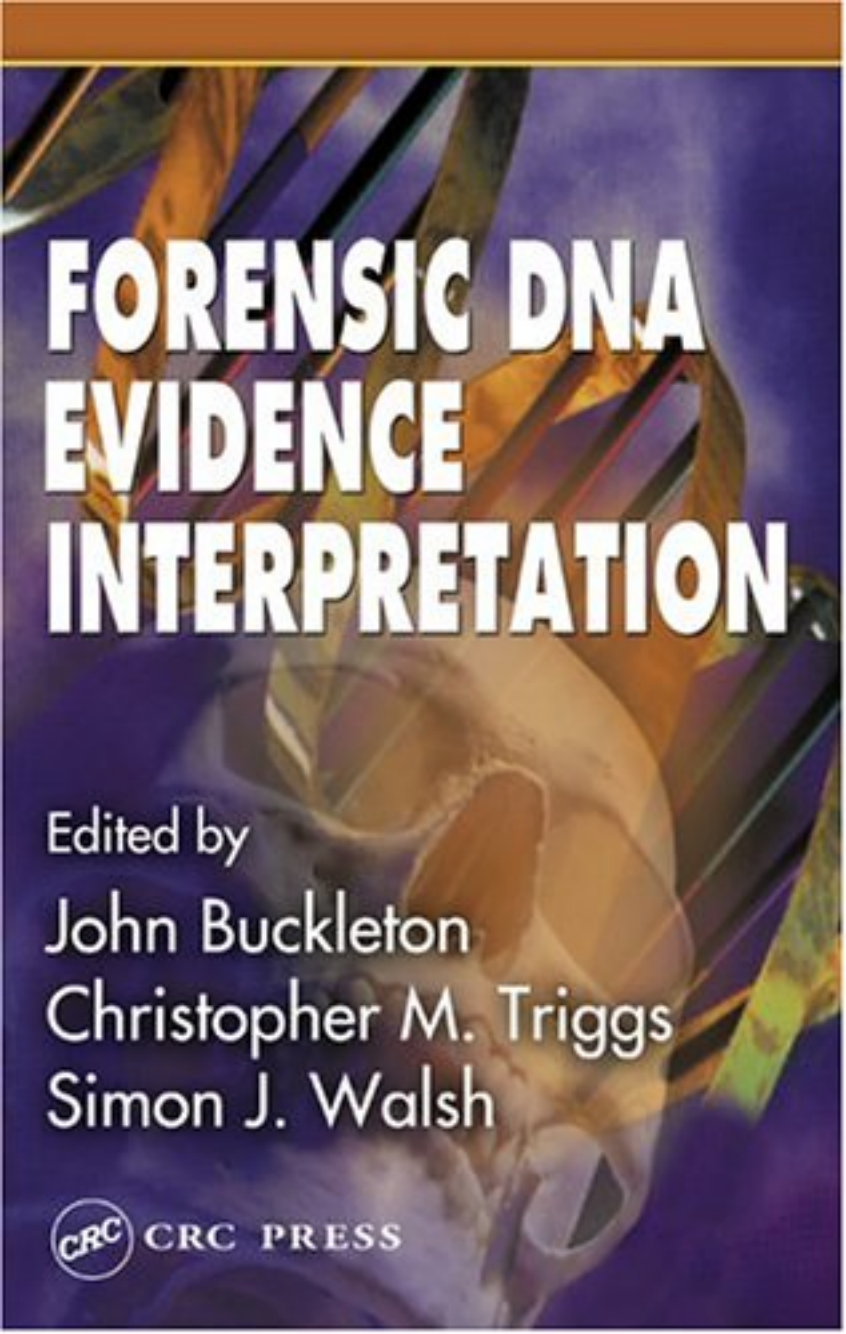

FORENSIC DNA
EVIDENCE
INTERPRETATION
Edited by
John Buckleton
Christopher M. Triggs
Simon J. Walsh
CRC PRESS

FORENSIC DNA EVIDENCE INTERPRETATION

Edited by

John Buckleton
Principal Scientist
ESR
Aukland, New Zealand

Christopher M. Triggs
Associate Professor
Department of Statistics
University of Aukland
Aukland, New Zealand

Simon J. Walsh
Lecturer in Forensic Biology
Centre for Forensic Science
University of Technology
Sydney, Australia

CRC PRESS

Boca Raton London New York Washington, D.C.

Library of Congress Cataloging-in-Publication Data

Catalog record is available from the Library of Congress

Visit the CRC Press Web site at www.crcpress.com

International Standard Book Number 0-8493-3017-3
Printed in the United States of America 1 2 3 4 5 6 7 8 9 0
Printed on acid-free paper

Preface

Forensic science is, to some extent, a derived science. It is happy to borrow technology and ideas from other sciences. There is, however, a "forensic mindset" and ethos that is peculiar to our science. When DNA technology was launched, the interpretation was attempted by forensic scientists such as Ian Evett and John Buckleton. Eventually it became clear, or indeed we had it rammed into our heads, that there was a great amount of classical population genetic work that needed to be considered by forensic scientists. This was brought to the world's attention by David Balding, Peter Donnelly, Richard Nichols, and Bruce Weir. Forensic science is very fortunate to have these fine minds working on their problems, and we are personally deeply indebted to Bruce Weir who has contributed so much to the field and several, otherwise unpublished sections to this book in areas that we could not solve ourselves.

Bruce Weir sought to bring a logical rigor to the interpretation of DNA evidence and carried this out through elegant papers, lectures, and eventually in his great textbook with Ian Evett. He has set the standard for forensic thinking and testimony.

This book is written to be Evett and Weir compatible. We have kept the nomenclature while including the developments in the intervening six years. This book is written from the perspective of less mathematically attuned caseworkers. We have also made some effort to review pertinent areas that have arisen during court proceedings.

This text is heavily referenced, and in many cases these references are "personal communications" or restricted material. This may be frustrating for the reader who wants to obtain these texts, and we have previously been criticized in reviews for doing this.

There are several reasons to reference a piece of work, whether it is published or not. One is to direct the reader to further reading. However, another is to give credit to the author of an idea. Therefore, in many cases we have tried to attribute an idea to the originator. Where we have failed to do this, we apologize and would welcome correction. We have also quoted the original texts extensively. Often the original authors stated the matter better than we possibly could, and it is often interesting to see how early

some illuminating comments were made. We would have included many more quotations from original publications if permissions had been more forthcoming.

We have also attempted to tabulate the formulae needed for routine forensic DNA casework. We have found many errors in the literature and have sought to correct these. However, no one is error free and we would welcome any corrections to our own tables.

We have had sections of texts read to us in court many times. To any caseworker who is read a section from this text, please direct the prosecutor or defense counsel to this preface. No author is perfect, and writing a text does not make one an authority. In many cases the caseworker has studied the case in question to an extent that advice from some "quoted authority" is completely irrelevant.

Above all, our goal is to provide a link between the biological, forensic, and interpretative (or statistical) domains of the DNA profiling field. It is a challenge for caseworkers to keep apace of the ever-changing technological and operational demands of their role and, additionally, to accurately assess the strength of the evidence under these fluctuating circumstances. We hope this book can act as a guide, or template, via which many of the complex issues can be tackled.

Acknowledgments

Many individuals and institutions have supported the production and compilation of this book. Specifically, we would like to mention the following noteworthy examples.

Naturally, we are thankful for the expert contributions of the authors, whose specific chapters have combined (we hope) to form a valuable collective work.

We would like to gratefully thank Ms. Catherine McGovern for the excellent proof reading of Chapters 1 and 2 and Dr. Jill Vintiner for an equally excellent proof reading of Chapters 3 and 4.

Dr. Ian Evett gave extensive comments upon the sections "Common source" and "O.J. Simpson."

Dr. SallyAnn Harbison, a co-author, assisted with many sections outside those directly attributed to her.

We have extensively used unpublished work by Professor Bruce Weir, Dr. James Curran, Associate Professor Christopher Triggs, and Dr. Ian Painter. In many cases, these were unpublished research results. Much material was developed in "conversations." We particularly mention Dr. Christophe Champod, Dr. Laszlo Szabo, Dr. Henry Roberts, Dr. Oscar Garcia, and Dr. Tim Sliter in this regard.

Mr. Wayne Chisnall, General Manager (Forensic) of ESR Ltd., the New Zealand Forensic Science Service, was extremely supportive throughout this undertaking. The venture of writing this book obviously made no money and cleared no casework. Nonetheless, he put both his moral and practical support behind this book. ESR has also supported the scientists in it by giving us two Ph.D. scholarships that were invested in gifted students in this field and by extensive support of research and development efforts. This is even more important for a small organization than it is for a large one.

Jason Ashton and Claire Winchester are our information specialists. They made a huge contribution and gave extensive support.

We would also like to acknowledge Becky McEldowney, our representative at CRC Press. We are thankful for your patience, your understanding, your open endorsements of our efforts, and the effective link that you provided between us and your organization.

Final versions of this work were significantly improved through the insightful comments of Professor Bruce Weir and those of our anonymous reviewer. Any errors that persist are our responsibility alone.

On many occasions we thought of throwing the venture in. We are each indebted to various individuals for their key comments of support and encouragement. We also appreciate the support we have given and received from each other in circumstances when we may have felt pressured, frustrated, and fatigued.

John S. Buckleton

Christopher M. Triggs

Simon J. Walsh

The Editors

John Buckleton is Principal Scientist at ESR in Auckland, New Zealand. He has more than 20 years experience in casework and testifying in New Zealand, the United Kingdom, and the United States where he has been employed for various governmental agencies on forensic work. Dr. Buckleton has co-authored more than 100 publications largely in the forensic field, holds four patents and has presented DNA training courses in the UK, USA, Australia, Asia, and New Zealand. He has delivered plenary addresses at speaking engagements worldwide, and is the holder of a number of awards.

Christopher Triggs is Professor in the Department of Statistics at the University of Auckland in New Zealand. He has been a respected statistician for 23 years, specializing in forensic statistics since the mid 1980s. Dr. Triggs has published more than 100 papers in a wide variety of scientific fields, and has three patents. His research interests include population genetics and the application of statistical methods in many fields of science, especially forensic science. He has lectured extensively on these subjects in Australasia, is often called on as a referee for journals in the field, and is a consultant to two major forensic organizations.

Simon J. Walsh is a lecturer in Forensic Biology at the Centre for Forensic Science at the University of Technology in Sydney, Australia. Prior to his involvement at the University of Technology, Simon J. Walsh was a practicing forensic scientist for eight years in Australian and New Zealand laboratories. Simon is credited with more than 15 publications. His research interests include the use of DNA databases, nonautosomal DNA markers, forensic statistics, population genetics, and intelligence-based uses of forensic case data. His current doctoral studies focus on evaluating the role and impact of forensic DNA profiling in the context of the criminal justice system.

Contributors

Peter Gill, B.Sc., Ph.D.
Forensic Science Service
Birmingham, England

Timothy Mark Clayton, Ph.D.
Forensic Science Service
Wetherby, England

James Michael Curran, Ph.D.
Department of Statistics
University of Waikato
Waikato, New Zealand

SallyAnn Harbison, B.Sc., Ph.D.
Science Leader ESR
Auckland, New Zealand

Contents

Biological Basis for DNA Evidence

1

PETER GILL AND JOHN BUCKLETON

Contents

This book deals in large part with the interpretation of DNA profiles, mixed or unmixed, after they have been collected, stored, transferred, and finally analyzed in the laboratory. The supposition throughout is that the earlier stages in the chain that leads to evidence in court have been undertaken correctly. The

inference at the final end of the chain is practically useless unless all these earlier aspects have been undertaken with due attention to continuity and integrity.[838]

This chapter gives a brief background to the biotechnology relevant to the interpretation of short tandem repeat (STR) samples. For an extended discussion, see the excellent work by Rudin and Inman.[678,679]

1.1 Historical and Biological Background

Modern forensic DNA "history" begins with the first DNA case that was processed by the then 34-year-old Professor Sir Alec Jeffreys from Leicester University, U.K. This case involved the murders of two 15-year-old girls, Lynda Mann and Dawn Ashworth.[816] Lynda had been raped and murdered in 1983 in the ancient Leicestershire village of Narborough. Her 5 feet 2 inches, 112-pound body was found on a frosty lawn by The Black Pad footpath, undressed below the waist and bleeding from the nose. In 1986 the scene was Ten Pound Lane not far away in the same village, but the story was similar. Dawn, like Lynda, was found naked from the waist down. DNA analysis of semen present on vaginal swabs from the two girls suggested that the same person had murdered them. In 1987 a man who had confessed to the second murder was arrested. He was subsequently charged with both murders. DNA profiling exonerated him but left the rape murders unsolved. The police, however, were convinced that the true perpetrator was a local man. Consequently, blood samples were requested from all males of a certain age group from three villages within the area of the two murders. These samples were analyzed using a combination of classical blood-typing techniques and multilocus probe DNA profiling. Colin Pitchfork, a cake decorator with a history for flashing, had asked various men to give his sample for him and finally convinced Ian Kelly, a work colleague, to do so. Bar room talk by Kelly on the subterfuge eventually got to police ears and led the police to Mr. Pitchfork, who confessed.[706]

This pioneering case demonstrated the potential of DNA profiling[341,434,435,830] and firmly pointed toward its future as the most important forensic investigative tool to be developed in the 20th century.

DNA is the genetic code of most organisms. The DNA of humans[673] and many other organisms such as cats, dogs,[547,601] sheep, cattle, tigers,[817] horses,[232] plants (e.g., cannabis),[204,356,419] and bacteria[622,704] has been used in forensic work. Human primers can also be used to amplify the DNA from some other primates.[4] Much of the work discussed here will focus on the analysis of modern human DNA. However, many of the principles apply to all organisms and to ancient DNA.[146]

Most human DNA is present in the nucleus of the cell. It is packaged in the 46 chromosomes of most cells. This DNA is termed nuclear DNA. However, a small portion of the DNA complement of each cell is housed in

the mitochondria. This mitochondrial DNA is inherited by a different mechanism and is treated differently in the forensic context. A separate section in a subsequent chapter is devoted to this topic.

Most human cells are diploid, meaning that they have two copies of each chromosome. Exceptions include sex cells (sperm or ova), which are haploid (having a single copy of each chromosome), and liver cells, which are polyploid. Diploid cells contain 46 chromosomes in 23 pairs (the count was given as 48 for over 40 years). The human chromosomes are numbered from 1 to 22, starting with the largest numbered 1 and the second largest numbered 2. The 23rd pair comprises the X and Y chromosomes, which dictate the sex of the individual. This pair may be referred to as "nonautosomal" or "gonosomal."

Each chromosome possesses a centromere. This structure is involved in organizing the DNA during cell division. It is always off center and hence produces the short arm and long arm of the chromosome.

A normal female has two X chromosomes whereas a normal male has one X and one Y chromosome. One of the female X chromosomes is deactivated in each cell, becoming a structure known as a Barr body visible through the microscope. Which X chromosome is deactivated may differ for each cell.[749] In mammals, possession of the Y chromosome determines that the organism will be male. In fact, possession of even a small section of the short arm of the Y chromosome will result in a male. Other orders of life, such as reptiles, determine sex using other mechanisms. One chromosome of each of the 23 pairs has been inherited from the mother and one from the father.

From an examination of a single individual, it was historically not possible to tell which chromosome came from which parent, with the exception that a Y chromosome must have come from a male individual's father and hence the X of a male must have come from his mother. However, there are recent reports utilizing paternally imprinted allele typing (PIA) that do suggest that this may be possible for some loci. In mammals, some genes undergo parental imprinting and either the maternal or paternal allele may be preferentially expressed in the offspring. The reason for this is currently unknown. Imprinting appears to be associated with differential methylation upstream from the allele. This difference gives the potential to determine the parental origin of some alleles in the vicinity of any imprinted genes. Thirty-nine human genes have been identified as undergoing paternal imprinting[572] (Sykes gives 50).[749]

When most individuals are DNA profiled, they show either one or two alleles at each locus. If they show one, we assume that they are homozygotic, meaning they have received two copies of the same allele, one from each parent. If an individual shows two alleles, he or she is usually assumed to be heterozygotic. In such cases, the individual has inherited different alleles from each parent. An exception is caused by null or silent alleles. Heterozygotic individuals bearing one silent allele may easily be mistaken for homozygotes. Silent alleles most

probably occur when an allele is actually present but the system is unable to visualize it. Alternative methods may in fact be able to visualize the allele. Hence the term "silent" allele is preferable to the use of the term "null."

There are a few genetic exceptions that may lead to people having more than two alleles. These include trisomy (three chromosomes), translocation of a gene (a copy of the gene has been inserted somewhere else on the genome), and somatic mutation (the individual has different genotypes in different cells).

It is thought that all humans, except identical twins, differ in their nuclear DNA. Even identical twins may differ in minor ways. There is no formal proof of this concept of underlying uniqueness, and it has little influence on forensic work as all technologies examine only a very few points or loci on the entire human genome. The areas of the human genome used for DNA STR profiling are largely intronic. This means that they are noncoding DNA segments between areas of DNA that code for proteins. They were initially presumed to be functionless; however, evidence is accruing that noncoding DNA may, indeed, have a function.[446,538,539] A function for some noncoding DNA regions may include regulating development in eukaryotes. Interestingly, large areas of noncoding DNA, many of which are not implicated in regulation, are strongly conserved between species. This may be strong evidence that they too are, indeed, functional.

Introns are peculiar to eukaryotes and are thought to have developed late in eukaryotic evolution. They have a propensity to contain polymorphic regions, which means they have many differing forms. This is thought to be because there is little or no selective pressure on some of these loci and hence different forms may persist in populations, side by side.

In most of the ensuing chapters it is assumed that the genotype of people does not change throughout their lives, and is the same in all their diploid cells. In general, the genotype of an individual is set at the moment of gamete fusion. Minor changes may occur during a lifetime as a result of somatic mutation, and an adult individual is expected to show some level of mosaicity. It is possible that some genetic changes may be encouraged by practices during a lifetime. For example, allelic alteration has been reported in the cells of oral or colorectal cancer patients and betel quid-chewers.[439,520,602] This may affect genotyping when a reference DNA sample is taken from a different body tissue to the scene sample.

1.1.1 DNA Profiling Technology

DNA profiling has gone through three major stages of technological advancement. Loosely speaking, these were the multilocus, single-locus, and STR stages.[a] Protocols for extracting DNA, and constructing single-locus and

[a] This list unfairly omits other PCR-based approaches such as the sequence polymorphisms targeted at the HLA-DQα and Polymarker loci and the VNTR locus D1S80. Each of these techniques has had a large impact on forensic science.

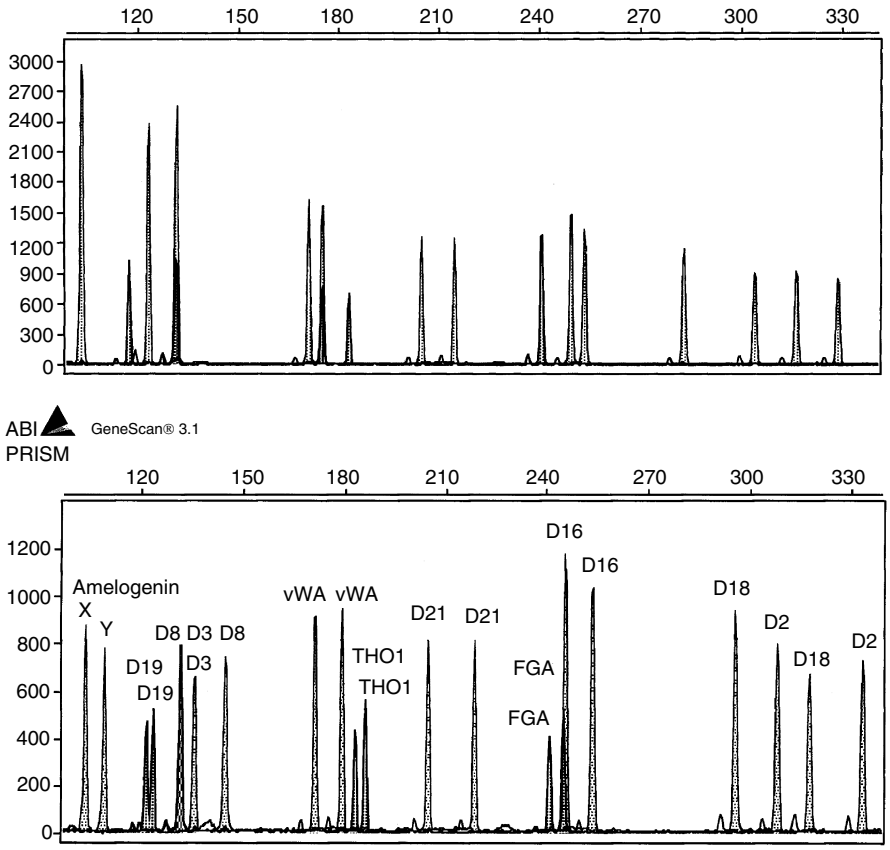


Figure 1.1 STR profiles of two different individuals. The profiles would normally have three colors — blue, green, and yellow — that help differentiate the loci.

STR profiles (see Figure 1.1) are described by Budowle et al.[130] These stages will be discussed here briefly.

1.1.1.1 Multilocus (Minisatellite) Testing

The first method developed for the forensic examination of samples was termed multilocus testing. Alec Jeffreys pioneered this approach — he discovered tandemly repeated DNA sequences or "minisatellites" that were variable between different individuals. Minisatellites were visualized by digesting the DNA with restriction enzymes to cut it into fragments of differing lengths that ranged between 1 and 20 kb in size. These fragments included relatively long minisatellites and their flanking regions of DNA sequence. The fragments were electrophoresed on a gel that separated them by size and then visualized using multilocus probes that hybridized many minisatellite loci at once. This made a pattern that looked a bit like a bar code. Jeffreys and

co-workers claimed that the probes produced patterns that were specific to an individual and coined the term DNA fingerprints.[435] The term "DNA fingerprinting" has been dropped in favor of the term "DNA profiling" because the analogy with fingerprints was not considered to be helpful. The issue of whether any particular DNA profile is unique to one individual is currently receiving considerable attention. However, at this time most forensic scientists prefer to stop short of such a claim.

There are various acronyms used to describe minisatellites, including restriction fragment length polymorphism (RFLP), a generic term that refers to any kind of DNA polymorphism that is based on length differences between restriction sites.

Minisatellite analysis could take many weeks to complete. Because of the relative slowness of methods employed during the early phase of DNA analysis, its use was restricted to violent or other serious crimes. Interpretation of these patterns was difficult due, in part, to the fact that it was not known how many loci were being visualized, which pairs were allelic, or which pairs were potentially linked. Also, large areas of the pattern could be a complete "blackout" and band intensity was highly variable and difficult to quantify.[342] A spot of blood larger than a quarter was required for effective analysis. This meant that the profile from the crime sample was often partial. The only publication to attempt to handle the partial nature of some crime sample profiles was that of Evett et al. [268] With hindsight, the assumption of independence for band presence or absence seems dubious. Today the use of multilocus probes in human forensic work is largely historic and is not discussed further in this book. The interpretation issues, however, were never fully settled. For a review of these and other issues surrounding RFLP analysis, see Donnelly.[235]

1.1.1.2 Single-Locus Probes

The next step in the development of forensic DNA work utilized the same RFLP technology; however, the probes used to visualize the product were altered to target only one locus at a time. These systems were referred to as single-locus probes (in the U.K. and New Zealand) and as variable number of tandem repeats (VNTR) systems in the U.S. As expected, most individuals showed one or two alleles at a locus. The use of the polymerase chain reaction (PCR)[568] for such loci was reported,[436] but not extensively implemented into casework. In the early 1990s, statistical interpretation of VNTR profiles was exclusively by use of the product rule.[276] The alleles were characterized by a measurement of their molecular weight. Each allele was an integer multiple of the repeat sequence plus the flanking DNA. However, the repeat length was small relative to the total fragment length and hence alleles separated by only one or a few repeat units could not be differentiated reliably using the agarose gel-based technology of the time. Although the underlying distribution was

discrete, the measurement of molecular weight was essentially continuous. Most implementations of the product rule treated this measurement as discrete, using floating bins (e.g., in the U.K., New Zealand, and parts of the U.S.[340]) or fixed bins (e.g., in most of the U.S.),[135,340,665] although elegant methods were suggested that avoided this step.[110,270,275,278,432,665,701] Argument at this time centered around the assumption of independence inherent in the use of the product rule[584]and some unfortunate details associated with the fixed-bin approach.[324]

While VNTR loci are still in use in some laboratories, they have largely been replaced by STR loci.

1.1.1.3 STR Analysis

In the mid-1990s, the technology changed to encompass the use of PCR of STR loci.[241] The PCR reaction has been likened to a molecular photocopier. It enables the exponential amplification of very small amounts of DNA. With the methods previously discussed, typically 500 ng was required for a successful test. With PCR, 1 ng or less could be analyzed. The STR loci selected had much smaller alleles, typically between 100 and 400 bp. Resolution of small fragments by polyacrylamide gel electrophoresis (PAGE) was much improved compared with previous methods that analyzed fragments of several kb. Consequently, the distance between STR alleles differing by one repeat was sufficient to allow unambiguous assignment of genotypes. This was perceived as a considerable advantage. Smaller alleles were also more suitable for the PCR reaction as it is more efficient with low molecular weight DNA fragments.

PCR involves a number of replication "cycles." Each cycle has the potential to double the amount of DNA, although actual amplification is slightly less than a doubling. In many cases, standard casework using STRs is undertaken at 28 cycles. At perfect amplification, this theoretically should amplify the starting template by a factor of 268,435,456. However, perfect amplification is not achieved.

Generally, PCR-based STR profiling is sensitive to approximately 250 pg; however, a template concentration in the order of 0.5–1.0 ng is commonly analyzed. To increase sensitivity to samples of DNA below this threshold, up to 34 cycles may be employed. This gives a theoretical amplification factor of 17,179,869,184 and can allow the analysis of samples that have only trace amounts of DNA present such as touched surfaces.

In fact, it is possible to amplify the DNA of a single cell.[294] The analysis of trace DNA evidence is described by the term "low copy number" (LCN) in the U.K. The suggested guidelines for reporting LCN evidence are different to "conventional" DNA profiling because of the increased uncertainty in the origin of the DNA and the increase in artifactual issues. This concept is dealt with separately in a subsequent chapter.

The introduction of PCR-based STR analysis was the major innovation that expanded the utility of DNA profiling. In summary:

- The development of PCR improved the sensitivity of the analysis.
- The time taken per analysis was reduced to less than 24 hours.
- The cost effectiveness of the method was greatly improved due to a reduction in the labor required.
- The shorter STR loci allow the analysis of degraded DNA samples, which are frequently encountered by forensic scientists. This was because these short segments of DNA stood a higher chance of being intact after degradation.
- STR loci can be multiplexed together using several different STR primer pairs to amplify several loci in one reaction. Multiplexing was further facilitated by the development of dye-labeled primers that could be analyzed on automated DNA sequencers.
- The collection of data was automated, and the analysis of data was partially automated.

1.1.1.3.1 Selection of STR loci for forensic multiplexing. STR loci consist of repeated segments of two to eight bases. These are termed dimeric, trimeric, and so on. Dimeric loci are not used for forensic applications because excessive slippage during amplification (termed stuttering) results in a large number of spurious bands that are difficult to interpret. Trimeric, tetrameric, and pentameric loci are less prone to this problem.

Several factors are considered when choosing candidate STR loci:

- A high level of variability within a locus is desired so that the locus has a low match probability.
- The length of alleles should be in the range 90–500 bp. Typically, the higher the molecular weight of the alleles, the lower the precision of their measurement. Smaller alleles are less affected by degradation and are therefore less likely to drop out.
- Loci may be selected based on chromosomal location to ensure that closely linked loci are not chosen. See Table 1.1 for the chromosomal location of some common STR loci. As a quick guide to the nomenclature of the locus locations, those that begin with, say, D5 are on chromosome 5.
- Robustness and reproducibility of results are essential.
- In order to ease interpretation, it is desirable that loci do not stutter excessively.

Early multiplexes were based on a few simple STR loci. The four-locus "quadruplex" was probably the first to be widely used for court reporting

Table 1.1 Loci Used in Certain Multiplexes

					PE Applied Biosystems					Promega			
Locus	Chromosomal Location	CODIS	ENSFI	SGM	SGM+	AmpF/STR Identifiler	AmpF/STR Profiler	AmpF/STR Profiler plus	AmpF/STR Cofiler	PowerPlex 1.1	PowerPlex 2.1	PowerPlex 16	PowerPlex ES
D16S539	16	*			*	*			*	*		*	
D7S820	7	*				*	*	*	*	*		*	
D13S317	13	*				*	*	*		*		*	
D5S818	5q21-31	*				*	*	*		*		*	
CSF1PO	5q33.3-34	*				*	*		*	*		*	
TPOX	2p13	*				*	*		*	*	*	*	
THO1	11p15.5	*	*	*	*	*	*		*	*	*	*	*
vWA	12p	*	*	*	*	*	*	*		*	*	*	*
FGA	4q	*	*	*	*	*	*	*			*	*	*
D21S11	21	*	*	*	*	*		*			*	*	*
D8S1179	8	*	*	*	*	*		*			*	*	*
D18S51	18	*	*	*	*	*		*			*	*	*
D3S1358	3	*	*		*	*	*	*	*		*	*	*
Amel	X,Y		*	*	*	*		*		*		*	*
Penta D	21											*	
Penta E	15										*	*	
D2S1338	2				*	*							
D19S433	19				*	*							
ACTBP2SE33	6												*

Amended from Gill[339] with kind permission from BioTechniques/Eaton Publishing.

purposes.[460] The match probability was high by modern standards, in the order of 10^{-4}; hence, initially the evidence was often supported by SLP evidence. In 1996, a six-locus STR system combined with the amelogenin sex test[741] was introduced.[722,723] This system, known as the "second-generation multiplex" (SGM), superseded SLP analysis in the U.K. and New Zealand. The SGM had more loci and included the complex STR loci HUMD21S11 and HUMFIBRA/FGA,[551] which are highly polymorphic. The expected match probability was decreased to approximately 1 in $\sim 50 \times 10^6$.

The introduction of SGM in 1995 narrowly preceded the launches of the U.K. (1995) and New Zealand (1996) national DNA databases.[390,851] More than two million samples are now stored in the U.K. database and a similar fraction of the population is in the New Zealand database. As databases

become much larger, it is necessary to manage and minimize the possibility of matches to innocent people (adventitious matches). This may be achieved by increasing the discriminating power of the STR systems in use. Such additional discrimination may be utilized either in the database itself or in post-hit confirmation.

To reduce the potential of adventitious matches, a new system known as the AMPF*l*STR®SGM Plus™ (SGM^+) was introduced in the U.K. in 1999, which comprised ten STR loci and amelogenin.[199] This replaced the previous SGM system. The estimated probability of a match between two unrelated people was approximately 10^{-10}–10^{-13}. For a full DNA profile, it is U.K. practice to report a default match probability of less than 10^{-9}. This figure is believed to be conservative.[311] To ensure continuity of the DNA database so that the new system can be used to match samples that had been collated in previous years, all six loci of the older SGM system were retained in the new SGM^+ system. Multiplexes with more loci and more discriminating power are becoming available (see Table 1.1).

Harmonization of STR loci used in forensic work has been achieved by collaboration at the international level. The European DNA profiling group (EDNAP) carried out a series of successful studies to identify and to recommend STR loci for the forensic community to use. This work began with an evaluation of the simple STR loci HUMTH01 and HUMVWFA.[461] Subsequently, the group evaluated the HUMD21S11 and HUMFIBRA/FGA loci.[352]

To date, a number of European countries have legislated to implement national DNA databases that are based upon STR loci. In Europe, there has been a drive to standardize loci across countries, in order to meet the challenge of cross-border crime. In particular, a European Community (EC)-funded initiative led by the European Network of Forensic Science Institutes (ENFSI) was responsible for coordinating collaborative exercises to validate commercially available multiplexes for general use within the EC.[347]

Based on the initial EDNAP exercises and on recommendations by ENSFI and the Interpol working party, four systems were defined as the European standard set of loci: HUMTH01, HUMVWFA, HUMD21S11, and HUMFIBRA/FGA. Recently, three further loci were added to this set: HUMD3S1358, HUMD8S1179, and HUMD18S51. A similar process occurred in Canada[320,804] and in the U.S.,[418] where standardization was based on 13 combined DNA index system (CODIS) loci. The 13 CODIS designated loci and the eight (the seven mentioned above plus amelogenin) ENSFI loci are marked in Table 1.1. These loci are included in the commercial multiplex systems manufactured by PE Applied Biosystems, Promega Corporation, and others.

There are currently seven loci that are in common use across both North America and Europe. The chromosomal positions of these loci are also shown. The short and long arms of a chromosome are designated as p and q,

respectively. Note, for instance, that among the CODIS set there are two loci on chromosome 5.

Bacher et al.[30] report that these loci are separated by 25 centiMorgans (cM). Penta D and HUMD21S11 are both on chromosome 21 and reported to be separated by 50 cM.[b,c]

1.1.1.3.2 STR locus nomenclature. Several different classes of STR loci have been defined. Urquhart et al.[792] classified different loci according to the complexity of their sequences. One of the most ubiquitous STR loci used is HUMTH01.[621,627] This consists of a simple repeating sequence $(TCAT)_{5–11}$ with a common nonconsensus allele $(TCAT)_4CAT(TCAT)_5$. Compound STR loci, such as HUMVWFA31,[459] consist of repeating sequences $(ATCT)_2(GTCT)_{3–4}(ATCT)_{9–13}$, whereas complex repeats such as HUMD21S11[707] are less uniform. Detailed information may, again, be obtained from STRBase.[138,680]

This nomenclature system has found widespread application. However, as technologies advance, deficiencies in the system are being found and we may see a revision in the future.[815]

These sequences are based on a tetrameric repeating sequence interspersed with invariant di- and trinucleotides. Complex hypervariable $(AAAG)_n$ repeats such as human beta-actin related pseudogene (ACTBP2)[793,819] are much more difficult to accommodate to a nomenclature based upon the number of tetrameric repeat sequences. This is because variant mono-, di-, tri-, and tetramers are scattered throughout the locus. These latter STRs have found limited use in a few European countries.

1.1.1.3.3 STR allele designation. The greatest advantage of fluorescence automated sequencer technology is the ability to detect several different

[b]Dr. Bentley Atchison directed us to a site http://www.gai.nci.nih.gov/CHLC that gives recombination information.

[c]Map distance, recombination fraction, and Kosambi distance by CM Triggs. A genetic map distance of 1 Morgan is that distance such that one crossover is expected to occur within it per gamete per generation. Typically, data are expressed in centiMorgans (cM) and in humans 1 cM is assumed to equal approximately 1000 kb.

The simplest relationship between distance and recombination fraction is due to Haldane.[383] Consider two loci, A and B, and denote the genetic distance between them as x, and their recombination fraction as R.

$$R = \frac{1}{2} \times (1 - e^{-2x}) \quad \text{(Haldane)}$$

Expressing x as a power series in R, we find that $x \approx R + 2R^2 + 4R^3 + 8R^4 + \cdots$.

Kosambi took into account the fact that the strands of the DNA molecule are to some extent rigid and hence that the occurrence of a crossover will inhibit the possibility of a second nearby recombination event. He gives the relationship between the recombination fraction, R, and the map distance by

$$R = \frac{1}{2} \times \frac{1 - e^{-4x}}{1 + e^{-4x}} \quad \text{(Kosambi)}$$

Expressing x as a power series in R, we find that $x \approx R + 4R^3 + \cdots$.

dye-labeled moieties. For example, current systems are able to detect five colors. The determination of DNA fragment sizes is dependent upon the use of two types of standard markers. In every sample that is electrophoresed, a series of dye-labeled DNA fragments of known size are included. This internal size standard may be composed of restricted bacteriophage labeled DNA (for instance, the Applied Biosystems GS 500 product) or, alternatively, artificial DNA concatamers (for instance, the Applied Biosystems HD 400 product).

The second kind of standard marker is the "allelic ladder" (Figure 1.2). This is comprised of all the common alleles for each locus and is compared with each lane on an electrophoretic run.[349] Allelic ladders should span the entire range of the common alleles of a locus. However, it is not necessary that every allele be represented in the ladder. Many rare alleles have been discovered and some of these are outside the range of the ladder. If possible, there should be no gap larger than four bases between the rungs of the ladder for tetrameric and dimeric STR loci. If the STR repeat is greater than four bases, then the maximum gap should be the size of the repeat.

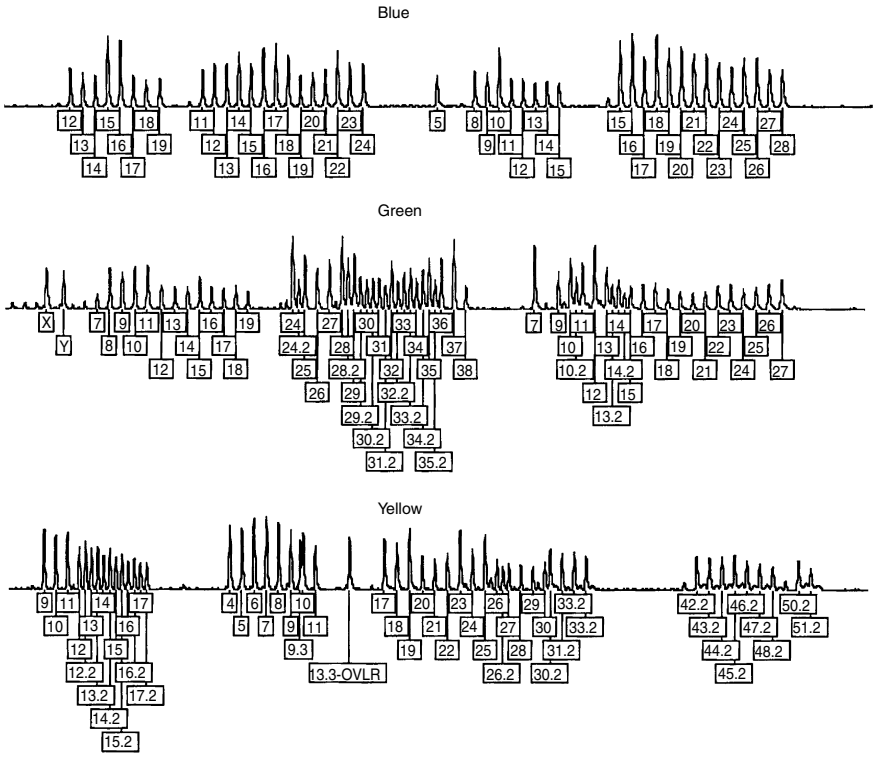


Figure 1.2 Allelic ladders from the AMPF*l*STR®SGM Plus™ system (PE Applied Biosystems, Foster City, CA).

The allelic ranges of some loci may overlap. These loci are labeled with different dyes, therefore allowing each locus to be identified. Loci that are labeled with the same dye have to be separated sufficiently to minimize the possibility of overlap of the allele ranges.

Allele sizes are measured relative to the internal size standard, often by using the Elder and Southern local method.[248,249] The size of the unknown alleles in the questioned sample is then compared with the size of the known alleles of the allelic ladder. The units are typically base pairs (bp) or bases.

Provided that a questioned allele is within ±0.5 bp of a corresponding ladder peak, allelic designation may be undertaken. In all electrophoretic systems, it is usual for a small amount of aberrant migration to occur such that the migration rate may be either slower or faster than expected. This is termed band shift. Band shift tends to be in the same direction for two alleles in the same lane. This can be measured to ensure consistency,[349] acting as an additional quality control check and also as a means to designate off-ladder or "rare" alleles.[343]

1.1.1.3.4 STR allelic nomenclature. The International Society of Forensic Genetics (ISFG) DNA Commission[48,353,594] has recommended an STR allelic nomenclature based upon the number of repeat sequences present in an allele. If a partial repeat sequence is present, then the size of the partial repeat is given in bases after a decimal point[d]; for example, the common allele HUMTH01 9.3 consists of nine repeats and a partial repeat of three bases. This method is suitable for typing simple STR loci.

Complex hypervariable repeats such as ACTBP2 (currently used in some European criminal DNA databases, e.g., Germany) do possess a simple repeating structure. The designation of complex STR repeats such as ACTBP2, D11S554, and APOAI1 follows from the size of specific alleles. The size is dependent upon the primers utilized, and hence different primers will produce a differently named allele. The allelic size may also be dependent upon the internal structure of the allele. Hence designations are prefixed with the term "type-."

The designation scheme to be used for a given locus is dependent upon the characteristics of the locus itself. If possible, the designation should follow the recommendations of the ISFG DNA Commission unless this approach is precluded by allelic structure at this locus.

Linking the allelic ladder and the nomenclature of STR loci provides the key to standardization. In principle, the platform used (capillary electrophoresis or PAGE) is not particularly important. Direct comparisons can be made between different instruments, provided that allelic sizing is consistent. In addition,

[d] Termed a decimal point by biochemists, but strictly it is just a dot. There is no hint of the decimal system in what comes after the dot.

comparisons can also be made between different multiplexes derived from different manufacturers using different primer sets. The allelic ladders act as control reference standard that enable laboratories using different hardware and multiplexes to compare results.

1.2 Understanding STR Profiles

In this section we begin the process of interpreting electropherograms. It is necessary to understand the effects of some genetic anomalies and the outputs of the PCR and electrophoresis systems to understand both simple unmixed profiles and more crucially mixtures. Some anomalies and outputs are introduced briefly here.

1.2.1 Genetic Anomalies

1.2.1.1 Trisomy and Gene Duplication

The first trisomy "discovered" was that associated with Down's syndrome at chromosome 21 reported by Lejeune in 1959.[749] Subsequently trisomies were discovered at chromosomes 13 and 18, but always associated with severe disorders. Trisomies appear more common in spontaneously aborted fetuses. Chromosomal duplication of ChrX appears to be more common and to have fewer effects, possibly due to the deactivation of all X chromosomes bar one.

Both chromosome and gene duplication affect all cells in an individual. In practice, it is impossible to tell the difference between these two phenomena without resorting to genetic analysis. If a deletion or insertion of a repeat unit accompanies duplication, then three bands of similar size are generated (see Figure 1.3).

If a gene is duplicated without additional mutation, then two bands are visible in a 2:1 ratio. In the example in Figure 1.4, an XYY individual has two copies of the Y chromosome. Note that the other loci are balanced and this argues against the possibility that this sample is a mixture. In the multiplex described by Sparkes et al.,[722,723] trisomy or gene duplication was observed rarely at each locus (see Table 1.2). Johnson et al.[445] report three gene duplication events in a sample of 525 males. Crouse et al.[207] report 18 three-banded patterns at HUMTPOX and one at HUMCSF1PO in over 10,000 samples. STRBase[138,680] gives up-to-date counts of three-banded patterns at some loci. Valuable reports continue to appear.[15,857]

1.2.1.2 Somatic Mutation

Somatic mutation occurs during embryological development or later in life. A mutation occurs in one line of cells and hence cells with two different genotypes coexist, leading to a three-banded profile (Figure 1.5) when samples of these cells are typed.

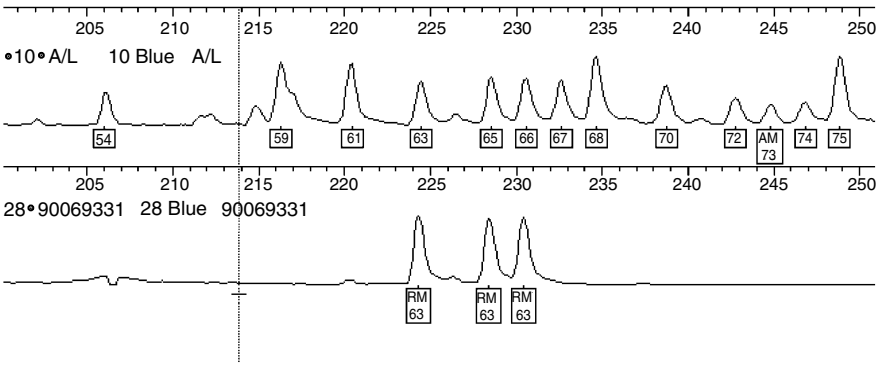


Figure 1.3 An example of a HUMD21S11 trisomy or translocation appears in the lower pane. Note that the bands are equivalent in size. The allelic ladder is in the upper pane. The nomenclature used to designate this sample follows the method of Urquhart et al.[792] Reproduced with the kind permission of BioTechniques/Eaton Publishing from Gill.[339]

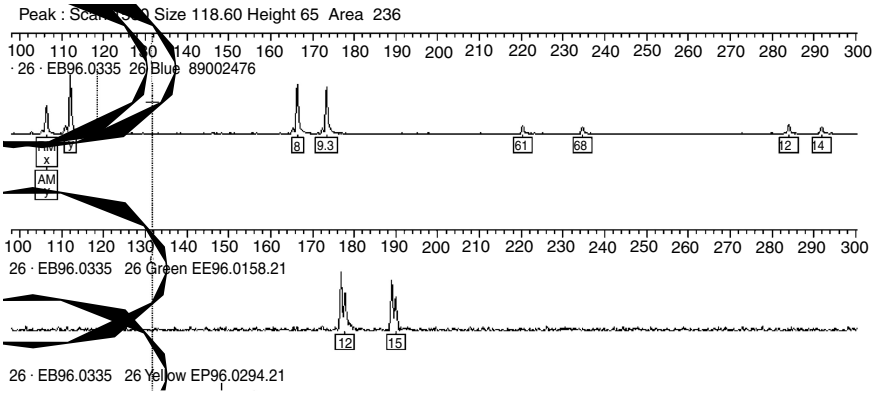


Figure 1.4 An XYY individual (upper pane left) showing a Y peak twice the size of the X peak. The remaining loci of the SGM multiplex are balanced. Reproduced with the kind permission of BioTechniques/Eaton Publishing from Gill.[339]

The peak areas will be dependent upon the relative proportion of the two cell types in the sample and need not be equivalent. This is arguably the most difficult condition to elucidate since it is possible that not all tissues will demonstrate somatic mutation. The incidence of somatic mutation varies

Table 1.2 Occurrence of Trisomy or Gene Duplication at Some STR Loci in ~600,000 Profiles

Locus	Count
Amelogenin	1191
HUMD21S11	9
HUMD18S51	7
HUMD8S1179	24
HUMFGA	12
HUMVWA	8
HUMTH01	1

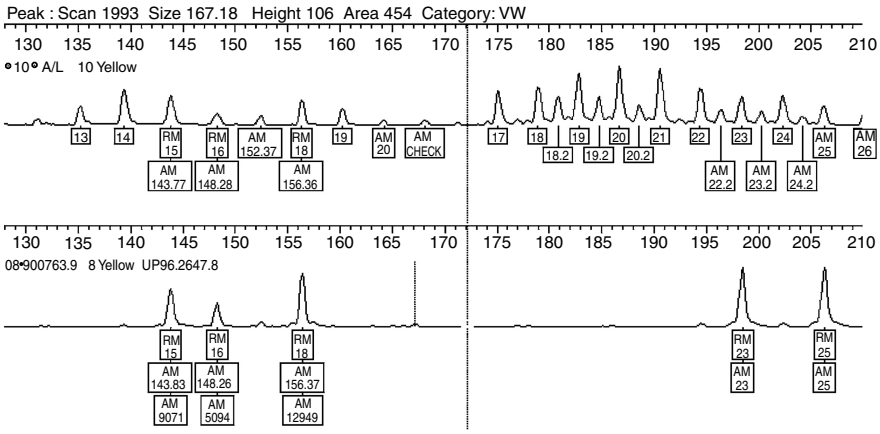


Figure 1.5 Somatic mutation at the HUMVWFA31 locus, lower left pane. Note that three peaks of different sizes are present. HUMFIBRA/FGA peaks are shown on the right side. The upper pane shows HUMVWFA31 and HUMFIBRA/FGA allelic ladders. Reproduced with the kind permission of BioTechniques/Eaton Publishing from Gill.[339]

between loci: out of 120,000 samples, no somatic mutations were observed at the HUMTH01 locus, whereas the incidence is approximately 1 in 5000 at the HUMD18S51 and HUMFIBRA/FGA loci. It is possible that some somatic mutations will not be distinguishable from stutters. Hence, these figures are probably underestimates since mutations are recorded only if they are unambiguous.

1.2.2 PCR Artifacts

1.2.2.1 Heterozygote Balance

There have been at least three terms proposed to define the phenomenon of heterozygote balance. These are heterozygote balance, heterozygote imbalance, and preferential amplification. Preferential amplification is probably an

inappropriate term for several reasons discussed later. Of the other two terms, "heterozygote balance" has historic precedence whereas "heterozygote imbalance" appears to more properly fit the phenomenon that we will describe graphically and mathematically.

Heterozygote balance (*Hb*) describes the area or height difference between the two peaks of a heterozygote (see Figure 1.6). An understanding of this phenomenon is vital to determining whether the profile may be a mixture or not.

There are several factors that may contribute to heterozygote balance. It is likely that the starting templates are sampled unequally. The pre-amplification sample is an aliquot from the post-extraction process and does not necessarily contain equal numbers of the two alleles for heterozygotes. This effect is likely to be more severe for samples with fewer templates *per se*.

There is also a natural variation in the PCR process. Accordingly, the two alleles of a heterozygote may be amplified unequally.

The effects of sampling and amplification variability combine to create the observed heterozygote balance. This heterozygote balance has been defined[854] previously in two differing ways: (i) as the ratio of the area of the smaller peak to the larger peak $Hb = \phi_{smaller}/\phi_{larger}$, and (ii) as the ratio of the area of the heavier

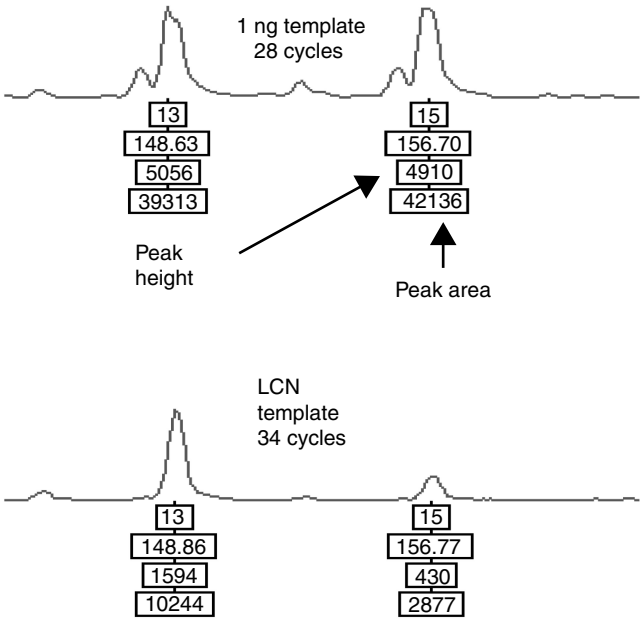


Figure 1.6 Profile morphology at the HUMD8S1179 locus in a sample amplified under normal (28 cycles/1 ng) and LCN (34 cycles/25 pg) conditions. Reprinted in altered form from Whitaker et al.[854] © 2001, with permission from Elsevier.

molecular weight peak to the lighter molecular weight $Hb = \phi_{HMW} / \phi_{LMW}$.[e] The latter definition contains more information and will be preferred here.[f] With this latter definition, a mean value of less than 1 for *Hb* suggests that the lighter molecular weight allele is being preferentially amplified.

A heterozygote is defined as extremely unbalanced if one of the alleles is greater in area than its counterpart by a specified amount. An approximate guideline of $0.6 \leq Hb \leq 1.67$ has been developed. Experimental observation suggests that the majority of data fall within these limits, but some outliers are observed. The intent of this guideline is to highlight to the reporting scientist that the sample may need further biochemical investigation. The distribution of *Hb* should be both asymmetric and conditioned on peak area and template quality. For instance, we expect more imbalance for profiles with small peaks. Imbalance may also be caused by rare genetic phenomena such as somatic mutation.

If we wish to really explore the phenomenon of heterozygote imbalance, we need to separate this effect from the effect of stuttering. This leads us to the concept of "total product." We consider the total PCR product from the two alleles. This product will include the stutter and allelic peaks at least. So we are, unfortunately, led to another definition:

$$Hb = \frac{(\phi_A + \phi_S)_{HMW}}{(\phi_A + \phi_S)_{LMW}}$$

where ϕ_A is the area of the allelic peak and ϕ_S is the area of the stutter peak.

Experimentally it has been observed that there is a small trend for the smaller molecular weight allele to amplify more efficiently; however, there is considerable scatter about this trend (see Figure 1.7 after Veth[800]; we follow Triggs and Patel[778]). As with most ratios, it is sensible to plot the logarithm of the ratio. It is reasonable to force the fitted line through (0,0) since we expect no preferential amplification for alleles that differ in no way.

This phenomenon is often referred to as "preferential amplification"; however, the term is inappropriate as it implies a strong preference for the low molecular weight allele to amplify more efficiently. In reality, when we consider total PCR product, a small trend is noted with considerable scatter about this trend. In addition the term implies that all imbalance is caused by differences in amplification and hence ignores sampling effects at the template level. We prefer the term heterozygote balance.

It is interesting to examine the variability about the fitted line (see Figure 1.8).

We could ask the question: Is the area normally distributed about the fitted line? This is investigated by the normal density quantile plot in

[e] It is very unfortunate that this also appears as its inverse in the literature, $Hb = \phi_{LMW} / \phi_{HMW}$, which complicates the situation further.
[f] You can calculate $Hb = \phi_{smaller} / \phi_{larger}$ from $Hb = \phi_{HMW} / \phi_{LMW}$, but not the reverse.

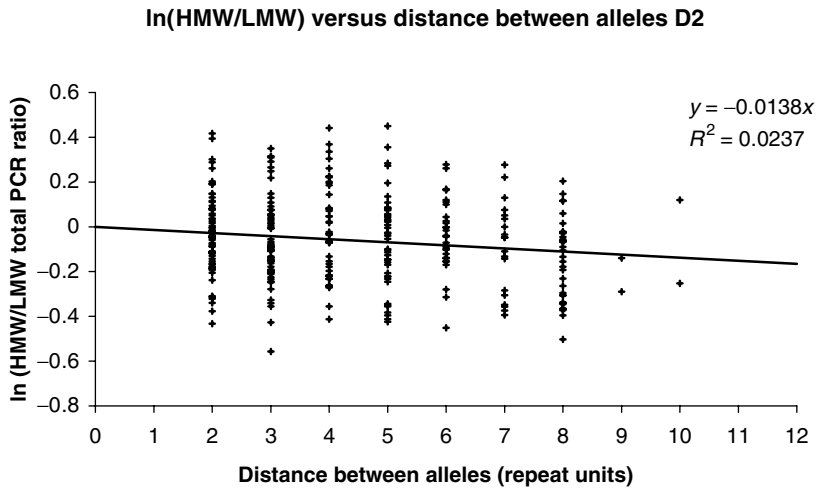


Figure 1.7 Total PCR product versus the distance between alleles for HUMD2S1338 in the SGM$^+$ system. The data have been developed from "simple heterozygotes," which we define as those separated by two or more repeat units or otherwise arranged that the stutter peak does not fall on an allelic peak. For such heterozygotes, we can calculate total PCR product more easily since the stutter peak is not masked by an allelic peak. Note the small downward trend of the line with considerable scatter above and below the line. This is a typical result. Data sourced with kind permission from Veth.[800]

Figure 1.8 Distribution about the fitted line. The prediction from the fitted line is set as 0 on the *x*-axis. The plot shows the distribution of ln $[(\phi_A+\phi_S)_{HMW} / (\phi_A+\phi_S)_{LMW}]$ about this fitted line. Data sourced with kind permission from Veth.[800]

Figure 1.9. If this line is perfectly straight, then those data fit exactly with a normal distribution.

To our subjective eye this is acceptably normal, and hence we model the natural log of total PCR product as normally distributed about the fitted line. The variance is easily calculated and both the variance and the slope are tabulated in Table 1.3 for the ten SGM$^+$ loci investigated by Veth.[800] The variance about the fitted line for all ten loci appears, visually, to be adequately modeled by a normal distribution (data not shown).

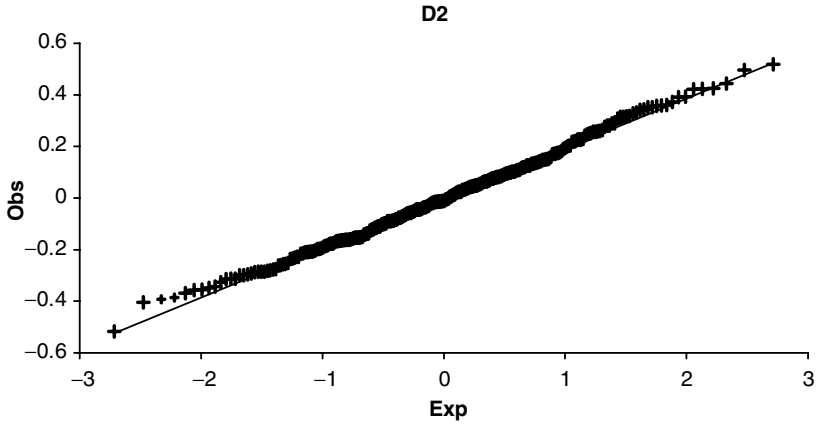


Figure 1.9 Normal density quantile plot for HUMD2S1338. Data sourced with kind permission from Veth.[800]

Table 1.3 Modeling Parameters for Allelic Area

Locus	Molecular Weight	Heterozygote Balance	
		Slope of Fitted Line (Recall the Intercept is Zero)	Standard Deviation About the Fitted Line
D19	106–140 bp	–0.014	0.14
D3	114–142 bp	–0.016	0.15
D8	128–172 bp	–0.022	0.17
vWA	157–209 bp	–0.017	0.18
THO1	165–204 bp	–0.018	0.15
D21	187–243 bp	–0.017	0.18
FGA	215–353 bp	–0.017	0.19
D16	234–274 bp	–0.018	0.19
D18	265–345 bp	–0.010	0.18
D2	289–341 bp	–0.014	0.19

Data sourced with kind permission from Veth.[800]

1.2.2.2 Allelic Dropout

Allele dropout is defined as the condition where an allele cannot be visualized. It is not yet clear whether it is an extreme form of variable amplification. It is most often observed when one allele of a heterozygote cannot be visualized. The phenomenon needs to be understood as it can lead to the false impression that the genotype is a homozygote. When allelic dropout is possible, for instance, when peak areas or heights are very low, it is wise to be cautious and utilize a genotype designation that recognizes this. This is typically, but not always, conservative. For instance, an apparent *16,16* homozygote with low peak height could be written as *16,F* as in the U.K. This designation stands for *16* "failed" and means that the genotype is assigned as *16* and any other allele. In many laboratories the height guideline where this designation should be applied is 150rfu[g] because experimental observations suggest that allele dropout does not occur above this level. We believe that this differs between laboratories and endorse recommendations as to internal validation of this and other parameters.

Findlay et al.[295,297] have also studied allelic dropout and suggest that it is a separate phenomenon to heterozygote balance, not simply an extreme form of it. They note that heterozygote balance shows a distribution terminating at nonextreme values for the fraction of product attributable to one allele. When investigating this statement, we need to carefully consider the methods available for assessing peak area or height and the various thresholds used. It is certainly possible that small heights or areas are simply "not measured" but are still present. Triggs and Patel (as above) went to some considerable effort to investigate this.[778]

1.2.2.3 Stuttering

Stuttering refers to the production of peaks at positions other than the parental allelic position. The term is reserved for loss of complete repeat units. For tetrameric loci, the loss of one unit is termed the $N-4$ stutter and the loss of two units is the $N-8$ stutter. Stuttering is presumed to be due to miscopying or slippage during the PCR process.

It is useful to define stutter ratio as $S_R = \phi_S / \phi_A$ and stutter proportion as $S_x = \phi_S / (\phi_A + \phi_S)$.

Larger alleles appear to stutter more. Specifically, strong supporting evidence has been given by Klintschar and Wiegand[467] for the hypothesis that the larger the number of homogeneous repeats, the larger the stutter peak.

When investigated, many loci do not give a straight line fit of log stutter ratio or logit stutter proportion to allele designation. Loci such as the

[g] Relative fluorescence units.

HUMD2S1338 locus (see Figure 1.10) have a compound repeat structure $(TGCC)_n (TTCC)_n$. Veth has observed similar deviations from a straight line at other SGM$^+$ loci. If the Klintschar and Wiegand hypothesis is correct, this may be the explanation for the kink in the graph for this locus and other loci showing the same effect. However, the pattern is not immediately obvious and warrants further investigation. Accordingly, it appears premature to offer a simple model for stutter based solely on a linear regression to allele number.

Whitaker et al.[854] investigated the ratio of the $N - 4$ stutter peak area to parental peak area. For 28-cycle PCR, the ratio was in the range between 0.05 and 0.10 with outliers ranging up to 0.15. This was true regardless of the peak areas of the parent peaks. No example of a stutter proportion greater than 0.15 was observed. The effect of stuttering, and in fact all PCR artifacts, was found to be greater in LCN work. A separate chapter is devoted to the interpretation of LCN profiles.

Frégeau et al.[322] report stutter ratios for casework samples using reduced reaction volumes (25 µl). Similar reaction volumes are in reasonably widespread use. They report that stutter peaks in the blue and green STR systems in Profiler Plus were all less than 0.16 of the parental peak. For the yellow STR system, they were less than 0.12 of the parental peak. Johnson et al. [445] give values for a Y-STR multiplex.

It is worthwhile at this point to return to the subject of preferential amplification. Combining the findings regarding heterozygote balance and stutter,

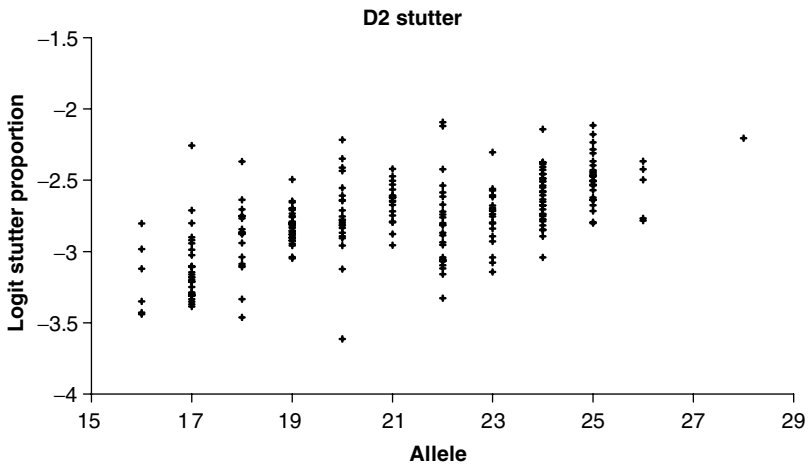


Figure 1.10 Stutter ratio or proportion for HUMD2S1338. The *x*-axis is the allele designation. The *y*-axis shows either $\ln(S_R)$ or logit of stutter proportion, which are equivalent. Logit refers to the function $\ln [S_x / (1 - S_x)]$ and is a function often used by statisticians when plotting proportions. Data sourced with kind permission from Veth.[800]

we see that larger alleles do amplify to slightly less total product but that more of the product is stutter. Hence, if we need to use the word "preferential" at all, it would be better to term this "preferential stutter."

1.2.2.4 Nonspecific Artifacts

Nonspecific artifacts are generated as a result of the priming of DNA fragments during the PCR process, possibly from degraded human or bacterial DNA. The band shift test described by Gill et al.[343] is particularly useful to identify peaks as nonspecific since they usually migrate atypically in the gel. This may be either because they have different sequences to STR alleles or because they are a partial renaturation of PCR products.

1.2.2.5 Pull-Up

One problem commonly observed in STR profiles is "pull-up." This typically occurs when a minor peak in one color corresponds to a major allelic peak in another color. Typically, a blue peak may "pull up" a green peak directly below it. This is only problematic if the minor peak coincides with the position of a potential allele. If such a possibility exists, options to consider include amplification of the locus under consideration by itself (single-plexing), re-PCR of the sample, or reapplication of the matrix or spectral calibration.

1.2.2.6 Poor Operator Technique

Leakage of a sample from one lane to another, commonly referred to as "lane-to-lane leakage," is a problem that may be encountered when loading samples into vertical acrylamide gels such as are used in the 377 Gene Sequencer. This can be detected by loading samples in a staggered fashion, either odd or even lanes first, with a short period of electrophoresis in between. Lane-to-lane leakage can then be detected by viewing the sample profiles by scan number in the Genescan analysis software (Applied Biosystems, Foster City, CA).

1.2.2.7 Suppression of Amplification Efficiency, Silent or Null Alleles

Peak-area asymmetry outside the normal range or the creation of a silent or null allele may occur because of a primer-binding site mutation. This has the effect of altering annealing and melting temperatures, which changes the amplification efficiency and decreases the resulting signal. If a substitution mutation occurs at the 3' end of the primer, a mismatch will result and amplification will fail completely, resulting in a silent allele. The closer the substitution to the 5' end of the primer, the lesser the effect on amplification efficiency.[14,118,378,809] Butler and Reeder[138] and Whittle et al.[857] report some silent allele frequencies shown in Table 1.4.

Table 1.4 Silent Allele Probabilities and Multibanded Patterns from Butler and Reeder[138] (Top When Two) or Whittle et al.[857]

	Silent Alleles	Multibanded Patterns %
CSF1PO	2/42,020	
	0/21,800	
D5S818	3/74,922	
	0/21,604	
D7S820	1/42,020	1/406
	6/32,120	
D13S317	52/62,344	
	0/21,394	
D16S539	3/52,959	0/1165
	2/21,498	
D21S11	1/203	
	2/20,600	
FIBRA(FGA)	2/1104	
	0/34,278	
THO1	2/7983	0/2646
	0/19,308	
TPOX	11/43,704	13/42,020
	0/21,884	
vWA	7/42,222	1/6581
	12/36,466	
F13B	0/21,964	
D3S1358	4/22,084	
F13A01	0/23,034	
D8S1179	6/33,110	
D10S1237	4/13,600	
FESFPS	10/30,906	
Penta E	0/8060	
D18S51	2/36,546	
D19S253	36/35,602	

Chang et al.[176] report high occurrences of silent alleles at the amelogenin locus in some populations that interfered with the efficiency of the test. They found a rate of 3.6% in an Indian population and 0.9% in a Malay population. Clayton et al.[185] identified a set of silent alleles at the HUMD18S51 locus associated with individuals of middle-eastern descent. They confirmed the nature of these alleles using alternative primers. At this locus, they found that the presumed primer-binding site mutation was associated in 12 of 15 instances with an *18* allele. The remaining instances were one each of a *17, 19,* and *20* allele. This supports the suggestion that the ancestral primer-binding site mutation was associated with an *18* allele and that the *17, 19,* and *20*

alleles have subsequently arisen by mutation from this *18* allele. Other valuable reports continue to appear.[15]

1.2.2.8 Promotion of Amplification Efficiency

DNA sequence differences in flanking regions near the PCR primer-binding site may improve amplification efficiency. In the SGM system,[722,723] at the HUMVWFA31 locus, a sequence polymorphism was found associated with most HUMVWFA31 *14* alleles, and to a much lesser extent with the HUMVWFA31 *15* allele. The polymorphism consisted of a substitution that is three bases from the 5' end of the primer-binding site (in the amplification region). This appears to enhance amplification and may result in a peak area ratio greater than 2:1. This phenomenon has not been observed in the SGM$^+$ system, presumably because the primers used are different.

1.3 Summary

The biological basis of contemporary forensic DNA profiling is linked to the processes of cell and human reproduction. From the many variations that subsequently exist on the human genome, STR's have emerged as the most suitable marker for current forensic identification. Standardising on this polymorphism has led to further harmonization with regard to the specific loci that are targeted and analysed in the international forensic community. As with any complex molecular technique however, the interpretation of data requires ongoing assessment and consideration.

A Framework for Interpreting Evidence

2

JOHN BUCKLETON[a]

Contents

This book is intended as a discussion of the interpretation of DNA evidence. However, there is nothing inherently different about DNA evidence that sets it qualitatively aside from all other forensic evidence or even all evidence.[167,663] Hence, it is important that DNA evidence is considered as one form of evidence and not as something completely separate. We come immediately to the issue of setting evidence into a framework that is appropriate for court. This has been the subject of entire books by more informed authors,[659] but it is by no means settled. The issue revolves around a basic contrast: the tools best fitted to interpret evidence coherently are also those that appear to be most problematic to explain to a jury or layperson. Does the

[a] I acknowledge many valuable discussions over the years with Drs Christopher Triggs and Christophe Champod, which have contributed to material present in this chapter.

system used by the scientist also have to be the one presented in court? This is a question that is only just beginning to be asked, let alone answered. Parts of this section follow Triggs and Buckleton[783] reproduced with the kind permission of Oxford University Press.

The interpretation of the DNA results has been a key area for debate in the DNA field ever since the inception of this form of evidence.

> The statistical interpretation of DNA typing results, specifically in the context of population genetics, has been the least understood and, by definition, the most hotly debated issue of many admissibility hearings. The perceived incomprehensibility of the subject, has led to a recalcitrance of the judicial system to accept DNA typing.[678,679]

This statement by Rudin and Inman is not only true but is also very interesting. DNA evidence is actually much simpler and more extensively studied than most other evidence types. Many evidence types, such as toolmarks and handwriting comparison, are so complex that at present they defy presentation in a numerical form. Numerical assessment is attempted in glass and fiber evidence in New Zealand and the U.K., but the issues in both these fields are far more complex than in the DNA field. It is the very simplicity of DNA evidence that allows it to be presented numerically at all. And yet, as Rudin and Inman point out, there is still much debate about how to present this evidence.

It could be argued that the presentation of scientific evidence should bend to conform to the courts' requirements. Indeed a court can almost compel this. There have been several rulings[b] on this subject by courts, which have been used to argue for or against particular approaches to the presentation of evidence. An instance of this could be the Doheney and Adams ruling.[201] More specifically the Doheney and Adams ruling has been, I believe erroneously, read as arguing against a Bayesian approach and for a frequentist approach (discussed further later).[c] However, a fairer and more impartial appraisal of the various methods offered for interpretation should proceed from a starting point of discussing the underlying logic of interpretation. Only as a second stage should it be considered how this logic may be presented in court or whether the court or jury have the tools to deal with this type of evidence. There is little advantage to the situation "wrong but understood."[658,660]

[b] For reviews of some court cases in Australia, New Zealand, the U.K. and the U.S., see References 319, 329, 401, 649, 660, 661, 662, and 663.

[c] Robertson and Vignaux[663] give a more eloquently worded argument in support of this belief.

> To be effective in the courtroom, a statistician must be able to think like a lawyer and present complex statistical concepts in terms a judge can understand. Thus, we present the principles of statistics and probability, not as a series of symbols, but in the words of jurists.[362]

What is proposed in this chapter is to consider the underlying claims of three alternative approaches to the presentation of evidence. These will be termed the frequentist approach, the logical approach, and the full Bayesian approach.[429] The first of these terms has been in common usage and may be familiar.[585] I have adopted a slightly different phrasing to that in common usage for the second and third approaches and this will require some explanation. This will be attempted in the following sections. The intention is to present the merits and shortcomings of each method in an impartial way, which hopefully leads the reader to a position where they can make an informed choice. Juries may misunderstand any of the methods described, and care should be taken over the exact wording. In fact, it is clear that care must be taken with all probabilistic work and presentation.[35,101,287,472,473,474,475,476,477,493,503,554,664,764,766,769] One famous statistician's evidence has been described as "like hearing evidence in Swahili, without the advantage of an interpreter."[604]

Comparisons of the potential impact on juries of the different methods have been published.[299,365,754,755] It is necessary to countenance a situation in the future where the desirable methods for interpretation of, say, a mixture by simulation are so complex that they cannot realistically be explained completely in court.

It is important that the following discussion is read without fear of the more mathematical approaches as this fear wrongly pressures some commentators to advocate simpler approaches. It is probably fair for a jury to prefer a method for the reason of mathematical simplicity, but it would be a mistake for a scientist to do so. Would you like your aircraft designer to use the best engineering models available or one that you can understand without effort?

2.1 The Frequentist Approach

At the outset it is necessary to make clear that the use of the frequentist approach in forensic science is related to, but not identical to the frequentist approach in probability theory.[519,652] The frequentist approach in forensic science has never been formalized and hence is quite hard to discuss. It appears to have grown as a logical framework by a set of intuitive steps. There are also a number of potential misconceptions regarding this approach,

which require discussion and will be attempted. To begin, the approach will be subdivided into two parts: the coincidence probability and the exclusion probability. A discussion of "natural frequencies," a concept introduced by Gigerenzer, will follow.[336]

2.1.1 Coincidence Probabilities

For this discussion, it is necessary to attempt to formalize this approach sufficiently. Use will be made of the following definition:

The coincidence approach proceeds to offer evidence against a proposition by showing that the evidence is unlikely if this proposition is true. Hence it supports the alternative proposition. The less likely the evidence under the proposition, the more support given to the alternative.

This is called the coincidence probability approach because either the evidence came from, say, the suspect or a "coincidence" has occurred.

There are many examples of evidence presented in this way:

- "Only 1% of glass would match the glass on the clothing by chance."
- "It is very unlikely to get this paint sequence match by chance alone."
- "Approximately 1 in a million unrelated males would match the DNA at the scene by chance."

We are led to believe that the event "match by chance" is unlikely and hence the evidence supports the alternative. At this stage let us proceed by assuming that if the evidence is unlikely under a particular hypothesis, then this supports the alternative.

This is strongly akin to formal hypothesis testing procedures in statistical theory. Formal hypothesis testing would proceed by setting up the hypothesis usually called the null, H_0. The probability of the evidence (or data) is calculated if H_0 is true. If this probability is small (say less than 5 or 1%), then the null is "rejected." The evidence is taken to support the alternative hypothesis, H_1.[305,579,612]

To set up a DNA case in this framework, we could proceed as follows. Formulate the hypothesis, H_0: the DNA came from a male not related to the suspect. We then calculate the probability of the evidence if this is true. We write the evidence as E, and in this context it will be something like:

E: The DNA at the scene is type α.

We assume that it is known that the suspect is also type α. We calculate the probability, Pr, of the evidence, E, if the null hypothesis H_0 is true, $\Pr(E|H_0)$. The vertical line, or conditioning sign, stands for the word "if" or "given."

Assuming that about 1 in a million unrelated males would have type α, we assign $\Pr(E|H_0)$ as 1 in a million. Since this is a very small chance, we

assume that this evidence suggests that H_0 is not true and hence is support for H_1. In this context, we might define the alternative hypothesis as:

H_1: The DNA came from the suspect.

Hence in this case, the evidence supports the hypothesis that the DNA came from the suspect. Later we are going to need to be a lot more careful about how we define hypotheses.

Hypothesis testing is a well-known and largely accepted statistical approach. The similarity between the coincidence approach and hypothesis testing is the former's greatest claim to prominence.

2.1.2 Exclusion Probabilities

The exclusion probability approach calculates and reports the exclusion probability. This can be defined as the probability that a random person would be excluded as the donor of this DNA, or the father of this child, or a contributor to this mixture. The details of these calculations will be discussed later. Again, the formal logic has not been defined; hence, it will be attempted here.

The suspect is not excluded. There is a probability that a random person would be excluded. From this it is inferred that it is unlikely that the suspect is a random person. Hence this evidence supports the alternative proposition that the suspect is the donor of the DNA. The higher the exclusion probability, the more support given to the alternative.

Examples are again common. For instance, the three phrases given previously can be reworked into this framework:

- "99% of windows would be excluded as a source of this glass."
- "It is very likely that a random paint sequence would be excluded as matching this sample."
- "Approximately 99.9999% of unrelated males would be excluded as the source of this DNA."

An advantage of the exclusion probability approach is that it can be easily extended beyond these examples to more difficult types of evidence such as paternity and mixtures:

- "Approximately 99% of random men would be excluded as the father of this child."
- "Approximately 99% of random men would be excluded as a donor to this mixture."

It was stated previously that the use of the frequentist approach in forensic science is related, but not identical, to the frequentist approach in probability theory. There are two common definitions of probability. These are

called the frequentist and the subjectivist definitions. It is not necessary to discuss these differences in any length here, as they have long been the cause of deep discussion in both philosophy and the theory of probability. Briefly, the frequentist approach treats probability as the expectation over a large number of events. For instance, if we roll a dice many times we expect about 1/6 of these rolls to be a "6." The subjectivist definition accepts that probability is a measure of belief, and that this measure will be conditional both on the information available and on the person making the assessment.[756] However, both the coincidence approach and the exclusion probability approach can be based on either frequentist or subjectivist probabilities. Proponents of the Bayesian or subjectivist school of probability criticize the frequentist definition. However, it is unfair to transfer this criticism of a frequentist probability to the frequentist approach to forensic evidence.

The coincidence and the exclusion probability approach do appear to be simple and have an intuitive logic that may appeal to a jury. Thompson[767] argued for their use in the O.J. Simpson trial apparently on the basis that they were conservative and more easily understood while accepting the greater power of likelihood ratios.

2.1.3 Natural Frequencies[d]

More recently, the argument has been taken up by Gigerenzer arguing that "*to be good it must be understood.*" He argues persuasively for the use of "natural frequencies." To introduce this concept, it is easiest to follow an example from Gigerenzer.[336]

> The expert witness testifies that there are about 10 million men who could have been the perpetrator. Approximately 10 of these men have a DNA profile that is identical with the trace recovered from the crime scene. If a man has this profile it is practically certain that a DNA analysis shows a match. Among the men who do not have this DNA profile, current DNA technology leads to a reported match in only 100 cases out of 10 million.[e]

Gigerenzer argues from his own research that this approach is more likely to be understood. He quotes that the correct understanding was achieved by 1% of students and 10% of professionals when using conditional probabilities. This rose to 40 and 70%, respectively, when "natural frequencies" were used.

[d] My thanks to Michael Strutt for directing me to this work.
[e] Gigerenzer is referring here to his estimate of error rates.

Of course, Gigerenzer's natural frequencies are nothing more than an example of the defense attorney's fallacy of Thompson and Schumann[769] or the recommendation of the Appeal Court regarding Doheney and Adams.[201,618]

I concede the seductive appeal of this approach. Let us accept at face value Gigerenzer's statement that they are more easily understood. I do, however, feel that this approach hides a number of serious issues.

First consider the assumption that N men could have been the perpetrator. Who is to make this decision? One would feel that the only people qualified and with the responsibility of doing this are the judge and jury. They have heard the non-DNA evidence and they can decide whether or not this defines a pool of suspects. Moreover, are we to assign an equal prior to all these men? Gigerenzer's approach has a tendency toward assigning equal priors to each of these men and to the suspect. This is a tenable assumption in some but not all circumstances. Essentially we have a partition of the population of the world into those "in" the pool of suspects and those "out" of it. Those "in" are assigned a prior probability of $1/N$. Those "out" are assigned a prior of 0.

What are we to do when the product of the match probability and the pool of possible suspects is very small? Let us take the case given above but reduce the match probability from 1 in a million to 1 in 10 million. This would lead to:

The expert witness testifies that there are about 10 million men who could have been the perpetrator. Approximately 1 of these men has a DNA profile that is identical with the trace recovered from the crime scene.

The witness will have to take great care that the jury understand this statement. There is a risk that they may assume that the suspect is this one man. What is needed is to explain that this is one man additional to the suspect and even then it is an expectation. There may be one man additional to the suspect, but there may also be 0, 2, 3, or more.

Let us take this case and reduce the match probability even further to 1 in a billion. This would lead to:

The expert witness testifies that there are about 10 million men who could have been the perpetrator. Approximately 0.01 of these men have a DNA profile that is identical with the trace recovered from the crime scene.

This will take some care to explain to the jury. Now suppose that the suspect has one brother in the set of 10 million men.

The expert witness testifies that there are about 10 million unrelated men and one brother who could have been the perpetrator. Approximately 0.01 of the unrelated men and 0.005 of the brother have a DNA profile that is identical with the trace recovered from the crime scene.

Taking the example further:

The expert witness testifies that there are about 10 million unrelated men and one brother who could have been the perpetrator. Approximately 0.002 of the unrelated Caucasian men, 0.004 of the unrelated African Americans, 0.004

of the unrelated Hispanics, and 0.005 of the brother have a DNA profile that is identical with the trace recovered from the crime scene.

If we accept the suggestion that it is more understandable, then it may have a use in those very simple cases where there is a definable pool of suspects, relatedness is not important, the evidence is certain under the prosecution hypothesis, and the product of the match probability times N is not small.

Outside this very restricted class of case, I would classify it in the "understood but wrong"[f] category even when it is understood. I really do doubt the usefulness of this approach. It is very difficult to see how to accommodate relatives, interpret mixtures, and report paternity cases within this framework. Gigerenzer has also subtly introduced the concept of 10 million replicate cases all with the same probability of error. This may be an acceptable fiction to lure the jury into a balanced view, but it would take a lot of thinking to reconcile it with my own view of probability. Even if we accept Gigerenzer's statement that natural frequencies are more easily understood and we decided to use this presentation method in court, it is important that forensic scientists think more clearly and exactly about what a probability is, what constitutes replication, and how probabilities may be assigned.

2.2 The Logical Approach

"We are all Bayesians in real day life." Bruce Budowle.[119]

"Bayes's theorem is a fundamental tool of inductive inference." Finkelstein and Levin.[300]

Frustrations with the frequentist approach to forensic evidence have led many people to search for alternatives.[105,258] For many, these frustrations stem from discussing multiple stains, multiple suspects, or from trying to combine different evidence types.[652,656] The foremost alternative is the logical approach (also called the Bayesian approach).[257,490,500,516,517,518] This approach has been implemented routinely in paternity cases since the 1930s.[255] It is however only in the later stages of the 20th century that it made inroads into many other fields of forensic science. It now dominates forensic literature, but not necessarily forensic practice, as the method of choice for interpreting forensic evidence.[6,8,170,171,173,214,334,585,659,663] Bär[47] gives an elegant review.

Let:

H_p be the hypothesis advanced by the prosecution,
H_d be a particular hypothesis suitable for the defense,
E represent the evidence, and
I represent all the background evidence relevant to the case.

[f] It is not the only error in this section by Gigerenzer. Professor Weir did not report likelihood ratios in the O.J. Simpson case and most laboratories and all accredited ones do undertake external QA trials.

The laws of probability lead to

$$\frac{\Pr(H_p|E,I)}{\Pr(H_d|E,I)} = \frac{\Pr(E|H_p,I)}{\Pr(E|H_d,I)} \times \frac{\Pr(H_p|I)}{\Pr(H_d|I)} \qquad (2.1)$$

This theorem is known as Bayes's theorem.[53] A derivation appears in Box 2.1. This theorem follows directly from the laws of probability. It can therefore be accepted as a logical framework for interpreting evidence.

To understand the workings of this formula, it is necessary to understand the workings of the conditioning sign. This is usually written as | and can be read as "if" or "given." This concept is little understood and is typically not taught well. The reader unfamiliar with it would be advised to work through the examples given in Evett and Weir.[267] A brief discussion is given in Box 2.2.

Equation (2.1) is often given verbally as

$$\text{posterior odds} = \text{likelihood ratio} \times \text{prior odds} \qquad (2.2)$$

The prior odds are the odds on the hypotheses H_p before DNA evidence. The posterior odds are these odds after DNA evidence. The likelihood ratio tells us how to relate these two. This would seem to be a very worthwhile thing to

Box 2.1 A Derivation of Bayes's Theorem

The third law of probability states:

$\Pr(a \text{ and } b|c) = \Pr(a,b|c) = \Pr(a|b,c)\Pr(b|c) = \Pr(b|a,c)\Pr(a|c)$

Rewriting this using H_p, H_d, E, and I

$\Pr(H_p,E|I) = \Pr(H_p|E,I)\Pr(E|I) = \Pr(E|H_p,I)\Pr(H_p|I)$

and

$\Pr(H_d,E|I) = \Pr(H_d|E,I)\Pr(E|I) = \Pr(E|H_d,I)\Pr(H_d|I)$

Hence

$$\frac{\Pr(H_p,E|I)}{\Pr(H_d,E|I)} = \frac{\Pr(H_p|E,I)\Pr(E|I)}{\Pr(H_d|E,I)\Pr(E|I)} = \frac{\Pr(E|H_p,I)\Pr(H_p|I)}{\Pr(E|H_d,I)\Pr(H_d|I)}$$

Hence

$$\frac{\Pr(H_p|E,I)\Pr(E|I)}{\Pr(H_d|E,I)\Pr(E|I)} = \frac{\Pr(E|H_p,I)\Pr(H_p|I)}{\Pr(E|H_d,I)\Pr(H_d|I)}$$

Cancelling $\Pr(E|I)$

$$\frac{\Pr(H_p|E,I)}{\Pr(H_d|E,I)} = \frac{\Pr(E|H_p,I)\Pr(H_p|I)}{\Pr(E|H_d,I)\Pr(H_d|I)} \qquad (2.1)$$

do, that is, to relate the odds before consideration of the evidence to those after the evidence. It also tells us how to update our opinion in a logical manner having heard the evidence.

