogi N, Vandenberg SL, Yener B, Bennett KP (2012a) TB-Lineage: an online tool for classification and analysis of strains of *Mycobacterium tuberculosis* complex. Infect Genet Evol 12(4):789–797.

doi:10.1016/j.meegid.2012.02.010, http://dx.doi.org/10.1016/j.meegid.2012.02.010

Shabbeer A, Ozcaglar C, Yener B, Bennett KP (2012b) Web tools for molecular epidemiology of tuberculosis. Infect Genet Evol 12(4):767–781. doi:10.1016/j.meegid.2011.08.019, http://dx.doi.org/10.1016/j.meegid.2011.08.019

Shrestha S, Knight GM, Fofana M, Cohen T, White RG, Cobelens F, Dowdy DW (2014) Drivers and trajectories of resistance to new first-line drug regimens for tuberculosis. Open Forum Infect Dis 1(2):ofu073. doi:10.1093/ofid/ofu073, http://dx.doi.org/10.1093/ofid/ofu073

Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, Schecter GF, Daley CL, Schoolnik GK (1994) The epidemiology of tuberculosis in San Francisco: a population-based study using conventional and molecular methods. N Engl J Med 330:1703–1709

Smith NH, Hewinson RG, Kremer K, Brosch R, Gordon SV (2009) Myths and misconceptions: the origin and evolution of *Mycobacterium tuberculosis*. Nat Rev Microbiol 7(7):537–544. doi:10.1038/nrmicro2165, http://dx.doi.org/10.1038/nrmicro2165

Stadler T (2011) Inferring epidemiological parameters on the basis of allele frequencies. Genetics 188(3):663–672. doi:10.1534/genetics.111.126466, http://dx.doi.org/10.1534/genetics.111.126466

Stadler T, Bonhoeffer S (2013) Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods. Philos Trans R Soc B: Biolog Sci 368(1614):20120198

Stadler T, Kouyos RD, von Wyl V, Yerly S, Böni J, Bürgisser P, Klimkait T, Joos B, Rieder P, Xie D, Günthard HF, Drummond A, Bonhoeffer S, the Swiss HIV Cohort Study (2012) Estimating the basic reproductive number from viral sequence data. Mol Biol Evol 29:347–357

Stadler T, Kuhnert D, Bonhoeffer S, Drummond AJ (2013) Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proc Natl Acad Sci U S A 110(1):228–233. doi:10.1073/pnas.1207965110, http://www.ncbi.nlm.nih.gov/pubmed/23248286

Steel M (2013) Consistency of bayesian inference of resolved phylogenetic trees. J Theor Biol 336:246–249

Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, Willery E, Savine E, de Haas P, van Deutekom H, Roring S et al (2006) Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. J Clin Microbiol 44(12):4498–4510

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10(3):512–526

Tanaka MM, Francis AR (2005) Methods of quantifying and visualising outbreaks of tuberculosis using genotypic information. Infect Genet Evol 5(1):35–43. doi:10.1016/j.meegid.2004.06.001, http://dx.doi.org/10.1016/j.meegid.2004.06.001

Tanaka MM, Rosenberg NA, Small PM (2004) The control of copy number of IS6110 in *Mycobacterium tuberculosis*. Mol Biol Evol 21(12):2195–2201. doi:10.1093/molbev/msh234, http://dx.doi.org/10.1093/molbev/msh234

Tanaka MM, Francis AR, Luciani F, Sisson SA (2006) Using approximate Bayesian computation to estimate tuberculosis transmission parameters from genotype data. Genetics 173(3):1511–1520. doi:10.1534/genetics.106.055574, http://dx.doi.org/10.1534/genetics.106.055574

Tavaré S (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. Lect Math Life Sci 17:57–86

To TH, Jung M, Lycett S, Gascuel O (2016) Fast Dating Using Least-Squares Criteria and Algorithms. Syst Biol 65(1):82–97. doi:10.1093/sysbio/syv068, http://www.ncbi.nlm.nih.gov/pubmed/26424727

Trauer JM, Denholm JT, McBryde ES (2014) Construction of a mathematical model for tuberculosis transmission in highly endemic regions of the Asia-Pacific. J Theor Biol 358:74–84. doi:10.1016/j.jtbi.2014.05.023, http://dx.doi.org/10.1016/j.jtbi.2014.05.023

Vogler AJ, Keys C, Nemoto Y, Colman RE, Jay Z, Keim P (2006) Effect of repeat copy number on variable-number tandem repeat mutations in Escherichia coli O157:H7. J Bacteriol 188(12):4253–4263. doi:10.1128/JB.00001-06, http://dx.doi.org/10.1128/JB.00001-06