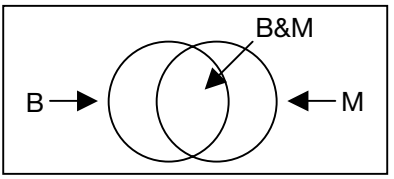
The prior odds, $\Pr(H_p|I)/\Pr(H_d|I)$, represent the view on the prosecution and defense hypothesis before DNA evidence is presented.[g] This view is something that is formed in the minds of the judge and jury. The information imparted to the jury is carefully restricted to those facts that are considered admissible and relevant. It is very unlikely that the prior odds are numerically expressed in the mind of the judge and jury and there is no need

Box 2.2. Conditional Probability

Several definitions are available for conditional probability. One proceeds from the third law of probability:

$$\Pr(a|b)=\frac{\Pr(a,b)}{\Pr(b)}$$

which can be interpreted quite well in set theory. For instance, evaluating $\Pr(a|b)$ involves enumerating the set of outcomes where event b is true and seeing in what fraction of these events a is also true.



Example: In a certain office there are ten men. Three men have beards (event B) and moustaches (event M). A further two have moustaches only. Say we were interested in $\Pr(B|M)$ we find the set of men where M is true: this has five members. Of these, three have beards. Hence $\Pr(B|M)=3/5$.

If we were interested in $\Pr(M|B)$ we find the set of men where B is true: this is three men. Of these, all three have moustaches. Hence $\Pr(M|B)=3/3=1$.[h]

[g] My wording is wrongly implying an order to events such as the "hearing of DNA evidence." In fact, the evidence can be heard in any order. The mathematical treatment will give the same result regardless of the order in which the evidence is considered.[659]

[h] In this simple example, we are making an assumption that each of the men is equally likely to be observed. This assumption may not be true in more general examples, but the principle behind the definition of the conditional probability remains valid.

that they should be numerical.[662,663] Strictly it is not the business of the scientist to form a view on the "prior odds" and most scientists would strictly avoid this (for a differing opinion, see Meester and Sjerps[543] and the subsequent discussion[220]). These odds are based on nonscientific evidence and it is the duty of judge and jury to assess this.[779,807]

The use of this approach typically reports only the likelihood ratio. By doing this the scientist reports the weight of the evidence without transgressing on those areas reserved for the judge and jury. This is the reason why the term "the logical approach"[i] has been used to describe this method. It has also been described elsewhere as "the likelihood ratio" approach. The term that is being avoided is "the Bayesian approach," which is the term used in most papers on this subject, including my own. This term is being avoided because, strictly, presenting a ratio of likelihoods does not necessarily imply the use of the Bayesian method. Most authors have intended the presentation of the likelihood ratio alone without necessarily implying that a discussion of Bayes' theorem and prior odds would follow in court. The intent was to present the scientific evidence in the context of a logical framework without necessarily presenting that framework.

However, the advantage of the logical approach is that the likelihood ratio can be put in a context of the entire case and in a consistent and logical framework. This advantage is somewhat lost if judge, jury, and scientist are reticent to use or even discuss Bayes' theorem in full.

Thompson[767] warns:

> Although likelihood ratios have appealing features, the academic community has yet fully to analyse and discuss their usefulness for characterising DNA evidence.

Pfannkuch et al.[616] describe their experiences teaching this material to undergraduate students:

> Bayes' theorem was the killer. There was an exodus of those mathematically unprepared and math-phobic students who were free to leave the course, supplemented by panic and agonised discussions with those who were trapped by their course requirements.

These professional scientists and teachers persisted and found good methods for teaching even math-phobic students because of the "*wealth of socially important problems*" that are best addressed by Bayes' theorem.

[i] I first had this distinction explained to me by Dr Christophe Champod.

Fenton and Neil[288] argue forcefully that Bayes' theorem is the method of choice for interpreting evidence, while giving the fair criticism that Bayesians have failed in their duty of communication. They quote the fact that many lawyers and other educated professionals misunderstand the subject.

Is there a lesson here? My own experience with practicing forensic scientists is that they can achieve an in-depth understanding of complex mathematical concepts and methods, especially when placed in a good learning environment and supported by colleagues and management. In this regard, I would like to commend the U.K. Forensic Science Service (FSS) practice of secluding scientists during training (in England we used excellent hotels in Evesham and the "Pudding club" somewhere south of Birmingham). The FSS also undertakes basic probability training and is considering putting in place a numerical competency in recruitment. This investment in people is repaid manyfold.

To gain familiarity with Equation (2.2), it is useful to consider a few results. What would happen if the likelihood ratio was 1? In this case, the posterior odds are unchanged by the evidence. Another way of putting this is that the evidence is inconclusive.

What would happen if the likelihood ratio was greater than 1? In these cases, the posterior odds would be greater than the prior odds. The evidence would have increased our belief in H_p relative to H_d. Another way of putting this is that the evidence supports H_p. The higher the likelihood ratio, the greater the support for H_p.

If the likelihood ratio is less than 1, the posterior odds would be smaller than the prior odds. The evidence would have decreased our belief in H_p relative to H_d. Another way of putting this is that the evidence supports H_d. The lower the likelihood ratio, the greater the support for H_d.

It has been suggested that a nomogram may be useful to help explain the use of this formulation. This follows from a well-known nomogram in clinical medicine. Riancho and Zarrabeitia[642] suggest the diagram that has been modified and presented in Tables 2.1 and 2.2. These tables are used by choosing a prior odds and drawing a line through the center of the *LR* value. The posterior odds may then be read directly. For example, assume that the prior odds are about 1 to 100,000 (against) and the likelihood ratio is 10,000,000; then we read the posterior odds as 100 to 1 (on).

The likelihood ratio (*LR*) is a numerical scale. One point can be hinged to words without argument; an *LR* of 1 is inconclusive. Other words may be attached to this scale to give a subjective verbal impression of the weight of evidence.[12,94,174,263,264] This association of words with numbers is subjective and necessarily arbitrary. One such scale used extensively in the FSS is given in Table 2.3.

The question of development of the prosecution and defense hypotheses was introduced above, but was not discussed in any depth. In fact, the defense are under no obligation to offer any hypothesis at all. An early discussion of

Table 2.1 Prosecutor's Nomogram

Prior		Likelihood Ratio	Posterior	
Probability	Odds		Odds	Probability
			100,000,000 to 1	99.999990%
0.001%	1 to 100,000		10,000,000 to 1	99.999989%
0.01%	1 to 10,000	10,000,000,000 1,000,000,000	1,000,000 to 1	99.9999%
0.1%	1 to 1000	100,000,000 10,000,000	100,000 to 1	99.999%
1%	1 to 100	1,000,000 100,000	10,000 to 1	99.99%
9%	1 to 10	10,000 1000	1000 to 1	99.9%
50%	1 to 1	100 10	100 to 1	99%
91%	10 to 1	1	10 to 1	91%
99%	100 to 1		1 to 1	50%

The prior and posterior probabilities associated with these odds are given next to the odds.

Reproduced and amended from Riancho and Zarrabeitia[642] with kind permission of the authors and Springer-Verlag who retain ownership of the copyright.

this appears in Aitken.[8] This is the subject of a large-scale project in the FSS called the Case Assessment Initiative.[193,194,272] The subject warrants separate treatment. Even though it has been introduced under the heading of "the logical approach," the development of propositions is actually universally important to evidence interpretation by any method (see Box 2.3).

2.3 The Full Bayesian Approach

The analysis given under the title of "the logical approach" works well if there are two clear hypotheses aligned with the prosecution and defense positions. However, regularly it is difficult to simplify a real casework problem down to two hypotheses.

To put this in context, consider a relatively simple STR case. We have a stain at the scene of a crime. Call this stain c and the genotype of this stain G_c, following the nomenclature of Evett and Weir.[267] A suspect comes to the attention of the police. Call this person s and the genotype G_s. The genotype of the suspect and the crime stain are found to be the same. We will write this as G_s=G_c.

Table 2.2 Defendant's Nomogram

Prior		Likelihood Ratio	Posterior	
Probability	Odds		Odds	Probability
0.1%	1 to 1000		100 to 1	99%
1%	1 to 100		10 to 1	91%
9%	1 to 10	10	1 to 1	50%
		1		
50%	1 to 1	1/10	1 to 10	9%
		1/100		
91%	10 to 1	1/1000	1 to 100	1%
		1/10,000		
99%	100 to 1	1/100,000	1 to 1000	0.1%
		1/1,000,000		
99.9%	1000 to 1	1/10,000,000	1 to 10,000	0.01%
		1/100,000,000		
99.99%	10,000 to 1	1/1,000,000,000	1 to 100,000	0.001%
			1 to 1,000,000	0.0001%

Reproduced and amended from Riancho and Zarrabeitia[642] with kind permission of the authors and Springer-Verlag who retain ownership of the copyright.

Table 2.3 A Verbal Scale

LR	Verbal Wording	
1,000,000+	Extremely strong	
100,000	Very strong	
10,000	Strong	Support for H_p
1000	Moderately strong	
100	Moderate	
10	Limited	
1	Inconclusive	
0.1	Limited	
0.01	Moderate	
0.001	Moderately strong	Support for H_d
0.0001	Strong	
0.00001	Very strong	
0.000001	Extremely strong	

Box 2.3 Which Way Up?

When we introduced Bayes's theorem, we wrote it as

$$\frac{\Pr(H_p|E,I)}{\Pr(H_d,E|I)}=\frac{\Pr(E|H_p,I)\Pr(H_p|I)}{\Pr(E|H_d,I)\Pr(H_d|I)} \tag{2.1}$$

Why did we write it this way up? What was wrong with

$$\frac{\Pr(H_d|E,I)}{\Pr(H_p|E,I)}=\frac{\Pr(E|H_d,I)\Pr(H_d|I)}{\Pr(E|H_p,I)\Pr(H_p|I)}\ ?$$

This approach would work just as well. High numbers would be support for H_d, typically the defense hypothesis. Is the reason we defined it with H_p on top an indication of subconscious bias? Is this the reason Balding, Donnelly, and Nichols[42] wrote their *LR*'s up the other way? Were they trying to help us see something?

Under the coincidence approach, this would be the match that is caused by the suspect being the donor of the crime stain or by a coincidence. To make the comparison with hypothesis testing, we would formulate

H_0: The DNA came from a male not related to the suspect.
H_1: The DNA came from the suspect.

We then calculate the probability of the evidence if this is true. Let us write this as $\Pr(G_c|G_s, H_0)$, which can be read as the probability of the genotype of the crime stain if the crime stain came from a person unrelated to the suspect (and the suspect's genotype is G_s). This is often written as f,[j] and taken to be the frequency of the crime genotype (or the suspect's genotype since they are the same). We assume that this frequency is small and hence there is evidence against H_0 and for H_1.

Under the "logical approach," we simply rename these hypotheses:

H_p: The DNA came from the suspect.
H_d: The DNA came from a male not related to the suspect.

We then calculate the probability of the evidence under each of these hypotheses.[5,6,8,9,10,11,37,257,267] $\Pr(G_c|G_s, H_p)=1$ since the crime genotype will be G_c if it came from the suspect who is G_s. Again we take $\Pr(G_c|G_s, H_d)=f$. Hence

$$LR=\frac{1}{\Pr(G_c|G_s,H_d)}=\frac{1}{f} \tag{2.3}$$

[j] This term will stand for two concepts in this text. This is unavoidable if we are to align with the published literature. In the context, it stands for the frequency of a profile. However, in population genetics f is often used for the within-population inbreeding parameter. When used in this latter context, it is synonymous with F_{IS}.

which is (typically) very much larger than 1 and hence there is evidence against H_d and for H_p.

But note that the following two hypotheses are not exhaustive:

H_p: The DNA came from the suspect.
H_d: The DNA came from a male not related to the suspect.

What about those people who *are* related to the suspect? Should they be considered? Genetic theory would suggest that these are the most important people to consider, and should not be omitted from the analysis. What we need is a number of hypotheses. These could be:

H_1: The DNA came from the suspect.
H_2: The DNA came from a male related to the suspect.
H_3: The DNA came from a male not related to the suspect.

Now consider H_2. What do we mean by "related"? Obviously there are many different degrees of relatedness. Suppose that the suspect has one father and one mother, several brothers, numerous cousins and second cousins, etc. We may need a multiplicity of hypotheses. In fact, we could envisage the situation where there is a specific hypothesis for every person on earth:

H_1: The DNA came from the suspect.
H_2: The DNA came from person 2, the brother of the suspect.
H_3: The DNA came from person 3, the father of the suspect.
⋮
H_i: The DNA came from person i related in whatever way to the suspect.
⋮
H_j: The DNA came from person j so distantly related that we consider the person effectively unrelated to the suspect.

What we need is a formulation that can handle from three to many hypotheses. Considering the enumeration given above, there would be about 6,000,000,000 hypotheses, one for each person on earth.

This is provided by the general form of Bayes' theorem (derived in Box 2.4).[37,38,42] This states that

$$\Pr(H_1|G_c, G_s) = \frac{\Pr(H_1)}{\sum_{i=1}^{N} \Pr(G_c|G_s, H_i)\Pr(H_i)}. \tag{2.4}$$

This equation is very instructive for our thinking but is unlikely to be directly useful in court, at least in the current environment. This is because the terms $\Pr(H_i)$ relate to the prior probability that the ith person is the source of the DNA. The introduction of such considerations by a forensic scientist is unlikely

Box 2.4

A comprehensive equation has been proposed[37] based on the general formulation of Bayes' rule. Following Evett and Weir:[267] for a population of size N, we index the suspect as person 1 and the remaining members of the population as 2, …, N. We will call the hypothesis that person i is the source of the DNA H_i. Since the suspect is indexed person 1, the hypothesis that the suspect is, in fact, the source of the DNA is H_1. The remaining hypotheses, H_2, …, H_N, are those hypotheses where the true offender is some other person. Before we examine the evidence, each person has some probability of being the offender $\Pr(H_i)=\pi_i$. Many factors may affect this, one of these being geography. Those closest to the scene may have higher prior probabilities while people in remote countries have very low prior probabilities. Most of the people other than the suspect or suspects will not have been investigated. Therefore, there may be little specific evidence to inform this prior other than general aspects such as sex, age, etc. The suspect is genotyped and we will call the genotype G_s. The stain from the scene is typed and found to have the genetic profile G_c, which matches the suspect. The remaining 2, …, N members of the population have genotypes G_2, …,G_N. These 2, …, N people have not been genotyped. We require the probability $\Pr(H_1|G_c, G_s)$. This is given by Bayes's rule as

$$\Pr(H_1|G_c,G_s)=\frac{\Pr(G_c|G_s,H_1)\Pr(G_s|H_1)\Pr(H_1)}{\sum_{i=1}^{N}\Pr(G_c|G_s, H_i)\Pr(G_s|H_i)\Pr(H_i)}$$

Assuming that $\Pr(G_s|H_1)=\Pr(G_s|H_i)$ for all i, we obtain

$$\Pr(H_1|G_s,G_c)=\frac{\Pr(G_c|G_s,H_1)\Pr(H_1)}{\sum_{i=1}^{N}\Pr(G_c|G_s,H_i)\Pr(H_i)}$$

We assume that the probability that the scene stain will be type G_c, given that the suspect is G_s and he contributed the stain, is 1. Hence,

$$\Pr(H_1|G_s,G_c)=\frac{\Pr(H_1)}{\sum_{i=1}^{N}\Pr(G_c|G_s,H_i)\Pr(H_i)} \tag{2.4}$$

$$=\frac{1}{1+\sum_{i=2}^{N}\dfrac{\Pr(G_c|G_s,H_i)\Pr(H_1)}{\Pr(H_i)}}$$

(*continued*)

Box 2.4 (*continued*)

$$= \frac{1}{1+\sum_{i=2}^{N}\frac{\Pr(G_c|G_s,H_i)\pi_1}{\pi_i}}$$

Writing $\pi_i / \pi_1 = w_i$, we obtain

$$\Pr(H_1|G_s,G_c) = \frac{1}{1+\sum_{i=2}^{N}\Pr(G_c|G_s,H_i)w_i}$$

which is the equation given on page 41 of Evett and Weir. Here w_i can be regarded as a weighting function that expresses how much more or less probable the *i*th person is than the suspect to have left the crime stain based on only the non-DNA evidence.

to be permitted in court.[k] However, such an approach may be possible if the court supplies its view of the prior. For instance, the terms "forensically relevant populations"[131] and "relevant subgroup"[201] provide inadvertent references to such a prior. The time may come when courts countenance this type of consideration. We could envisage the situation where a court instructs the witness to consider only the subgroup "Caucasian sexually active males in the Manchester area," which is, in effect, setting a prior of zero outside this group.

In the likely absence of courts providing such priors, it is suggested that this unifying equation should be used to test various forensic approaches and to instruct our thinking. However, there is so much benefit in the use of this equation that research into how it could be used in court would be very welcome.

2.4 A Possible Solution

There is a "halfway house" between the likelihood ratio approach and the unifying equation that has neither been published previously nor tested, but has some considerable merit. Using the same nomenclature as above, we rewrite the likelihood ratio as

$$LR = \frac{\Pr(G_c|G_s,H_p)}{\sum_{i=2}^{N}\Pr(G_c|G_s,H_i,H_d)\Pr(H_i|H_d)} \tag{2.5}$$

where $H_2, \ldots, H_N$ is an exclusive and exhaustive partition of H_d (following Champod,[169] we will call these subpropositions). The advantage of this

[k] Meester and Sjerps[543] argue to the contrary.

approach is that it only requires priors that partition the probability under H_d. There is no requirement for the relative priors on H_p and H_d. This may be more acceptable to a court.

2.5 Comparison of the Different Approaches

The very brief summary of the alternative approaches given above does not do full justice to any of them. It is possible, however, to compare them. In the most simplistic overview, I would state that:

- The frequentist approach considers the probability of the evidence under one hypothesis.
- The logical approach considers the probability of the evidence under two competing hypotheses.
- The full Bayesian approach considers it under any number of hypotheses.

If we turn first to a critique of the frequentist approach, the most damning criticism is a lack of logical rigor. In the description given above, you will see that I struggled to define the frequentist approach and its line of logic with any accuracy. This is not because of laziness but rather that the definition and line of logic has never been given explicitly, and indeed it may not be possible to do so.

Consider the probability that is calculated. We calculate $\Pr(E|H_0)$ under the frequentist view. If it is small, we support H_1.

First note that because $\Pr(E|H_0)$ is small, this does not mean that $\Pr(H_0|E)$ is small. This is called the fallacy of the transposed conditional.[769]

Second note that simply because $\Pr(E|H_0)$ is small does not mean that $\Pr(E|H_1)$ is large. What if it was also small? Robertson and Vignaux[659] give a thought-provoking example adapted here slightly: Consider a child abuse case. Evidence is given that this child rocks and that only 3% of nonabused children rock. It might be tempting to assume that this child is abused since the evidence (R: rocking) is unlikely under the hypothesis (H_0: This child is nonabused). But we may be wrong to do so. Imagine that we now hear that only 3% of abused children rock. This would crucially alter our view of the evidence. We see that we cannot evaluate evidence by considering its probability under only one hypothesis. This has been given as a basic principle of evidence interpretation by Evett and Weir [267]and Evett et al.[281]

The logical flaws in the frequentist approach are what have driven many people to seek alternatives. Fortunately for justice and unfortunately for the advance of logic in forensic science, this flaw does not manifest itself in most simple STR cases. This is because the evidence is often certain under the

alternative H_1. In such cases, the frequentist approach reports f' and the logical approach LR = $1/f$. Critics of the logical approach understandably ask what all the fuss is about when all that is done in simple cases is calculate one divided by the frequency. Other criticisms have been offered. Effectively, these relate to reasonable criticisms of the difficulty of implementation and less reasonable criticisms arising largely from a lack of understanding of the underlying logic.[657,864] This brings us to a critique of the logical approach.

If we start with difficulty of implementation, one reasonable criticism of the logical approach is the ponderous nature of a statement involving a likelihood ratio. Contrast **A** with **B**:

A: The frequency of this profile among unrelated males in the population is less than 1 in a billion.
B: This evidence is more than a billion times more likely if the DNA came from the suspect than if it came from an unrelated male.

Many people would prefer **A** over **B**, and in fact studies have demonstrated that there are serious problems with understanding statements like **B**.[754,755] Some respondents described **B**-type statements as "*patently wrong*." This is not to imply that there is no prospect of misunderstanding a frequentist statement because there clearly is, but rather to suggest that the likelihood ratio wording is more ponderous and will take more skill and explanation to present.

We next move on to the fact that the very advantage of the "logical approach" is that the likelihood ratio can be placed in the context of a logical framework. This logical framework requires application of Bayes' rule and hence some assessment of priors. However, the legal system of many countries relies on the "common sense" of jurors and would hesitate to tell jurors how to think.[201,618,660] Forcing jurors to consider Bayes's theorem would be unacceptable in most legal systems. It is likely that application of common sense will lead to logical errors, and it has been shown that jurors do not handle probabilistic evidence well. However, there is no reason to believe that these logical errors would be removed by application of a partially understood logical system, which is the most likely outcome of trying to introduce Bayes' theorem into court. If we recoil from introducing Bayes' theorem in court, then the likelihood ratio approach forfeits one of its principal advantages although it certainly retains many others in assisting the thinking of the scientist.

This is not a fatal flaw as likelihood ratios have been presented in paternity evidence since the mid-1900s. In this context, they are typically termed paternity indices and are the method of choice in paternity work.

Inman and Rudin[429] note that: "While we are convinced that these ideas are both legitimate and useful, they have not been generally embraced by the practising community of criminalists, nor have they undergone the refinement that only comes with use over time." This is fair comment from a U.S. viewpoint.

The considerations given above are real issues when applying the logical approach. There are a few more objections that arise largely from a misunderstanding of the underlying logic. These would include criticisms of conditioning on I and H_p and the arbitrariness of the verbal scale. For an elegant discussion, see Robertson and Vignaux.[652]

Forensic scientists are raised in a culture that demands that they should avoid any bias that may arise from ideas seeded into their minds by the prosecution (or anyone else). This has led to the interpretation that they should consider the evidence in isolation from the background facts or the prosecution hypothesis. This idea is a misconception or misreading of the use of the conditioning in the probability assessment. In essence all probabilities are conditional, and the more relevant the information used in the conditioning, the more relevant the resulting probability assignment. Failure to consider relevant background information would be a disservice to the court. An example given by Dr. Ian Evett considers the question: What is the probability that Sarah is over 5 feet 8 inches? We could try to assign this probability, but our view would change markedly if we were told that Sarah is a giraffe. Ignoring the background information (Sarah is a giraffe) will lead to a much poorer assignment of probability. This is certainly not intended to sanction inappropriate information and conditioning.

The second argument is a verbal trick undertaken in the legal context. Consider the numerator of the likelihood ratio. This is $\Pr(E|H_p, I)$, which can be read as: the probability of the evidence given that the prosecution hypothesis is correct and given the background information. The (false legal) argument would be that it is inconsistent with the presumption of innocence to "assume that the prosecution hypothesis is true." This again is a misconception or a misreading of the conditioning. When calculating the likelihood ratio we are not assuming that the prosecution hypothesis is true, which indeed would be bias. What we are doing is weighing the prosecution and defense hypotheses against each other by calculating the probability of the evidence *if* these hypotheses were true. This is an instance where the verbal rendering of Bayes' rule can be misconstrued (possibly deliberately) to give a false impression never intended in the logical framework.

I have also heard the following erroneous argument: If H_1 and H_2 are independent, then $\Pr(H_1 \text{ and } H_2)$ is less than $\Pr(H_1)$ or $\Pr(H_2)$.

This part of the statement is correct. It is actually correct whether or not the events are independent. However, sometimes it is extended to "in a trial in which the case for the prosecution involves many propositions that must be jointly evaluated the probability of the conjunction of these hypotheses will typically drop below .5, so it would seem that a probabilistically sophisticated jury would never have cause to convict anyone."[759]

Of course this is an erroneous attack on probability theory *per se*, not specifically Bayesian inference. But let us examine the argument. Suppose we have events:

B: The suspect had intercourse with the victim.
C: The intercourse was not consensual.

Let us assume that there is some evidence, *E*. We seek the probability of guilt, *G*, given the evidence, Pr(*G*|*E*). Usually a court would require both *B* and *C* to be very probable to conclude *G*. It is logically certain that Pr(*B* and *C*|*E*) is less than or equal to Pr(*B*|*E*) and it is also less than or equal to Pr(*C*|*E*). However rather than being logically worrying, this is actually the correct conclusion. If there is doubt about *B* or *C* or collectively doubt about *B* and *C*, then *G* is not a safe conclusion, and I would be very concerned about any inference system that did not follow these rules. Robertson and Vignaux[659,662,663] argue eloquently that any method of inference that does not comply with the laws of probability must be suspect.

However, I am unsure whether this was the point that was being advanced. Let us assume that the *propositions* are something like:

A: The blood on Mr. Simpson's sock is from Nicole Brown.
B: The blood on the Bundy walk is from Mr. Simpson.
C: LAPD did not plant the blood on the sock.
D: LAPD did not plant the blood on the Bundy walk.

Suppose that guilt is established if *A*, *B*, *C*, and *D* are true. Indeed then Pr(*A*, *B*, *C*, *D*) would be less than the probability of any of the individual events. However, in my view guilt may also be established if *A*, *B* and *C* are true but *D* is false (there are other combinations).

Let us assume that guilt is certain if one of the following combinations of events held:

True	**False**
A,B,C,D	
A,B,C	*D*
A,B,D	*C*
A,C	*B,D*
B,D	*A,C*

Guilt may also be true under other combinations that are not listed, but in such a case none of these events, *A, B, C,* or *D,* would be evidence for it.

Let us conservatively assign the probability of guilt under these alternatives as zero. Then

$$\Pr(G) = \Pr(A, B, C, D) + \Pr(A, B, C, \bar{D}) + \Pr(A, B, \bar{C}, D) + \Pr(A, \bar{B}, C, \bar{D}) + \Pr(\bar{A}, B, \bar{C}, D)$$

The probabilistic argument can be extended rather easily to any number of events or to include instances where guilt is not certain but probable given certain events. In fact rather than being problematic, I find the laws of probability rather useful.

Regarding the arbitrariness of the verbal scale, this point must be conceded except with reference to the point labeled inconclusive. However, any verbal scale, Bayesian or otherwise, is arbitrary. The problem really relates to aligning words that are fuzzy and have different meanings to different people to a numerical scale that possesses all the beauty that is associated with numbers. This problem will be alleviated in those rare cases where the logic and numbers are themselves presented and understood in court.

This brings us to the full Bayesian approach. There is little doubt that this approach is the most mathematically useful. Most importantly, it can accommodate any number of hypotheses, which allows us to phrase the problem in more realistic ways. It is the underlying basis of Bayes' nets, which will certainly play a prominent part in evidence interpretation in the future. However, it is impossible to separate out the prior probabilities from this formulation, and hence implementation would be possible only in those unlikely cases where the court was prepared to provide its prior beliefs in a numerical format. At this time, the approach must be considered as the best and most useful tool for the scientist to use, but currently not presentable in court. The unanswered question is whether the compromise approach given above is an acceptable solution to the courts.

When weighing these approaches against each other, the reader should also consider that the vast majority of the modern published literature on evidence interpretation advocates the logical or full Bayesian approaches. There is very little published literature advocating a frequentist approach, possibly because the lack of formal rigor in this approach makes publication difficult.

Throughout this book we will attempt to present the evidence in both a frequentist and a likelihood ratio method where possible. There are some situations, such as missing persons' casework, paternity, and mixtures, where only the likelihood ratio approach is logically defensible.

2.6 Evidence Interpretation in Court

2.6.1 The Fallacy of the Transposed Conditional

Initially I had not planned to write anything on the famous fallacies and especially the fallacy of the transposed conditional also known as the prosecutor's fallacy. What was left to say after so many publications on the subject?[7,35,58,97,148,201,239,261,267,288,291,334,399,506,618,637,638,654,655,658,659,660,661,686,769,841] However, I discovered in 2003 that there was still much uncertainty about the subject and indeed that groups of people with important responsibilities in the criminal justice system had never heard of the issue.

What can I add to a debate that is already well written about? I will again explain it here for those readers for whom the fallacies are new. I will also add a section that attempts to assess the mathematical consequences of this error and gives some tips on how to avoid making a transposition. Many of these tips come from my experiences working with colleagues at the Interpretation Research Group of the FSS in the U.K.: Champod, McCrossan, Jackson, Pope, Foreman, and most particularly Ian Evett. Few forensic caseworkers have written on the subject, although most have faced it.

The fallacy of the transposed conditional is not peculiar to the logical approach. It can occur with a frequentist approach as well. Opinion is divided as to whether the fallacy is more or less likely when using the logical approach. In essence, it comes from confusing the probability of the evidence given a specific hypothesis with the probability of the hypothesis itself. In the terms given above, this would be confusing $\Pr(E|H_p)$ with $\Pr(H_p)$, $\Pr(H_p|E)$, or $\Pr(H_p|E, I)$.

Following a publication by Evett,[256] we introduce the subject by asking "What is the probability of having four legs IF you are an elephant?" Let us write this as $\Pr(4|E)$ and we assign it a high value, say, 0.999.

Next we consider "what is the probability of being an elephant IF you have four legs?" Write this as $\Pr(E|4)$ and note that it is a very different probability and not likely to be equal to 0.999. This example seems very easy to understand both verbally and in the symbolic language of probability. But the fallacy seems to be quite tricky to avoid in court.

Imagine that we have testified in court along the lines of one of the statements given below:

- The probability of obtaining this profile from an unrelated male member of the New Zealand population is 1 in 3 billion.
- The frequency of this profile among members of the population of New Zealand unrelated to Mr. Smith is 1 in 3 billion.
- This profile is 3 billion times more likely if it came from Mr. Smith than if it came from an unrelated male member of the New Zealand population.

The first two are frequentist statements and the last is a statement of the likelihood ratio. Let us work with the first. We are quite likely in court to face a question along the lines: "In lay terms do you mean that the probability that this blood came from someone else is 1 in 3 billion?"

This is the fallacy of the transposed conditional. It has led to appeals and retrials. It appears to be very natural to make this transposition however incorrect. Every newspaper report of a trial that I have read is transposed and I suspect that many jurors and indeed judges make it.

How can a scientist who is testifying avoid this error? The answer involves training and thinking on one's feet. But I report here Stella's Spotting Trick (named after Stella McCrossan) and Ian's Coping Trick (named after Ian Evett).

Stella's spotting trick: The key that Stella taught was to ask oneself whether the statement given is a question about the evidence or hypothesis. Probabilistic statements about the hypothesis will be transpositions. Those about the evidence are likely to be correct. The moment that you notice the statement does NOT contain an IF or a GIVEN you should be cautious. Consider the sentence given above: "In lay terms do you mean that the probability that this blood came from someone else is 1 in a billion?" Is this a statement about a proposition or the evidence? The proposition here is that the blood came from someone else. And indeed the statement is a question about the probability of the proposition. Hence it is a transposition.

Ian's coping trick: The essence of this trick is to identify those statements that you are confident are correct and those that you are confident are incorrect. This is best done by memory. There will be a few standard statements that you know to be correct and a few transpositions that you know to be incorrect. Memorize these. Then there is the huge range of statements in between. These may be correct or incorrect. The prosecutor may have transposed in his/her head and is trying to get you to say what he/she thinks is a more simple statement. That is his/her fault not yours (if you are a forensic scientist reading this). He/she should have read and studied more. In this circumstance I suggest you say something like:

> I have been taught to be very careful with probabilistic statements. Subtle misstatements have led to appeals in the past. I am unsure whether your phrasing is correct or incorrect. However I can give some statements that I know are correct.

These will include the numerical statement of type 1, 2, or 3 given above or the verbal statements given in Table 2.3.

Of course, care by the scientist is no guarantee that the jury, judge, or press will not make the transposition themselves. For instance, Bruce Weir

had gone to great trouble with the wording in the report for his testimony in the O.J. Simpson case. Weir was careful and correct in his verbal testimony as well. As an example, he reported that there was a 1 in 1400 chance that the profile on the Bronco center console would have this DNA profile IF it had come from two people other than Mr. Simpson and Mr. Goldman. This was transposed by Linda Deutsh of the Associated Press (June 26, 1995) to "a chance of 1 in 1400 that any two people in the population could be responsible for such a stain." To quote Professor Weir: "It is incumbent on both the prosecution and defense to explain the meaning of a conditional probability of a DNA profile."[835]

I found another transposition in an interesting place. Horgan[415] was warning about errors in the Simpson case and went on to commit the prosecutor's fallacy while explaining the error of the defender's fallacy! "Given odds of 1 in 100,000 that a blood sample came from someone other than Simpson, a lawyer could point out that Los Angeles contains 10 million people and therefore 100 other potential suspects. That argument is obviously specious…" All the students in the 2003 (University of Auckland, New Zealand) Forensic Science class spotted the error when given it as an assignment!

Mathematical consequences of transposition: The transposition is of no consequence if the prior odds are in fact 1. This is because the answer arrived at by transposition and the "correct" answers are the same in this circumstance. The issue only occurs if the prior odds differ from 1. If the odds are greater than 1, then the transposition is conservative. Table 2.4 gives some posterior probabilities for differing prior probabilities. The table shows, as is known, that for a high likelihood ratio (a low match probability) the practical consequences are negligible. The practical consequences, if they occur at all, are for lower likelihood ratios and where there is little "other" evidence against the defendant or where there is evidence for the defendant.[95,833]

2.6.2 Establishing the Propositions

The concept of a hierarchy of propositions was first introduced by Aitken[8] and greatly developed by Cook et al.[193] and Evett et al.[272] Propositions are classified into three levels: offense, activity, or source. The top of the hierarchy is taken to be the offense level, where the issue is one of guilt or innocence. An example of this could be "the suspect raped the victim." It is often held that this level of proposition is for the courts to consider and above the level at which a forensic scientist would usually operate. The next level is taken to be the activity level. An example would be "the suspect had intercourse with the victim." This differs from the offense level in that it talks about an activity (intercourse) without making a comment about its intent (rape) that would need to consider other matters such as consent. The lowest level is taken to be

Table 2.4 Consequences of Transposing Assuming that DNA Evidence Gives a Match with Match Probability 1 in a Billion and My Subjective View of This

Prior Odds	Meaning	Posterior Probability with Transposition	Posterior Probability without Transposition	My Subjective View
4,000,000:1 against	The defendant is as likely as anyone else in New Zealand to be the donor	0.999999999	0.996015936	No practical consequence
4000:1 against	The defendant is more likely than a random person in New Zealand to be the donor	0.999999999	0.999996000	No practical consequence
1:1	The suspect is vastly more likely than a random person in New Zealand to be the donor	0.999999999	0.999999999	No practical consequence
Any odds on	The suspect is vastly more likely than a random person in New Zealand to be the donor			No practical consequence

the source level. At this level, we consider questions of the type "did this semen come from the suspect?" Considerations at this level do not directly relate to activity, in this example intercourse, which would involve issues such as from whence the sample was taken, drainage, and contamination.

It has become necessary to add another level below the source level. This has been termed sublevel 1. This has arisen because it is not always certain from what body fluid the DNA may have come. For instance, when considering the source level proposition "the semen came from the suspect," the sublevel 1 proposition would be "the DNA came from the suspect."

The further down the hierarchy the scientist operates, the more the responsibility for interpreting the evidence is transferred to the court.

It would be reasonable to leave the interpretation of such matters to the court if that were the best body to undertake this interpretation. However, if the matter requires expert knowledge regarding such matters as transfer and persistence, it would seem wise for the scientist to attempt interpretation at a higher level in the hierarchy, or at least to warn and equip the court to make such an attempt. The evidence must eventually be interpreted at the offense level by the court. If the evidence cannot be put in the context of the offense, then it is, in itself, irrelevant to the court.

Let us assume that the scientist can make a decision as to which level in the hierarchy the propositions should be formulated. The next step is to attempt to formulate one hypothesis for the prosecution and one for the defense. The defense are under no obligation to provide a hypothesis and, in fact, the defendant may have given a "no comment" interview. McCrossan et al. (in draft) ask:

> Is it the role of the forensic scientist to formulate the defense proposition when "no comment" is given?
> If the scientist does formulate a proposition on behalf of the defense, how should the implications of this action be highlighted/exposed in the statement?

One issue here is the consideration of the obvious alternative:

H_d: The suspect had nothing to do with the …(activity associated with the crime)

tends to maximize the *LR* and hence has a tendency to maximize the apparent weight of the evidence.

There is an issue as to whether the defense must choose only one proposition or whether they can have many. In fact, it is worthwhile considering what happens if the prosecution and defense hypotheses are not exhaustive. Let us assume that there could be three hypotheses H_1, H_2, and H_3. H_1 aligns with the prosecution view of the case, H_2 is the hypothesis chosen for the defense, and H_3 is any hypothesis that has been ignored in the analysis but is also consistent with innocence.

Set hypothetically:

Hypothesis H_i	**$\Pr(E\|H_i)$**
H_1	0.1
H_2	0.000001
H_3	1

Let us assume that we proceed with the "logical approach" and calculate

$$LR = \frac{\Pr(E|H_1)}{\Pr(E|H_2)} = \frac{0.1}{0.000001} = 100{,}000$$

which would be described as very strong support for H_1. Is this acceptable? Well, the answer is that it is only acceptable if the prior probability for H_3 is vanishingly small and if the three hypotheses exhaust all possible explanations. The approach of McCrossan et al. to hypothesis formation suggests

that all propositions for which there is a reasonable prior probability should be considered, either directly by the scientist or after the defense have made the scientist aware of such a possibility. Under these circumstances, there should be no risk of the likelihood ratio being misleading. The future may entertain a more comprehensive solution based on the general form of Bayes' theorem.

2.6.3 Errors in Analysis[1]

There have been justified complaints that most discussions, including our own, start from the premise that the typing has been completed in an error-free way.[57,201,501,509,583,618,665,765,767] Other than this brief section and a section in Chapter 8 on low copy number analysis, we will also assume that the analysis is error free.

However, there is clear evidence that errors do occur. For a brief review, see Thompson et al.[770] and the following correspondence.[179,200,771] The rate of such errors is probably low and quality assurance goes some way to reassuring the court and public that the error rate is not high. But it must be admitted that a good estimate of the rate is not available. Nor could one rate be applied fairly to different cases, different laboratories, or even different operators. There have been calls for monitoring of this rate (reviewed again in Thompson et al.; see also Chakraborty[159]). The error rate would be a very hard parameter to estimate and there are clear practical difficulties. This may have forestalled any large-scale effort to estimate this rate. A more likely explanation is the quite legitimate wish of forensic scientists that whenever an error is found, they do not want to count it; rather, they want to eliminate the possibility of its future reoccurrence. However, we endorse efforts to investigate the error rate. One reason for this is that all forensic scientists we know are honest, dedicated persons and any investigation such as this will be used primarily to improve methods.

Despite these barriers, there are modern collaborative exercises that take a very responsible approach to assessing the rate, the source of errors and that make suggestions for their reduction. Parson et al.[607] give the outcome of a very large mitochondrial DNA collaborative exercise. They report 16 errors. Ten of these errors were clerical, two were sample "mix-ups," one was assigned as contamination, and the remainder were assigned as arising from interpretational issues.

Errors can be of several types. Clearly, false exclusions and false inclusions have differing consequences. The most serious errors would be sample swapping or sample contamination. However, the most common "error" of which

[1] This section on error is provided by Christopher Triggs and John Buckleton.

we are aware is the assumption that a heterozygote is a homozygote because an allele is undetected. It is difficult to see how there could be too serious a consequence for this.

The presence of completely random contamination, from, say, plasticware, in a normal corroborative case is unlikely to lead to a false identification implicating the suspect. This type of contamination may be identified by the negative controls if the contamination is repeated. The same contamination in a database search case, if missed by the controls, could have far more serious consequences, for example implicating a worker in a plastic factory who is on the database. The risks of contamination from mortuary surfaces,[682] from scene officers,[683] and the presence of third-party DNA after simulated strangulation[681] have been discussed.

If a scene sample is contaminated with suspect DNA, then the suspect is at great risk. Forensic scientists are aware of these risks and treat them very seriously, but complacency should be rigorously opposed.

Other risks are run whenever subjective judgement is involved. This is slowly diminishing in forensic DNA work with the advent of automation but still remains in some areas. Risinger et al.[648] and Saks et al.[685] give a very well argued examination of the risks of observer effects in forensic science. Observer effects are errors in observation, recording, or decision making that are affected by the state of mind of even the most honest and diligent observer. Observers have been making this warning for some time:

> When you employ the microscope, shake off all prejudice, nor harbour any favourite opinions; for, if you do, 'tis not unlikely fancy will betray you into error, and make you see what you wish to see.[32]

A famous example is the count of human chromosomes. Early visualization techniques were rudimentary and counting was very difficult. In 1912, Hans von Winiwater reported 47 chromosomes in men and 48 in women (the Y chromosome is very small). In 1923, Theophilus Painter confirmed the count of 48 after months of indecision. This was despite his clearest views only showing 46. Despite improvements in the preparation and dyeing of chromosomes in the intervening 30 years, it was not until 1956 that Levan gave the correct count of 46. Levan was a plant biologist and did not "know" that humans had 48 chromosomes.[749]

> Men generally believe quite freely that which they want to be true.[141]

Thompson et al. argue, correctly, that such effects are widely considered in other fields of science, and protocols to deal with them are in place.[648,685,770] These include such well-known experimental methods as the double blind testing mechanism in much medical research. Why not, then, in forensic science? We recommend the Risinger et al. and Saks et al. discussion as necessary reading for all forensic scientists and recommend that it be included in

their basic training as well as the relevant sections on bias, overinterpretation, and "how much should the analyst know" in Inman and Rudin[429] (for additional comments, see also USA Today[21] and King[464]).

Other possibilities for error include deliberate or accidental tampering from external persons. The FSS reported 156 security breaches in the year ending June 2002, a 37% decrease on the previous year. The report outlined two of these as examples. They involved theft of property such as computers and credit cards rather then evidence tampering.[78]

Without a good estimate of the error rate, we are left with speculation. The error rate is clearly greater than zero. No forensic scientists would claim that it is zero. This is obviously a legitimate avenue for defense examination, and we would recommend that all prosecution witnesses should treat it as a legitimate form of examination, and should not react in a hostile or defensive manner.

We now come to the issue of combining the error rate and the match probability. This has been suggested (see again Thompson et al.[770] for a review) but never, to our knowledge, applied. If we assume that both the error rate and the match probability are known and constant, then the mathematics are trivial. Below we reproduce the common form in which this is given, but either the full Bayesian approach (Equation (2.4)) or the compromise approach (Equation (2.5)) could handle this easily by introducing a subproposition of contamination. Taroni et al.[757] discuss the problem using Bayes' nets and demonstrate this point.

We have two profiles of interest: G_c, the true type of the profile recovered at the crime scene; and G_s, that of the suspect. We will assume that the profile G_s is always determined without error.

As usual, we have two hypotheses:

H_p: The suspect is the donor of the DNA in the crime sample.
H_d: The suspect is not the donor of the DNA.

We further consider the event, $\exists$, that the profile produced in the electropherogram is not a true representation of the type of the DNA in the crime sample; that is, an error in typing has occurred.

Its complementary event, $\overline{\exists}$, is that the profile produced in the electropherogram is a true representation of the type of DNA in the crime sample. We follow Thompson et al.[770] and assume that $\exists$ and $\overline{\exists}$ are not conditional on H_p or H_d. If we write the error rate as e, then we can take

$$\Pr(\exists) = e \text{ and } \Pr(\overline{\exists}) = (1-e)$$

$$\Pr(\exists | H_p) = \Pr(\exists | H_d) = \Pr(\exists) = e$$

$$\text{and } \Pr(\overline{\exists} | H_p) = \Pr(\overline{\exists} | H_d) = \Pr(\overline{\exists}) = 1-e$$

We have the four probabilities:

$\Pr(G_c \mid G_s, \overline{\exists}, H_p)$	1
$\Pr(G_c \mid G_s, \exists, H_p)$	The probability of a false positive match given that an error has occurred, k
$\Pr(G_c \mid G_s, \overline{\exists}, H_d)$	the match probability, f
$\Pr(G_c \mid G_s, \exists, H_d)$	The probability of a false positive match given that an error has occurred, k

The likelihood ratio becomes

$$LR = \frac{\Pr(G_c \mid G_s, \overline{\exists}, H_p)\Pr(\overline{\exists} \mid H_p) + \Pr(G_c \mid G_s, \exists, H_p)\Pr(\exists \mid H_p)}{\Pr(G_c \mid G_s, \overline{\exists}, H_d)\Pr(\overline{\exists} \mid H_d) + \Pr(G_c \mid G_s, \exists, H_d)\Pr(\exists \mid H_d)}$$

Thompson et al.,[770] Weir,[829] and Buckleton and Triggs (this text) give three different formulae for this likelihood ratio:

	LR
Buckleton and Triggs	$\dfrac{1-(1-k)e}{f(1-e)+ke}$
Thompson et al.	$\dfrac{1}{f+ke(1-f)}$
Weir	$\dfrac{1-ke}{f(1-2ke)+ke}$

Thompson et al. explicitly make the approximation that, in their notation, $\Pr[R|M] = 1$, a fuller treatment could take this probability as $1 - e + ke$. The formula for the Thompson et al. likelihood ratio would then agree with the Buckleton and Triggs formula.

We see that the Thompson et al. *LR* will always exceed the Buckleton and Triggs *LR* and for fixed values of the match probability, *f*, and error rate, *e*. The value of the false positive rate *k* that maximizes this difference depends on the relative magnitude of *f* and *e*. For those cases where the error rate *e* is much greater than the match probability *f*, the difference is maximized for values of *k* close to, but greater than 0. For example, if $f = 10^{-9}$ and $e = 10^{-4}$ the maximum difference between the two values of the likelihood ratio is 0.00994% and occurs when the false positive rate $k = 0.03152$.

While accepting that Thompson et al. have made an explicit approximation, it is instructive to look at the value of the likelihood ratio under certain limiting boundary conditions. We note the peculiar results for Thompson et al. and Weir in the fifth column of Table 2.5 when there is an unrealistically high error rate, *e*.

Table 2.5 Comparison of Approaches to Incorporate Error Rate by Buckleton and Triggs (BT), Thompson et al. (TTA), and Weir (W)

	LR	Profile common f→1	No error, $e \to 0$	Error certain, $e \to 1$	False positive probability $k=0$	False positive probability, $k=1$
BT	$\frac{1-(1-k)e}{f(1-e)+ke}$	1	$\frac{1}{f}$	1	$\frac{1}{f}$	$\frac{1}{f(1-e)+e}$
TTA	$\frac{1}{f+ke(1-f)}$	1	$\frac{1}{f}$	$\frac{1}{f(1-k)+k}$	$\frac{1}{f}$	$\frac{1}{f(1-e)+e}$
W	$\frac{1-ke}{f(1-2ke)+ke}$	1	$\frac{1}{f}$	$\frac{1-k}{f(1-2k)+k}$	$\frac{1}{f}$	$\frac{1-e}{f(1-2e)+e}$

To make any exploration of the likelihood ratio $(1-(1-k)e)/(f(1-e)+ke)$, we need to postulate an error rate. If this is larger than the match probability, then it will completely dominate the equation and hence $LR \approx 1 / ke$

This shows that the error rate and the match probability can be mathematically combined. But should they be? The arguments for and against have occurred in the literature (reviewed in Thompson et al. and indeed in court, e.g., *Regina v Galli*[637]). Those making the "for" argument would comment, correctly, that the jury may not be able to weigh the respective contributions of error rate and match probability. Those supporting the "against" argument would argue that an error rate is not known and hence the equation is not implementable. The error rate relates to many things. The arguments given above are phrased largely in the context of a single reference sample and a single stain sample. In many cases, there are multiple samples collected and perhaps typed at differing times. All of this would affect the probability of an error and that subset of errors that represent false inclusions. Lynch[526] makes the interesting point that eyewitness evidence is known to be fallible. Juries have been asked to evaluate this "eyewitness" risk on a case-by-case basis for a long time and no explicit combination is made of the error rate with the weight of evidence. Of course, eyewitness evidence is not presented numerically at all and this may be a fundamental difference.[526]

Our view is that the possibility of error should be examined by the judge and jury on a per case basis and is always a legitimate defense explanation for the DNA result. The two possible hypotheses that are consistent with "innocence" should be explored in court. This argument however does not answer the complaint that the jury may be unable to weigh the two hypotheses consistent with innocence (one numerical and the other not) and may give undue weight to the match probability.

Let us assume that we intend to develop a case-specific estimate of the probability of an error as Thompson et al. suggest, following Thompson:[768]

> … it makes little sense to present a single number derived from proficiency tests as THE error rate in every case, … I suggest that it be evaluated case-by-case according to the adequacy of its scientific foundation and its helpfulness to the jury.

Even assuming that this case-specific error rate is accurately estimated, there still is an objection to the combination of the probability of an error and that of a coincidental match. The likelihood ratio is uncertain in all cases because it is based on estimates and models. It is normal to represent this type of uncertainty as a probability distribution. If we add the possibility of error, then this distribution has a point mass at 1 and a continuous distribution around high values for the *LR*. In Figure 2.1 we give a hypothetical distribution of this sort. The Thompson et al. equation suggests we report the *LR* signified by the arrow. This value is in the void between the two modes, in a region where there is no density, and may be viewed by many as a very poor summary of the distribution. However, the large mode at the right of the figure, if reported without mention of error, could also be viewed as an equally poor summary.

The innocent man who has been implicated by an error or indeed by a coincidental match is at great risk of a false conviction and it is generally accepted that a false conviction is the worst outcome that can occur in the judicial system. The Thompson et al. formula, if applied, may very occasion-

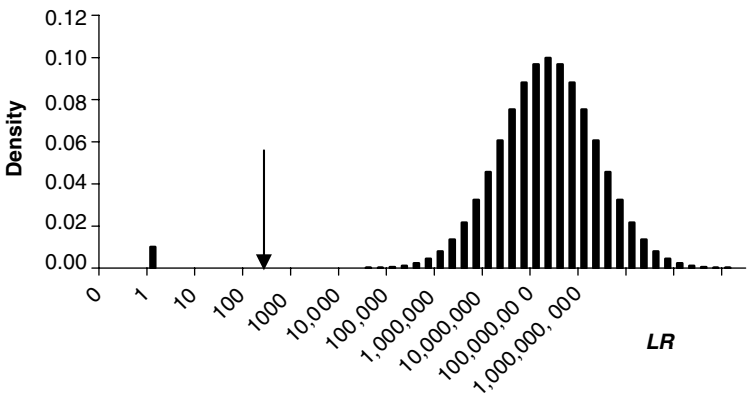


Figure 2.1 Hypothetical distribution for *LR*. The small mode at 1 represents the *LR* if an error has occurred. The larger mode centered at about 1,000,000,000 represents the *LR* if no error has occurred. The Thompson et al. equation would suggest that we report a value somewhere in the area signified by the arrow.

ally help such a man. In effect, the formula reduces the likelihood ratio and it may do so to the point where a jury will not convict on the DNA evidence alone. The reality, in our view, is that most often the wrongly implicated man will almost have to prove his innocence by establishing that an error HAS happened (it is difficult to see how unless alternative uncontaminated samples are available) or to produce very strong alibi evidence. Unless the wrongly accused man can produce considerable evidence in his favor, it is possible or even likely that he will be convicted. However, there is very little that statistics can do to help him. The reduction of the likelihood ratio affects both the correctly accused and the wrongly accused equally. We suspect that it is of some, but probably inadequate, help to the wrongly accused man and a false benefit to the correctly accused. The answer lies, in our mind, in a rational examination of errors and the constant search to eliminate them. The forensic community would almost universally agree with this.

Findlay and Grix[299] make the reasonable point that the very respect given to DNA evidence by juries places an obligation on scientists to maintain the highest standards and to honestly explain the limitations of the science in court.

It is appropriate to end this section with an appeal for higher standards of the already considerable impartiality in forensic laboratories. We recommend that all forensic scientists read the account by the father of the victim of a miscarriage caused by wrongful fingerprint evidence[541] or the call for standards by Forrest in his review of the Sally Clark case.[316] Most forensic scientists aspire to a position of impartiality but unconscious effects must be constantly opposed. In our view, language is one tool that can be utilized. The words "suspect" and "offender" have specific meanings but are often used interchangeably. In our view, both should be avoided. Both have too many emotional associations: Would you buy a "suspect" car? The preferable term is Mr. or Ms. We also object to the placing of the "suspect's" name in capitals as required by the procedures in some laboratories such as our own. Why is it "Detective Smith" but the suspect is termed "Mr. JONES?" All emotive terms or terms with unnecessary implications should be avoided.

The matter is one of culture. Everyone in a laboratory needs to cooperate in developing the culture of impartiality. People lose their life or liberty based on our testimony and this is a considerable responsibility.

2.6.4 The Absence of Evidence

Special attention is given in this section to interpreting the "absence of evidence." This is largely because of a widespread misunderstanding of the subject despite excellent writing on the matter (see, for instance, Inman and Rudin[429]). This misunderstanding has been fostered by the clever but false saying:

The absence of evidence is not evidence of absence.

We assume a situation where some evidence has been searched for but not found. Call this event $\bar{E}$. Bayes', theorem quickly gives us a correct way to interpret this evidence.

$$LR = \frac{\Pr(\bar{E}|H_p)}{\Pr(\bar{E}|H_d)}$$

The issue then is simply one of estimating whether the finding of no evidence was more or less likely under H_p than H_d. Unless some very special circumstances pertain, then the finding of no evidence will be more probable under H_d, and hence the absence of evidence supports H_d. Often, in real casework, this is only weak support for the hypothesis, H_d.

The special circumstances that could pertain would be those that made no evidence very likely under H_p but the finding of evidence likely under H_d. Situations involving such circumstances take a little bit of thinking to suggest.

This (correct) mathematical argument is not accepted by many forensic scientists and lawyers, but is universally accepted by interpretation specialists. The counter argument is that one can often think of an explanation for the absence of evidence. For instance, let us imagine that a fight has occurred where one person was stabbed and bled extensively. A suspect is found and no blood is found on his clothing. How is this to be interpreted? Many forensic scientists will observe that the suspect may have changed clothes, washed his clothes, or contact may have been slight in the first place. These observations are correct, but are more along the lines of explanations of the (lack of) evidence. It is better to look at this problem from the point of view of propositions. What was the probability that the suspect would have blood on him *if* he were the offender? Let us imagine that we do not know whether or not the suspect has changed or washed his clothes. Further, let us imagine that we have some information about the fight, but that this is inexact or unreliable. From this we must accept that it is uncertain whether we expect to find blood on the clothing or not, even if the suspect is, indeed, the offender. However, we must feel that this probability is not zero. There must have been some probability that we would have found blood on the clothing; why else were we searching for it? Only if this probability is essentially zero is the evidence inconclusive. Otherwise, if this probability is in any real way larger than zero, it will be larger than the probability if the suspect is not the offender, and hence the evidence will support the defense hypothesis.

Clearly this area is not well understood, nor is there widespread agreement. Further discussion in the literature would be most welcome. Research on transfer and persistence of evidence is also seen to be of great importance.

2.7 Summary

This chapter has reviewed options for a framework for interpretation. Subsequent chapters will focus on details of DNA interpretation. It is, however, very important to understand this fundamental structure for interpretation before proceeding to detailed analysis.

Additional reading: Inman and Rudin[429] give an elegant discussion of many aspects of evidence interpretation. This book would serve very well as part of all training courses in forensic science.

Robertson and Vignaux[651,656] consider both the legal concerns regarding this type of analysis, and more specifically the situation where the evidence itself is both multiple and uncertain. This is a level of complexity above and beyond anything considered in this chapter. They also introduce the useful concept of Bayesian networks that are being extensively researched as a method for forensic interpretation.

Population Genetic Models

3

JOHN BUCKLETON

Contents

3.1 Introduction

This chapter will discuss those population genetic models used for assigning a profile or preferably a match probability. Three models — the product rule, the subpopulation correction, and an admixture model — will be discussed.

The interpretation of DNA evidence often requires the assignment of a probability to the chance of observing a second copy of a particular genotype in a certain population. Implicit in this apparently simple statement, there are many questions about what this probability should be, how it should be assessed, and upon what other information, if any, should it be conditioned.

In common usage, the word "frequency" is often substituted for "probability." Hence a genotype probability will become a genotype frequency. This is a slight loss in accuracy in the use of nomenclature, but it allows us to slip into common usage. A frequency really should have a numerator and a denominator, e.g. 3 in 25, where we have counted 3 particular outcomes out of the 25 possible. Since most genotype probabilities are very small, they are not estimated by direct counting. Hence, strictly, they are not frequencies.

The frequentist approach to interpreting evidence will report this genotype frequency, f.

Under the logical approach for these hypotheses:

H_p: The DNA came from the suspect, and
H_d: The DNA came from a male not related to the suspect,
the likelihood ratio

$$LR = \frac{1}{\Pr(G_c|G_s,H_d)} = \frac{1}{f} \tag{2.3}$$

The standard response to our inability to directly assess these frequencies has been to attempt to model them using a population genetic model. However, certain cautions should be considered with the concept of a true genotype probability. First among these cautions is to consider what would represent a "fair and reasonable" assignment of probability. It would be tempting to suggest that a fair and reasonable assignment would be one that was near the true value. If we consider the values of probabilities that will be generated by 13-locus CODIS or 10-locus SGM$^+$ multiplexes, we realize that they are very small. It would be very difficult, if not impossible, for us to determine their true values. In fact, this would typically require the genetic typing of the whole population of the world, and the values would change constantly as individuals were born or died.

Is there a requirement for the fair and reasonable probability assignment to equal the true value? Interestingly, the answer is no. If we consider the profile probabilities of 13-locus CODIS profiles (or 10-locus SGM$^+$ profiles), it is certain that most genotypes do not exist. There are more possible genotypes (about 10^{23}) at these 13 loci than there are people. Therefore, only about 1 profile in 10^{14} can exist.[842] True profile frequencies will take the values

$$\frac{n}{6{,}000{,}000{,}000}$$

where $n=0, 1, 2\ldots$ and the population of the world at a given instant is taken for illustration as 6 billion.

Most genotypes will thus have true frequencies of 0. These are of no interest to us because they do not exist and will not occur in casework. It is the remaining ones that are of interest. For those that do exist we know that the suspect has this genotype, but we must remember that we are interested in the probability of obtaining this genotype from someone other than the suspect.

All our probability assignments will differ from the true frequencies. Even if we move to the superior conditional probabilities, these will typically be small numbers whereas the actual frequencies are 0, 1, 2, or more in 6 billion. We distinguish between the actual frequency of a genotype and its probability. The frequency of a genotype will be a probability only if we could conceive of carrying out an experiment of randomly sampling, with replacement, individuals chosen from the population of the world at a given instant.

The assignment of a probability to a multilocus genotype is an unusual activity. Few other fields of science require such a probability assignment. The field of genetics is well established, but largely concerns itself with things such as allele probabilities or genotype probabilities at one or a very few loci. Therefore, the attempt by forensic scientists to assign probabilities to multilocus genotypes is a relatively novel experiment peculiar to forensic science. It may be based on genetics and statistics, but it is a new extension of previous methods, broadly speaking attempting to go where no science has gone before.

These probabilities cannot be directly measured by any mechanism that we can envisage. Ian Evett has discussed his view of whether these probabilities can be considered estimates at all:

> Probability is a personal statement of uncertainty. In the DNA context, I take some numbers (that are estimates of things like allele proportions and F_{ST}'s) and stick them into a formula. Out comes a number and on the basis of that I assign… a probability.

> That is a personal, subjective probability, which incorporates a set of beliefs with regard to the reliability/robustness of the underlying model. So, whenever you talk about estimating a probability, I would talk about assigning a probability.
>
> Thus I would not say, as you do… that the probabilities are "untestable estimates." I would ask — "is it rational for me to assign such a small match probability?"

We cannot directly compare our probability assignments to true values. We may be able to test the process by which these probabilities are assigned, but in casework we will be unable to test the final probability assignment. This makes it most important that these inherently untestable probabilities are assigned by the most robust methods.

In this chapter, the options currently in use to assign these genotype probabilities are discussed. In addition, we consider a third option that has been suggested by Bonnie Law. This model was designed to cope with the phenomenon of admixture.

3.2 Product Rule

This is the simplest of the available population genetic models. It is deterministic as opposed to stochastic.[211] This means that it assumes that the populations are large enough that random effects can be ignored. It was the first model implemented in forensic DNA analysis, having previously been used for a number of years in blood group analysis. It is based on the Hardy–Weinberg law and the concept of linkage equilibrium.[805,806] Both these concepts have been extensively discussed. However, it is worthwhile making a few comments that are specifically relevant to forensic science.

3.2.1 Hardy–Weinberg Law

This concept was first published in 1908,[392,826] although simplified versions had been published previously.[151,611,878] This thinking developed naturally following the rediscovery of Mendel's work.[546] It concerns the relationship between allele probabilities and genotype probabilities at one locus. In essence, the Hardy–Weinberg law is a statement of independence between alleles at one locus.

The Hardy–Weinberg law states that the single-locus genotype frequency may be assigned as the product of, allele probabilities

$$P_i = \begin{cases} p_{i1}^2, & A_{i1} = A_{i2} \\ 2p_{i1}p_{i2}, & A_{i1} \neq A_{i2} \end{cases} \tag{3.1}$$

for alleles A_{i1}, A_{i2} at locus *i*. This will be familiar to most in the form

$$\begin{cases} p^2 & \text{homozygotes} \\ 2pq & \text{heterozygotes} \end{cases}$$

This law will be exactly true in all generations after the first if a number of assumptions are met. It may also be true or approximately true under some circumstances if these assumptions are not met. The fact that the equilibrium genotype frequencies are obtained after one generation of random mating means that we do not need to enquire into the deep history of a population to describe the genotype frequencies at one locus[211] if these requirements are met. It also means that any perturbation from equilibrium is likely to be rectified rapidly. This is not exactly true for populations with overlapping generations, such as humans, where equilibrium is achieved asymptotically as the parental population dies. A few other exceptions to the rule that equilibrium is achieved in one generation are given in standard population genetic texts such as Crow and Kimura.[211]

The assumptions that make the Hardy–Weinberg law true are that the population is infinite, randomly mating, and there are no disturbing forces. Inherent in this law is the assumption of independence between genotypes: specifically, that the knowledge of the genotype of one member of a mating pair gives no information about the genotype of the other. Consider what would happen if the population was finite, as indeed all populations must be. The knowledge of the genotype of one member of a mating pair slightly reduces the probabilities for these alleles in the other member, since one or two copies of these alleles have been "used up." This effect is very minor indeed unless the population is quite small or the locus extremely polymorphic. Most human populations may be numbered in tens of thousands or more individuals.

The assumption of random mating assumes that the method of selection of mates does not induce dependence between genotypes. This is often translated comically and falsely along the lines "I did not ask my spouse his/her genotype before I proposed." When the assumption of random mating is questioned, no one is suggesting that people who are genotype *ab* deliberately go and seek partners who are type *cd*. What is suggested is that geography, religion, or some other socioeconomic factors induce dependence. This will be discussed later, but the most obvious potential factor is that the population is, or more importantly has been in the past, divided into groups that breed more within themselves than with other groups.

A consequence of the assumption of an infinite population and random mating is that the allele proportions are expected to remain constant from one generation to the next. If the population is infinite, randomly mating, and the allele proportions do not change, then the Hardy–Weinberg law will

hold in all generations after the first. This is true whether or not the Hardy–Weinberg law holds in the first generation, the parental one. It therefore describes an equilibrium situation that is maintained indefinitely after the first generation. Note that it does take one generation of random mating to achieve this state. Such a stable state would describe an equilibrium situation and hence this state is often called Hardy–Weinberg equilibrium (HWE).

There are, however, a number of factors that can change allele proportions. These are referred to as disturbing forces. The term is derived from the fact that they change genotype proportions from those postulated by HWE. These factors include selection, migration, and mutation. There are comprehensive texts available describing the effect of these forces on both allele proportions and on HWE, and they will not be discussed at length here. In this chapter we will simply consider how close the Hardy–Weinberg assumptions are to being fulfilled, and what the probable consequences of any failure of these assumptions may be. Remember a model may be useful even though it is not an exact description of the real world.

3.2.2 Linkage and Linkage Equilibrium

HWE describes a state of independence between alleles at one locus. Linkage equilibrium describes a state of independence between alleles at different loci.

The same set of assumptions that gives rise to HWE plus an additional requirement that an infinite number of generations has elapsed also lead to linkage equilibrium. This result was generalized to three loci by Geiringer,[331] and more generally to any number of loci by Bennett.[54]

However, recall that HWE is achieved in one generation of random mating. Linkage equilibrium is not achieved as quickly. Strictly the state of equilibrium is approached asymptotically, but is not achieved until an infinite number of generations have elapsed. However, the distance from equilibrium is halved with every generation of random mating for unlinked loci or by a factor of $1-r$, where r is the recombination fraction, for linked loci. Population subdivision slows this process.[421]

It is worthwhile discussing the difference between linkage equilibrium and linkage, as there is an element of confusion about this subject among forensic scientists. Linkage is a genetic phenomenon and describes the situation where one of Mendel's laws breaks down. It was discovered in 1911 by Morgan[555,556] working on *Drosophila*. The discovery was a by-product of his team's studies of inheritance that had largely led to the confirmation of the chromosomal theory of inheritance. The first paper on gene mapping appeared in 1913.[740]

Specifically, the phenomenon of linkage describes when alleles are not passed independently to the next generation. The physical reason for this phenomenon had been identified by 1911 and related to the nonindependent segregation of alleles that are sufficiently close on the same chromosome.[597]

The state of linkage can be described by the recombination fraction or by the distance between two loci. Typical data for distance may be expressed in centiMorgans (cM) or in physical distance in bases. In humans, 1 cM is assumed to equal approximately 1000 kb.

The physical distance may be converted to a recombination fraction by standard formulae.[a] Recombination fractions tend to be different for each sex. Distances may be given separately or sex-averaged.

Linkage disequilibrium is a state describing the relationship between alleles at different loci. It is worthwhile pointing out that linkage disequilibrium can be caused by linkage or by other population genetic effects such as population subdivision. This will be demonstrated later.

Therefore, it is incorrect to advance the following line of logic.

A: The loci are on different chromosomes or well separated on the same chromosome.

Which implies that

B: There is no linkage.

Which implies that

C: There is no linkage disequilibrium.

Modern genetic understanding would state that the progression from statement A to statement B is logical and grounded on experimental observation. However, the progression from statement B to statement C is not supportable without additional data.

Linkage disequilibrium has been noted for very closely linked loci. For example, Gordon et al.[366] investigated 91 unrelated Afrikaners and observed linkage disequilibrium between pairs of loci separated by 0.00, 0.00, 0.54, 2.16, 2.71, 3.68, 5.28, and 5.51 cM on chromosomes 1, 2, 5, 11, 20, and 21. Such linkage disequilibria have been used to estimate the time since the last bottleneck for various populations[522] and may give interesting anthropological information. Deka et al.[227] investigated linkage disequilibrium and identified Samoans as an interesting study group plausibly because of a recent bottleneck. Szibor et al.[750] investigated linkage disequilibrium between alleles at loci on the X chromosome for a sample of 210 males. The loci investigated contained three linkage groups from a total of 16 loci. They observed disequilibrium only for alleles at the loci DXS101 and DXS7424. This is an example of the well-known phenomenon that linkage does not necessarily imply linkage disequilibrium.

[a] See Chapter 1, footnote c.

The CODIS loci HUMCSF1PO and HUMD5S818 are both located on chromosome 5 and are reported to be separated by 25 cM.[30] This translates to a recombination fraction (Haldane) of 0.39. This would be expected to have no effect at the population level, but in restricted circumstances may have a moderate effect in paternity testing or disaster victim identification.

The most likely causes of linkage disequilibrium for unlinked or loosely linked loci are population genetic effects such as population subdivision or admixture.[154,421] These will be discussed in some detail later.

If the population is in linkage equilibrium, then a multilocus genotype probability (P) may be assigned by the product of single-locus genotype probabilities (P_i):

$$P = \prod_i P_i \tag{3.2}$$

3.2.3 Consideration of the Hardy–Weinberg and Linkage Equilibrium Assumptions

There are five assumptions for the Hardy–Weinberg law to hold and one additional assumption for linkage equilibrium to hold. In this section each of these assumptions will be considered with regard to whether or not they are true, and in particular to how far from true they may be.

3.2.3.1 Infinite Population

This assumption is clearly violated to greater or lesser extents, depending on the size of the population. In addition, there is ample evidence for the existence of population bottlenecks in the past. The effect on disturbing the equilibrium in the present is likely to be very limited for most realistic populations unless a relatively recent bottleneck is suspected. Recall that one generation of random mating is sufficient to restore HWE. Any effect is most likely to occur for rare alleles.

Crow and Kimura[211] give

$$\Pr(A_i A_i) = p_i^2 - p_i(1 - p_i) f$$
$$\Pr(A_i A_j) = 2p_i p_j(1 + f)$$

where N is the number of individuals and $f = 1/(2N - 1)$ We see that any departure from equilibrium is expected to be very small for most realistic values of N.

3.2.3.2 No Mutation

One of the assumptions for Hardy–Weinberg and linkage equilibrium is that there is no mutation at the loci in question. With regard to the commonly used STR loci, this assumption is clearly violated. In fact, we believe

that the STR loci are mutational "hot spots," with mutation rates above much of the coding DNA but probably less than the VNTR loci or mitochondrial DNA.

Various treatments have been offered that deal with change in allele frequencies due to mutation or to the effects of mutation and selection.[267] If, however, we accept that these loci are selectively neutral, then the most realistic situation that we need to consider is the situation of mutation and genetic drift. The effect of mutation, of the type observed at STR loci, on a divided population is that it tends to oppose the effect of drift. If drift is tending to remove genetic variation from separated subpopulations, mutation tends to reintroduce it. When a mutation occurs at an STR locus, it tends to add or subtract a single repeat, with mutational losses or gains of multiple repeats being much more rare (see Chapter 10 for a summary of mutation references). This mode of mutation fits well with a theoretical model, the stepwise mutation model, that was first proposed by Kimura and Ohta.[462]

If we consider two populations that have become separated or isolated, then they begin to evolve separately and their respective allelic frequencies tend to drift apart. This process will be associated with an increase in relatedness within the separated subpopulations and can be quantified by an increase in the inbreeding coefficient θ. The effect of stepwise mutation to alleles already present is to lower relatedness and hence θ.[285,671,672] This may seem odd. The people are still related, but their alleles can no longer be identical by descent as they are no longer identical. The equilibrium situation that may result is given by Evett and Weir.[267] Whether drift or mutation is the dominant factor depends on the product $N\mu$, where N is the population size and μ the mutation rate. If $N\mu \ll 1$, the population will typically be moving toward fixation for one allele, which means that genetic drift forces are dominant. If $N\mu \gg 1$, then mutation is the dominant force and multiple alleles will be present.[577]

This effect can be elegantly demonstrated using simulation software. Two programs have been offered by forensic programmers — Gendrift (Steve Knight and Richard Pinchin, FSS) or Popgen (James Curran, University of Waikato[b]) — and there are others in the population genetics community.

It would be unwise, however, to assume that mutation is a completely benign phenomenon from the perspective of decreasing associations between individuals. The exact nature of the mutational process does have a serious effect on the departures that may be observed and the validity of models to correct for them. This is discussed briefly later.

[b] The latter program is available free from James Curran's website: http://www.stats.waikato.ac.nz/Staff/curran/Popgen95.zip.

3.2.3.3 No Migration Into or Away from the Population

Allele probabilities will change if migration occurs into or away from the population. Emigration from a moderately sized population has very little effect since the subtraction of a few alleles from the gene pool alters the allele probabilities very little. Immigration of alleles into the population from a different population can have a much more marked effect. Such gene migration is often accompanied by physical migration of people, but this is not necessarily a requirement.

To consider this issue, it is critical to consider the interaction of migration and our definition of population. Most of our current definitions of population have both an ethnic and a geographical basis. Consider the New Zealand population. We currently subdivide this arbitrarily into Caucasian, Eastern Polynesian (Maori and Cook Island Maori), Western Polynesians (Samoans and Tongans), and Asians. The physical migration of a British person to New Zealand would represent migration of alleles into the New Zealand Caucasian gene pool. The intermarriage of Caucasians and Maori would represent migration of Caucasian genes into the Eastern Polynesian gene pool without necessarily involving any physical migration of people. The fact that this is treated as a migration of genes INTO the Eastern Polynesian gene pool is dependent on how we intend to (arbitrarily) define the ethnicity of the resulting progeny.

The effect of migration on equilibrium is dependent on the difference in allele frequencies between the donor and recipient populations.[267] Hence the physical migration of British people to New Zealand is likely to have a very small effect on the equilibrium situation of New Zealand Caucasians since the allele frequencies in the two populations are similar. However, the migration of Caucasian genes into the Eastern Polynesian gene pool is much more likely to disturb the equilibrium since the populations have more differing allele probabilities.

3.2.3.4 No Selection

It is difficult to find experimental data that bear directly on the issue of whether or not there is selection at the STR loci used for forensic work. This is clearly an area that warrants further scrutiny. The general field is very active in human population genetics. At this stage, most of the argument in favor of there being little or no selection at STR loci relates to the fact that these loci are noncoding and hence do not produce any gene products. Theoretically then, any mechanism for selection would have to operate by an indirect route, say by hitchhiking on other advantageous or disadvantageous genes, or by affecting DNA packing, replication, or repair.

The STR loci are intronic. Introns are thought to have "invaded eukaryotic genes late in evolution, after the separation of transcription and translation."[538,539] When first studied, these DNA sections were thought to be

nonfunctional and were termed "junk DNA." Mattick[538,539] argues convincingly for a role for at least some intronic products in gene expression and postulates that they were a critical step in the development of multicellular organisms.

Makalowski[532] discusses the origin of the phrase "junk DNA" and reinforces the modern conception that this DNA may have important functions. If this theory is eventually accepted, as would seem inevitable, then the question would arise as to whether there is a function for the specific intronic segments used in forensic work.[149]

The observation of greater microsatellite diversity among Africans[448] is consistent with the out of Africa event and a selectively neutral model. However, greater diversity among Africans is certainly not proof of selective neutrality.

Mitochondrial DNA shows a deviation from selective neutrality; however, this is postulated to be the result of a selective sweep in modern humans outside Africa.

Selection is a proven phenomenon in some blood group systems such as ABO and Rhesus.[152] A mechanism has been proposed for the selective interaction between ABO and Haptoglobin.[567] However, these genes are clearly coding and produce important gene products. Hence direct selective mechanisms are expected.

Selection by association with disease loci is a mechanism that may possibly affect STR loci. Such associations at other loci are known.[596] The effect of a selective sweep caused by the appearance of an allele favored by selection at a nonforensic locus has not been considered in detail in the forensic literature. However, unless such a sweep is recent, this is unlikely to have much effect on the modern state of equilibrium (although it may have had an effect on modern allele probabilities).

Neuhauser[577] compares random drift and selection and notes that if $Ns \ll 1$, where N is the population size and s is the selective advantage of one allele over another, for a two-allele locus, then selection does not have much effect, and the locus acts almost as if it were neutral.

A theoretical model for estimating mutation rates at di-, tri-, and tetranucleotides from the distributions of their allele sizes was given by Chakraborty et al.,[166] who note the departure of the predictions of the model from directly observed values. This led Chakraborty et al. to an interesting discussion of whether there is any evidence of constraints in the number of DNA repeats at a locus, which may be evidence for the existence of selection. They conclude that the shape of modern allele distributions is inconsistent with the existence of constraints.

In summary, there are reasonable theoretical reasons to believe that these loci are selectively neutral or nearly so. No direct evidence for strong selection at forensic loci has been reported, but how hard have we looked for it? Equally, there is little direct experimental evidence for selective neutrality.

3.2.3.5 *Random Mating*

Of the various assumptions given, this is the one that has deservedly attracted the most attention. It is clear that we do not select our mates on the basis of their DNA genotypes at the STR loci. Most of us do not even know our own genotype at these loci. We also believe that these genotypes have no physical manifestation, which is to say that they do not affect the phenotype of an individual. Hence we should be unable to detect these genotypes by looking at a person. This should preclude some inadvertent selection of genotypes. However, it would be wrong to assume from this that random mating is a fair assumption.

Crow and Kimura[211] discuss the two main types of nonrandom mating: inbreeding and assortative mating. Assortative mating is not discussed here. There is considerable evidence that it does occur in humans. For instance, an intelligent person is more likely to marry another intelligent person. Jared Diamond[231] discusses this in some detail in his popular science book *The Rise and Fall of the Third Chimpanzee*. In the STR context, I believe that the issue of importance is inbreeding.

What is alleged is that the population is made up of subpopulations[510,511] whose members preferentially mate within their subpopulation, possibly for religious, language, or other reasons, but more probably just because of geographical proximity (for an excellent review, see Excoffier[283]). This is termed inbreeding. In the past, people traveled a lot less than they do now. The notion of marrying the "girl or boy next door" is not universal nor is it totally unknown. It is important to note that there is no suggestion that subpopulations are completely isolated from each other. All that is required is any departure from a completely random choice of mates. The more isolated the subpopulations, the larger the effect, but partial isolation will also lead to some subpopulation effects.

In lectures on DNA around the world, I have performed a trial with the various classes. Unfortunately I have not kept the results, which would make an interesting section. However, the general flavor of them can be reported. What was asked was for people to give the "ethnicity" of their four grandparents. Table 3.1 gives the results for the area around my desk at the laboratory at the FSS at Trident Court in Birmingham, U.K. Each cell represents one individual's self-declared ethnicity for their four grandparents.

This experiment would not meet minimum survey standards; however, let us treat them as a demonstration rather than as evidence. First let us note that this arrangement does not look random. Too many ethnicities occur together. For instance, there are four Chinese entries and four Indian entries together. Let us assume that we separated these two individuals out as being of a different "race." What we are left with still does not look like a random arrangement. For instance, there are four Greek Cypriots and two Iraqis together. Let us assume further that we take these out and put them into different categories.

Table 3.1 Self-Declared Ethnicity of Some Staff at the FSS Laboratory, Trident Court in 2002

Irish, Irish, Irish, Irish	Swiss, Swiss, Swiss, Swiss
English, English, English Irish	English, English, English English
English, English, English, English	Chinese, Chinese, Chinese, Chinese
Welsh, English, English, Scottish	English, English, English, English
Scottish, Scottish, English, English	English, English, Irish, Scottish
English, English, English, English	English, English, English, Scottish
Hungarian, Scottish, Scottish, English	English, English, English, Scottish
English, English, English, English	Greek Cypriot, Greek Cypriot, Greek Cypriot, Greek Cypriot
English, English, English, English	Irish, Irish, Iraqi, Iraqi
English, English, English, Scottish	Indian, Indian, Indian, Indian

Still, what we are left with does not look random. There are too many Irish and Swiss together. If we could peer deeper into the past, we might find that the people reporting "English" have differing amounts of Celtic, Scandinavian, or Saxon heritage.

This experiment has worked wherever I have tried it: in New Zealand, Australia, the United States of America, and the United Kingdoms of Great Britain and Northern Ireland. I, personally, do not believe that the modern human population is the result of random mating. I do believe that we are the result of an evolutionary process whereby our ancestors mated in groups to a greater or lesser extent. This is breaking down in modern times, but the process is far from complete.

This leads us to the classical consideration of the Wahlund principle.[801] Assume that a certain area is made up of two or more subgroups that breed within each group but not to any large extent between the two groups. Further assume that there are some allele probability differences between these groups. Then even if the subpopulations themselves are in HWE, the full population will not be. An example is given in Table 3.2.

First we note that the mixed population is not in HWE even though each subpopulation is. Next we note the classical Wahlund effect in which all the probabilities for homozygotes are increased above Hardy–Weinberg expectation. The total heterozygote probabilities are generally decreased, although individual heterozygotes may be above or below expectation. Note that in this example two of the heterozygotes are below expectation, whereas one is above. The total for all the heterozygotes will always be down (which is really the same as saying the total of the homozygotes is always up).[267,836]

The same subpopulation phenomenon will induce between locus dependence, that is, it will induce linkage disequilibrium. This is more complex but not harder to demonstrate. In Table 3.3 we give a numerical demonstration.

Table 3.2 An Example of the Wahlund Effect

Allele	*a*	*b*	*c*
Subpopulation 1	0.7	0.2	0.1
Subpopulation 2	0.2	0.1	0.7

Genotype	Subpopulation 1	Subpopulation 2	1:1 Mix	Hardy–Weinberg expectation
aa	0.49	0.04	0.2650	0.2025
bb	0.04	0.01	0.0250	0.0225
cc	0.01	0.49	0.2500	0.1600
ab	0.28	0.04	0.1600	0.1350
ac	0.14	0.28	0.2100	0.3600
bc	0.04	0.14	0.0900	0.1200

This table shows the "correct" genotype proportions and two incorrect calculations. The first incorrect calculation proceeds by combining the two subpopulations and then using the population allele probabilities — this incorrectly assumes Hardy–Weinberg and linkage equilibrium in the population. This is the type of error (although greatly exaggerated) that would occur if we assumed that a structured population was homogeneous. The second incorrect calculation (again carried out on the combined population) proceeds as if we had performed some sort of testing and had abandoned the assumption of HWE, but instead had used observed genotype proportions and then multiplied across loci. This approach is a better method to assign probabilities as it corrects for Hardy–Weinberg disequilibrium; however, it fails to account for linkage disequilibrium.

The third approach was adopted, incorrectly, by Buckleton and Weir in some of their early recommendations, but is now abandoned. It appears later in this chapter as the "Cellmark wrinkle" in the descriptions of the O.J. Simpson case. It persists in recommendations by other authors but should be superseded.

Inspection of these numbers shows that the "correct" probabilities for two loci cannot be determined if the population structure is ignored. Proceeding from either the population allele probabilities or the population genotype probabilities will give incorrect answers.

The demonstration that the multiplication of population genotype probabilities gives an incorrect answer shows that linkage disequilibrium can be induced by population substructure whether or not the loci are physically linked. Loci that are on different chromosomes may, therefore, be in disequilibrium[514,576,590,591] and expressions have been derived to estimate the magnitude of the disequilibrium.[267,836] In fact, almost any instance of disequilibrium in the forensic literature involves loci that are on different chromosomes.

Table 3.3 Two-Locus Genotype Probabilities for a Population Consisting of Two Subpopulations in Equal Proportions

Allele	Subpopulation 1	Subpopulation 2
Locus 1		
a	0.7	0.2
b	0.2	0.1
c	0.1	0.7
Locus 2		
d	0.5	0.2
e	0.2	0.4
f	0.3	0.4

	dd	*ee*	*ff*	*de*	*df*	*ef*
	1:1 Mix Correct					
aa	0.062	0.013	0.025	0.052	0.077	0.036
bb	0.005	0.002	0.003	0.005	0.007	0.004
cc	0.011	0.039	0.040	0.040	0.041	0.079
ab	0.036	0.009	0.016	0.031	0.045	0.023
ac	0.023	0.025	0.029	0.036	0.043	0.053
bc	0.008	0.012	0.013	0.015	0.017	0.025
	1:1 Mix from Alleles					
aa	0.025	0.018	0.025	0.043	0.050	0.043
bb	0.003	0.002	0.003	0.005	0.006	0.005
cc	0.020	0.014	0.020	0.034	0.039	0.034
ab	0.017	0.012	0.017	0.028	0.033	0.028
ac	0.044	0.032	0.044	0.076	0.088	0.076
bc	0.015	0.011	0.015	0.025	0.029	0.025
	1:1 Mix from Genotypes					
aa	0.038	0.027	0.033	0.048	0.061	0.058
bb	0.004	0.003	0.003	0.005	0.006	0.006
cc	0.036	0.025	0.031	0.045	0.058	0.055
ab	0.023	0.016	0.020	0.029	0.037	0.035
ac	0.030	0.021	0.026	0.038	0.048	0.046
bc	0.013	0.009	0.011	0.016	0.021	0.020

Some of the most common causes of disequilibrium are population genetic effects, such as the existence of subpopulations, and such disequilibria occur for the same reasons as the Wahlund effect.[484,485]

This disequilibrium phenomenon is sufficiently understood that decay rates for linkage disequilibrium for nonlinked loci have been calculated and appear in standard texts.[267 (pp. 127–129),421,836] The dependency effects are not expected to be large for loci with low mutation rates. There is a slight tendency for the dependencies to rise with the number of loci.[488,843]

We give examples later using the ESR data for Eastern Polynesians. Analysis of these data suggests disequilibrium regardless of the chromosomal position of the loci. In this particular case, the most likely explanation is not population subdivision but the effects of admixture with Caucasians. The population in the U.S. described as Hispanics may be showing the same admixture effects or this may be the result of subpopulations, or both.[c] The Hispanic population is often subdivided into South-Eastern and South-Western Hispanic.

Conversely, loci that are closely linked on the same chromosome may be in equilibrium (or near it). In fact, there is no absolute relationship between the position on a chromosome and the state of independence between loci. However, as a generalization, Hudson[421] notes "loosely linked loci are typically observed to be near linkage equilibrium in natural populations…. In contrast…very tightly linked loci often show some signs of linkage disequilibrium."

There is growing evidence of a block-like structure to linkage disequilibrium. This implies that some regions of the genome are closely linked and others are unlinked. This structure can, obviously, be produced by recombination hot spots, but interestingly can also be produced without such hot spots.[882]

In summary, a lack of random mating, in particular the existence of subpopulations with different allele probabilities, will cause Hardy–Weinberg and linkage disequilibrium. The proportions of the different subpopulations and the differences in their allele probabilities will affect the magnitude of this disequilibrium. The larger the differences in the allele probabilities between the differing subpopulations, the larger the resulting disequilibria. Excoffier[283] notes that population subdivision will also produce a larger number of observed alleles, with an excess of rare alleles.

The first human populations that came under intense scrutiny by the forensic community were the Caucasian populations of the U.K. and the U.S. These populations comprise subpopulations arising from different areas of the U.K. and Europe. Studies have suggested that there are only minor differences between these Caucasian subpopulations in Europe or the U.K. *per se*. Although these differences are real,[79,152,566] they are small and hence they give rise to very small disequilibrium effects. The effect of these disequilibria is a very mild bias in the product rule toward the assignment of a genotype probability that is too low.

3.2.3.6 An Infinite Number of Generations

Loci that are on different chromosomes or well separated on the same chromosome will assort in a Mendelian manner. The linkage disequilibrium associated with such loci is expected to halve with every generation,[267] and hence

[c] This possibility appears to have received recent acceptance from Budowle and Chakraborty at least in the published literature, both previously strong supporters of the use of the product rule.[60]

will approach equilibrium asymptotically, but never quite get there, if the disturbing force is removed. Linked loci will also approach equilibrium but more slowly, depending on the rate of recombination between the loci. An example of very tightly linked loci that are near equilibrium is given by Mourant, when he discusses the Rhesus blood group (a set of three linked loci) in Australian Aborigines.[566]

3.2.3.7 Summary

It was a pity that the first population extensively studied by the forensic community was the Caucasian population. This is because this population is probably one of those nearest to Hardy–Weinberg and linkage equilibrium of the large modern human populations. Hence it was the least likely to educate us on departures from equilibrium and how to manage these. At that time we did not understand the weakness of our independence tests, and this contributed to our misunderstandings. We return to this subject in Chapter 5.

This section is closed with a quote from Wild and Seber: "What often happens is that, in the absence of knowledge of the appropriate conditional probabilities, people assume independence. … this can lead to answers that are grossly too small or grossly too large — and we won't know!" [865] The situation in DNA is probably not this bad, but the warning is real nonetheless.

3.2.4 How Big Is the Potential Departure If We Use the Product Rule?

It has become accepted wisdom that the error induced by ignoring subpopulation effects may be of the order of a factor of 10. This was based on the comparison of the product rule estimator using various databases as the source of the allele probability estimates. Budowle et al.[127,128] and Hartmann et al.[396] compared the product rule assessment calculated from different subpopulation databases and demonstrated that over 80% of assignments were within a factor of 10 of each other. This approach compares an estimate with an estimate. There has been considerable discussion about the bias inherent in this analysis due to sampling effects,[691] but we have difficulty deciding how much can be read into the results of these discussions.

The conclusions arising from these studies require further validation. It is not totally different to the situation where two students give the same answer in a test. It would be unwise to assume that because they gave the same answer they are both correct.

In addition, we must expect an effect from the number of loci and the populations under consideration. The more the loci, the larger the potential effect of population subdivision. Certain populations are expected to show larger departures than others.

Much later, Gill et al.[355] investigated the magnitude of this bias and refined Budowle's method. Using this modified approach, Gill et al. calculated the product rule assignment for a ten-locus genotype using allele probabilities from the relevant subpopulation and this probability when estimated from an averaged European database (see Table 3 of Gill et al.). They found that the difference between these two estimates may be of the order of two, three, or even four orders of magnitude. Further, they show that almost any of the available adjustment methods, such as a subpopulation correction or even the use of minimum allele probabilities, if applied sensibly, will compensate in part or in full for this effect.

The comparison of an estimate with an estimate is interesting, and would give us some confidence that the effect of changing the database is minor. However, it does not show that either estimate is within a factor of 10 of the true value. It is the latter question that is of forensic interest: How far is our estimate from the true value? The suggestion that the difference between the product rule estimate and a hypothetical true value is a factor of 10 must be taken as a hypothesis with some empirical support. It cannot be taken as proved as we cannot know the true value. Even the simulations by Curran et al.[d] described later in this chapter do not truly compare this estimate to a true value. They simply compare the difference between the product rule assignment and that which would occur under certain population genetic events. It is a simple fact that we cannot measure the difference between the product rule estimate and a true value. Nor can we measure this difference for any other population genetic model. The simulations seek to bring evidence to bear on this matter, but they are, in my opinion, a long way short of scientific proof.

It is often assumed that cosmopolitan populations do not exhibit subdivision. While this may be true, there are also instances where it may not. If the population is old and well mixed, there should be very little, if any, population subdivision. However, a cosmopolitan population may be something like that of London or New York, which consist of people with very different genetic backgrounds who live in the same area. This is exactly the situation where we expect subpopulation effects.

3.2.5 Populations Separating By Genetic Drift

If we accept that the loci that we consider in forensic applications are selectively neutral, then we expect the main evolutionary force producing differences between separated populations to be the random drift of allele

[d] This follows a set of concepts discussed between Mulligan J. and myself during *R v Karger*.[639] I am indebted to His Honour for sharing his insight in this matter, which is often hard to convey in a court situation.

probabilities. This is an extensively researched subject and is only covered very superficially here.

Even if all other evolutionary forces were absent, the allele probabilities in one generation would still differ slightly from the previous one. This difference is caused by the random transmission of alleles to the new generation. For large populations, this effect is very small and takes a long time to be observable. However, for smaller populations the effect may be quite rapid.

The difference between populations that are diverging by drift is often characterized by a parameter θ or F_{ST}, which may be treated as synonyms for the purposes of this text. This parameter is often termed the between-person coancestry coefficient. It is a very useful parameter for characterizing the subpopulation effect; however, it is both difficult to visualize and to measure. For the purposes of this section, it will be adequate to consider it as a measure of the genetic distance between subpopulations. The larger the distance between subpopulations, the longer we assume that they have been separated and the higher θ will be.

It turns out that θ may also be considered as a measure of the relatedness between people in the subpopulation. If this subpopulation has been separate from others for some time, then people in this subpopulation will be more related to each other than they would be to a person taken from a different subpopulation. To help give a feel for the size of θ values, consider that first cousins would have $\theta = 0.0625$.

A formula relating θ to the time since separation is given in many standard texts:[836]

$$\theta_t = 1 - \left(1 - \frac{1}{2N}\right)^t$$

where t is the time since separation in generations and N is the effective size of the population (strictly a monoecious population in which selfing is allowed). Evett and Weir[267] discuss the avoidance of selfing and show that the above model is a close approximation. Crow and Kimura[211] give

$$\frac{1}{N_e} = \frac{1}{4N_m} + \frac{1}{4N_f}$$

for the effective size of the population (N_e) when separate sexes of number N_m and N_f are present. When the sexes are present in equal numbers, $N_m = N_f = N/2$ and hence $N_e = N$. Crow and Kimura discuss the effect of differing numbers of progeny on N_e.

If mutation of the infinite alleles type is added to the model, then the opposing forces of drift and mutation may form an equilibrium state, given in several texts:[267,836]

$$\hat{F} \approx \frac{1}{1 + 4N\mu}$$

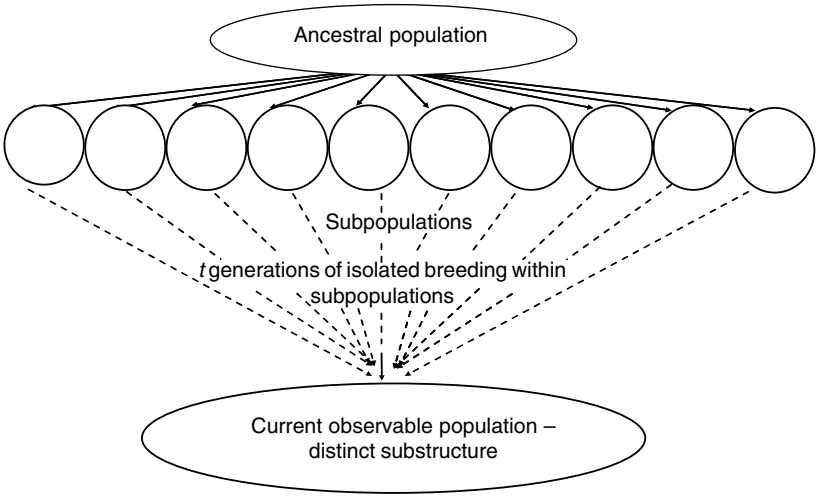


Figure 3.1 Simplified population model. Reproduced in amended form from Curran et al.[215] © 2003, with permission from Elsevier.

where $\hat{F}$ is the equilibrium value of the between-person inbreeding coefficient and μ is the mutation rate.

3.3 Simulation Testing

3.3.1 Product Rule

Curran et al.[215] consider the question: How wrong could the product rule estimator be if the population was subdivided into ten subpopulations and the θ value was approximately 0.03? A computer simulation that allowed the liberty of using the true match probability referred to as the "Gold Standard" examined this question. Populations with known amounts of substructure were produced by dividing a population and allowing it to breed by random mating only within the subpopulations for a suitable number of generations to create the required amount of structure (see Figure 3.1). The ratio of the product rule estimator to the true match probability was then compared. This simulation demonstrated the subpopulation effect but it does not include the effect of mutation. Nor can we truly claim that this is the true match probability. It is certainly the probability if the populations satisfy certain genetic assumptions, but how accurately these assumptions apply to the human condition is the real question.

The Curran et al. results are reproduced in Figure 3.2. Data points above the line given by ratio = 1 indicate that the assignment is conservative with

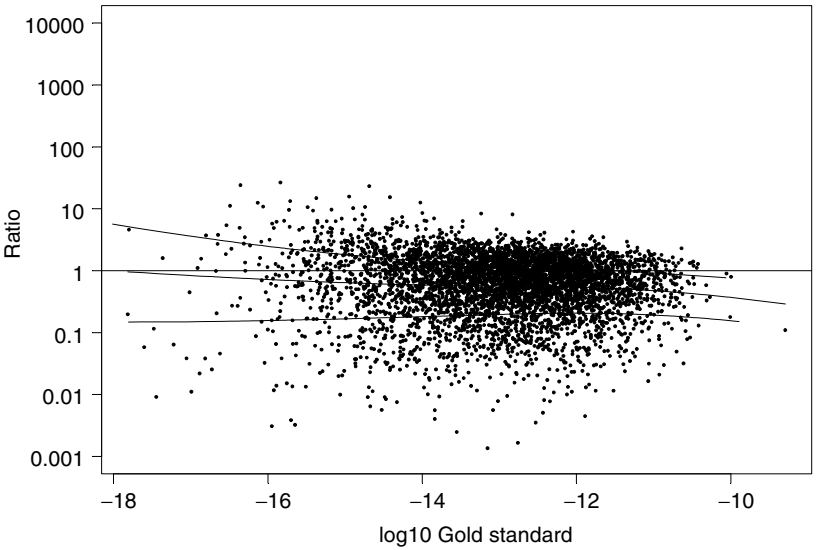


Figure 3.2 Ratio of the naïve product rule profile frequency to the true profile frequency for a population with true inbreeding coefficient $\theta = 0.03$. The median and quartile trend lines are fitted. 64.7% of samples have values less than 1. Reproduced in amended form from Curran et al.[215] © 2003, with permission from Elsevier.

respect to the true value. Data points below this line indicate that the estimate is nonconservative. The product rule assignment is seen to be nonconservative for 64.7% of the 50,000 simulated profiles (given the above conditions). The first thing that we note is that this number is greater than 50%. In other words, the product rule estimator has a mild bias in favor of the prosecution if the population is subdivided. This effect is most pronounced when the profile is common. The simulation is for ten loci. The effect would be greater for more loci and less for fewer loci.

In 14.7% of simulated profiles, the estimate was less than one tenth of the true value. By this we are saying that in 14.7% of cases the product rule estimator is incorrect and favors the prosecution by more than a factor of 10. Indeed, a number of estimates differ by more than a factor of 100. This effect is not a result of sampling error because the simulation has been set up to remove all effects of sampling error. Sampling error would add additional uncertainty to these estimates and would spread the results up and down on the graph. We emphasize that usually the subpopulation effect is mild and we do not wish to overemphasize it. The result could be viewed as not substantially different from the conclusion of Budowle et al.: that 80% of estimates were within a factor of 10 of each other.[122,127,128]

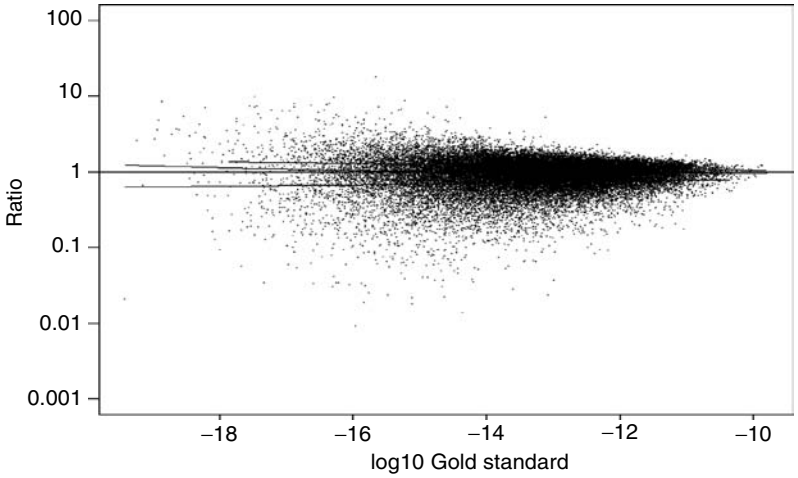


Figure 3.3 Ratio of the naïve product rule profile frequency to the Gold Standard Profile Frequency for a population with true inbreeding coefficient $\theta = 0.01$. The median and quartile trend lines are fitted. 51.8% of samples have values less than 1. Reproduced in amended form from Curran et al.[215] © 2003, with permission from Elsevier.

The choice of 3% as a value for θ is somewhat arbitrary and would be excessive for Caucasian populations in the U.S. However, it may be more appropriate for Hispanic populations and may, indeed, be an underestimate for Amerinds. The subpopulation effect would be smaller for smaller θ.

In Figure 3.3, we reproduce the equivalent graph with the subpopulations bred to $\theta = 0.01$. In this case, 51.8% of samples returned values less than 1, compared with 64.7% for $\theta = 0.03$. The bias is seen to be very small in this instance. (Do not be deceived by the mean trendline being above 1 at the left. This is expected and is more than compensated for by it being slightly below 1 at the right hand end.) Only a few values lie outside a factor of 10 of the true answer.

It can be seen from these experiments that the product rule estimator has a very small bias in favor of the prosecution in most cases where the population is subdivided. The magnitude of this bias is not large, and it is important not to overemphasize it. However, it is real and is not the result of sampling uncertainty. It will be larger for strongly subdivided populations and smaller for less subdivided populations. The effect may be more than a factor of 10. This finding adds an important verification relative to a true match probability.[e] It does put into perspective comments such as "implementation of the product rule is a reasonable best estimate,"[395,486,509] which must be qualified

[e] Of course this is not a "true match probability" either, but it is the true match probability under THIS model.

with our current understanding that the product rule is unlikely to be an unbiased estimator.

The Curran et al. simulations do not include a specific consideration of mutation. Consideration of an infinite allele mutational process has suggested that this may have a significant effect on the estimation process:

> The product rule probability always underestimates the two-locus match probability. For highly mutable minisatellite loci, these probabilities can differ by an order of magnitude or more… the degree of underestimation worsens for more loci.[488]

This statement is for an infinite allele mutation model and may not be appropriate for a stepwise mutation model. However, it does suggest that further research is warranted if the product rule is to be used.

3.3.2 NRC II Recommendation 4.1

NRC II recommendation 4.1 offered a correction for Hardy–Weinberg disequilibrium caused by the Wahlund effect. It was suggested that a correction upward in frequency be applied to correct for the expected upward bias produced by population subdivision, and further that this correction should be applied only to homozygotes. No correction was recommended for heterozygotes since, on average, these should have a downward bias (recall that individual heterozygotes may be displaced from expectation in either direction). This comment is generally true for the event of population subdivision, but would be untrue for populations undergoing admixture. In admixing populations, the number of heterozygotes is likely to be elevated.

The recommendation suggests that

$$P_i = \begin{cases} p_{i1}^2 + p_{i1}(1 - p_{i1})F, & A_{i1} = A_{i2} \\ 2p_{i1}p_{i2}, & A_{i1} \neq A_{i2} \end{cases} \tag{3.3}$$

where F is the within-person inbreeding coefficient and not the between-person inbreeding coefficient, θ, as written in NRC II.

This recommendation is a logical way of correcting for Hardy–Weinberg disequilibrium, but makes no attempt to correct for linkage disequilibrium. It will suffer from the same approximations that are revealed in Table 3.2 for the 1:1 mix from genotypes. Hence it will still have a very mild tendency to underestimate multilocus genotype probabilities.

Curran et al. tested recommendation 4.1 by comparing this assignment with the "Gold Standard Profile Frequency" for a population with a true inbreeding coefficient $\theta = 0.03$ created by simulation. This is reproduced in Figure 3.4. In this simulation, 54.4% of values are less than 1 (reduced from

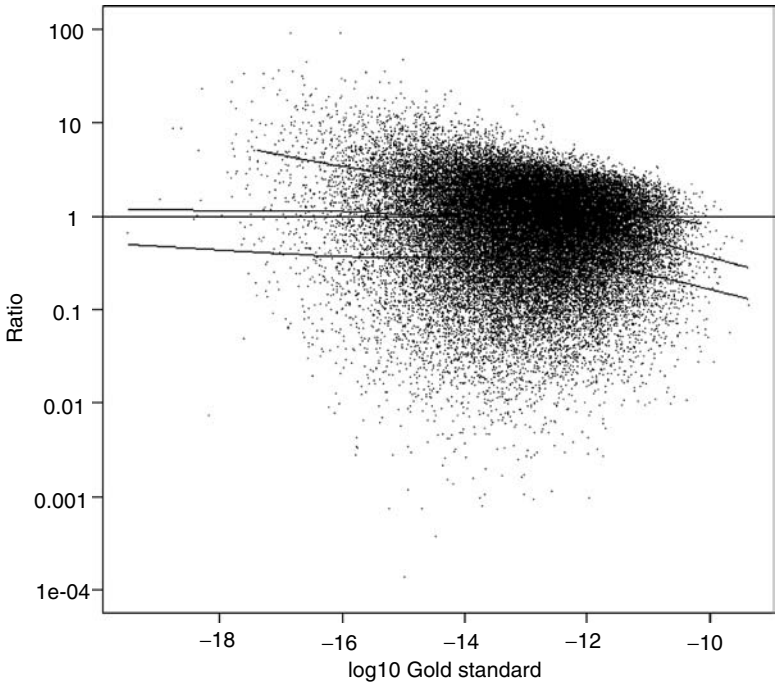


Figure 3.4 Ratio of the Recommendation 4.1 profile frequency ($\theta = 0.03$) to the Gold Standard Profile Frequency for a population with true inbreeding coefficient $\theta = 0.03$. The median and quartile trend lines are fitted. 54.4% of samples have values less than 1. Reproduced in amended form from Curran et al.[215] © 2003, with permission from Elsevier.

64.7% for no correction). We see that this estimator still has a small prosecution bias and some undesirable variance properties.

3.3.3 The Subpopulation Formulae

If it is difficult to calculate the genotype probability in the population due to the effects of population subdivision, can we calculate it in the subpopulation of the suspect? We note that the subpopulation of the suspect may not be known, may not be easily defined, and almost certainly has not been sampled.

A potential solution has been offered by Balding and Nichols and has found widespread acceptance both in the forensic and the legal communities. The formulae[29,36,41,267,585] calculate the conditional probability of a second profile matching the stain from the subpopulation of the suspect given the profile of the suspect.

These formulae follow from a formal logic given initially by Balding and Nichols and appearing as Equations (4.10) in NRC II and (4.20) in Evett and

Weir, but they date back to the work of Sewall Wright[873] in the 1940s. A reasonably gentle derivation appears in Balding and Nichols.[39]

$$P_i = \begin{cases} \dfrac{[3\theta+(1-\theta)p_{i1}][2\theta+(1-\theta)p_{i1}]}{(1+\theta)(1+2\theta)}, & A_{i1}=A_{i2} \\ \dfrac{2[\theta+(1-\theta)p_{i1}][\theta+(1-\theta)p_{i2}]}{(1+\theta)(1+2\theta)}, & A_{i1}\neq A_{i2} \end{cases}$$

$$P = \prod_i P_i \tag{3.4}$$

Let us call the profile found at the scene of a crime profile C with genotype G_c. We will write the probability that the offender has this profile as $\Pr(G_c)$. Such a probability is called a profile probability, as the probability is not conditioned on any other information. Recommendation 4.1 is an attempt to calculate this probability.

However, let us consider whether the probability of a second copy of a certain genotype is raised slightly if one other person is known to have this genotype. There are many reasons why this may be true. But initially we will merely assume that it is true. If we had no knowledge as to whether or not this genotype had ever been found previously in an individual, then, indeed, we would be required to resort to a profile probability and Recommendation 4.1 may be an appropriate method. The "true" value of most of these profile probabilities would be 0 as discussed in Chapter 2.

However, we invariably have the information that at least one copy of the profile exists. We have seen it in the suspect. In other words, we are not talking about the vast majority of profiles that do not exist, we are talking about one of the few that do, indeed, exist in the real world.[842] Let us call the genotype of the suspect G_s, and we note that G_s and G_c are the same. In other words, the suspect could be the source of the stain at the scene. We are interested, however, in calculating the probability that a second person has this profile given that the suspect has it. This is written as $\Pr(G_c|G_s)$ and is called a match probability. It will be the same as the profile probability $\Pr(G_c)$ only if the knowledge that one person has the profile has no impact on our assessment that a second person has the profile. This is the assumption of independence discussed at the start of this chapter.

For the various population genetic reasons given above, we expect the assumption of independence to nearly hold, but to be violated in a minor way, in real populations. The main reason for this is population subdivision and relatedness. The fact that one person has the profile slightly increases the probability that his/her relatives or other members of his/her subpopulation have the profile. We are therefore led to the consideration of match probabilities.

It has been assumed that application of these formulae requires an assumption of independence between loci.[280, 311] This follows from the way that the single locus probability assignments are assembled into a multilocus probability assignment. Indeed these are multiplied and this gives the impression of an assumption of independence.

However, this is not true and was explicitly stated in Balding and Nichols' original paper:[36]

> Further, we have restricted attention to the suspect's sub-population and hence concerns about the Wahlund effect and correlations among loci can be ignored. Therefore the whole profile match probability is, to a close approximation, the product of the single-locus probabilities.

For those who prefer to investigate this statement in an algebraic way, some formative thoughts are given in Box 3.1. The subpopulation formulae of Balding and Nichols were designed to give an estimate of the match probability in the same subpopulation as the suspect. Most implementations of this approach apply this correction (in an overly conservative manner) to the whole racial group to which the suspect belongs rather than simply applying it to the subpopulation of the suspect. This is an understandable response to the difficulties in defining the subpopulation of the suspect, which most often is unknown, and not definable even if known. Equally the proportion of this subpopulation in the population is likely to be unknown. However, the approach of applying the correction to the whole "race" usually results in the correction becoming an "overcorrection" and hence gives rise to considerable conservativeness (or even performs in an overly conservative manner[f]) in the probability assignments.

Over the years I have received a lot of adverse criticism to the use of this correction regarding the difficulties in defining the subpopulation of the suspect. The difficulties can be demonstrated by taking almost any person and considering the question: "To what subpopulation does he belong?" Consider a Caucasian resident of New Zealand, born in London to New Zealand parents. He has Irish, Scottish, Norwegian, and English ancestors. It is almost impossible to define a subpopulation for him. This would be true of most people. This is termed a "population-centered approach" and it can be depicted graphically (see Figure 3.5). In this arbitrary graphic are placed circles depicting the Irish, Scottish, and English subpopulations. These all overlap in differing ways. Where should we now place Norwegian? Nor have we

[f] Clearly the term "overly conservative" used here has no objective definition. Rather it is a subjective term used to imply a very strong bias in favor of the defendant.

Box 3.1 Linkage Equilibrium and Conditional Probabilities (J.S. Buckleton and C.M. Triggs)

Consider two loci (locus 1 and 2). The crime stain has genotype G_c^i at locus i. The suspect matches and hence has genotype G_s^i at this locus. We note that $G_c^i = G_s^i$ for each of the loci, i, examined. We require $\Pr(G_c^1, G_c^2 | G_s^1, G_s^2)$. Using the third law of probability,

$$\Pr(G_c^1, G_c^2 | G_s^1, G_s^2) = \Pr(G_c^1 | G_c^2 | G_s^1, G_s^2) \Pr(G_c^2 | G_s^2, G_s^2)$$

Balding and Nichols' equation (Equation (3.4)) approximates this as

$$\cong \Pr(G_c^1 | G_s^1) \Pr(G_c^2 | G_s^2)$$

This is not an assumption of independence between G_c^1 and G_c^2.
One condition that will make this true is if

$$\Pr(G_c^1 | G_c^2, G_s^1, G_s^2) = \Pr(G_c^1 | G_s^1) \quad \text{and} \quad \Pr(G_c^2 | G_s^1, G_s^2) = \Pr(G_c^2 | G_s^2)$$

Looking at the first equality, we note that this does not imply independence between G_c^1 and G_c^2 unconditionally but rather implies that G_c^1 is independent of G_c^2 and G_s^2 in the presence of G_s^1. In other words, G_c^2 and G_s^2 provide no further information about G_c^1 given G_s^1. The truth of this assumption depends on our belief in the population genetic model.

The second equality requires that G_c^2 is independent of G_s^1 in the presence of G_s^2. The Balding and Nichols' equations are not a simple assumption of independence between loci.

The model upon which Balding and Nichols' equations (Equations (3.4)) are based assumes Hardy–Weinberg and linkage equilibrium at the subpopulation level (as well as some other assumptions). This is an explicit assumption of disequilibrium both within a locus and between loci at the population level. It is therefore seen that Balding and Nichols' formulae correct for that component of linkage disequilibrium that is caused by population subdivision.

really been specific enough. Should we have said "Graham" rather than Scottish? Hence the argument goes: subpopulations are indefinable.

However, the problem is illusionary. This can be shown by a similar graphic. Consider the same population but from a suspect-centered approach. The suspect has a number of close relatives: siblings, parents, and children. He also has more distant relatives: uncles, cousins. Further out he

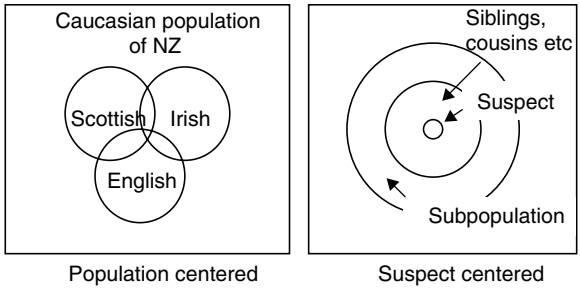


Figure 3.5 Diagrams depicting the population centered and suspected centered views of defining a subpopulation.

has second cousins and so forth. Beyond this there are a number of people to whom he is related more remotely. He may not know these people and there is probably no collective name for them. These are his subpopulation.[99,g]

Curran et al. use this same simulation approach to test how the "correction" advocated by Balding and Nichols[36] would perform.

Figures 3.6 and 3.7 reproduce the ratio of the "Balding and Nichols' θ corrected probability" to the true match probability for populations with true inbreeding parameters $\theta = 0.01$ and $\theta = 0.03$, respectively. In this experiment, Curran et al. have used the correct θ value created by the simulation when they applied Balding and Nichols' formula and have applied it to the whole population. In other words, there is no inherent conservativeness in the θ value *per se*, but there is a conservancy in that the correction is applied to the whole population rather than the subpopulation of the suspect alone. We can see that "θ corrected probability" has a strong bias in favor of the defendant, as expected. Few values lie below the ratio = 1 line and most are strongly conservative especially at the "rare" end on the graph.

This approach should remove any tendency of the product rule or Recommendation 4.1 to underestimate the genotype probability from population subdivision, but could potentially leave unaccounted subdivision of the subpopulation, possibly called sub-subpopulation division. The above simulations suggest that there is a substantial bias in the subpopulation formulae toward the direction of overestimation of the genotype probability. Since it is likely that sub-subpopulation effects will be markedly less than

[g] Subpopulations do not end, they fade out. We could envisage persons who are progressively more and more remotely related to the suspect. This could be approximated, if necessary, by bands of persons with differing θ values or better by the use of the general formulation whereby each pair of persons has a θ appropriate for their relationship. For this diagram, we take an arbitrary boundary to the subpopulation. The further out we push the boundary, the more people who are included in the subpopulation but the smaller the average value of θ.

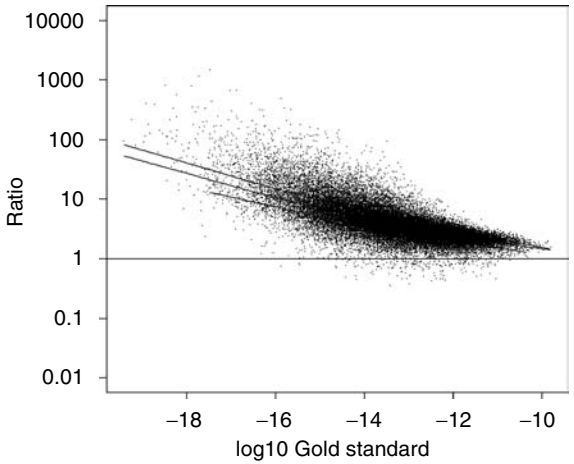


Figure 3.6 Ratio of the Balding and Nichols' profile frequency ($\theta = 0.01$) to the Gold Standard Profile Frequency for a population with true inbreeding coefficient $\theta = 0.01$. 0.5% of samples have values less than 1. The median and quartile trend lines are fitted. Reproduced in amended form from Curran et al.[215] © 2003, with permission from Elsevier.

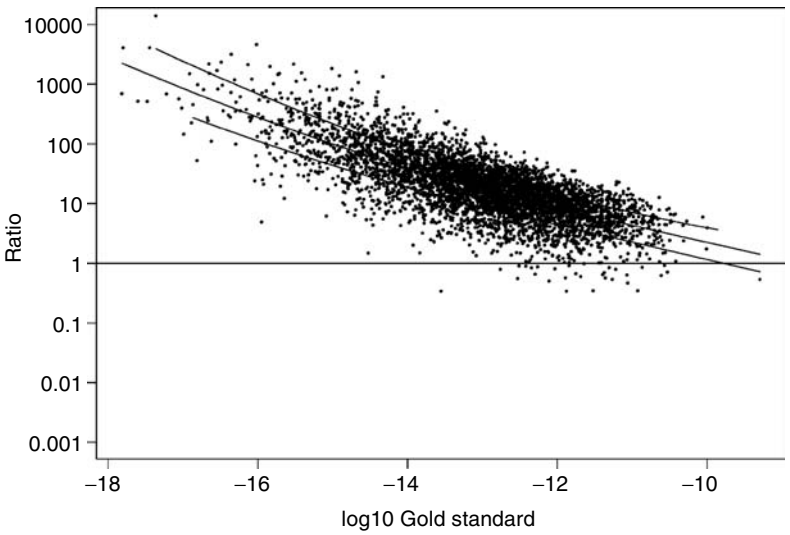


Figure 3.7 Ratio of the Balding and Nichols' profile frequency ($\theta = 0.03$) to the Gold Standard Profile Frequency for a population with true inbreeding coefficient $\theta = 0.03$. 0.8% of samples have a ratio of less than 1. The median and quartile trend lines are fitted. Reproduced in amended form from Curran et al.[215] © 2003, with permission from Elsevier.

subpopulation effects, it seems very unlikely that there will be any remaining bias toward underestimation.

Most laboratories actually exceed this level of conservativeness in that they tend to use a conservative value for θ. For example, the U.K. Forensic Science Service use a value of 0.02, whereas 0.005 could probably be justified for the Caucasian population of the U.K. Curran et al., using the simulation approach, also tested this. Figures 3.8 and 3.9 give the results from these simulations where the true population inbreeding coefficient $\theta = 0.005$, but 0.01 or 0.02 was used in the Balding and Nichols' correction.

This added level of conservativeness, that is, using a conservative value of θ, simply introduces increased conservativeness in the performance of the Balding and Nichols' estimator.

A criticism of this approach points out that this conditional probability is the probability assignment for a certain genotype in the same subpopulation as the defendant, not in the population as a whole.[121,129] This is indeed correct. It is sometimes suggested that these formulae are, therefore, only applicable if it is known that the true offender, if not the suspect, must be from the same

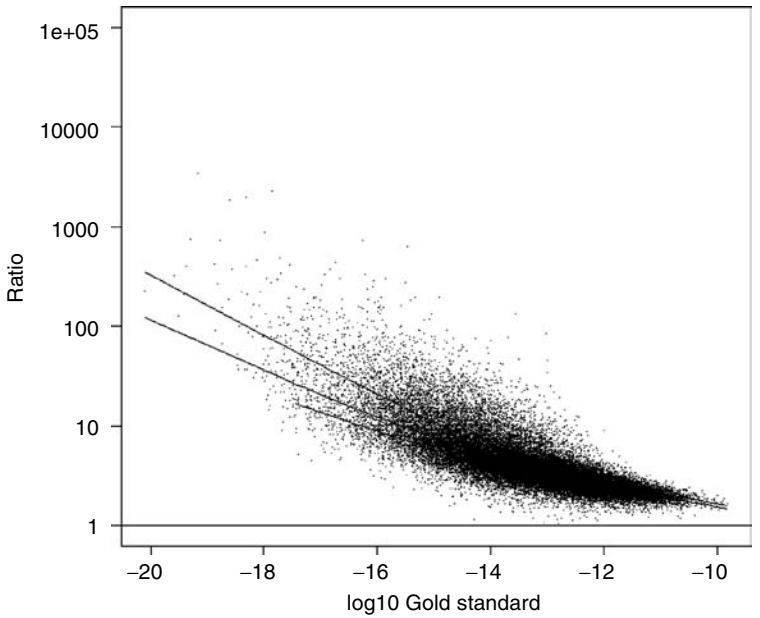


Figure 3.8 Ratio of the Balding and Nichols' profile frequency ($\theta = 0.01$) to the Gold Standard Profile Frequency for a population with true inbreeding coefficient $\theta = 0.005$. The median and quartile trend lines are fitted. 0% of samples have values less than 1. Reproduced in amended form from Curran et al.[215] © 2003, with permission from Elsevier.

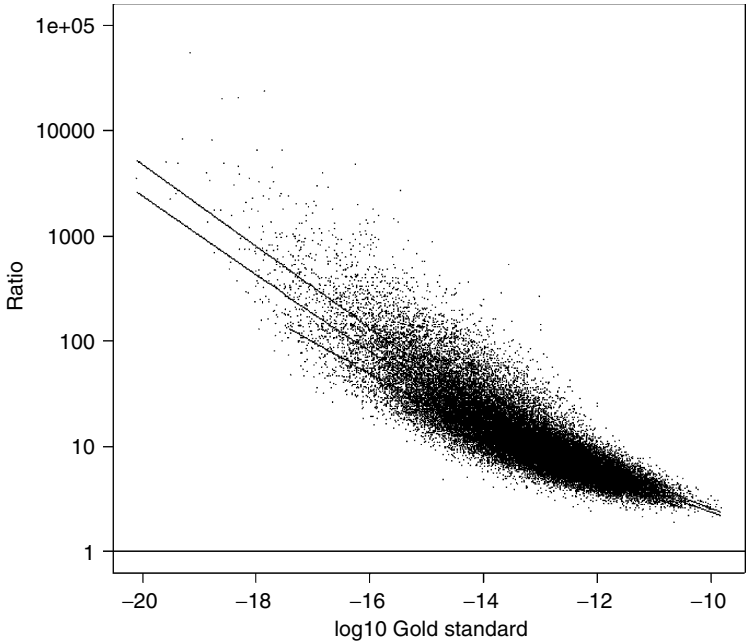


Figure 3.9 Ratio of the Balding profile frequency ($\theta = 0.02$) to the Gold Standard Profile Frequency for a population with true inbreeding coefficient $\theta = 0.005$. The median and quartile trend lines are fitted. 0% of values are less than 1. Reproduced in amended form from Curran et al.[215] © 2003, with permission from Elsevier.

subpopulation as the suspect. This argument can be easily examined by simple mathematical exploration. But before we do that we ask: can the product rule be used only if it is known that all possible offenders are *not* from the same subpopulation as the suspect or are *not* related to the suspect? This is the logical corollary of the argument of Budowle et al.[121,129] If we pursue this line, we will eliminate all possible estimators.

We will assume arbitrarily that each person is as likely as any other to be the true offender if the suspect is innocent. This assumption is very unlikely to be realistic in practice for many reasons, not the least of which is that those people close to the crime scene have a higher chance of being the offender, and persons in remote locations have a lesser chance. Assume further, for example, a population of which 10% are in the same subpopulation as the suspect.

To demonstrate these effects, we generated simulated allele proportions randomly between 0.02 and 0.20 (Table 3.4) and examined the relative contribution to the estimated match probability. In this simulation, 11 loci were

Table 3.4 Simulation of Allele Proportions Randomly between 0.02 and 0.20, and Relative Contribution to the Estimated Match Probabilty

Locus	Pr(Allele 1)	Pr(Allele 2)	Product Rule	Subpopulation	Ratio
1	0.15	0.19	0.0564	0.0724	1.3
2	0.03	0.05	0.0027	0.0085	3.1
3	0.08	0.16	0.0254	0.0384	1.5
4	0.06	0.15	0.0184	0.0305	1.7
5	0.16	0.08	0.0267	0.0398	1.5
6	0.20	0.04	0.0159	0.0297	1.9
7	0.12	0.11	0.0256	0.0380	1.5
8	0.03	0.07	0.0040	0.0110	2.8
9	0.15	0.03	0.0101	0.0212	2.1
10	0.19	0.03	0.0097	0.0227	2.3
11	0.08	0.10	0.0173	0.0281	1.6
12	0.09		0.0082	0.0240	2.9
13	0.18		0.0310	0.0551	1.8
Assigned probability			1.32E−24	6.31E−21	4780
Weighted probability assignment				6.33E−22	

set as heterozygotes and two as homozygotes. The product rule and the subpopulation corrected probability assignments were calculated. For the subpopulation correction, we used $\theta = 0.03$. If we assume that the product rule relates to the 90% of the population who are not members of the subpopulation, and the subpopulation correction relates to the 10% who are members of this subpopulation, we arrive at a weighted probability assignment given.

We see that the weighted probability assignment is different to both the product rule and the subpopulation corrected estimate. But it is almost totally dominated by the contribution of the 10% of the population who are in the same subpopulation as the suspect. The contribution from the product rule is almost irrelevant. In fact, a reasonable approximation could be obtained by simply multiplying the subpopulation probability estimate by its fraction in the population, completely ignoring the product rule contribution. However, if the correction is applied to the whole population rather than simply the subpopulation, as is customary, this is likely to result in an "overcorrection," as previously discussed and demonstrated by simulation. Hopefully this simple example can settle the discussion on the subject of product rule or subpopulation correction. We have a choice: Do we want to be slightly under or more substantially over with our estimate?[h]

[h] Bear in mind that we do not know the true answer. Hence the words "over" and "under" are relative to the "gold standard" which, in itself, is the result of a model.

3.4 Discussion of the Product Rule and the Subpopulation Model

If we are able to show by population genetic studies that the effects of population subdivision are so minor that we are prepared to ignore them, then it is permissible to use the product rule as a first-order approximation provided that it is understood that it is probably slightly biased in favor of the prosecution. A useful review of various approaches is made by Gill et al.[355]

The belief on which the use of the product rule is based can arise only from well-constructed population genetic examinations[600] that assess the population genetic subdivision at the genetic level. This is vital rather than assessment at the geographical level, which may be peripheral, especially in countries settled largely by recent colonization. This is because geographic samples in, say, the U.S., taken from Caucasians from different states or cities, are unlikely to express the underlying genetic diversity. Suppose that we took two samples each of, say, 33% Scottish, 33% English, and 33% Italian. The allele frequencies demonstrated by these two samples will probably be very similar. However, if we compare comparable samples drawn separately from the Scottish, English, and Italian populations, we will find small but real differences between them.

A common and reasonable response is that the difference between the product rule estimate and a fair and reasonable assignment of the evidential value is not forensically significant.[127,128] This is probably true in many instances; however, there is divergent evidence. For instance, in the identification of war victims from the 1991–1995 war in Croatia, Birus et al.[69] found an unexpectedly high number of false matches between skeletal remains and the relatives of missing persons. They attribute this to substructure in Croatia and warn:

> Although genetically and statistically sound and widely accepted, calculations that we perform today produce numbers that might not be fully applicable in all situations. One of the factors not included in these calculations (the product rule) is the effect of local inbreeding.

It remains important to understand that the commonly applied approach of independence testing in no way measures the extent of departure from equilibrium, and cannot be used to estimate the difference between the product rule assignment and a fair and reasonable assignment.[230, 503, 504, 511, 584, 665]

Therefore, the statement that the potential error is not forensically significant, if true at all, cannot be based on independence testing. Again it can only be investigated at all, and certainly not proved, by a population genetic

model or perhaps by experiments of the type pioneered by Tippett[315] in the case of less vanishingly small probabilities.

It may be interesting to note the expected behavior of these two approaches, if indeed the requirement of independence is not fulfilled. If we pick a genotype at random, irrespective of whether it is known to exist or not, then recommendation 4.1 is likely to provide a fair and reasonable probability assignment (note that although it is fair and reasonable, it is not necessarily the true value). However, if we now add the additional information that one person, the suspect, has this profile, then we have two options.

First, we could ignore this additional information and still proceed with Recommendation 4.1. This is no longer an unbiased approach. In fact, using Recommendation 4.1 the probability assignment is likely to have a small bias in favor of the prosecution because the knowledge that we have ignored increases the probability that a second copy of this genotype exists. The extent of this bias is dependent on how large or small are the dependence effects.

Second, we could follow the logical Bayesian approach, which does, in fact, lead to consideration of the conditional probabilities such as $\Pr(G_c|G_s)$ discussed above. These have a remarkable robustness to deviations both from Hardy–Weinberg and linkage equilibrium and as such, we believe, represent a more fair and reasonable probability assignment. However, we accept that, as implemented, they appear to represent an overcorrection. For a discussion on implementation in the U.K., see Foreman et al.[313] (unfortunately not generally available).

This difference between the two approaches is as fundamental as the difference between unconditional probabilities and conditional ones.[267,840] An approach based on mathematical logic leads us to the conditional probabilities. In fact, it would appear that some former major proponents of the validity of the product rule have now modified their position in the face of increasing data.[60,120,121,134,154,743]

There is no possibility of experimentally verifying probability assignments this small. They represent, in multilocus cases, extrapolation way beyond anything that can be experimentally examined.

It must be accepted that, like the product rule, the subpopulation formulae rely on a population genetic model, albeit one that is more robust and concedes doubt correctly to the defendant. Whereas it is possible to say that the product rule is mildly biased towards the prosecution, it is not possible to state whether or not the subpopulation formulae are also biased. It is at least theoretically possible that they are conservative, and the experimental evidence given here suggests that this is so.

A discussion of the ethics of this debate is given by Beyleveld,[62] who also discusses some of the pressures that have been brought to bear on independent bodies, when considering these issues.

3.4.1 Effect of Mutation

The effect of mutation on the assessment of multilocus genotype probabilities has recently been considered. Laurie and Weir[488] warn of the consequences of mutation of the infinite allele type on the estimation process. This model may be a reasonable model for minisatellites, although a consensus has not yet been developed.

Laurie and Weir suggest that the assumption of independence understates the two-locus match probabilities for such loci. The effect increases with increasing mutation rate. For loci with high mutation rates, the two-locus probabilities may differ substantially from the product of single-locus probabilities. They show that these dependency effects accumulate across loci: "These results indicate a potential concern with using the product rule to compute genotypic match probabilities for highly mutable loci."[488]

In loci with high mutation rates, alleles stand an increased chance of being recent and rare. "Hence, if two individuals share alleles at one locus, they are more likely to be related through recent pedigree, and hence more likely to share alleles at a second locus."[488]

This conclusion may hold for the infinite alleles model. This model is unlikely to be applicable to STRs and the effect of mutation on between-locus dependencies at these loci has yet to be settled.

If we restrict ourselves to the question — Do the Balding and Nichols' formulae give an adequate assignment of the match probability in the subpopulation of the suspect? — we again must accept the impossibility of experimentally testing such multilocus estimates.

We are left with examining the validity of the assumptions of the model and simulation results. This matter is elegantly considered by Graham et al.,[370] who point out that the assumptions of the Balding and Nichols' model include a steady-state population and a mutation model in which the allelic state after mutation is independent of the state prior to mutation. Both of these assumptions are untenable. Graham et al.[370] investigate the consequences of a generalized stepwise model and conclude: "[the Balding and Nichols] theory can still overstate the evidence against a suspect with a common minisatellite genotype. However, Dirichlet-based estimators [the Balding and Nichols' formulae] were less biased than the product rule estimator, which ignores coancestry."

Laurie and Weir finish with the conclusion:

> The method of adjusting single-locus match probabilities for population structure [the Balding and Nichols' equations] when multiplied across loci has been shown empirically to accommodate the dependencies we have found for multiple loci.

3.4.2 Admixture

Previously we have described a population genetic model designed to cope with population subdivision. This may describe the evolutionary event where one population continually splits into two or more populations that subsequently evolve separately. Human history is more complex than this and no pretence was ever made by the authors of these approaches that they were exact descriptions of the evolution of actual human populations.

What happens when the rate of gene flow into a population becomes very large?

This may describe the modern evolutionary events in many populations. Populations such as the New Zealand Maori were once much more isolated than they are now. However, they were never completely isolated as the Polynesians were great navigators and there is considerable evidence of extensive trading networks across large distances in the Pacific. With the large-scale settlement of Aotearoa (New Zealand) by Pakeha (Caucasians), gene flow of Caucasian genes into the Maori population was initiated and seems to have been sudden and considerable. The modern New Zealand Maori population is thought to contain no full-blood Maori.[491]

This is a different evolutionary event to the small-scale migration treated in modifications of the subpopulation model. It warrants separate treatment with a different population genetic model. We will refer to this model as the "admixture model."

Admixture in the Americas is common, with individuals having ancestors who may be Caucasians, Native Americans, Asians, or Africans.[688] It has been estimated that 15–25% of the African-American gene pool is derived from the Caucasian population.[606]

Chakraborty and Kidd[161] suggested that estimation of profile frequencies using average allele frequencies and the product rule may be recommended as the number of individuals in the population with mixed ancestry increased. This is partially because random mating in the admixed population restores the within-locus disequilibrium in the population and the between-locus disequilibrium is halved after each generation.[160] However, this thinking applies more to a future equilibrium situation and not to the transitional state that most admixing human populations demonstrate. In the transitional state, there is pronounced correlation between loci, whether the admixed population is defined to exclude pure blood individuals or not. This can be demonstrated by extreme examples such as Table 3.5. Note that in the crossed offspring, every individual is genotype *abcd* and hence this population is in Hardy–Weinberg and linkage disequilibrium. Real examples will show much milder effects.

Law[491] describes an alternative and preferable model for this situation. This model is based on the concept that alleles are independent within and between loci conditional on the pedigree (essentially an assumption of Mendelian

Table 3.5 Hypothetical Admixture Between Two Populations

	Allele probabilities	Pop 1	Pop 2
	Locus 1		
	Allele *a*	1	0
	Allele *b*	0	1
	Locus 2		
	Allele *c*	1	0
	Allele *d*	0	1
Genotype probabilities	*aa*	*ab*	*bb*
	Pop 1 × Pop 1		
cc	1	0	0
cd	0	0	0
dd	0	0	0
	Pop 1 × Pop 2		
cc	0	0	0
cd	0	1	0
dd	0	0	0
	Pop 2 × Pop 2		
cc	0	0	0
cd	0	0	0
dd	0	0	1

segregation). The model allowed for differing mating patterns, number of parental populations, and genetic distance between populations. Comparisons were made with the product rule estimate using average allele frequencies.

Law concludes that "as the genetic distance and the number of parental populations increases, the difference between the match probability calculated using (the Law admixture model and) the product rule increases. The maximum difference can be larger than (a) factor of more than 10,000 for a six loci genotype."

The Law model can also be compared with the estimate that would be produced if the substructure model of Balding and Nichols were used for a population undergoing admixture. This analysis suggests that a conservative estimate of θ could be used in Balding and Nichols' equation along with the allele frequencies from the whole admixed population. Since we are using a model where the inbreeding coefficient θ does not have its usual interpretation, it is better to rename it as the "equivalent inbreeding parameter q" and to understand that we are simply seeking that value for q which gives us approximately equal estimates when compared with the admixture model.

Law concludes that:

> …there are genotypes which require an equivalent inbreeding coefficient that is greater than the genetic distance between the

parental populations especially when there are three or more parental populations. However, the spread of the estimated equivalent inbreeding coefficients is reasonably large as different genotypes may be affected by admixture to differing degrees depending on the difference in allele frequencies. Using the maximum estimated equivalent inbreeding coefficients is likely to overestimate the match probability since such an extreme estimate is most likely (to be) due to rare alleles in one of the parental populations. The 95th percentile of the equivalent inbreeding coefficient may provide a more appropriate value of q.

This analysis suggests that the use of a value for q that is the same as the genetic distance between the parental populations may be an adequate compensation for admixture effects (see Table 3.6). If a more accurate estimation is required, the Law algorithm is preferred.

3.4.3 Allelic Dropout

Occasionally the situation occurs when one allele can be reliably scored but it is ambiguous whether or not there is a second allele. This situation is handled using the "F" designation in the U.K. and the "N" designation in New Zealand. Using the product rule the, say, 16, F genotype is assigned a frequency $2p_{16}$ (strictly this should be $p_{16}(2 - p_{16})$[102]). This approximation has been referred to extensively as the "2p rule." Using the subpopulation

Table 3.6 Median, Upper Quartile, 90th, 95th Percentiles, and the Maximum for q

Number of Parental Populations	Admixture Proportions	Genetic Distance	50%	75%	90%	95%	max
2		0.03	0.01	0.02	0.02	0.02	0.04
	Equal	0.05	0.01	0.02	0.02	0.03	0.06
		0.10	0.03	0.04	0.05	0.06	0.10
		0.03	0.01	0.01	0.02	0.02	0.05
	Unequal	0.05	0.01	0.02	0.03	0.03	0.06
		0.10	0.04	0.05	0.07	0.08	0.13
3		0.03	0.02	0.02	0.03	0.03	0.05
	Equal	0.05	0.03	0.04	0.05	0.05	0.08
		0.10	0.08	0.10	0.12	0.14	0.23
		0.03	0.01	0.02	0.02	0.03	0.05
	Unequal	0.05	0.02	0.03	0.04	0.05	0.08
		0.10	0.05	0.07	0.09	0.10	0.14

Reproduced with the kind permission of Dr. Law.

Table 3.7 Conditional Probabilities for the "θ" Designation Assuming Two Different Conditioning Situations

Genotype of stain	Suspect	"2p Equivalent"	"$p_{16}(2-p_{16})$ Equivalent"
16θ	16,16	$2\dfrac{2\theta+(1-\theta)p_{16}}{1+\theta}$	$\dfrac{2\theta+(1-\theta)p_{16}}{1+\theta}\left(2-\dfrac{3\theta+(1-\theta)p_{16}}{1+2\theta}\right)$ that is always less than the 2p equivalent
		$2\dfrac{\theta+(1-\theta)p_{16}}{1+\theta}$	$\dfrac{\theta+(1-\theta)p_{16}}{1+\theta}\left(2-\dfrac{2\theta+(1-\theta)p_{16}}{1+2\theta}\right)$
	16, x		that is always less than the 2p equivalent

correction, the probability assignment depends on the genotype of the suspect and any other conditioning genotypes. To demonstrate this, we condition only on the suspect's genotype below, and the extension to conditioning on additional genotypes follows by the same method (see Table 3.7). However this approach does not adequately model drop out. It is preferable to use the models discussed in Chapter 8.

3.4.4 Arbitrary Limits

Foreman and Evett[311] have suggested that "case specific match probabilities should not be calculated as a matter of principle." Instead they suggest the use of "general figures." Below we give the calculated figures for the most common profiles for an SGM$^+$ 10-locus match and the suggested reported value:

- 1 in 8300 for siblings which they suggest reporting as 1 in 10,000.
- 1 in 1.3 million for parent/child reported as 1 in a million.
- 1 in 27 million for half siblings or uncle/nephew reported as 1 in 10 million.
- 1 in 190 million for first cousins reported as 1 in 100 million.
- 1 in 2.4 billion for members of the same subpopulation reported as 1 in a billion.
- 1 in 5 billion for unrelated persons also reported as 1 in a billion.

This is an extension of an older Metropolitan Police Forensic Science Laboratory policy of truncating match probabilities at 1 in 10 million.[382]

This approach is or has been accepted practice in the FSS and at Forensic Alliance in the U.K. Foreman and Evett motivate their approach by stating that "the independence assumptions are sufficiently reliable to infer probabilities that are of the order of 1 in tens of millions" but that SGM$^+$ case specific match

probabilities would "invoke independence assumptions to a scale of robustness which we could not begin to investigate by statistical experiment. ..."

I admit the pragmatism and intuitive appeal of this approach; however, it really is a long way away from my own philosophy. My objections would range from the practical to the philosophical and will be mentioned briefly here.

- The relative reliance upon independence assumptions and Mendel's laws differs markedly between the calculations for siblings to the use of the product rule. For siblings, most of the procedure leading to a probability assignment is based on the assumption that alleles assort in a Mendelian fashion and only to a very small extent on independence assumptions within a population. Hence these calculations are much less affected by uncertainties about independence.
- If we can support probability assignments of 1 in tens of millions using Tippett testing (see Chapter 5) but not lower, how are we to support assignments of 1 in a billion?
- The probability assignments that are advocated in this chapter are really based on belief in a model. They are not based on independence testing or Tippett tests at all.
- A limit of 1 in a billion is not likely to induce further refinements of the model, or simulate further sampling and study.
- What would we do if we added more loci?

In general, I would vastly prefer to assign a probability, whatever it may be, without a limit but to accept and make explicit that very low probabilities cannot be verified experimentally.

3.4.5 Same Source?

The reasonable question has arisen: when can a DNA profile match be considered proof that two DNA samples have come from the same source? The FBI announced a policy on this in November 1997.[410] The term "same source" is used in this discussion to describe this situation as it best approximates the underlying forensic question. Other terms such as "uniqueness," "source attribution," and "individualization" have been used elsewhere. This has led to considerable discussion of the use of these terms, which has also produced useful philosophical debates about their meaning. I cannot do justice to these arguments and simply direct the reader to the well-written work by Champod and Evett[173] on the equivalent subject in the area of fingerprints (see also the response by McKasson[540] and the more balanced commentary by Crispino[206] or the excellent writing of Inman and Rudin[429]).

The question of whether we can ever base a conclusion of common source on a probabilistic argument has also been examined, most notably by Stoney,[734,735] Champod,[168] and Evett et al.[280] In the DNA context we can see that, using the current population genetic models, the more loci we add, the smaller are the match probabilities produced by our model. There are three important points with regard to this. First, that the estimated match probability derived from the model can approach zero but never actually equal zero. Second, that estimates of very small match probabilities arising from models cannot be directly tested. They are as reliable or unreliable as the models themselves. Third, we recognize that we are considering an extreme extrapolation using these models. We are not operating near the center of their prediction range where they are more testable and tested. The models have been extensively tested in this central range and there is some considerable reason to believe that they are robust there, but they are still models and the probabilities produced by them are still untestable.[i]

To conclude the same source from a probabilistic model, someone has to decide that the probability estimate produced by that model at this extreme end of extrapolation is sufficiently reliable that it can be trusted and the probability is sufficiently small that it can be ignored. Stoney[735] terms this the "leap of faith."

Inman and Rudin[429] describe this situation, "at some subjective point, most qualified authorities would agree that, for practical applications, the likelihood … is so small that it can be ignored." In the text following this quote, they very clearly set out the subjective nature of this decision.

There has been considerable argument about whether a scientist should do this or leave the matter to the court. Certainly in England and Wales, the court direction appears to be that the scientist should not be the person who decides whether the probability is small enough to ignore.[201]

Inman and Rudin[429] agree:

> It is the purview of the fact finder to draw inferences from circumstantial evidence, and, of course, potentially individualizing physical evidence is circumstantial evidence. However, there are pieces of information that only science can legitimately provide to the fact finder, such as population frequencies, transfer and persistence data, and limitations of the evidence and the test.

It is unclear whether the scientists should even be the persons who decide on the reliability of the model. It is regrettable to me that, as we add more loci, we extrapolate the model further and further, but little new experimental

[i] To quote Ian Evett: "is it rational for me to assign such a small match probability?"

data into the reliability of the model at this extreme are being produced. Robertson and Vignaux[659] complained about a similar lack of fundamental research in the area of fingerprints:

> In these cases it seems that the expert is giving evidence of identity when, and only when, in his judgement the probability of getting the evidence assuming the alternate hypothesis is so small that it does not matter what the numerator or even the prior odds are. At what point this is reached seems to be a matter of judgement and experience and there most writers on expert evidence are content to let the matter rest. This may have had the unfortunate effect of removing the incentive to carry out the basic research to build appropriate models. Intellectually, this is unsatisfactory and further work is required to understand the processes involved in making these decisions. In the meantime the proposal that all forms of scientific evidence be given in the form of a likelihood ratio is a counsel of perfection.

Returning to DNA profiling, Budowle et al.[129] make the reasonable distinction between the judgement in one particular case and the judgement in all potential cases. We could imagine a criterion that was considered reasonable in an individual case and Budowle et al. suggest "99% confidence."[j] They go on to suggest that this may correspond with the term a "reasonable degree of scientific certainty." This term has been selected because of its legal implications.

> From the medical model has come the phrase "to a reasonable scientific certainty." Both the judicial system and some experts have latched onto this phrase as a convenient way to render an opinion as fact. As convenient as it might be, it is a non sequitur. As we have repeatedly discussed throughout this book, the notion of scientific certainty does not exist. In our opinion, scientific experts should refrain from resorting to that phraseology in expressing their opinions.[429]

Budowle et al.'s method stems from a suggestion by NRC II who discussed the use of the formula $p_x \leq 1 - (1 - \alpha)^{1/N}$, where p_x is the match probability, N is the size of the suspect population, and $1 - \alpha$ is the confidence interval. They give an example using a 99% confidence interval $(1 - \alpha) = 0.99$ implying $\alpha = 0.01$ and $N = 260{,}000{,}000$, the approximate population of the U.S. This suggests

[j] This term needs thought. There is a distinction between the use of the words confidence and probability.

a match probability of $p_x = 3.9 \times 10^{-11}$. It is suggested that the estimated p_x be decreased by a factor of 10 to provide additional conservativeness. Weir[840] correctly points out the flaws in this approach which unreasonably assumes independence of trials.

Also included in the original publication is a brief mention of relatedness. In particular, they recommend typing of relatives. The typing approach to dealing with relatedness is admirable, but is applied only rarely in the U.S., the U.K. or New Zealand. In the absence of typing, they suggest that the match probability for brothers be calculated or that calculations should be performed (when required) for three classes of people: unrelated, subpopulation members, and relatives. They do not give a specific formulation of how to amalgamate the contribution from relatives and unrelated people, directing the reader, correctly, to Balding.[34]

This division of the population into unrelated, subpopulation, and related persons is akin to the coarse division undertaken by Balding. The unifying formula suggests that it is the weighted sum of all three contributions that should be considered and not simply one or the other of these probabilities.

The unifying formula will assign a posterior probability to the hypothesis that the suspect is the donor of the stain material. This appears to be the probability that is desired in "source attribution." However, the unifying formula will require an assignment of prior probabilities and this cannot be avoided. This may appear as a fatal flaw and indeed it is worrying. It is central to the concerns about the concept of "source attribution" and "a reasonable degree of scientific certainty." We see therefore that any approach to assigning a posterior probability involves a prior. This is, of course, not an original insight and was reported as long ago as 1983[257] in forensic science and much earlier in other sciences.

There is an interesting interplay between the prior for the suspect and the probability that someone else possesses this profile. Balding and Donnelly[37] explained this:

> Finally, we remark that the magnitude of the size biasing effect… is related to the prior distribution. Intuitively, the effect occurs because, under the hypothesis of innocence, two distinct τ-bearers[k] have been observed. Such an observation stochastically increases the number of τ-bearers, thus decreasing the strength of the evidence against the suspect and decreasing the probability of guilt. Decreasing the prior probability of guilt increases the chance that the suspect and criminal are distinct, hence increasing the

[k] This is the term used to describe the people carrying the matching profiles: in this case, the defendant and the true perpetrator.

> effect of size biasing. (David Balding and Peter Donnelly quoted with the kind permission of CRC Press)

This effect can easily be illustrated. Suppose that we have a certain profile at a crime scene and that this matches a suspect. But the suspect, for whatever reason, cannot have been the donor (his prior is 0). Then the probability that someone else possesses this profile goes from whatever value it was before to 1.

> Consider a crime scene DNA profile which is thought to be so rare that an expert might be prepared to assert that it is unique. Suppose that, for reasons unrelated to the crime, it is subsequently noticed that the crime scene profile matches that of the Archbishop of Canterbury. On further investigation, it is found to be a matter of public record that the Archbishop was taking tea with the Queen of England at the time of the offense in another part of the country. (You may consider your preferred religious leader, beverage, and head of state in place of those named here.) A reasonable expert would, in light of these facts, revise downwards any previous assessment of the probability that the crime scene profile was unique. However, this is just an extreme case of the more general phenomenon that any evidence in favour of a defendant's claim that he is not the source of the crime stain is evidence against the uniqueness of his DNA profile.[34] (David Balding, quoted with the kind permission of Science and Justice)

The supposition that the Budowle et al. approach is necessarily conservative is of concern. An appeal is often made at this point to the increase in the frequency assignment by a factor of 10 and the relatively large value chosen for N (260 million). The factor of 10 was intended to compensate for potential sampling error or subpopulation effects or both. Examination of the unifying formula suggests that it may be inadequate especially when many loci are considered. It is also likely to be inadequate to compensate for both subpopulation effects and sampling error, and certainly cannot compensate for the effect of uneliminated brothers.

Budowle et al. make it clear that this approach is designed for a case-by-case application. If we misapply this method to the question of "are such profiles unique in the U.S.," we will soon be embarrassed. There are 3.38×10^{16} pairs of people in the U.S. If we use the estimated match probability suggested for the 99% confidence interval $p_x = 3.9 \times 10^{-11}$ and assume that the factor of 10 recommended as additional conservativeness was included, then $p_x = 3.9 \times 10^{-12}$. If this match probability is exactly correct (recall that it is only an estimate), then there will be an expectation of about 132,000 matching pairs

Table 3.8 Size of Databases That Give the Expectation of One Match

	$\theta = 0.00$	$\theta = 0.03$
U.S. African-Americans	43,000,000	11,000,000
U.S. Caucasians	34,000,000	9,300,000
U.S. South-Western Hispanics	21,000,000	5,900,000

of unrelated people in the U.S. In fact, a database of about 716,000 profiles all with a match probability of $p_x = 3.9 \times 10^{-12}$ would have an expectation of about 1 match. In reality, full CODIS profiles produce match probability estimates less than this. Bruce Weir[844] estimates that we would expect a full CODIS match among unrelated people if the databases were of the size shown in Table 3.8.

Despite the careful words in the paper of Budowle et al., my suspicion is that it will be read as providing a declaration of uniqueness among all people and hence such an adventitious match will cause public embarrassment. Certainly the view is developing among the public that DNA profiles are unique.

The situation is probably slightly worse when we consider relatives. The expected number of matches when relatives are included in the population or database will be larger. It is likely that there are a number of pairs of persons matching at the 13 CODIS loci in the whole U.S. population of 260 million. Many of these matching sets will be brothers. The chance that two of these are involved in the same crime is small, but the matches will eventually be revealed as the sizes of databases increase and will embarrass forensic science if we have declared such profiles unique.

Findlay and Grix[299] have studied juries and report a strong preexisting prejudice that is pro-DNA. It is likely that many jury members wrongly believe that all DNA findings represent certain identification. It would be worrying to foster this belief.

My feeling is that we would be unwise to conclude the same source because it is not our place to do so. If we do so, I would prefer the standard to be much higher than previously suggested AND I would like us to make transparent that we have subjectively decided to round a probability ESTIMATE off to zero. On balance I cannot see much positive coming from a policy of declaring a common source.

3.4.6 Animal and Plant DNA

We are starting to see the use of animal DNA in criminal proceedings (for an excellent review, see Halverson[386]) when, say, blood from a shot dog may have been transferred to an offender. Animal and plant DNA is extensively used in

wildlife and conservation science to investigate illegal hunting and other risks to protected species. The population genetic arguments given above apply to all species, except that in many cases subpopulation effects and inbreeding are more severe outside humans.[52]

3.5 A Complex Case Example — DNA Evidence and Orenthal James Simpson[1]

In June 1994, O.J. Simpson was 47 years old. He was one of the most respected sportsmen in the U.S. and he had just been charged with the double murder of his estranged wife Nicole Brown Simpson and her friend Ronald Goldman. This precipitated a trial with media coverage unprecedented in U.S. history. DNA evidence was about to be center stage.

In his early sporting career, O.J. had been the star running back for the University of Southern California, winning the Heisman Trophy in 1968. His professional career was with the Buffalo Bills until his retirement in 1979. That same year his first marriage to Marguerite Whitley, his teenage sweetheart, ended. The couple had three children, a son Jason, daughter Arnelle, and a second daughter, Aaren, who accidentally drowned at the age of two.

O.J. had met Nicole Brown in 1977. She was aged 18, he 30 at the time. Nicole had been born in Frankfurt, Germany to a German mother and a U.S. military serviceman father.

O.J. was inducted into the football hall of fame in 1985, his first year of eligibility.[498] He had married Nicole the same year and the couple later had two children: Sydney born in 1986, and Justin in 1988. However by 1992, Nicole had left O.J. after what was presented at the trial as a history of abuse and violence. In 1993 police were summoned to Nicole's residence after the now estranged O.J. had kicked in the door, screamed obscenities, and had beaten her Mercedes-Benz car with a baseball bat. Official records listed 62 separate incidents of physical and mental abuse by Simpson toward his wife. One of these incidents occurred in 1985 and involved Detective Mark Furhman, who was to feature prominently later in the investigation and trial. Furhman later recalled that this incident was "indelibly pressed" into his memory.[507]

At 10:20 PM on Sunday, June 12, 1994, there was the sound of a dog barking at 875 South Bundy Drive in the Brentwood district of LA. Shortly before midnight, Akita, Nicole's dog, paws splashed with blood, had led neighbors to the scene of the murders. Nicole, aged 35, was face down with her throat slashed almost through. To her right lay the body of a male later identified as Ronald Goldman, aged 25, a waiter at the fashionable Mezzaluna restaurant.

[1] This section was written by John Buckleton and Christopher Triggs.

Nicole and Ronald had known each other for six months, but there was no suggestion of a romance between them. On the night of the murders, he had been delivering Nicole's mother's reading glasses, which had been left in the restaurant. Next to the bodies were keys, a blue knit cap, a beeper, a blood-spattered white envelope, and, nearer to Nicole's body, a bloodstained left-hand leather glove. Bloody shoeprints and spots led from the bodies toward the back of the property.

This book deals largely with the interpretation of DNA after it has been analyzed in the laboratory. It neglects the huge and vitally important fields of evidence collection, recording, and handling. This section seeks in a small way to redress this imbalance. There have been understandable complaints that we forensic scientists have not learnt the lessons necessary from this and other similar cases.

The autopsy was performed on June 14 by Dr. Irwin Golden. It showed injuries to both of Nicole's hands, which suggests that she had defended herself. From the cut to the throat, the pathologist concluded that the attacker was right handed.

Mr. Goldman had been clubbed from behind and stabbed 19 (or 28[498]) times.

White towels had been used by the detectives to soak up blood[498] to allow easier approach to the bodies. *This is an unwise practice*.[m]

Detective Mark Furhman was the 17th officer to sign in at the scene. After initial inspections, instructions had been issued that O.J. should be told personally of the tragedy. Furhman had volunteered. He knew Mr. Simpson's house was two miles from Nicole's from the previous visit. At the trial, the defense claimed that Mr. Simpson was "targeted" by the police. However, it would be normal for an ex-husband to be a suspect early in an investigation and this would not be an issue as long as an open mind was maintained. Detective Vannatter, the head of the team of detectives at the scene, has subsequently insisted that O.J. was not being treated as a suspect at this time. However, events suggest that he was. For instance, the Goldmans were not informed personally of their son's death although O.J. had been. Forensic staff were initially called to O.J.'s Rockingham house. *But a valid complaint would relate to the sending of any personnel from one crime scene to another potential scene. The issue of cross contamination would immediately arise and should have been stringently guarded against. If the same staff must go to both scenes then strict precautions must be taken, such as overalls (clean or disposable), overshoes, and fresh gloves.*

[m] We use the italics here to signify personal commentary as opposed to the historical narrative.

Two cars containing detectives went to O.J.'s residence at the junction of Rockingham Avenue and Ashford Street. Outside, parked badly, was Mr. Simpson's 1994 white Ford Bronco. Detective Furhman pointed out to a fellow officer what appeared to be blood inside the Bronco near the driver's door handle. This supposedly represented reasonable cause and so Detective Furhman climbed over the wall and unlocked the gate. *If this evidence represented probable cause, then this was now a crime scene and different personnel should have been summoned.* Dr. Henry Lee, a criminalist employed by the defense, presented arguments that Furhman must have opened the vehicle[498] since the blood was not visible with the door closed.

There had been no response from the front door. At the back of the house were three guest bungalows. Brian "Kato" Kaelin, a friend of Nicole's, was staying in one. Arnelle Simpson, O.J.'s daughter, in another. She let the officers into the main house. No one was present. O.J. Simpson had taken the 11:45 flight to Chicago to attend a convention of the Hertz Rental Company scheduled for the next day. He appeared in advertisements for this company and his presence at this conference had been expected. The flight had been booked well in advance.

Mr. Kaelin was then interviewed. He and O.J. had been together for dinner at a McDonalds in Santa Monica and had returned to the house at 9:40 PM. At 10:45 Kaelin had heard three banging noises from the rear of the building near an air-conditioning unit. He went outside to inspect the source of these noises and had seen a limousine parked outside the gate. This was the vehicle previously ordered by Simpson to take him to LA airport to catch the prebooked flight to Chicago. A few minutes later O.J. had appeared and Kaelin had helped Allan Park, the chauffeur, to load some bags into the vehicle. O.J. had insisted on holding onto a small black bag.

Allan Park later testified that he had been instructed to arrive at Rockingham no later than 10:45. He had arrived early and first called on the buzzer at 10:40. He received no answer. At 10:50 he spotted a tall, well-built, black man who had hurried up to the house from the Rockingham gate. He had tried the buzzer again and had spoken with Simpson who came down 10 minutes later carrying a bag. Park testified that Simpson was sweating and that he requested that the air-conditioning in the limousine be turned on. Park also testified that it was a cool night.

Furhman returned to the house with the news that he had found a blood-stained right glove in a dark narrow walkway between the bungalows.[498] He had already started the search of this secondary scene. There were blood spots leading out of the west gate into Rockingham. Other red marks were present inside the Bronco on the driver's door and the console near the passenger's side. Another trail of blood spots led up to the front door of the house (see Figure 3.10).

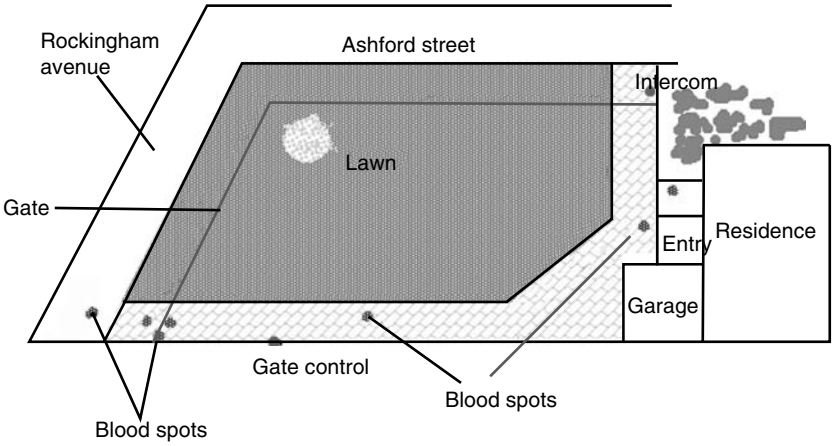


Figure 3.10 The layout of part of Mr. Simpson's Rockingham residence. Reproduced with kind permission from Professor Douglas Linder, University of Missouri Kansas City School of Law.

Detective Vannatter had instructed Furhman to drive to Bundy to check whether the glove at Rockingham matched the one beside the bodies. Furhman did this and then returned to Rockingham. Later the defense would raise the suggestion of the deliberate planting of evidence. *These two episodes certainly create the potential for cross contamination and as such these actions were inviting criticism.*

At this time the detectives had called O.J. at his hotel in Chicago. They reported that his reaction to the news was puzzling, in that he did not ask for details of the deaths. *Highly subjective comments like this are of debatable value. They are unlikely to be admissible in court, nor should they be admissible.*

At 07:10 on June 13, Dennis Fung, an LAPD criminalist, and his assistant Andrea Mazzola, a trainee, arrived at Rockingham. Of subsequent interest in the trial was that they were called to the secondary scene first and not the primary scene at Bundy. *Not every laboratory has the resources to send different teams to different scenes. Many forensic scientists have examined multiple scenes from the same case.*[n] *However, a policy of different personnel for different scenes is clearly advisable especially if one is the crime scene and the other a suspect's domicile.*

At the murder scene a blanket from elsewhere at the scene had been thrown over Nicole's body, presumably by detectives, to protect her from photographers. The motivation was to allow Nicole dignity in death, but the

[n] John Buckleton admits that on more than one occasion he has examined multiple scenes of the same case.

evidential implications had not been well thought through. *Many scene of crime investigators now carry sterile plastic coverings for purposes such as this.*

Later that day O.J. returned from Chicago. He was detained and led to Bundy in handcuffs. This was recognized as improper treatment, and he was released. Detective Vannatter, while unlocking the handcuffs, noted that the middle finger of Simpson's left hand was bandaged. Simpson — reported by detectives as confused — stated that he had cut himself in LA while retrieving his cellphone from his Bronco vehicle. He had reopened the wound in Chicago on a broken glass in the sink. When his hotel room at O'Hare Plaza Hotel was checked, a broken glass was present in the bathroom sink.

At this time, a blood sample had been taken from O.J. and passed to Vannatter and then to Fung at Rockingham. *It has been questioned why the reference sample was taken to the scene and not directly to the laboratory.* Fung and Mazzola had by now bagged and tagged a pair of navy blue bloodstained socks found in the master bedroom at Rockingham.

The 90-minute chase of O.J.'s Ford Bronco on the 17th of June was viewed by an estimated 95 million people. The vehicle was televised driving slowly down LA freeway 405 followed by numerous police vehicles. Al Cowlings, O.J.'s friend and former teammate, was at the wheel. Simpson had a .357 Magnum pressed to his own head in the back seat. In a bag were $US8000, his passport, a fake moustache, and beard. Earlier that day he had failed to appear for arraignment on charges of double murder. Eventually the vehicle had driven sedately back to Rockingham. The following day Mr. Simpson was charged with double murder.

3.5.1 The Evidence

Hairs had been found on Mr. Goldman's shirt and inside the knit cap. These were described in evidence as consistent with having come from O.J. Hairs on the glove found at Rockingham (the Rockingham glove) were consistent with having come from Nicole and Ronald.

Fibers in the Ford Bronco matched fibers on the Rockingham glove and the Bundy knit cap. Blue/black cotton fibers on Ronald's shirt matched the socks that had been found in O.J.'s bedroom. Cashmere fibers from the knit hat matched the glove lining. One glove with this type of lining was at the scene anyway, so the finding of the fibers was not *per se* a connection between Bundy and Rockingham.

The dark brown leather, cashmere-lined, size extra-large gloves had been manufactured by Aris Gloves. This style of glove had only been sold by Bloomingdale's in New York City. Between 1989 and 1992, 240 pairs had been sold, two of these, on December 20, 1990, to Nicole. Photographs were produced of Mr. Simpson wearing gloves of this type in 1993 and 1994. Richard

Rubin, the vice president and general manager of Aris Gloves, testified that he had measured Mr. Simpson's hands as size extra-large.

The police had surmised that the single set of shoeprints at Bundy implied a single killer. The shoeprints showed a waffle-type pattern and were later identified as Italian-made Bruno Magli shoes selling for US$160. They had been sold in 40 stores across the U.S. and 300 pairs of size 12 shoes had been sold in total. Simpson wore size 12 shoes, but our literature search has been unable to ascertain if the prints at the scene were definitely identified as size 12. Simpson later denied ever owning a pair of Magli shoes. However, a photograph was eventually produced of him wearing this type of shoe at a stadium in New York in September 1993.

Henry Lee visited the Bundy scene 13 days after the murder. He found extra shoeprints on a piece of paper, an envelope, and in photographs of Goldman's blood-soaked jeans. This undermined the prosecution's single-killer premise. William Bodziak demonstrated, using photographs taken on the 13th of June, that the extra shoeprints were not there on the day after the murder. Presumably they had occurred after the scene was released. Lee's version of these events does not appear in his book (but he did present photographs of the shoeprints and marks). He has offered to provide his view by correspondence, but it was not available at the time of writing. Without having heard his response, it would be unwise to draw a conclusion.

The DNA profiles of 45 bloodstains were typed and subsequently presented in court. In many cases, these stains were divided and analyzed by two or three separate laboratories. Only the most superficial summary of the results of the typing is presented here. The most important results are considered below.

DNA on the Rockingham glove was consistent with being a mixture of DNA from Mr. Simpson and the two victims. In total, 11 subsamples from this glove were typed. The most extensively typed subsample, item 9:G3, taken from the inside back of the ring finger is discussed in detail. This subsample was typed at eight RFLP and two PCR loci and was found to match Ronald Goldman. Other subsamples on this glove were found to match O.J. Given the subsequent "planting" defense, the presence of blood matching O.J. on the glove is of interest.

Samples of blood, items 47–50 and 52, had been taken from what became known as the Bundy walk. These matched O.J. The samples were taken by Fung and Mazzola on the 13th before O.J.'s blood had been sampled. The first PCR result became available on the 14th. Tampering, if it occurred, had to occur in this window. The five control samples for this batch of items had been unaffected by contamination. Item 52 was the most fully typed, portions of the analysis having been done at one or more of the Los Angeles Police Department laboratory, California Department of Justice, and

Cellmark. Eventually this item was typed at a total of five RFLP and seven PCR loci. The match probability for the PCR loci was estimated as 1 in 240,000 and 1 in 170 million for the VNTR loci. To many this was the strongest evidence.

The defense argued that the items had been incorrectly stored in plastic bags in a hot truck. The parent DNA had completely degraded and the result matching O.J. had come from contamination in the laboratory, allegedly from Mr. Yamauchi's gloves as he had prepared the samples. The defense further argued that the control samples could not be relied on in such a laboratory. A defense explanation of the RFLP results was required, but never given. The RFLP technique requires much more DNA to obtain a result, typically from a stain about the size of a quarter; hence, to explain O.J.'s profile being present, it is necessary to posit gross contamination. Spot 49 (but not 52) in the sequence of five spots had also been tested by conventional serological methods. These would also require the grossest of contamination to register a false result. The match probability for these serological tests was approximately 1 in 200.

The blood on the Rockingham socks was consistent with having come from Nicole. This blood was typed at 14 RFLP and 7 PCR loci. The RFLP match produced a match probability[o] estimate of 1 in 4.4 billion for the Cellmark RFLP set and 1 in 4×10^{10} for the California DOJ set. The PCR result was 1 in 45,000. The two RFLP numbers cannot be simply multiplied as they share two loci but, as Weir points out, numbers are barely necessary.

The defense presented considerable evidence to support their "planting" suggestion. The blood on the socks had not been noticed by Fung when he collected them on June 13, by the defense when they examined them on June 22nd, nor by an LAPD criminalist doing an evidence inventory on June 29. The defense presented evidence that EDTA, a substance used as a preservative for the blood sample tubes drawn from people, was present in the sample of blood recovered from the sock, suggesting that it may have come from Nicole's reference sample. The FBI disputed this finding.

Stains were collected from the rear gate at Bundy on July 3rd and matched O.J. Fung had presumably overlooked these. These were typed by the LAPD using RFLP producing a 1 in 57 billion match probability and by the California DOJ using PCR producing a 1 in 520 match probability. Due to the late collection of this sample, it came under attack as potentially planted. The question of whether these stains contained EDTA and why they were in better condition than the samples taken much earlier was hotly debated.

[o] Quoting the most common result calculated from the five databases used.

Stains were collected from the Bronco on June 14 and later on August 26, over two months after the murders. These stains were consistent with having come from O.J. Simpson, Nicole, or mixtures of blood from O.J. and Nicole, from O.J. and Ronald, or from all three. Various defense arguments weakened much of the probitive value of these findings. Had the stain recovered from the partial shoeprint on the carpet in the Bronco which matched Nicole been transferred by Furhman who had travelled to Rockingham from the Bundy scene? The controls had failed for item 31, which was consistent with having come from O.J. and Ronald. According to the defense, the stains collected on August 26 had been planted and this supposition was bolstered by the fact that a theft had occurred from the vehicle while it was in police custody, reinforcing the view that it was not securely stored.

Match probability statistics were produced for each of these bloodstains and many others.[835,838] One of the authors, John Buckleton, was working with Bruce Weir at this time.

> My part in the saga involved the statistics and the statistics themselves were barely central. Weir and I had advised Cellmark on their data and processes and they presented the statistics we recommended. Weir had presented evidence himself. I was his assistant and my part was repeating his calculations, in the U.S. initially and later in the U.K., after I relocated due to contract obligations. I am the colleague he refers to later, along with Richard Pinchin, Steve Knight, and Ian Evett. I reproach myself for not being in LA and being of more use in the checking.

At this time, match probabilities were still calculated using the product rule except for the "Cellmark wrinkle." This was used at one locus that had failed independence testing. At this locus, the observed genotype probabilities were used. A 99% confidence interval for the match probability was estimated by bootstrapping. All the match probabilities were very small.

> Should we just say that it was O.J.'s blood at Bundy or Nicole's blood on the sock? Weir and I debated it. In the end we didn't. This moral high ground led to a complex report. I still find it hard, today, to amalgamate the information from all the different items and different laboratories. At the time we were unsure whether or not we should multiply the results for different loci from different laboratories where the databases and protocols were different and where independence testing of the various loci between the different laboratories had not been done. We were expecting a severe challenge.[242,415,457,583,767]

I would do things differently now. I routinely use the subpopulation correction and appropriate values for θ. I would still apply sampling error estimation but use the Bayesian posterior rather than the bootstrap. I have no qualms about multiplying results from different laboratories where independence testing had not been done. This latter is largely because I have abandoned any faith that independence testing informs at all about the population genetic model.

3.5.2 The Trial

The trial lasted 133 days, produced 50,000 pages of transcript, called 126 witnesses, and produced 857 pieces of evidence. The defense team, eventually dubbed the "dream team," included Robert Shapiro, Barry Scheck, Johnny Cochran, Peter Neufeld, and William Thompson. Appearing for the prosecution were Marcia Clark, Christopher Darden, Rockne Harmon, and George "Woody" Clark.

The defense hired a jury consultant who found that black, middle-aged women were Mr. Simpson's strongest supporters. Of the 200 African-Americans polled, 44% stated that they had been treated unfairly by the LAPD at least once.[759] The jury included eight blacks, most of them middle-aged women.

Forensic scientists should not allow themselves a view on guilt or innocence. But some of the evidence looked strong. The defense soon undermined much of that.

Detective Furhman was questioned about racism:

Bailey: "You say under oath that you have not addressed any black person as a nigger or spoken about black people as niggers in the past 10 years, Detective Furhman?"

Furhman: "That's what I'm saying."

Later, on September the 5th, a 10-year-old set of tapes made by a North Carolina writer researching racism in the LAPD was played. Furhman could be heard using the word "nigger." Worse, the tapes were littered with gloating admissions that he and other officers had often planted evidence on suspects to secure convictions. There were 42 instances of "nigger" and 18 instances admitting participation in police misconduct in order to incarcerate criminals, including planting evidence. Furhman bragged about stopping interracial couples for no reason, he spoke of his desire to put black people in a pile and burn them, and that he was against having women in the police force because they would not engage in cover-ups.[507] On September 6, Furhman invoked his 5th Amendment rights.