Volz EM (2012) Complex population dynamics and the coalescent under neutrality. Genetics 190(1):187–201. doi:10.1534/genetics.111.134627, http://dx.doi.org/10.1534/genetics.111.134627

Volz EM, Kosakovsky Pond SL, Ward MJ, Leigh Brown AJ, Frost SDW (2009) Phylodynamics of infectious disease epidemics. Genetics 183(4):1421–1430. doi:10.1534/genetics.109.106021, http://dx.doi.org/10.1534/genetics.109.106021

Vynnycky E, Fine PEM (1997) The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. Epidemiol Infect 119(2):183–201. doi:10.1017/S0950268897007917, <GotoISI>://WOS:A1997YD95900010

Waaler H, Geser A, Andersen S (1962) The use of mathematical models in the study of the epidemiology of tuberculosis. Am J Public Health Nat Health 52(6):1002–1013

Wakeley J (2009) Coalescent theory: an introduction. Roberts and Co., Greenwood Village

Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, Eyre DW, Wilson DJ, Hawkey PM, Crook DW et al (2013) Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. Lancet Infect Dis 13(2):137–146

Warren RM, van der Spuy GD, Richardson M, Beyers N, Borgdorff MW, Behr MA, van Helden PD (2002) Calculation of the stability of the IS6110 banding pattern in patients with persistent *Mycobacterium tuberculosis* disease. J Clin Microbiol 40(5):1705–1708

Warren RM, Victor TC, Streicher EM, Richardson M, Beyers N, Gey van Pittius NC, van Helden PD (2004) Patients with active tuberculosis often have different strains in the same sputum specimen. Am J Respir Crit Care Med 169(5):610–614. doi:10.1164/rccm.200305-714OC, http://dx.doi.org/10.1164/rccm.200305-714OC

Weniger T, Krawczyk J, Supply P, Harmsen D, Niemann S (2012) Online tools for polyphasic analysis of *Mycobacterium tuberculosis* complex genotyping data: now and next. Infect Genet Evol 12(4):748–754

WHO (2016) Global tuberculosis report 2016. Tech. rep., World Health Organization

Wigginton JE, Kirschner D (2001) A model to predict cell-mediated immune regulatory mechanisms during human infection with *Mycobacterium tuberculosis*. J Immunol 166(3):1951–1967

Wirth T, Hildebrand F, Allix-Béguec C, Wölbeling F, Kubica T, Kremer K, van Soolingen D, Rüsch-Gerdes S, Locht C, Brisse S, Meyer A, Supply P, Niemann S (2008) Origin, spread and demography of the *Mycobacterium tuberculosis* complex. PLoS Pathog 4(9):e1000160. doi:10.1371/journal.ppat.1000160, http://dx.doi.org/10.1371/journal.ppat.1000160

Yang Z (1994a) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J Mol Evol 39(3):306–314. doi:10.1007/BF00160154, http://www.ncbi.nlm.nih.gov/pubmed/7932792

Yang Z (1994b) Statistical properties of the maximum likelihood method of phylogenetic estimation and comparison with distance matrix methods. Syst Biol 43(3):329–342

Yang Z (1996) Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol Evol 11(9):367–372. doi:10.1016/0169-5347(96)10041-0

Yang Z (2014) Molecular evolution: a statistical approach. Oxford University Press, Oxford

Zainuddin Z, Dale J (1990) Does *Mycobacterium tuberculosis* have plasmids? Tubercle 71:43–49. doi:10.1016/0041-3879(90)90060-L, http://www.sciencedirect.com/science/article/pii/004138799090060L

Zheng N, Whalen CC, Handel A (2014) Modeling the potential impact of host population survival on the evolution of M. tuberculosis latency. PLoS One 9(8):e105721. doi:10.1371/journal.pone.0105721, http://dx.doi.org/10.1371/journal.pone.0105721

Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. Evol Genes Proteins 97:97–166

Zwerling A, Gomez GB, Pennington J, Cobelens F, Vassall A, Dowdy DW (2016) A simplified cost-effectiveness model to guide decision-making for shortened anti-tuberculosis treatment regimens. Int J Tuberc Lung Dis 20(2):257–260

Index

S. Gagneux (ed.), *Strain Variation in the* Mycobacterium tuberculosis *Complex: Its Role in Biology, Epidemiology and Control*, Advances in Experimental Medicine and Biology 1019,
DOI 10.1007/978-3-319-64371-7