U.S.A. Today and Gannett News Service had previously published a survey from legal and media databases itemizing 85 instances since 1974 of prosecutors knowingly or unknowingly using tainted evidence.[455]

Furhman was not the only detective to come under scrutiny. Detective Vannatter's statement that O.J. was not a suspect on the night of the 12th or the morning of the 13th stretches credibility. He also stated that he had entered Simpson's home without a warrant because of the risk that there was another victim. Vannatter had access to the blood sample from O.J. taken on the 13th that had been handed to him. He had carried it around rather than logging it as police procedures required.[507] The suggestion that O.J.'s blood was planted was strengthened by the "missing" 1.5 ml. Thano Peratis had testified that he had drawn 7.9–8.1 ml. In all, 1 ml was used for DNA testing and the toxicology department measured the remainder on receipt in their section as 5.5 ml. Peratis, by this time too ill to come to court, altered his testimony in a video. This process denied the defense the right of cross examination.

Mr. Fung's testimony lasted three weeks. He had 11 years of forensic experience and had examined 500 scenes. He was questioned about the blanket used to cover Nicole (there was never a suggestion that he had personally placed it over Nicole). Could hairs from the blanket have transferred onto Nicole? He was shown a crime scene photo with his hand ungloved when it should have been gloved. He was questioned about taking only representative samples from the Bronco and the incorrect placing of blood samples into plastic bags where they could deteriorate.

Mazzola, Fung's trainee assistant, was cross examined by Neufeld. She had collected most of the blood samples without supervision. Videotape showed her resting a hand on a dirty footpath, wiping tweezers with a dirty hand, and dropping several blood swabs.

Evidence was produced that the blood on the socks had occurred by "compression transfer," implying that the blood had not got there while O.J.'s foot was inside the sock. There was also the disputed finding of EDTA in the blood from the sock.

Finally, Darden asked Simpson to put on the gloves. To guard against contamination and hazard to Simpson, he donned latex gloves, and then the leather glove. Simpson stated: "They're too tight."

The RFLP technology was not seriously questioned by the defense.

Mullis, the Nobel Prize winning inventor of PCR, stated that he felt the technology was not ready for forensic application. Listing his interests as drug taking, womanizing, and surfing, he was eventually not called by the defense but other witnesses more than adequately spoke of the contamination risks.

Alan Dershowitz, who advised the defense, stated on TV that the probability of a known wife beater actually killing his wife was very small (1/10,000). This statement is somewhat misleading as pointed out by Good.[360,361] Let

B: the event that a man beats his wife.
M: the wife was murdered.

The statistic quoted on TV was close to $\Pr(M \mid B) = 0.0001$.

But the actual evidence is that Mr. Simpson beat his wife AND his wife was murdered. We are interested in the probability that Mr. Simpson is the murderer GIVEN that Nicole was murdered AND Mr. Simpson had beaten her. Let the event that Mr. Simpson is the murderer be *G*. We require $\Pr(G \mid M, B)$, which is quite different to $\Pr(M \mid B)$. Hibbert[405] and Good[360,361] give this as approximately 0.33 revised later to approximately 0.9.

Professor Weir had either advised others or produced much of the statistics himself. In the end, working in his hotel room he produced three and four person calculations for the mixtures in the Bronco.

> The "hard times" referred to in the title of this column apply to what happened next. The time for my scheduled testimony was moved forward two weeks, and I was called to Los Angeles before completing my mixture calculations. I was able to extend my computer program there to handle three unknown contributors instead of the two that had ever been considered before, but was unable to fax my results to colleagues in England for checking because the hotel's computer would not recognize a change in area codes in the United Kingdom. On the afternoon of Thursday June 22, Judge Ito ordered me to perform additional calculations for four unknown contributors before I could testify the following morning! Another late night session with my lap-top computer in a hotel room, and no opportunity for careful checking. In my written report to both prosecution and defense it was obvious that I had left out a term in the calculations — a term that I had correctly included in the calculations I did in my office during normal waking hours. Reviewers of a scientific paper would have noted such an inconsistency and simply called for a correction, but opposing lawyers in a trial are free to use such errors to discredit an expert. Never mind that the errors concerned only a very small number of the calculations, and did not alter the overwhelming evidentiary strength of the matching DNA profiles in all those bloodstains which came from only one person. Subsequently I have developed the algebraic treatment that circumvents the need for those hurried computer calculations…. I do not believe that statisticians should agree to perform detailed analyses in hotel rooms, especially if they are going to be on national TV the next day. (Weir[837] reprinted with permission

from Stats. © 1996 by the American Statistical Association. All rights reserved)

Neufeld cross examining Weir: "The numbers on the board are biased against Mr. Simpson, isn't that correct?"

Weir: "As it turns out, it looks that way."

The number in question was for a mixed stain on the Bronco steering wheel (item 29). The relevant number went from 1 in 3900 to 1 in 1600. Weir had sighted the error himself, made the correction himself, and put the matter before the court. The complexity of four person calculations was substantial.

> I could not repeat the calculations by hand and Steve Knight and Richard Pinchin had to write software to enumerate the large number of possibilities. It took us a long time to repeat Weir's calculation and the relevant exchanges in court were over before we had done this. In the context of the trial the observation that we were fallible counted more than the number itself. To me, of course, this is not news but it does emphasise the value of independent checking. [John Buckleton]

The profiles from item 29, the Bronco steering wheel and the three reference samples, are given in Table 3.9. The 1.3 allele was not observed in item 29; hence, putatively, Ronald was excluded. This implied the presence of an unknown DNA source. However, the spot from the 4 allele was weak and the issue of whether the 1.3 allele had "dropped out" arose. If we accept that the 4 allele is not from Ronald Goldman, then this is the only allele out of 400 from 45 stains not included in one of the principals' profiles.

The press statements were not flattering:

LA Times[578]

> Dry as sand and just as digestible. (Peter Arenella, UCLA law professor)

Table 3.9 Profiles Considered from Item 29

Locus	Item 29	OS	NB	RG
DQα	1.1,1.2,4	1.1,1.2	1.1,1.1	1.3,4
LDLR	AB	AB	AB	AB
GYPA	AB	BB	AB	AA
HBGG	ABC	BC	AB	AA
D7S8	AB	AB	AB	BB
Gc	ABC	BC	AC	AA

Reproduced from Weir[835] with the kind permission of Nature and Professor Weir.

> More mind-numbing statistics of all sizes with little real meaning to the case, even assuming jurors had any clue about their significance. (Myrna Raeder, Professor of Law, Southwestern University)

In the end, the matter was settled largely on other considerations. The prosecution summation included emotive sections such as

> … it is because he hit her in the past. And because he slapped her and threw her out of the house and kicked her and punched her and grabbed her around the neck … and it's because he used a baseball bat to break the windshield of her Mercedes back in 1985. And it's because he kicked her door down in 1993 … It's because of a letter he wrote her … June the 6th, talking about the IRS. It's because he stalked her … and the fuse is burning. … the fuse is getting shorter, the fuse is getting shorter, and there is about to be an explosion, he is about to lose control, and he is about to lose control like he did on those earlier occasions. And sure he didn't kill her on those earlier occasions in October '93 or in 1989. But that was then and back then the fuse was a lot longer. But now the fuse is way short and it is awfully short … . how do we evaluate this, when a man takes a baseball bat to his wife's car and beats the "F" out of it? If nothing else, it sends a message to her. It instills fear, wouldn't you agree? And would you agree it suggests to her that this can happen to you, that maybe you'll be next? That fuse is burning. It's burning in 1985 … the fuse is lit. It's burning, but it's a slow burn. (Darden, closing argument, reprinted with kind permission from Cotterill[197])

Perhaps the best metaphor from the defense alluded to the glove in particular and the evidence in general: "If it doesn't fit, you must acquit."[197] Mr. Simpson was acquitted on October 3, 1995.[835]

Thagard[759] has studied possible lines of reasoning by which the jury may have reached this verdict. He mentions the inference from Nicole's history of cocaine use that drug dealers may have been involved.

Bayesian inference in the hands of Thagard[759] and JavaBayes gives a posterior of 0.72 that Mr. Simpson was guilty and 0.29 to the alternative that drug dealers were the killers. It also assigns a posterior of 0.99 to the proposition that the LAPD framed Mr. Simpson. 0.72 is well below our subjective level for "beyond reasonable doubt," and in our opinion is entirely consistent with acquittal. Three of the jurors, Cooley, Bess, and Rubin-Jackson, described their conclusions as based on reasonable doubt. "I'm sorry, O.J. would have had to go if the prosecution had presented the case differently, without a doubt. As a black woman it would have hurt me. But as a human being, I would have to do what I had to do." (Juror Carrie Bess)[759]

Relatedness

4

JOHN BUCKLETON AND CHRISTOPHER TRIGGS

Contents

4.1 Introduction

In this chapter we discuss the evaluation of the joint and conditional probabilities of obtaining various genotypes for two people who are related, and the effect of this relatedness on the interpretation process. Formulae are given for some common relationships. Most of this work has appeared elsewhere, for instance, in Evett and Weir.[267] Elegant algorithms have been published[86,88,762] that perform these and far more complex analyses. Such probabilities have many uses outside the specific forensic context.

In our forensic work we will often need to consider relatedness. This can occur because of a specific defense such as "my brother committed the crime,"[260] but is becoming increasingly relevant even in the absence of such a specific defense. There are several probabilities that we may be interested in regarding relatives. These would include answers to such questions as:

- What is the probability that a brother would "match?"
- Given that these two individuals match, "What is the probability that they are brothers?"

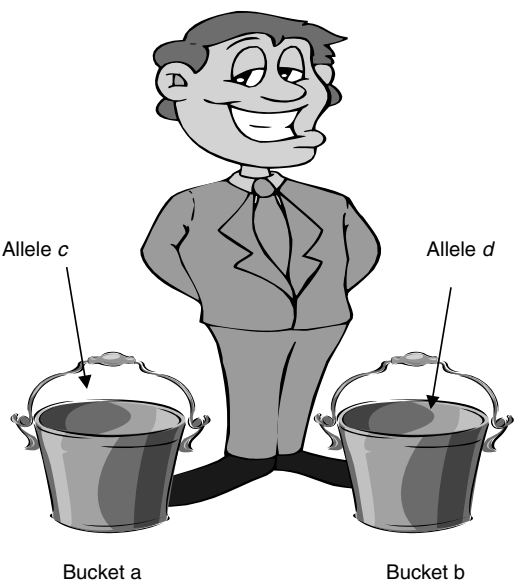
The same methods can be used to evaluate the probabilities associated with these two questions, although they are applied slightly differently. There

are at least three methods of which we are aware. All give the same result. The approach published by Evett[260] is cumbersome and is not widely used. We will discuss here the remaining two methods, both of which utilize the concept of identity by descent (IBD) initially introduced in 1940 by Cotterman[198] and extended by Malecot,[534] Li and Sacks,[512] and Jacquard.[188, 302, 384, 431, 872] Two alleles are said to be IBD if they are the same BECAUSE they are copies of the same ancestral allele.

Consider two people *X* and *Y* (termed a "dyadic" relationship).[761] We can label the alleles at a specific locus for person *X* as (*ab*) and person *Y* as (*cd*). This does not imply that person *X* has genotype *ab*, but rather that we have labeled his two alleles *a* and *b*. In such a case, the labels "*a*," "*b*," "*c*," and "*d*" are referred to as placeholders. The actual allele in place *a* is denoted by an italicized label. See Box 4.1.

Box 4.1 Buckleton's Buckets

This term was coined by Bruce Weir. The distinction between the actual allele and the label of the allele, or placeholder, is one that, in our experience, many readers, students, and teachers find difficult to either understand or communicate. However, it is vitally important to clearly understand the distinction. We may illustrate the concept of placeholders using the following figure. A useful visual metaphor for the label or placeholder is that of a bucket.



This person has two buckets; a and b; they contain the alleles *c* and *d*.

4.2 Conditional Probabilities

These are the probabilities that an untyped relative will have a certain genotype, G_2, given that a typed relative has genotype G_1. Such probabilities can be used to answer most questions of forensic interest, such as "what is the probability that a brother of the matching suspect would also match?" Such conditional probabilities may be developed in two ways, either directly or via the joint probability and the definition of a conditional probability,

$$\Pr(G_2 \mid G_1) = \frac{\Pr(G_2, G_1)}{\Pr(G_1)}$$

Either method has its merits and drawbacks. Both the method of Balding and Nichols[36] and that due to Weir[267] can be used to evaluate the conditional probabilities.

4.2.1 The Method of Balding and Nichols

Any two people possess four alleles at a specific locus. If we consider one of these people, then they may have 0, 1, or 2 alleles IBD with the other person.[a] Following Balding and Nichols, we consider the events:

Z_0: 0 alleles are identical by descent, with probability $\Pr(Z_0)$.
Z_1: 1 allele is identical by descent, with probability $\Pr(Z_1)$.
Z_2: 2 alleles are identical by descent with probability $\Pr(Z_2)$.

Consider the relationship between a parent and their child (Figure 4.1). The child has two alleles, which we have labeled *a* and *b* (completely arbitrarily). By the principle of Mendelian inheritance, we can see that we expect one of these alleles to be IBD with one of the alleles from the parent.

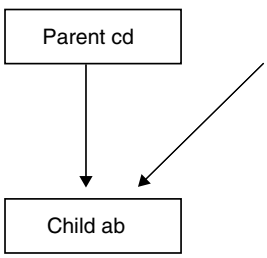


Figure 4.1 A pedigree for a parent and child.

[a] Of course the relationship is reflexive. Balding and Nichols also make the assumption that the two alleles within an individual are not IBD. Thus, their method can only be applied to dyadic relationships and hence cannot handle those situations where one or more of the founders are inbred.

Thus, we can see that for a parent/child relationship $\Pr(Z_1) = 1$ and $\Pr(Z_0) = \Pr(Z_2) = 0$.

Consider now a pair of siblings (Figure 4.2). Each sibling will receive an allele from his or her father. With probability $\frac{1}{2}$, these will be copies of the same allele, and thus IBD. Thus with probability $\frac{1}{2}$, they will not be IBD. Similarly, the probability that the two copies of the maternal allele will be IBD will also be $\frac{1}{2}$. Therefore, both will be IBD with probability $\Pr(Z_2) = \frac{1}{4}$. There are two ways in which a pair of siblings can have one pair of IBD alleles. They may share only the maternal or only the paternal allele and hence $\Pr(Z_1) = \frac{1}{2}$. Similarly, it follows that the probability that they have 0 IBD alleles, $\Pr(Z_0) = \frac{1}{4}$. Similar arguments lead to Table 4.1, which gives the values of $\Pr(Z_0)$, $\Pr(Z_1)$, and $\Pr(Z_2)$ for some of the forensically important relationships between two individuals.

To demonstrate the use of this table, we calculate the conditional probability that a person has genotype *aa* given that his sibling has genotype *ab*. We will write this as $\Pr[aa \mid ab, \text{siblings}]$. We will omit the conditioning on "siblings" for simplicity, except where the omission may cause ambiguity.

Using the law of total probability,[b] this can be written as

$$\Pr[aa|ab] = \Pr[aa|ab, Z_2]\Pr(Z_2) + \Pr[aa|ab, Z_1]\Pr(Z_1) + \Pr[aa|ab, Z_0]\Pr(Z_0)$$

If the two siblings share two pairs of IBD alleles, then they must have the same genotype. Since you cannot obtain the *aa* genotype from an *ab* genotype with two alleles IBD, then $\Pr[aa \mid ab, Z_2] = 0$.

If the two siblings share one pair of IBD alleles, then with probability $\frac{1}{2}\Pr(Z_1)$ the *a* allele in the conditioning genotype is IBD, and we need the other bucket to be filled with the *a* allele by chance. Hence we assign

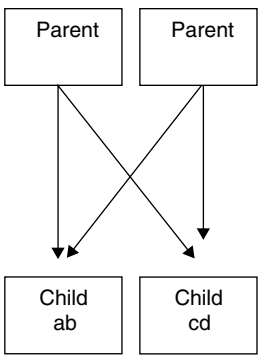


Figure 4.2 A pedigree for siblings.

[b] We assume that the IBD state and the genotype of the conditioning individual are independent, $\Pr[Z_i \mid ab] = \Pr[Z_i]$.

Table 4.1 Probabilities that Two Individuals with a Given Relationship Share 0, 1, or 2 Pairs of IBD Alleles

Relationship	$\Pr(Z_0)$	$\Pr(Z_1)$	$\Pr(Z_2)$
Parent/Child	0	1	0
Full-siblings	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
Half-siblings	$\frac{1}{2}$	$\frac{1}{2}$	0
Grandparent/grandchild	$\frac{1}{2}$	$\frac{1}{2}$	0
Uncle/nephew	$\frac{1}{2}$	$\frac{1}{2}$	0
First cousins	$\frac{3}{4}$	$\frac{1}{4}$	0

$\Pr[aa|ab, Z_1]=p_a$. With probability $\frac{1}{2}\Pr(Z_1)$, the b allele in the conditioning genotype is IBD, but here we have $\Pr[aa|ab, Z_1]=0$ since you cannot obtain the aa genotype if the b allele is IBD.

If the two siblings share zero pairs of IBD alleles, then $\Pr[aa|ab, Z_0] = p_a^2$ since both buckets in the individual whose genotype is in front of the conditioning bar are unconstrained, that is, not determined by any IBD state, and each bucket must be filled with separate copies of the allele a.

This calculation can be set down in a general stepwise process. Table 4.2 illustrates the general algorithm to evaluate $\Pr[G_1|G_2]$, which can easily be implemented in a spreadsheet by following six steps. First lay out a table with four rows, one for each of the cases of two or zero pairs of IBD alleles and two for the case of one pair of IBD alleles. In a column write $\Pr(Z_2)$, $\frac{1}{2}\Pr(Z_1)$, $\frac{1}{2}\Pr(Z_1)$, $\Pr(Z_0)$ with the corresponding values for the "relationship" from Table 4.1. In the next column write the probabilities of observing genotype G_1 given genotype G_2 and the corresponding IBD state. For this column the probability in the Z_2 row will have either a 0 or a 1 in it depending on whether or not the persons before and after the conditioning bar have the same genotype. When the genotype G_2 behind the conditioning bar is a heterozygote, we use two rows for the Z_1 event to account for each allele in G_2 being the allele involved in the IBD pair. When G_2 is homozygous, these two rows will contain the same value. The Z_0 row describes the event when the two genotypes have no IBD alleles. Initially we use the "product rule" to evaluate $\Pr[G_1|G_2, Z_0]$. In the final column of the table, form the product of the previous two columns. Sum the final column to give the required probability.

Application of this method leads to the formulae given in Table 4.3. The multilocus probability estimate is calculated by multiplying the single-locus probabilities.

Table 4.2 Calculation of Pr [*aa* | *ab*] for Siblings

Pairs of IBD Alleles	$\Pr[Z_i]$	$\Pr[Z_i]$ Siblings	$\Pr(ab \mid aa, Z_i)$	Product
2	$\Pr(Z_2)$	$\frac{1}{4}$	0	0
1	$\frac{1}{2}\Pr(Z_1)$	$\frac{1}{4}$	p_a	$\frac{p_a}{4}$
	$\frac{1}{2}\Pr(Z_1)$	$\frac{1}{4}$	0	0
0	$\Pr(Z_0)$	$\frac{1}{4}$	p_a^2	$\frac{p_a^2}{4}$
Sum				$\frac{p_a(1+p_a)}{4}$

Table 4.3 Conditional Probabilities for the Genotype of an Untyped Person G_1 Given the Genotype G_2 of a Typed Relative

Genotype of Typed Person, G_2	Genotype for the Untyped Relative, G_1	Postulated Relationship		
		Full-Siblings	Cousins	Half-Siblings, Uncle/Nephew, Grandparent/ Grandchild
aa	*aa*	$\frac{(1+p_a)^2}{4}$	$\frac{p_a(1+3p_a)}{4}$	$\frac{p_a(1+p_a)}{2}$
	bb	$\frac{p_b^2}{4}$	$\frac{3p_b^2}{4}$	$\frac{p_b^2}{2}$
	ab	$\frac{p_b(1+p_a)}{2}$	$\frac{p_b(1+6p_a)}{4}$	$\frac{p_b(1+2p_a)}{2}$
	bc	$\frac{p_b p_c}{2}$	$\frac{3p_b p_c}{2}$	$p_b p_c$
ab	*aa*	$\frac{p_a(1+p_a)}{4}$	$\frac{p_a(1+6p_a)}{8}$	$\frac{p_a(1+2p_a)}{4}$
	ab	$\frac{1+1\,p_a+1\,p_b+2p_a\,p_b}{4}$	$\frac{p_a+p_b+12p_a p_b}{8}$	$\frac{p_a+p_b+4p_a p_b}{4}$
	ac	$\frac{p_c(1+2p_a)}{4}$	$\frac{p_c(1+12p_a)}{8}$	$\frac{p_c(1+4p_a)}{4}$
	cc	$\frac{p_c^2}{4}$	$\frac{3p_c^2}{2}$	$\frac{p_c^2}{2}$
	cd	$\frac{p_c p_d}{2}$	$\frac{3p_c p_d}{2}$	$p_c p_d$

Table 4.4 Conditional Calculation for Pr[*aa*|*ab*] for Brothers Including the Subpopulation Correction

Pairs of IBD Alleles	$\Pr[Z_i]$	$\Pr[Z_i]$ Siblings	$\Pr(ab\|aa, Z_i)$	Product
2	Z_2	$\frac{1}{4}$	0	0
1	$\frac{1}{2}Z_1$	$\frac{1}{4}$	$\frac{\theta + (1-\theta)p_a)}{1+\theta}$	$\frac{\theta + (1-\theta)p_a)}{4(1+\theta)}$
	$\frac{1}{2}Z_1$	$\frac{1}{4}$	0	0
0	Z_0	$\frac{1}{4}$	$(\theta + (1-\theta)p_a)$ $\times \frac{(2\theta + (1-\theta)p_a)}{(1+\theta)(1+2\theta)}$	$(\theta + (1-\theta)p_a)$ $\times \frac{(2\theta + (1-\theta)p_a)}{4(1+\theta)(1+2\theta}$
Sum				$\frac{\theta + (1-\theta)p_a)}{4(1+\theta)}$ $\times \left(1 + \frac{(2\theta + (1-\theta)p_a)}{(1+2\theta)}\right)$

To drop the assumption of independence in the calculation of $\Pr[G_1|G_1, Z_i]$, we can introduce the conditional probability[36] at step 3. The example above, evaluating $\Pr[G_1 = aa|G_2 = ab]$ for a pair of siblings and involving the subpopulation correction, is given in Table 4.4.

This method of calculation leads to the formulae given in Table 4.5.[c]

4.2.2 The Method of Weir

A more precise nomenclature was given by Weir.[267,836] It was based on four allele descent measures. This requires a labeling of the alleles in each person. We name the alleles in person 1 as ab, and in person 2 as cd (Table 4.6), where a, b, c, and d are placeholders (buckets). These allele designations must be tied to the pedigree to give the values given by Weir. Consider, for example, the case of half-siblings (Figure 4.3). In this figure we assign allele a as coming from person *G*, allele b as coming from person *H*, c from *H*, and d from *I* as indicated. If we assume that *H* is the father and *G* and *I* are mothers, then we are labeling b and c as paternal alleles, and a and d as maternal alleles. This labeling of the alleles is arbitrary and troubles many people. However, any other arrangement can be used and produces the same result. We need to consider 15 possible IBD states for the four alleles. For example, the term δabcd represents the probability that all the alleles a, b, c, and d are IBD.

[c] We thank Lindsey Foreman for checking these formulae.

Table 4.5 Conditional Probabilities for Some Relatives Including the Subpopulation Correction

Genotype of Typed person, G_2	Genotype for the Untyped Sibling, G_1	Siblings
aa	*aa*	$\frac{1}{4}\left(1+\frac{2(2\theta+(1-\theta)p_a)}{1+\theta}+\frac{(2\theta+(1-\theta)p_a)(3\theta+(1-\theta)p_a)}{(1+\theta)(1+2\theta)}\right)$
	bb	$\frac{(1-\theta)p_b(\theta+(1-\theta)p_b)}{4(1+\theta)(1+2\theta)}$
	ab	$\frac{(1-\theta)p_b}{2(1+\theta)}\left(1+\frac{2\theta+(1-\theta)p_a}{(1+2\theta)}\right)$
	bc	$\frac{(1-\theta)^2 p_b p_c}{2(1+\theta)(1+2\theta)}$
ab	*aa*	$\frac{\theta+(1-\theta)p_a}{4(1+\theta)}\left(1+\frac{(2\theta+(1-\theta)p_a)}{(1+2\theta)}\right)$
	ab	$\frac{1}{4}\left(1+\frac{2\theta+(1-\theta)(p_a+p_b)}{(1+\theta)}+\frac{2(\theta+(1-\theta)p_a)(\theta+(1-\theta)p_b)}{(1+\theta)(1+2\theta)}\right)$
	ac	$\frac{(1-\theta)p_c}{4(1+\theta)}\left(1+\frac{2(\theta+(1-\theta)p_a)}{(1+2\theta)}\right)$
	cc	$\frac{(\theta+(1-\theta)p_c)(1-\theta)p_c}{4(1+\theta)(1+2\theta)}$
	cd	$\frac{(1-\theta)^2 p_c p_d}{2(1+\theta)(1+2\theta)}$

Genotype of Typed person, G_2	Genotype for the Untyped Sibling, G_1	Half Siblings, Uncle/Nephew, Grandparent/Grandchild
aa	*aa*	$\frac{2\theta+(1-\theta)p_a}{2(1+\theta)}\left(1+\frac{3\theta+(1-\theta)p_a}{1+2\theta}\right)$
	bb	$\frac{(1-\theta)\,p_b(\theta+(1-\theta)p_b)}{2(1+\theta)(1+2\theta)}$
	ab	$\frac{(1-\theta)p_b}{2(1+\theta)}\left(1+\frac{2(2\theta+(1-\theta)p_a)}{(1+2\theta)}\right)$

	bc	$\frac{(1-\theta)^2 p_b p_c}{(1+\theta)(1+2\theta)}$
ab	*aa*	$\frac{\theta+(1-\theta)p_a}{4(1+\theta)}\left(1+\frac{4\theta+2(1-\theta)p_a}{(1+2\theta)}\right)$
	ab	$\frac{1}{4(1+\theta)}\left(2\theta+(1-\theta)(p_a+p_b)+\frac{4(\theta+(1-\theta)p_a)(\theta+(1-\theta)p_b)}{(1+2\theta)}\right)$
	ac	$\frac{(1-\theta)p_c}{4(1+\theta)}\left(1+\frac{4(\theta+(1-\theta)p_a)}{(1+2\theta)}\right)$
	cc	$\frac{(\theta+(1-\theta)p_c)(1-\theta)p_c}{2(1+\theta)(1+2\theta)}$
	cd	$\frac{(1-\theta)^2 p_c p_d}{(1+\theta)(1+2\theta)}$

Genotype of Typed person, G_2	Genotype for the Untyped Sibling, G_1	Cousins
aa	*aa*	$\frac{2\theta+(1-\theta)p_a}{4(1+\theta)}\left(1+\frac{3(3\theta+(1-\theta)p_a)}{(1+2\theta)}\right)$
	bb	$\frac{3}{4}\left(\frac{(1-\theta)p_b(\theta+(1-\theta)p_b)}{(1+\theta)(1+2\theta)}\right)$
	ab	$\frac{(1-\theta)p_b}{4(1+\theta)}\left(1+\frac{6(2\theta+(1-\theta)p_a)}{(1+2\theta)}\right)$
	bc	$\frac{3(1-\theta)^2 p_b p_c}{2(1+\theta)(1+2\theta)}$
ab	*aa*	$\frac{\theta+(1-\theta)p_a}{8(1+\theta)}\left(1+\frac{6(2\theta+(1-\theta)p_a)}{(1+2\theta)}\right)$
	ab	$\frac{1}{8(1+\theta)}\left(2\theta+(1-\theta)(p_a+p_b)+\frac{12(\theta+(1-\theta)p_a)(\theta+(1-\theta)p_b)}{(1+2\theta)}\right)$

(*Continued*)

Table 4.5 (*Continued*)

ac	$\frac{(1-\theta)p_c}{8(1+\theta)}\left(1+\frac{12(\theta+(1-\theta)p_a)}{(1+2\theta)}\right)$
cc	$\frac{3}{4}\left(\frac{(\theta+(1-\theta)p_c)(1-\theta)p_c)}{(1+\theta)(1+2\theta)}\right)$
cd	$\frac{3(1-\theta)^2 p_c p_d}{2(1+\theta)(1+2\theta)}$

Table 4.6 Four Allele Descent Measures Following Weir

Alleles IBD If the alleles Are Not Mentioned They Are Not IBD	Probabilities			
	Term	Full-Siblings	Cousins	Half-Siblings
None	δ0	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{1}{2}$
a≡b	δab			
c≡d	δcd			
a≡c	δac	$\frac{1}{4}$		
a≡d	δad			
b≡c	δbc		$\frac{1}{4}$	$\frac{1}{2}$
b≡d	δbd	$\frac{1}{4}$		
a≡b≡c	δabc			
a≡b≡d	δabd			
a≡c≡d	δacd			
a≡b, c≡d	δab, cd			
a≡c, b≡d	δac, bd	$\frac{1}{4}$		
a≡d, b≡c	δad, bc			
a≡b≡c≡d	δabcd			

The sign ≡ is often used to signify that two alleles are identical by descent, IBD. Hence a≡b means that the alleles with labels a and b are IBD. Consider the relationship of a parent and a child (Figure 4.1). We have labelled the alleles in the parent c and d and those in the child a and b. The laws of Mendelian inheritance state that one of the alleles labeled a or b must be a copy of one of the alleles labeled c or d. The actual allele that is a copy of the parental allele is IBD.

a ≡ b implies that alleles a and b are IBD but not IBD with alleles c or d, nor is c IBD with d.

The same procedure for developing the conditional probability directly as given under Balding and Nichols' method also applies here, except that part

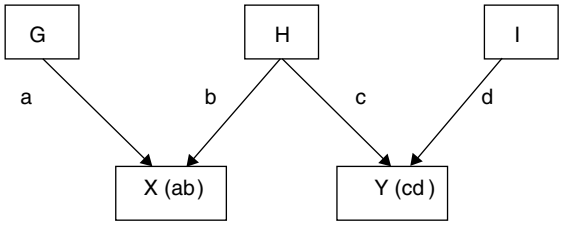


Figure 4.3 A pedigree for half-siblings (X and Y).

of the ambiguity regarding Z_1 is resolved using the Weir approach. Table 4.7 reproduces the calculation for Pr($aa|aa$) for full-siblings. The method is analogous to the approach given by Balding and Nichols, but is more versatile.

Even using the Weir nomenclature, there is a wrinkle that one needs to be aware of when only one pair of alleles is IBD. This occurs if we try to write down the conditional probability directly, but does not occur if we proceed via joint probabilities (see later) and only when the conditioning profile is heterozygotic. This wrinkle can be demonstrated using an example involving cousins. Consider the probability Pr($ac|ab$) for cousins (Table 4.8).

Following Weir,[267,841,844] we describe the process. It is necessary to consider the cell marked with "?". For this cell, we know that the alleles marked by the b and d placeholders are IBD. But this does not inform us whether the "a" allele is the one involved in this IBD state. If the "b" allele is involved, then we cannot obtain an "ac" genotype. The "a" allele is the one involved $\frac{1}{2}$ of the time. This results in a value of

$$\frac{\Pr(c \mid ab, \text{unrelated})}{2} = \frac{(1-\theta)p_c}{2(1+\theta)}$$

for the "?" cell. By multiplying across the rows and adding downward, Pr($ac \mid ab$) for cousins is

$$\Pr(ac \mid ab, \text{cousins}) = \frac{3\ \Pr(ac \mid ab, \text{unrelated})}{4} + \frac{1}{4} \times \frac{\Pr(c \mid ab, \text{unrelated})}{2}$$

$$= \frac{3 \times 2(1-\theta)p_c\,[\theta+(1-\theta)p_a]}{4(1+\theta)(1+2\theta)} + \frac{(1-\theta)p_c}{8(1+\theta)}$$

$$= \frac{(1-\theta)p_c}{8(1+\theta)}\left(1 + \frac{12(\theta+(1-\theta)p_a)}{(1+2\theta)}\right)$$

We do not pursue this approach in depth here as it is much more extensively and better described in Evett and Weir.[267] The use of such descent measures permits us to consider much more complex pedigrees than merely dyadic relationships. However, it leads to the same results as the Balding and Nichols' method for the simple pedigrees considered here.

Table 4.7 Calculation of Pr(*aa*|*aa*, siblings) Following the Weir Methodology

Alleles IBD	Term	Probabilities	
		Full-Siblings	
None	δ0	$\frac{1}{4}$	Pr(*aa*|*aa*, unrelated)[d]
a≡c	δac	$\frac{1}{4}$	Pr(*a*|*aa*, unrelated)
b≡d	δbd	$\frac{1}{4}$	Pr(*a*|*aa*, unrelated)
a≡c, b≡d	δac, bd	$\frac{1}{4}$	1
Sum of the products		$\frac{1}{4}\left(1+\frac{2(2\theta+(1-\theta)p_a)}{1+\theta}+\frac{(2\theta+(1-\theta)p_a)(3\theta+(1+\theta)p_a)}{(1+\theta)(1+2\theta)}\right)$	

Table 4.8 Calculation of Pr(*ac*|*ab*, cousins) for Cousins Following Weir Methodology

Alleles IBD	Term	Probabilities	
		Cousins	
None	δ0	$\frac{3}{4}$	Pr (*ac*|*ab*, unrelated)[e]
b≡d	δbd	$\frac{3}{4}$	?
Sum of the products			

Example 4.1. Suppose that we have typed a suspect at, say, six loci. Suppose we were interested in the question: "What is the probability that a brother of the suspects would also match. The calculation is laid out in Table 4.9.

Therefore, we expect that the probability that a brother of the suspect would have this profile is 0.0198 if we estimate using the product rule and 0.0204 if we incorporate a subpopulation correction into the brother's formula ($\theta = 0.01$).

[d] Unrelated is shorthand for "unrelated members of the same subpopulation."
[e] This is Pr(*ac* | *ab*) for unrelated members of the same subpopulation. This is written as

$$\frac{2(1-\theta)p_c\,[\theta+(1-\theta)p_a]}{(1+\theta)(1+2\theta)}$$

Table 4.9 An Example of the Use of the Brother's Formulae

Locus	Suspect's Profile	Formula for a Matching Brother Assuming Independence of Alleles $\theta = 0$	Formula for a Matching Brother Using the θ Correction $\theta > 0$
LDLR	A, B	$\frac{1 + p_A + p_B + 2p_A p_B}{4}$	$\frac{1}{4}\left(1 + \frac{2\theta + (1-\theta)(p_A + p_B)}{(1+\theta)} + \frac{2(\theta + (1-\theta)p_A)(\theta + (1-\theta)p_B)}{(1+\theta)(1+2\theta)}\right)$
GYPA	B, B	$\frac{(1 + p_B)^2}{4}$	$\frac{1}{4}\left(1 + \frac{2(2\theta + (1-\theta)p_B)}{1+\theta} + \frac{2(\theta + (1-\theta)p_B)(3\theta + (1-\theta)p_B)}{(1+\theta)(1+2\theta)}\right)$
HBGG	B, C	$\frac{1 + p_B + p_C + 2p_B p_C}{4}$	$\frac{1}{4}\left(1 + \frac{2\theta + (1-\theta)(p_B + p_C)}{(1+\theta)} + \frac{2(\theta + (1-\theta)p_B)(\theta + (1-\theta)p_C)}{(1+\theta)(1+2\theta)}\right)$
D7S8	A, C	$\frac{1 + p_A + p_C + 2p_A p_C}{4}$	$\frac{1}{4}\left(1 + \frac{2\theta + (1-\theta)(p_A + p_C)}{(1+\theta)} + \frac{2(\theta + (1-\theta)p_A)(\theta + (1-\theta)p_C)}{(1+\theta)(1+2\theta)}\right)$
GC	B, C	$\frac{1 + p_B + p_C + 2p_B p_C}{4}$	$\frac{1}{4}\left(1 + \frac{2\theta + (1-\theta)(p_B + p_C)}{(1+\theta)} + \frac{2(\theta + (1-\theta)p_B)(\theta + (1-\theta)p_C)}{(1+\theta)(1+2\theta)}\right)$
DQα	1.1, 1.2	$\frac{1 + p_{1.1} + p_{1.2} + 2p_{1.1}p_{1.2}}{4}$	$\frac{1}{4}\left(1 + \frac{2\theta + (1-\theta)(p_{1.1} + p_{1.2})}{(1+\theta)} + \frac{2(\theta + (1-\theta)p_{1.1})(\theta + (1-\theta)p_{1.2})}{(1+\theta)(1+2\theta)}\right)$

Locus	Profile	Allele Probabilities		Numerical value for a brother using Cellmark Diagnostic's African American allele probabilities $\theta = 0.00$	$\theta = 0.01$
LDLR	A, B	0.296	0.704	0.604	0.604
GYPA	B, B	0.531		0.586	0.594
HBGG	B, C	0.260	0.332	0.441	0.444

(*Continued*)

Table 4.9 (*Continued*)

Locus	Mr. Simpson's Profile	Formula for a Matching Brother Assuming Independence of Alleles $\theta = 0$		Formula for a Matching Brother Using the θ Correction $\theta > 0$	
D7S8	A, C	0.679	0.321	0.609	0.609
GC	B, C	0.673	0.214	0.544	0.545
DQα	1.1, 1.2	0.163	0.276	0.382	0.386
Product				0.0198	0.0204

4.3 Joint Probabilities

In this section we consider the probability that a pair of persons would have genotypes G_1 and G_2. These joint probabilities are useful for answering questions of the type: "We have a match on the database between two males; what is the support for the suggestion that they are brothers?" Consider that we have two hypotheses such as:

B: These two males are brothers.
U: These two males are unrelated.

Then we can approach the problem by calculating the likelihood ratio

$$LR = \frac{\Pr\,(G_1,G_2\,|B)}{\Pr(G_1,G_2|U)}$$

which leads us to the set of joint probabilities under the related and the unrelated conditions. Li et al.[513] describe a similarity index approach, which is not discussed here since a likelihood ratio approach is almost universally preferred.

Tables of joint probabilities have been produced[267,836] that typically give the joint probability for "unordered" pairs. Examples are given below (Table 4.10). These generally agree with equivalent formulae given in Evett and Weir,[267] except for the entry for the joint probability of two brothers with genotypes *aa*, *bc* where we differ by a factor of 2.

Table 4.10 refers to pairs regardless of order. Hence the genotype pairs *aa*, *ab* and *ab*, *aa* are equivalent. However, care must be taken when progressing to multilocus genotypes, and it is our opinion that working with ordered pairs is safer.

If we think of G_1 and G_2 as an unordered pair of multilocus genotypes, we note that

$$\Pr(G_1, G_2) = 2\prod_l \Pr(G_1^l, G_2^l)$$

Table 4.10 Joint Probabilities of Observing Genotypes G_1, G_2 Under Specified Relationships Between Individuals Without Regard to Order and Utilizing the Product Rule

G_1	G_2	Pr[G_1, G_2|full-siblings]	Pr[G_1, G_2|half-siblings]	Pr[G_1, G_2|unrelated]
aa	*aa*	$\frac{1}{4}p_a^2(1+p_a)^2$	$\frac{1}{2}p_a^3(1+p_a)$	p_a^4
aa	*ab*	$p_a^2p_b(1+p_a)$	$p_a^2p_b(1+2p_a)$	$4p_a^3p_b$
aa	*bb*	$\frac{1}{2}p_a^2p_b^2$	$p_a^2p_b^2$	$2p_a^2p_b^2$
aa	*bc*	$p_a^2p_bp_c$	$2p_a^2p_bp_c$	$4p_a^2p_bp_c$
ab	*ab*	$\frac{1}{2}p_ap_b(1+p_a+p_b+2p_ap_b)$	$\frac{1}{2}p_ap_b(p_a+p_b+4p_ap_b)$	$4p_a^2p_b^2$
ab	*ac*	$p_ap_bp_c(1+2p_a)$	$p_ap_bp_c(1+4p_a)$	$8p_a^2p_bp_c$
ab	*cd*	$2p_ap_bp_cp_d$	$4p_ap_bp_cp_d$	$8p_ap_bp_cp_d$

where $\Pr(G_1^l, G_2^l)$ is the probability of the ordered pair of genotypes. If the formulae for the unordered pairs at each locus are used such as given on p. 206 of Weir[836] or on p. 116 of Evett and Weir,[267] there are likely to be too many factors of 2. This typically does not affect any *LR* calculated from these terms as the same excess of 2's appears in the numerator and denominator. Table 4.11 gives the probability of ordered pairs of genotypes evaluated using the product rule.

Example 4.2. In this example, we also consider a pair of male profiles where we are interested in the two propositions:

B: These two males are brothers.
U: These two males are unrelated.

Loucs	G_1	G_2	$\Pr(G_1^l, G_2^l\|B)$ at Locus l Using the Ordered Sets	$\Pr(G_1^l, G_2^l\|U)$ at Locus l Using the Ordered Sets
1	*aa*	*aa*	$\frac{p_a^2(1+p_a)^2}{4}$	p_a^4
2	*cd*	*cd*	$\frac{p_cp_d(1+p_c+p_d+2p_cp_d)}{2}$	$4p_c^2p_d^2$
3	*ef*	*eg*	$\frac{p_ep_fp_g(1+2p_e)}{2}$	$4p_e^2p_fp_g$
4	*hh*	*hi*	$\frac{p_h^2p_i(1+p_h)}{2}$	$2p_h^3p_i$

If we proceed to calculate the joint probability given that they are (ordered) brothers, then it is given by the product of terms in the column

Table 4.11 Genotypes G_1 and G_2 in Order

G_1	G_2	Pr(G_1, G_2|brothers)	Cousins
aa	*aa*	$\frac{p_a^2(1+p_a)^2}{4}$	$\frac{p_a^3(1+3p_a)}{4}$
	bb	$\frac{p_a^2p_b^2}{4}$	$\frac{3p_a^2p_b^2}{4}$
	ab	$\frac{p_a^2p_b(1+p_a)}{2}$	$\frac{p_a^2p_b(1+6p_a)}{4}$
	bc	$\frac{p_a^2p_bp_c}{2}$	$\frac{3p_a^2p_bp_c}{2}$
ab	*aa*	$\frac{p_a^2p_b(1+p_a)}{2}$	$\frac{p_a^2p_b(1+6p_a)}{4}$
	ab	$\frac{p_ap_b(1+p_a+p_b+2p_ap_b)}{2}$	$\frac{p_ap_b(p_a+p_b+12p_ap_b)}{4}$
	ac	$\frac{p_ap_bp_c(1+2p_a)}{2}$	$\frac{p_ap_bp_c(1+12p_a)}{4}$
	cc	$\frac{P_aP_bP_c^2}{2}$	$\frac{3p_ap_bp_c^2}{2}$
	cd	$p_ap_bp_cp_d$	$3p_ap_bp_cp_d$
G_1	G_2	Pr(G_1, G_2|siblings)	Unrelated
aa	*aa*	$\frac{P_a^3(1+P_a)}{2}$	p_a^4
	bb	$\frac{p_a^2p_b^2}{2}$	$p_a^2p_b^2$
	ab	$\frac{p_a^2p_b\,(1+2p_a)}{2}$	$2p_a^3p_b$
	bc	$p_a^2p_bp_c$	$2p_a^2p_bp_c$
ab	*aa*	$\frac{p_a^2p_b(1+2p_a)}{2}$	$2p_a^3p_b$
	ab	$\frac{p_ap_b(p_a+p_b+4p_ap_b)}{2}$	$4p_a^2p_b^2$
	ac	$\frac{p_ap_bp_c(1+4p_a)}{2}$	$4p_a^2p_bp_c$
	cc	$p_ap_bp_c^2$	$2p_ap_bp_c^2$
	cd	$2p_ap_bp_cp_d$	$4p_ap_bp_cp_d$

$\Pr(G^l_1,G^l_2|B)$. Equally the joint probability given that they are unrelated (ordered) is given by the product of terms in the column $\Pr(G^l_1, G^l_2|U)$. If the unordered pairs from Evett and Weir are used, it is likely that both terms (in this example) will be too large by a factor of 2.

The correct likelihood ratio is obtained using the formulae for unordered pairs of genotypes because both the numerator and denominator will be incorrect by the same factor. Rather than use "joint probabilities," the same problem can be solved by noting

$$LR = \frac{\Pr(G_1,G_2|B)}{\Pr(G_1,G_2|U)} = \frac{\Pr(G_1|G_2, B)}{\Pr(G_1|U)} = \frac{\Pr(G_2|G_1, B)}{\Pr(G_2|U)}$$

and utilizing the conditional probabilities.

4.4 The Unifying Formula

The models that were discussed in Chapter 3 (the product rule and the subpopulation model) are attempts to calculate a match probability within large groups of unrelated or loosely related people. Neither of these models takes direct account of close relatives. As we add more and more loci, the effect of close relatives becomes increasingly important. It becomes necessary to amalgamate the estimates given by the population genetic models with those for close relatives.[359,501,502,665]

This can be examined by examining the probability $\Pr(H_1|G_c, G_s)$, the derivation of which was given in Box 2.4.

$$\Pr(H_1|G_s,G_c) = \frac{1}{1 + \sum_{i=2}^{N} \Pr(G_c|G_s, H_i)w_i} \qquad (2.4) \text{ (repeated)}$$

Here w_i can be regarded as a weighting function that expresses how much more or less probable the ith person is than the suspect to have left the crime stain based on the non-DNA evidence only. Equation (2.4) is very instructive for our thinking, but it is unlikely to be directly useful in court. This is because the terms w_i relate to the ratio of the prior probability that the ith person is the source of the DNA to the prior probability that the suspect is the source of the DNA. As discussed in Chapter 2, the introduction of such considerations into court testimony presented by a forensic scientist is unlikely to be permitted. However, such an approach may be possible if the court supplies their view of the prior.

In the likely absence of such priors, we suggest that this unifying equation should be used to test various forensic approaches and to instruct our thinking. We start by considering every person in the population to have the same

prior. This means that, before examining the DNA evidence, every person is equally likely to be the source of the DNA. As discussed previously, this assumption is very unlikely to be true, but is likely to be neutral or conservative except in some cases, which will be discussed later. This assumption gives

$$\Pr(H_1|G_s,G_c) = \frac{1}{1 + \sum_{i=2}^{N} \Pr(G_c|G_s, H_i)} \tag{4.1}$$

It can be seen that the posterior probability that the suspect is the source of the DNA $\Pr(H_1|G_s, G_c)$ will be close to 1 whenever

$$\sum_{i=2}^{N} \Pr(G_c|G_s, H_i)$$

is small. We propose to examine this latter term.

The people $P_2, \ldots, P_N$ are the other people in the population who are not the suspect. Following Balding,[34] let us assume a coarse division of the population such that one of these is a brother of the suspect, 6 are cousins, 10 million are members of the subpopulation, and the remaining 240 million are unrelated persons.[f]

Table 4.12 gives a spreadsheet layout to examine the behavior of the term

$$\sum_{i=1}^{N} \Pr(G_c|G_s, H_i)$$

for $\theta = 0.03$. In this table we have simulated typical allele probabilities and calculated the probability assignment for different degrees of relatedness. The central portion of the table is removed for formatting reasons.

It can be seen that in this example the contribution of the single brother is by far the largest, and therefore largely determines the weight of the evidence against the suspect. This is true for almost any configuration of allele probabilities. In almost all cases, the contribution from unrelated people and cousins is very minimal.

This formula is based largely on concepts regarding Mendelian inheritance, and to a lesser extent on population genetic considerations. It is therefore not reliant to any serious extent on the database. We therefore see that a coherent approach to interpreting DNA evidence is based largely on our understanding of formal mathematics and genetics, particularly when possible brothers are included in the interpretive framework.

[f] The customary assumption, for instance stated on p. 41 of Evett and Weir, that each person has the same prior is unnecessary if we treat the population as being coarsely divided into groups of people to whom we assign the same match probability. In this case, we only need to assume that the average prior for each group is equal, not the specific prior for each individual.

Table 4.12 Examining the Contribution of the Term to the Posterior Probability for Different Degrees of Relatedness

Locus	Pr(Allele 1)	Pr(Allele 2)	Product Rule	Subpopulation	Cousin	Brother	
1	0.07	0.18	0.026	0.038	0.051	0.320	
2	0.20	0.12	0.045	0.057	0.073	0.339	
				⋮			
12	0.03		0.001	0.009	0.008	0.265	
13	0.09		0.009	0.025	0.030	0.300	
Probability assignment			4.69E−24	1.95E−20	3.85E−19	2.67E−07	
Number of persons			240,000,000	10,000,000	6	1	$\sum_{i=2}^{N} \Pr(G_c \mid G_s, H_i)$
Product of the number of persons and the probability Contribution to the term			1.12E−15	1.95E−13	2.31E−18	2.67E−07	2.67E−07
$\sum_{i=2}^{N} \Pr(G_c \mid G_s, H_i)$			0.0000004%	0.0000729%	0.0000000%	99.9999267%	

If in the example above, the single brother of the suspect is eliminated by, say, genetic typing, then the subpopulation contribution assumes the position of the major contributor and the product rule contribution is again largely irrelevant. This is shown in Table 4.13. Again a set of simulated allele probabilities is used and the middle portion of the table is removed for formatting reasons.

Cases where this approach is used are likely to be nonconservative: The examination of the unifying formula given above assumed that every person in the population had the same prior probability of being the offender. It was earlier stated that this assumption is neutral or conservative, except in a few circumstances. These circumstances largely occur when a sibling or siblings have a high prior or when the suspect himself has a low prior.

Examination using this formula directs our attention strongly in the interpretation of multilocus DNA evidence toward a noneliminated sibling whether or not there is any particular reason to suspect a sibling. If the suspect has no sibling or if his siblings are eliminated, then attention is focused on other members of the subpopulation whether or not there is any special reason to suspect these persons.

V. Effect of Linkage: We consider here the effect of linkage on the joint and conditional probabilities for relatives. No extended derivation is given here, as for the purposes of this chapter all that is required are the results. The derivation is to appear in a future paper.[109] We report only the matching pairs: that is, sibs or half-sibs matching at both loci. The mismatching pairs may be necessary in some applications and can be developed by the method that is to appear.

4.4.1 Full-Siblings

Consider two siblings who are both PQ at locus 1 and UV at another linked locus, 2. Let the allele probabilities for allele P be p, Q be q, U be u, and V be v. Let the recombination fraction be R. Then the joint probability is

$$\frac{pquv}{2}\begin{Bmatrix}[2+(p+q)(u+v)+8\,pquv][R^2+(1-R)^2]^2\\ +4[4pq+(p+q)(u+v)+4uv]R^2(1-R)^2\\ +4[(p+q)\,(2uv+1)+(2pq+1)(u+v)]R(1-R)\,[R^2+(1-R)^2]\end{Bmatrix}$$

The conditional probability that the second sibling is PQ at one locus and UV at another linked locus GIVEN that the first sibling is this genotype is

$$\frac{1}{8}\begin{Bmatrix}[2+(p+q)(u+v)+8\,pquv][R^2+(1-R)^2]^2\\ +4[4pq+(p+q)(u+v)+4uv]R^2(1-R)^2\\ +4[(p+q)(2uv+1)+(2pq+1)(u+v)]R(1-R)[R^2+(1-R)^2]\end{Bmatrix}$$

Table 4.13 Contribution to the Term If Brother/Sibling Is Eliminated

Locus	Pr(Allele 1)	Pr(Allele 2)	Product Rule	Subpopulation	Cousin	Brother	
1	0.16	0.17	0.054	0.066	0.082	0.346	
2	0.03	0.08	0.004	0.011	0.017	0.279	
			⋮				
12	0.07		0.005	0.019	0.022	0.288	
13	0.06		0.004	0.016	0.017	0.281	
Probability assignment			1.17E−25	7.87E−22	4.64E−20		
Number of persons			240,000,000	10,000,000	6		$\sum_{i=2}^{N} \Pr(G_c \mid G_s, H_i)$
Product of the number of persons and the probability			2.80E−17	7.87E−15	2.78E−19		5.51E−13
Contribution to the term $\sum_{i=2}^{N} \Pr(G_c \mid G_s, H_i)$			0.354%	99.643%	0.004%		

Consider two siblings who are both *PP* at locus 1 and *UU* at locus 2. The joint probability is

$$\frac{p^2u^2}{4}[(pu+1)(R^2+(1-R)^2)+2(p+u)R(1-R)]^2$$

and the conditional probability is

$$\frac{1}{4}[(pu+1)(R^2+(1-R)^2)+2(p+u)R(1-R)]^2$$

Consider two siblings who are both *PP* at locus 1 and *UV* at locus 2. The joint probability is

$$\frac{p^2uv}{2}\left\{\begin{array}{l}[1+p\,(u+v)+2p^2uv][R^2+(1-R)^2]^2\\ +4[p(u+v)+p^2+2uv]R^2(1-R)^2\\ +2[2p(1+2uv)+(1+p^2)(u+v)]R(1-R)[R^2+(1-R)^2]\end{array}\right\}$$

and the conditional probability is

$$\frac{1}{4}\left\{\begin{array}{l}[1+p\,(u+v)+2p^2uv][R^2+(1-R)^2]^2\\ +4[p\,(u+v)+p^2+2uv]\,R^2(1-R)^2\\ +2[2p(1+2uv)+(1+p^2)\,(u+v)]R(1-R)[R^2+(1-R)^2]\end{array}\right\}$$

Consider two siblings that are both *PQ* at locus 1 and *UU* at locus 2. The joint probability is

$$\frac{pqu^2}{2}\left\{\begin{array}{l}[1+(p+q)u+2pqu^2][R^2+(1-R)^2]^2\\ +4[2pq+(p+q)u+u^2]R^2(1-R)^2\\ +2[2(1+2pq)u+(p+q)(1+u^2)]R(1-R)[R^2+(1-R)^2]\end{array}\right\}$$

and the conditional probability is

$$\frac{1}{4}\left\{\begin{array}{l}[1+(p+q)u+2pqu^2][R^2+(1-R)^2]^2\\ +4[2pq+(p+q)u+u^2]R^2(1-R)^2\\ +2[2(1+2pq)u+(p+q)(1+u^2)]R(1-R)[R^2+(1-R)^2]\end{array}\right\}$$

4.4.2 Match Probabilities for Half-Siblings

Consider two half-siblings who are both PQ at locus 1 and UV at locus 2. The joint probability is

$$\frac{pquv}{2}[((p+q)(u+v)+16pquv)(R^2+(1-R)^2) + 8\ [pq(u+v)+(p+q)uv)\ R(1-R)]$$

and the conditional probability is

$$\frac{1}{8}((p+q)(u+v)+16pquv)\ (R^2+(1-R)^2) + (pq(u+v)+(p+q)uv)\ R(1-R)$$

Consider two half-siblings who are both PP at locus 1 and UU at locus 2. The joint probability is

$$\frac{p^3u^3}{2}\ [(1+pu)(R^2+(1-R)^2)+2\ (p+u)\ R\ (1-R)]$$

and the conditional probability is

$$\frac{pu}{2}[(1+pu)(R^2+(1-R)^2)+2\ (p+u)\ R\ (1-R)]$$

Consider two half-siblings that are both PP at locus 1 and UV at locus 2. The joint probability is

$$\frac{p^3uv}{2}\ \{[u+v+4puv][R^2+(1-R)^2]+2\ [p\ (u+v)+4uv]R\ (1-R)\}$$

and the conditional probability is

$$\frac{p}{4}\ \{[u+v+4puv][R^2+(1-R)^2]+2\ [p\ (u+v)+4uv]R\ (1-R)\}$$

Consider two half-siblings who are both PQ at locus 1 and both UU at locus 2. The joint probability is

$$\frac{pqu^3}{2}\ \{[p+q+4pqu][R^2+(1-R)^2]+2\ [(p+q)u+4pq]R\ (1-R)\}$$

and the conditional probability is

$$\frac{u}{4}\ \{[p+q+4pqu][R^2+(1-R)^2]+2\ [(p+q)u+4pq]R\ (1-R)\}$$

4.4.3 Numerical Effect of Linkage for Full-Siblings and Half-Siblings

Two of the 13 CODIS loci, HUMCSF1PO and HUMD5S818, are located on chromosome 5. Bacher et al.[30] report that these loci are separated by 25 centiMorgans (cM). This equates to a recombination fraction of $R = 0.197$. The loci Penta D and HUMD21S11 are both on chromosome 21 and reported to be separated by 50 cM, which equated to a recombination fraction of $R = 0.316$ (see also Table 1.1). Both of these loci are contained in the 16-locus Promega PowerPlex 16 multiplex DNA profiling system.

Tables 4.14 and 4.15 give an overview of the magnitude of the numerical effect of this degree of linkage on allele probabilities for full- and half-siblings. The tabulated values are the ratios of the value accounting for linkage divided by the value if the loci were unlinked. Hence, in Table 4.14 for instance, if the recombination fraction $R = 0.197$ and the allele probabilities are all 0.10, then the probability that a full-sibling of a *PQ* heterozygote at locus 1 and a *UV* heterozygote at locus 2 will also be a *PQ* heterozygote at locus 1 and a *UV* heterozygote is 1.54 times larger when we account for linkage. The effect appears modest to us. It may become more serious when multiple sets of linked markers are used, as may be the case when technology moves to single nucleotide polymorphisms (SNPs).

Table 4.14 Numerical Effect of Linkage for Full-Siblings

Allele probabilities	*R*												
	0.00	0.05	0.10	0.15	0.197	0.20	0.25	0.30	0.316	0.35	0.40	0.45	0.50
Both *PQ* Heterozygotes at Locus 1 and *UV* Heterozygotes at Locus 2													
0.05	3.29	2.75	2.31	1.96	1.69	1.67	1.45	1.28	1.23	1.15	1.07	1.02	1.00
0.10	2.74	2.34	2.01	1.74	1.54	1.52	1.35	1.22	1.18	1.12	1.05	1.01	1.00
0.15	2.32	2.02	1.77	1.57	1.41	1.40	1.27	1.17	1.14	1.09	1.04	1.01	1.00
0.20	1.98	1.76	1.58	1.43	1.31	1.31	1.21	1.13	1.11	1.07	1.03	1.01	1.00
0.25	1.73	1.57	1.43	1.32	1.23	1.23	1.15	1.10	1.08	1.05	1.02	1.01	1.00
0.30	1.53	1.41	1.32	1.23	1.17	1.17	1.11	1.07	1.06	1.04	1.02	1.00	1.00
0.35	1.38	1.30	1.23	1.17	1.12	1.12	1.08	1.05	1.04	1.03	1.01	1.00	1.00
Both *PP* Homozygotes at Locus 1 and *UV* Heterozygotes at Locus 2													
0.05	3.30	2.76	2.32	1.96	1.69	1.67	1.45	1.28	1.23	1.15	1.07	1.02	1.00
0.10	2.76	2.36	2.03	1.75	1.54	1.53	1.36	1.22	1.19	1.12	1.05	1.01	1.00
0.15	2.35	2.05	1.80	1.59	1.43	1.42	1.28	1.18	1.15	1.10	1.04	1.01	1.00
0.20	2.03	1.81	1.61	1.46	1.33	1.32	1.22	1.14	1.12	1.08	1.03	1.01	1.00
0.25	1.78	1.61	1.47	1.35	1.26	1.25	1.17	1.11	1.09	1.06	1.03	1.01	1.00
0.30	1.59	1.46	1.36	1.27	1.20	1.19	1.13	1.08	1.07	1.05	1.02	1.00	1.00
0.35	1.44	1.35	1.27	1.20	1.15	1.14	1.10	1.06	1.05	1.03	1.02	1.00	1.00
Both *PP* Homozygotes at Locus 1 and *UU* Homozygotes at Locus 2													
0.05	3.31	2.77	2.32	1.96	1.69	1.68	1.45	1.28	1.23	1.15	1.07	1.02	1.00
0.10	2.79	2.38	2.04	1.76	1.55	1.54	1.36	1.23	1.19	1.12	1.05	1.01	1.00
0.15	2.39	2.08	1.82	1.61	1.44	1.43	1.29	1.18	1.15	1.10	1.04	1.01	1.00
0.20	2.09	1.85	1.65	1.48	1.35	1.35	1.23	1.15	1.12	1.08	1.04	1.01	1.00
0.25	1.85	1.67	1.51	1.38	1.28	1.28	1.19	1.12	1.10	1.07	1.03	1.01	1.00
0.30	1.66	1.52	1.41	1.30	1.22	1.22	1.15	1.09	1.08	1.05	1.02	1.01	1.00
0.35	1.52	1.41	1.32	1.24	1.18	1.17	1.12	1.08	1.06	1.04	1.02	1.00	1.00

Table 4.15 Numerical Effect of Linkage for Half-Siblings

Allele probabilities	*R*												
	0.00	0.05	0.10	0.15	0.197	0.20	0.25	0.30	0.316	0.35	0.40	0.45	0.50
Both *PQ* Heterozygotes at Locus 1 and *UV* Heterozygotes at Locus 2													
0.05	1.67	1.54	1.43	1.33	1.25	1.24	1.17	1.11	1.09	1.06	1.03	1.01	1.00
0.10	1.44	1.36	1.28	1.22	1.16	1.16	1.11	1.07	1.06	1.04	1.02	1.00	1.00
0.15	1.29	1.23	1.19	1.14	1.11	1.10	1.07	1.05	1.04	1.03	1.01	1.00	1.00
0.20	1.18	1.15	1.12	1.09	1.07	1.07	1.05	1.03	1.02	1.02	1.01	1.00	1.00
0.25	1.11	1.09	1.07	1.05	1.04	1.04	1.03	1.02	1.02	1.01	1.00	1.00	1.00
0.30	1.06	1.05	1.04	1.03	1.02	1.02	1.02	1.01	1.01	1.01	1.00	1.00	1.00
0.35	1.03	1.03	1.02	1.02	1.01	1.01	1.01	1.00	1.00	1.00	1.00	1.00	1.00
Both *PP* Homozygotes at Locus 1 and *UV* Heterozygotes at Locus 2													
0.05	1.74	1.60	1.47	1.36	1.27	1.27	1.19	1.12	1.10	1.07	1.03	1.01	1.00
0.10	1.55	1.44	1.35	1.27	1.20	1.20	1.14	1.09	1.07	1.05	1.02	1.01	1.00
0.15	1.40	1.32	1.25	1.20	1.15	1.14	1.10	1.06	1.05	1.04	1.02	1.00	1.00
0.20	1.29	1.23	1.18	1.14	1.10	1.10	1.07	1.05	1.04	1.03	1.01	1.00	1.00
0.25	1.20	1.16	1.13	1.10	1.07	1.07	1.05	1.03	1.03	1.02	1.01	1.00	1.00
0.30	1.13	1.11	1.09	1.07	1.05	1.05	1.03	1.02	1.02	1.01	1.01	1.00	1.00
0.35	1.08	1.07	1.05	1.04	1.03	1.03	1.02	1.01	1.01	1.01	1.00	1.00	1.00
Both *PP* Homozygotes at Locus 1 and *UU* Homozygotes at Locus 2													
0.05	1.82	1.66	1.52	1.40	1.30	1.29	1.20	1.13	1.11	1.07	1.03	1.01	1.00
0.10	1.67	1.54	1.43	1.33	1.25	1.24	1.17	1.11	1.09	1.06	1.03	1.01	1.00
0.15	1.55	1.44	1.35	1.27	1.20	1.20	1.14	1.09	1.07	1.05	1.02	1.01	1.00
0.20	1.44	1.36	1.28	1.22	1.16	1.16	1.11	1.07	1.06	1.04	1.02	1.00	1.00
0.25	1.36	1.29	1.23	1.18	1.13	1.13	1.09	1.06	1.05	1.03	1.01	1.00	1.00
0.30	1.29	1.23	1.19	1.14	1.11	1.10	1.07	1.05	1.04	1.03	1.01	1.00	1.00
0.35	1.23	1.19	1.15	1.11	1.09	1.08	1.06	1.04	1.03	1.02	1.01	1.00	1.00

Validating Databases

5

JOHN BUCKLETON

Contents

5.1 Introduction

This chapter is concerned with the issue of validating population databases for forensic work. The issue of independence testing is discussed. It is worthwhile considering here a quote from Weir[829] reproduced with permission: "Arguments have arisen that could have been avoided if the deliberate pace with which scientific investigation proceeds had been applied to the forensic uses of DNA evidence." The situation has improved since 1992, but there is an unfortunate reluctance in some areas to adopt continuous improvement due to entrenched views, fear of complexity, and fear of retrospective review of past cases.

Open publication of data and analysis, and the open debate on the conclusions that may be drawn from these data, represent a sound scientific approach to alleviating this type of problem. In 1995 Strom[737] complained of

> the refusal by the FBI laboratory of outside inspection and data verification is troubling, especially when I have been called upon to testify in support of its findings. Regardless of the reasons for this policy, I believe that the FBI laboratory should be held to the same standards and requirements as other laboratories. (Reproduced with the kind permission of *Nature* and Dr. Strom)

This situation appears to have been remedied in part by the placement of the FBI data into the public domain,[133] which is an admirable policy that should be widely implemented. That is: Let's have the arguments out of court — not in it. This involves openness by the government agencies responsible for the majority of forensic work. In fact, it involves actively supporting independent or defense reanalysis whether this support is reciprocated or not.

5.2 Which Is the Relevant Population?

As discussed in Chapter 3, profile or match probabilities are estimated with reference to a population. This raises the question of "what is a population" or "which population?"[265,472,501,502,503,505,508,510,665,847] Two options briefly discussed here are the race to which the suspect belongs, and the "general population." A brief discussion is provided here as on balance the use of the "unifying formula" discussed in Chapter 3 is preferred, which does not require the definition of a relevant population. Nonetheless, there has been much debate on the appropriateness of one or the other of the two options outlined above. Consider the question: Why are we doing a calculation? Typically, the answer would be to assess the evidence if the suspect is not the contributor, or under the Bayesian framework to assess the evidence under H_d.[167] It is clear then that the race of the suspect does not define the relevant population. This is defined more by the circumstances of the crime or other evidence such as eyewitness evidence.[111,265,508,784,806] The circumstances or evidence may point to one ethnic group wholly or partially, or the location of the crime may suggest which set of persons had opportunity.[779,807] Using the race of the suspect is typically conservative; however, it is not necessarily a reasonable representation of the relevant population. Hence it is more appropriate to attempt to model a population defined by the crime. This is typically multiracial. In a later section we discuss how to combine estimates from separate races into one appropriate for a multiracial population.

5.3 Population Databases

Population databases are distinct from intelligence databases. They are used to estimate the rarity of a profile in a population in order to give an indication to a court of the strength of the DNA evidence.

Consider the highly stylized diagram of the human population given in Figure 5.1. This figure is not intended to imply that there is some objective definition of the term "race." This term has come to be viewed as increasingly arbitrary as our understanding of human population genetics and evolution has improved. Rather it simply implies that there is some structure to the human population. This is more properly viewed as a continuum, but most

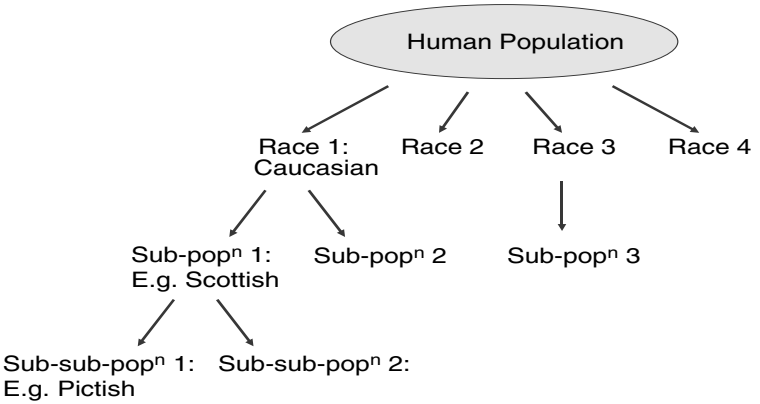


Figure 5.1 A highly simplified and stylized diagram of the structure of the human population.

models treat it as a hierarchy of partitions. For reasons that were discussed in Chapter 3, the estimation process will have a small error if we ignore the structure in the human population. This and other issues relating to the validation of population databases are discussed in this chapter.

5.3.1 Validating Population Databases

It is general practice for a laboratory to validate the population database it intends to use before proceeding to court with estimates of evidential weight. This is a desirable feature, as is publication or deposition of the details in the public domain. Typically there may be a challenge to the use of a database in the first few court cases undertaken by a particular laboratory. The challenges may include issues about the size of the database, the method of selection of the samples, and dependence effects. During my time working with Bruce Weir, we "validated" a number of U.S. databases. I have also validated databases in New Zealand, Australia, and the U.K. One would have thought that having performed quite a few validations I would have a good idea of what was required; however, the more I look at the subject, the more I ask "What is the process of validating a database?" There appears to be no published advice on this. This question was thrown into a particular light by a case in South Australia: *R v Karger* before the Honourable Justice Mulligan.[639] The effect of this case and questions from His Honour have led me to believe that it is not the database that should be validated but rather a system of interpretation. There is an inherent interplay between what is expected of a database and the mode by which the testimony will be developed. The concept of "fitness for purpose" is closely akin to this process. The key questions are: What are you going to do with the database? Is it fit for this purpose? In this

regard, noted authors such as Brinkmann[91] write: "For their application in forensic casework, extensive studies have to be carried out. This would include population studies … A minimum of 200 unrelated individuals and/or 500 meioses have to be investigated for each STR system to study allele frequencies, the Mendelian inheritance, and whether significant numbers of mutations exist."

The number 200 has become the *de facto* standard for the size of the database. This size certainly suffices for estimation of allele probabilities especially if sampling error is considered. Smaller samples may also suffice, again especially if sampling error is considered. This is discussed further in Chapter 6. However, a sample of 200 profiles will not inform us much with regard to "population studies" if our plan is to investigate deviations from Hardy–Weinberg and linkage equilibrium by hypothesis testing. Other population genetic studies, such as comparisons with other populations, appear to be more informative in this regard.

5.3.2 Sampling Races or Populations?

Let us begin with the question: Should we survey general populations, say the population of the State of Victoria, Australia, or should we divide this population according to some partition, which for the purposes of this section we will call race? Both approaches are currently used. It is normal practice in the U.S., U.K., and New Zealand to divide according to race. However, Australia, until recently, took mixed race samples of the "general" population. Consider, again, the highly simplified model of the human population given in Figure 5.1. There has already been considerable argument on the question "How many races are there?" or even "What is a race?" Many studies suggest that there is little evidence of a clear subdivision of humans into races but rather that variation is essentially continuous.[670] From the genetic point of view, "race" is simply an arbitrary partition of the total human diversity.

However, these somewhat arbitrary partitions of the human population do correspond in some way to our view of recent human evolution. The more we clump groups together (i.e., the higher we go in this hierarchy), the higher the resulting dependence effects within and between loci. The examples illustrating the Wahlund effect (Tables 3.2 and 3.3) showed that the effects were larger the more the allele frequencies differed between the groups. If we mix "races" to create a "general" population, we create larger dependence effects. Conversely, the more we subdivide the population into genetic groups, the lower the remaining dependence effects should be, since the remaining subdivisions of these groups should be more similar to each other.[805,806]

A common compromise implemented in many laboratories is to subdivide the population as far as races but not further. It is possible to recombine these

estimates into a general population estimate if required and the preferred option is stratification (see Box 5.1).[114,784]

5.3.3 Source of Samples

There has been considerable discussion in the courts along the lines that the samples used in forensic work are not collected by random sampling

Box 5.1 Stratification

Consider an area with races R_1, R_2, …, R_N resident. We consider the prior probability that a person from each race is the donor of the stain. As a first approximation, we take these priors to be simply the fraction of the population that these races represent. Suppose these are in the proportions $\Pr(R_1)$, $\Pr(R_2)$ … $\Pr(R_N)$ in the population that we consider relevant. Suppose that the probability of the evidence (E) depends on which race is the donor; we could write $\Pr(E \mid R_1)$, $\Pr(E \mid R_2) \ldots \Pr(E \mid R_N)$
Then if these partitions are exclusive and exhaustive,

$$\Pr(E) = \sum_{i=1}^{N} \Pr(E \mid R_i)\Pr(R_i)$$

which suggests a fairly easy way to combine different estimates. However, use of the general form of Bayes's theorem is superior (see Chapters 2 and 4; under the full Bayesian approach to interpretation).

The National Research Council's (NRC) second report in 1996,[585] stated that the subgroup to which the suspect belongs is irrelevant. The logic followed the line that we desire to estimate the probability of the evidence if the suspect is innocent and that instead a random individual committed the crime. This is substantially correct, but overlooks the issue that the presence of some members of the subpopulation among the group of potential offenders may have a significant impact on the total weight of the evidence (see Chapter 4).

To answer the question of the magnitude of the diversity between subpopulations, fairly extensive studies have been carried out estimating the genetic differences between different groups of people. These are reviewed later in this chapter. In general, these studies support the notion that differences between subpopulations are small.

Variation between subpopulations can be accommodated by the use of a correction factor (F_{ST} or θ)[36,41] discussed in Chapter 3. Since differences between subpopulations are typically minor, inferences for a subpopulation for which a database is not available can be accommodated by using a general database so long as the θ correction is incorporated.

methods.[502] The accusation stems from standard statistical theory. Statisticians differentiate between experimental and observational data. Experimental data arise when they are obtained using some sort of deliberate sampling. In contrast, observational data arise when the data have been collected for some other purpose. Experimental data are taken to be superior to observational data in that they have much less chance of being affected by various types of bias in the sampling.

Most forensic samples are convenience samples. This means that they have come to the laboratory by some "convenient" way, such as from blood donors, staff, or offender databases. As such they do not comprise random samples. An incorrect response is to say that they were not selected on the basis of their genotypes and hence no bias is expected. Indeed they are not selected on the basis of their genotypes, but the accusation is that the bias is inadvertent.

I argue that this is not a major issue. However, to set the scene let me postulate some extreme examples. Imagine that we use as our sample the staff of the Informatics Department of North Carolina State University. This consists, on this day, of two New Zealanders, a Canadian, three Chinese, and a few U.S. nationals. The U.S. nationals come from many states, but none come from North Carolina. If this sample were to be used to model the North Carolina population, it would be very unrepresentative. This is not because we have deliberately made an unrepresentative sample by knowing the genotypes of the candidates, but rather that our sampling strategy has an in-built bias (in this case, to people who have relocated).

Real situations are likely to show a much less pronounced bias. We could imagine that blood donors and staff overrepresent some groups and underrepresent others. It is harder to argue that offender databases are unrepresentative as they certainly seem close to a representative sample of "alternate offenders."[111] To summarize the statistical argument: Only random sampling can guarantee a representative sample.

To turn now to the counter argument, it is wise to admit that we cannot guarantee that our samples are representative. This is for two reasons: (i) we do not undertake random sampling, and (ii) we do not always know what group we are trying to represent.

Consider crimes in one of the states of the United States. In some cases we may want to represent small rural populations, and in others large cosmopolitan populations. In other cases, there may be evidence from, say, eyewitnesses that direct us toward a particular group. The very act of defining a population of alternate offenders is very difficult (and unnecessary and unhelpful if we use the unifying formula of Balding).[265,508]

Consider then our surveying requirements if we wished to meet the strictest statistical standards. First we must define our population of offenders, next we need to randomly sample from these, and last we need to do this for every crime.

If we concede that we cannot guarantee databases that are truly random samples, where does this lead us? Many defense analysts would argue that it leaves us nowhere and that all future arguments are built on an insecure foundation. However there really is quite a body of population genetic evidence that suggests that, although we might have slightly unrepresentative samples, the effect is likely to be minor. Fung[323] provides important experimental support for this. There is also a theory, the subpopulation theory, available to attempt to accommodate this unrepresentativeness.

How bad could our sample be? Let us imagine that we intend to sample race *x* in a specific locality. We take samples by self-declaration of that race at, say, a blood donation clinic. We could imagine the following biases:

A bias caused by the self-declaration process. This will be dealt with separately.
A bias caused because one subpopulation overdonates and others underdonate.
Systematic typing bias. This will also be dealt with separately.

We are left, in this section, with the task of assessing the possible effect of the bias caused by one subpopulation overdonating and others underdonating. The pertinent questions are:

How much do subpopulations differ?
How much could one subpopulation overdonate?
Do we intend to make any compensation for nonrepresentativeness in our testimony?

So we come to the first task when validating a database. How much bias could the sampling process have induced, *and* will we compensate for it?

5.3.4 Self-Declaration

Most laboratories obtain samples for their DNA database from volunteers or from offender databases. These are typically separated into races by self-declaration. Self-declaration is taken to be the process by which people nominate their own race. More occasionally, other methods are used such as "surname." The issue has been raised often in court as to whether the self-declaration (or surname) process introduces any unacceptable bias.[835]

There are many instances of possible self-declaration bias. Wild and Seber[865] note: "In recent U.S. censuses there has been a big upsurge in the census counts of American Indians that could not be explained by birth and death statistics."

From my own experience I can confirm that there are, at least, errors in the self-declaration process. In the New Zealand subpopulation databases, we have historically investigated all "matches" (incidents of duplicate STR profiles). Most

often this has occurred because the same person has been sampled twice. There have also been instances of identical twins on the database. It is not uncommon for the declaration to be different for the different occasions an individual is sampled or for each member of the pair of twins. This is typically a difference of detail, such as a claim of $\frac{1}{2}$ Maori on one occasion and $\frac{1}{4}$ Maori at another.

Does this render the process useless? The evidence suggests not. For New Zealand, the Maori and Samoan data were further divided into subsets of varying levels of ethnicity, representing the dilution of the selected ethnic subpopulation largely by Caucasians. For example, New Zealand Maori samples were distributed into six subsets: full-blood, $\frac{3}{4}$, $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, and $\frac{1}{16}$. Similarly, Samoan samples from the database were distributed into four subgroups: full-blood, $\frac{3}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$. An estimate has been made of the pairwise genetic distance between the self-declared ethnicity for the New Zealand STR data. This was possible through a self-declaration process based on ancestral information over four generations.[814]

The results of the genetic distance estimates (Tables 5.1 and 5.2) show that the genetic distance, θ, from the Caucasian population increases as the level of self-declared Maori or Samoan ethnicity increases. This matrix of genetic distances was also represented using principal coordinates (Figure 5.2) and the same pattern can be seen. This provides significant support for the effectiveness of self-declaration as a means of segregating reference samples by ethnicity. There is no claim that it is error free, yet it cannot be totally random or we would not get this logical pattern.

The points corresponding to small reported fractions of Maori and Samoan ancestry are closer to each other than they are to the point representing the Caucasian population. Walsh et al.[814] suggested that this is because the admixture is complex, and a person reporting a small fraction of, say, Samoan ancestry may also have some Maori as well as Caucasian ancestors.

Rosenberg et al.[671] typed 377 autosomal microsatellite loci in 1056 individuals from 52 populations. Without using any prior information, they identified

Table 5.1 Distance for New Zealand Maori from the Caucasian Population

Self-Declared Ethnicity Level	Distance from Caucasian
Full Maori	0.037
$\frac{3}{4}$ Maori	0.030
$\frac{1}{2}$ Maori	0.023
$\frac{1}{4}$ Maori	0.014
$\frac{1}{8}$ Maori	0.010
$\leq \frac{1}{16}$ Maori	0.003

Following Walsh et al.[812,814] © 2003 ASTM International. Reprinted with permission.

Table 5.2 Distance for Samoans from the Caucasian Population

Self-Declared Ethnicity Level	Distance from Caucasian
Full Samoan	0.038
$\frac{3}{4}$ Samoan	0.021
$\frac{1}{2}$ Samoan	0.014
$\leq \frac{1}{4}$ Samoan	0.001

Following Walsh et al.[812,814] © 2003 ASTM International. Reprinted with permission.

six main genetic clusters, five of which corresponded to major geographical regions, and subclusters that often corresponded to individual populations. There was a general agreement of this "genetically determined" origin with self-reported ancestry. This is, again, important confirmation of the usefulness of self-reported ancestry (subsequent discussions[285,672] relate to mathematical treatments of the data and do not affect this conclusion).

5.3.5 Systematic Mistyping or Systematic Nontyping

These are two potential sources of bias in any population survey. The first is far more dangerous than the second, although both may be important.

Systematic mistyping describes the situation where one or more genotypes are systematically mistyped as a different genotype. An instance could be that some heterozygotes are systematically mistyped as homozygotes because of allelic dropout or severe heterozygote imbalance. The result will

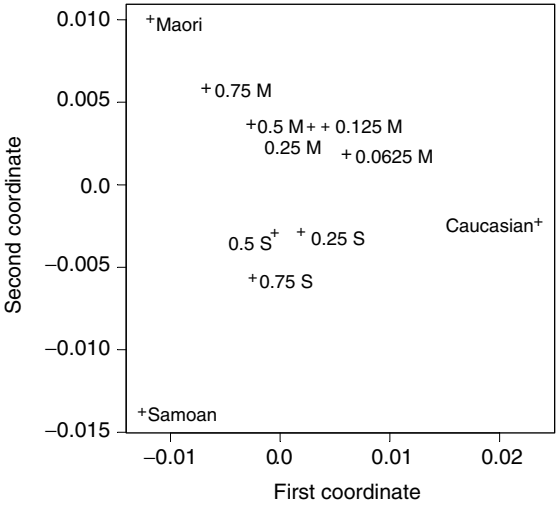


Figure 5.2 Principal coordinate representation of the interpopulation genetic distances following Walsh et al.[812,814] © 2003 ASTM International. Reprinted with permission.

be the appearance of slightly too many homozygotes that may be detected during the statistical analysis. Experience suggests that this does occur and some anomalies that have been highlighted during statistical analysis are a consequence of this effect. Another possible mistyping is to designate a common homozygote or heterozygote as a nearby rare option because of band shift.[343] For instance, analysis of the New Zealand subpopulation data detected one instance of a 10,10 genotype at the TH01 locus for Caucasians that, when reexamined, was found to be the common 9.3,9.3. This datum was noticed because 10,10 should be rare. Mistyping between common genotypes is unlikely to be detected during statistical analysis.

Hence the validation of the database by statistical analysis may lead to the detection of some mistyping. It is unwise, however, to assume that all instances of mistyping would be found, as it is likely that statistical analysis will fail to detect all but the most obvious. The integrity of the remaining samples does not rely on statistical testing but on the quality standards of the laboratory doing the typing.

Statistical examination cannot, in any meaningful way, guarantee the correctness of the data. That relies principally on the quality standards of the laboratory.

Systematic nontyping refers to a situation where certain alleles or genotypes are less likely to "type" or "be called." This is realistic if, say, the larger alleles are harder to amplify. Another possibility is that low peak area homozygotes are classed as, say, "11,?." By this, the operator means that the genotype has the 11 allele, but is uncertain whether or not another allele may be present. This type of data is often present in files that I receive. It is difficult to process and as such it is often omitted. Thus some homozygotes could systematically be removed from the data.

This effect is akin to a "nonresponse" bias in classical sampling terminology. It could lower the apparent frequency of those alleles or genotypes that are hard to type and hence raise the relative frequency of the others. The check for this is to see how many genotypes are classified as "missing," for example, "11,?." Obviously if there is little or no "missing data," then there can be no bias from systematic nontyping.

What should be done if there is a substantial amount of missing data? Let us say that at a locus there is of the order of 10% of the data missing due to nontyping. This opens the possibility of systematic nontyping bias, but it does not prove that such a bias exists. If the nontyping is random, that is, if it is evenly spread among all the alleles, then this will have no effect. The only clues as to whether the nontyping has affected one allele predominantly would be a comparison with a closely related population.

Many laboratories, understandably, perform their statistical survey at the implementation phase of new technology. This is the time that they are most

prone to mistyping and nontyping. This leads us to another possible task when validating databases: check for the possibility of nontyping bias.

5.3.6 Size of Database

How large should a database be to be valid? This must be the most prevalent question asked of the statistician either by laboratories or in court. It is an entirely reasonable question because in statistical sampling size does matter. However, it is surprisingly difficult to answer in a logical way. Once again, the answer comes down to "fitness for purpose." The two key factors in this assessment of fitness are:

Whether or not the database is meant to inform on choice of population genetic model.
Whether or not the testimony will include sampling error estimates.

Most published attempts at validation of databases suggest that they, in some way, inform the choice of population genetic model, in particular that they somehow validate the product rule. If we intend to validate the use of the product rule on the basis of this database rather than base the validation on all the literature on the subject, then the database has to be enormous. In essence, to set the size of the database to "validate" the product rule, we need some "acceptance" criteria for the product rule. In particular, we need to answer the question "How wrong are we prepared to be?" Do not assume from this statement that we will be able to produce the "correct" answer from our pocket, but that we do need to think about tolerable limits for "error." To date, this question, "How wrong are we prepared to be?," has never been answered. In fact, it may actually never have been asked in this way. We could further discuss whether this is a decision for a scientist or a court.

Let us assume that we intend to tackle the question "How wrong are we prepared to be?" Do we want these limits to be phrased as a ratio, for instance "this estimate could be high or low by a factor of 10?" The scientist (or court) may be more tolerant of an estimate that is likely to err in favor of the defendant than one that may err in favor of the prosecution. This may result in them being inclined to give limits for "acceptance" that are asymmetric. Do we become more tolerant of "error" as the estimates get smaller? For instance, do we need a higher level of accuracy for estimates in the area of 1 in a million, but after 1 in a billion can we tolerate more uncertainty? This suggests some sort of sliding scale that may be definable on a logarithmic scale. For instance, do we want the log of the estimate to be within a factor of, say, $\pm$ 17%? Embedded in the argument above is a concept of defining "wrong." The obvious answer is to use the concept of a true answer that we unfortunately

do not know. This approach to defining an acceptance criterion is set up to fail.

What would suit us best would be if we could define something like "I will use the product rule if I can be reasonably certain that θ is less than 1%." This would allow us to perform power studies and determine how large a database would need to be so that we can be 90% sure of finding dependence if $\theta = 0.01$. Weir[842] gives an estimate for the chi-square test, which suggests that a database of size 105,000 should be sufficiently large to detect a value of f[a] of 0.01.

Clearly most databases are not of this size, and hence they could not validate the population genetic model under this criterion. It is argued below that no database *per se* can realistically validate the population genetic model. Any validation must rely on population genetic studies. Examination of one database of the order of hundreds or a few thousand samples cannot validate the product rule, nor can it validate any other population genetic model. If advice is to be given on the choice of population genetic model, it should be based on an understanding of the population genetics of the populations in question.

If we are not going to validate the population genetic model, then all we are going to use the database for is to determine allele probabilities. As long as we make a consideration of sampling error, then almost any size database will do.

If a sampling error correction is used, then there are almost no restrictions on how large or small a database needs to be.

What if no consideration of sampling error is to be made? Then we are back to the question "How wrong are we prepared to be?" Fortunately this time, we have a body of statistical theory that allows us to estimate the expected sampling error for a given database size. So if we are informed as to how wrong the analyst is prepared to be, we can give them an approximate estimate of how large the database needs to be. This is necessarily approximate as it depends on the number of loci and the separate allele probabilities for each particular genotype. A sample of size 200 has become the *de facto* standard, but this is more by common acceptance rather than by forceful scientific argument that this is the correct number.

If no consideration of sampling error is to be made for each case, then it is wise to assess the probable uncertainty arising from sampling during validation.

5.4 Validating the Population Genetic Model

As a simple instance, suppose that a laboratory intends to go into court using the product rule. Are they obliged to "prove independence?"

[a] f is the within-population inbreeding parameter and is not synonymous with θ.

Let us assume that the act of validating a database also includes the necessity to validate the population genetic model that is to be used in court. The proof of the validity of the product rule can proceed from population genetic considerations completely independently of the existence of a genetic database. It is at least theoretically feasible that the laboratory could study mating patterns with their population and the other requirements for Hardy–Weinberg and linkage equilibrium, and conclude that the product rule was a valid approximation without ever typing a DNA sample. However, the assumptions for Hardy–Weinberg and linkage equilibrium are never exactly fulfilled in real human populations, and hence it will not be possible to conclude exact correspondence to the product rule from a purely population genetic argument. In fact, the reality of population genetics would lead us to doubt the validity of exact adherence to Hardy–Weinberg and linkage equilibrium in the first place.

Can the existence of a database save us from this dilemma by its examination by independence testing? The answer is no. We now turn to a discussion of independence testing.

5.4.1 Independence Testing

There are a number of tests that may be undertaken to investigate departures from genetic equilibrium.[880] The recommended test for STR data is Fisher's exact test and it is used in most situations. This and a number of other options are discussed. This section follows Law[491] extensively.

5.4.1.1 The Exact Test

Genotype counts are expected to follow a multinomial distribution and hence depend on the unknown true allele frequencies. To avoid the requirement for the unknown allele frequencies, the exact test for the hypothesis of allelic independence is conditioned on the observed allelic counts.[377] The exact test has been reported to have better power when compared to alternative testing strategies.[531,842]

Following Law et al.,[492] we write:

P_c: the conditional probability of the genotype counts,
n_g: the genotype counts,
n_{lj}: the allelic counts, and
$H = \sum_l \sum_g H_{gl}$: the total number of heterozygotic loci in the sample

Then

$$P_c = \frac{n!2^H}{\Pi_g n_g!} \prod_l \frac{\Pi_j n_{lj}!}{(2n)!} \tag{5.1}$$

The exact test compares P_c calculated from the observed sample with the values in all genotype arrays with the same allele counts as the observed sample.

The *p*-value of the test is the proportion of arrays with a probability no more than that for the observed sample.

It is typically impossible to enumerate all possible genotype arrays, as there are too many of them. An approach attributed to Felsenstein (in Guo and Thompson,[377] who proposed a Markov chain approach) is to take a sample of all possible genotype arrays. The alleles at each locus are permuted separately to form new multilocus genotypes. The proportion of permuted data sets that give rise to a smaller P_c value than the original data is noted and serves as the empirical *p*-value of the test.

When a permutation approach is used, a portion of the statistic is invariant and can be omitted.

Hence, instead of calculating

$$P_c = \frac{n!2^H}{\Pi_g n_g!}\prod_l \frac{\Pi_j n_{lj}!}{(2n)!}$$

as in Equation (5.1), the simpler quantity

$$\frac{2^H}{\Pi_g n_g!}$$

can be used. This is no longer the conditional probability of the genotype counts, but it is proportional to that probability.

Zaykin et al.[880] showed that the power of the exact test increases when more loci are used in the testing procedure. However, Law et al.[492] show that this is only true for certain population genetic events such as substructure but not for, say, admixture.

5.4.1.2 Total Heterozygosity Test

Total heterozygosity may be used as a test for some types of disequilibrium. It should be noted however that this is not a test for independence *per se* as there are types of dependency that do not affect total heterozygosity. Under allelic independence, the total heterozygosity is

$$H^e = L - \sum_l \sum_j p_{slj}^2 \tag{5.2}$$

This allows reduction of the genotype array to two categories: heterozygous and homozygous genotypes.

This gives the total heterozygosity test statistic:

$$X^2 = \frac{(H - H^e)^2}{H^e} + \frac{(H - H^e)^2}{nL - H^e} = \frac{nL(H - H^e)^2}{H^e(nL - H^e)} \tag{5.3}$$

where nL is the total count (where n is the sample size and L is the number of loci).

The statistic X^2 has a chi-square distribution with one degree of freedom under the hypothesis of within-locus allelic independence. This is a two-sided test and rejects the hypothesis of allelic independence for both large and small values of H in the data. However, the true allele frequencies in the population are generally unknown and need to be estimated from the data when H^e is calculated. Consequently, X^2 will no longer have a chi-square distribution under the null hypothesis due to the use of estimated allele frequencies, and hence there is expected to be a loss of power for the test.

Alternatively, permutation methods may be used instead of the chi-square test to estimate the distribution of the test statistic, H, under the null. The empirical p-value of the test is calculated from the proportion of permuted genotype arrays with fewer (H^- test) or more (H^+ test) heterozygous genotypes than the original data.

5.4.1.3 Variance of Heterozygosity Test

Brown and Feldman[96] and Chakraborty[157] suggested that the variance of the number of heterozygous loci for each genotype in the sample could be used as a test for between-locus associations. They give the variance of the heterozygosity test statistic as

$$V = \frac{\sum_g H_{g\cdot}^2 n_g}{n-1} - \frac{H^2}{n(n-1)}$$

where $H_{g\cdot}$ is the number of heterozygous loci in genotype G_g. The test statistic V is the variance of the number of heterozygous loci for each genotype in the population.

5.4.2 Performance of the Tests

Law[491] and Law et al.[492] used simulation to examine the properties of the tests (see Tables 5.3 and 5.4).

5.4.2.1 Population Substructure

Zaykin et al.,[880] Law,[491] and Law et al.[492] showed that as the number of loci increases, for a substructured population the empirical power for both the exact and the total heterozygosity test increases.

As noted above, when a permutation test is used, a portion of the exact test statistic is invariant and can be omitted. The simpler quantity

$$\frac{2^H}{\Pi_g n_g!}$$

may be used instead. When the data become sufficiently sparse such that all n_g are 0 or 1 in the data or any permutation, then the product

$$\prod_g n_g! = 1$$

and this term also becomes invariant. We are left with a statistic based almost solely on total observed heterozygosity.

Accordingly, the power of the exact test and the H^- test when more than one locus is used in the test becomes very similar. The empirical power of

Table 5.3 Power of P_c, H^-, and V to Detect Departure from Equilibrium due to Drift and Drift Followed by One Generation of Admixture

θ	Loci	Substructure			Admixture		
		P_c	H^-	V	P_c	H^-	V
0.00	1	0.05 (0.01)	0.05 (0.01)	0.06 (0.01)	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)
	2	0.06 (0.01)	0.05 (0.01)	0.05 (0.01)	0.05 (0.01)	0.03 (0.01)	0.04 (0.01)
	3	0.06 (0.01)	0.06 (0.01)	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)
	4	0.04 (0.01)	0.04 (0.01)	0.05 (0.01)	0.04 (0.01)	0.04 (0.01)	0.06 (0.01)
	10	0.05 (0.01)	0.05 (0.01)	0.05 (0.01)	0.07 (0.01)	0.07 (0.01)	0.04 (0.01)
0.01	1	0.07 (0.01)	0.10 (0.01)	0.06 (0.01)	0.05 (0.01)	0.03 (0.01)	0.04 (0.01)
	2	0.14 (0.02)	0.14 (0.02)	0.07 (0.01)	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)
	3	0.19 (0.02)	0.19 (0.02)	0.06 (0.01)	0.04 (0.01)	0.04 (0.01)	0.06 (0.01)
	4	0.20 (0.02)	0.20 (0.02)	0.08 (0.01)	0.04 (0.01)	0.04 (0.01)	0.05 (0.01)
	10	0.31 (0.02)	0.30 (0.02)	0.08 (0.01)	0.04 (0.01)	0.04 (0.01)	0.06 (0.01)
0.03	1	0.12 (0.01)	0.26 (0.02)	0.11 (0.01)	0.05 (0.01)	0.04 (0.01)	0.05 (0.01)
	2	0.42 (0.02)	0.42 (0.02)	0.13 (0.01)	0.06 (0.01)	0.04 (0.01)	0.06 (0.01)
	3	0.55 (0.02)	0.55 (0.02)	0.14 (0.02)	0.03 (0.01)	0.03 (0.01)	0.05 (0.01)
	4	0.67 (0.02)	0.67 (0.02)	0.14 (0.02)	0.05 (0.01)	0.05 (0.01)	0.05 (0.01)
	10	0.94 (0.01)	0.94 (0.01)	0.15 (0.02)	0.05 (0.01)	0.05 (0.01)	0.06 (0.01)
0.05	1	0.30 (0.02)	0.58 (0.02)	0.16 (0.02)	0.05 (0.01)	0.05 (0.01)	0.04 (0.01)
	2	0.80 (0.02)	0.80 (0.02)	0.20 (0.02)	0.06 (0.01)	0.05 (0.01)	0.06 (0.01)
	3	0.91 (0.01)	0.92 (0.01)	0.22 (0.02)	0.04 (0.01)	0.04 (0.01)	0.07 (0.01)
	4	0.96 (0.01)	0.96 (0.01)	0.24 (0.02)	0.04 (0.01)	0.04 (0.01)	0.06 (0.01)
	10	1.00 (0.00)	1.00 (0.00)	0.28 (0.02)	0.05 (0.01)	0.05 (0.01)	0.08 (0.01)
0.10	1	0.77 (0.02)	0.89 (0.01)	0.38 (0.02)	0.04 (0.01)	0.02 (0.01)	0.05 (0.01)
	2	1.00 (0.00)	0.99 (0.00)	0.41 (0.02)	0.09 (0.01)	0.04 (0.01)	0.07 (0.01)
	3	1.00 (0.00)	1.00 (0.00)	0.44 (0.02)	0.04 (0.01)	0.04 (0.01)	0.08 (0.01)
	4	1.00 (0.00)	1.00 (0.00)	0.49 (0.02)	0.05 (0.01)	0.05 (0.01)	0.09 (0.01)
	10	1.00 (0.00)	1.00 (0.00)	0.56 (0.02)	0.06 (0.01)	0.06 (0.01)	0.29 (0.02)

Samples of size 200 were used with 1, 2, 3, 4, and 10 loci, each with 10 alleles per locus (reproduced with the kind permission of Law[491]).

Table 5.4 Power of P_c, H^-, and V to Detect Departure from Equilibrium due to Drift and Drift Followed by one Generation of Admixture

θ	Loci	Substructure			Admixture		
		P_c	H^-	V	P_c	H^-	V
0.00	1	0.04 (0.01)	0.03 (0.01)	0.06 (0.01)	0.06 (0.01)	0.04 (0.01)	0.05 (0.01)
	2	0.05 (0.01)	0.04 (0.01)	0.06 (0.01)	0.07 (0.01)	0.04 (0.01)	0.06 (0.01)
	3	0.03 (0.01)	0.03 (0.01)	0.06 (0.01)	0.05 (0.01)	0.04 (0.01)	0.05 (0.01)
	4	0.04 (0.01)	0.05 (0.01)	0.05 (0.01)	0.04 (0.01)	0.04 (0.01)	0.06 (0.01)
	10	0.04 (0.01)	0.04 (0.01)	0.06 (0.01)	0.04 (0.01)	0.04 (0.01)	0.05 (0.01)
0.01	1	0.09 (0.01)	0.16 (0.02)	0.06 (0.01)	0.05 (0.01)	0.04 (0.01)	0.04 (0.01)
	2	0.28 (0.02)	0.29 (0.02)	0.07 (0.01)	0.06 (0.01)	0.03 (0.01)	0.04 (0.01)
	3	0.32 (0.02)	0.33 (0.02)	0.09 (0.01)	0.04 (0.01)	0.03 (0.01)	0.06 (0.01)
	4	0.35 (0.02)	0.35 (0.02)	0.11 (0.01)	0.05 (0.01)	0.05 (0.01)	0.05 (0.01)
	10	0.55 (0.02)	0.55 (0.02)	0.13 (0.02)	0.06 (0.01)	0.06 (0.01)	0.06 (0.01)
0.03	1	0.25 (0.02)	0.45 (0.02)	0.18 (0.02)	0.05 (0.01)	0.04 (0.01)	0.04 (0.01)
	2	0.69 (0.02)	0.77 (0.02)	0.23 (0.02)	0.08 (0.01)	0.03 (0.01)	0.04 (0.01)
	3	0.88 (0.01)	0.90 (0.01)	0.27 (0.02)	0.05 (0.01)	0.05 (0.01)	0.05 (0.01)
	4	0.97 (0.01)	0.97 (0.01)	0.30 (0.02)	0.07 (0.01)	0.06 (0.01)	0.07 (0.01)
	10	1.00 (0.00)	1.00 (0.00)	0.32 (0.02)	0.07 (0.01)	0.07 (0.01)	0.08 (0.01)
0.05	1	0.49 (0.02)	0.77 (0.02)	0.35 (0.02)	0.05 (0.01)	0.03 (0.01)	0.04 (0.01)
	2	0.99 (0.00)	0.98 (0.01)	0.41 (0.02)	0.10 (0.01)	0.05 (0.01)	0.06 (0.01)
	3	1.00 (0.00)	1.00 (0.00)	0.43 (0.02)	0.04 (0.01)	0.04 (0.01)	0.07 (0.01)
	4	1.00 (0.00)	1.00 (0.00)	0.49 (0.02)	0.05 (0.01)	0.05 (0.01)	0.07 (0.01)
	10	1.00 (0.00)	1.00 (0.00)	0.56 (0.02)	0.04 (0.01)	0.04 (0.01)	0.10 (0.01)
0.10	1	0.98 (0.01)	1.00 (0.00)	0.68 (0.02)	0.06 (0.01)	0.07 (0.01)	0.05 (0.01)
	2	1.00 (0.00)	1.00 (0.00)	0.75 (0.02)	0.35 (0.02)	0.05 (0.01)	0.06 (0.01)
	3	1.00 (0.00)	1.00 (0.00)	0.80 (0.02)	0.09 (0.01)	0.07 (0.01)	0.09 (0.01)
	4	1.00 (0.00)	1.00 (0.00)	0.85 (0.02)	0.07 (0.01)	0.07 (0.01)	0.14 (0.02)
	10	1.00 (0.00)	1.00 (0.00)	0.87 (0.02)	0.05 (0.01)	0.05 (0.01)	0.58 (0.02)

Samples of size 500 were used with 1, 2, 3, 4, and 10 loci, each with 10 alleles per locus (reproduced with the kind permission of Law[491]).

these two tests becomes identical when the "sparseness" condition (all n_g are 0 or 1) is met. At this point, the exact test is no longer a test for independence, but simply a test for excess homozygosity.

5.4.2.2 Randomly Mating Admixed Populations

Recall that one generation of random mating restores Hardy–Weinberg equilibrium. Therefore, an admixed population that has undergone one generation of random mating should exhibit within-locus equilibrium. However, linkage disequilibrium will still be present in the admixed population, although it is expected to halve for unlinked loci with every generation of

random mating. Law[491] and Law et al.[492] investigated the ability of the various tests to detect such disequilibrium.

The empirical power of the exact test, the variance of the heterozygosity test, and the total heterozygosity test (obviously) was low for most admixture situations studied.

5.4.3 Misuse of Independence Testing

It is possible to misinterpret the results of independence testing, and it is wise to consider a few warnings before proceeding.[103,104,113,266,781] Most of these warnings represent standard statistical thinking and are well known though often overlooked in the forensic literature.

> The Hardy–Weinberg "Law" is not a physical law, like Newton's Law of gravitation, which might be supposed to hold with absolute precision. It is only an approximation to reality, describing with fair accuracy the behaviour of real populations; but one which cannot be expected to hold completely exactly, even if only because of chance fluctuations in the sample numbers. There is therefore little point in testing whether a population obeys the formula, since we are reasonably sure that it does not. It would be much more sensible to ask what upper and lower bounds to the deviation from Hardy–Weinberg are reasonably consistent with the observations. In many cases the lower bound will be zero, i.e. the observations would be consistent with no deviation, which we can interpret as meaning that the deviation is too small to be detected statistically in a sample of the size actually available.[b] (Reproduced from Smith[719] with the kind permission of the *Annals of Human Genetics*)

There have been considerable moves in the medical sciences to "tidy up their act" with respect to statistical testing, and this would be welcome in forensic science. This section often follows the principles expressed by Sterne[727] or Nickerson.[581] For differing views or debate, see Chow[178] and the subsequent discussion.

The concept of independence testing was first introduced by Fisher. He appears to have viewed the *p*-value as an indicator of discrepancy between the data and the null hypothesis.

> If *P* is between 0.1 and 0.9 there is certainly no reason to suspect the hypothesis tested. If it is below 0.02 it is strongly indicated that the hypothesis fails to account for the whole of the facts. We shall not often be astray if we draw a conventional line at 0.05.[303]

[b] I am grateful to Dr. Karen Ayres for bringing this material to my attention.

However, it does seem that Fisher did not intend 5% (or 1%) to be the division line between "significant" and "not-significant" in the strict way that it has become. This latter approach was developed by Neyman and Pearson,[579] and led to the decision thresholds that are in common use today.

> No test based upon a theory of probability can by itself provide any evidence of the truth or falsehood of a hypothesis. … Without hoping to know whether each separate hypothesis is true or false, we may search for rules … [to ensure] that … we shall not often be wrong.

This quote should send warnings to the large number of forensic authors who perform independence testing for the exact purpose of determining whether Hardy–Weinberg or linkage equilibria are the case. However, papers continue to appear commenting on the correctness or otherwise of the hypothesis after significance testing.[60]

Fisher never agreed with the stance of treating hypothesis testing as a decision rule in the absence of considerations of all the other information:

> The attempts that have been made to explain the cogency of tests of significance in scientific research, by reference to supposed frequencies of possible statements, based on them, being right or wrong, thus seems to miss the essential nature of such tests. A man who "rejects" a hypothesis provisionally, as a matter of habitual practice, when the significance is 1% or higher, will certainly be mistaken in not more than 1% of such decisions. However the calculation is absurdly academic, for in fact no scientific worker has a fixed level of significance at which from year to year, in all circumstances, he rejects hypotheses; he rather gives his mind to each particular case in the light of his evidence and his ideas.[306]

> The Neyman–Pearson approach to hypothesis testing has in practice been simplified into acceptance or rejection of the null hypothesis without consideration of the alternative hypothesis or the power of the study … our teaching of statistical inference should continue to move away from decisions based on statistical significance and towards interpretation of results based on both the statistical analysis … and wider considerations…a common and serious mistake … is to misinterpret a large *p*-value as meaning "the null hypothesis is true." This is a particular problem with small samples. Because a small sample provides very little information. (Sterne.[727] © 2002 John Wiley & Sons Limited. Reproduced with permission)

With these pertinent and salutary comments, let us turn to a more detailed commentary of some pitfalls associated with independence testing in forensic DNA work.

5.4.3.1 Weakness of the Tests

It is important to note that while Fisher's exact test is the method of choice for investigating departures from independence in this situation, it has limited power unless databases are very large.[831,838,842] There is now widespread acceptance that the power of independence testing to find realistic disequilibrium is very restricted,[121] although this knowledge has existed for some time. "At least since 1970 there have been many studies suggesting that reasonable levels of departure from HWE are practically impossible to detect (with high power) with data from a single population, unless the sample sizes are prohibitively large."[121] In Table 5.5, we reproduce some power estimates. This table may be interpreted as follows. If the disequilibrium in the population was characterized by an inbreeding coefficient of size $\theta = 0.03$ (say), and a sample of size 200 was drawn from this population, then we would expect a significant (<0.05) p-value only 10.7% of the time. In other words, most of the time we would not find disequilibrium of this size with samples of size 200. Sample sizes in much forensic work are of the order of 200, although data sets range from as many as a few thousand to as few as 80.[133] There is nothing inherently wrong with these smaller surveys unless they are overinterpreted. The most obvious overinterpretation is to suggest that they can validate a population genetic model.

> It is not likely that Hardy–Weinberg disequilibrium, at the level thought to exist in human populations, will be detected with samples of 1000 or less. Weir[842]

If the power to find disequilibrium is weak, how can such tests ever be used to infer that disequilibrium is not there? The answer is that they cannot. In fact, independence testing of relatively small data sets does very little to inform us whether or not independence exists in the DNA context.

Table 5.5 Power Estimates for the Exact Test

θ	Sample Size	
	80	200
0.00	4.8%	4.9%
0.01	5.8%	5.6%
0.03	8.1%	10.7%

Following Buckleton et al.[113] © 2003 ASTM. Reprinted with permission.

Therefore, it is logical that the validation of the product rule cannot proceed from independence testing of a database. Why then has an independence testing industry developed? (I would have to include myself in this industry.)

The most prevalent type of misinterpretation is to assume that this testing somehow proves independence. Statements such as "provide little evidence of departures from HWE," or that "based on these observations, the data do not support any significant departure from independence between pairs of loci in any sample population" are strictly correct but deeply misleading. It is often claimed after independence testing that "the application of the product rule is valid for estimating the rarity of a multiple loci profile for these tests." Formal testing procedures, when applied to databases of at most a few hundred individuals, do not have sufficient power to show that the underlying population is not in Hardy–Weinberg equilibrium, or is substructured, or is admixed. This fundamental misconception is well known but still prevalent. See, for instance, Nickerson[581] (p. 260). The reader may be led to ask "How big should the samples be in order to validate the product rule?" The answer is very large indeed, of the order of many thousands, if you plan to proceed by independence testing. However, there is a way to investigate these models that relies less on huge samples but more on population genetics. This will be discussed later.

I conclude this section with a quote from Bruce Weir,[838] which may lighten this dour narrative and contextualize it against the realities of the adversarial legal system. "My own involvement in the U.S. courts ended after a case in Colorado when the defense objected that my use of Fisher's exact test for independence was hearsay and that the prosecution needed to call Mr. Fisher to the stand. It was clear that the public defender was not at all interested in independence testing and had no idea that [Sir] Ronald Fisher had died in 1962. He was simply interrupting my testimony. I was wasting my time."

5.4.3.2 Assuming That Independence Testing Measures Departure from Independence

It is important to note that independence testing does not measure departure from independence. Again, this is a well-known misconception.[581] For instance, a large p-value (close to 1) in the exact test is not proof of independence, nor does it prove that the population must be close to independence. Furthermore, in a large data set we expect to find small departures from Hardy–Weinberg equilibrium. However, the abundance of data means that these effects may be detected (even though they are small) and hence exact tests will assign significant p-values. Small data sets will give p-values anywhere in the range [0…1] almost independently of the amount of departure. It is a fundamental and elementary statistical error to equate the p-value with the extent of departure. The only conclusion that we can make from independence testing of a few hundred individuals is that gross departures of an extent not expected in reality should have been found. This is a barely useful statement.

5.4.3.3 Extrapolation from Single-Locus Tests

Another potential trap is discussed here. Consider the 13 CODIS loci. Assume that independence testing on, say, 200 individuals has "passed" most or all loci for the Hardy–Weinberg test, and has also passed most or all of the 78 pairs of loci. It is tempting to assume that all loci are in Hardy–Weinberg equilibrium and that all 13 loci are collectively in linkage equilibrium. Indeed, NRC II did suggest this line of logic. However, this approach contains an array of extrapolations that cannot be substantiated. Consider the following exaggerated anecdote to demonstrate the point.

Once a young man decided to work out if the world was flat. On a very calm day, he went to sea and took his one meter ruler. He placed the ruler on the sea and observed that it was level. Then he rowed a meter further on and did the same again. In fact, he did a series of level one meter sections until he was back where he started and declared that the world was flat.

This inference is much the same as assuming that we can accept that Hardy–Weinberg and linkage equilibrium exist from investigations at single locus and at pairs of loci. To infer therefore that the model is robust across 13 loci is an extreme extrapolation and does not appear to be justified by simulation experiments. There are many population genetics effects that would give only small indications of departure at one or a pair of loci, but that would accrue to a moderate effect across many loci. In fact, most realistic population genetic effects display this property.

5.4.3.4 Assuming That Other Weak Tests Support This Weak Test

Given the knowledge that independence tests are weak, it would be correct to turn to other sources of data to decide which population genetic model is most suitable. Can we somehow combine many different tests to conclude that, although each has low power, collectively they support the concept of independence? This is a more reasonable approach, but we should tread carefully. Note that the sum of the conclusions is not necessarily the conclusion of the sum. It could be suggested that there is ample evidence that human populations are in equilibrium, and indeed there are a large number of publications in this area, most of which can show that human populations are close to equilibrium; however, occasionally these studies give indications of departure. However, while it is true that human populations are close to equilibrium taken in total, the population genetic evidence suggests that they cannot be exactly in equilibrium. There is no study ever published that proves that a human population is in exact equilibrium.

However, as an exercise we can still examine the suggestion. Below is a histogram of the *p*-values, whether Hardy–Weinberg or linkage, from all the Caucasian STR databases that happened to be on my laptop (most of my files are archived and are not included). These included databases from the

U.K., New Zealand, and Australia, of varying sizes from 200 to several thousand. If independence is true, they should distribute themselves evenly between 0 and 1, that is, they should have a uniform distribution, $U[0,1]$. Figure 5.3 gives the results. The reader should decide whether they are uniformly distributed between 0 and 1 or whether there is a slight aggregation toward low p-values. Remember this trial was attempted among Caucasians, which is one of the races closest to equilibrium. Evidence for much more marked disequilibrium exists in other data sets for different races that we have studied.

5.4.3.5 Post Hoc Rationalization

One practice that is particularly dangerous, but prevalent, is the practice of post hoc rationalization when testing data sets. There are a number of post hoc data treatments that can occur. The process is often to perform independence testing and then find which data are causing the departures. Scrutiny of such data may reveal a legitimate reason to remove them, such as a typographical error or a reason to doubt the ethnicity of the sample.

It is disappointing for the person examining the data to find that most of the departure in an analysis is due to one datum. However, this is quite often the case. In most cases, highlighting this datum leads to some reason to doubt its correctness and it is often either corrected or removed. However, what chance is there really of finding disequilibrium if every time we find it we also find some reason to remove the data causing it? All we are left with are those instances where disequilibrium is broadly based on a large number of genotypes. My colleagues and I have personal experience of quite a few of these as well, although mainly in non-Caucasian populations.

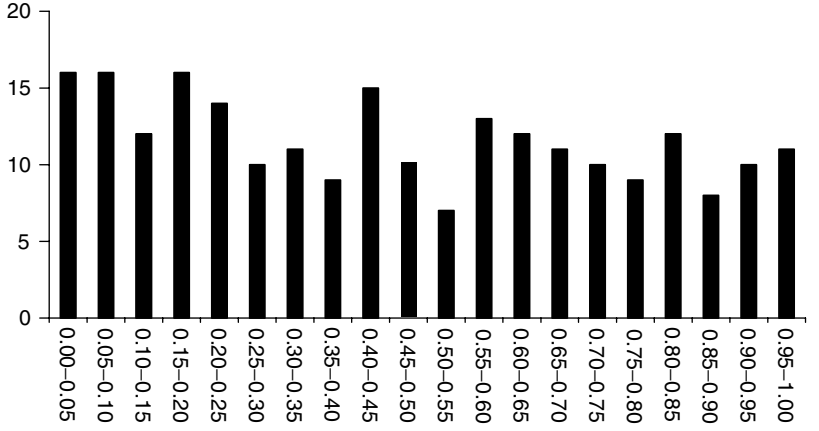


Figure 5.3 Histogram of assorted p-values from some independence tests on Caucasians.

5.4.3.6 Multitesting

Another problem arises from the multitesting nature of the problem. If we examine, say, 13 loci, there will be 13 Hardy–Weinberg tests and $\frac{1}{2}[N(N-1)]=78$ tests between pairs of loci. A significance test at the 5% level would be expected to give a few significant results even if the null hypothesis were true. Weir[842] discusses what should be done, or what should not be done, when a test does cause rejection. "To ignore single rejections on that (the multitesting) basis calls into question the logic of performing the test in the first place." Weir points out the shortcomings of the Bonferroni correction, which requires each of x tests to meet a $0.05/x$ significance level in order to declare rejection, describing it as "unduly conservative." Note that the word "conservative" in this sense does not have the desirable properties that it has in much of the rest of DNA work.

An elegant way to deal with multiple tests like this is described by Zaykin et al.,[880] who follow Fisher. This involves forming the sum of $-2\ln(p)$ across, say, x independent tests.[c] This is expected to have a chi-square distribution with $2x$ degrees of freedom and is known as the truncated product method. An example is given in Box 5.2.

Another useful way is treat the multiple comparison problems by examination using a p–p plot (Figure 5.4 and Box 5.3). In this examination the p-values are expected to be uniformly distributed between 0 and 1 if the null hypothesis (independence between alleles) is true, and therefore should lie along the diagonal. The envelopes of values that would be expected to enclose 99% of the points in the null case are superimposed.

Figure 5.4 provides clear evidence for an excess of low p-values for the Eastern and Western Polynesian populations.[d,e] There is a possible indication of deviation from linkage equilibrium in the Caucasian population.[f] Due to the smaller size of the Asian data set,[g] we would not expect to find disequilibria whether or not they were present.[812,814]

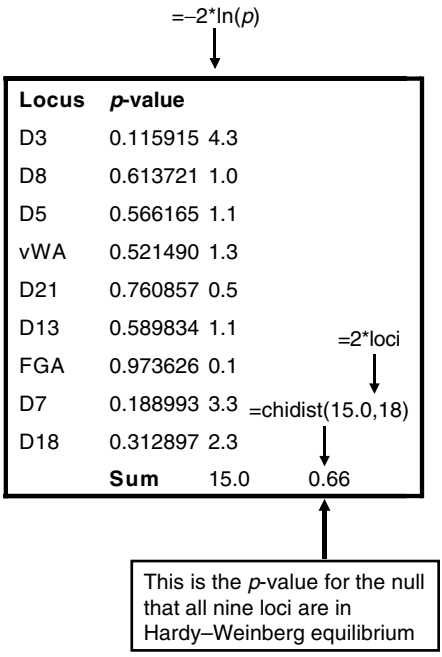
5.4.3.7 Effect of Silent Alleles

Significance tests appear useful to detect genetic phenomena that lead to excess homozygosity, especially primer–dimer mutations. If there is a mutation at the 3′ flanking region of the primer binding site, then PCR can be completely inhibited. The result is that the genotype will appear to be a homozygote. An excellent example is given by Budowle et al.,[136] who observe that binding

[c] The tests in a DNA database analysis are not all independent. For example, if we have tested the pairs HUMD3S1358–HUMD8S1179 and HUMD3S1358–HUMD5S818, then we have some information about the pair HUMD8S1179–HUMD5S818. However, the approach is useful nonetheless.
[d] N=4222 SGM and 1815 SGM$^+$ profiles and N=828 SGM and 477 SGM$^+$ profiles, respectively.
[e] $p<0.001$ on omnibus tests using the truncated product method.[881]
[f] p=0.01, N=2309 SGM and 1001 SGM$^+$ profiles.
[g] N=59 SGM and 114 SGM$^+$ profiles.

Box 5.2 An Example Using the Truncated Product Method for Hardy–Weinberg Tests on Nine Loci

This approach could also be used on the 36 (or any number) linkage equilibrium tests for nine loci (not shown). It is set up in EXCEL but any equivalent package should suffice.



This approach assumes that the tests are independent, which they are not. It is a useful approach nonetheless.

site-induced allelic dropout was present at HUMD8S1179 in Chamorros ($n = 568$) and Filipinos ($n = 574$). Thirteen individuals typed with the PowerPlex 16 kit were heterozygotes, whereas only single alleles were observed with the Profiler Plus kit. The observation of a deviation from Hardy–Weinberg proportions at this locus in these two populations suggested that further investigation was merited, and indeed led to the discovery of the silent alleles.

Such silent alleles would be expected at low frequency at all loci. This will increase counts of homozygotes, albeit mildly.

5.4.3.8 Misuse of Genetics

In Chapter 3 we discussed the assumptions that underlie the Hardy–Weinberg law and the state of linkage equilibrium. Often validation of the use of the

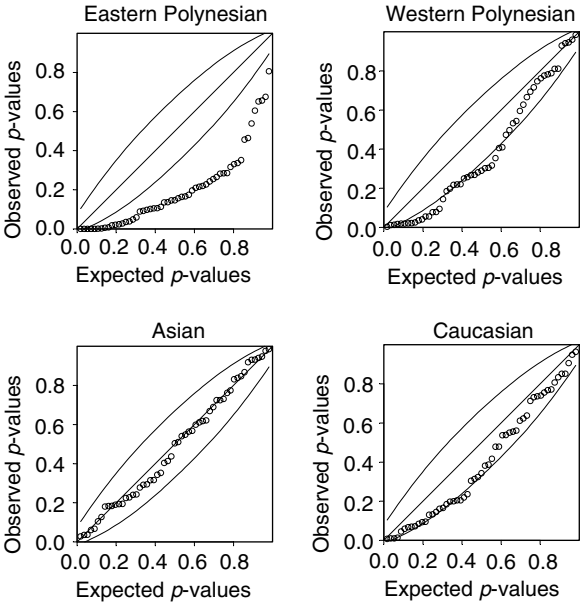
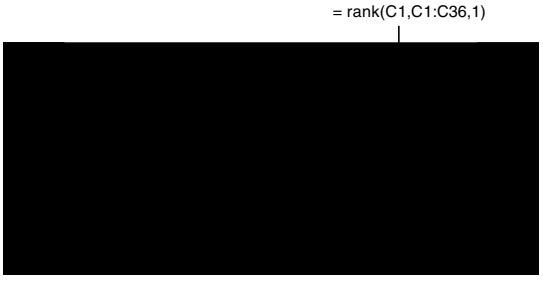


Figure 5.4 *p–p* plots investigating deviations from Hardy–Weinberg and linkage for each of the four major subpopulations in New Zealand.[780,812]

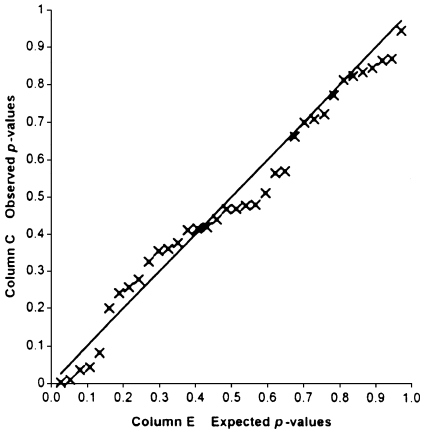
product rule contains an amalgam of statistical and genetic logic. In principle, this approach is perfectly reasonable but the practice is often wayward. For instance, it would be wrong to point out that some genetic conditions leading to equilibrium are present but not to point out that others are not.

Box 5.3 Hypothetical Example Showing the Creation of a *p–p* Plot for the 36 Linkage Equilibrium Tests for Nine Loci (Again it is set up in EXCEL)



This table continues for the 36 values… .

Plot column C on the *y*-axis and column E on the *x*-axis. It helps to add the line of slope 1. A deviation from equilibrium is shown by a deviation from the diagonal line.



More sophisticated packages can place the 95% or 99% confidence interval on the *p*–*p* line.

It is misleading to suggest that validation of the product rule has much, if anything, to do with allelic segregation. In particular, it would be wrong to suggest that linkage implies linkage disequilibrium and that a lack of linkage implies linkage equilibrium. The most likely causes of linkage disequilibrium are not genetic linkage but population genetic effects such as population substructure.

The linkage state of two loci will suggest the rate at which equilibrium will be established after a disturbing event if all the disturbing forces are removed. For example, allelic associations decay by a factor of $1-R$ each generation, where R is the recombination fraction. For unlinked loci $R = \frac{1}{2}$, and hence any allelic associations halve every generation but only in the absence of any disturbing forces. Where disturbing forces such as selection or continuing population subdivision and admixture continue, they will maintain the disequilibrium even for unlinked loci. In this context, Thompson[763] terms them "maintaining forces."

5.5 Estimating θ

A logical approach informs us that the validity of the product rule does not follow from independence testing of a database but from belief in a population genetic model. Independence testing is virtually irrelevant to the interpretation of DNA evidence because it does not realistically inform us about the validity of any model that we may apply. Our belief in the validity of

predictions by the product rule or any other model is based largely on our belief as to the correctness of the model.

These beliefs can only arise from well-constructed population genetic examinations[600] that assess the population subdivision at the genetic level.

One logical way to examine the validity of the population genetic model is to estimate the parameter often called θ, or F_{ST}. This approach is markedly superior to testing hypotheses about independence (which is equivalent to stating that $\theta = 0$).

In cosmopolitan Caucasian populations in the U.K., a θ value near 0.005 may be appropriate. At this level, the estimates from the subpopulation formulae and the product rule start to converge. Therefore, the practical difference between the two approaches becomes minor. However, an important application difference still remains. The product rule is still based on assumptions of independence and its use is likely to lead the court into irrelevant discussion of independence testing. The subpopulation formulae, even when used with a very low θ, should not lead a court to discussions of independence testing but rather to discussions of the population genetics of the population in question. This latter topic is likely to be far more productive for understanding and the course of justice. Again, my argument relies on an appeal to sounder logic rather than any large-scale practical significance, although Birus et al.[69] warned of significant effects that they had noticed while examining war victims in Croatia.

The parameter θ has been defined in a number of ways, but the most intuitively pleasing is the definition based on identity by descent. Under this definition, θ is taken to be the probability that two alleles in different people in the same subpopulation are identical by descent. It therefore describes the increase in relatedness in the subpopulation relative to the population. Another intuitively pleasing way to view θ is as a measure of the genetic distance between two subpopulations that are diverging by drift.

The logic of this approach would be to estimate θ between the subpopulations making up the population in question[712,842] and to decide if it is "near enough" to zero to be ignored, or preferably to make an appropriate correction. "Given that human populations do have nonzero values of θ, there is some appeal to making probability statements about θ lying in a certain range rather than simply failing to reject the false hypothesis that it is zero."[842]

As previously stated, θ is a difficult parameter to estimate.[834] There are issues relating to both the selection of samples and to the method of estimation of the parameter. Three methods that have been employed in forensic work are summarized here (Table 5.6). The method of Weir and Cockerham[846] gives a point estimate, whereas the two Bayesian methods, that of Balding and Nichols[39] and Foreman et al.,[312] give a probability distribution for the parameter, θ. Many other methods exist.[580] For a review see Excoffier,[283]

Table 5.6 A Summary of Three Approaches to Estimating θ

Method of Estimation	Sample Requirements
Weir and Cockerham[846]	The Weir and Cockerham method requires two or more samples from the subpopulations making up the population. The representativeness of the final estimate will depend on how well the samples represent the subpopulation diversity in this population. Methods are available which proceed from either genotype data or summary allele probabilities. Testing of this approach by simulation verifies that it accurately estimates θ if the samples are correctly drawn from subpopulations.
Balding and Nichols[39]	The Balding and Nichols approach requires data from only one subpopulation and a "reference" population. The reference population should reflect the population-wide "average." This approach is very scientifically appealing. It produces output that allows the selection of a range of "supported" values for θ. Practically this estimation procedure has been found to be very sensitive to the choice of reference population and due care must be taken with this aspect.
Foreman et al.[312]	This method requires only one population and seeks to estimate a parameter by investigating partitions in the data. Trials[603] suggest that it works adequately if the number of partitions is known. Doubts have been raised that the parameter that it seeks to estimate is indeed the genetic parameter θ.

who points out that "Even though the analysis of microsatellite data has exactly the same form as that of other data types, estimators based on such data have a much larger associated variance…" due largely to the single-step mutation processes thought to occur at these loci. The consequence of this is that large samples do not necessarily produce consistent estimates and a very large number of independent loci are needed to obtain meaningful estimates.

The issue of sample selection is also absolutely crucial. The comparison of geographically based samples from, say, different cities in the U.S. is unlikely to reveal the underlying genetic diversity that is to be measured. Societies in the "new world," especially in countries settled largely by recent colonization, are likely to be mixes of immigrant populations.[117] Suppose that we took two samples of, say, 33% Scottish, 33% English, and 33% Italian. These two samples will probably be very similar. Whereas if we compared separate Scottish, English, and Italian samples, we will find small, but real, differences. The result will be underestimates of θ. Even in the comparison of geographical samples in the "old world," it is preferable to get ethnically defined samples.

Comparison of different ethnic groups from Europe, Asia, or Africa is more likely to reveal the underlying genetic diversity now embedded and partially diluted in the U.S., Australia, and other countries populated by immigration.

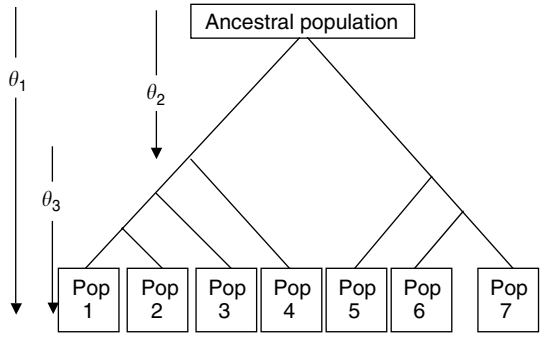


Figure 5.5 Stylized description of a set of populations. Adapted from Weir[839] with the kind permission of Kluwer Academic Publishers.

Weir[839] gives an illuminating description of a highly idealized human population using a tree diagram such as Figure 5.5. This diagram depicts a population diverging by drift with a series of bifurcations. The value of θ obtained is influenced strongly by the selection of populations.

For instance, a comparison of population 1 and population 5 is likely to attempt to estimate θ_1. A comparison of population 1 and population 4 is likely to attempt to estimate θ_3. Imagine that our sampling procedure produces a mixed sample from populations 1–4 and compares this with a mixed sample from populations 5–7. This is likely to attempt to estimate a parameter probably closest to θ_2.

Weir further points out that the use of, say, θ_3 in the calculation of the conditional probabilities requires the allele probabilities from the ancestral population that formed populations 1 and 4. These probabilities are not available for many reasons, but not the least of which is that this population no longer exists. The allele probabilities for this ancestral population may, or may not, be adequately approximated by the average of populations 1–4, but it would be difficult to know how good or poor this approximation may be.

It should also be remembered that when we are considering 13 locus CODIS profiles, we are dealing with a polynomial in θ containing terms in up to the 26th power. The average of terms in such high powers is dominated by the contribution from a few large values. It may be appropriate to take a value for this parameter, θ, at the high end of the plausible distribution.[36] Therefore, when examining the estimates given later in this chapter, it may be wise to err on the high side.

5.5.1 Historic and Modern Consanguineous Marriage

A summary of historic consanguinity does not have any direct implication for modern θ values; however, it may imply that there was a high level of

inbreeding in historic human populations. The question of whether or not this has been erased in modern times is unanswered. However, the prevalence of consanguinity in ancient times and its considerable persistence into modern times is worthy of discussion in the context of forensic DNA interpretation. This is because the mating patterns of humans in the past are likely to impact on modern θ values. In particular, we do not need to consider simply consanguineous marriage, which is a relatively extreme form of inbreeding, but we also need to consider any restriction of the mating choice. The most obvious noncultural restriction is geographical availability.

It is worth noting that there are several cultural forces driving consanguinity. These include maintenance of family property and bride wealth. In societies where inheritance is divided among progeny, there is a considerable incentive to consanguineous marriage. In addition, the economic burden of bride wealth is an important driver of consanguinity. Under these circumstances, consanguineous marriage is economically the most feasible option where culturally permissible.

Bittles and Neel[71] present a review specifically with the intent of assessing the load of lethal recessives in the population and their impact on modern consanguineous marriages. "As the great religions emerged (Buddhism, Confucianism, Islam, Hinduism, Judaism, Christianity) only one — Christianity — had specific proscriptions against nonincestuous consanguineous marriage, and even those were not enunciated until the Council of Trent in the mid-sixteenth century."[71] These authors conclude that the relatively low modern load of lethal recessives is strong evidence for ancient inbreeding.

Much of the subsequent discussion follows Bittles.[70] "'Western' opinion tends to be that whether or not consanguineous marriage was present in the past it has largely disappeared in modern society. This is untrue and it is worthwhile reviewing the evidence for both the relatively recent prohibition of consanguinity and the persistence of it into our generation." Prohibition on second- and third-cousin marriages was formally rescinded by the Roman Catholic Church in 1917. Specific dispensation remains a prerequisite for Roman Catholic marriages between first cousins, who wish to have their marriage recognized by the church.[70] First-cousin marriages are criminal offenses in eight of the 50 United States and are subject to Civil sanction in a further 31 under statutes introduced from the mid-19th century onward.[598] Exceptions have been incorporated for specific communities. For instance, uncle–niece marriage is permissible for Jews in Rhode Island. It may be of interest to the reader to note that Charles Darwin was married to his cousin Emma Wedgewood.

In North America and Western Europe, the rates of first-cousin marriage are about 0.6%,[190,495] with Japan at about 3.9%.[428,709] Bittles et al.[72] give estimates in the range of 20–50% for marriages between second cousins or closer

in the Muslim countries of North Africa, Central and West Asia, and in most parts of South Asia. The preferred type of marriage in Muslim society is for a male to marry his father's brother's daughter. Uncle–niece marriage is prohibited by the Koran. In the primarily Hindu states of southern India, 20–45% of marriages are consanguineous, with uncle–niece and mother's brother's daughter being especially popular.

Information from modern China is not available. However, before the Second World War, first-cousin marriage of the type mother's brother's daughter was the accepted custom among the Han, who make up approximately 90% of the population.

Bittles concludes that consanguinity is not an unusual or rare phenomenon, but rather was and is the preferred or prescribed form of marriage in much of the world. The expected reduction in consanguineous marriage in the latter half of the 20th century does not appear to be happening universally. Bittles also points out that many immigrant groups maintain these cultural norms when they move to Europe or North America and the inbreeding may even be enhanced by the reduction of choice in smaller communities.

5.5.2 Summary of Published Estimates of θ

This section summarizes some published estimates of θ. The absolutely crucial nature of selecting samples so as to expose the underlying genetics can be clearly seen from this comparison.

> F_{ST} values, reflecting the contribution of differences among subpopulations to the total divergence, strongly depend on how these subpopulations are defined within the total population.[886]

Comparison of different estimates is complicated by a number of factors. One of these is that the two Bayesian approaches produce distributions rather than point estimates. These distributions need to be summarized in some way and various papers have opted for different methods of summarization. In most cases, the modal value of the posterior distribution is quoted; however, these distributions are typically skewed, and a value from the right-hand tail may be more appropriate. The mode may be a very poor summary of the full distribution.

Balding and Nichols[41] discuss whether or not we should expect θ to be the same at all loci. It is indeed possible that each locus has a different evolutionary history and hence a different θ. One obvious factor that may affect certain loci more than others is mutation rate. High mutation rates are expected to lower θ. However, we expect θ to be similar across loci within broad classes, such as STR and VNTR loci, and determined more by the history of the populations compared.

The largest compilation discussed below is that of Cavalli-Sforza et al.[152] published in 1994. They survey a very large number of well-defined populations. Due to the publication date, none of the data are based on STR loci and many of them are based on conventional blood group polymorphisms; however, it is understood that an update is in preparation. Some at least of these conventional loci are known to be affected by selection. The purpose of the compilation was not to furnish forensic θ estimates but, rather, to study human history. Accordingly, the θ estimates are given between pairs of populations and often built into trees. The forensic problem requires a different distance: that between one population and some form of average represented by the database to be used. However, the compilation does represent a very comprehensive collection of genetic distances between well-defined populations. It is not clear how much inference about STR θ values can be taken from this study; however, in broad terms it must represent the most extensive current compilation. The data are so comprehensive that it is not possible to summarize them in any but the briefest way here.

5.5.2.1 Caucasian

In their compilation, Cavalli-Sforza et al.[152] prefer to work with the geographical term European rather than the term Caucasian. They study 88 genes (all non-STR) and identify four major outlier populations: Lapps, Sardinians, Basques, and Icelanders. If this is borne in mind and the θ tree and table (Figure 5.5.1 and Table 5.5.1 of Cavalli-Sfoza et al.[152]) are examined, a reasonable estimate can be made. This would suggest that the major subpopulation components of the population in consideration, say U.S. Caucasians, should be determined and a value selected from this graphic that captures most or all of the expected variation. Subjectively, a value of 0.02 would appear to capture most of the variation present in Europe, and the majority would still be "captured" by a value of 0.01, which is the value recommended in NRC II.[585]

Foreman et al.[314] apply Balding and Nichols' Bayesian method to various geographical subgroups of the U.K., such as Glasgow, Neath, and London, utilizing the six autosomal STR loci of the SGM system. This seems a reasonable approach to attempting to get at the source of genetic variability in the U.K., and could only be bettered by extensive pedigree examination on each donor to try to obtain people who had a long family history in one region. They report the posterior modes for θ that vary between 0.0001 and 0.0036. The highest value occurs for the comparison of Merthyr with the FSS Caucasian database as reference.

Balding and Nichols[41] give distributions using their Bayesian method for Finnish, Italian, Norwegian, Spanish, and Turkish samples using a general Caucasian reference database, for five VNTR loci. They do not appear to

summarize their distributions themselves, but from the graphs the modes appear to range from about 0.004 to about 0.020. This study effectively rebuts the assertions[122] that there is no difference between subpopulations, and supports the position of Sawyer et al.[691] and Krane et al.[479] This point was conceded by Budowle[154] (p. 434). The same approach was applied by Balding et al.[43] to four autosomal STR loci (the FSS Quadruplex set) using a mixed British Caucasian database as reference against the population samples of Caucasians from Derbyshire, Dundee, Northern Ireland, and Strathclyde. Modal values (read from the graphs) are of the order of 0.001–0.004. For the set of populations of Greeks, Greek Cypriots, and Italians, the modal values are approximately 0.004, 0.021, and 0.002, respectively. Another set of estimates for two Greek populations, Sarakatani and Vlachi, is given by Balding and Nichols[41] and also in Overall.[599] Posterior modes lie between 0.005 and 0.010 as reported by Balding and Nichols and close to 0.010 as reported by Overall. Overall also gives estimates for Helston (Cornwall), Welshpool (Wales), and Ullapool (Scotland) from people with known ancestry in the area. Posterior modes were 0.000, 0.002, and 0.018, respectively. Overall[599] gives an additional estimate for Tuam (Eire) of 0.005.

Gill and Evett[354] give estimates ranging from −0.0004 to 0.0029 for the set U.K.(FSS), U.K.(MPFSL), U.K.(Derbyshire), U.S.(Foster City), U.S.(Army), and Sweden for the six autosomal STR loci of the SGM set.

In Poland[869] estimates ranging between 0.0003 and 0.0044 for various Polish populations, and 0.004 for south Poland[870] have been reported. A value of 0.003 was given for the Byelorussian minority of northeastern Poland.[613]

Weir[832] gives estimates for three geographically defined Caucasian databases from Florida, Texas, and California for six VNTR loci. The estimates found using the method of Weir and Cockerham vary from −0.001 to 0.004. However, this study is unlikely to have segregated the groups into genetically defined subpopulations and hence must be seen as a lower bound.

Budowle[117] gives results using the method of Weir and Cockerham for Caucasians drawn from the FBI data, Roche, Alabama, Switzerland, Israel, and French Basques using the six loci of the Polymarker and DQα set. The first three samples are presumably all U.S. Caucasians. The values reported are 0.0034 for the full set of five subpopulations and 0.0015 if the French Basques are omitted. Later Budowle et al.[120] gave data for 13 STR loci for nine U.S. Caucasian populations as −0.0005 and for 11 European populations as 0.0028. The first set will suffer from the sample selection issues discussed. Budowle et al.[120] give a value of 0.001 for Omani and Dubai Arabs for nine STR loci. NRC II[585] gives an estimate of 0.0015 for D2S44 for the Canadian, Swiss, French, and Spanish groups. Sun et al.[743] give a value of 0.0022 for G_{ST}

for the Caucasian set; German, Spanish, United Arab Emirates, and Brazilian.[h]

Greenhalgh and Buffery[371] take a substantially different approach, in which they directly investigate the initial Balding and Nichols[36] hypothesis. Specifically, it has been suggested that the population of suspects may include an overrepresentation of the subpopulation to which the defendant himself belongs. They report two large investigations, one targeting Afro-Caribbeans and the other Caucasians, in each of which over 200 "suspects" were included. The distribution of the VNTR types from these two investigations was compared to that of the general population and very little difference could be seen. Unfortunately their work does not result in an estimate of θ, but it does suggest that any effect, if present, is small, and that pools of suspects may not be greatly different from their associated racial group in general. With hindsight we now understand that an overrepresentation of the defendant's subpopulation in the pool of suspects is not a prerequisite for there to be a subpopulation effect. The subpopulation effect can occur if there are any members of the defendant's subpopulation in the pool of suspects. However, this was a valuable and insightful investigation of one aspect of the initial hypothesis.

Other evidence for the survival of ancient population diversity comes from anthropological investigations. An example is Y chromosome data from samples taken in a transect across Britain. This study suggests the survival of Celtic populations in the north of Wales and populations with a larger Anglo-Saxon[i] component across England.[822]

5.5.2.2 Black African, Caribbean

Again, great care must be taken with the use of these broad racial terms. For instance, it is thought that the current population of Africa is affected "by the relatively ancient presence of Caucasoids in the northern strip along the Mediterranean, and additions from West Asia."[152] Cavalli-Sforza et al.[152] study 49 populations for an average of 48 genes (no STR loci). The division

[h] G_{ST} is expected to be numerically similar to the preferred F_{ST}.

[i] Following Morris,[561] we note that Anglo-Saxon is actually a relatively modern term. Post Roman British called themselves *Cives* in Latin, meaning fellow countrymen, *Combrogi* in ancient British, *Cymry* in modern Welsh, and *Cumber* in modern English. The English knew the ancient British by both ancient names, but added a third, calling them foreigners, *Wealh* or *Wylisc* in Old English, and *Welsh* in modern English. The newcomers from Germany were from many different nations, but in Britain they adopted a collective name: *Engle* or *Englisc*, which Latin writers wrote as *Angli*. The word *Angle* is a modern transliteration of the ordinary Latin word for English. *Saxon* was the term applied to all of the immigrants and not simply the West, East, South, and middle Saxons. Eighth-century writers coined the term *Angli Saxones* meaning English Saxons to distinguish those of Britain from those of Germany. It was not until the 20th century that the unhappy hybrid Anglo-Saxon prevailed. Morris[561] gives further detail on the origin of the terms Irish, Scot, English, and Welsh.

of sub-Saharan Africa into Northern and Eastern populations is clearly demonstrated. It is becoming accepted that human genetic diversity is highest among Africans, and this is again demonstrated in the study. It is interesting to consider the mix of African populations that are now represented in, say, African-American or Afro-Caribbean populations. There has been an amount of historical study of the slave trade that could inform this, or perhaps mitochondrial typing of modern African-Americans may lead to an answer. Again bearing this in mind, the θ tree and subsequent tables (Figure 3.5.1 and Table 5.5.1 of Cavalli-Sforza et al.[152]) may be examined and a reasonable estimate can be made. Table 3.9.1 of Cavalli-Sforza et al. gives some distances for "West Africans." These values suggest that a θ value of 0.01 as recommended by NRC II[585] would not cover the full genetic diversity from West Africa. The pivotal question would then be what level of mixing has occurred in the U.S. itself?

Zietkiewicz et al.[886] gives a value of 0.072 for Africans. However, the removal of M'Buti pygmies, a known deep division, lowered this to 0.027.

Foreman and Lambert[310] again apply Balding and Nichols' Bayesian method to various groups such as Cardiff Africans and West Midlands Caribbean against the FSS Afro-Caribbean database as reference, for the six autosomal STR markers of the SGM system. It is difficult to see how this sampling could have been improved; however, it can be argued that it may have missed the underlying diversity in Afro-Caribbeans if these groups are now regional conglomerates of genetically diverse groups that had existed separately in Africa. This has to be left as a theoretical objection at this stage until we understand the history of immigration of Afro-Caribbeans into the U.K. in more depth. As a counterpoise to this potential objection, Foreman and Lambert include a South African ethnic group in which the modal values for θ vary between 0.0029 and 0.0082. The highest value occurs for the comparison of the South African samples.

Gill and Evett[354] give estimates ranging from -0.0004 to 0.0009 for the set U.K.(FSS), U.K.(MPFSL), and U.S.(Army) for the six autosomal STR loci of the SGM set.

Weir[832] also gives estimates using the method of Weir and Cockerham for three geographically defined Black databases from Florida, Texas, and California for six VNTR loci. The values vary from -0.002 to 0.001. Again, this study is unlikely to have segregated the groups into genetically defined subpopulations and hence must be seen as a lower bound.

Budowle[117] gave results using the method of Weir and Cockerham for three African-American groups. The value reported is 0.0023 for the six loci of the Polymarker and DQα set. Later, Budowle[120] also gave a value of 0.0006 for 11 African-American samples analyzed at 13 STR loci. It is difficult to see a better way of sampling African Americans in the U.S.; however, it is unlikely that this approach would reveal the full genetic diversity present in Africa or

even in those areas of Africa from which the modern U.S. African-American population was drawn. Sun et al.[743] give a value of 0.0018 for G_{ST} for the African populations; Sudanese, Nigerian, Benin, and South Carolina Black.

5.5.2.3 Asian (Indo-Pakistani)

Cavalli-Sforza et al.[152] present a specific section on the Indian subcontinent, giving information on, for instance, 28 populations studied at an average of 47 genes (although again none are STR loci). This study does suggest that θ values of the order of 0.02 or 0.03 could easily be supported for this population.

Foreman and Lambert[310] again apply Balding and Nichols' Bayesian method to various groups defined by religion, locality, and ethnicity from within the U.K., but including Doabi and Miripuri data. This must be seen as an admirable way to find any underlying genetic diversity. Again they utilize the six autosomal STR markers of the SGM system. Modal values for θ vary between 0.0001 and 0.0063. The highest value occurs for the comparison of the Midlands with the Hindu/Sikh data as a reference. The Foreman and Lambert study may include the data given by Overall[599] for Jullunduri and Mirpuri, giving a maximum likelihood estimate of θ of 0.005 and 0.007, respectively. This conclusion contrasts strongly with the result reported by Zhivotovsky et al.,[884] who examine three co-resident Pakistani populations who favor consanguineous marriage. Using the method of Weir and Cockerham, they obtain a mean value for θ of 0.13 for ten autosomal dinucleotide markers. Requests for the data and pedigrees of the individuals are unanswered at the time of writing.

Gill and Evett[354] give estimates ranging from 0.0002 to 0.0023 for the set U.K.(FSS), Hindu, Sikh, and Bangladesh for the six autosomal STR loci of the SGM set.

5.5.2.4 Hispanic and Latin American

Weir[832] also gives estimates using the method of Weir and Cockerham for two geographically defined Hispanic databases from Florida and Texas for six VNTR loci. The values vary from 0.002 to 0.009. Budowle[117] gives results using the method of Weir and Cockerham for two regionally defined Hispanic populations: South Eastern and South Western. The value reported is 0.0142 for the six loci of the Polymarker and DQα set. Cerda-Flores et al.[155] give G_{ST} values of 0.031 for D1S80 and 0.067 for DQα for three Mestizo populations. Budowle[120] gives 0.0021 for eight Hispanic populations for the 13 CODIS STR loci.

Our understanding of the composition of the Hispanic populations in the U.S. suggests that these studies stand a higher chance of having segregated the groups into genetically defined subpopulations than the same studies on

Caucasians and Blacks. Budowle[120] gives a value of 0.0053 for five population groups in Latin America for 15 STR loci.

5.5.2.5 Amerinds

Kidd et al.[456] investigate three Amerind groups and discuss the phenomenon of drift among Amerinds. A θ estimate is not given in this paper, but allele frequencies are reported which would enable such an estimate to be calculated. This paper has been much debated in court because of the possible presence of a seven-locus match between two of the Karitaina and a six-locus match between a Mayan and a Suri.

Cavalli-Sforza et al.[152] again present a specific section on Amerinds. Their study does suggest high levels of genetic diversity between Amerind tribes. Values of θ of the order of 0.05–0.12 could easily be supported for this population, depending on which Amerind groups may be involved.

Budowle et al.[134] report a value of 0.0309 for three Alaskan populations. Separating on linguistic grounds gives a figure of 0.0167 for the two remaining Alaskan groups. This leaves Athabaskans who were compared with Apaches and Navajos on the basis of a similarity in their language. This yielded a figure of 0.018 for θ. Later Budowle[120] gave a value of 0.0282 for seven Native American samples.

Mesa et al.[548] give a value of 0.068 for G_{ST} for five Native American populations from Columbia for nine autosomal STR markers. Their evidence suggests that all these populations have undergone admixture with Caucasians.

Sun et al.[743] give a value of 0.0407 for G_{ST} for Native Americans, but appear to still conclude that "the entire set of nine loci are mutually independent in all populations." Discussion with the authors appears to suggest that this statement was unintended, but rather that the authors feel that the loci are acceptably close to equilibrium. This amended statement was, correctly, not based on independence testing.

5.5.2.6 East Asian

Budowle[117] gives results using the method of Weir and Cockerham for two samples: Japanese and Chinese. The value reported is 0.0024 for the six loci of the Polymarker and DQα set. Wei et al.[825] give a value of 0.0039 for four Taiwanese populations. Sun et al.[743] give a value of 0.0048 for G_{ST} for "Asian" and 0.027 for "Oceanic." Budowle[136] gives 0.0090 for nine STR for Chamorros and Filipinos.

5.5.3 Dealing with Diverse Populations

What should we do if a part of our population includes individuals from a race that we have not sampled extensively or if the suspect belongs to a race we have not sampled? NRC II[585] Recommendation 4.3 addresses this issue. "If the person

who contributed the evidence sample is from a group or tribe for which no adequate database exists, data from several other groups or tribes thought to be closely related to it should be used. The profile frequency should be calculated as described in Recommendation 4.1 for each group or tribe." Of course, the population to be modeled is not necessarily dictated by the group or tribe of the defendant, but that aside this is a sensible approach to the problem. However, it should be accepted that, on average, a person's genotype tends to be more common in their own group than others. Hence we should expect a person's genotype estimate to be more common on average in their own group than even in several closely related groups. This effect is likely to be very minor. However, an effective mechanism for making estimates of groups that have not been sampled is provided by Balding and Nichols'[36] equations with a suitable value of θ. Using this approach, we would sample the closely related groups and then use Balding and Nichols' formulae with an appropriate estimate of θ. This approach is a preferable refinement of Recommendation 4.3.

5.6 Tippett Testing

A method for investigating the magnitude and consequence of random matches has been championed by Dr. Evett and is colloquially called "Tippett testing." Examples of Tippett plots appear on pp. 213–215 of Evett and Weir[267] and large-scale Tippett-type experiments are reported by Weir.[843] The tests originate from an experiment by Tippett et al.[774] on paint. Dr. Evett has applied the same technique to data from both glass and DNA. In the DNA context, we imagine that we have a database of N profiles. First we perform the "within" experiment. We compare each person in the database with himself. There will be N such comparisons. Obviously, each person matches himself, and we then calculate a likelihood ratio, or any other statistic assessing the strength of the match. In the STR DNA context, this section of the experiment is relatively straightforward. However, in other evidence types such as single-locus probes, paint, or glass, it is much more demanding and revealing. This shows the range of likelihood ratios expected when the suspect is in fact the true offender.

Next we compare each person in the database with every other person. This is called the "between" experiment, and will give us the distribution of the likelihood ratio if the suspect is not the true offender. We will have the results from $N(N-1)/2$ possible comparisons. For demonstration, consider the comparison of 1401 Caucasian FSS Quadruplex genotypes undertaken by Evett et al.[279] For this set there are 980,700 pairwise comparisons. (Note also that not all these comparisons are independent, although the consequences of this are probably negligible.) In almost all of these comparisons, the profiles will be different. In such cases the likelihood ratio is zero. On 118 occasions there was

a four-locus match, and for such occasions a likelihood ratio was calculated. This is as close to a direct measurement of the average match probability as we are going to get. We can say that in 118/980,700=1 in 8311 comparisons we will obtain a match between different people for this multiplex. If the database was constructed from unrelated Caucasians, then we have the estimate that 1 in 8311 unrelated pairs of Caucasians will match at these four loci.

Several things need to be noted about the general Tippett approach. First it makes very few assumptions and hence does not rely to any large extent on models. It is therefore our best approach to directly measuring average match probabilities. However, match probabilities are usually quoted for the profile in question in court. This approach yields an average match probability across all N profiles that exist in the database.

The next thing that may be done with these data is to shuffle the alleles in the database. This effectively imposes Hardy–Weinberg and linkage equilibrium on the data by breaking any possible association between alleles, whether or not it was there originally. We can then perform the between experiment over and over again and obtain the distribution of the number and magnitude of matches expected if independence was true. What we typically note is that this distribution contains the number of matches that we observed in the unshuffled data. Does this test the assumption of independence? Are we entitled to say something like: The number of matches observed is consistent with the assumption of independence?

It turns out that this would be a misleading conclusion. Making databases with known amounts of disequilibrium, possibly by simulation, and performing the experiment can show this. Often enough the databases with deliberately made disequilibrium also pass this test, that is: The number of observed matches and their relative magnitude is also consistent with the assumption of dependence. Hence the Tippett-type tests cannot really distinguish between databases that are in equilibrium and those that are not (there is no current method to do this on databases of realistic size), and consequently they cannot measure the extent of departure. What they do show is that the presence of relatively large amounts of disequilibrium has very little effect on the number and magnitude of matches.

To demonstrate this method, Curran and Buckleton (unpublished results) considered an example given by Foreman et al.[315] They investigated the performance of this concept under two genetic models. One population is in Hardy–Weinberg and linkage equilibrium, and the second is a substructured population characterized by an inbreeding coefficient $\theta = 0.03$ created by simulation. Databases were simulated many times from these populations and the number of matches counted. The method of this simulation is given in Appendix 5.1.

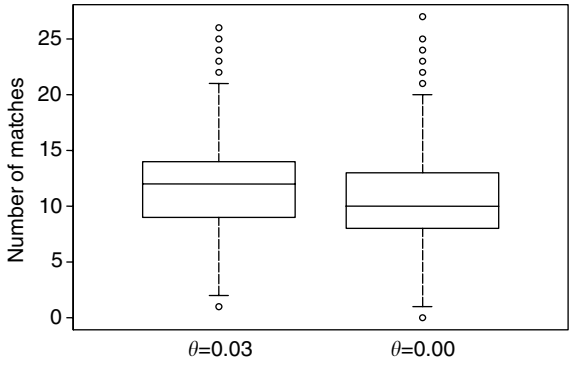


Figure 5.6 Number of pairwise matches per database generated by the Tippett test.

Figure 5.6 shows the distribution of the number of matches under each model. It is immediately apparent that the inclusion of a relatively large amount of disequilibrium has very little effect on the number of matches in each simulation. A direct consequence of this is that the number of matches is a very poor tool to use to distinguish between the independence model ($\theta = 0.00$) and the model with a value of $\theta = 0.03$. This had been previously shown algebraically by Weir.[827]

Next Curran and Buckleton calculated a likelihood ratio for each match to produce the typical "Tippett" graph, shown in Figure 5.7. First we note that the 90% confidence intervals have substantial overlap for the two models tested. It is again apparent that the inclusion of a relatively large amount of disequilibrium has very little effect on the number and magnitude of matches in each simulation.

The conclusion is that the Tippett test is the best way currently available to directly measure average match probabilities. It also demonstrates that the

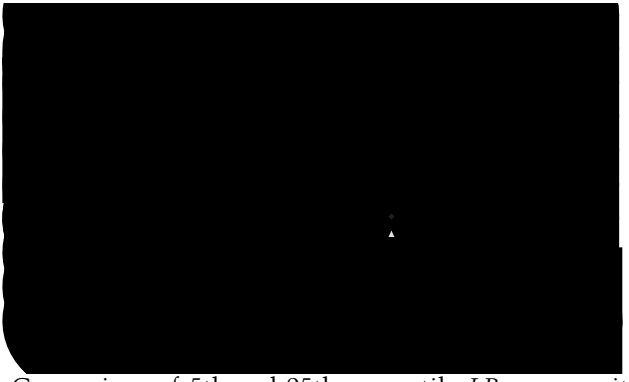


Figure 5.7 Comparison of 5th and 95th percentile *LR* curves with $\theta = 0$ and $\theta = 0.03$.

practical consequences of disequilibrium, if it exists, are not large and offers a way to measure these effects. However, it is a very poor tool for finding disequilibrium or measuring it if present, and is hence a poor tool for validating population genetic models.

5.7 Descriptive Statistics for Databases

Most laboratories that are validating their databases publish the data. This is a very desirable activity. This policy is supported by the editors of many journals who have developed sections such as "For the record." In these sections, short announcements of population data may be published.

These publications typically give a summary of the allele probabilities. This is useful for other scientists who may wish to use these data or to compare them to their own or other populations.

Many publications also include some summary statistics used to describe the data. This move has been facilitated by the provision of software such as Powerstats.[758] The purpose of this section is to give the mathematical definitions of some of these statistics and to make recommendations for their use.

When describing descriptive statistics, it is vital to draw a careful distinction between the value of population parameter and the estimate of that parameter calculated from the sample actually observed.

5.7.1 Heterozygosity

This term is applied to a measure of the fraction of heterozygotes[836] in the population. Let us term the population heterozygosity at locus l in population q, h_l^q. To avoid the proliferation of superscripts and subscripts, we will use h as shorthand, but it must be recalled that this is at a certain locus and in a certain population.

The simplest estimate of this parameter is the fraction in the sample, $\tilde{h}_l^q$, where the $\sim$ is used to signify the sample value. Again we will occasionally shorten this to $\tilde{h}$. This is assessed by taking the count of heterozygotes in the database at this locus, n_l, and dividing by the total samples, N_l:

$$\tilde{h}_l^q = \frac{n_l}{N_l} \text{ (heterozygosity at locus } l \text{ in population } q)$$

The observed sample proportion, $\tilde{h}_l^q$, is expected to be an unbiased estimator of the true parameter, h_l^q; $\varepsilon\tilde{h}_l^q = h_l^q$ with variance

$$\text{Var } \tilde{h}_l^q = \frac{h_l^q(1-h_l^q)}{n_l}$$

if the sampling of individuals is independent.

Weir[836] gives the average over m loci as the simple arithmetic average:

$$\tilde{h}^q = \frac{1}{m}\sum_{l=1}^{m}\tilde{h}_l^q = \frac{1}{m}\sum_{l=1}^{m}\frac{n_l}{N_l} \quad \text{(average heterozygosity across } m \text{ loci in population } q\text{)}$$

This overall average will rarely be required in forensic work as it is customary to report locus specific heterozygosities. Weir states that this remains an unbiased estimator, but that the variance requires cognizance to be taken of the covariance between the estimates at each locus.

5.7.2 Homozygosity

Homozygosity, H, can be similarly defined as the fraction of homozygotes in the sample:

$$\tilde{H}_l^q = \frac{p_l}{N_l}$$

where p_l is the count of homozygotes in population q at locus l. We see directly that

$$\tilde{H}_l^q = 1 - \tilde{h}_l^q$$

5.7.3 Gene Diversity (Often Termed Hex)

Heterozygosity is estimated as the fraction of heterozygote genotypes at a locus in a sample. An alternative strategy would be to proceed from allele frequencies. Let $\tilde{p}_{lu}^q$ be the frequency of the uth allele at locus l in a sample from population q of size n individuals. For simplicity, we drop the subscript for locus and the superscript for populations. The maximum likelihood estimator for gene diversity is given by

$$\hat{D} = 1 - \sum_u \tilde{p}_u^2$$

where the summation is over all alleles at this locus. Weir notes that

$$\mathcal{E}\hat{D}_l = \left(1 - \frac{1+f}{2n}\right)D_l$$

Hence there is a small downward bias, $(2n-1)/2n$, for non-inbred populations ($f = 0$) and a slightly larger one for inbred populations. He also gives expressions for the variance and covariance of $\hat{D}_l$. This suggests the use of

$$\hat{D} = \frac{2n}{2n-1}\left(1 - \sum_u \tilde{p}_u^2\right)$$

to compensate for this bias,[164,575] where n is the number of individuals sampled.

Gene diversity is expected to be similar to, but not exactly the same as, heterozygosity. The difference will be larger for samples that differ from Hardy–Weinberg proportions markedly. Gene diversity should have smaller sampling variance than heterozygosity.

Nei[575] also suggests that the quantity $1 - \hat{D}$ that he terms gene identity may be useful.

5.7.4 Match Probability

The probability of a match at locus l, PM_l, was first described from genotype data. Fisher[304] gave

$$PM_l = \sum_i \tilde{G}_{il}^2$$

where $\tilde{G}_i$ is the sample frequency of the ith genotype at locus l. Jones[447] suggests setting

$$PM_l = \frac{\sum_{i=1}^{n} \tilde{G}_{il}^2 - 1/N_l}{1 - 1/N_l} \approx \sum_{i=1}^{n} \tilde{G}_{il}^2$$

where the first part of this equation is for a sample of size N_l at locus l. An alternative does exist which would proceed from allele probabilities.

$$\begin{aligned} PM_l &= \sum_i \tilde{p}_{il}^4 + 2 \sum_{i \neq j} \tilde{p}_{il}^2 \tilde{p}_{jl}^2 \\ &= \left(\sum_i \tilde{p}_{il}^2 \right)^2 + \sum_{i \neq j} \tilde{p}_{il}^2 \tilde{p}_{jl}^2 \\ &= \left(1 - \hat{D} \right)^2 + \sum_{i \neq j} \tilde{p}_{il}^2 \tilde{p}_{jl}^2 \end{aligned}$$

Across k loci, Jones gives

$$PM = 1 - \prod_{j=1}^{k} PM_j$$

5.7.5 Power of Discrimination

The power of discrimination is often given as

$$1 - PM = 1 - \left(1 - \prod_{j=1}^{k} PM_j \right) = \prod_{j=1}^{k} PM_j$$

5.7.6 Polymorphism Information Content

Botstein et al.[77] give

$$PIC = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_i^2 p_j^2$$

where n is the number of alleles and p_i is the allele probability of the ith allele.

5.7.7 Probability of Excluding Paternity

Ohno et al.[589] give the probability of excluding paternity:

$$Q_n = \sum_{i=1}^{n} p_i(1-p_i)^2\ (1-p_i+p_i^2) + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} p_i p_j (p_i+p_j)(1-p_i+p_j^2)$$

where n is the number of alleles and p_i is the allele probability of the ith allele.

5.7.8 Average Paternity Index

Brenner and Morris[84] give

$$\overline{PI} = \frac{1}{1-\overline{A}}$$

for fathers and nonexcluded nonfathers, where $\overline{PI}$ is the average paternity index and $\overline{A}$ is the mean exclusion probability. They further give approximations $\overline{A} \approx h^2$ or more accurately $\overline{A} \approx h^2\,(1 - 2hH^2)$ (see Nijenhuis[582]). The posterior probability of paternity, W, would be

$$W = \frac{1}{2-\overline{A}}$$

(referenced in Brenner and Morris to Morris J.W., pp. 267–276 of the same volume).[582] As usual, prior odds of 1:1 are assumed.

5.7.9 Haplotype Diversity

$$H = \frac{n}{n-1}\left(1 - \sum_{i=1}^{k} p_i^2\right)$$

where H is the "haplotype diversity."[575,697]

5.7.10 Summary

The purpose of a descriptive statistic is to give an overall impression of the usefulness of a locus for forensic, paternity, or other purposes. Most of the measures differ very little in their utility. Since the allele frequencies have a value in themselves, we assume that they will be presented in such publications. Many of the statistics may be derived directly from these

allele probabilities. The presentation of multiple summary statistics seems excessive.

The most informative statistics are the two simple measures: heterozygosity and gene diversity. The first cannot be checked against the data, the second can. It would be better for the forensic community to agree on one of these as the summary statistic of choice.

The *p*-value from the exact test for Hardy–Weinberg proportions is also valuable and should be presented without comment as to whether or not the value is significant. The *p*-values from linkage equilibrium tests are also valuable. They would not fit onto the type of table currently published, but could be deposited as supplementary data or placed in a separate table.

Appendix 5.1 (by James Curran and John Buckleton)

A "population" of 10,000 individuals was generated under independence assumptions using allele frequencies for New Zealand Caucasians ($N = 936$) from four forensic loci: HUMvWA31A, HUMTH01, HUMD8S1179, and HUMFIBRAFGA. The population was then divided into ten homogeneous subpopulations of equal size ($N_S = 1000$). To achieve a desired level of inbreeding or coancestry θ, the subpopulations were allowed to randomly breed (with no migration or mutation) for a fixed number of generations, t, dependent on the subpopulation size and θ, where

$$t = \frac{\ln(1-\theta)}{\ln(1-1/2N_s)}$$

When $\theta = 0.03$, $t \approx 61$ generations are required. To simulate the act of constructing a forensic database (many times) 10,000 random samples of size 1400 were taken (without replacement) from the population comprising the recombined subpopulations. Each member of the database was then compared to every other member of the database. If the two members had identical profiles, then the likelihood ratio (LR) was calculated for the matching profile; otherwise the LR was set to zero. The LR was calculated according to the Balding and Nichols formulae given in NRC II (Equation (3.4)):

$$LR = \begin{cases} \dfrac{(1+\theta)(1+2\theta)}{[2\theta+(1-\theta)\gamma_i][3\theta+(1-\theta)\gamma_i]} & \text{for homozygotes } (A_i A_i) \\ \\ \dfrac{(1+\theta)(1+2\theta)}{2[\theta+(1-\theta)\gamma_i][\theta+(1-\theta)\gamma_j]} & \text{for heterozygotes} (A_i A_j) \end{cases}$$

The allele frequencies were estimated according to the approximate Dirichlet multinomial (size bias correction) product moments, where

$$\gamma_i = \begin{cases} \dfrac{x_i + 4}{2N_s + 4} & \text{for homozygous profiles} \\ \dfrac{x_i + 2}{2N_s + 4} & \text{for heterozygous profiles} \end{cases}$$

and x_i is the count of the ith allele in the database. These simulations were carried out with $\theta = \{0.00, 0.03\}$ for the breeding and $\theta = 0.03$ for calculation. This represents calculation of the *LR* under the independence model using a value of θ known to be conservative (since $\theta = 0.00$ in the independence simulation and 0.03 is used in the *LR* calculation). Further we perform the calculation of *LR* under the substructure model (where $\theta = 0.03$ in the simulation and 0.03 is used in the *LR* calculation). In both cases, following Foreman et al.[312,315] the size bias correction is applied.

The exclusion power of the four loci selected for the experiment (under assumptions of Hardy–Weinberg equilibrium and linkage equilibrium) is 99.9988%; therefore, in 979, 300 ($= {}^{1400}C_2$) comparisons, we expect to find about 11–12 matches (compared with 118 for Foreman et al., the difference arising from the different loci used in our simulation) between two loci.

6 Sampling Effects

JOHN BUCKLETON AND JAMES CURRAN

Contents

6.1 Introduction

It is usual to attach a numerical weight to a match between DNA obtained from a crime sample and DNA taken from a sample given by a suspect. In Chapter 2, we discussed the possibilities of using a frequency, an exclusion probability, or a likelihood ratio for this purpose. None of these methods

returns an exact answer but, rather, each produces an estimate. The fact that an estimate is given and not an exact answer leads to the following question: "Should this numerical estimate be a 'best' estimate or should some consideration be given to the uncertainty in this estimate?" This is a matter where opinions in the forensic community differ.

Part of the uncertainty in the estimate is often referred to as sampling error. The word "error" does not refer to an analytical error but rather the variation that would occur if a different sample of individuals were taken to create the population database. Nor does the word "sampling" have anything to do with the physical act of taking a DNA sample, but rather the selection process whereby someone's DNA profile ends up in the database used for statistical calculations. This selection process induces sampling error in the resulting estimate regardless of the population genetic model used to construct it or the size of the reference database.

The argument for the assessment of sampling error is best made with an example. Take a situation involving a single suspect and a single stain from a scene. The questioned item (the stain from the scene) and the known sample (from the suspect) are sent to the laboratory. After DNA testing, on the basis of whatever typing system is in use, it is decided that there is a match between the suspect and the scene. It remains for us to assess the statistical weight of the evidence. Based on a database of genotypes and a population genetic model, the estimate for the frequency of this genotype[a] is one in a billion (10^{-9}). This is obviously compelling evidence. Now let us add, unrealistically, an additional fact: we are somewhat uncertain about this estimate. In fact, rather than being exactly one in a billion it may be anything from one in ten (10^{-1}) to one in a billion billion (10^{-18}). We believe that most people would now regard the original estimate of one in a billion somewhat differently.

Brenner[87] offers the following allegory for our consideration:

> Suppose you plan to drive to some point in the desert and must carry enough fuel for the round trip. Your best estimate is that ten gallons will be enough, but you know that this estimate carries some uncertainty, and there is, let us say, a 1% chance that you really will need 15 gallons. So 15 gallons is the "98% (or may be 99%) upper confidence estimate," and you may well judge it prudent to carry this amount of gas, rather than the "point estimate" of 10 gallons. (Reproduced with the kind permission of Dr. Brenner)

In our DNA example, the uncertainty about the statistical evidence varies from moderate (10^{-1}) to extremely strong (10^{-18}). Sampling error has been investigated by a number of authors,[33,165,217,253,267,584,585] and in reality the

[a] In fact, this is an estimate of the probability that someone who is unrelated to the suspect has this same genotype, which happens to be numerically identical to the frequency of the genotype in the population when substructure is not an issue.

variability in a DNA estimate is not as large as this (at least for a database of moderate size $\approx$ 200 individuals). One may ask whether it is sufficient to rely on these published results. While we believe that the results generalize to any DNA database, we encourage forensic scientists to think about what should be accepted practice in their own laboratories and institutions, and consider the following questions.

Should sampling error be assessed in every case?
Should it be done once for a sample of profiles from a database and published?
Should it be never done at all?

We attempt to address these questions here.

Not all commentators believe that an assessment of sampling error is necessary. Brenner[87] makes explicit his doubts of the usefulness of assessing sampling uncertainty with the following challenge:

> Will someone tell me, please, what rational difference it ever can make to know the confidence limits in addition to knowing the best point estimate? Specifically, can you give premises under which, for a fixed point estimate, the decision to convict or not to convict would depend on the size of the confidence interval? (Reproduced with the kind permission of Charles Brenner)

There is a lot of substance to Brenner's challenge. However, these comments may not have taken full account of the cross-examination process, in which any uncertainty or doubt should be, and often is, explored at length. An analyst who has prepared for such a cross examination will definitely present better evidence to the court than one who chooses to answer "would it make any difference?" Furthermore, and perhaps more importantly, it is accepted in adversarial systems that all reasonable uncertainty should be conceded to the defendant.

Commenting on statistical evidence in general, rather than DNA in particular, Good[362] stated: "The court expects us to provide both an average based on our sample and some measure of the accuracy of our average."

Almost any measurement in science has an associated measure of uncertainty. Well-prepared lawyers correctly investigate this avenue of questioning. In our experience, this is most commonly done by asking a question along the lines: "Is your database of 180 individuals big enough?"

Is there any reason why DNA evidence should be exempt from this line of questioning? The position advocating a consideration of sampling uncertainty is also taken by many authors.[33,123,158,165,217,267,585] In most cases, even with the inclusion of an estimate of sampling uncertainty, the final answer is not vastly different to the point or "best" estimate. However, we would argue that the

analyst is undoubtedly more prepared for ensuing courtroom presentation, and is also being more "scientifically honest" in doing so.[b] The admission that there is sampling error in our estimates is not a flaw but merely a statement of fact.

Please do not construe from this discussion that we are suggesting that sampling uncertainty is the only source of uncertainty. There are at least two sources: sampling uncertainty and our uncertainty about the appropriateness of the population genetic model. We feel that both warrant consideration. In fact, sampling uncertainty may be the lesser; but we cannot really tell. Some authorities advocate considering one or another of these items, "asserting" that the correction introduced is sufficient to cover both sources of uncertainty. We would see it as reasonable to consider both by using separate appropriate methods. We conclude this chapter with an extended discussion of this point; however, no uniform consensus exists in the scientific literature.

6.2 Bounds and α-Level

We assume that the sampling uncertainty will be described simply by its confidence bounds. Weir at al.[848] and Curran et al.[217] discuss one- and two-sided intervals and α-levels in the DNA context. This is a simple extension of classical theory. A two-sided, say 95%, confidence interval would allow the following type of statement to be made: 95% of intervals constructed in this way will contain the true frequency. The one-sided equivalent is: In 95% of intervals constructed in this way, the true frequency will be higher than this value.

The practical differences between a one-sided and two-sided interval are that the upper limit changes slightly and that the one-sided confidence limit has an upper bound rather than both a lower and an upper bound. The philosophical differences are larger. The two-sided bound attempts to bracket the "true value" above and below. The one-sided bound attempts to give a value above the "true value" for frequencies or below it for likelihood ratios. The argument for a two-sided interval is that it is more scientific and balanced to bound above and below. The one-sided advocates, who include us, argue that there is one fewer number to give in court. The court may be more interested in the weakest that the evidence could reasonably be, and the one-sided limit corresponds more with this.

It is not acceptable to substitute the word probability for confidence in statements regarding confidence intervals. "… A report issued by the NRC[584] that contains (p. 76) 'the traditional 95% confidence limit, whose use implies the true value has only a 5% chance of exceeding the upper bound' must lose credibility with statisticians."[829] The report in question wrongly confuses a confidence interval with a probability interval. Strictly speaking, any particular

[b] We do *not* assert that a scientist who does not choose to do this is dishonest.

confidence interval either contains the true value or it does not, but 95% of intervals should contain the true value. We cannot say that "It is 95% probable that this confidence interval contains the true value." The difference appears academic but could easily lead to difficulty in court.

The Bayesian posterior method, given by Curran et al.,[217] would allow the following statement: It is 95% probable that the true frequency is not more than 1 in 1.1 billion. This latter statement seems easier to understand but can only be made using Bayesian methods.

6.3 Methods for Assessing Sampling Uncertainty

We briefly review the suggested methods below to allow comparison of their accuracy and the relative ease of their implementation and use. For the more widely used mathematical methods, we include a discussion of their derivations.

6.3.1 Method: NRC I

Requirements: Pen and paper.[584]

Applicability: Cases where a suspect matches a simple unmixed stain from the scene.

Comment: "There is no need to discuss further the first NRC suggestion of replacing $\tilde{p}_{ij}$ by its binomial-based confidence limit $\tilde{p}_{ij} + 1.96\sqrt{\tilde{p}_{ij}(1-\tilde{p}_{ij})/2n_i}$ as that is clearly invalid. Confidence limits for products are not obtained as products of confidence limits."[848] See also Weir.[831]

Implementation: This approach has no statistical credibility and its implementation is not discussed here.

6.3.2 Method: Factor of 10

Requirements: Pen and paper.[585]

Applicability: This method applies to cases where a suspect matches a simple unmixed stain from the scene. This method was not developed for mixtures, relatives, paternity or missing person cases, and its performance in these instances is unknown.

Comment: "Similarly, the second NRC suggestion of constructing the interval $(\hat{P}/10, 10\hat{P})$ has limited theoretical validity. There must at least be an effect of sample size."[848]

"The 'Factor of 10' approach has little to recommend it from a theoretical standpoint, and we must prefer the other methods which have a more firm statistical backing. As mentioned previously there must be an effect of sample size and number of loci. . . . However, our own simulation may be viewed as further empirical support for the 'Factor of 10' approach. In general the

'Factor of 10' approach performed in a broadly similar way to the other methods or was excessively conservative."[217]

Implementation: Suppose that the match probability is estimated as, say, 1 in a billion. This approach would suggest that the bounds for uncertainty (sampling and population model) are 1 in 100 million to 1 in 10 billion.

6.3.3 Method: Asymptotic Normality of the Logarithm Extended by NRC II to Allow for Population Structure

Requirements: May be easily applied by an EXCEL™ spreadsheet. Laszlo Szabo of the Tasmanian Forensic Science Laboratory has developed one that is in extensive use throughout Australia.[165,585]

Applicability: This method may be applied to cases where a suspect matches a simple unmixed stain from the scene. "We cannot as yet see how to extend this method generally for other formulations such as mixtures or paternity cases, although the methodology should be applicable."[848]

Comment: Curran et al.[217] scored the performance of this method as adequate in most situations except for small database sizes, reporting that "for single-contributor stains, such as those considered here, it does appear that these normal methods are completely suitable." However, this method does not easily extend to mixed stains, nor does it generalize to other situations. This shortcoming limits its overall usefulness.

Implementation: An excellent explanation is given in NRC II.[585 p. 122] Here we present a spreadsheet layout to apply this approach with the subpopulation correction, θ (Figure 6.1). If the product rule is desired, the more simple formulae in NRC II may be used or this spreadsheet may be utilized with θ set to zero.

Theoretical basis (following Weir[848]): Methods based on asymptotic normality have the advantage of leading to relatively simple analytical expressions. Let P be the match probability and $\hat{P}$ the estimate. The procedure assumes that $\ln(\hat{P})$ is normally distributed, so that a 99% confidence interval for P is $\hat{P}/C, C\hat{P}$, where $\ln(C) = 2.57\sqrt{\mathrm{Var}[\ln(\hat{P})]}$. The task is to compute the variance of $\ln(\hat{P})$. Assuming independence between loci, this is approximated by

$$\mathrm{Var}\left[\ln(\hat{P})\right] = \mathrm{Var}\left[\sum_i \ln(\hat{P}_i)\right] \approx \sum_i \mathrm{Var}(\hat{P}_i)/\hat{P}_i^2$$

As θ is generally assigned a numerical value (NRC II,[585 p. 122]) such as 0.03 in this context, rather than being estimated from sample data, it will be assumed to be constant.

For a homozygous profile at locus i

$$\mathrm{Var}(\hat{P}_i) \approx \left(\frac{\partial \hat{P}_i}{\partial \tilde{p}_{i1}}\right)^2 \mathrm{Var}(\tilde{p}_{i1})$$

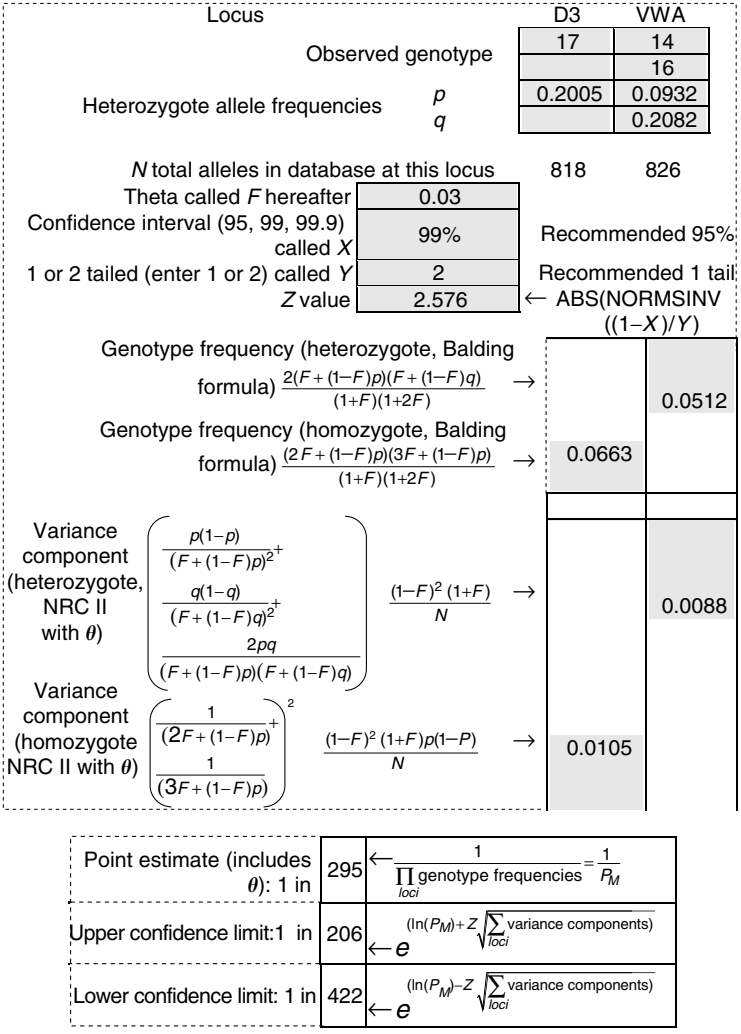


Figure 6.1 Spreadsheet layout to implement the asymptotic normality of the logarithm method.

and for a heterozygous profile

$$\mathrm{Var}(\hat{P}_i) \approx \left(\frac{\partial \hat{P}_i}{\partial \tilde{p}_{i1}}\right)^2 \mathrm{Var}(\tilde{p}_{i1}) + \left(\frac{\partial \hat{P}_i}{\partial \tilde{p}_{i2}}\right)^2 \mathrm{Var}(\tilde{p}_{i2}) + 2\left(\frac{\partial \hat{P}_i}{\partial \tilde{p}_{i1}}\right)\left(\frac{\partial \hat{P}_i}{\partial \tilde{p}_{i2}}\right) \mathrm{Cov}(\tilde{p}_{i1}, \tilde{p}_{i2})$$

When $\theta = 0$, these expressions become

$$\mathrm{Var}[\ln(\hat{P}_i)] = \begin{cases} \dfrac{2(1-p_{i1})}{n_i p_{i1}}, & A_{i1} = A_{i2} \\ \dfrac{p_{i1} + p_{i2} - 4\,p_{i1}\,p_{i2}}{2n_i p_{i1} p_{i2}}, & A_{i1} \neq A_{i2} \end{cases}$$

as given previously by Chakraborty et al.[165] and NRC II,[585 p. 146]. Use was made of the binomial variances and covariances of allele proportions:

$$\mathrm{Var}(\tilde{p}_{ij}) = \frac{p_{ij}(1-p_{ij})}{2n_i}$$

$$\mathrm{Cov}(\tilde{p}_{ij}, \tilde{p}_{ij'}) = -\frac{p_{ij}\,p_{ij'}}{2n_i}$$

in the derivation of $\mathrm{Var}[\ln(\hat{P}_i)]$.

6.3.4 Method: Bootstrap

Requirements: This method requires a purpose written programme.[848] A very adaptable one that can handle any forensic situation has been developed by Curran.[213]

Applicability: Simple stains, mixtures, paternity cases, missing persons, all forensic casework.

Comment: Curran et al.[217] scored the performance of this method as adequate in most situations except for small database sizes. As they explained, "… the relatively poor performance of normal-based limits or bootstrap limits for small sample sizes [small databases] is a consequence of specific alleles not appearing in these samples. The problem disappears when θ is assigned a non-zero value."

Implementation: Consider a database of individuals indexed $1 \ldots N$. We wish to assess, say, the genotype probability across 13 loci. The steps are as follows: Assess the multilocus genotype using whichever formula is preferred. This could be the product rule, a theta correction, the brother's formula, a paternity calculation, or whatever. Select an individual at random between 1 and N and put a copy of this genotype into the "new database," but do not remove this individual from the original database (i.e., we sample genotypes — or individuals — from the database with replacement). Repeat these processes N times. We now have a new database of N individuals. Some of the initial individuals may be represented twice or thrice, some once, or some not at all. Recalculate the allele frequencies using our new

database. Recalculate the formula of interest using these allele frequencies. Repeat this approximately 1000 times. Sort the 1000 results of the formula into ascending order. The 25th and 976th results represent the bounds of the 95% two-sided confidence interval. The 951st represents the 95% one-sided confidence interval.

6.3.5 Balding's Size Bias Correction Corrected in Evett and Weir

Requirements: Pen and paper.[33,267]

Applicability: This method was developed for simple unmixed stains: "there appears to be no simple general extension to mixed stains and paternities."[217]

Comment: Curran et al.[217] score the performance of this method as poor if the intent is to assess sampling error. "An unfortunate consequence of Balding's discussion of 'conservative Bayesian estimates' is that some forensic agencies have taken to presenting only point estimates based on sample allele proportions calculated by adding crime and suspect profiles to a database. These modified point estimates do not address sampling error. As sample size increases they are more likely to provide intervals (bounded above by these values) that do not contain the true values. It would be misleading to regard them as acting like confidence limits … we are disappointed in the lack of scientific rigor both in its derivation and application. However for small databases or rare profiles it is probably acceptable as a coarse correction for sampling error. For larger databases and common profiles it performs more as a mean estimate."[848] The performance of this method may be adequate when applied to very small databases and rare profiles. It is difficult to predict this method's performance in any given case.

The theoretical support for this method is given under the headings "Support Intervals" and "Uniform Allele Prior Distribution." Also present is a discussion of the possible use of nonuniform priors.

Implementation: Suppose that we have a database of size $2N$ alleles. To calculate the probability of observing an *aa* homozygote, take the count, x, of the *a* allele in the database and use

$$\hat{P}_{aa} = \left(\frac{x+4}{2N+4}\right)^2$$

To calculate the probability of observing an *ab* heterozygous profile, count the number of *a* and *b* alleles in the database, x_a and x_b, and use

$$\hat{P}_{ab} = 2\left(\frac{x_a+2}{2N+4}\right)\left(\frac{x_b+2}{2N+4}\right)$$

6.3.5.1 *Theoretical Support (following Weir et al.[848])*

6.3.5.1.1 Support intervals. Suppose that the population probability of allele j at locus i is p_{ij}, and a sample from the population contains x_{ij} copies of that allele. For locus i, the likelihood function for the alleles in the profile is

$$L(\{p_{ij}\}) \propto \begin{cases} p_{i1}^{x_{i1}}(1-p_{i1})2^{n_i-x_{i1}}, & A_{i1}=A_{i2} \\ p_{i1}^{x_{i1}}\,p_{i2}^{x_{i2}}(1-p_{i1}-p_{i2})2^{n_i-x_{i1}-x_{i2}}, & A_{i1}\neq A_{i2} \end{cases}$$

where the sample has n_i individuals scored at locus i. The likelihood has a maximum value of

$$L(\{\tilde{p}_{ij}\}) \propto \begin{cases} \left(\dfrac{x_{i1}}{2n_i}\right)^{x_{i1}}\left(\dfrac{2n_i-x_{i1}}{2n_i}\right)^{2n_i-x_{i1}}, & A_{i1}=A_{i2} \\ \left(\dfrac{x_{i1}}{2n_i}\right)^{x_{i1}}\left(\dfrac{x_{i2}}{2n_i}\right)^{x_{i2}}\left(\dfrac{2n_i-x_{i1}-x_{i2}}{2n_i}\right)^{2n_i-x_{i1}-x_{i2}}, & A_{i1}\neq A_{i2} \end{cases}$$

Balding[33] considered support intervals obtained by constraining the multinomial proportions to give a fixed profile probability P_0, introducing a Lagrange multiplier λ to maximize the expression

$$\sum_i [\ln L(\{p_{ij}\})] + \lambda\left[\sum_i \ln(P_i) - \ln P_0\right]$$

The constrained solutions are

$$P_{0i} = \begin{cases} \dfrac{(x_{i1}+2\lambda)^2}{(2n_i+2\lambda)^2}, & A_{i1}=A_{i2} \\ \dfrac{2(x_{i1}+\lambda)(x_{i2}+\lambda)}{(2n_i+2\lambda)^2} & A_{i1}\neq A_{i2} \end{cases} \tag{6.1}$$

which correspond to the addition of λ copies of the profile to the sample. The ratio of the constrained to the unconstrained maximum likelihoods is

$$R_i(\lambda) = \begin{cases} \dfrac{(1+2\lambda/x_{i1})^{x_{i1}}}{(1+2\lambda/2n_i)^{2n_i}}, & A_{i1}=A_{i2} \\ \dfrac{(1+\lambda/x_{i1})^{x_{i1}}(1+\lambda/x_{i2})^{x_{i2}}}{(1+2\lambda/2n_i)^{2n_i}} & A_{i1}\neq A_{i2} \end{cases}$$

which differs from equation 15 of Balding[33] in detail. Over the whole profile, the likelihood ratio is

$$R(\lambda) = \prod_i R_i(\lambda)$$

A 99% profile likelihood interval is found by choosing those two values of λ such that $\ln R(\lambda) = -(1/2)\chi^2_{1;0.99} = 3.317$.

The example does not support Balding's claim that the λ value for the upper support limit should be two, corresponding to the addition of the crime stain profile and the suspect's profile to the sample. Without this appealing way to choose λ, there seems little numerical advantage to the use of profile likelihood support intervals over conventional confidence intervals. The bounds of λ will depend on the number of loci in the profile, the population probabilities of the profile alleles, and the size of the sample, as well as on the probability ordinate used to construct the interval. Adding the crime stain profile and the suspect's profile to the sample before constructing allele sample frequencies neither accounts for the sampling variation induced by finite samples, nor corrects for the presence of population substructure.

6.3.5.1.2 Uniform allele prior distribution. Balding[33] also considered a Bayesian approach by assuming a Dirichlet prior distribution for allele probabilities. The probability density for values p^*_{ij} in a particular population was taken to be

$$\pi\left(p^*_{i1}\right) \propto \left(p^*_{i1}\right)^{\gamma_{i1}-1}\left(1-p^*_{i1}\right)^{\gamma_{\bar{1}}-1}, \qquad A_{i1}=A_{i2}$$

$$\pi\left(p^*_{i1,}\, p^*_{i2}\right) \propto \left(p^*_{i1}\right)^{\gamma_{i1}-1}\left(p^*_{i2}\right)^{\gamma_{i2}-1}\left(1-p^*_{i1}-p^*_{i2}\right)^{\gamma_{\bar{1}\bar{2}}-1}, \quad A_{i1}\neq A_{i2}$$

The quantities $\gamma_{i\bar{1}}$ and $\gamma_{i\bar{1}\bar{2}}$ are the Dirichlet parameters for the allelic classes "not 1" and "not 1 or 2" at locus *i*. Combined with multinomial sampling of alleles from the population, the posterior profile probability (i.e., the expectation over populations of $\left(p^*_{i1}\right)^2$ for $A_{i1}\,A_{i1}$ profiles and of $2p^*_{i1}\,p^*_{i2}$ for $A_{i1}\,A_{i2}$ profiles) is

$$\mathrm{E}(P_{0i}) = \begin{cases} \dfrac{\Gamma(\gamma_{i1}+\gamma_{i\bar{1}}+2n_i)\,\Gamma(\gamma_{i1}+x_{i1}+2)}{\Gamma(\gamma_{i1}+\gamma_{i\bar{1}}+2n_{i1}+2)\;\Gamma(\gamma_{i1}+x_{i1})} & A_{i1}=A_{i2} \\ \dfrac{2\Gamma(\gamma_{i1}+\gamma_{i2}+\gamma_{i\bar{1}\bar{2}}+2n_i)\,\Gamma(\gamma_{i1}+x_{i1}+1)\,\Gamma(\gamma_{i2}+x_{i2}+1)}{\Gamma(\gamma_{i1}+\gamma_{i2}+\gamma_{i\bar{1}\bar{2}}+2n_i+2)\,\Gamma(\gamma_{i1}+x_{i1})\,\Gamma(\gamma_{i2}+x_{i2})}, & A_{i1}\neq A_{i2} \end{cases} \tag{6.2}$$

Balding then simplified this to allow for uniform priors: $\gamma_{i\bar{1}} = \gamma_{i\bar{1}} = 1$ for homozygotes and $\gamma_{i1} = \gamma_{i2} = \gamma_{i\bar{1}\bar{2}} = 1$ for heterozygotes. Then the expected

profile posterior probabilities are

$$\mathrm{E}(P_{0i}) = \begin{cases} \dfrac{(x_{i1}+2)(x_{i1}+1)}{(2n_i+3)(2n_i+2)}, & A_{i1}=A_{i2} \\ \dfrac{2(x_{i1}+1)(x_{i2}+1)}{(2n_i+4)(2n_i+3)}, & A_{i1}\neq A_{i2} \end{cases} \tag{6.3}$$

which again differ slightly from the expressions given by Balding. An advantage of the Dirichlet distribution is that all moments have simple expressions, so that with a uniform prior the expectations of $(p_i^*)^4$ and $(2p_{i1}^* p_{i2}^*)^2$ give

$$\mathrm{E}(P_{0i}^2) = \begin{cases} \dfrac{(x_{i1}+4)(x_{i1}+3)(x_{i1}+2)(x_{i1}+1)}{(2n_i+5)(2n_i+4)(2n_i+3)(2n_i+2)}, & A_{i1}=A_{i2} \\ \dfrac{4(x_{i1}+2)(x_{i2}+1)(x_{i1}+2)(x_{i2}+1)}{(2n_i+6)(2n_i+5)(2n_i+4)(2n_i+3)}, & A_{i1}\neq A_{i2} \end{cases}$$

(which also differ slightly from the expressions given by Balding). The ratio of these two expectations can be regarded as the probability of the profile occurring twice given that it has occurred once (i.e., the match probability):

$$\frac{\mathrm{E}(P_{0i})}{\mathrm{E}(P_{0i}^2)} = \begin{cases} \dfrac{(x_{i1}+4)(x_{i1}+3)}{(2n_i+5)(2n_i+4)}, & A_{i1}=A_{i2} \\ \dfrac{2(x_{i1}+2)(x_{i2}+2)}{(2n_i+6)(2n_i+5)}, & A_{i1}\neq A_{i2} \end{cases} \tag{6.4}$$

This is almost the expression that would result for the simple product rule if the profile in question was added twice to the database.

Balding suggested this method as a way to incorporate sampling effects into estimated match probabilities, in the sense that the sample database was allowed to modify the assumed uniform prior for allele probabilities. Weir et al.[848] believe, however, that these estimates are posterior means for match probabilities when the prior means for the probabilities are one third for homozygotes and one sixth for heterozygotes, as can be seen by setting $x_{i1} = x_{i2} = n_i = 0$ in Equations (6.3).

6.3.5.1.3 Nonuniform allele prior distribution. As Balding remarks, it is illogical to assign uniform priors for allele probabilities that differ with

the number of alleles (1 or 2) that occur in the evidentiary profile. There is population-genetic logic in the nonuniform Dirichlet that invokes the parameter θ via

$$\gamma_{ij} = \frac{(1-\theta)p_{ij}}{\theta}$$

as discussed by Balding and Nichols[41] and Evett and Weir.[267] The posterior match probabilities are then

$$\frac{\mathrm{E}(P_{0i})}{\mathrm{E}(P_{0i}^2)} = \begin{cases} \dfrac{[(3 + x_{i1})\,\theta + (1 - \theta)p_{i1}][(2 + x_{i1})\theta + (1 - \theta)p_{i1}]}{[1 + (1 + 2n_i)\theta][1 + (2 + 2n_i)\theta]}, & A_{i1} = A_{i2} \\ \dfrac{2[(1 + x_{i1})\theta + (1 - \theta)p_{i1}][(1 + x_{i2})\theta + (1 - \theta)p_{i2}]}{[1 + (1 + 2n_i)\theta][1 + (2 + 2n_i)\theta]}, & A_{i1} \neq A_{i2} \end{cases} \tag{6.5}$$

There is a problem in knowing what values to use for the unknown allele probabilities p_{ij}. Simply using sample proportions from current populations appears to ignore the variation that the Dirichlet distribution is designed to incorporate, although the problem is lessened when the x_{ij} values are large. Balding does not comment on the fact that the sample of size n_i individuals in Equations (6.5) is from the specific subpopulation relevant to the crime. It is not a sample that would furnish an estimate of the population-wide frequencies p_{ij}, further explaining why there is no simple interpretation of these results in terms of adding copies of the matching profile to the database.

Note that as the sample size n_i increases, Equations (6.5) reduce to

$$\frac{\mathrm{E}(P_{0i})}{\mathrm{E}(P_{0i}^2)} = \begin{cases} p_{i1}^2, & A_{i1} = A_{i2} \\ 2p_{i1}p_{i2}, & A_{i1} \neq A_{i2} \end{cases} \tag{6.6}$$

which are just the product rule expressions for the relevant subpopulation. The product rule expressions are also obtained when $\theta = 0$ because there is then no distinction between subpopulations and the whole population. When there are no data from the relevant subpopulation, $x_{i1} = x_{i2} = n_i = 0$, and Equations (6.6) are recovered.

If only the two samples from the crime stain and the suspect are available from the relevant subpopulation, $n_i = 2$ and $x_{i1} = 4$ for homozygous profiles

$A_{i1}A_{i1}$ or $x_{i1} = x_{i2} = 2$ for heterozygous profiles $A_{i1}A_{i2}$:

$$\frac{\mathrm{E}(P_{0i})}{\mathrm{E}(P_{0i}^2)} = \begin{cases} \dfrac{[7\theta + (1-\theta)p_{i1}][6\theta + (1-\theta)p_{i1}]}{[1+5\theta][1+6\theta]}, & A_{i1} = A_{i2} \\ \dfrac{2[3\theta + (1-\theta)p_{i1}][3\theta + (1-\theta)p_{i2}]}{[1+5\theta][1+6\theta]}, & A_{i1} \neq A_{i2} \end{cases} \tag{6.7}$$

6.3.6 Method: Posterior Density[c]

Requirements: This method requires a purpose written program. A very adaptable one that can handle any forensic situation has been developed by Curran.[213]

Applicability: All forensic casework.

Comment: Curran et al.[217] scored the performance of this method as adequate in most situations.

Implementation: This is the most mathematically intimidating of the various approaches, but in concept it is the most familiar and most intuitive. It helps to start by thinking about the problem without the hindrance of the mathematics.

One way "into" the problem is to think about a situation where we have no alleles of type *a* in our database, but have just done a case where the suspect and crime stain have this allele. Our allele probability estimate from our database is zero (please ignore minimum allele probabilities at this point). But we have just seen one copy of allele *a* (in the suspect). So we certainly no longer believe that the frequency is zero.

Next we ask ourselves why we are calculating a frequency at all. It is to assess the chance of this evidence if the suspect did not leave the stain. Hence the whole calculation of a frequency is based on the assumption that the suspect did not leave the stain. Now if the suspect did not leave the stain, someone else did. Hence we have two, not one, observations of allele *a*. Thinking of this type led Scranage and Pinchin[701] to add the "suspect and offender" to the database when they wrote the groundbreaking program DNASYS.

This is what the Bayesian approach does. It starts from a position, observes the database and the suspect and the offender. This results in an estimate and the variability in that estimate. We feel that the court would also respond well to an explanation that we had "updated" our view of allele probabilities based on the suspect and offender profiles. A program is required to implement this approach but one is available.

[c] We attribute this method to the suggestion of Dr. Ian Painter. It is an extension of the method of Professor David Balding.

6.3.6.1 *Explanation of the Bayesian Highest Posterior Density*

This approach is not explained in simple terms elsewhere in the literature. We attempt this here.

6.3.6.1.1 Bayes theorem and Bayesian estimation. In forensic applications the odds form of Bayes's Theorem is used to show how the likelihood ratio can be combined with the prior odds on guilt to give us the posterior odds on guilt. In Bayesian estimation, we are interested in the value of an unknown population parameter such as an allele probability. To estimate this parameter, we combine our prior probability about the possible values for this parameter with the data that have been observed to get the posterior probability on the possible values the parameter may take.

Bayes's theorem tells us how to do this. We write

$$\Pr(\lambda|\text{data}) = \frac{\Pr(\text{data}|\lambda)\Pr(\lambda)}{\int \Pr(\text{data}|\lambda)\Pr(\lambda)\,d\lambda}$$

or

$$\Pr(\lambda|\text{data}) \propto \Pr(\text{data}|\lambda)\Pr(\lambda) \tag{6.8}$$

where λ represents the parameter(s) of interest. In forensic casework, λ is likely to be an allele probability. In words, Equation (6.8) states: "the probability of the parameter given the data is proportional to the probability of the data given the parameter times the probability of the parameter." The first equation shows the "scaling" factor that we need to calculate the probability.

We start with some belief about a parameter. Possibly we have no knowledge at all. This can be modeled by various functions. For instance, "no knowledge at all" is often modeled by a function that assigns all values between 0 and 1 the same probability. An experiment is performed to collect some information about the parameter. In our case, this is the database and the suspect and offender profiles. Then the prior belief and the data are combined to give an updated idea about the parameter. The equation can be broken down into the posterior, $\Pr(\lambda|\text{data})$, the likelihood, $\Pr(\text{data}|\lambda)$, and the prior, $\Pr(\lambda)$. The likelihood is usually straightforward to compute and is suggested by the problem. Choice of the prior can be very problematic.

6.3.6.1.2 Prior probabilities. Assume that we wish to assess the frequency (probability) of various alleles at a locus. Furthermore, let us assume that this particular locus has only alleles A and B. Since people can only have A or B alleles, then $\Pr(A) + \Pr(B) = 1$ or $\Pr(B) = 1 - \Pr(A)$. Therefore, it suffices to estimate the probability of allele A, denoted by π_A.

A sample of people is taken and typed. Our maximum likelihood estimate for the probability of the A allele is

$$f_A = \hat{\pi}_A = \frac{\#\text{ of } A\text{'s}}{2N}$$

where N is the number of individuals in our database. The hat "^" is used to indicate that this is an estimate. Imagine that in a sample of, say, ten people, there were seven type A alleles; then our estimate is $7/20 = 0.35$. However, before this sample was taken what did we know about π_A? Regardless of what we assume, we need a way of representing our knowledge. We do this by saying how probable we think certain values of π_A are. For example, we might say that there is a 10% chance that π_A is less than 0.2 ($\Pr(\pi_A < 0.2) = 0.1$) and a 10% chance that π_A is greater than 0.9 ($\Pr(\pi_A > 0.9) = 0.1$), and an 80% chance that π_A is between 0.2 and 0.9 ($\Pr(0.2 < \pi_A < 0.9) = 0.8$). Together these probabilities add up to one, and what we have described is called a cumulative density function (CDF). They describe the area under a curve called a probability density function. The key fact is that the area, and not the height of the curve, measures probability.

The proportion may have been estimated as 0.53 using a database in Scotland so it might be similar in Ireland. Or, we may choose to say that we know nothing — all values of π_A are equally likely. We can choose prior densities that have these probabilities. Typically these are chosen (to simplify the mathematics) from a family of probability density functions with well-known properties. In the case of a single proportion, this family of curves is called the Beta family. The shape of distributions in the Beta family is defined by two parameters, a and b. Any choice of a and b that differs from 1 gives substantial shape to the curve. This of course will affect the posterior distribution, so some people would say that if a and b are not 1, then we have chosen an informative prior. If we set a and b to 1, we have an uninformative prior and are assuming that all values of π_A are equally likely.

A convenient property of the Beta distribution is that if our prior is Beta(1, 1), and we observe x A alleles in a sample of $2N$, then the posterior distribution is Beta($x + 1$, $2N - x + 1$).

6.3.6.1.3 Posterior probabilities. Using the posterior distribution, we can answer questions such as: "If we specify a probability p, what are the points l and u such that $\Pr(l < \pi_A < u) = p$?" For example, if $p = 0.95$, then what are l and u? These may turn out to be 0.18 and 0.57. Therefore, we can say that the probability that π_A is between 0.18 and 0.57 is 0.95, or that we are 95% sure that π_A lies between 0.18 and 0.57. This is very much like a

confidence interval, but we would not be able to make such a statement with a confidence interval. You would have to say you are 95% confident that π_A lies between 0.18 and 0.57. The word confident here translates into "on average, 95% of intervals generated in the same way, would contain the true value π_A."

6.3.6.1.4 Highest posterior density intervals. The interval described above is the highest posterior density interval or region. It is an interval for which the posterior density is the highest. This means that the values in the interval are the most likely in the whole range of possible values. What remains is to explain how we use this approach to assess sampling error in a likelihood ratio calculation.

First, we need to extend the theory from a two-allele locus to a k-allele locus. The Beta distribution has a natural multivariate analog called the Dirichlet distribution. The Dirichlet distribution allows us to model a whole set of proportions, which add up to one instead of just a single proportion. More specifically, if we need to model the probabilities for k alleles, which we denote π_{A_i}, for $i = 1, \ldots, k$, then we can do this by using a Dirichlet distribution such that $(\pi_{A_1}, \pi_{A_2}, \cdots, \pi_{A_{k-1}}) \sim Dirichlet\ (\alpha_1, \alpha_2, \ldots, \alpha_k)$. Note that the subscripts on the π_{A_i}'s only go up to $k-1$. This is because

$$\pi_{A_k} = 1 - \sum_{i=1}^{k-1} \pi_{A_i}$$

You should also note that if k is 2, then this is actually just a Beta distribution, that is, $\pi_A \sim Dirichlet\,(\alpha_1,\ \alpha_2) = Beta\,(\alpha_1,\ \alpha_2)$. The Dirichlet distribution works in the same way as the Beta distribution, in that an uninformative Dirichlet prior (where all the α's are 1), and a count of x_i alleles of type A_i in a sample of $2N$ gives the posterior density

$$Dirichlet\left(x_1 + 1, x_2 + 1, \ \ldots,\ 2N - \sum_{i=1}^{k-1} x_i + 1\right)$$

Consider that we may have a likelihood ratio that contains a large number, say 20 or 26, of unknown allele probabilities. We do, however, know something about the allele probabilities from the database, and we can combine these with our prior beliefs about the probabilities to get the posterior densities of the allele probabilities. We can then generate a random sample from each of these densities and insert the values into our likelihood ratio calculation. If we do this many times, we will begin to build up a distribution of likelihood ratios. This is very much like the bootstrap, but with the added advantages that (a) it is very fast and (b) it lends itself to a much more natural way of explaining the resulting interval.

6.4 Minimum Allele Probabilities

The concept of a minimum allele probability replaces zero or very small allele probabilities derived by counting from some database with some minimum probability. This avoids the genotype probability estimate being zero and stems largely from the concern that these small allele probabilities are very poorly estimated. Minimum allele probabilities are unnecessary when either the Bayesian support interval (Balding's size bias correction) or the highest posterior density interval are used as these methods can "handle" a zero estimate. Consider Balding's size bias estimator for a homozygote,

$$\hat{P}_{aa} = \left(\frac{x+4}{2N+4}\right)^2$$

This returns a nonzero value for $\hat{P}_{aa}$ even when the count in the database, x, is zero.

When a nonzero value is assigned to θ, the genotype estimate will be nonzero even when the count of the allele or alleles is zero in the database. However the bootstrap, the factor of 10, and the assumption of asymptotic normality of the logarithm will not correctly estimate sampling variation in these circumstances. The typical solution has been to apply a minimum allele probability.

Budowle et al.[123] discuss two options for the $1-\alpha$ upper confidence interval: (i) Following Chakraborty,[158]

$$p_{min} = 1 - [1 - (1 - \alpha)^{1/c}]^{1/2n}$$

where p_{min} is the minimum allele probability, c is the number of common alleles, and n is the number of individuals in the database. (ii) Following Weir,[828] $p_{min} = 1 - \alpha^{1/2n}$. Chakraborty's approach typically gives a higher minimum allele probability and behaves in an unusual manner. We wonder if it has any merit.

6.5 Discussion of Appropriateness of Sampling Uncertainty Estimates — Buckleton

To conclude, I review some published opinions as to the appropriateness of various sampling uncertainty corrections. No uniform consensus exists and there are some quite polarized views. My own opinion is directly ascribed.

It is worthwhile to begin this discussion by considering the potential sources of uncertainty in determining a match probability. The larger ones relate to errors in laboratory work or in assigning genotypes. We begin the

mathematical consideration by assuming that these functions have been correctly carried out. However, it must always be remembered that everything from here onwards is conditional on the profiles being correctly assigned.

Probably the next largest source of uncertainty would be the existence of a monozygotic twin or other close relatives in the population of potential suspects. This is becoming more important as more loci are added. The addition of further loci focuses attention away from unrelated persons and onto close relatives and members of the same subpopulation. This was discussed in Chapter 4.

Next we come to uncertainties in the appropriateness of the population genetic model, sampling uncertainty, and minimum allele probabilities. These last two are manifestations of the same thing. We are aware of opinion that supports the use of:

minimum allele probabilities and the product rule;
minimum allele probabilities, the product rule, and sampling uncertainty assessment;
minimum allele probabilities, a conservative θ correction, and Balding's size bias correction;
a conservative θ correction and sampling uncertainty assessment.

A key question is: Are we seeking the best estimate or a conservative one with known properties? The best estimate (that is defined here as the one with least total bias either way) may be the product rule and Balding's size bias correction. I have formed this opinion from simulation studies. This is the only way that has been developed since we do not know the true answer. This approach would be highly appropriate in civil cases where the least biased answer is required.

A problem in criminal cases when giving only the "best estimate" is that it immediately leads to legitimate debate in court about the uncertainty inherent in that estimate. If the analyst is unprepared for this debate, then he/she may be in for a rough time and may appear unscientific and underprepared. In order to be prepared, he/she must assess both types of uncertainty: that arising from the population genetic model and that arising from sampling uncertainty (the latter includes minimum allele probabilities).

The next fact that needs consideration is that robust methodology exists that can handle both sampling uncertainty and minimum allele probabilities. These methods have been summarized in this chapter. The behavior of these methods is known, and there is a large body of support for their use.

The two intermediate approaches of which we are aware are laboratories using a conservative value for θ and a mean estimator for allele probabilities. This appears to be a hybrid of mixed philosophies. Why be excessively

conservative in one regard but aim for the mean in another? The stated reason is that the conservative θ is "enough." But how could one know unless one could measure the conservativeness induced by both the large θ value and the sampling uncertainty? We can measure the sampling uncertainty, but we cannot as yet be completely confident of the conservativeness induced by a large θ. The laboratories undertaking this approach do not even measure the sampling uncertainty; hence, in our opinion they assert, perhaps correctly, that the conservative θ is sufficient.

The reciprocal approach is to use the product rule and minimum allele probabilities either with or without a sampling uncertainty assessment but no θ correction. These laboratories assert the reciprocal, that their correction is "enough."

The most scientifically sound approach, in my opinion, is to measure both forms of uncertainty using the best available tools and report the sum of the uncertainty. This approach can be easily implemented by biologists who are not mathematical specialists. This has been evidenced numerous times by skilled caseworkers. These biologically trained caseworkers have often commented to me on the professional pride that they take in attempting to give the best scientific evidence possible at this time. Even then it is important to make apparent that the resulting figure is the outcome of an estimation process that cannot be fully calibrated against a "true answer."

Mixtures

TIM CLAYTON AND JOHN BUCKLETON

7

Contents

7.1 Introduction

The analysis of forensic stains will inevitably lead to mixtures of DNA from different individuals resulting from the mixing of body fluids and secretions. The recognition, resolution, and statistical evaluation of such mixtures are therefore integral and vital parts of forensic casework. The subject is relatively complex and requires experience and judgement. It is often treated as a separate competency by forensic organizations such as ESR and the FSS. Often scientists move to mixture interpretation after experience with simple stains. It is desirable that a formal training and testing program is associated with this transition.

The typing of mixed samples may be undertaken using autosomal DNA or with Y chromosome or mitochondrial analysis. Each has advantages.[156] A number of methods have been developed to facilitate the evaluation of evidence from mixed profiles. These differ according to whether or not they employ a Bayesian or a frequentist approach, and whether or not they utilize quantitative aspects of the data (peak heights or areas) as well as qualitative aspects (which alleles are present) (see Figure 7.1).

In order to facilitate a discussion of these methods, it is necessary to digress slightly and introduce various notations, nomenclature, and naming conventions peculiar to the analysis of mixed stains. First, it is important to realize that a mixture can contain DNA from any number of contributors, *N*. Experience has indicated that most mixtures encountered in forensic casework appear to consist of DNA from just two individuals. However,

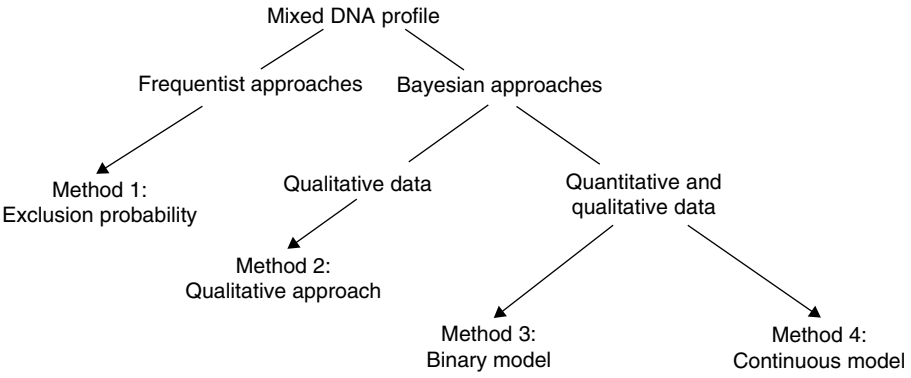


Figure 7.1 Methods used in the interpretation of mixtures.

mixtures with $N>2$ are encountered. These are referred to as "higher-order" mixtures.

Second, one of the most important terms in mixture analysis (at least where the quantitative aspects of the data are being considered) is the mixing proportion (M_x, see Table 7.5). For a two-person mixture, this can take any value between 0 and 1. Practitioners often prefer to use the mixture ratio as this is intuitively easier to estimate from a visual inspection of the profile. In this text we will use mixture proportions as the mathematics flows more easily.

Where the mixing proportion is such that in the judgement of the scientist the unambiguous profile of one of the contributors is discernibly larger than the others, then practitioners generally refer to this as the major component of the profile. The remaining component(s) are referred to as minor component(s). In artificially created two-person mixtures, a good estimate of the mixing proportion is known in advance. In forensic stains however, it is not, and the scientist must attempt to deduce this from the extant data. Necessarily, this will be conditional on which particular genotypes are being considered.

Third, when peak area data are being considered, the symbol ϕ is used to denote this area. Findlay and Grix[299] give a warning of the potential in court to use an unassigned minor peak to foster doubt in the mind of a jury by postulating that the unknown minor represents the true assailant. This seems a timely warning to us and suggests that full mixture analysis may be warranted more often than we had previously considered.

Mixture evaluation can proceed via a process of calculating a likelihood ratio or by calculating a probability of exclusion. Most of the rest of this chapter will be devoted to calculating likelihood ratios that are accepted as being more powerful. However, we will briefly introduce the frequentist method of calculating the probability of exclusion. We largely follow Budowle.[119]

7.2 The Frequentist Approach

7.2.1 Method 1 — Exclusion Probabilities

In the mixture context, the exclusion probability is defined as "the probability that a random person would be excluded as a contributor to the observed DNA mixture." When considered as a frequency, it may be used to answer the question: "How often would a random person be excluded?" This is the reason that it is often referred to as the Random Man Not Excluded (RMNE) approach.

If the mixture has alleles $A_1 \ldots A_n$, then the exclusion probability at locus l (PE_l) is

$$PE_l = 1 - \left(\sum_{i=1}^{n} p(A_i) \right)^2$$

if Hardy–Weinberg equilibrium is assumed. By writing

$$\sum_{i=1}^{n} p(A_i) = p$$

we can obtain $PE_l = 1 - p^2$.

If Hardy–Weinberg equilibrium is not assumed, Budowle gives

$$PE_l = 1 - \left(\sum_{i=1}^{n} p(A_i)\right)^2 - \theta \sum_{i=1}^{n} p(A_i)\left(1 - \sum_{1=l}^{n} p(A_i)\right)$$

We can write this as $PE_l = 1 - p^2 - \theta p(1-p)$. This expression follows from the use of

$$p(A_i A_i) = p_i^2 + \theta p_i (1 - p_i)$$
$$p(A_i A_j) = 2(1 - \theta) p_i p_j$$

The proof appears in Box 7.1 and was due to Professor Bruce Weir.

The use of the equivalent of NRC II Recommendation 4.1 leads to

$$PE_l = 1 - p^2 - \theta \sum_{i=l}^{n} p_i (1 - p_i)$$

which differs slightly from the expression based on Recommendation 4.2.

Box 7.1 Provided by Professor Bruce Weir

Consider a mixture that has alleles $A_1, \ldots, A_n$ present. We require the exclusion probability at locus l (PE_l). We start by considering the sum of all homozygotes and heterozygotes that are entirely within the mixture:

$$\text{sum of the homs} \sum_{i=1}^{n} (p_i^2 + \theta p_i (1 - p_i)) + \text{sum of the hets} \sum_{i \neq j} (1-\theta)\, p_i\, p_j$$

$$= \sum_{i=1}^{n} (p_i^2 + \theta p_i (1 - p_i)) + \sum_{i \neq j} (1-\theta)\, p_i\, p_j$$

$$= \sum_{i=1}^{n} p_i^2 + \sum_{i \neq j} p_i\, p_j + \sum_{i=1}^{n} \theta p_i (1 - p_i) - \sum_{i \neq j} \theta p_i\, p_j$$

$$= \left(\sum_{i=1}^{n} p_i\right)^2 + \theta \sum_{i=1}^{n} p_i - \theta\left(\sum_{i=1}^{n} p_i^2 + \sum_{i \neq j} p_i\, p_j\right)$$

$$= \left(\sum_{i=1}^{n} p_i\right)^2 + \theta \sum_{i=1}^{n} p_i - \theta\left(\sum_{i=1}^{n} p_i\right)^2$$

Next we write

$$p = \sum_{i=1}^{n} p_i$$

so the sum above becomes

$$= p^2 + \theta p - \theta p^2$$
$$= p^2 + \theta p (1 - p)$$

Since the exclusion probability $PE = 1-$ the sum of the homs $+$ the sum of the hets,

$$PE = 1 - p^2 - \theta p (1 - p)$$

as given by Budowle.

However, applying the rationale of NRC II Recommendation 4.1 gives the result as follows[a] (also provided by Professor Weir):

$$\text{sum of the homs } \sum_{i=1}^{n} (p_i^2 + \theta p_i (1-p_i)) + \text{sum of the hets } \sum_{i \neq j} p_i p_j$$
$$= \sum_{i=1}^{n} (p_i^2 + \theta p_i (1 - p_i)) + \sum_{i \neq j} p_i p_j = p^2 + \theta \sum_{i=1}^{n} p_i (1 - p_i)$$

Hence

$$PE = 1 - p^2 - \theta \sum_{i=1}^{n} p_i (1 - p_i)$$

The *PE* across multiple loci is calculated as

$$PE = 1 - \prod_{l} (1 - PE_l)$$

The advantages of the exclusion probability approach are often cited as simplicity and the fact that the number of contributors need not be assumed. In Box 7.2 we give a well-worded argument for the use of this approach that was provided by Laszlo Szabo of the Tasmanian Forensic Science Laboratory.

NRC II comments on a similar exclusion approach (also advocated by NRC I) by saying that the "… calculation is hard to justify, because it does not make use of some of the information available, namely, the genotype of the suspect."

[a] No factor of 2 is required since the summation is over $i \neq j$ rather than $i < j$.

Box 7.2 Arguments for the Use of the RMNE approach by Laszlo Szabo (Tasmania Forensic Science Laboratory)

As the defendant has a right to silence, we will usually never know what the defense hypothesis is, and to impose one on the court from a myriad of LR options may be unwarranted (it might be the wrong one).

Given that both RMNE and the likelihood ratio are valid approaches to the mixtures problem, the defendant may well prefer RMNE, as it is generally much more conservative than the likelihood ratio for the same data. The difficulty here occurs when a forensic laboratory quotes the most compelling likelihood ratio for a complex mixture (say around several billion), but does not report less impressive numbers (say around a million) for other likelihood ratio scenarios (even though these calculations may appear in the case file), and the RMNE calculation comes in at 1 in 20,000 for the same data.

The RMNE approach allows the evidential value of a crime scene profile to be estimated without reference to a suspect's DNA profile. This is important in cases without a suspect, where the investigator can be given some indication as to the potential usefulness of the DNA evidence from the crime scene.

Similarly, RMNE finds application in the Tasmanian DNA database, where all profiles (including partial and mixed crime scene profiles) have a calculation associated with them, so that we can see at a glance the strength of any DNA hits. So if we put a suspect on the DNA database and obtain a number of hits to complex crime scene mixtures, we can see immediately if these are "good matches" or not. We also have a policy of not putting a mixture on the DNA database unless the RMNE calculation is at least as good as 1 in 50. These approaches require RMNE, which is independent of knowledge of the suspect's profile.

Intuitively, RMNE is easier to explain to a jury and express in reports than the likelihood ratio, and is probably closer to what the court wants — e.g., the suspect matches the mixture, but what if this is the wrong person — then what is the probability that someone else in the population would also match the mixture (i.e., not be excluded as a contributor).

Weir[838] suggests, correctly, that exclusion probabilities "often rob the items of any probative value."

Brenner[81] gives a brilliant explanation of the shortcomings of the probability of exclusion. We follow him here. The evidence has two parts: (1) blood types of the suspect and (2) blood types of the mixed stain. Together, this information would let us infer that: (3) the suspect is not excluded.

Brenner points out that (3) can be deduced from (1) and (2). But (1) cannot be deduced from (2) and (3), or from (1) and (3). Hence the use of (1) and (3), or (2) and (3) is a loss of information. The likelihood ratio is a summary of the information in (1) and (2), whereas an exclusion probability is a summary of the evidence in (2) and (3). He concludes:

> In … a mixed stain case the exclusion probability usually discards a lot of information compared to the correct, likelihood ratio, approach. But still the exclusion probability may be acceptable sometimes.

There are occasional debates in court and among scientists about the merits of the RMNE approach versus the alternatives about to be described. Each method is in use in some jurisdictions. Generally, worldwide the move is away from RMNE toward likelihood ratios. It may be worthwhile briefly summarizing the pros and cons of RMNE before we move on. There are two pros: (1) It makes one fewer assumption in that it does not require an assumption of the number of contributors to a mixture. This is a fairly weak advantage since the assumption of the number of contributors is firmly grounded when a highly discriminatory multiplex is used. (2) It is easier to explain in court.

The cons are: (1) It wastes the information contained in the genotype of the suspect and hence makes weaker use of the available information. Hence it often robs the evidence of its true probative power. (2) The likelihood ratio approaches are developed within a consistent logical framework.

7.3 Bayesian Approaches

7.3.1 Models Employing Qualitative Approaches

Before the advent of automated fluorescent techniques (which provide quantitative data such as peak height and area), mixtures were interpreted without taking account of quantitative aspects of the data. The development of this qualitative style of interpretation commenced during the single-locus probe (SLP) era.[277] It is the method supported by NRC II who said the "… correct approach (the likelihood ratio approach), we believe, was described by Evett et al."

This approach has also received judicial sanction in the U.S. When Professor Weir presented the evidence regarding the mixed stains in the Bronco automobile during the trial of O.J. Simpson, he advocated the use of likelihood ratios. The defense preferred to argue for exclusion probabilities as suggested by NRC I. Judge Ito commented, "I find that

the analysis offered by Dr Weir is the more accurate and closest to what the evidence truly represents" (transcript page 33, 345), quoted in Weir.[835,838]

The likelihood ratio approach has been implemented in various guises ranging from use of the simple product rule to inclusion of sampling error and subpopulation corrections through to a refined treatment that accounts for all of these factors. In any manifestation it is superior to a probability of exclusion.

As with any Bayesian application, a key step is the formulation of the hypotheses. In fact, this is often the most difficult step, and relies on an understanding of the pertinent questions that are before the court and on what background information may be agreed on.

One of the most important factors that may be decided from the circumstances of the case or by agreement between prosecution and defense is whether any persons may be assumed to be present in the mixture.

To put this into context, consider a case in which fingernail clippings have been taken from a woman who has been assaulted and claims to have scratched her attacker. Suppose that a mixed DNA profile is obtained, which appears to consist of DNA from two individuals and can be fully explained by the presence of DNA from both the woman and her suspected attacker. The expectations from this type of sample and the circumstances of the case suggest that DNA from the complainant is likely to be present irrespective of whether there is any DNA from her attacker. Furthermore, the assumption seems wholly justified as there is prima facie evidence, from the mixed profile itself, of a contribution of DNA from the donor herself. Therefore, it may not be in contention that the profile of the complainant is present. Under these circumstances, it seems reasonable to form the following two hypotheses:

H_p: The nail clippings contain the DNA of the complainant and the suspect.

H_d: The nail clippings contain the DNA of the complainant and an unknown unrelated person.

The presence of DNA from the complainant under both hypotheses effectively allows the scientist to "condition" on the presence of her DNA. In practical terms, this allows much or all of the profile of the other contributor to be deduced straightforwardly. Those alleles that are not attributable to the complainant must be from the other contributor.

At this point it will be noted that the resolution of the mixture has assumed that there were exactly two contributors. This assumption is unnecessary and the formulation can be generalized to any number of contributors under H_p or

H_d with an associated significant increase in complexity. However, it simplifies the manual analysis appreciably if one can make this assumption. Under many circumstances this type of assumption can be strongly justified. If each locus in a highly discriminating multiplex has only 1–4 alleles, it seems very likely that there are only two contributors.[b] One could state that there is no evidence to indicate a contribution of DNA from a third individual and, given the context of the case, there is no need to invoke the presence of DNA from a third individual to explain the observed result. The argument regarding this assumption is in fact no different from that involving an apparently single source stain. Strictly, that profile may be a mixture of DNA from two individuals, but the scientist assumes, justifiably, that it emanates from a single individual.

If we call the evidence of the alleles in the stain E and the genotypes of the complainant and the suspect G_v and G_s, respectively,[c] we require

$$LR = \frac{\Pr(E|G_s, G_v, H_p)}{\Pr(E|G_s, G_v, H_d)}$$

If we assume that E is independent of G_s under H_d (this is in effect the assumption of Hardy–Weinberg and linkage equilibrium), then

$$LR = \frac{\Pr(E|G_s, G_v, H_p)}{\Pr(E|G_v, H_d)}$$

We will now work through a series of examples based on common casework scenarios.

Example 7.1. Consider the situation where the fingernail clipping in the case described above has been typed at a locus and found to contain the alleles A_1, A_2, A_3, and A_4. The complainant is type A_1, A_2 and the suspect is type A_3, A_4.

First consider the situation under H_p. This hypothesis states that the mixture is made from the suspect (type A_3, A_4) and the complainant (type A_1, A_2); hence, we expect the swab always to be type A_1, A_2, A_3, A_4. Therefore, barring laboratory error, $\Pr(E \mid G_s, G_v, H_p) = 1$.

[b] Strictly, this statement is a transposition. A more correct statement would be that we can infer that most of the posterior density lies on those propositions that have two contributors.
[c] General usage is to use the terms "victim" and "suspect" here, hence G_v and G_s. However, the word "victim" has implications that may be unwarranted. For example, the defense may be that the sex was consensual or indeed the matter may be a "false complaint" altogether, in which case it would be interesting to argue who the victim is. We will attempt to use the word "complainant" in the text. However, we have persisted with G_v in the equations to avoid the potential confusion with G_c, which we have already used for crime stain and to keep aligned with previous publications.

Now consider the situation under H_d. This hypothesis states that the mixture is made from the complainant (type A_1, A_2) and a random person. This random person must be type A_3, A_4, which happens with probability $2p_3p_4$ (using the product rule, and writing the probability of allele A_i as p_i). Hence $\Pr(E|G_v, H_d) = 2p_3p_4$ and so

$$LR = \frac{1}{2p_3p_4}$$

A more general consideration of the Bayesian approach to two-person mixtures shows that most propositions encountered in casework fall into one of three families of propositions:

(i) H_p: The mixture contains the DNA of the complainant and the suspect.

H_d: The mixture contains the DNA of the complainant and an unknown unrelated person.

Or

H_p: The mixture contains the DNA of the suspect and the complainant.

H_d: The mixture contains the DNA of the suspect and an unknown unrelated person.

(ii) H_p: The mixture contains the DNA of suspect 1 and suspect 2.

H_d: The mixture contains the DNA of two unknown unrelated people.

Or

H_p: The mixture contains the DNA of complainant 1 and complainant 2.

H_d: The mixture contains the DNA of two unknown unrelated people.

(iii) H_p: The mixture contains the DNA of the suspect and an unknown unrelated person.

H_d: The mixture contains the DNA of two unknown unrelated people.

Or

H_p: The mixture contains the DNA of the complainant and an unknown unrelated person.

H_d: The mixture contains the DNA of two unknown unrelated people.

Example 7.1 was drawn from the first family of propositions (i) and involved the consideration of a four-allele pattern. The remainder of this section involves examining different combinations of alleles and hypotheses.

Example 7.2. Consider a situation where a semen-stained vaginal swab has been typed and found to contain the alleles A_1, A_2, and A_3. It is thought that

DNA from the complainant is present due to the failure to completely separate the spermatozoa from vaginal material. Suppose that the complainant has genotype A_1, A_2 and the suspect has genotype A_3, A_3. Again, use of the first family of propositions (i) seems most appropriate:

H_p: The mixture contains the DNA of the complainant and the suspect.

H_d: The mixture contains the DNA of the complainant and an unknown unrelated person.

Under H_p, the mixture is made from the suspect (type A_3, A_3) and the complainant (type A_1, A_2); hence we expect the swab to be type A_1, A_2, A_3. Barring laboratory error, we evaluate this as

$$\Pr(E|G_s, G_v, H_p) = 1$$

Under H_d, the mixture is made from the complainant (type A_1, A_2) and a random person. This random person can be one of three possible genotypes A_1, A_3, A_2, A_3, or A_3, A_3 which, using the simple product rule, occurs with probability $2p_1p_3 + 2p_2p_3 + p_3^2$; hence

$$LR = \frac{1}{2p_1p_3 + 2p_2p_3 + p_3^2} = \frac{1}{p_3(2p_1 + 2p_2 + p_3)}$$

Example 7.3. Consider a situation where two men have allegedly raped a woman. A crime stain has been typed and found to contain the alleles A_1, A_2, A_3, and A_4. One suspect (S_1) has been arrested and is determined to have genotype A_1, A_2. The complainant is genotype A_5, A_6 and therefore cannot have contributed to the observed mixture. In this situation the third (iii) family of propositions seems most appropriate. Namely

H_p: The mixture contains the DNA of the suspect and an unknown unrelated person.

H_d: The mixture contains the DNA of two unknown unrelated people.

Under H_p, the mixture is made from the suspect (type A_1, A_2) and a random person who must therefore be type A_3, A_4. Hence $\Pr(E \mid G_s, G_v, H_p) = 2p_3p_4$.

Under H_d the mixture is composed of DNA from two random persons. There are six ways in which two individuals can contribute four alleles. Thus,

the potential combinations are:

Person 1	Person 2	Joint Probability
A_1, A_2	A_3, A_4	$4p_1p_2p_3p_4$
A_1, A_3	A_2, A_4	$4p_1p_2p_3p_4$
A_1, A_4	A_2, A_3	$4p_1p_2p_3p_4$
A_2, A_3	A_1, A_4	$4p_1p_2p_3p_4$
A_2, A_4	A_1, A_3	$4p_1p_2p_3p_4$
A_3, A_4	A_1, A_2	$4p_1p_2p_3p_4$
		Sum $= 24p_1p_2p_3p_4$

Hence, using the simple product rule,

$$LR = \frac{2p_3p_4}{24p_1p_2p_3p_4} = \frac{1}{12\ p_1p_2}$$

Suppose that sometime later a second suspect (S_2) is arrested who has genotype A_3, A_4. In theory, one could now consider using the second family of propositions (ii) and evaluating the joint evidence against S_1 and S_2. Under these circumstances:

H_p: The mixture contains the DNA of S_1 and S_2.

H_d: The mixture contains the DNA of two unknown unrelated people.

Under H_p, the mixture is made from suspect 1 (genotype A_1, A_2) and suspect 2 who must be genotype A_3, A_4. Hence

$$\Pr(E|G_{S_1}, G_{S_2}, H_p) = 1$$

and the denominator remains unchanged so

$$LR = \frac{1}{24p_1p_2p_3p_4}$$

Tables 7.1–7.3 provide the *LR* expressions using the simple product rule for each family of propositions for different combinations of alleles.

7.3.1.1 The General Formula

Weir et al.[849] give a general approach that allows calculation of the likelihood ratio for any mixture in those cases where peak area or height data are ignored. They also correct the typographical error regarding silent alleles that appeared in the original Evett et al. paper.[277] The nomenclature is provided in Table 7.4.

Table 7.1 Family (i)

Crime Sample	Complainant	Suspect	*LR*
A_1, A_2, A_3, A_4	A_1, A_2	A_3, A_4	$\frac{1}{2p_3 p_4}$
A_1, A_2, A_3	A_1, A_2	A_1, A_3 or A_2, A_3 or A_3, A_3	$\frac{1}{p_3(2p_1 + 2p_2 + p_3)}$
A_1, A_2, A_3	A_1, A_1	A_2, A_3	$\frac{1}{2p_2 p_3}$
A_1, A_2	A_1, A_2	A_1, A_1 or A_1, A_2 or A_2, A_2	$\frac{1}{(p_1 + p_2)^2}$
A_1, A_2	A_1, A_1	A_1, A_2 or A_2, A_2	$\frac{1}{p_2(2p_1 + p_2)}$
A_1, A_1	A_1, A_1	A_1, A_1	$\frac{1}{p_1^2}$

Table 7.2 Family (ii)

Crime Sample	Suspect 1	Suspect 2	*LR*
A_1, A_2, A_3, A_4	Any combination that contains all the alleles in the crime sample		$\frac{1}{24p_1 p_2 p_3 p_4}$
A_1, A_2, A_3			$\frac{1}{12p_1 p_2 p_3 (p_1 + p_2 + p_3)}$
A_1, A_2			$\frac{1}{2p_1 p_2 (2p_1^2 + 3p_1 p_2 + 2p_2^2)}$
A_1			$\frac{1}{p_1^4}$

The likelihood ratio is

$$LR = \frac{\Pr(E|H_p)}{\Pr(E|H_d)} = \frac{P_x(U_p|E, H_p)}{P_x(U_d|E, H_d)}$$

where

$$P_x(U|E, H) = T_0^{2x} - \sum_j T_{1;j}^{2x} + \sum_{j,k} T_{2;j,k}^{2x} - \sum_{j,k,l} T_{3;j,k,l}^{2x} + \cdots$$

Table 7.3 Family (iii)

Crime Sample	Known Suspect	*LR*
A_1, A_2, A_3, A_4	A_1, A_2	$\frac{1}{12p_1p_2}$
A_1, A_2, A_3	A_1, A_2	$\frac{2p_1 + 2p_2 + p_3}{12p_1p_2(p_1 + p_2 + p_3)}$
A_1, A_2, A_3	A_1, A_1	$\frac{1}{6p_1(p_1 + p_2 + p_3)}$
A_1, A_2	A_1, A_2	$\frac{(p_1 + p_2)^2}{2p_1p_2(2p_1^2 + 3p_1p_2 + 2p_2^2)}$
A_1, A_2	A_1, A_1	$\frac{2p_1 + p_2}{2p_1(2p_1^2 + 3p_1p_2 + 2p_2^2)}$
A_1, A_1	A_1, A_1	$\frac{1}{p_1^2}$

Table 7.4 Nomenclature for the General Formulation of Weir et al. for a Mixture *LR* without Accounting for Peak Area, Peak Height, or Subpopulation Effects

E	The set of alleles in the crime stain
H	A shorthand simply to avoid writing H_p or H_d
x	The number of unknown contributors to the stain under H
U	The set of alleles in E not carried by the known contributors under H
ϕ	The empty set. This may be necessary if there are no alleles in U not carried by the known contributors
$P_x(U\|E, H)$	The probability that x unknown contributors carry the alleles in U but no alleles outside E under H. We differ from Weir et al. here by specifying the hypothesis. This was implicit in the original publication. The difference is pedantic and cosmetic
T_0	The sum of all allele probabilities in E
$T_{1;j}$	The sum of all allele probabilities in E except the jth allele in U
$T_{2;j,k}$	The sum of all allele probabilities in E except the jth and kth allele in U

Mortera et al.[564] give an elegant implementation of this approach for any number of known or unknown contributors.

7.3.2 Models Employing Quantitative and Qualitative Data

The Evett et al.[277] paper was the progenitor of what we now term the binary model. In these manual methods, the scientist uses expert judgement, together with a number of numerical parameters, to decide which, if any, of all the possible genotype combinations at a locus can be excluded. Effectively

this assigns a probability of 1 or 0 to each of the genotypes based on whether the analyst considers them to be possible or impossible. This is based on expert judgement or whether the quantitative information falls within certain predefined parameters for the multiplex system. Strictly, partitioning of all possible contributing genotypes as included (weight 1) or excluded (weight 0) will never lead to the correct likelihood ratio since all possible contributing genotypes should have some weight between 0 and 1. This has naturally led to the development of alternative continuous probability models. These approaches attempt to take account of the need to weigh all of the possible genotypes but generally require software support, as they are computationally intensive. Foremost among these is the model that is based on treating peak areas as random variables and determining the probability of these peak areas for any given set of contributing genotypes.[271] These probabilities can be shown to act as the weights discussed above.

We will first discuss the binary model at some length here as it is the practice currently implemented in many laboratories internationally and then, more briefly, the continuous models. The binary model can be applied with or without a subpopulation correction.

7.3.2.1 Quantitative Data — Peak Areas or Heights?

Automated sequencers display quantitative information giving both allele peak height and peak area. The question of the use of peak height or area has arisen, although little has been published on this matter. Theoretical considerations would tend to favor area as more accurately reflecting the amount of DNA present. This is most likely to be true if peaks have differing shapes, as peak area should more accurately adjust for differing peak morphology. For instance, the higher molecular weight alleles might be broader and squatter.

Due to the fact that peaks are approximately triangular, one would not expect to see a doubling of height to equate to a doubling of area (a doubling of height should lead to a quadrupling of area if the shape remains the same and is triangular). On the other hand, when peak heights are low, apparently arbitrary decisions by the software as to where to delineate the boundaries of a peak can lead to the baseline contributing to the total peak area — the so-called "plinth effect." Thus, when peaks are low, height may be preferable to area. These theoretical considerations aside, limited trials by the authors on heterozygote balance suggest that area has no practical advantage or disadvantage over height. No preference is given here, although consistency is important. The answer to the height or area question therefore seems to be — both!

7.3.3 The Binary Model

The binary model is a manual method for the resolution of two-person mixtures. As it is applied manually, it is not readily extended to higher order

mixtures as the mathematics become complex and unwieldy. It relies upon the experience of the expert together with the application of a number of numerical guidelines. The rationale was outlined by Clayton et al.[182]

Interpretation of mixtures cannot proceed without an understanding about how nonmixtures behave. Before interpreting a potential mixture, it is important to investigate the behavior of the multiplex system of interest. First, it is necessary to have an appreciation for the normal variation in peak area between the two alleles of a known heterozygote (termed heterozygote balance, *Hb*). Second, it is necessary to understand the incidence and extent of stuttering. Such data can easily be collected from the study of single source reference samples during laboratory validation (see, for example, Gill et al.[343]). From such studies, it has become clear that loci can behave somewhat differently from each other, depending on which multiplex is being considered. The frequency and extent of stuttering can vary between individual alleles within a locus, depending on their size and internal sequence structure (a discussion of this can be found in Chapter 1). Other artifactual peaks such as the $N/N+1$ bands or primer–dimer very rarely confuse the interpretation.

It is possible to construct some general guidelines for all multiplexes, although it is better that the specific behaviors of each system are obtained during laboratory validation. Heterozygote balance is rarely found outside the range $0.6 \leq Hb \leq 1.67$. Stutter peaks are seldom more than 0.15 times the size of the parent allele.

By a consideration of these parameters, one can infer that pairs of alleles where $0.6 \geq Hb$ or $Hb \geq 1.67$ should not be from a heterozygote unless one or more peaks overlap another contributor. Second, we infer that in mixtures, if a band in a stutter position approaches or exceeds 0.15 of the area of the main allele, then it should be considered, at least at some stage, as a minor allele. Further instructive observations can be gained from the study of artificially created DNA mixtures using individuals of known genotype. Such experimentation should also form part of the laboratory validation program. From such validation experiments, it has become clear that where alleles were shared between two contributors, post-PCR, the area of the shared allele was approximately the sum of the two contributions.

Varying the mixture proportion in such experiments was also instructive. From this it could be demonstrated that, post-PCR, the peak areas of the alleles from the two contributors were approximately in the same proportion to the initial amount of DNA from the two contributors. Thus, an approximate mixing proportion could be estimated from the relative peak areas of the alleles for each contributor.

Mixture proportion, estimated in this way, was found to vary between the different loci from the same amplification. We have assumed that this is the

corollary of heterozygote imbalance. If the two alleles of a heterozygote can amplify differently, the alleles of one contributor could amplify differently to those from another contributor.

7.3.3.1 Application of the Binary Model

The interpretation of mixtures according to the binary model follows a series of steps, which will be outlined here.[182]

Step 1: Identification of a mixed profile. A mixture can only be interpreted in full if the alleles of the minor component are above the background noise. In practice, the threshold represents approximately a 1:10 mixture ratio for autosomal STR multiplexes. Ratios of at least 1:50 are reported to have been interpretable for Y STR multiplexes. If a locus is observed with more than two peaks, or if severe imbalance is present, then there is an inference that more than one individual is present in the profile. Occasionally, extra bands or unbalanced peaks may have either a genetic or artifactual origin. The following alternatives should be considered:

- Allelic artifacts — such as stutters: The great majority of artifacts can be excluded as nonallelic.[343] However, it is not always possible to exclude stutters since they are indeed allelic products, often differing structurally from the associated allele by just one repeat unit. It follows that alleles from a low-level contributor and stutters associated with the major contribution may be approximately equivalent in size and difficult or impossible to distinguish.
- Nonspecific artifacts.
- Software-induced artifacts — e.g., pull-up peaks.
- Poor operator technique — e.g., lane-to-lane leakage.
- Masking effects: A mixture may not always be evidenced by the presence of three or more bands at every locus. This would occur in cases where the contributors to a mixture actually share alleles at a particular locus. It is quite difficult for masking to occur at every locus of a polymorphic multiplex. However, masking may be more likely when the contributors are relatives or with less polymorphic multiplexes. Masked alleles may be evidenced by severe peak imbalance. Consider two individuals sharing the same alleles at, say, HUMD18S51. Contributor 1 is genotype 14, 14 while contributor 2 is genotype 14, 15. If the mixture ratio is 1:1, then the ratio of the 14:15 peak areas will be approximately 3:1, respectively, and pronounced peak imbalance will be observed.
- Suppressed amplification of an allele — putatively due to primer binding site mutation.
- Promoted amplification of an allele — putatively due to flanking region mutation.

- Multiple banded or unbalanced profiles generated as a result of genetic phenomena such as trisomy, translocation, and somatic mutation.

If the presence of any of the genetic phenomena such as trisomy, gene duplication, or primer binding site mutations is suspected, these phenomena should also appear in the reference sample. An exception may be a tissue-specific somatic mutation. In the latter case, confirmation may depend upon a reference sample, which has the same origin as the case stain. Plausibly we cannot completely rule out the possibility that the appearance of somatic mutations could vary over time within tissues such as the buccal lining, which consists of rapidly dividing cells.

Step 2: Designation of allelic peaks. Once an assessment of the mixed profile has been made, designation of peaks as allelic or possibly allelic follows the principles described by Gill et al.[343,344] The positions of peaks (typically measured in base pairs, bp) are compared with allelic ladders and scored only if they are within the following guidelines:

1. Alleles should be within ± 0.5 bp of the designated control allelic ladder marker.
2. The band shift for each allele should be approximately consistent.

Step 3: Identification of the potential number of contributors. Once the most supported explanation for multiple allelic peaks and peak imbalance has been attributed to the profile being a mixture, the next step is to estimate the minimum number of contributors. The maximum number of alleles at any locus for a simple two-person mixture is four (given that no genetic phenomena are involved). Care must be taken not to confuse stutters with true alleles. Consideration of the circumstances of a case is often important in assessing the number of potential contributors: for example, in a vaginal swab it would not be surprising to find a minor component from the complainant.

Torres et al.[775] present an impressive four-year casework study:

> In our own and other authors' experience two person mixtures account for the overwhelming majority of mixtures encountered during casework, but occasionally mixtures of three or more persons are seen with more than four alleles at some loci. Eight of the 163 mixed samples corresponded to such higher-order profiles.

The simple methods described above will suffice for highly polymorphic loci. Superior methods have been suggested for determining the number of

contributors that may come into increased use for, say, SNPs where determination of whether or not a sample is a mixture at all may be quite problematic. Egeland et al.[245] proposed calculating the likelihood of the profile for differing numbers of contributors and then taking the maximum over the number of contributors. This approach could also be extended to include the effect of persons known to be contributors or when relatives are suspected of being co-contributors.

Step 4: Estimation of the mixture proportion or ratio of the individuals contributing to the mixture. There are a few terms that are particularly useful when considering mixtures. These are set out in Table 7.5.

It has been demonstrated that if DNA templates are mixed, then this ratio will be approximately, but not exactly, preserved throughout all loci.[350] Kirkham[466] observed that the estimated mixture proportion at some loci could differ by as much as 0.35 from the known mixture proportion. Understanding of this variation would benefit from further research studies. Typically we have found that forensic caseworkers prefer to work with the mixture ratio, whereas the mathematics flows slightly better working with the mixture proportion. A simple relationship exists between the two parameters (but not necessarily their estimates):

$$M_x = \frac{M_R}{1 + M_R}, \quad M_R = \frac{M_x}{1 - M_x}$$

The mixture proportion can range from the contributors being approximately equal to each other, to one being in great excess. It is helpful to classify the mixture as shown in Table 7.6.

Either the mixture proportion or the ratio can be estimated relatively easily when there are no shared alleles (Figure 7.2).

Consider the profile of HUMD18S51 in Figure 7.2. It is possible to pair the alleles into the minor contributor (14, 15) and the major contributor (16, 18). The mixture proportion can be estimated from peak areas (ϕ):

$$\hat{M}_x = \frac{\phi_c + \phi_d}{\phi_a + \phi_b + \phi_c + \phi_d} = \frac{1375 + 1465}{1375 + 1465 + 2867 + 2281} = \frac{2840}{7988} = 0.36$$

$$1 - \hat{M}_x = \frac{\phi_c + \phi_d}{\phi_a + \phi_b + \phi_c + \phi_d} = 0.64$$

Alternatively, a mixture ratio can be estimated as

$$\hat{M}_R = \frac{\phi_a + \phi_b}{\phi_c + \phi_d} \cong 0.55$$

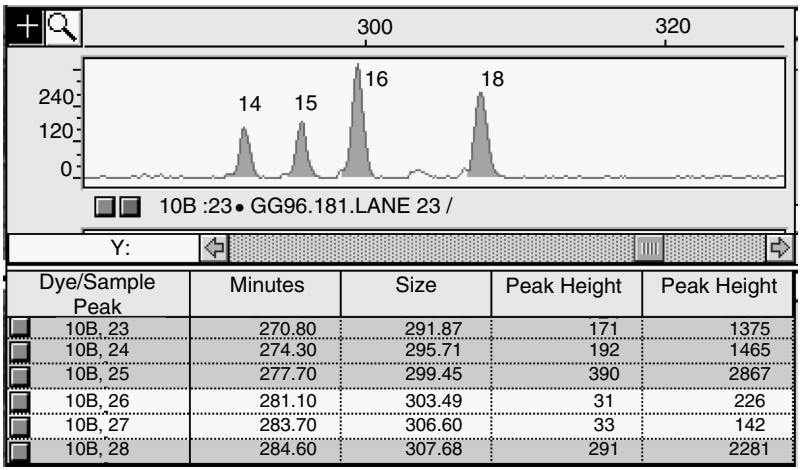
which we may write as approximately 1:2.

Table 7.5 Mixture Nomenclature

Term	Definition
M_R The true but unknown mixture ratio	We assume this to be a preamplification term. Under many circumstances, it will be possible to assume that it is constant across all loci preamplification. Such a circumstance would be when we have a large number of undegraded diploid cells in the template, and sampling effects in the preamplification stage may be ignored
$\hat{M}_R^l \mid ab{:}cd$ The estimated postamplification mixture ratio at locus l, conditional on the genotypes being ab and cd	The estimated ratio of the components postamplification at locus l in the mixture postulated to be ab and cd. We explicitly allow $\hat{M}_R^l \mid ab{:}cd$ to vary across loci. Whether or not the mixing ratio is constant across loci preamplification, there is no requirement for it to be constant postamplification. For instance, $\hat{M}_R^l \mid ab{:}cd = 1$ means that the mixture is comprised of a 1:1 mix of two components ab and cd at locus l, $\hat{M}_R^l \mid ab{:}cd = 2$ means that the mixture is comprised of a 2:1 mix of two components ab and cd at locus l, and $\hat{M}_R^l \mid ab{:}cd = \frac{1}{2}$ means that the mixture is comprised of a 1:2 mix of two components ab and cd at locus l. Hence we expect $\hat{M}_R^l \mid ab{:}cd = \dfrac{1}{\hat{M}_R^l \mid cd{:}ab}$
$\hat{M}_R$ Estimated mixture ratio abbreviation	For brevity, we may drop the conditioning on the genotypes and the nomination of locus. However, strictly the mixture ratio can only be estimated if the genotypes are known. The ratio postamplification is likely to vary across loci even if it is constant preamplification
M_x The true but unknown mixture proportion	Again we assume this to be a preamplification term
$\hat{M}_x^l \mid ab{:}cd$ The estimated postamplification mixture proportion of the genotype ab in the mixture of ab and cd at locus l	The estimated postamplification proportion of one component in the mixture at locus l. This will usually be taken to be the minor component; however, this distinction is not necessary. For instance, $\hat{M}_R^l \mid ab{:}cd = 0.5$ means that genotype ab is estimated to represent 50% of the total area in the epg. $\hat{M}_x^l \mid ab{:}cd = 0.33$ means that genotype ab is estimated to represent 33% of the total area in the epg
$\hat{M}_x$ The estimated mixture proportion abbreviation	For brevity, we may drop the conditioning on the genotypes and the nomination of the locus. However, strictly the mixture proportion can only be estimated if the genotypes are known. The ratio postamplification is likely to vary across loci even if it is constant preamplification
a,b,c etc	The name of the allelic peak present in the mixture
ϕ_a	The area of peak a
$\phi_{a\&b}$	An abbreviation for $\phi_a + \phi_b$
p_a	The allele probability for allele a in some relevant population

Table 7.6 A Useful Classification Scheme for Mixtures

Classification	Description	Approximate Definition
Type A	No clearly defined major contributor	$0.33 \leq \hat{M}_x \leq 0.67$ $0.5 \leq \hat{M}_R \leq 2$
Type B	Clearly defined major and minor contributors	$0.13 \leq \hat{M}_x \leq 0.33$ $0.15 \leq \hat{M}_R \leq 0.5$
Type C	A mixture containing a low-level minor	Strictly this is any mixture where the minor is sufficiently low that it could be confused with an artifact such as a stutter
Type D	The majority of peaks are below 150 rfu in height	



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Height
10B, 23	270.80	291.87	171	1375
10B, 24	274.30	295.71	192	1465
10B, 25	277.70	299.45	390	2867
10B, 26	281.10	303.49	31	226
10B, 27	283.70	306.60	33	142
10B, 28	284.60	307.68	291	2281

Figure 7.2 Analysis of HUMD18S51 showing a mixed profile from two different heterozygotic individuals (run on an ABI 377 Gene Sequencer, Applied Biosystems, Foster City, CA).

Note, however, the conditional nature of this estimate in that it is based upon the genotypes of the contributors being 16, 18 and 14, 15, respectively. The formulae required are outlined in Table 7.7. There has been some argument that the mixture proportion (M_x) cannot be estimated. This appears to be a misunderstanding of the conditional nature of the mixing proportion estimate. It is indeed correct that the true mixing proportion is unknown, but this is also true of almost any other parameter in DNA or any other statistical problem. In addition, the process of estimating a mixture proportion requires a knowledge of the genotypes, and hence any mixture proportion estimate is

Table 7.7 Estimating the Mixture Proportion and Mixture Ratio

Proposed Genotype Combination			$\hat{M}_x$	$\hat{M}_R$
4-allele loci	1,2	3,4	$\hat{M}_x = \frac{\phi_1 + \phi_2}{\phi_1 + \phi_2 + \phi_3 + \phi_4}$	$\hat{M}_R = \frac{\phi_1 + \phi_2}{\phi_3 + \phi_4}$
3-allele loci	1,1	2,3	$\hat{M}_x = \frac{\phi_1}{\phi_1 + \phi_2 + \phi_3}$	$\hat{M}_R = \frac{\phi_1}{\phi_2 + \phi_3}$
	2,3	1,1	$\hat{M}_x = \frac{\phi_2 + \phi_3}{\phi_1 + \phi_2 + \phi_3}$	$\hat{M}_R = \frac{\phi_2 + \phi_3}{\phi_1}$
	1,2	1,3	$\hat{M}_x = \frac{\phi_2}{\phi_2 + \phi_3}$ [d]	$\hat{M}_R = \frac{\phi_2}{\phi_s}$
2-allele loci	1,1	2,2	$\hat{M}_x = \frac{\phi_1}{\phi_1 + \phi_2}$	$\hat{M}_R = \frac{\phi_1}{\phi_2}$
	1,2	2,2	$\hat{M}_x = \frac{2\phi_1}{\phi_1 + \phi_2}$	$\hat{M}_R = \frac{2\phi_1}{\phi_2 - \phi_1}$
	1,1	1,2	$\hat{M}_x = \frac{\phi_1 - \phi_2}{\phi_1 + \phi_2}$	$\hat{M}_R = \frac{\phi_2 - \phi_1}{2\phi_1}$
	1,2	1,2	No information is present	
1-allele loci	1,1	1,1	No information is present	

always conditioned on the genotypes. This means that different combinations of genotypes will result in different mixture proportion estimates. This is an integral part of the theory.

Further criticism has been leveled when M_x is subsequently used to help decide which genotypes are present at a locus on the basis that there is a circularity to the argument. That is, the assumed genotypes are used to estimate M_x and then M_x is used to decide the genotypes. We accept this criticism of the circularity in this argument, but in essence the binary model procedure simultaneously estimates those genotypes and mixing proportions that have high support. This approach is directly analogous to standard statistical iterative methods such as the EM algorithm. Once the mixture proportion or

[d] We have previously offered the more complex estimator

$$\hat{M}_x = \frac{3\phi_2 - \phi_3 + \phi_1}{2(\phi_2 + \phi_3 + \phi_1)}$$

Professor Bruce Weir[845] argues in favor of the simple estimator as the area of the shared peak is not proportional to the mixing proportion. We accept the force of his argument.

ratio has been estimated, it is then possible to calculate the expected peak areas for any given genotype combination.

It is possible to automate the process of M_x estimation using a software algorithm based on the concept of least squares residuals. This has been programmed at the FSS into a software package called PENDULUM. This software will be discussed later.

Step 5: Consideration of all possible genotype combinations. Having obtained an estimate of M_x from one or more loci, the next step is to enumerate all possible genotype combinations for the locus under consideration and generate a list of every possible combination. The number will vary according to whether the allele pattern has 4, 3, 2, or just 1 allele. These combinations are given in Table 7.8.

Note that all calculations are conditional on the hypothesized genotypes. Taking each locus separately, every genotype in the list is then considered in turn. Two parameters are next estimated: heterozygote balance (*Hb*) and mixture proportion (M_x). First, the *Hb* guidelines are applied. These are written as a set of mathematical rules and appear in Table 7.8. These "rules" need to be applied with care as no simple mathematical rule can match human judgement.

Second, M_x should be similar at each locus of the mixture and so should show consistency with the previously estimated mixture proportion.

Those combinations that are not supported on the basis of the guidelines are considered very unlikely[e] and are removed. In this way, the list may be reduced to leave only those allelic combinations that are well supported by the quantitative data.

Step 6: Compare reference samples. Once a list of all the well-supported genotypes has been generated and recorded, a comparison is made with the reference samples. If the genotype of an individual is such that he/she matches one of the well-supported combinations at each locus, then there is evidence that may support the suggestion that this person has contributed to the mixture and a statistical evaluation of that evidence may be warranted. Conversely, if at one or more loci the individual either lacks one or other or both alleles for either combination, then this is evidence supporting noncontribution. Even if the individual possesses the necessary alleles, but that combination was excluded during application of the rules, then this evidence also may support noncontribution.

7.3.3.2 Automation of M_x Estimation: PENDULUM

The FSS has developed these approaches[350] into a program called PENDULUM[535,619] for use on the 11-locus SGM$^+$ set (ten autosomal loci plus

[e] Again, strictly this is a transposition. The better term would be to consider these combinations to have a low posterior probability.

Table 7.8 Allele Combinations for Differing Loci and Heterozygote Balance (*Hb*) Rules

Individual 1		Individual 2	Guideline
Four-peak locus-6 combinations			
ab		*cd*	All hets in this block may
ac		*bd*	be checked using the
ad		*bc*	simple het guideline
bc		*ad*	
bd		*ac*	
cd		*ab*	
Three-peak locus-12 combinations			
aa	No test	*bc*	Simple het
ab	Shared het	*ac*	Shared het
ab	Shared het	*bc*	Shared het
ab	Simple het	*cc*	No test
ac	Simple het	*bb*	No test
ac	Shared het	*bc*	Shared het
bc	Simple het	*aa*	No test
ac	Shared het	*ab*	Shared het
bc	Shared het	*ab*	Shared het
cc	No test	*ab*	Simple het
bb	No test	*ac*	Simple het
bc	Shared het	*ac*	Shared het
Two-peak locus-7 combinations			
aa	Het hom	*ab*	Het hom
aa	No test	*bb*	No test
ab	Het hom	*bb*	Het hom
ab	Het hom	*aa*	Het hom
bb	No test	*aa*	No test
bb	Het hom	*ab*	Het hom
ab	Simple het	*ab*	Simple het
One-peak locus-1 combinations			
aa	No test	*aa*	No test

Hb is the accepted heterozygote balance level, say 0.6.
Het is an abbreviation for heterozygote.
Hom is an abbreviation for homozygote.
Simple het guideline

$$Hb \leq \frac{\phi_1}{\phi_2} \leq \frac{1}{Hb}$$

where peaks 1 and 2 are the low and high molecular weight peaks of the heterozygote, respectively.
Shared het guideline

$$Hb \leq \frac{\phi_s}{\phi_1 + \phi_2} \leq \frac{1}{Hb}$$

where s is the shared peak and peaks 1 and 2 are the peaks of the different heterozygotes.
Het Hom guideline $\phi_1 \geq Hb\phi_s$.

Amelogenin). The theory has been implemented for SGM$^+$, but is in principle extendable to any STR multiplex. This program performs two approaches to mixture analysis.

First the approach enumerates every possible genotype combination for the 11-locus major and minor contributors to a two-person mixture. This is typically many millions of possible combinations. Then, for each combination it finds the M_x value that minimizes the sum of the squared residuals across all loci simultaneously, by varying M_x until the minimum is found (hence the name PENDULUM, although strictly a hill climbing algorithm is used). Expected peak areas are calculated following the principles of consistency of mixture proportion across loci and the additivity of differing contributions to peak area. Hence this approach does not take account of heterozygous balance and stutter.

Each genotype combination is then ranked according to its residual sum of squares, the smallest sum being assumed to give the best fit to the data. It is important that this "best fit" is not interpreted as being the most probable combination. It is likely that the ranked list is in approximate probability order, but it should not be assumed that the sums of squares relate in any direct way to probability.

Limited trials suggest that the correct major/minor combination appears in the top 500 choices 75% of the time (remember that the number of possible combinations may be many millions). Examination of those instances where the best fit was not in the top 500 choices suggests that this is a result of inherent variability in the PCR process. Misses of the correct choice occurred when, for instance, the major had, contrary to expectation, made the two smaller peaks at one locus. The disappointing performance of this aspect of the program in terms of picking contributors is likely to improve if the repeatability of the PCR process improves. However, the program is limited in that it does not account for heterozygous balance, stutter, and furthermore the method cannot be extended into a probabilistic model. This does not detract from the overall performance of the program as a second interpretation strand is employed.

The second portion of PENDULUM applies the rules outlined in this chapter for preferential amplification (Table 7.8) and calculates ($\hat{M}_x^l$|postulated genotypes) for each locus, *l*. The postulated genotypes are only "passed" if the ($\hat{M}_x^l$|postulated genotypes) is within ± 0.35 (but may be user defined) of the PENDULUM average for the top 500 hits. Lastly, PENDULUM applies a simplified set of "F" rules (see later in this chapter). This whole approach has been programmed into a very professional package by Mr. Marcus Healy.

Linear mixture analysis (LMA): Recently, Perlin and Szabady[615] also used the minimization of the sum of squared deviations across all loci in their LMA algorithm. This approach used a more elegant mathematical procedure to search the space of possible genotypes, and in addition it deterministically

removed stutter before proceeding. We have some concerns about the deterministic removal of stutter. In addition, the constraint that ($\hat{M}_x^l$ | postulated genotypes) is constant across all loci postamplification is too strict. It is the preamplification M_x that may be assumed to be constant under some, but not all, circumstances. This stated, the method appears to be elegant and efficient.

7.3.3.3 Assumptions of the Binary Model

The binary model makes a number of assumptions. These include the following: (1) The mixture proportion is approximately constant across loci. (2) The peak area is proportional to the amount of DNA. (3) The area of "shared" peaks is the sum of the contribution of the two contributing individuals.

7.3.3.4 Allowing for Stutters

In some cases it will not be possible to decide whether or not a minor peak in a stutter position is composed entirely of area derived from stuttering or whether it does contain an allelic contribution. It is, however, possible to proceed on the basis that the peak may or may not be in part allelic.

In the example shown in Figure 7.3, peak *b* is in a stutter position. If this is a two-person mixture, the genotype of the minor contributor could therefore be *aa*, *ab*, *ac*, or *ad* if the peak at position *b* really is entirely stutter, or genotype *ab* if the peak at position *b* is not entirely due to stutter.

Gill et al.[344] suggested that such a situation can be treated using the assumption that the peak at position *b* may be all stutter (event *S* with probability $p(S)$) or not all stutter, in which case some must be allelic (event $\overline{S}$ with probability $p(\overline{S})$). This leads to a bound for the *LR*:

$$LR = \frac{1}{p(\overline{S})2p_a p_b + p(S)\{p_a^2 + 2p_a p_c + 2p_a p_d\}}$$

$$\geq \frac{1}{\{p_a^2 + 2p_a p_b + 2p_a p_c + 2p_a p_d\}}$$

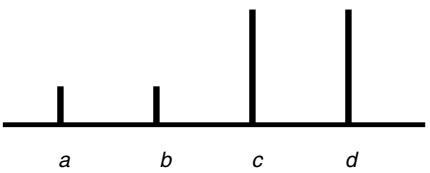


Figure 7.3 Stylized rendition of an electropherogram. This style of diagram is often referred to as a "Madonna Plot." It shows a profile comprising two minor bands *a, b* and two major bands *c, d*. The minor bands are <15% the area of the major bands and the distance between *b* and *c* is one repeat unit. Hence the *b* band could be a stutter.

7.3.3.5 *Allowing for Incomplete Representation*

To this point, the approaches for resolving mixtures have assumed full representation of alleles. However, for low peak area the possibility of allelic dropout must be considered. Most laboratories have a threshold above which it is assumed to be safe to designate a homozygote. Typically this is somewhere between 75 and 150 rfu.

There are at least two possible solutions to the problem: (1) The binary method is not implemented when the peak areas are sufficiently low that alleles may not be represented. (2) A rule-based system is invoked to deal with the possibility that an allele may not have been detected.

The use of a designation to denote a potentially missing allele such as "F" in the U.K. or "O" in New Zealand is an example of such a system. This system is implemented in the PENDULUM program. Consider the situation with a three-banded profile *a*, *b*, *c* with heights of 300, 300, and 100 rfu, respectively. Assume that the combination *ab*:*cc* is under consideration. Is it safe to assign the genotype of the minor component as a *cc* homozygote? In such a case, it is recommended that the genotype of the minor component be assigned as *cO*, by which we mean the *c* allele and any other allele (including the *cc* homozygote).

In some instances, there will be no visible minor peaks at one or a few loci. The question arises as to whether the minor peaks are masked or have dropped out. In such a case, this decision needs to be based on the other loci. It may be useful in such cases to calculate predicted areas. This does, however, presuppose consistency of mixing ratio across loci. This latter assumption needs to be checked on a case-by-case basis as it may be affected by poor template, preferential degradation, or other detectable biochemical phenomena. Predictions will necessarily be made on an area basis. If this approach is to be applied manually by a trained analyst, they should also consider any pattern of degradation that exists between high and low molecular weight loci.

These caveats made, consider the situation where we have a minor component with approximate mixture proportion M_x. Given that we accept that there is variability in M_x between loci, we require a realistic lower bound on M_x at a locus that we denote min M_x^l. Then the expected peak areas are

$$E[Area] = \begin{cases} \min M_x^l T & \text{for homozygotes} \\ \dfrac{\min M_x^l T}{2} & \text{for heterozygotes} \end{cases}$$

where T is the total peak area at the locus.

If the expected peak areas are below the threshold where dropout may occur, then it is wise to consider using the *O*-designation. For three-banded

profiles, it may be necessary to consider the addition of one *O*-designation, whereas there are circumstances for one- and two-banded profiles where it may be advisable to consider two, or even three, *O*-designations.

7.3.3.6 Reliability of the Binary Model.[f, g]

The reliability of the binary model has been questioned in the literature. Specifically:

> There are also semi-intuitive methods that discount selected genotypic combinations, but these methods must be used with extreme care as they can give misleading results. (Evett,[267] p. 205).

Given the widespread use of "judgement"-based techniques and the binary model, and given the existence of this published criticism, it is worthwhile examining under what circumstances the binary model will be unreliable.

It will be necessary for this consideration to extend to multiple loci; however, let us initially simply consider the one locus shown in Figure 7.4. In addition, factors such as stutter and preferential amplification may be important but are ignored here for simplicity. We have formative models for quantitatively modeling stutter and preferential amplification; however, they are not implemented. Inspection of such an epg would suggest that the most supported combinations of genotypes for two contributors would be *ad*:*bc* and the reverse *bc*:*ad*. By this we mean that one individual is *ad* whereas the other is *bc*. However, the combinations *ab*:*cd*, *ac*:*bd*, *cd*:*ab*, and *bd*:*ac* would not be excluded. Next suppose that this case is a double rape. The complainant's DNA does not appear in the mixture. One suspect is tested whose genotype is *ab*. Should we be concerned? The suspect's genotype does not

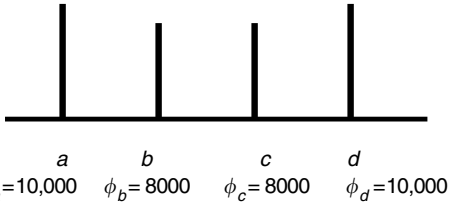


Figure 7.4 Madonna plot of a hypothetical electropherogram of a mixed profile at an STR locus.

[f] We acknowledge the contribution to the following approach made by Dr. Mark Perlin during discussions of the likelihood function in Perlin and Szabady.[615]

[g] We are indebted to two anonymous referees of a paper that we did not pursue. These referees greatly facilitated the writing of this section. The original paper was to have been authored by Christopher Triggs, Mark Perlin, Peter Gill, and John Buckleton.

appear in the most supported combination; however, it certainly is not excluded. This is the issue that Evett raised. It is not without some force. The binary model does not make full use of the quantitative information, although it does use it in some ways. The issue, if it exists at all, comes from the all or nothing partition of the binary model. It would be better if every genotype combination were given some weight. As a general observation, every time we fail to utilize information in forensic casework, we run a risk of being less accurate than is possible. However, it is important to remember Box's famous statement that all models are wrong, but it is not necessary for them to be correct; all they need to be is useful. We turn to an attempt to quantify the potential effect.

We denote the vector of observed peak areas by $\underline{d}$. Pursuing the example shown in Figure 7.4, this would be the areas of the four peaks:

$$\underline{d} = \{\phi_a, \phi_b, \phi_c, \phi_d\} = \{10{,}000; 8000; 8000; 10{,}000\}$$

As usual, we define two hypotheses. We need H_p to specify those individuals deemed to be in the mixture. Additional unknown individuals may be required. These specified individuals constrain the combinations of genotypes that are permissible under H_p. Sometimes there is only one combination that fulfils this requirement, but we will investigate instances where there are multiple possibilities. Suppose that there are p combinations, $S_1,\ldots, S_p$, of possible genotypes under H_p. Each combination, S_i, will contain m genotypes $\underline{G}_{i1}\ldots\underline{G}_{im}$. In the example discussed in Figure 7.4, there is one known contributor with genotype ab, and so we have two combinations ($p=2$) containing two genotypes[h] ($m=2$). m need not be constant over combinations. Hence S_1 declares that $G_{11}=ab$, $G_{12}=cd$ are contributors and S_2 declares that $G_{21} = cd$, $G_{22}=ab$ are contributors. In case the nomenclature has clouded the issue, it may be worthwhile restating that the above means that H_p has declared that there are two contributors. One is the suspect who is ab, and the other is unknown but must be cd.

Similarly, H_d specifies a number of known and unknown contributors. These constrain the combinations, q, of possible genotypes, each combination containing n genotypes $\underline{G}_{j1}\ldots\underline{G}_{jn}$, where n need not be constant over combinations.

Usually there will be more combinations possible under H_d than H_p. In the example under consideration, there are $q = 6$ combinations of $n = 2$ genotypes:

$$\begin{array}{ll} G_{11} = ab, G_{12} = cd & G_{21} = cd, G_{22} = ab \\ G_{31} = ac, G_{32} = bd & G_{41} = bd, G_{42} = ac \\ G_{51} = ad, G_{52} = bc & G_{61} = bc, G_{62} = ad \end{array}$$

[h] We represent a genotype as a vector with two elements. Thus, person i has the genotype $\underline{G}_i$ with alleles a_{ir1}, a_{ir2} at locus r.

Typically the prosecution combination should appear in the defense set. If not, then effectively the evidence suggests an exclusion. In this example, we note that the prosecution combinations do, indeed, appear. It is necessary at this point to step to multiple loci.

We see that at locus 2 the most supported combination would be *eh*:*fg* or *fg*:*eh*. However, now that we have two loci, we see that the combinations *adeh*:*bcfg* and *bcfg*:*adeh* are more supported than *adfg*:*bceh* and *bceh*:*adfg*, although these latter combinations would not be eliminated. This example is intended to highlight the difference between the locus-by-locus approach and the whole-genotype approach. Most forensic scientists approach mixtures on a locus-by-locus basis. This is to say that they consider one locus at a time albeit they "assess" the mixture ratio across loci. It is useful from a mathematical point of view not to do this. If the full multilocus components of the mixture are considered, then the mixture proportion and the probability of the areas can be considered in a more meaningful way. However, the number of possible full multilocus combinations can be huge and beyond enumeration by hand.

As a simple example, Table 7.9 enumerates the six one-locus and the 36 full two-locus genotype combinations for the epg given in Figure 7.5.

We define w_k as the mixture proportion for person k. For a two-contributor mixture, there is therefore only one nonredundant fraction w_1, which we will write as w. In such a case, $w \equiv M_x$. We will restrict our attention to two-contributor mixtures under both H_p and H_d. In addition, we will only consider full multilocus genotypes. Hence the summations, particularly in the denominator, will typically be over a very large number of combinations.

We require

$$LR = \frac{\Pr(\underline{d}|H_p)}{\Pr(\underline{d}|H_d)}$$

Introducing the genotype combinations S_j,

$$LR = \frac{\sum_{j=1}^{p} \Pr(\underline{d}|S_j, H_p)\Pr(S_j|H_p)}{\sum_{k=1}^{q} \Pr(\underline{d}|S_k, H_d)\Pr(S_k|H_d)}$$

At this time there are only formative models to calculate $\Pr(\underline{d}|S_i, H_x)$. These are discussed in the section on continuous methods. Currently this term is assessed subjectively and based on judgement. Potential combinations S_i are assigned as included or excluded based on whether $\Pr(\underline{d}|S_i, H_x)$ is judged to be high or low.

It is interesting to note that explicit enumeration of the alleles present in a mixture would not be required if we could explicitly calculate $\Pr(\underline{d}|S_i, H_x)$.

Table 7.9 Enumeration of Genotype Combinations for Two Loci Following the Locus-By-Locus and the Multilocus Approaches

Locus 1		Locus 2	
Person 1	Person 2	Person 1	Person 2
Locus-by-locus approach			
ab	*cd*	*ef*	*gh*
ac	*bd*	*eg*	*fh*
ad	*bc*	*eh*	*fg*
bc	*ad*	*fg*	*eh*
bd	*ac*	*fh*	*eg*
cd	*ab*	*gh*	*ef*

Person 1	Person 2	Person 1	Person 2	Person 1	Person 2
Full multilocus approach					
abef	*cdgh*	*adef*	*bcgh*	*bdef*	*acgh*
abeg	*cdfh*	*adeg*	*bcfh*	*bdeg*	*acfh*
abeh	*cdfg*	*adeh*	*bcfg*	*bdeh*	*acfg*
abfg	*cdeh*	*adfg*	*bceh*	*bdfg*	*aceh*
abfh	*cdeg*	*adfh*	*bceg*	*bdfh*	*aceg*
abgh	*cdef*	*adgh*	*bcef*	*bdgh*	*acef*
acef	*bdgh*	*bcef*	*adgh*	*cdef*	*abgh*
aceg	*bdfh*	*bceg*	*adfh*	*cdeg*	*abfh*
aceh	*bdfg*	*bceh*	*adfg*	*cdeh*	*abfg*
acfg	*bdeh*	*bcfg*	*adeh*	*cdfg*	*abeh*
acfh	*bdeg*	*bcfh*	*adeg*	*cdfh*	*abeg*
acgh	*bdef*	*Bcgh*	*adef*	*cdgh*	*abef*

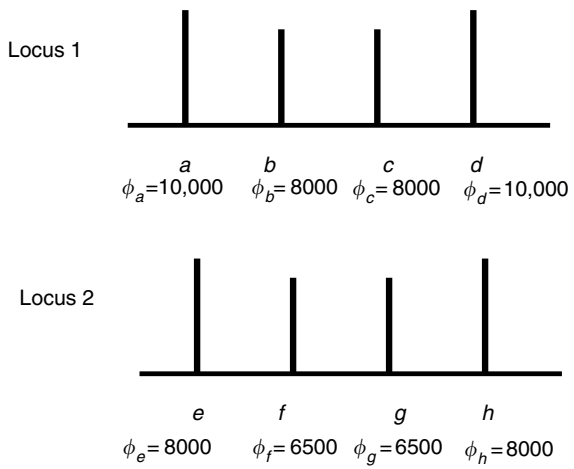


Figure 7.5 Madonna plot for two loci.

It is completely feasible to permute all possible combinations of genotypes either over all allele positions showing any area above background or even over all allele combinations regardless of area. In the example above, we could, for example, include in the list under H_d the combination *ab*:*ce*, where *e* is an allelic position where no area was observed at all. Let us call this combination S_7. Obviously, we would expect that $\Pr(\underline{d}|S_7, H_d) \approx 0$, and hence this combination effectively removes itself (strictly it will add a very small term to the denominator of the likelihood ratio). This approach would be very computationally expensive, but it would present some advantages in certain circumstances. These are not discussed here, but they do include the ability to handle allelic dropout and an increased potential for automation.

By introducing the mixing fraction w, and considering the density of w given either hypothesis H_p or H_d, $p(w|H_p)$, $p(w|H_d)$ respectively, we obtain

$$LR = \frac{\sum_{j=1}^{p} \int p(\underline{d}|S_j, w, H_p) p(S_j|w, H_p) p(w|H_p)\, \mathrm{d}w}{\sum_{k=1}^{q} \int p(\underline{d}|S_k, w, H_d) p(S_k|w, H_d) p(w|H_d)\, \mathrm{d}w}$$

We can simplify this expression by making a number of assumptions. First we assume that the mixing fraction w is not conditioned on either H_p or H_d so that $p(w|H_p) = p(w|H_d) = p(w)$. We further assume that the probability of observing a set of genotypes S_i in the population is independent of the mixing proportions so that $p(S_i|w, H_x) = p(S_i|H_x)$ for both H_p and H_d. We make the final assumption that $p(\underline{d}|S_i, w\, H_p) = p(\underline{d}|S_i, w, H_d) = p(\underline{d}|S_i, w)$ and hence obtain

$$\begin{aligned} LR &= \frac{\sum_{j=1}^{p} \int p\left(\underline{d}|S_j, w, H_p\right) p(S_j|w, H_p) p(w|H_p)\, \mathrm{d}w}{\sum_{k=1}^{q} \int p\left(\underline{d}|S_k, w, H_d\right) p(S_k|w, H_d) p(w|H_d)\, \mathrm{d}w} \\ &= \frac{\sum_{j=1}^{p} p\left(S_j|H_p\right) \int p(\underline{d}|S_j, w) p(w)\, \mathrm{d}w}{\sum_{k=1}^{q} p(S_k|H_d) \int p(\underline{d}|S_k, w)\, p(w)\, \mathrm{d}w} \end{aligned} \tag{7.1}$$

In all nonexcluded cases, at least one of the combinations $S_1, \ldots, S_p$ specified by H_p will also appear in the list of combinations specified by H_d. Within a combination, the same allele may appear in both genotypes, and hence a set S_i may contain two genotypes contributing to the observed peak at any given allelic position.

Under the binary model, the likelihood ratio LR_G is given by

$$LR_G = \frac{\sum_j p(S_j|H_p)}{\sum_k p(S_k|H_d)}$$

We characterize those circumstances where the use of LR_G is conservative, that is, when LR_G is less than LR.

Example 7.4 (*Two-person mixture where one combination under H_p fully accounts for the alleles in the stain*). We consider the special case when under H_p there are no peaks in the mixture that are not present in the genotypes of known contributors to the mixture profile. Thus, there is only one combination of genotypes, S_1, in the numerator and $p(S_1|H_p) = 1$. One of the combinations of genotypes under H_d will be S_1 or we must suspect an exclusion. Therefore,

$$LR = \frac{\int p(\underline{d}|S_1, w)p(w)\mathrm{d}w}{p(S_1|H_d)\int p(\underline{d}|S_1, w)p(w)\mathrm{d}w + \sum_{k=2}^{q} p(S_k|H_d)\int p(\underline{d}|S_k, w)p(w)\,\mathrm{d}w}$$

Rearranging, we have

$$LR = \frac{1}{p(S_1|H_d) + \sum_{k=2}^{q} p(S_k|H_d)\dfrac{\int p(\underline{d}|S_k, w)p(w)\,\mathrm{d}w}{\int p(d|S_1, w)p(w)\,\mathrm{d}w}}$$

Since we will use integrals of the form $\int p\ (\underline{d}|S_i, w)\ p(w)\ \mathrm{d}w$, often we will denote it by r_i. Hence

$$LR = \frac{1}{p(S_1|H_d) + \sum_{k=2}^{q} p(S_k|H_d)r_k/r_1}$$

In this situation, the binary LR is

$$LR_G = \frac{1}{p(S_1|H_d) + \sum_{k=2}^{q} p(S_k|H_d)}$$

If applied by hand or by programs such as PENDULUM, the binary approach would include all S_k in the denominator where r_k is large for at least some w.

This means that all genotype combinations are included in the denominator if there is a reasonable fit to peak area for at least some values of w.

We can therefore see that the binary approach will be conservative provided that $r_1 \geq r_k$ for all S_k and when all allelic combinations S_k with significantly positive values of r_k are included in the binary list. In other words, the binary approach will be an underestimate of LR whenever the set of alleles under H_p is the closest fit to the peak areas and when all reasonable alternative allele sets have been included under H_d.

Example 7.5 (*Two-person mixture where there is one combination under H_p and only one fully constrained unknown contributor*). Suppose that under H_p there are peaks in the mixture that are not present in the genotypes of known contributors, but only one unknown is required whose genotype is fully constrained. This would occur when, for example, the unknown was the major and can be unequivocally "called." Thus, there is again only one combination of genotypes, S_1, in the numerator. This combination of genotypes will appear under H_d. Therefore,

$$LR = \frac{p(S_1|H_p)r_1}{p(S_1|H_d)r_1 + \sum_{k=2}^{q} p(S_k|H_d)r_k}$$

$$= \frac{p(S_1|H_p)}{p(S_1|H_d) + \sum_{k=2}^{q} p(S_k|H_d)r_k/r_1}$$

In this situation, the binary LR is

$$LR_G = \frac{p(S_1|H_p)}{p(S_1|H_d) + \sum_{k=2}^{q} p(S_k|H_d)}$$

We can therefore see that the binary approach will again be conservative provided that $r_1 \geq r_k$ for all S_k, and all allelic combinations S_k with significantly positive values of r_k are included in the binary list.

Example 7.6 (*Two-person mixture with multiple possible combinations of genotypes under H_p*). This argument may also be attempted when there is more than one combination of possible alleles under H_p. We begin from Equation (7.1):

$$LR = \frac{\sum_{j=1}^{p} p(S_j|H_p)r_j}{\sum_{k=1}^{p} p(S_k|H_d)r_k} \text{ and } LR_G = \frac{\sum_{j=1}^{p} p(S_j|H_p)}{\sum_{k=1}^{q} p(S_k|H_d)}$$

Hence

$$\frac{LR}{LR_G} = \frac{\sum_{j=1}^{p} p(S_j|H_p) r_j \sum_{k=1}^{q} p(S_k|H_d)}{\sum_{k=1}^{q} p(S_k|H_d) r_k \sum_{j=1}^{p} p(S_j|H_p)}$$

If we assume that all the p prosecution combinations appear in the defense set,

$$\frac{LR}{LR_G} = \frac{\sum_{j=1}^{p} p(S_j|H_p) r_j \left(\sum_{k=1}^{p} p(S_k|H_d) + \sum_{k=p+1}^{q} p(S_k|H_d) \right)}{\left(\sum_{k=1}^{p} p(S_k|H_d) r_k + \sum_{k=p+1}^{q} p(S_k|H_d) r_k \right) \sum_{j=1}^{p} p(S_j|H_p)}$$

The value of this ratio varies considerably, depending on the values r_k. However, some comments can be made directly: If all r_i are equal or approximately so, then the ratio $LR/LR_G \approx 1$. This situation would only occur in rare instances, possibly when two contributors are present in very nearly equal proportions. If all r_i in the prosecution set are equal and greater than the remaining combinations, we write the prosecution r value as $r_{prosecution}$:

$$\frac{LR}{LR_G} = \frac{\sum_{k=1}^{p} p\,(S_k|H_d) + \sum_{k=p+1}^{q} p(S_k|H_d)}{\sum_{k=1}^{p} p(S_k|H_d) + \sum_{k=p+1}^{q} p(S_k|H_d)\; r_k/r_{prosecution}} > 1$$

Hence, use of the binary LR_G will be conservative. If one prosecution combination is clearly the best fit, then a possible strategy emerges. If the forensic scientist includes only this possibility under H_p but all reasonable possibilities under H_d, then the binary model is also likely to be conservative. Let S_1 be the single best fit. Therefore, r_1 is larger than all the other r's. We can state that

$$\sum_{j=1}^{p} p(S_j|H_p) r_j = p(S_1|H_p) r_1 + \sum_{j=2}^{p} p(S_j|H_p) r_j$$

Hence

$$\frac{LR}{LR_G} > \frac{r_1 \sum_{k=1}^{q} p(S_k|H_d)}{p(S_1|H_d) r_1 + \sum_{k=2}^{q} p(S_k|H_d) r_k}$$

$$> \frac{p(S_1|H_d) + \sum_{k=2}^{q} p(S_k|H_d)}{p(S_1|H_d) + \sum_{k=2}^{q} p(S_k|H_d) r_k/r_1} > 1$$

Example 7.7 (*Three contributors*). Here we have two independent mixture contributions w_1 and w_2.

$$LR = \frac{\sum_{j=1}^{p} p(S_j|H_p)\iint p(\underline{d}|S_j, w_1, w_2)\, p(w_1)\, p(w_2)\, \mathrm{d}w_1, \mathrm{d}w_2}{\sum_{k=1}^{q} p(S_k|H_d)\iint p(\underline{d}|S_k, w_1, w_2)\, p(w_1)\, p(w_2)\, \mathrm{d}w_1, \mathrm{d}w_2}$$

If we again write $r_i = \iint p(\underline{d}|S_i, w_1, w_2)\, p(w_1)p(w_2)\, \mathrm{d}w_1, \mathrm{d}w_2$, then all the considerations given above remain unchanged.

Is it possible that $r_k/r_1 \geq 1$ for some S_k? This will occur when there are some allele combinations under H_d that have a closer fit to the peak areas than the set specified under H_p. We have undertaken trials with simple least-squares fitting using artificially created mixtures of known genotypes. This is a fairly naive way to assess goodness of fit as it does not account for preferential amplification and stutter. However, these trials highlight the fact that the amount of template is crucial to whether the correct genotype combination appears as the closest fit or not. At 2 ng DNA we find that the correct genotypes are often the closest fit, but at less than 0.5 ng DNA the correct genotypes are often lower on the least-squares list. In casework, the amount of DNA template is unknown or estimated by quantitation. In many circumstances, the forensic scientist may be able to subjectively assess the amount of template from the quality of the electropherogram. More experimental work would be valuable to test this issue.

If the combination of allelic vectors specified by the prosecution is not the closest fit to the areas (strictly this should be "if the set of allelic vectors specified by the prosecution does not give the highest probability for the peak areas"), it is not possible to guarantee that the binary approach is conservative. It still may be a reasonable estimate as long as this ratio is not markedly larger than 1 for many allelic combinations. How would the analyst know that $r_k/r_1 \geq 1$ for some S_k? We need to assume that the judgement of the analysts is tuned sufficiently so that they can perceive which combination is the best fit and that they are able to assess this ratio in some intuitive way. The casework mixture should be assessed without reference to controls from the complainant and suspect. This means that the initial assessment is blind and this simple procedure eliminates the bias that could otherwise result. Consequently, if the closest fit to the peak areas is markedly different to that specified by H_p, then it is likely that the analyst would have noted this fact and may have had misgivings about interpreting the case at all, or may have asked for reanalysis. But what would happen, as is more likely, if the allelic vector specified under H_p is the best fit at a number of loci but is, say, the second best or lower fit at a few loci? Under these circumstances, the binary *LR* may or

may not be an overstatement of the evidential value. This will be very difficult to assess intuitively.

The issue seems partially alleviated by the presence of conditioning profiles. In some cases, the presence of conditioning profiles forces the choice of genotype combination under both H_p and H_d. Under such circumstances, there is no risk in utilizing the binary model.

It is important to assess the risk of continued use of the binary model and to consider the probable consequences. When considering these two approaches, it is necessary to remember that uncertainty arising in the assessment of peak area is only one of several sources of uncertainty in assessing DNA evidence. The other sources of uncertainty include, but are not restricted to, the population genetic model, sampling error, and the presence of close relatives as possible offenders. At this point, these other areas of uncertainty can be quantified to some extent using the binary approach. It is theoretically possible to include these factors in a more refined nonbinary or continuous approach. Implementation would however require further research.

We can identify a small number of relatively easily observed conditions under which we expect that judgement-based techniques will be reliable. These conditions typically include that the combination of alleles specified under H_p either is the single closest fit to the peak areas or, if multiple sets are included under H_p, then only the best fit is included in the numerator of the likelihood ratio, and all reasonable alternative allele combinations have been included under H_d. Dr. Champod has made the very reasonable suggestion that these considerations should be taught and tested in forensic scientists working on mixtures. The potential benefit arising from the simplicity of the binary approach should be offset against the greater objectivity and the expectation of greater accuracy in more comprehensive approaches that take full account of peak areas.

At this point we have no reason to believe that the continuous methods currently on offer have an advantage over the judgement of a well-trained forensic scientist. However, in the future this will change. Once reliable continuous methods become available, the binary method will have to be viewed as "second best" and will become obsolete.

7.3.4 Continuous Methods

Two alternative approaches to the binary model have been considered. Neither of these is in current use in any forensic laboratory. Both rely on modeling assumptions that would benefit from considerable refinement.

7.3.4.1 Normal Approximation-Based Methods

The first approach was proposed by Triggs and Buckleton[777] and Evett et al.[271] It was based on assumptions of normality, and was embodied in the program BETAMIX.[271] The Betamix approach could not assess the effect of stuttering

and heterozygote balance was modeled inappropriately. Foreman undertook considerable refinement of the program subsequent to publication. In summary it was a promising approach, but it was unproven whether it has any advantage over the judgement of the forensic scientist in its current embodiment. Such proof would require extensive investigation of the modeling assumptions and an assessment of the risk posed by the program not taking into account stuttering.

7.3.4.2 Monte Carlo and Monte Carlo Markov Chain (MCMC) Methods

The second approach based on Monte Carlo methods appears to have greater promise.[115] As yet, it is also untried. These ideas were subsequently taken up by Evett to produce a prototype: mixtures full Monte (MFM).[274] Again, this approach makes use of some very formative modeling assumptions. It is envisaged that simulation methods would have superior ability to deal with the complex interaction of preferential amplification, stuttering, other PCR artifacts, dropin, dropout, or sampling effects.

Some of the assumptions of the Monte Carlo and MCMC models include the following: (1) The preamplification mixture proportion is approximately constant across loci. (2) The peak area is proportional to the amount of DNA. (3) The area of shared peaks is the sum of the contribution of the two contributing individuals. (4) The contribution from stuttering can be added to the contribution from other sources.

These approaches require a model for heterozygote balance and stuttering that is in turn dependent upon the total amount of DNA present in the mixture. As before, we specify the set of prosecution and defense hypotheses and these in turn specify the possible combinations of contributing alleles. As before:

H_p specifies a number of combinations, S_1, …, S_p, of possible genotypes. Each combination, S_i, will contain m genotypes $\underline{G}_{i1}$, …,$\underline{G}_{im}$, similarly.

H_d specifies a number of combinations, q, of possible genotypes, each combination containing n genotypes $\underline{G}_{j1}$,…,$\underline{G}_{jn}$.[i]

We start from

$$LR = \frac{\sum_{j=1}^{p} p(S_j|H_p) \int p(\underline{d}|S_j, w)\, p(w)\, \mathrm{d}w}{\sum_{k=1}^{q} p(S_k|H_d) \int p(\underline{d}|S_k, w)\, p(w)\, \mathrm{d}w} \tag{7.1}$$

This equation contains the repeated assessment of an integral of the type

$$\Pr(S_i|H_x) \int p(d|S_i, w)\, p(w)\, \mathrm{d}w$$

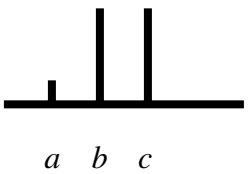
[i] In principle, the number of contributors for each combination can vary.

We focus here on the assessment of

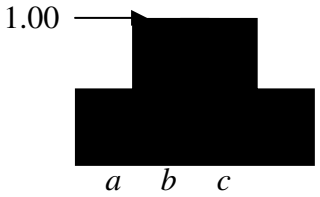
$$\int p(d|S_i, w, H_x)\, p(w)\, \mathrm{d}w$$

which is the critical term whether we intend to proceed by this method or any other. Since this integral is analytically intractable, we propose integration by numerical or simulation means.

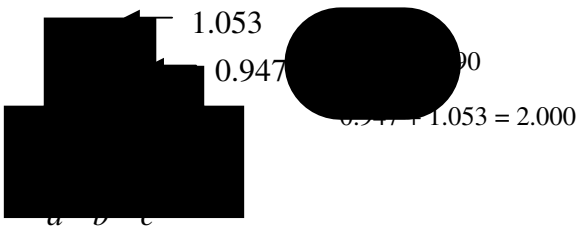
We begin by demonstrating the approach using diagrams. It is easiest to start with an unmixed, or single contributor, stain. Imagine that the epg looks something like:



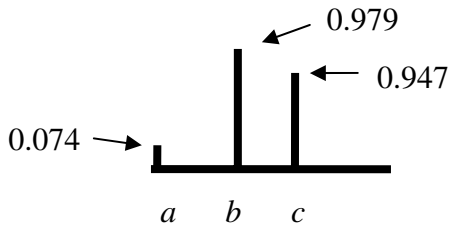
We start by assessing the probability of these data *if* the genotype is *bc*. Imagine a perfect amplification of a *bc* heterozygote. This would have two peaks at *b* and *c* and no stutter peaks at all. The two peaks would be exactly the same size. It is convenient to mark these peaks as having area 1.00 in arbitrary units.



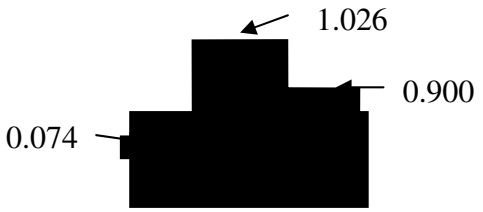
Next we sample a preferential amplification factor from an experimental distribution of preferential amplifications. In this case we suppose that the alleles are separated by 4 bp or one repeat unit so we might sample from the distribution for heterozygotes that are separated by one repeat. Although the area has been arbitrarily set to 1.00, we also account for the actual areas. Hence we use a distribution that allows for more variation in heterozygote balance if the peak areas are low. Let us say that this sampled factor is 0.90. We apply this factor to our perfect amplification.



Next we apply a stutter proportion to the *b* peak by selecting from the stutter distribution for the *b* allele. Let us imagine that we select 0.07. We apply this to the *b* allele. This suggests that the stutter peak should be $0.07 \times 1.053 = 0.074$ in area and the parent peak reduced to $1.053 - 0.074 = 0.979$ in area.

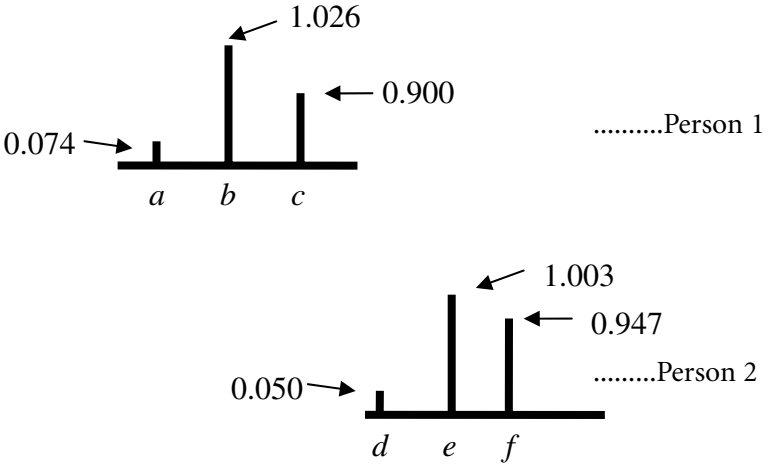


Next we apply a stutter proportion to the *c* peak by selecting from the stutter distribution for the *c* allele. Let us imagine that we select 0.05. We apply this to the *c* allele. This suggests that the stutter peak should be $0.05 \times 0.947 = 0.047$ in area. This area would add to the *b* allele, giving an area of $0.979 + 0.047 = 1.026$ for this peak. The parent *c* peak would be reduced to $0.947 - 0.047 = 0.900$ in area.

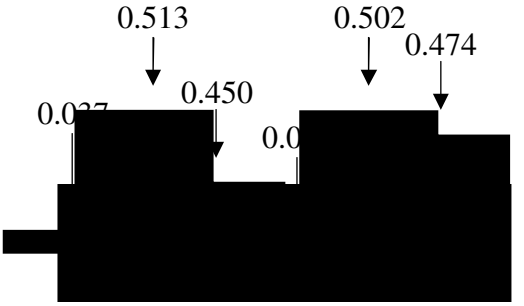


Last we score the closeness of this realization to the epg that we have actually observed. This scoring is problematic. At this point, let us assume that if the peak areas at the *a*, *b*, and *c* positions are "close enough," we would score this as a "yes." In the case discussed above, the simulated realization of the epg is different to the actual epg; hence we would score this as a "no." To complete the simulation, we repeat this many times and count how often we reproduce the epg under consideration. This is the estimated probability of the data given the genotype. This approach may appear unnecessary for single-contributor stains, but it does possess the potential to model allelic dropout.

Next we move from a consideration of a single-contributor stain to one where there are two (or more) contributors. Again we are merely sketching out the process — the formal mathematics will follow. Imagine that our simulation generates two single-contributor epgs.



Next we imagine that we combine these two profiles. Let us assume that we generate a mixture proportion of exactly $M_x = 0.50$.



Again we score whether this is close to the epg observed or not and count the fraction of these simulations that are close.

Since the mixing proportion (M_x) is not known, there is a need to consider multilocus genotypes explicitly. Hence we need to sample from the space of all multilocus genotypes. For each sampled allelic set (S_i), we envisage simulating heterozygote balance, stutter, and any other artifacts considered important. For instance, for low copy number (LCN) analysis, this could include "dropin." The M_x is either sampled from its prior distribution or numerically integrated. We assume that peak area is linearly proportional to the amount of DNA over the range to be modeled and that contributions from different sources combine in a simple additive manner.

Each simulation produces a datum at each allelic position at each locus. Since we seek $\int (p(d|S_i, w, H_x)p(w)\,\mathrm{d}w$ which is strictly a probability density function (PDF), we have two options. First, and probably easiest,[274] is to

substitute a proxy for this by scoring the simulated data vector against the actual data vector by distance (possibly Euclidean) in multidimensional space. This is allowable for the calculation of a likelihood ratio because the same proxy can be used in the numerator and denominator. If a proxy is used, a substantial filtering is made on the allelic space, and a coarse grid is used for the integration of the mixing proportion, it is possible that this integral may be assessed by Monte Carlo methods. For example, the allelic space could be confined to those peaks deemed to be allelic by a scientist, or to those peaks with area above background. This reduction in allelic space comes at the cost of reduced automation and the loss of the ability to model dropout. The prototype, MFM, was created based on a heterozygote balance model, an assumed stutter model, and a known mixture proportion with no stochasticity in M_x. This is clearly inadequate for casework purposes, but is an interesting proof of the original concept.

A second potential implementation of the approach involves solving this integral by MCMC methods.[115] The final output from the model will be the value of a likelihood function, or strictly the ratio of two likelihood functions conditional upon the observed data. To evaluate using MCMC methods, we construct samples from the full conditional densities of the parameters of the model, given the data, to form a Markov chain that converges to the full joint posterior density function of the parameters. Given the joint posterior density of the parameters, one may then evaluate probabilities relating to events of interest using the full joint posterior, or by integrating out nuisance parameters to obtain marginal posterior densities. Both approaches are amenable to searching databases and are patented by Buckleton and Pinchin.[108]

It is of interest to consider the performance of the Monte Carlo or MCMC method under conditions that have been described as preferential degradation. Typically, when DNA degrades, the high molecular weight fragments are the first to be affected. This results in a profile that is biased toward the expression of low molecular weight loci. This is most easily defined for two-person mixtures, but is applicable to n-person mixtures. We consider the situation that can be viewed as the mixing ratio for these two people varying in some manner according to locus. This is described as the preferential degradation of one component of the mixture (preferential in relation to the other component); however, the exact cause is immaterial as it is the observed phenomena *per se* of an apparently changing mixing ratio that we seek to model.

Consider a system of n loci ($L_1, \ldots, L_n$). We envisage that the phenomenon of preferential degradation can be modeled by the function $w_{Li} = m\chi_{Li} + c$, where w_{Li} is the mixing ratio at locus L_i, χ_{Li} is a function of the locus (possibly

molecular weight) that makes most preferential degradation curves approximately linear, and m and c are constants.

The purpose of creating such a model is to enable subsequent interpretation. We have proposed to evaluate the integral $\int p(d \mid S_i, w)p(w)\mathrm{d}w$ by numerical integration or an MCMC approach. The extension to preferential degradation involves rewriting this as

$$\int_c \int_m p(d|S_i, m, c)\ p(m|c)\ p(c)\ \mathrm{d}m\ \mathrm{d}c$$

If we can assume that m and c are independent, we can write

$$\int_c \int_m p(d|S_i, m, c)\ p(m)\ p(c)\ \mathrm{d}m\ \mathrm{d}c$$

which will require a prior on both c and m. It may be reasonable to assume $c \sim U[0, 1]$. For the prior distribution for m, it seems more likely that a distribution centered on 0 will be appropriate. This could be informed from experimentation. This approach is patented by Buckleton and Gill.[107]

7.4 Statistical Evaluation of Mixtures

To this point we have considered only the calculation of *LR* using the product rule model. However, it is possible to allow for population substructure. The addition of the subpopulation correction to mixtures is based on the application of the sampling formula introduced by Balding and Nichols.[36,39] It is useful to introduce the sampling formula for a simple single stain case first. Here we follow Harbison and Buckleton,[389] who were assisted in their development by David Balding.

7.4.1 The Sampling Formula

If x alleles are type a out of a total of n alleles sampled from the subpopulation, then the probability that the next allele observed will be of type a is given by the sampling formula

$$\frac{x\theta + (1-\theta)p_a}{1+(n-1)\theta} \tag{7.2}$$

where θ is the coancestry coefficient and p_a is the probability of allele a in the population. This is the basis for Equations (4.10) in the NRC II report and (4.20) in Evett and Weir.[267,585] For a stain at a scene that is typed as containing only the alleles a and b, matching a suspect of type ab, we require the

probability that the unknown true offender is of type *ab* given that the suspect is of type *ab*. We write this as Pr(*offender* is *ab*|*suspect* is *ab*). In such a case, the likelihood ratio

$$LR = \frac{1}{\Pr(\textit{offender}=ab|\textit{suspect}=ab)}$$
$$= \frac{\Pr(\textit{suspect}=ab)}{\Pr(\textit{offender}=ab \ \& \ \textit{suspect}=ab)}$$

which can be written as

$$LR = \frac{1}{\Pr(ab|ab)} = \frac{\Pr(ab)}{\Pr(ab \ \& \ ab)}$$

Application of the sampling formula follows from this last expression. Starting with the numerator, we consider each allele in turn (the order is unimportant). Take allele *a* first.
Application of the sampling formula gives

$$\frac{(1-\theta)p_a}{1-\theta} = p_a$$

($n = 0$ and $x = 0$, no alleles have previously been sampled and no alleles of type *a* have previously been seen). Next we take allele *b* ($n = 1$ and $x = 0$, one allele has been sampled but no *b* alleles have yet been seen), which gives

$$\frac{(1-\theta)p_b}{1} = (1-\theta)p_b$$

This gives the numerator as $2(1-\theta)p_a p_b$, a factor of 2 being included for each heterozygote.

A similar argument follows for the denominator. We have a factor of 4 because we are considering two heterozygotes. Turning to the alleles, we consider the first *a* ($n = 0$ and $x = 0$), the first *b* ($n = 1$ and $x = 0$), the second *a* ($n = 2$ and $x = 1$), and the second *b* ($n = 3$ and $x = 1$), giving the denominator as

$$\frac{4(1-\theta)p_a(1-\theta)p_b[\theta+(1-\theta)p_a][\theta+(1-\theta)p_b]}{(1-\theta)(1)(1+\theta)(1+2\theta)}$$

$$= \frac{4(1-\theta)p_a p_b[\theta+(1-\theta)p_a][\theta+(1-\theta)p_b]}{(1+\theta)(1+2\theta)}$$

Hence the likelihood ratio follows by dividing numerator by denominator:

$$LR = \frac{(1 + \theta)(1 + 2\theta)}{2[\theta + (1 - \theta)p_a][\theta + (1 - \theta)p_b]}$$

Most practising forensic scientists do not proceed in this manner. There are a number of shortcut rules that have been developed that allow the formulae to be written down directly.[98] These proceed from the conditional form of the probability. They can be seen to follow from the recursive formula given by Balding and Nichols.[39]

7.4.1.1 Shortcut Rules

The shortcut rules are demonstrated by way of examples given below. These are not really formal derivations but are just a set of rules that allow the answer to be written down. With practice, this becomes second nature. We begin by writing the probability in the conditional form. In front of the conditioning bar, we place the genotype(s) of the possible offender(s). Behind the bar we place the conditioning genotype(s). This should always include the suspect, but in some circumstances other profiles may also be included. This has become an area of some debate, which is covered in a short section later in the chapter.

Example 7.8 Calculation of Pr(*aa*|*aa*). Although our purpose is to demonstrate the application of this process to mixed stains, it is easiest to start with a simple example of a case where the stain at the scene is unmixed and shows the genotype *aa*. The suspect is *aa*. Hence we see that the only genotype for possible offenders is *aa* and the only potential conditioning profile is that of the suspect, also *aa*. Accordingly in this example, we consider the calculation of the conditional probability Pr(*aa*|*aa*) shown figuratively in Figure 7.6. The following three steps are required to obtain the formula: (1) Apply a factor of 2 if the possible offender is heterozygous. The possible offender will be the term in front of the conditioning bar. In this example, the possible offender

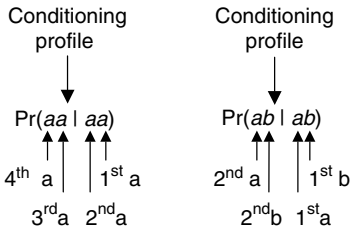


Figure 7.6 A diagrammatic representation to assist evaluation using the shortcut rules.

is the homozygote *aa*; therefore, no factor of 2 is required. (2) Counting from the back toward the front, label each allele as the first of this type seen, second of this type seen, and so on. Replace each of the possible offender's alleles with the terms given in Table 7.10. It is necessary to proceed from one or the other end of the offender's genotype. For instance, in the calculation of $\Pr(aa|aa)$ we see that the homozygote *aa* in front of the conditioning bar is treated as the 3rd and 4th *a* alleles. (3) Divide by a correction term based on the number of alleles in front of and behind the conditioning bar shown in Table 7.11.

This yields the familiar formula

$$\Pr(aa|aa) = \frac{(3\theta + (1-\theta)p_a)(2\theta + (1-\theta)p_a)}{(1+\theta)(1+2\theta)}$$

Example 7.9 Calculation of $\Pr(ab|ab)$. Consider the calculation of $\Pr(ab|ab)$ shown diagrammatically in Figure 7.6. Application of the rules leads quickly to the familiar formula

$$\Pr(ab|ab) = \frac{2(\theta + (1-\theta)p_a)(\theta + (1-\theta)p_b)}{(1+\theta)(1+2\theta)}$$

Example 7.10. As a more practical example, consider the following where the complainant (of race 1) has been genotyped as *ab*, the suspect (of race 2) has been genotyped as *cc*, and a semen-stained swab taken from the

Table 7.10 Conversion of Terms Using the Shortcut Rules

1st allele *a*	$(1-\theta)p_a$
2nd allele *a*	$\theta + (1-\theta)p_a$
3rd allele *a*	$2\theta + (1-\theta)p_a$
4th allele *a*	$3\theta + (1-\theta)p_a$
…	

Table 7.11 Correction Terms

2 alleles in front, 2 behind	$(1+\theta)(1+2\theta)$
2 in front, 4 behind	$(1+3\theta)(1+4\theta)$
2 in front, 6 behind	$(1+5\theta)(1+6\theta)$
4 in front, 2 behind	$(1+\theta)(1+2\theta)(1+3\theta)(1+4\theta)$
4 in front, 4 behind	$(1+3\theta)(1+4\theta)(1+5\theta)(1+6\theta)$
4 in front, 6 behind	$(1+5\theta)(1+6\theta)(1+7\theta)(1+8\theta)$
N in front, *M* behind	$[1+(M-1)\theta]\ldots[1+(N+M-3][1+(N+M-2)\theta]$

complainant after an alleged assault has been genotyped as *abc*. In the absence of any quantitative information, the genotype of the offender could be *ac*, *bc*, or *cc*.

Complainant	Race 1	Typed as *ab*
Suspect	Race 2	Typed as *cc*
Swab		Typed as *abc*

It is unreasonable to assume that the complainant and the suspect are from the same subpopulation if they are of different races. This assumption follows from a rigid application of a hierarchical population/subpopulation approach. However, subpopulations from different races could share alleles that are identical by descent (IBD) by recent admixture, in which case this simplification may not be valid. Following the arguments of Nichols and Balding,[36] the suspect and offender are assumed to be from the same subpopulation.

The likelihood ratio uses the probabilities of the offender's type conditional on the suspect's type (the complainant's type is ignored as having come from a different population):

$$LR = \frac{1}{\Pr(ac|cc) + \Pr(bc|cc) + \Pr(cc|cc)}$$

since

$$\Pr(ac|cc) = \frac{2(1-\theta)p_a[2\theta + (1-\theta)p_c]}{(1+\theta)(1+2\theta)}$$

$$\Pr(bc|cc) = \frac{2(1-\theta)p_b[2\theta + (1-\theta)p_c]}{(1+\theta)(1+2\theta)}$$

$$\Pr(cc|cc) = \frac{[3\theta + (1-\theta)p_c][2\theta + (1-\theta)p_c]}{(1+\theta)(1+2\theta)}$$

$$LR = \frac{(1+\theta)(1+2\theta)}{(2\theta + (1-\theta)p_c)(3\theta + (1-\theta)(2p_a + 2p_b + p_c))}$$

Substitution of $\theta = 0$ recovers the product rule formulae given in Table 7.1

$$LR = \frac{1}{p_c(2p_a + 2p_b + p_c)}$$

and provides a useful check.

7.4.1.2 When Should a Genotype Be Used in the Conditioning?[j]

The subpopulation model works best when those people who share the same subpopulation as the suspect are used in the conditioning. There are many complicating factors in this. These include the following: (1) The subpopulation of the suspect may be both undefinable and unknown. (2) The subpopulation of any other typed person may be both undefinable and unknown.

Clearly the suspect is a member of his/her own subpopulation whether or not we know that or can define it. But who else is? In many cases, this is unanswerable. The inclusion of additional genotypes in the conditioning if they are not members of the suspect's subpopulation essentially adds an unwelcome random element. Such an addition is not expected to improve the estimation process at all but rather adds variance about the true value. The addition of such people tends to give a more conservative *LR* when the person and the suspect share many alleles. It tends to give a less conservative *LR* when the person and the suspect share few or no alleles. It had been supposed that the addition of random persons was conservative on average. We are uncertain whether this is true, but even if true it applies on average over a number of cases rather than in each case. Accordingly, we consider that the addition of random genotypes to the conditioning set may make the *LR* more or less conservative but does not improve the process of obtaining the best estimate.

The effect of adding random genotypes is to randomize the answer.

As a first approximation, we suggest that only those persons known or reasonably assumed to share the subpopulation of the suspect should be added to the conditioning. This knowledge will very rarely be available in casework, and hence most often only the suspect's genotype will appear behind the conditioning.

If the forensic scientist wishes to report the more conservative estimate, we cannot think of anything better at this time than calculating the likelihood ratio both ways and reporting the smaller.

Example 7.11. Consider the following case, using the CTT system and not accounting for peak area:

Sample	TH01	TPOX	CSF1PO
Penile swab	6, 9.3, 10	8, 9, 10, 11	10, 11, 12
Complainant	6, 9.3	8, 10	10, 11
Suspect	9.3, 10	9, 11	10, 12

[j] This matter was brought to our attention by a senior caseworker in New Zealand, Sue Vintiner. It has been constructively discussed in meetings in New Zealand and in conversations with Robert Goetz, Manager of the Forensic Biology Laboratory of the Division of Analytical Laboratories, NSW, Australia.

H_p: The swab contains the suspect's and complainant's DNA.

H_d: The swab contains the suspect's DNA and the DNA of some other female.

If we know that complainant and suspect are from the same subpopulation, then the analysis follows as described below.

TH01: The possible genotypes of the second contributor to the crime stain, assuming that the suspect's own DNA is present, are: 6,9.3 or 6,10 or 6,6 leads to

$$LR = \frac{(1 + 3\theta)(1 + 4\theta)}{[\theta + (1 - \theta)p_6][8\theta + (1 - \theta)(2p_{9.3} + 2p_{10} + p_6)]}$$

TPOX: The second contributor, if not the complainant, must have type 8, 10:

$$LR = \frac{(1 + 3\theta)(1 + 4\theta)}{2(\theta + (1 - \theta)p_8)(\theta + (1 - \theta)p_{10})}$$

CSF1PO: The same rationale as for TH01 gives

$$LR = \frac{(1 + 3\theta)(1 + 4\theta)}{[\theta + (1 - \theta)p_{11}][8\theta + (1 - \theta)(2p_{10} + 2p_{12} + p_{11})]}$$

7.4.1.3 The General Formula

Curran et al.[216] and Evett and Weir[267] extended the general formula of Weir et al.[849] to allow calculation of the likelihood ratio in those cases where peak area or height data are ignored, but subpopulation effects are to be considered. This approach has been applied in Melbourne, Australia and is taught in lectures by some commentators in the U.S. It can also be used to calculate the likelihood ratio for those cases where peak area or height data are ignored and subpopulation effects are also ignored.

We outline the use of the general formula. Relevant nomenclature is listed in Table 7.12. Consider the crime profile. It has C_g distinct alleles. Since we are not considering area or height, each of these alleles may have originated from 1, 2, or more contributors and may have come from heterozygotes or homozygotes. Each hypothesis H_p and H_d declares some people to be contributors, some to be noncontributors, and may require some unknown contributors. For example, H_p is likely to declare that the suspect is a contributor whereas H_d is likely to declare him to be a noncontributor. Both H_p and H_d may declare the complainant to be a contributor. H_d is likely to have at least one unknown contributor, the "random man" who replaces the suspect as a

Table 7.12 Notation for Mixture Calculations

Alleles in the evidence sample

C	The set of alleles in the evidence profile, comparable to E in Table 7.4
C_g	The set of distinct alleles in the evidence profile
n_C	The known number of contributors to C
h_C	The unknown number of heterozygous contributors
c	The known number of distinct alleles in C_g
c_i	The unknown number of copies of allele A_i in C: $1 \le c_i \le 2n_c, \sum_{i=1}^{c} c_i = 2n_c$

Alleles from typed people that H declares to be contributors

T	The set of alleles carried by the declared contributors to C
T_g	The set of distinct alleles carried by the declared contributors
n_T	The known number of declared contributors to C
h_T	The known number of heterozygous declared contributors
t	The known number of distinct alleles in T_g carried by n_T declared contributors
t_i	The known number of copies of allele A_i in T: $1 \le t_i \le 2n_c, \sum_{i=1}^{c} t_i = 2n_T$

Alleles from unknown people that H declares to be contributors

U	The sets of alleles carried by the unknown contributors to C, also U in Table 7.4
x	The specified number of unknown contributors to C: $n_C = n_T + x$, also x in Table 7.4
$c - t$	The known number of alleles that are required to be in U
r	The known number of alleles in U that can be any allele in C_g, $r = 2x - (c - t)$
n_x	The number of different sets of alleles U, $\dfrac{(c + r - 1)!}{(c - 1)!r!}$
r_i	The unknown number of copies of A_i among the r unconstrained alleles in U, $0 \le r_i \le r, \sum_{i=1}^{c} r_i = r$
u_i	The unknown number of copies of A_i in U, $c_i = t_i + u_i, \sum_{i=1}^{c} u_i = 2x$ If A_i is in C_g but not in T_g: $u_i = r_i + 1$ If A_i is in C_g and also in T_g: $u_i = r_i$

Alleles from typed people that H declares to be noncontributors

V	The set of alleles carried by typed people declared not to be contributors to C
n_V	The known number of people declared not to be contributors to C
h_V	The known number of heterozygous declared noncontributors
v_i	The known number of copies of A_i in V: $\sum_i v_i = 2n_v$

Reproduced in amended form from Curran et al.[216] © 1999 ASTM International. Reprinted with permission.

contributor. However, both H_p and H_d may require unknown contributors in cases of, say, multiple rape when one suspect has been developed.

The alleles from people declared to be noncontributors may appear irrelevant; however, they do appear in the conditioning and affect the result if the subpopulation correction is applied.

Sometimes it will help to refer to the hypotheses H_p or H_d as H simply to avoid writing "H_p or H_d."

Each hypothesis therefore declares a total number of contributors to the stain, C. Let this be n_C, and hence the total number of alleles in the stain is $2n_C$. These $2n_C$ alleles may not be distinct since contributors may share alleles or may be homozygotes. Consider one particular allele A_i from the set C_g of alleles in the stain. There may be 1, 2, or more copies of this allele declared to be present. Let the number of copies of A_i be c_i. We can count the number of copies of A_i contributed by the declared contributors. However, there may be additional copies of A_i contributed from the unknown contributors.

At least one of the hypotheses and possibly both will require unknown contributors. The unknown contributors must carry those alleles in the crime stain that are unaccounted for by the declared contributors. Beyond this constraint, the remaining alleles of the unknown contributors may be any of the alleles in the set C_g (recall that we are not accounting for area or height). These unassigned alleles are free and may be assigned to any of the alleles in the set C_g. We need to permute these unassigned alleles over all possibilities by assigning each possible number r_i of them to each allele A_i and sum the result. We need to account for the possible ways of ordering the alleles into genotypes and keep track of the factors of 2 for the heterozygotes. Using the product rule and specifying x unknown contributors, this leads to

$$P_x(T, U, V|C_g) = \sum_{r_1=0}^{r} \sum_{r_2=0}^{r-r_1} \cdots \sum_{r_{c-1}=0}^{r-r_1-\ldots-r_{c-2}} \frac{2^{h_T+h_V} n_T!(2x)!n_v!}{\prod_i u_i!} \prod_{i=1}^{c} p_i^{t_i+u_i+v_i}$$

To introduce the subpopulation correction, we replace

$$\prod_{i=1}^{c} p_i^{t_i+u_i+v_i}$$

with

$$\frac{\Gamma(\gamma.)}{\Gamma(\gamma. + 2n_i + 2x + 2n_v)} \prod_{i=1}^{c} \frac{\Gamma(\gamma_i + t_i + u_i + v_i)}{\Gamma(\gamma_i)}$$

where

$$\gamma_i = \frac{(1-\theta)p_i}{\theta}$$

and

$$\gamma_{\cdot} = \sum_{i=1}^{c} \gamma_i$$

The likelihood ratio is evaluated by calculating this probability under H_p and H_d and dividing the two results. When this is done, we note that the number of typed individuals and the alleles that they carry are the same under each hypothesis. Hence the term $2^{h_r + h_v}$ will cancel as will portions of the terms in the products.

7.4.1.4 Mixtures and Relatives

The possibility exists that a forensic scientist may be asked to consider hypotheses where a relative of the accused is implicated. This is the logical extension of the subpopulation considerations given above and can be accommodated by a similar approach. All that is required is that the conditional probabilities given above be assessed using the relatives formulae (see Chapter 4) when the possible offender and a conditioning profile are hypothesized to be related. Once again, we will demonstrate our approach by a worked example.

Example 7.12 Incorporating relatedness. We demonstrate the incorporation of relatedness in this example. The complainant (of race 1) has been genotyped as *ab*, the suspect (of race 2) has been genotyped as *bc*, and a semen-stained swab taken from the complainant after an alleged assault has been genotyped as *abc*. Possible offenders are considered to be *ac*, *bc*, and *cc*. We consider the following hypotheses:

H_p: The suspect is the donor of the stain.

H_d: A brother of the suspect is the donor of the stain.

We see that we will need the probabilities of the untyped brother being *ac*, *bc*, or *cc* given that the typed brother is *bc*. These can be obtained from Table 4.5.

Conditional Probability Required	Brother with Subpopulation Correction
$\Pr(ac\|bc)$	$\frac{(1-\theta)p_a}{4(1+\theta)}\left(1 + \frac{2(\theta+(1-\theta)p_c)}{(1+2\theta)}\right)$
$\Pr(bc\|bc)$	$\frac{1}{4}\left(1 + \frac{2\theta+(1-\theta)(p_b+p_c)}{(1+\theta)} + \frac{2(\theta+(1-\theta)p_b)(\theta+(1-\theta)p_c)}{(1+\theta)(1+2\theta)}\right)$
$\Pr(cc\|bc)$	$\frac{\theta+(1-\theta)p_c}{4(1+\theta)}\left(1 + \frac{2\theta+(1-\theta)p_c}{1+2\theta}\right)$

The likelihood ratio is 1 divided by the sum of these probabilities. For $p_a = p_b = p_c = 0.10$ and $\theta = 0.02$, the likelihood ratio is 2.5. This compares with a likelihood ratio of 16.7 if the possible offenders had been treated as unrelated members of the suspect's subpopulation.

7.4.2 The Two-Trace Problem and the Factor of 2

A two-trace transfer problem where two stains of different types were left at the scene of a crime was published by Evett.[259] In this example, the investigator is satisfied that both of the stains were left during the commission of the offense. A single suspect is produced who matches one of the stains. If the second stain is explained in both H_p and H_d as, say, having come from the complainant, then there is no further issue. Here we deal only with the situation where the second stain is unexplained. The likelihood ratio is $LR = 1/2f_X$ (where f_X is the probability of the matching stain in the appropriate population) for the following pair of propositions:

H_p: The suspect was one of the two men who left the crime stains.

H_d: Two unknown men left the crime stains.

Aitken[8] and Evett and Weir[267] give the same result.

The central conclusion of these publications is to reduce the weight of evidence represented by the likelihood ratio by a factor of 2. This procedure has not been universally adopted. In addition, it leads to the reasonable question: In the interests of the defense, should an investigator look for a second stain after finding a matching stain?[177]

Triggs and Buckleton,[782] Dawid,[220] and Meester and Sjerps[543] correct an error in the derivation published by Evett, and show that the key question is not how many traces were transferred or found on examination, but how many perpetrators took part in the commission of the offense.

If this finding is taken in conjunction with that of Stoney,[736] it can be seen that both the relevance and the number of traces at the scene as well as the number of offenders are important. These factors may be unknown or difficult to evaluate in an actual casework situation.

7.4.2.1 The Issue of Relevance and the Factor of 2

Intuitively we would expect that a second, but irrelevant, stain would not require the down-weighting of the likelihood ratio by a factor of 2. However, the issue of relevance is quite a difficult one to define. If, post hoc, all stains that do not match are defined as irrelevant, then, indeed, the factor of 2 is unnecessary. However, this is inappropriate.

We need to consider the meaning of the word "relevance." The most suitable definition is: A relevant stain is one that was transferred during the offense. The state of "relevance" or "irrelevance" is not known in all cases. In fact, there is often considerable uncertainty. It may be possible to make a good estimate of which stains are recent or likely to have been transferred during a crime from a scene examination. Equally it may be possible to determine with reasonable confidence that a stain is old and thus probably irrelevant. For this analysis, the key consideration appears to be the probability that may be assigned to the event that the stain was transferred during the offense. This is essentially the definition used by Stoney[736] when he considered a one- or two-trace problem with one offender.

7.4.2.2 Propositions

The analysis of any multiple trace problem (including mixtures) appears to be critically interconnected with the formulation of the propositions. In the case of multiple trace problems, it is critically important to determine what, if any, background information may reliably be used in the interpretation of the evidence. Suppose that the evidence consists of a garment recovered from a complainant on which two stains (1 and 2) are found. Neither stain matches the complainant. Let us assume that we are considering a case of a violent attack against a woman who is able to say that two men attacked her. The presence of two attackers turns out to be crucial to later stages of our argument, and loosening this assumption will be discussed.

Given this information, we can create at least three different pairs of propositions.

1. H_p: The suspect is one of the two men who assaulted the woman.

   H_d: Two other men assaulted the woman.

2. H_p: The suspect is one of the two men who were in contact with the woman during the offense.

   H_d: Two other men were in contact with the woman during the offense.

3. H_p: Stain 1 came from the suspect.

   H_d: Stain 1 came from someone else.

Use of the third set of propositions leads directly to a likelihood ratio $LR_C = 1/f_X$, where f_X is the probability of the matching stain. This is equivalent to the standard expression for a single stain.

However, it could be correctly argued that we have avoided evaluating the nonmatching evidence in a selective way. In fact, we would need to form our

propositions after the typing of the stains, as before that we would not know which stain matched. Does using the second set of propositions alleviate this problem? For this set of propositions, the likelihood ratio was given by Triggs and Buckleton:[k,782]

$$LR_B \approx \frac{1}{2f_X} \frac{P_0 t_1 t_2 (1-2q) + P_1^2 t_1 (1-t_2)(1+f_X) + P_1^1 (1-t_1) t_2 f_X + 2P_2^{1,2} (1-t_1)(1-t_2) f_X}{P_0 t_1 t_2 (1-2q) + P_1^2 t_1 (1-t_2) + P_1^1 (1-t_1) t_2 + P_2^{1,2} (1-t_1)(1-t_2)}$$

where q is the probability that both stains were transferred from perpetrator 1.

Triggs and Buckleton consider some special cases, when the transfer probabilities, t_i, either tend to 1 or 0. In other words, transfer of the stain i during commission of the offense is almost certain or is extremely unlikely. If t_1 tends toward 1, stain 1 becomes more relevant, while if t_2 tends toward 0, stain 2 becomes less relevant.

$t_1 \to 1$ $t_2 \to 1$	$t_1 \to 1$ $t_2 \to 0$	$t_1 \to 0$ $t_2 \to 1$	$t_2 \to 0$ $t_2 \to 0$
$LR_B \to \frac{1}{2f_X}$	$LR_B \to \frac{1+f_X}{2f_X}$	$LR_B \to \frac{1}{2}$	$LR_B \to 1$

The result in the first column shows that the standard treatment of the two-trace problem is a special case of a more general situation. When both hypotheses H_p and H_d postulate two perpetrators, and when the evidence consists of stains of two distinct genotypes both assumed to have been definitely transferred during the offense, the likelihood ratio takes the form published initially by Evett. If both stains are relevant, then the weight of evidence due to the matching stain should be down-weighted by a factor of 2.

The result in the fourth column is, to some extent, self-verifying. If both stains are extremely unlikely to have been transferred during the commission of the crime, the evidence is almost completely irrelevant (whether they match or not), and the likelihood ratio should indeed be close to 1. This result was previously given by Stoney.[736]

In the second and third columns, we have the situation where only the first and second stains respectively are relevant. The second column considers specifically the case where we assume that there were two offenders, but that the second stain was not transferred during commission of the crime. Thus, any physical, chemical, or genetic properties of the second stain are

[k] Here we must consider the transfer probabilities, t_i, and the four probabilities P_0, P_1^1, P_1^2, and $P_2^{1,2}$, P_1^1, for example, is the probability that one stain of type 1 is already present on the garment, and $P_2^{1,2}$ is the probability that two stains, one of each of types 1 and 2, are already present on the garment.

irrelevant in weighing the evidence against the suspect. We see that the likelihood ratio in this case is similar to the situation in column one, where the second stain was transferred during the commission of the crime.

The logic that leads to the results in the first three columns has led us to believe that the key consideration is that there are two offenders, and not that there were two stains. In fact, this result generalizes easily to n offenders. This observation also allows us to relax the constraint that we have been told how many offenders there were. We could allow for uncertainty in the number of offenders by simply modeling a probability distribution on the number of offenders.

These results also answer the question: Should the investigator continue looking for stains that may not match, as this would benefit the defense? We see that the examination of irrelevant or very low relevance stains has no effect on the likelihood ratio (whether they match or not). This allows us to begin consideration of mixed stains where both components are of potential evidential interest. Whittaker[856] noted that clear major/minor stains, if they can be unambiguously assigned as clear major/minors, can effectively be treated as if they were two distinct stains and the considerations given above apply. Therefore, in summary:

For the pair of hypotheses:

H_p: The suspect is one of the two persons in the mixture, and

H_d: Two random persons are in the mixture.

If the suspect matches the major stain, then

$$LR = \frac{1}{2\Pr(\textit{Major stain})}$$

If the suspect matches the minor stain, then

$$LR = \frac{1}{2\Pr(\textit{Minor stain})}$$

For two different hypotheses:

H_p: The suspect is the donor of the major stain (or the minor), and

H_d: Another person is the donor of the major stain.

$$LR = \frac{1}{\Pr(\textit{Major stain})}$$

However, we do feel that this set of hypotheses is abdicating a large part of the responsibility for interpretation to the jury as they need to know to add the factor of 2.

If amelogenin, or any other locus, can be used to give a clear assignation of major and minor and this locus is not used to calculate the frequency, then the need for the factor of 2 is avoided. We believe, however, that the prevalence of clear major/minor stains is much exaggerated. In reality, we believe that many forensic scientists take an (admirably) conservative stance by "allowing" a large number of combinations at each locus and between loci. By doing this, however, they do not treat the mixture as a clear major/minor. In such cases, the question of whether there should be a factor of 2 or not is very complex.

7.4.2.3 Locus-by-Locus Approach[1]

The following discussion applies to an analysis that proceeds locus by locus. By this we mean that genotypes are assigned as belonging to person 1 or person 2 on a locus-by-locus basis. To proceed with this discussion, we introduce the concept of a reciprocal combination. Hence the reciprocal of *ab*:*cd* is *cd*:*ab*. The summary of the situation when proceeding on such a basis is:

1. If every combination at each locus has its reciprocal combination, then the factor of 2 is already accounted for and any further factors are unnecessary.
2. If no combinations have their reciprocal, then a factor of 2 is required once, either at one locus, or once across the whole likelihood ratio.
3. If some combinations have their reciprocal and others do not, then the situation is in between. The likelihood ratio is very hard to get correct (with regard to the factor of 2) in such situations.

Several options exist:

1. Either the factor of 2 can be added on the assumption that it is conservative.
2. Or combinations can be inserted to restore "symmetry", by which we mean that each combination has its reciprocal.
3. Or the locus-by-locus approach can be dropped in favor of a "full-genotype" approach.

7.4.2.4 The Full-Genotype Approach

The nature of the assumption that M_x is approximately constant across loci makes it advantageous to determine a set of possible multilocus genotypes for the major, each paired with a multilocus genotype for the minor. This

[1] This section owes a lot to discussions with Dr. Christophe Champod during our time at the FSS in Birmingham.

approach is virtually impossible to implement by hand (except for clear major/minors), but could be attempted by a computer program. In such cases, only those combinations (of person 1 and person 2) that do not have their reciprocal (person 2 and person 1) somewhere in the list of possibilities require the factor of 2.

7.4.2.5 Unknown Number of Contributors and Ethnicity

There are often a number of unknowns in mixture analysis. These include the number of contributors and the ethnicity of any unknown contributors. This can lead to a large number of potential combinations. The approach of presenting all of these in a table, as undertaken in the O.J. Simpson case by Weir and Buckleton, has been justifiably criticized as excessive.[767,838] Brenner et al.[85] suggest that the maximum value be found for the denominator and this likelihood ratio be reported. They point out that this maximum for the denominator will usually correspond to the minimum number of persons sufficient to explain all the bands. This result is extended by Lauritzen and Mortera,[489] who show how to calculate this maximum in generality.

Buckleton et al.[112] attempted to deal with the issue of the multiplicity of ethnic origins for unknown persons. A bound can be calculated in those circumstances where there is either no prior knowledge of the ethnicity of the unknowns or these priors are equal under H_p and H_d.

Low Copy Number

8

JOHN BUCKLETON AND PETER GILL

Contents

This chapter deals with the forensic analysis of trace material. This is often achieved by an increase in the number of PCR amplification cycles, although modern multiplexes appear to have much improved sensitivity even at the normal 28 cycles. We will discuss changes in profile morphology that occur when there are very few starting templates, and the issue of contamination. In addition, we will consider the formative ideas that have been offered on interpretation of profiles that have been or may have been affected by the morphological changes and contamination risks associated with ultra-trace work.

8.1 Introduction

The analysis of such trace evidence was anticipated in 1988 by Jeffreys et al.[436] Subsequently, the successful typing of single cells was first reported in the mid-1990s.[292–294,296–298,720,795] The typing of DNA from fingerprints,[794,795,797] touched paper,[46] latex gloves,[620] debris from fingernails,[391,862]epithelial cells,[859,860] single hairs,[406,791] clothing,[89,698] cigarettes,[821] and biological material on strangulation tools[861] has been reported. Full or partial profiles of transferred

DNA were obtained up to 10 days after simulated strangulation.[681] The author of the simulated strangulation work also noted the presence of DNA from a third party in some experiments. Sutherland et al.[744] report the typing of fingerprints after development with fingerprint powder, but warn, wisely, that the brush and powder may be a source of carryover from previous brushings since the brush is usually placed back into the powder multiple times when brushing previous prints. Two strategies to increase sensitivity have been employed: increasing the number of cycles[294] or by nested PCR.[738,739,751] To this we should add the use of reduced reaction volumes.[322,497] Reducing reaction volumes can be utilized to save on reagents or to preserve sample for subsequent reanalysis. However, for the purpose of this chapter, the significant finding is that reducing the reaction volume but maintaining the template DNA quantity leads to increased signal-to-noise ratio and hence enhanced sensitivity.

The approach to using enhanced sensitivity in DNA analysis originated largely in Australia and has led to a field of forensic work that was subsequently termed low copy number (LCN) in the U.K. It was launched into casework in January 1999 in the U.K. LCN broadens the potential impact of forensic science, but also brings with it additional challenges with regard to interpretation issues. These issues include substantial changes to profile morphology, such as extreme peak imbalance, and the almost unavoidable appearance of contaminant alleles. Standard methods for interpretation appear inadequate for some LCN work, and new methods and safeguards are desirable.[132,338] In addition, the presentation of this evidence requires extreme care. For example, it should be made explicit to the court that LCN work cannot be viewed in the same way as standard DNA work, and that issues of transfer and relevance, which are always pertinent, come into even greater prominence.[523,858] We will discuss these issues and suggest ways of coping with them.

Increasing the number of PCR amplification cycles allows the amplification of less than 100 pg of total genomic DNA.[294,348] It is expected that 3.5 pg contains one haploid DNA template; therefore, 100 pg should equate to approximately 30 templates. This increase in sensitivity facilitates the examination of a whole new range of evidence types that previously could not be analyzed because of the very low amounts of DNA recoverable from the sample. Variables affecting the recovery of DNA were discussed by Abaz et al.[1]

The amount of DNA deposited by contact between a donor and an object appears to depend on the donor. Some people are observed to be "good shedders," which means that their DNA is easily transferred to objects simply by touching them. There appears to be day-to-day variability in this property. Others people are "poor shedders" who deposit little DNA, with a range of propensities in between.[482,523,795,796,860] (for a different result, see

Balogh et al.,[46] who found no difference for their four subjects). The reason for these differences between people are, as yet, unknown but are not thought to be sex dependent.

Secondary transfer has also been observed. In such a case, DNA was seen to transfer from a good shedder to the hand of a poor shedder and then subsequently onto a touched object, usually as part of a mixed profile including both the good and poor shedder (see Lowe et al.[524] for an excellent review). This finding was not replicated when Ladd et al.[482] repeated the work.

Balogh et al.[46] found that the time of day that a person touched paper or whether or not they had finished playing sport did not affect the amount of DNA detected. They also confirm van Oorschot and Jones'[795] observation that the strongest signal in a mixture is not necessarily from the last person to touch the paper.

Along with the increased sensitivity has come a greater need for strict efforts to reduce contamination both at the scene, in the consumables, and in the laboratory (again, Balogh et al.[46] have dissenting results finding no contamination issue in their study). Rutty et al.[683] discuss the use of protective clothing and masks at scenes, both of which reduced, but did not eliminate, the possibility of transfer from the scene officers. Talking and coughing increased transfer; hence, it is advisable to wear face masks and full protective clothing in the laboratory. They discussed the utility of a "no talk" policy in the laboratory. Negative controls may have to be introduced from the scene onwards.

Increasing the number of amplification cycles has a very significant effect on profile morphology when very few templates are amplified. Peaks from a heterozygote may be very imbalanced, with one or the other of the alleles giving a much larger peak. This may be sufficiently extreme that the heterozygote appears to be a homozygote — usually this is termed "dropout." An increased occurrence of stutters, artifacts, and preferential amplification was observed that significantly reduced the quality of the DNA profile. In addition, there are increased risks of laboratory-based or consumable-based contamination. To reduce this risk, preparation and extraction of samples must be carried out under stringent conditions of cleanliness and consumables need to be monitored.

Such changes to profile morphology may be due to sampling effects in the template, stochastic effects in the PCR reaction, or other as yet unidentified phenomena. For example, in a sample of 20 sperm it is quite unlikely that the two chromosomes are present in exactly the balanced 10:10 ratio. Equally, an early PCR "event" such as a stutter in the first cycle when there are very few templates might theoretically have a large effect on subsequent peak areas (although in practice, stutters are not excessively common at such low levels of DNA).

The effect of varying the number of amplification cycles has been investigated.[348] A compromise was sought where sensitivity was maximized while the deleterious effects on profile morphology were tolerable.

Heterozygote imbalance increased as the amount of DNA template was reduced and as the number of amplification cycles was increased. The number and size of stutter peaks increased at all loci as the cycle number increased and as the amount of template DNA decreased. When samples were overamplified, this markedly increased the number and size of stutters observed. At 28 cycles and DNA concentrations below 100 pg, neither SGM nor SGM$^+$ produced alleles. However, at 34 cycles, full profiles could be obtained down to approximately 25–50 pg — the equivalent of 7–14 templates. Below this level of DNA template, allelic dropout occurred. This may be because at these levels a full genetic complement was not present in the available template.

When less than 100 pg of DNA was analyzed, there was no advantage in using more than 34 cycles because the presence of artifacts increased and the profile became increasingly imbalanced. Hence when the amount of DNA was less than 100 pg, the subjectively assessed best compromise (following Whitaker et al.[854]) was obtained using 34 amplification cycles. It is certainly conceivable that this compromise would be different for different multiplexes or may change in the future as systems evolve.

There was some evidence to suggest that more alleles could be observed using nested singleplex primers.[738,739] This method utilizes two sets of primers in two separate PCR reactions. In the first reaction, the STR and adjacent flanking regions are amplified. The primers used in the second round are designed to amplify a smaller product using an aliquot from the first round of PCR as the DNA template. Nested PCR reduced the amount of nonspecific or artifactual products. This process has been demonstrated to analyze the contents of a single cell,[720] but suffers from the disadvantage of necessitating transfer of PCR product into a separate tube, thereby increasing contamination risks.

8.2 Changes in LCN Profile Morphology

We briefly discuss some of the differences in profile morphology that may be expected in LCN work. Strictly, LCN refers to low copy number and may be the situation when employing either 28 or 34 cycles. We largely follow Whitaker et al.[854] or Taberlet et al.[751]

8.2.1 Heterozygote Balance

The tendency for the low molecular weight allele to amplify preferentially was accentuated in 34 cycle amplifications (dissenting data appear in

Table 8.1 Median Heterozygote Balance $Hb = \phi_{HMW}/\phi_{LMW}$ and IQR for the Various Loci of the SGM$^+$ System Following Whitaker et al[854]

	34 cycles				28 cycles			
	Peak Area < 10,000		> 10,000		< 10,000		> 10,000	
	Median	IQR	Median	IQR	Median	IQR	Median	IQR
Amel	0.96	0.60	0.85	0.37	1.05	0.20	1.02	0.09
THO1	0.95	0.56	1.56	n/a	1.00	0.15	1.00	0.13
D21	1.22	0.68	0.99	0.61	0.95	0.17	0.96	0.10
D18	0.74	0.71	0.88	0.66	0.94	0.22	0.94	0.12
D8	0.94	0.61	0.95	0.45	0.93	0.15	0.96	0.08
VWA	0.94	0.63	0.87	0.62	0.91	0.14	0.93	0.09
FGA	0.93	0.79	0.85	n/a	0.94	0.16	0.95	0.10
D2	0.88	0.77	0.93	0.65	0.90	0.23	0.93	0.15
D3	0.97	1.08	1.00	0.55	0.90	0.14	0.93	0.09
D19	0.94	0.95	0.91	0.31	0.95	0.15	0.93	0.13
D16	1.01	1.65	1.06	0.67	0.96	0.20	0.96	0.11
Mean	0.95	0.82	0.99	0.54	0.95	0.17	0.96	0.11

Findlay et al.[297]). An example of a heterozygote displaying peak asymmetry is shown in Figure 1.6 (Chapter 1).

At 34 cycles the mean heterozygote balance, *Hb*, was, as expected, typically slightly less than one; however, individual samples varied greatly (see, for example, the larger interquartile ranges (IQRs)[a] in Table 8.1). It was concluded that peak area was significantly less useful or even not at all useful to form inferences when using 34 cycles with poor templates. For the ABI 3700 system (Applied Biosystems, Foster City, CA), as the peak area of the larger peak decreased from greater than 10,000 rfu to less than 5000 rfu, the heterozygous imbalance of individual samples tended to vary from 1 more frequently. Once the peak area of the larger peak was less than 5000 rfu, *Hb* varied to such a degree that it was concluded that peak area was virtually uninformative.

8.2.2 Allelic Dropout

Findlay et al.[295,297] suggested that allelic dropout is a separate phenomenon to heterozygote balance, not simply an extreme form of it. Their evidence stems from the observation that heterozygote balance shows a distribution terminating at nonextreme values for the fraction of product attributable to one allele. They suggest that allelic dropout is caused by several phenomena that

[a] IQR is the difference between the first and third quartile of the distribution. It is a measure of the spread of the distribution often used in preference to the standard deviation for skewed distributions.

are not mutually exclusive. When a PCR amplification is prepared, only a portion of the extract is analyzed; hence, some template alleles may be absent from the reaction. Findlay et al.[295] note that failure to transfer the cell, degradation or loss of the target sequence, and/or problems associated with the PCR could be possible causes of allelic dropout. This list appears almost exhaustive to us. Whitaker et al.[858] report that allelic dropout was not observed when the area of the larger peak was greater than 10,000 rfu. However, at lower peak areas it became increasingly prevalent.

8.2.3 Stutter

After 28 cycles, most stutter ratios,[b] when a stutter was observed at all, ranged from 0.05 to 0.10 with outliers ranging up to 0.15. For LCN templates, the mean stutter ratio remained in the 0.05–0.10 range; however, there was a longer right tail to the distribution. Stutter ratios up to 0.20 were recorded where the peak area of the associated allele was greater than 10,000 rfu and 0.40 when the associated allele was less than 10,000 rfu in area. We note that the apparent outliers may not be stutters at all, but could be attributed to somatic mutation. No observations of a stutter peak in the absence of the associated allele peak were recorded in this data set. However Whitaker et al.[854] accept that theoretically this could occur.

Stuttering is thought to be a simple branching process where at each cycle a template can do one of three things: replicate correctly, stutter, or make no product.

This basic model explains PCR behavior at least to a coarse approximation. A stutter early in the PCR process would result in proportionally more of the final DNA product population being affected. If we accept that stuttering is an inherently stochastic process, then fewer starting templates would lead to a higher variance in product peak areas and consequently in heterozygote balance and stutter ratios. However, in practice, LCN stutters appear to be uncommon.

8.2.4 Spurious Alleles

The word "contamination" is a bit of a "catch all." Clearly DNA that "got into" the sample in the laboratory would be described, correctly, as contamination. However, DNA could also be added by persons at the scene, such as attending officers, ambulance personnel, or forensic scene investigators. This DNA could be their own, in which case it is fairly innocuous. In principle, the profiles of every investigator or medical attendant could be checked. What must

[b] Stutter ratio $S_R = \phi_S/\phi_A$ and stutter proportion $S_x = \phi_S/(\phi_A + \phi_S)$, where ϕ_S is the area of the stutter peak and ϕ_A is the area of the allelic peak.

be guarded against is any possibility of transfer of DNA between crime scenes, or between items associated with the suspect and the crime scene. It is also possible that DNA may have been present on items before the crime occurred. This DNA is not strictly contamination but could more properly be called "background." We concentrate in this section on laboratory-based contamination.

Spurious alleles were observed in profiles even when DNA extractions were carried out in facilities designed to minimize the chance of laboratory contamination.[348] Contaminants were more often associated with the low molecular weight loci.

Negative controls are designed to detect contamination within the reagents (such as distilled water) that are used for DNA extraction and PCR. Historically, we have assumed that if contamination of a reagent occurs, then the same spurious alleles should be observed in every sample processed. However, within the context of LCN it appears that the negative control does not act as an indicator of minor contamination within associated samples of the same batch. This is because the method is sensitive enough to detect a single contaminant molecule of DNA. Single events such as this only affect one tube. Contaminants may be tube-specific, and transfer could occur via minute dust particles, plasticware, or other unknown processes. Hence the negative control cannot operate in the traditional sense. This has led to the "drop-in" concept that envisages single alleles from degraded or fragmented DNA that can randomly affect tubes, whether they contain casework samples or are negative controls. The consequence of this is that casework samples could be affected by laboratory-based contaminants that do not appear in the negative control and vice versa.

Nevertheless, the negative controls serve an important function as a "health check" of the process. They indicate the rate of appearance of spurious alleles. This rate needs to be kept to a minimum. In the context of LCN, replication of extraction negative controls is recommended in order to determine if laboratory-based contaminants are reproducibly amplified. If not, we suggest that there is no *a priori* reason to suppose that any alleles observed in the negatives have affected the associated extracted samples.

When working with LCN templates, it is necessary to accept that it is not possible to avoid laboratory-based contamination completely. This issue is ameliorated by replication. In LCN casework, it is typically not possible to carry out more than three separate tests of a DNA extract because of the limited size of the initial sample. If the sample were more generous (i.e., greater than 100 pg of DNA), then an LCN approach would not be needed. However, if the contaminant events are truly random single events, then the chance of a contaminant appearing in each of two replicates is small. The risk that spurious alleles could be duplicated and reported in the consensus results has been

estimated by pairwise comparisons of samples.[348] In this study, four double-contaminant events out of 1,225 comparisons were observed. Similarly, it is useful to compare profiles against operator controls to guard against the possibility of gross contamination of a sample from an operator.

Table 8.2 gives an example of negative control data. We see that many controls do have an allele or alleles present. In order to determine the impact of contaminants on the evidential value, it is important that we understand the mechanism by which these alleles appear in the sample. This area warrants further study.

We have initially hypothesized that "alleles" appear singly using the "drop in" model and that these "fall" into samples independently of each other. We term this the "alleles snowing from the ceiling" model. If this model is correct, then the number of alleles in the negative controls should follow a Poisson distribution. Table 8.2 illustrates the fit of a Poisson distribution with parameter equal to the sample mean. Subjectively the fit is poor. We focus on the rows from five alleles onwards. There is a small (in absolute terms) excess of observed over expectation. This suggests that two processes are present. It seems necessary to treat the occurrence of multiallele profiles as a separate phenomenon to the "drop in" model. We therefore initially proposed to treat contamination as two different phenomena: first as "drop in" and second as more complete partial profiles (i.e., multiple bands from a single source).

The examination given in the examples to follow suggests that the biological model is remarkably robust to the first type of contamination with the proviso that a few situations are taken into consideration. There is probably no statistical approach that can effectively militate against the second type of contamination. This therefore places a strong emphasis on reducing

Table 8.2 Comparison of the Observed Number of Contaminant Alleles in a Sample with Predictions from a Poisson Model[106]

Number of Contaminant Alleles	Observed Probability of This Number of Contaminant Alleles	Expected Probability of This Number of Contaminant Alleles
0	0.44	0.23
1	0.24	0.34
2	0.15	0.25
3	0.07	0.12
4	0.04	0.05
5	0.02	0.01
6	0.01	0.0033
7	0.0087	0.0007
8	0.0065	0.00013
9+	0.0185	0.000025

this type of contamination. Such contamination would be seen as an excess over Poisson expectations at a high number of alleles per profile. We do not imply by this that contamination by single alleles should be ignored, but rather that it is a "lesser evil."

Gill et al.[348] assumed that contaminant alleles appearing singly should occur in the frequency with which they appeared in the population. Their origin is unknown, but may come from laboratory personnel or consumables for example. It does not seem unreasonable that these alleles would appear in numbers close to their population frequency if they were sampled from a large number of people reasonably independently. However, this assumption would benefit from experimental investigation.

A conclusion from this reasoning is that if an allele is found in a crime sample that does not match the suspect, this does not necessarily lead to an exclusion. The same confidence cannot be placed in the presence or absence of each allele in LCN work as in normal (greater than 100 pg) work. This is likely to lead to considerable debate in court and is one of the drawbacks of LCN work at this stage of its development. The most balanced approach accepting all these difficulties is to interpret the data using a Bayesian model. Formative models have been suggested and will be reviewed here.

8.3 Interpretation of LCN Profiles

In this section, profiles will be described as having spurious bands or having undergone dropout. These statements would be accurate only if we knew the true genotype that should appear in the profile. In real casework, this will never be known. The correct phrasing would be to describe bands as spurious or having dropped out *if* the prosecution hypothesis, H_p, was true. This phraseology is cumbersome and will be dropped in the remainder of this discussion. However, we emphasize that bands are not known to be spurious or to have dropped out. We are simply naming them as such.

8.3.1 The "Biological" Model

Guidelines for reporting LCN profiles have been developed and published.[348] These guidelines are based on the concept that all LCN samples will be analyzed in duplicate, at least. The first step in this process is to compare duplicate profiles and note which alleles are present in both profiles. These alleles are scored in the "consensus" profile. After this process, there may be 0, 1, 2, or more alleles at each locus. It can be very difficult to determine whether the resulting consensus profile represents one or more individuals. The method applied after normal 28-cycle casework relies on the expectation

that most alleles will be visualized. In normal 28-cycle casework, we conclude that if there are more than two alleles scored at a locus then there is more than one individual present. In LCN casework, these rules break down. Three alleles may be the result of a reproduced contaminant, and 0, 1, or 2 alleles at each locus may still represent two individuals with allelic overlap and dropout. Hence the decision as to how many individuals are present in the consensus profile is one of the more problematic decisions in LCN casework. Currently, this decision[c] is based on a subjective assessment of the quality of the profile and peak areas, but this decision is always tentative.

Let us assume that the consensus profile is thought to be from a single individual. Any loci that have zero alleles in the consensus profile are inconclusive and do not affect the likelihood ratio. Under this model the loci with two alleles, both of which match, may be interpreted as normal, although it is important to remember that LCN work cannot be reported with the same confidence as normal work and important caveats should be made. The loci with one allele will have been denoted by, say, 16F, indicating the presence of the 16 allele and "anything else."

These intuitive guidelines are based on biological principles. However, it is useful to compare them against a statistical model. Our aim is to discover whether the results derived from the two methods are reasonably concordant. Of course, concordance does not mean that the estimates are correct — they may both be wrong — but it does increase our confidence in them.

8.3.2 A Formative Bayesian Model

Since we accept that profiles from LCN templates can be affected by "dropout," the appearance of spurious alleles (termed "drop in"), high levels of heterozygote imbalance, and stuttering, we believe that they require a different method of interpretation. If we consider the procedure for interpreting "normal" single-contributor DNA profiles, it proceeds via a series of steps, which include:

1. Consider which hypotheses are relevant in this case.
2. Assign the alleles present in the profile; this step will involve assigning peaks as allelic or artifactual.
3. Determine whether the comparison is a "match" or a "nonmatch."
4. Estimate a match probability or preferably a likelihood ratio for this evidence.

[c] Again this issue is markedly ameliorated by a Bayesian approach. A "decision" as to how many individuals are present is not needed. In fact, the consensus approach is not required at all. Rather the probability of the replicates is calculated conditional on each of all the realistic possibilities.

When we discussed mixture cases (Chapter 7), we introduced the well-acepted concept that there may be ambiguity in the genotypes present. For instance, a locus showing four alleles *a*, *b*, *c*, and *d* of approximately the same peak areas may be a mix of *ab* and *cd*, or *ac* and *bd*, or any other combination of alleles. The inability to explicitly nominate which genotypes are present in a mixture is not an impediment to interpreting the mixture as long as reasonable safeguards are followed. This is especially facilitated by the Bayesian approach. At least with many mixtures we can nominate which peaks are allelic and which are not. However, with mixtures where one component is present in low proportions, it may be impossible to decide, say, whether a peak is allelic or is due to stuttering from a larger adjacent peak. Again, this inability to nominate peaks as allelic or not, is not a fatal impediment to interpretation, and methods to handle this ambiguity were given in Chapter 7. When we consider LCN templates, we must accept that we have even further levels of ambiguity: Specifically: (1) Peaks that are present may be allelic, stuttering, or spurious, or (2) alleles that should be present may have dropped out.

Under these circumstances, it may be unrealistic to unambiguously assign a genotype to a sample. It may, in fact, be very difficult to determine whether the evidence supports the presence of one or more contributors. These factors suggested that a completely different approach to interpreting profiles was warranted. Rather than try to assign genotypes from the sample, we are better served to consider the probability of the electropherogram (EPG) *given* various genotypes. In the following section, we will outline an approach to do this. However, the Reporting Officer must be warned that not only will the evidence take greater skill to interpret and explain but also the very concept that the scene profile and the genotype of the suspect may differ will undoubtedly be a source of serious questioning in court. The key to interpreting and understanding this type of evidence is to think of the probability of the evidence *given* a genotype not the other way around. The following Bayesian model has not been presented in court at the time of writing. At the moment, it serves as a partial theoretical backing for the biological model. It is patented by Buckleton et al.[345] and is being extended and programmed by James Curran for the FSS.

The statistical model that we are about to present contains a number of assumptions. It is coarse and approximate. However, it has allowed us to examine the behavior of the biological model in a useful way. The likelihood ratios produced by this statistical model suffer from all the uncertainties associated with the estimation process for any DNA profile plus a few new and larger uncertainties. Hence likelihood ratios calculated for LCN templates will be less exact and more reliant on modeling assumptions than those calculated for ordinary template levels.

Gill et al.[348] considered modeling the presence or absence of peaks as follows:

C: The event that an allele appears at a given locus in the crime sample as a result of contamination.[d]
D: The event that a given allele drops out. We model the dropout of homozygotes (both alleles) and heterozygotes with the same probability estimate.
St: The event that an allele stutters to form an appreciable peak.

For simplicity, we model the continuous events of contamination, dropout, and stutter as discrete. Under this simplified model, an allele either drops out or it does not, an allele either stutters or it does not, contamination either happens or it does not. Ignoring peak area in this way leads to a loss of information. However, this information loss may not be as large in LCN work as in normal high template work as peak area is less informative for LCN templates. We suggest that the probabilities of the events *C*, *D*, and *St* should be estimated from experimental data.[e]

8.3.2.1 A Coarse Statistical Theory[f]

Suppose that n replicates have been analyzed. We term these $R_1, R_2, \ldots, R_n$. Hypotheses are defined as:

H_1: The DNA in the crime stain is from the suspect.
H_2: The DNA in the crime stain is from someone else.

We can write

$$LR = \frac{\Pr(R_1 R_2 \ldots | H_1)}{\Pr(R_1 R_2 \ldots | H_2)} \tag{8.1}$$

[d] Contamination in this sense is taken to mean additional genetic material added to one replicate during laboratory processing. Genetic material in the sample *per se* is treated as a component of a mixture. Strictly speaking, the probability of a contaminant allele may vary from locus to locus; however, we will make the approximation that the probability is equal across loci. In addition, we will also consider the event $\overline{C}$ by which we mean that no contaminant allele is present. Strictly, we should allow for the possibility that a contaminant allele has appeared in a position of an existing allele (and therefore is not seen). However, for simplicity we have not done this. This should be a minor error using highly polymorphic loci. For example, if a locus shows alleles *ab* in the crime stain and the suspect is an *ab* genotype, then we write $\Pr(\overline{C})$ meaning that we do not need to postulate that contamination has occurred. Strictly, we should allow for the possibility that contamination has occurred at position *a* or *b* and is not seen because it is masked by the presence of true allelic peaks at *a* and *b*. This would suggest a formulation $\Pr(\overline{C}) + \Pr(C)(p_a + p_b)$. Since, in most real cases p_a and p_b are small, this should be a minor error.
Under this model the probability of observing a given contaminant allele at a given locus is taken to be $\Pr(C)\, p_a$, where p_a is the frequency of the allele in the population. Strictly, this should be the population of alleles from contaminant sources. This population may differ from the population proportions if, for instance, most alleles were from operators. It is probable that different laboratories will have different levels of $\Pr(C)$ and that this may differ within a laboratory from time to time or even operator to operator.

[e] The tables we will give later in the examples become very hard to format. For the sake of typographical convenience, we use *D* rather than $\Pr(D)$, and *C* rather than $\Pr(C)$ for the probability of dropout, contamination, and similarly for other events.

[f] Hereinafter referred to as the "Bayesian model."

We assume that the events in one replicate are independent of those in the others. We accept that this assumption is unlikely to be true, but it is useful nonetheless. First we note that it is the events of contamination, stutter, and dropout that we assume to be independent and not the profiles themselves. For instance, a contaminant band in R_1 has no effect on the presence or otherwise of one in R_2. This is a consequence of the "alleles snowing from the ceiling" model. However, the presence of a contaminant in one replicate might increase the chance of one in another. The logic for this is that a laboratory contaminant band has a reasonable chance of occurring in both duplicates. Therefore, the probability of contamination occurring in R_2 conditional on it appearing in R_1 is higher than $\Pr(C)$. Our assumption of independence is likely to underestimate the numerator more than the denominator and hence is likely to be conservative. We examine this issue in more detail later.

Using this assumption, we obtain

$$\Pr(R_1 \ldots | M_j) = \prod_i \Pr(R_i | M_j) \tag{8.2}$$

It is necessary to specify the "components" of H_2. By this we mean that the "expert" determines a set of possible "random man" genotypes worth considering, $M_1, M_2, \ldots, M_n$. These will be exclusive but not necessarily exhaustive. This step is actually unnecessary as it is possible to postulate all possible "random man" genotypes and sum over all possibilities. It does, however, markedly simplify manual implementation if the list can be shortened by judgement.

$$LR = \frac{\Pr(R_1 R_2 \ldots | H_1)}{\sum_i \Pr(R_1 R_2 \ldots | M_j) \Pr(M_j)} \tag{8.3}$$

Hence

$$LR = \frac{\prod_i \Pr(R_i | H_1)}{\sum_i \prod_i \Pr(R_i | M_j) \Pr(M_j)} \tag{8.4}$$

We evaluate the likelihood ratio using Equation (8.4). To calculate $\Pr(R_i|M_j)$, we assume that the events C, D, and St are independent both of each other and of M_j. It is convenient to analyze the components separately in tabular format. This will be demonstrated by example.

Example 8.1 (A heterozygotic suspect and replicates showing apparent dropout). We begin by considering spurious peaks and dropout only. Later we will consider stuttering. Suppose a crime stain was analyzed in three separate replicates (R_1, R_2, and R_3) and three different results were observed at the HUMD18S511 locus: $R_1 = 12$; $R_2 = 16$; $R_3 = 12, 16$. The suspect (S) was 12, 16.

For the hypotheses

H_1: The DNA in the crime stain is from the suspect, and
H_2: The DNA in the crime stain is from someone else

we calculate *LR*, using the format of Table 8.3, and proceed as follows.

Step 1: Assess the reasonable random man genotypes from the information in the replicates. List these in column M_j.
Step 2: Calculate $\Pr(M_j)$ in the second column. We are using the product rule for simplicity, but the subpopulation formulation may be substituted.
Step 3: Calculate $\Pr(R_i \mid M_j)$ in the rows R_1, R_2, and R_3.
Step 4: Calculate the products of each row.
Step 5: Sum the products $= 2p_{12}p_{16}\overline{D}^2 D\overline{C}[\overline{D}^2 D\overline{C}^2 + p_{12}p_{16}C^2]$.
Step 6: The numerator is the product $\Pr(R_i|M_j)$ corresponding to the genotype of the suspect. In the example (Table 8.3), this appears as part of the term at the right-hand side of the second row corresponding to the genotype 12,16 but without the frequency terms, $\overline{D}^4 D^2 \overline{C}^3$.
Step 7: Assemble *LR* by taking the appropriate terms in the numerator and the appropriate sum in the denominator:

$$LR = \frac{1}{2p_{12}p_{16}} \times \frac{1}{\left[1 + \dfrac{p_{12}p_{16}C^2}{D\overline{D}^2\overline{C}^2}\right]} \tag{8.5}$$

Table 8.3 A Layout for Example 8.1

Possible "Random Men," M_j	$\Pr(M_j)$	$\Pr(R_1=12\|M_j)$	$\Pr(R_2=16\|M_j)$	$\Pr(R_3=12,16\|M_j)$	Product
12, 12	p_{12}^2	Allele 12 not dropped No contamination $\overline{D}\overline{C}$	DCp_{16}	$\overline{D}Cp_{16}$	$p_{12}^2 p_{16}^2 \overline{D}^2 DC^2 \overline{C}$
12, 16	$2p_{12},p_{16}$	Allele 12 not dropped Allele 16 dropped No contamination $\overline{D}D\overline{C}$	$\overline{D}D\overline{C}$	$\overline{D}^2\overline{C}$	$2p_{12}p_{16}\overline{D}^4 D^2 \overline{C}^3$
16, 16	p_{16}^2	Allele 16 dropped Cont. by allele 12 DCp_{12}	$\overline{D}\overline{C}$	$\overline{D}Cp_{12}$	$p_{12}^2 p_{16}^2 \overline{D}^2 DC^2 \overline{C}$

Provided that C is small (<0.3) and neither D nor $\overline{D}$ are small, the expression in the denominator $1 + p_{12}p_{16}C^2/D\overline{D}^2\overline{C}^2$ approximately equals 1; hence

$$LR \approx \frac{1}{2p_{12}p_{16}} \tag{8.6}$$

The biological model would have given a consensus profile of 12, 16 and would have reported a likelihood ratio of

$$LR \approx \frac{1}{2p_{12}p_{16}}$$

Hence we see that the biological model is slightly nonconservative relative to the Bayesian model in this case, but the effect is mild if C is small (<0.3) and neither D nor $\overline{D}$ are small. However, the omitted term is always greater than 1, so the approximate LR is non conservative.

Example 8.2 (Apparently single-banded profiles). We consider the situation where only one replicate is available. An apparent single-banded homozygote is encountered in a crime stain ($R_1 = a$) and the peak area is small. This may mean that allelic dropout has occurred. The genotype may in fact be heterozygous and one peak is missing. This possibility should be considered whenever the remaining peak is below a threshold level. At low peak areas, experimental observation confirms that the probability of allele dropout D is high. If the allele in the crime stain is type a and the suspect is type ab, then it would seem reasonable to limit M_j to aa, ab, or aF, where F stands for any allele other than type a or type b. The biological model would not report this since no alleles have been duplicated. However, let us assume that this case was reported and the approximate likelihood ratio $LR \approx 1/2p_a$ was given. We test this approximation for hypothetical purposes to see what would occur if the duplication rule was overlooked. The Bayesian model gives:

Possible Random Men, M_i	$\Pr(M_i)$	$\Pr(R_1 = a \mid M_j)$	$\Pr(M_i) \times \Pr(R_1 = a \mid M_j)$
a, a	p_a^2	No drop, no contamination $\overline{D}\,\overline{C}$	$p_a^2\,\overline{D}\,\overline{C}$
a, b	$2p_ap_b$	a allele not dropped, b allele dropped, no contamination $\overline{D}D\overline{C}$	$2p_ap_b\overline{D}D\overline{C}$
a, F	$2p_a(1-p_a-p_b)$	a allele not dropped, F allele dropped, no contamination $\overline{D}D\overline{C}$	$2p_a(1 - p_a - p_b)\overline{D}D\overline{C}$
Sum			$p_a\,\overline{D}\,\overline{C}[p_a + 2(1 - p_a)D]$

The numerator of LR is $\overline{D}D\overline{C}$, and the denominator is $p_a\overline{D}\overline{C}[p_a + 2(1 - p_a)D]$; hence

$$LR = \frac{\overline{D}\overline{C}D}{p_a\overline{D}\overline{C}\left[p_a+2(1-p_a)D\right]}$$

$$= \frac{D}{p_a\left[p_a+2(1-p_a)D\right]}$$

$$= \frac{1}{2p_a} \times \frac{1}{\left[\frac{p_a}{2D} + (1 - p_a)\right]}$$

$$= \frac{1}{2p_a} \times \frac{1}{\left[1 + p_a\frac{(1 - 2D)}{2D}\right]} \tag{8.7}$$

$$\text{The scaling function } \frac{1}{1 + \frac{1-2D}{2D}\,p_a} \leq 1.0$$

provided that $D < 0.5$, which may be reasonable when the peak is close to the background. Hence the approximation of the reporting of $LR \approx 1/2\,p_a$ may be conservative relative to the Bayesian model as long as dropout is probable based on the peak area of the single peak.

Example 8.3 (Apparently single-banded profiles — the effect of one additional replicate). Extending the previous example, we now consider the advantages of replication. Consider that an additional aliquot (R_2) of the same DNA extract is separately amplified. Suppose that the second replicate yields a heterozygote *ab* profile that matches the suspect's profile (suspect = *ab*; $R_1 = a-$; $R_2 = ab$).

We take account of two possible explanations for the evidence: either M_1=*aa* homozygote or $M_2 = ab$ heterozygote. If the first explanation is true, then the *b* allele must be a spurious band. In this example, the likelihood ratio would be reported as $LR = 1/2p_a$ because only the *a* allele was duplicated. The Bayesian model would give:

Possible Random Men, M_i	$\Pr(M_i)$	$\Pr(R_1 = a \mid M_j)$	$\Pr(R_1 = ab \mid M_j)$	$\Pr(M_i) \times \Pr(R_1 = a \mid M_j)$
a, a	p_a^2	No drop, no contamination $\overline{D}\,\overline{C}$	*a* allele not dropped cont. by allele *b* $\overline{D}\,C\,p_b$	$p_a^2 p_b \overline{D}^2 C\,\overline{C}$
a, b	$2p_a p_b$	*a* allele not dropped, *b* allele dropped, no contamination $\overline{D}D\overline{C}$	*a* allele not dropped, *b* allele not dropped, no contamination $\overline{D}^2\overline{C}$	$2p_a\, p_b \overline{D}^3 D \overline{C}^2$
Sum				$2p_a p_b \overline{D}^2\,\overline{C} \times \left[\frac{1}{2} p_a C + \overline{D} D\,\overline{C}\right]$

The numerator of LR is $\overline{D}^3\,D\,\overline{C}^2$, and the denominator is $2p_a p_b \overline{D}^2\,\overline{C} \times \left[\frac{1}{2} p_a C + \overline{D} D \overline{C}\right]$; hence

$$LR = \frac{\overline{D}^3 D \overline{C}^2}{2p_a p_b \overline{D}^2 \overline{C} \times \left[\frac{1}{2} p_a C + \overline{D} D \overline{C}\right]}$$

$$= \frac{1}{2p_a p_b} \times \frac{1}{\left[1 + \frac{p_a C}{2\overline{D} D \overline{C}}\right]} \tag{8.8}$$

This expression is always less than $1/2p_a p_b$, but the reported approximation $1/2p_a$ is conservative whenever

$$p_b \left[1 + \frac{p_a C}{2D\overline{DC}}\right] \leq 1.0$$

We expect this to be true for many reasonable estimates of C and D.

Example 8.4 (Additional replicates matching the suspect) Extending the two previous examples, we consider the effect of an additional (n) replicates that have been analyzed and demonstrated to all correspond to the genotype of the suspect (*ab* in this example). The suspect is *ab*; $R_1 = a$; $R_2 \ldots {}_{n+1} = ab$ (i.e., a total of $n + 1$ replicates were analyzed). In the above table, we append $n - 1$ copies of the column headed $\Pr(R_1 = ab \mid M_j)$.

The numerator of LR is, $\overline{D}^{2n+1}\, D\overline{C}^{n+1}$, and the denominator is $2p_a p_b \overline{D}^{n+1}\, \overline{C}\left[\frac{1}{2}p_a {p_b}^{n-1} C^n + D\,\overline{D}^n\, \overline{C}^n\right]$; hence

$$LR = \frac{1}{2p_a p_b} \times \frac{1}{\left[1 + p_a p_b^{n-1} \dfrac{C^n}{2D\overline{D}^n \overline{C}^n}\right]} \tag{8.9}$$

Provided that n is greater than or equal to 2, the biological model guideline would allow the reporting of $LR = 1/2p_a p_b$ because both alleles were duplicated. The likelihood ratio approximated using Equation (8.9) will always be less than this, but is nevertheless a very close approximation for most intermediate values of D. Concern would be raised when D is small. This would occur when the a band in R_1 is large. The difference between $n = 2$ and $n = 3$ appears to be minor. Differing values for C appear to have very little effect on the final estimate provided that C is less than 0.6, which it should always be if contamination is kept under reasonable control.

Example 8.5 (Dropout in both replicates). Next we consider the situation where some alleles have dropped out of both replicates. The profile appearing in the consensus will be an apparent homozygote. This locus will carry the F designation. Imagine that both replicates have the allele a. The suspect is ab as before.

Suspect	ab
Replicate 1	a
Replicate 2	a

The biological model again reports $1/2\, p_a$. The LR model would report:

Possible Random Men, M_i	$\Pr(M_i)$	$\Pr(R_1 = a \mid M_j)$	$\Pr(R_2 = a \mid M_j)$	$\Pr(M_i) \times \Pr(R_1 = a \mid M_j) \times \Pr(R_2 = a \mid M_j)$
a, a	p_a^2	$\overline{D}\overline{C}$	$\overline{D}\overline{C}$	$p_a^2 \overline{D}^2 \overline{C}^2$
a, b	$2p_a p_b$	$\overline{D}D\overline{C}$	$\overline{D}D\overline{C}$	$2p_a p_b \overline{D}^2 D^2 \overline{C}^2$
a, F	$2p_a p_F$	$\overline{D}D\overline{C}$	$\overline{D}D\overline{C}$	$2p_a(1 - p_a - p_b)\overline{D}^2\, D^2\, \overline{C}^2$
Sum				$p_a \overline{D}^2\, \overline{C}^2 \times [p_a + 2(1 - p_a)\, D^2]$

The numerator of LR is $\overline{D}^2 D^2\, \overline{C}^2$, and the denominator is $p_a\, \overline{D}^2\, \overline{C}^2 \times [p_a + 2(1 - p_a)D^2]$; hence

$$LR = \frac{1}{2p_a} \times \frac{1}{\left[1 + p_a \dfrac{(1 - 2p(D)^2)}{2p(D)^2}\right]}$$

Inspection of this function suggests that the biological model is likely to be nonconservative in this instance, especially when the probability of dropout is small.

Example 8.6 (An allegedly contaminant band). Suppose that the suspect is *ab*; $R_1 = abc$ (where *c* is a supposedly contaminant allele under H_1) and $R_2 = ab$. We limit the possible (M_j) genotypes to *ab*, *ac*, or *bc* and we evaluate against the biological guideline that would yield $LR \approx 1/2\, p_a\, p_b$. Evaluated against $p(C) = 0.3$, the approximation is reasonable provided that p_a is less than 0.10 and $p(D)$ is less than 0.50.

$$LR = \frac{1}{2p_a p_b} \times \frac{1}{\left[1 + \dfrac{(p_a + p_b)DC}{\overline{D}\,\overline{C}}\right]} \tag{8.10}$$

Example 8.7 (An allegedly contaminant band is observed in conjunction with apparent allelic dropout). The next example is more extreme than those previously discussed. Suppose that a replicate (R_1) matches the suspect (*ab*) at one allele (*b*), but has an additional allele (*c*) that is not found in the suspect. Furthermore, we assume that there is no trace of allele *a*. We assess the condition where the suspect is *ab*; R_1 is *bc*; R_2 is *ab* by consideration of the genotypes (M_j) *ab*, *ac*, *bc*, and *bb*:

$$LR = \frac{1}{2p_a p_b} \times \frac{1}{\left[2 + \dfrac{p_b C(1 + 2D^2)}{2\overline{D}D\overline{C}}\right]} \tag{8.11}$$

The biological reporting guideline would only allow reporting of the duplicated *b* allele; hence the reported likelihood ratio would be $LR = 1/2p_b$. There was very little effect contributed by *D* since the scaling function was always greater than 1.0 even when *C* was moderately high. This supports the conservative nature of the biological model reporting guideline relative to the Bayesian model.

Example 8.8 (Dropout in both replicates and a spurious allele). Consider the situation where some alleles have dropped out of both replicates, but a spurious allele has appeared at one locus. The profile appearing in the consensus will again be an apparent homozygote and the locus will carry the *F* designation. Imagine that both replicates have the allele *a* but one also has allele *c*. The suspect is *ab* as before.

Suspect	*ab*
Replicate 1	*ac*
Replicate 2	*a*

The biological model again reports LR $\approx 1/2p_a$. The Bayesian model would report

$$\begin{aligned} LR &= \frac{p_c D^2 \overline{D}^2 C \overline{C}}{p_a^2 p_c \overline{D}^2 C \overline{C} + 2p_a p_c \overline{D}^3 D \overline{C}^2 + \text{other small terms}} \\ &= \frac{D^2 C}{p_a^2 C + 2p_a \overline{D} D \overline{C} + \text{other small terms}} \\ &< \frac{1}{2p_a} \times \frac{D^2 C}{\overline{D} D \overline{C} + \frac{1}{2} p_a C} \end{aligned}$$

Inspection of this function suggests that it is very likely that the biological model will be seriously nonconservative in this instance. We have considered simply not including such loci in the calculation. However, this would only be reasonable if we could be certain that the contribution to the likelihood ratio from this locus was greater than or equal to 1. The above analysis actually suggests that the contribution to the likelihood ratio may be significantly less than 1. It would seem desirable if this situation is encountered that the entire profile is either not reported or that this locus or the whole profile is reported using the Bayesian statistical model. It seems likely that a Reporting Officer would have noted this locus and be concerned about it without any warning from us.

Example 8.9 (Evaluation where stuttering may be an explanation). We now consider the scenario where the suspect is *ac*, $R_1 = abc$ where *b* is in a stutter position of allele *c* and $R_2 = ac$. We limit the possible M_j genotypes to *ab*, *ac*, and *bc* (recall that this limitation is for convenience and simplicity only and that the approach can be applied to a complete list of any possible genotypes).

$$LR \approx \frac{1}{2p_a p_c} \times \frac{1}{\left[1 + \frac{p_b(p_a + p_c) DC^2}{\overline{D}\overline{C} \times (St\overline{C} + p_b C)}\right]} \tag{8.12}$$

The *LR* is always less than $1/2p_a p_c$ (since the scaling function is always less than 1), but it gives a reasonable approximation provided that the probability of stuttering, *St*, is high (greater than 0.5) and the probability of contamination *C* is less than 0.3. Although the *LR* increases as the probability of dropout *D* decreases, the effect is small provided *St* is high. Hence the key to interpreting this situation, unsurprisingly, is: "What is the probability of the area at *b* in R_1 arising as stutter from *c*?"

Example 8.10 (Multiple spurious alleles in an otherwise matching profile) The Bayesian model described above assumes independence in the appearance of spurious alleles. This assumption appears to be violated at least in part. Further, the event C describing laboratory-based contamination was dealt with at each locus as a binary event; i.e., there was either one contaminant band or there were none. If we accept the "alleles snowing from the ceiling" model, we can see that there could be 0, 1, 2, etc. contaminant alleles per locus.

As we have previously stated, we believe that there is no statistical way to militate against the effects of contamination of substantial partial profiles (≥ 5 alleles). The presence of these, and especially the duplication of these in a pair of replicates, must be guarded against by laboratory-based techniques.

It is possible, however, to investigate the consequences of any violation of the assumption of independence in the appearance of single contaminant alleles. This can be done by considering the effect of this violation on the Bayesian model. As with any other statistical model, neither the biological nor the Bayesian models are claimed to be a "correct" model and hence we cannot claim to be investigating the deviation from some "true answer."

Consider a pair of replicates, both giving a full profile matching a suspect but with one of the "replicates" showing nine spurious alleles at nine of 11 loci. We assume that nine loci have one spurious band, no loci have two or more spurious bands, and $22 - 9 = 13$ loci (from two replicates) have no spurious bands. The consensus profile will be a complete profile and none of the spurious alleles will appear in this consensus. The reader will note that this is an extreme test of the model. For simplicity, we make this profile heterozygous at each locus so that this discussion is not confused with the use of the F designation. The likelihood ratio reported by the biological model would be $1/\Pr(G_s)$, where G_s is the genotype of the consensus (and the suspect). This case would not be treated as a mixture.

For this extension of the Bayesian statistical model, we require the terms $\Pr(C_0)$ and $\Pr(C_1)$: the probability of none or one spurious allele appearing at a locus, respectively. This definition differs slightly from the Gill et al.[348] definition; previously, these were C and $\overline{C}$. We assume, as previously, that these alleles appear at their population frequencies. Hence we assign these alleles the probabilities $p_1,\ldots,p_9$. This would give a likelihood ratio of the type

$$LR = \frac{\Pr(C_1)^9 \Pr(C_0)^{13} \overline{D}^{44} \prod_{i=1}^{9} p_i}{\Pr(G_s)\Pr(C_1)^9 \Pr(C_0)^{13} \overline{D}^{44} \prod_{i=1}^{9} p_i + \text{other small terms}}$$

$$= \frac{1}{\Pr(G_s)} \times \frac{1}{1 + \text{other small terms}}$$

We see that the Bayesian model gives a likelihood ratio close to but smaller than the biological model. The biological model is nonconservative by an amount determined by the size of the sum of the "other small terms." Each of these small terms arise from the alternatives under the defense hypotheses ($H_{2\ldots i}$) that the true contributor may have a genotype different to the consensus, but incorporating one or more of the spurious alleles. It is certainly possible that the sum of many small terms could have a substantial effect on the likelihood ratio. The larger the sum of these, the less conservative the biological model relative to the Bayesian model. These terms are more numerous, the larger the number of spurious bands. In this example, there are $3^9 - 1 = 19{,}682$ alternative "heterozygote only" profiles without even considering potential homozygote profiles and more spurious bands. Therefore, even if these terms are small, there are a number of them.

Next we imagine that the spurious alleles are positively correlated; i.e., the presence of one tends to increase the chance of another. Hence $\Pr(C_1)^9$ is likely to be an underestimate of the appearance of nine spurious bands in one of a pair of replicates. We see that this correlation has no effect on the relative size of the numerator and the first term in the denominator, both of which are larger. The "other small terms" postulate different sets of spurious alleles. These will also be affected by this lack of independence and are also likely to be underestimated. Hence the "other small terms" are likely to be larger if this correlation exists than if it does not. It is difficult to weigh these effects without some very substantial experimental data.

8.3.2.2 Dealing with Multiple Spurious Alleles

We have discussed the situation where spurious bands appear independently. There may be more than one of them in the profile, but they have arrived separately. The analysis given above suggests that the biological model and the Bayesian statistical model have a robustness to this type of contamination. As long as the total rate of contamination is kept within limits, and the situation described above as "dropout in both replicates and a spurious allele" is considered, the appearance of spurious bands singly should be tolerable.

We have discussed above that contamination by the "alleles snowing from the ceiling" mechanism strongly suggests that the appearance of nine or more extra alleles in a profile is very unlikely. These profiles are sufficiently complete to be searchable on the database and reportable. This would occur if the partial or full contaminant profile appeared in both duplicates of a casework sample. No type of statistical analysis can safeguard against this situation. It is necessary to accept that in a fraction of cases a profile may be detected and reported that has nothing to do with the crime sample. It may have occurred from laboratory-based contamination from staff, from consumables brought into the laboratory such as plasticware, or from other sources. It may also

have occurred resulting from DNA profiles being transferred at the crime scene independently of the crime event.

The analysis of minute samples of DNA must also lead to a consideration of the relevance of the evidence. It may be difficult to associate a DNA profile with a body fluid, and also to give information about the possibilities of secondary transfer or persistence. It is possible that mixtures may arise as a result of unconnected events, which may be associated with a crime event, or have nothing to do with the crime event.

8.3.2.3 Mixture Analysis

The method for the interpretation of STR profiles derived from mixtures of DNA has been well documented[182,350] for standard 28-cycle work, but is markedly more complex for LCN templates. The stochastic effects experienced with LCN 34-cycle PCR analysis means that a completely different set of guidelines are required for the latter.

The lower the peak area, the lower the inferential value of peak area in mixture analysis. When the peak area is less than 5000 rfu the relative sizes of allele peaks are virtually uninformative. Above 10,000 rfu it can be inferred that $Hb_{x\ min} = 0.2$. Since the value of peak area information is markedly less for LCN work, the emphasis is moved toward approaches based on methods that do not take explicit account of area.[277,849]

LCN analysis offers the potential to generate useable DNA profiles from a greater range of evidence types and circumstances. However, by operating at the extremities of PCR detection limits, many troublesome characteristics appear in the resultant DNA profiles. These must be understood and included in interpretative frameworks. To some degree, this brings about a challenging corollary. Scientists must be prepared to subjugate some long-held attitudes toward issues such as contamination and reproducibility. However, they must also explore the boundaries of the technology through experimentation and embrace complex inferential logic. LCN is a frontier for forensic autosomal STR profiling from which the benefits could be considerable.

Nonautosomal Forensic Markers

9

JOHN BUCKLETON, SIMON WALSH AND SALLYANN HARBISON

Contents

9.1 Introduction

In this chapter we describe some of the molecular biology and the inheritance and interpretation models relevant to the forensic analysis of nonautosomal markers. Specifically we will discuss the DNA profiling of mitochondrial DNA (mtDNA) and the two gonosomes: the X and Y chromosomes.

9.2 Forensic Mitochondrial DNA Typing

The analysis of mtDNA is of considerable use in forensic science, and has been extensively used in archeology[186,629] and anthropology.[75,150,530,548,647,747,748] The increased copy number and stability of mtDNA provided by the additional protective mitochondrial membrane layers and closed circular double-stranded DNA structure leads to increased success in the analysis of samples from old and otherwise compromised samples. An example of this is the successful recovery of mtDNA from the Pleistocene skeletal remains of an indigenous Australian Aboriginal, putatively dating the skeleton beyond 60,000 years before the present.[3] The matrilineal inheritance of mtDNA coupled with its enhanced stability has been used extensively in anthropological studies and has been used to estimate, for example, that New Zealand was settled by a founding population that included a mere 56 women.[416] Analysis of mtDNA is also used increasingly in the investigation of victims of mass disasters[364] and mass graves.[363]

In forensic science, mtDNA analysis is often used to provide evidence where nuclear DNA fails to give a result, or when distant relatives must be used as reference samples. Typically, tissue almost devoid of nuclear DNA is utilized. This can be because there was little nuclear DNA present originally (for example, bone and hair shafts) or because the sample has been subjected to such severe environmental insult that the mtDNA survives when the nuclear DNA does not (for example, in burnt remains). The prevalent approach involves the determination of sequence variants, and there are a number of approaches to achieve this ranging from conventional PCR and sequencing technology based on the chain termination method originally described by Sanger et al.[687] to denaturing high-performance liquid chromatography.[49,481] Hallmark cases have included the identification of the last Russian Tsar, Nicholas II, and part of his family (the Tsarevich and one of the Grand Duchesses are missing),[351] the exclusion of Anna Anderson as Grand Duchess Anastasia,[240,733] the identification of Air Force 1st Lt. Michael Blassie as the soldier buried in the Vietnam Tomb of the Unknown Soldier,[515] and the identification of the remains of Jesse James[731] and Martin Bormann.[20]

Mitochondria are the remnants of symbiotic α-purple bacteria that were ingested by a eukaryotic cell with no cell wall. The existence of separate DNA

in the mitochondrion was first suspected because the inheritance of some mitochondrial genes was non-Mendelian. It is now understood that mitochondria contain their own DNA, which, in common with bacteria, is a double-stranded closed circular molecule. There is evidence that some of the genes from the ancestral bacteria have been transferred to the nuclear DNA; however, about 16,569 bases remain within the mitochondrion. The molecule contains coding and noncoding areas. The coding regions code for the 37 remaining genes of the mitochondria, which comprise 22 transfer RNAs, 2 ribosomal RNAs, and 13 protein enzymes. Deletions, duplications, and substitutions in this region have been linked to diseases.[411]

Of greatest forensic interest is the control region also termed the D-loop (displacement loop), a structure visible by electron microscopy during replication. The control region is 1125 bases long and flanks the origin of replication. It is noncoding, although it does contain the light (L-) and heavy (H-) strand promoters, transcriptional regulatory elements, binding sites for mitochondrial transcription factors, the origin of H-strand replication, and the termination associated sequence (TAS).[13,145,175,205,234,307,308,409,569,587,592,745,746] These elements would be expected to be under greater selective pressure than areas with no known function, although the lack of functionality of noncoding areas is coming under increasing scrutiny.

Two portions of the control region have been found to be the most variable between individuals, termed hypervariable regions 1 and 2 (HV I and HV II, respectively). HV I extends from position 16,024 to approximately 16,365. HV II extends from approximately position 73 to approximately 340. More recently, a third region, HV III, has been added to this pair. This region shows less polymorphism than HV I and II, but may resolve some cases where additional discrimination is desired.[64,254,525] An excellent review is given by Tully et al.[786] We follow them here.[a]

9.2.1 Matrilineal Inheritance and Recombination

Varying degrees of uniparental inheritance have been suggested for both chloroplasts and mitochondria and Birky[66] warns that "strict uniparental inheritance is probably not as common as is generally believed. The inheritance of mtDNA in interspecific crosses of mice was believed to be strictly uniparental until a more sensitive technique (PCR amplification) was used to detect low levels of paternal mtDNA." Birky suggests that we should treat uniparental inheritance as a quantitative trait. Rates for this type of inheritance vary between species in the range from 0 to 100%. For instance, paternal inheritance has been reported in mice[381] at a frequency of 0.0001 relative to the maternal contribution.

[a] Sections that are reproduced in full from Tully et al.[786] (©1999) appear with kind permission from Elsevier.

It is supposed that the ancient mechanism of inheritance in the first mitochondria-carrying cells was biparental and that uniparental inheritance has arisen subsequently. Birky[66] also comments on possible evolutionary processes that may have led to uniparental inheritance. Mechanisms leading to full or partial uniparental inheritance are thought to vary. They may be as simple as fixation of the more populous maternal mtDNA; however, it does appear that there may be more complex and effective mechanisms, for instance, in mammals.[411] Species-specific exclusion of sperm DNA has been observed in mice.[710] Experimental injection of liver and sperm mtDNA suggests that the mechanism is specific to sperm mtDNA.[711] Recognition of sperm mitochondria by embryonic cells and inhibition of the inheritance of paternal mtDNA has been reported.[212] Possible mechanisms are reviewed by Birky,[67] who discusses the role of ubiquitination. This is a process by which a protein is thought to bind to sperm mitochondria and mark them for subsequent degradation by the 26S proteasome.

Failure of the mechanism has been reported in abnormal embryos[726] but persists following intracytoplasmic sperm injection.[417] A possible replicative advantage for deleteriously mutated mtDNA has also been reported[877] and may be a factor in increasing the fraction of paternal mtDNA in some human cases.[411,699,877]

Uni- or biparental inheritance and recombination are different things and should not be confused. Because there are numerous organelles in each cell, it is possible for there to be biparental inheritance with no recombination. In this process, different organelles contain copies of the DNA from only one parent, but the total population contains examples of each parent. Recombination of chloroplasts has been noted in fungi[65,68,674,871] and slime mold.[450] In contrast, no recombinants were noted in screens of blue mussel even though both genomes "have been present in the fertilised egg and germ line cells of embryos in every generation for over five million years." [66]

Maternal passage of mtDNA in humans has been demonstrated through multiple extended lineages with rare evidence of paternal contribution.[609] Recent papers have suggested that there may be some biparental inheritance and recombination leading to hybrid mtDNA molecules in humans.[26,286] Birky[67] reports that some of this evidence was based on a data set with multiple errors and that the evidence is much weaker when these errors are corrected. Sykes[747] goes to some effort to rebut the claims of recombination in humans in his popular science book *The Seven Daughters of Eve*. The consensus appears to be emerging that recombination in humans is either minimal or does not occur[608] (for excellent reviews, see Birky[67] or Holland and Parsons[411]). For practical purposes, in identification cases maternal inheritance, alone, is usually assumed, and we will follow that practice here.

The consequences of a lack of recombination are quite marked. First the entire mtDNA genome is treated as a haplotype. This means that the whole DNA sequence is treated as a unit and not as the sum of its parts. Frequency estimates are typically made by "counting" the number of occurrences of this haplotype in a database rather than by recourse to population genetic models. The mtDNA is expected to be under severe selection. Any deleterious mutation cannot be repaired and cannot be masked by a dominant gene on a homologous chromosome (since there is none). Since the segments do not recombine, deleterious or advantageous mutations are always linked to the positions that we are investigating forensically. Hence, even if we consider the D-loop as selectively neutral, we must consider the whole mtDNA haplotype to be under selection. However, since we do not invoke either the Hardy–Weinberg or linkage equilibrium assumption, there is no requirement for selective neutrality and accordingly this selection has no implications for the interpretation.

9.2.2 Mutations and Heteroplasmy

MtDNA has a much higher mutation rate than nuclear DNA, about 20-fold higher according to one estimate, although arguments persist as to the exact rate (for a review, see Gibbons[335]). The estimates of mutation rate from short pedigrees or mother–child pairs appear to be higher than that obtained from phylogenetic studies. An explanation for this is still required, but such discrepancies have also been noted for the Y chromosome. Mutations and heteroplasmy are relatively common occurrences. The reasons advanced for the higher mutation rate include the relatively high turnover both in mitotic and post-mitotic cells, and the susceptibility of mitochondrial structures to oxidative stress.[874] The functioning of the mitochondrion produces the superoxide anion, O_2^-, that is implicated in the promotion of mutation. However, the female germline mitochondria are present in the eggs of females that have undergone a mere 24 divisions early in embryonic development. There is also evidence that the mitochondria of ova have been "shut down" and that these cells are respiring anaerobically. This helps to preserve the mtDNA of the female germline (recall that the male germline does not pass on its mtDNA).

The most common type of mutation found within the forensically significant D-loop region are single base substitutions with transitions outnumbering transversions by approximately 40:1. Small insertions and deletions are common in the two homopolymeric polycytosine (poly(C)) regions between positions 302 and 310 and between positions 16,183 and 16,194. Insertions and deletions can hinder the interpretation of mixed samples as the sequence of the contributing individuals is thrown out of register. The chance of observing insertions and deletions at homopolymeric regions

increases as the length of the uninterrupted homopolymer stretch increases. Other alterations include large deletions and short duplications.[874]

Stability varies along the mtDNA genome as some base positions appear to be very stable while others are highly mutable.[284,358,397,398,411,471,550,802,866] Furthermore, some body tissues, such as hairs, tend to show more variability in their mtDNA sequence.[868] This could be due to differential segregation of preexisting heteroplasmic variants, to the accumulation of new somatic mutations, or to a combination of both phenomena. There is still some debate over whether the number of point mutations increases with age or whether or not it is implicated in the aging process.[874] Mutations are passed between generations in varying ratios and segregate during development and later life. Mutations also accumulate and segregate during the lifetime of an individual (reviewed in Holland and Parsons[411]). This results in mixtures of mtDNA molecules that characteristically differ from each other at one or more bases. This is known as heteroplasmy.

Heteroplasmy can be either sequence or length based. It can occur in an individual in essentially three different ways.[49]

A single tissue type from an individual may have more than one mtDNA type, or different tissues from the same person may exhibit different mtDNA types or some tissues may have more than one mtDNA type but another tissue from the same person only one mtDNA type. Heteroplasmy probably exists in all individuals, although it is often at such a low level that it cannot always be detected by the routine sequencing techniques presently used. In order for a mutation to be detected above background by sequencing, it must currently be present at a level approaching 20%. In addition, the chance of detection of heteroplasmy is dependent upon the sequencing chemistry used. Detection may be more efficient at certain nucleotide positions than at others, and differences in detection may also be observed between the two DNA strands. The interrelationship between observation of heteroplasmy and amplification strategy is discussed by Brandstatter and Parson.[80]

A report of exceptionally high levels of heteroplasmy in hair roots[374] led to a considerable discussion of these influences. This discussion included the suggestion that a high starting template, high cycles, and the use of nested PCR may have given rise to the high observation of heteroplasmy.[125,126] D'Eustachio[228] argued that high template levels should not have an effect. Tully and Lareu[785] raised the possibility of contamination. The original work had been undertaken using nested PCR. Reanalysis of the same samples by Grzybowski et al.[373] using a non-nested approach supported the suggestion that the nested approach gives higher indications of heteroplasmy. Tully et al.[787] give a sensitive method for detecting low levels of heteroplasmy (see also Holland and Parsons[411]) and suggest that there are heteroplasmy "hot spots" in the genome.

Budowle et al.[125,126] argued that the location of heteroplasmy should correlate with sites known to show high variation. Stoneking[732] found evidence for this by direct comparison of new mutations with hypervariable sites, and Holland and Parsons agree.[411]

Once a mutation has led to significant heteroplasmy in the germline, the offspring will either be heteroplasmic, or fixation on one mitotype will have occurred, so that heteroplasmy can no longer be detected. This suggests that heteroplasmy may be an intermediate state in the evolution from one mitotype to another. Holland and Parsons[411] review the evidence that suggests that the mitochondria are reduced in number during oogenesis. This bottleneck has the potential to alter the ratio of heteroplasmic variants or to lead to fixation on one mitotype. An interesting investigation of this looked at 180 twin pairs and found matching heteroplasmy in four instances. Analysis of other family members showed differing ratios of the heteroplasmic variants. Various authors have estimated the size of the bottleneck and these estimates vary widely.

There is some residual debate as to whether or not mutations accumulate during an individual's lifetime. For example, Calloway et al.[144] investigated the variation of heteroplasmy with tissue and age. They found the highest levels in muscle and found that heteroplasmy increased with age. Bai et al.[31] present supporting data from a study of a large deletion that is associated with ageing and deafness. Lagerstrom-Fermer et al.[483] present limited data that would support the opposite conclusion. Review articles strongly favor the former position.[411,786]

A consequence for forensic science of the presence of an elevated mutation rate and heteroplasmy is that the evidential (crime) sample and reference samples may exhibit sequence differences even when the two are, in reality, from the same individual or lineage. This strongly suggests use of the logical interpretation approach, which will be discussed later.

9.2.3 Nomenclature

Conventions for mtDNA nomenclature have been published[50,786] and we again follow them here.

The first published complete mtDNA sequence[17] is referred to as the "Anderson sequence" or as the "Cambridge Reference Sequence (CRS)." The nomenclature standard for mtDNA has emerged from this. The Anderson sequence does not originate from a single individual. At some bases it contains the rarer variant, notably at positions 263 and 315.1. The sequence contains several sections of bovine and HeLa sequence. However, it is preferred for use when describing the control region, due to its widespread acceptance.

Starting at an arbitrary position near the origin of heavy strand (H-strand) replication, each of the bases in the mitochondrial genome was given a

consecutive number from 1 to 16,569. Any sequence may be described as a list of differences from the Anderson sequence. Of the two bases on each of the complementary strands, it is the base on the cytosine-rich light strand (L-strand) that is quoted. For example, in the Anderson sequence at position 280 the L-strand is C. If the sequence to be reported differs at this position, then it is reported. Suppose that the base of the L-strand in this sequence position is T; this is reported as 280T.

Tully et al.[786] recommend that for sequence data outside the control region the "Revised Cambridge Reference Sequence[18]" should be used. This sequence corrected errors and rare polymorphisms in the original Anderson sequence. In any instance, a clear statement of which reference system is used is important.

9.2.3.1 Insertions and Deletions

Deletions are reported by nominating the deleted base with a "d." For example, if the base between 245 and 247 were deleted, this would be listed as 246d.

Insertions are reported at the insertion position with a "0.1" after the lower numbered base. For example, 245.1A informs us that an adenosine base was inserted after base 245. When an insertion occurs within a homopolymeric tract, the exact location of the insertion is ambiguous. By convention, the insertion is labeled as occurring at the high end of the homopolymeric tract. For example, an insertion in the poly(C) stretch between positions 302 and 310 would be designated as 309.1C; two insertions here would be listed as 309.1C and 309.2C.

9.2.3.2 Point Heteroplasmy

There are several options for describing heteroplasmy: (1) The appropriate IUB designation may be used. For example, heteroplasmy at position 152 would be designated as 152Y. (2) Alternatively, the designation T∼C could be used. When the level of heteroplasmy is comparable in the two reactions and one base is at a substantially higher proportion than the other, notation of the type C > T may be employed. When the presence of two bases cannot be confirmed by the second sequencing reaction, the position would be designated as ambiguous (N) or by use of the IUPAC codes[b] in the relevant extract.

Hühne et al.[423] have described an approach in which a specific threshold level is defined for heteroplasmy. Implementation of such a threshold is hampered due to different laboratories employing different sequencing chemistries.[787] Tully et al. favor interpreting each sequence individually, taking into account the background "noise" over the entire sequence.

[b] R = (A or G), Y = (C or T), K = (G or T), M = (A or C), B = (C,G, or T), D = (A,G, or T), H = (A,C, or T), V = (A,C, or G), S = (G or C), W = (A or T).

9.2.3.3 *Length Heteroplasmy*

Length heteroplasmy is a particular concern in the poly(C) stretches of both HV I (between 16,183 and 16,194) and HV II (between 302 and 310).[411] If an "out of register" sequence (indicating the presence of more than one length variant) is observed in one sequencing reaction after a poly(C) stretch, the second sequencing reaction would confirm the observation. In the case of poly(C) stretches, the second sequencing reaction would often be of the same strand as the first because of the commonly encountered difficulty in sequencing through long poly(C) tracts. If the number of cytosines present can thus be confirmed, nomenclature of the type 309.1 > 309.2 could be employed. However, if it were not possible to confirm the presence of a mixed number of cytosines by two amplification and sequencing reactions, the number of cytosine bases would be reported as ambiguous.

Rasmussen et al.[632] describe a method to obtain reliable mtDNA sequences downstream of a length heteroplasmy in the homopolymeric regions.

9.2.4 Interpretation — The Logical Approach

The various options for interpretation that were discussed in Chapter 2 are equally applicable to mtDNA, except that there are some situations where the logical approach has additional advantages.

Before progressing to a discussion of "matches and exclusions," it is best to set up the logical framework. This departs from the approach taken in previous chapters where the frequentist approach was given first. The reason for this is that the terms "match" and "exclusion" can be ambiguous in mitotyping. The correct line of thinking is best laid down before approaching these complexities. We move to the development of two hypotheses. As always, care must be taken. We consider that we have typed a crime stain or material, C, and produced its mitotype G_C. The prosecution allege that this stain or material comes from a person, S. If this person, or a sample from him (we consistently use him for simplicity), is available, we may have the mitotype for S, G_S. It is unnecessary to assume that the mitotype is a single sequence. It may be two or more if the samples show heteroplasmy.

Often we may not have a sample from S and may be using, as a reference, material from a (maternal) relative, K, with mitotype G_K. Consider two alternative source level propositions:

H_p: the evidential sample, C, originates from S, or
H_d: the evidential sample, C, originates from a different individual to S.

We need to differentiate the situations where S is available and when he is not. Let us begin with the situation where we have the mitotype for S. As

usual, following the logical approach, we require

$$LR = \frac{\Pr(G_C|G_S, H_p)}{\Pr(G_C|G_S, H_d)}$$

having made the assumption that $\Pr(G_S|H_p) = \Pr(G_S|H_d)$.

If we assume that the mitotype from two samples from the same person will always be the same, then it would be possible to write $\Pr(G_C|G_S, H_p) = 1$. However, it may be necessary to consider somatic variations such as heteroplasmy. If so, it may not always be possible to assume that two samples from the same person, but from different tissues, or at different times will give the same mitotype. Equally, if the mitotype is a combination of two or more sequences, these may appear in differing ratios from sample to sample, even when they are from the same source. What is needed is a probabilistic model to assess the joint probability of two samples, or the probability of one sample conditional on another when both have come from the same individual. This would be a profitable area for interpretation research.

Next consider that we may not have *S* available and are working with a reference sample from *K*. We have available the background information, *I*, which includes the information on the relationship between *K* and *S*.

Following the logical approach, we require

$$LR = \frac{\Pr(G_C|G_K, H_p, I)}{\Pr(G_C|G_K, H_d, I)}$$

having made the assumption that $\Pr(G_K|H_p, I) = \Pr(G_K|H_d, I)$.

Now we need to consider whether or not the mitotype from two samples from the two different people related by *I* will always be the same. If they always were the same, then $\Pr(G_C|G_K, H_p, I) = 1$. However, it may be necessary to consider mutation between generations and somatic variation. It will not always be possible to assume that two samples from two different but matrilineally related people would give exactly the same mitotype. Again, what is needed is a probabilistic model to assess the joint or conditional probability.

It is well established that the mutation rate for mitochondrial DNA is substantially higher than that encountered with nuclear DNA. Consequently, it is not uncommon for differences to be observed in the DNA sequence when comparing close maternal relatives (such as a mother and a child).[609] Substitution has also been observed in somatic tissues, presumably due to the segregation of existing heteroplasmy within the individual. This means that differences may be observed between different hairs or tissues within an individual.[144,742,868] Consequently, if there are mismatches between G_C and G_S or G_K, this does not automatically exclude the hypothesis that *C* and *S* are from the same individual or that *C* and *K* are matrilineally related.[49]

If *C* and *S* or *K* do not match at a base position, it is clear that the strength of the evidence will be dependent upon the inherent mutability of that base. The current level of knowledge regarding substitution rates at each position within the mtDNA noncoding region is limited. Nevertheless, estimation of substitution rates in the control region is assisted by (1) observations of both heteroplasmy and substitutions from germline and intraindividual mutation studies[144,411,423,424,609,714,742,787,868] and (2) observations of the occurrence of heteroplasmy in casework and research inferences from phylogenetic studies.[284,358,397,398,471,550,802,866]

For example, long poly(C) stretches (between positions 302 and 310 and between 16,183 and 16,194) are extremely mutable since substitutions are observed in at least one out of ten individuals (unpublished observations). At the opposite end of the spectrum, position 73 is thought to be relatively stable.[866] Thus, a difference at position 73 may provide stronger evidence for exclusion of *K* as the matrilineal origin of *C* than two base differences between *K* and *C* if these are at positions 309.2 (homopolymeric region) and 16,093 (an apparent mutation "hot spot" that has been observed to vary in studies from several laboratories).[787]

However, despite such estimates, precise values for substitution rates are difficult to determine for the following reasons: (1) segregation of mutations will occur at different rates in different tissues; (2) there may be sequence context specific variations in substitution rates; and (3) paternal inheritance and recombination, if they occur at appreciable rates, may bias inference from phylogenetic studies.

In applying a logical approach (for example, as given by Holland and Parsons[411]), the assessment of the term $\Pr(G_C|G_K, H_p, I)$ is informed by these considerations. If there are many differences between G_C and G_K, the probability $\Pr(G_C|G_K, H_p, I)$ is effectively zero, and hence an exclusion can be reported. Conversely, if G_C and G_K share no differences, the probability $\Pr(G_C|G_K, H_p, I)$ approaches, but does not quite reach, 1. In the case of sequences that differ at one or two base positions, the value of the numerator term, $\Pr(G_C|G_K, H_p, I)$ is intermediate but likely to be low.

Because the mtDNA is inherited as a haplotype, the sequence is interpreted as a single haploid locus; it is invalid to estimate evidential strength by multiplying the population frequencies for each of the bases in the sequence.

Currently, most laboratories use the counting method whereby the number of matching sequences in the database of size *N* is reported. This method presents the evidence in a purely factual way and is a statement of observation; there are no assumptions related to population genetics. This may seem an advantage, but it does overlook some significant issues.

All of the above discussion assumes that the frequencies of all mtDNA haplotypes in the relevant reference population are known. This is not the case.

However, since we are working with haplotypes the issue does not relate so much to the application of a population genetic model for estimation of a genotype probability as it did for nuclear DNA. Rather the issue relates to the fit between the estimate from the database and the parameter in the population of interest in any particular case. The following considerations are therefore important.

9.2.4.1 Distribution of Mitochondrial DNA Types in Populations

We suppose that a laboratory has a database of mitotypes taken probably from convenience samples across their particular catchment area. Each laboratory may seek to report the frequency of a mitotype by simply counting (and possibly correcting) the frequency of this haplotype in their database. Large population databases also exist (for example, MITOMAP[153] and EMPOP[607]).

For this to be a sensible statistical practice, we require that the database is a representative sample from the population defined by the crime under H_d. This is most likely to be true in general if the migration rate has been high enough to effectively randomize mitotypes within the population. This helps ensure that the database is representative of the population of interest in a particular case. If we can be confident that the database is representative of the population of interest under H_d, then we can, as a first approximation, assign the probability of this mitotype as the sample frequency in the database (the counting rule[411]). Since we expect our database to be the result of multinomial sampling this count is expected to be an unbiased estimate.

Evidence is reported[545] that some randomization has occurred in European cosmopolitan populations since θ is low (< 0.01). Since no multiplication is to be attempted in the estimation of the haplotype probability, this value for θ will not accrue an "error across loci" as it would with nuclear DNA. Hence the argument for the use of a θ correction in mitochondrial DNA and Y chromosomal testing is much weaker. However, if the probability of the mitotype and an STR profile are to be combined, then this heterogeneity should be carefully considered.

Although he does not report a θ value, Sykes's[747] data would appear to differ since he is able to suggest areas where there are high frequencies of certain descendants of his seven "Eves". Equally, the data of Richards et al.[647] would strongly support inhomogeneity. Mesa et al.[548] give a value of 0.18 for G_{ST} for mtDNA in five Columbian populations undergoing admixture. In Asian (Indian subcontinent) populations, θ can be up to 0.10,[544] which is large when compared with nuclear DNA (see Chapter 5).

There is some information on small village populations in Europe where mobility may have been severely restricted over centuries. These data suggest the remnants of ancestral heterogeneity. Most of the existing forensic databases cannot address this question since they were collected over a wide area across a wide range of subpopulations (for example, the U.K. database comprises

individuals from all over the United Kingdom). Smaller or relatively isolated populations need to be analyzed in order to improve the understanding of population genetics of mtDNA at a local level. For some populations, for example in the South Pacific, founding population sizes are estimated to be small with an equivalently restricted number of mitotypes.[570] Additional studies on small or isolated populations are occurring.[617]

The fact that anthropological studies do produce meaningful results suggests that there is real heterogeneity in the human population. This variation can often be detected on a very local scale. This suggests that even more care must be given to sample choice when we are creating mtDNA databases than for autosomal ones.

Most of the anthropological surveys target very carefully selected individuals whose families are suspected of having been long-standing local residents. These surveys often reveal ancient heterogeneity and hence are fascinating from an anthropological point of view. However, the structure of such anthropological surveys may make them unsuitable for direct use in forensic science. They may underestimate the frequency of common modern haplotypes by deliberately seeking ancient diversity. This should be borne in mind if such data are used for forensic interpretation.

From the forensic point of view, we really require less structured and more random surveys of these same small isolated populations. The size and construction of such surveys is likely to be problematic and suggests consideration of whether the expected results of such surveys could be simulated from existing data.

It is interesting to note that anthropological comparisons of Y chromosome and mtDNA distributions appear to reflect cultural practices such as patrilocality, whereby the woman moves to live with the man. This would suggest more localization of Y chromosome types and a slightly larger geographical distribution of mitotypes. This is supported by studies in Europe,[705] Melanesia, and Australia.[453] Matrilocal societies show the opposite pattern.[595] Similar anthropological studies appear to suggest that the histories of the mtDNA and the Y chromosome may be different for quite different social reasons. For instance, there appears to be considerable evidence that those areas settled by Vikings had a larger contribution of Scandinavian males than females.[400] More of the females appeared to be of Gaelic ancestry.

Gusmão et al.[380] detected clines across Europe by clustering Y haplotypes with their "one-step" mutation neighbors. Such clines have implications for population structure. Fernandes and Brehm[290] note a difference between the population of the Açores Islands and those of Madeira Island and North Portugal. They use this to argue correctly for appropriate local databases.

Therefore, the assumption of homogeneity of mitotypes within the population may hold approximately for large and ancestrally mobile populations, but is unlikely ever to be exactly correct and is also less appropriate for any

small, historically localized, populations. Provided that the data could be developed, the frequency of a mitotype in smaller, less mixed populations could be estimated by applying appropriate θ corrections.

We can find no report in the literature yet of how to apply the θ correction for mtDNA chromosome haplotypes should one decide that it is necessary. Applying the logic of Balding and Nichols[36] leads to

$$LR \approx \frac{1}{\theta+(1-\theta)\Pr(G_C)} \tag{9.1}$$

where $\Pr(G_C)$ is the probability of this mitotype.

Numerical examination of Equation (9.1) suggests that its use with large values for θ would reduce the likelihood ratio substantially. Intuitively it would seem to reduce it too far. We conclude that further investigation into how to compensate for population subdivision at mtDNA loci is warranted urgently. In the absence of new theory, it is imperative that every effort should be made to use appropriate local databases and hence no correction or a low value for θ.

9.2.4.2 Heteroplasmy

If both the crime and reference samples have the same degree of heteroplasmy, this will increase the strength of the evidence.[411] This occurs because the denominator term $\Pr(G_C|G_K, H_d, I)$ is low, since the occurrence of a certain heteroplasmic mitotype is expected to be low.[430]

Holland and Parsons[411] discuss the situation where a blood sample from a suspect is known to exhibit heteroplasmy. A single hair alleged to be from this person is examined and found to exhibit a homoplasmic sequence for one of the types present in the blood. Clearly if multiple hairs from the suspect could be examined, a probabilistic approach could be taken. Imagine that a fraction x of hairs exhibited a matching homoplasmic mitotype. Then the numerator of the likelihood ratio is x. The denominator of the likelihood ratio may be taken as the population frequency of the homoplasmic variant. However, Holland and Parsons point out that this detailed knowledge of the fraction of hairs with this variant would seldom be available.

9.2.4.3 Sampling Uncertainty

The frequency of a mitotype, p, is uncertain in a population because of population heterogeneity and sampling variation.[c] With relatively small databases, and a large number of haplotypes, it is inevitable that new variants that do not appear in the database will frequently be observed. Balding[33] suggested an estimator in

[c] We will use p for the true, but unknown population proportion. $\bar{p}$ will be the sample proportion and $\hat{p}$ will be used for an estimate of the population parameter.

the nuclear DNA context. The analogous estimator in the haploid case is

$$\hat{p} = \frac{x + 1}{n + 2}$$

where x is the number of matching samples in that database and n is the total number of samples in the database. This follows from taking a uniform beta prior and assuming binomial sampling, and corrects the typographical error in Tully et al.[786] where $\hat{p}$ was described as a conservative Bayesian estimator. In fact, it is the posterior mean as described by Evett and Weir[267] (p. 69). Also following Weir, we obtain

$$LR = \frac{n + 3}{x + 2}$$

It should not be assumed that this estimator compensates for sampling variation in a meaningful way.

When there are one or more matching samples in the database, then the use of confidence intervals has been recommended based on a normal approximation to the binomial.[786] Holland and Parsons[411] suggest the use of the large sample normal approximation to give a two-sided 95% confidence interval:

$$\tilde{p} \pm 1.96 \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}$$

where $\tilde{p} = x/n$ is the sample proportion. These authors correctly note that when $\tilde{p}$ is small the normal approximation is known to be poor. However, $\tilde{p}$ is likely to be small in all mitochondrial casework. They suggest that a better approximation may be achieved by assuming the log of the proportion to be normally distributed. This is similar to the approach taken by Chakraborty et al.[165] for nuclear DNA evidence. For a mitotype previously unseen (a zero sample proportion), these authors suggest an interval based on $\hat{p} = 1 - \alpha^{1/n}$ where α is set to 0.05 for a 95% confidence interval (see also Budowle et al.,[123] Chakraborty,[158] and Weir[828]). Tully et al.[786] report a comparison of the 95% confidence limits from a zero sample proportion and the (incorrect) Bayesian estimator $(x + 2)/(n + 2)$. This revealed little difference between the two methods for the instance of a zero sample count. There is likely to be very little difference between the incorrect Bayesian estimator and the corrected version.

The above argument relies on the assumption that each person has only one mitotype and that people show no somatic mosaicism. Consider, now, the situation where we compare a single hair recovered from a scene with a blood or buccal sample from a suspect. Assume that the mitotype of the hair and the suspect show no differences. Under these circumstances, we may be prepared to assign $\Pr(G_C|G_S, H_p) \cong 1$. But what is $\Pr(G_C|G_S, H_d)$? Consider the probability of obtaining G_C if the evidential hair did not come from the suspect but from someone else. Evaluation of this probability requires an estimate of the

frequency of occurrence of a G_C-type hair in the relevant population. Because mtDNA frequency databases are composed primarily of sequences from blood or saliva samples, we need to consider the possibility that a hair mitotype may differ slightly from the blood or saliva mitotype.

For example, if only one individual in a population of 100 has the sequence G_C, but ten individuals have a sequence, G_i, that differs from G_C only at position 309.1, then the probability of obtaining G_C if the hair did not come from the suspect may be higher than would be estimated from the occurrence of G_C alone in the database.

This suggests a procedure such as

$$\begin{aligned}\Pr(G_C|G_S, H_d) &= p \\ &= \sum_i \Pr(G_C|G_i, H_d)\Pr(G_i|H_d)\end{aligned}$$

where G_i is the genotype of the ith person in the population. If we now assume that the database is a representative sample of the population of interest, we write

$$\hat{p} = \frac{\sum_k \Pr(G_C|G_k)}{n}$$

where G_k is the genotype of the kth person in the database of size n. Until we have a model to estimate the probability that a hair will differ from the blood in a certain way at a certain position, this equation cannot be implemented.

The probability suggested by this equation is likely to be substantially less than the combined frequency of the genotype with no differences, G_C, and those differing at one base. The increasing body of data on mtDNA mutation rates and segregation will enable the significance of such a correction factor to be assessed. However, as a first approximation it is likely that the frequency of a sequence in a database of mtDNA sequences from blood samples closely approximates the frequency of that sequence among hairs, in which case such a correction is unnecessary.

9.2.4.4 Examples of the Logical Approach in Practice

Although the current data regarding mtDNA substitution rates and population genetics are limited, the general likelihood ratio formulation is a useful framework upon which an assessment of evidential significance can be made. Examples of how this framework could be used in commonly encountered situations are given below. In each, the likelihood ratio framework is used to assess the evidence.

Example 9.1 (*Matching S and C mtDNA sequences*). The known, *S*, and questioned, *C*, sequences match exactly. There are no observations of the same or similar sequences in the database. Under H_p the samples are alleged

to have come from the same person. H_d alleges that they come from different people. We seek to assess

$$LR = \frac{\Pr(G_C|G_S, H_p)}{\Pr(G_C|G_S, H_d)}$$

The numerator term $\Pr(G_C|G_S, H_p)$ relates to the chance of obtaining a matching sample from the same person as the reference. Hence issues such as the type of bodily samples used for questioned and known profiles and somatic mosaicism are relevant. However, this term is likely to be close to 1. The denominator term $\Pr(G_C|G_S, H_d)$ relates to the chance of obtaining a matching sample from a different person to the reference sample. The counting method, or a frequency (corrected for sampling error and the distribution of mtDNA types in the population if applicable) could be used to inform this term.

Example 9.2 (*Matching K and C mtDNA sequences*). The known, K, and questioned, C, sequences match exactly and there are no observations of the same or similar sequences in the database. Under H_p, the samples C are alleged to have come from a person S who is not available. However, his maternal grandmother, K, has provided a sample. H_d alleges the samples C have come from a different person to S and hence are not necessarily maternally related to K. We seek to assess

$$LR = \frac{\Pr(G_C|G_K, H_p, I)}{\Pr(G_C|G_K, H_d, I)}$$

The numerator term $\Pr(G_C|G_K, H_p, I)$ relates to the chance of obtaining a matching sample from a person who is the grandson of the donor of the reference sample. Hence issues such as "what is the chance of no mutations in two generations?" and "what body samples were used to obtain the questioned and known profiles?" are relevant. Somatic mosaicism must also be considered. Once again, this term is likely to be near 1. The denominator term $\Pr(G_C|G_K, H_d, I)$ relates to the chance of obtaining a matching sample from a different person to the grandson of the donor of the reference sample. Again, the counting method (or estimating a frequency corrected for sampling error and the distribution of mtDNA types in the population if applicable) could be used to inform this term.

Example 9.3 (*MtDNA sequences K and C differ by a single base*). The known, K, and questioned, C, sequences differ by a single base that is known to mutate frequently. Neither sequence has been observed previously in the database, but there are other sequences in the database differing by a single base. Under H_p the samples C are alleged to have come from a person S who is not available. We have a sample from his maternal grandmother, K. H_d alleges the samples C have come from a different person to S. As usual, we

seek to assess

$$LR = \frac{\Pr(G_C|G_K, H_p, I)}{\Pr(G_C|G_K, H_d, I)}$$

The numerator term $\Pr(G_C|G_K, H_p, I)$ relates to the chance of obtaining a matching sample from a person who is the grandson of the donor of the reference sample. Hence issues such as "what is the chance of this particular mutation at this site in two generations?" must be assessed. This term is likely to be small but nonzero. The denominator term $\Pr(G_C|G_K, H_d, I)$ relates to the chance of obtaining a sample matching C from a different person to the grandson of the donor of the reference sample. Again, the counting method could be used to inform this term.

Each case should, of course, be treated on its own merits, using all available data regarding the frequency of the sequences in question, mutation rates, somatic mosaicism, and the structure of the relevant populations. In rare instances, where population structure has been well studied, differences between sequences are at well-characterized nucleotide positions, and sizeable frequency databases for the relevant population are available, the likelihood ratio estimate may be sufficiently accurate to warrant a numerical statement. In the vast majority of current cases, this is unlikely to be true.

9.2.5 Interpretation — The Frequentist Approach

The logical approach does not require a statement of "inclusion" or "exclusion." The nearest equivalent to these terms would be a likelihood ratio greater or less than one, respectively. The frequentist approach to interpreting this evidence, however, does require such a definition. An inclusion is usually defined by nominating the maximum number of bases at which two samples may differ and yet be deemed to match. An exclusion is the converse. Logically, this should relate to whether the reference sample is alleged to be the same person (S) or a maternal relative (K), and if a maternal relative then it should be conditional on the pedigree information (I). As described above, this criterion should also consider exactly which bases differ. This is because different mutation rates can be expected for different sequence positions. However, in practice the criterion usually specifies simply the number of bases that are allowed to differ.[284,550,802] Budowle et al.[124,125] suggest:

1. If the two sequences are the same at all sites either both heteroplasmic or both homoplasmic report: failure to exclude.
2. If one sequence shows heteroplasmy and the other is homoplasmic sharing bases report: failure to exclude.
3. If both sequences are homoplasmic but differ by one nucleotide report: inconclusive.

Bär et al.[49] suggest a similar scheme with the following exception. If both sequences are homoplasmic but differ by one nucleotide report: inconclusive. However, these authors discuss situations where this may be evidence supporting different maternal origins. In particular, they emphasize this when both samples are from the same tissue, such as blood.

Logically, this criterion also specifies a "window" for the calculation of the frequency. This frequency should be calculated as the sum of the frequencies of all sequences that would be deemed to be "not excluded."

If the sequence of interest has not been observed in the database, it is incumbent on the forensic scientist to ensure that the court is not left with the impression that the sequence could be as rare as an autosomal STR profile.

9.2.6 Combining Mitochondrial DNA Estimates with Nuclear DNA Estimates

We have been unable to find any published suggestions on this matter. If the nuclear DNA estimate has been formed by the product rule and the mtDNA estimate is the count in a database, then multiplying them is likely to be a reasonable first estimate of the joint probability. However, such an estimate is likely to have a bias toward too low a probability, and hence against the defendant — an outcome we have consistently opposed.

Another suggestion (Buckleton) would be to develop the nuclear DNA estimate using the subpopulation formulation of Balding and Nichol's and an appropriate nuclear DNA estimate of θ. The mtDNA estimate could be developed using

$$\theta_M + (1 - \theta_M)\frac{(count + 2)}{N + 3}$$

where θ_M refers to an appropriate mitochondrial θ estimate. The two estimates could then be multiplied. However, this is something of an intuitive guess and we are unable to produce any formal mathematical foundation for it.

9.3 Forensic Y Chromosome Analysis

9.3.1 Introduction

Each male has only one Y chromosome, a condition known as hemizygosity. The human Y chromosome represents about 2% of the total human genome and is approximately 60 Mb in length. Ohno[588] was the first to postulate that the human Y chromosome emerged from a severely degenerate X chromosome. This presumably arose when an ancestral mammal developed a sex determining gene on one X chromosome. It is assumed that prior to that sex had been

determined by factors such as temperature, mechanisms that persist in reptiles and many other animal orders. The male determining gene (SRY gene) resides in the male specific (MSY) or nonrecombining (NRY) region near the distal end of the short arm.

A recent paper by Skaletsky et al.[716] reviews the current understanding of the genetic structure of the Y chromosome, in particular the region comprising about 95% its length where there is no X–Y crossing-over (see Figure 9.1) known as the nonrecombining region.[675] Recombination occurs only at the distal portions of the short and long arms of the chromosome, the pseudoautosomal

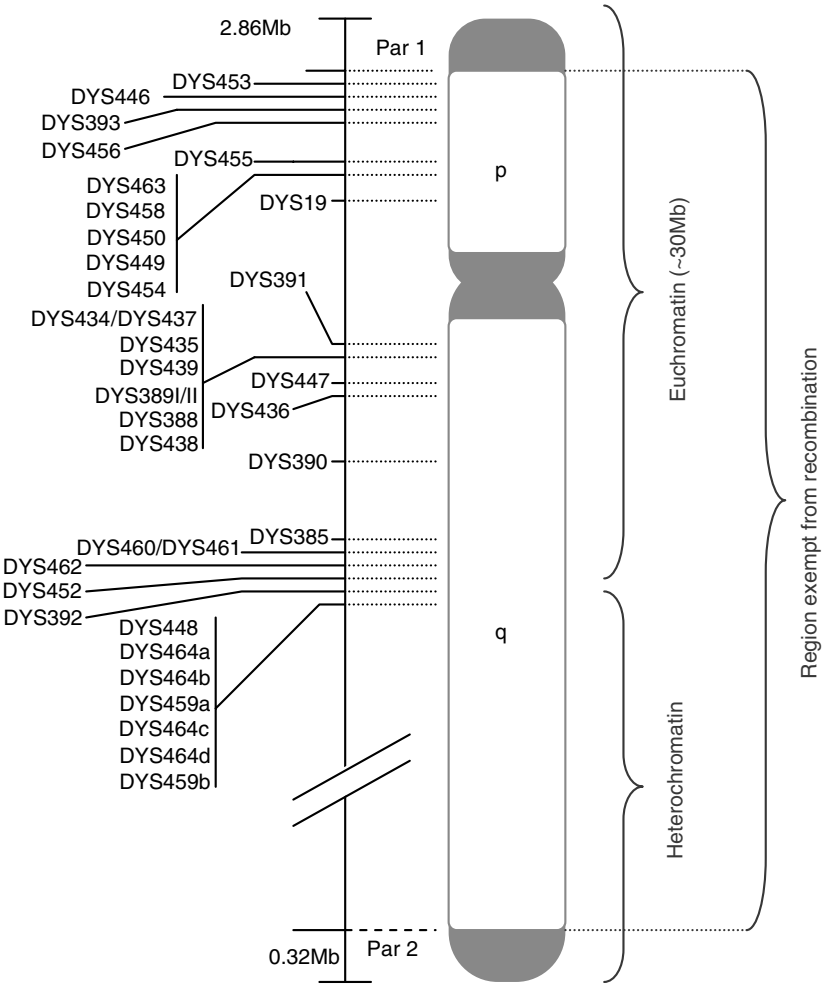


Figure 9.1 Schematic of the Y chromosome showing approximate locations of microsatellite loci (not drawn to scale).

regions (PAR 1 and 2).[630] The recombination between the ends of the Y chromosome and the X chromosome has been observed microscopically and is part of the original evidence that identified these as a pair.

There is growing evidence that the MSY region of chromosomes may have some ability to recombine within itself. Large segments of DNA are repeated in reverse along the length of the Y chromosome. These sequences are termed palindromes, as they read the same forward or reverse, and may have some ability to recombine (perhaps more appropriately termed "gene conversion"). This is very different from normal recombination with a homologous chromosome, but may have some implications for DNA repair in the Y chromosome. There is also extensive evidence that such gene shuffling is implicated in massive deletions and consequent infertility.

The lack of ability to recombine is thought to be the reason why there are few functional genes on the Y chromosome. There are 78 protein coding sequences in the MSY that encode 27 distinct proteins. This is a low density of genes compared with the autosomes. Any deleterious mutation that "knocks out" a gene has little chance of repair and cannot be removed by recombination. High population frequencies have been observed for some deleterious mutations[640,788] and are supposed to be the result of repetitive recreation by mutation or support from advantageous mutations elsewhere on the Y chromosome. It is generally expected that all functional genes on the Y chromosome will slowly be deactivated by mutation, including eventually the male determining switch. Unless these genes have relocated elsewhere, they will be lost with implications that Sykes[749] graphically explores.

Skaletsky et al.[716] detail the characterization of the sequence of the MSY region. It had been previously determined that the region contained repetitive sequences called amplicons, comprising long intrachromosomal repetitive sequences, each of which possessed little sequence variation.[692] However, these small sequence variations were used to determine a physical map of the MSY[773] that has now been developed into the complete DNA sequence of the MSY. The unusual chromosomal structure of the Y chromosome coupled with a lack of crossover and a high apparent rate of gene conversion suggests that Y chromosome sequence variation might differ markedly from other chromosomes.

There is a large array of different polymorphisms that are suitable for PCR-based analysis, including single nucleotide polymorphisms (SNPs), microsatellites, and minisatellites. Jobling[441] estimated that thousands of base substitutions probably exist in the contemporary human population. This estimate was formed by extrapolating from the research of Hammer,[388] who had observed three base substitutions in a 2.6 kb region among 16 individuals.

The characteristics of a high degree of polymorphism, the ability to multiplex using PCR, and the ease of analysis makes the analysis of short tandem repeats (STRs) the current method of choice for forensic DNA profiling

favored over SNPs. Roewer et al.[666] reported the first STR on the Y chromosome now known as DYS19 in 1992. Since then the potential use of Y STR analysis for forensic casework has been recognized and well documented.[61,140,402,413,442,445,452,713] Y STR typing is useful when there are mixtures of DNA from two or more males, when there is a low level of male DNA in comparison to the amount of female DNA present in a sample, or when a study of family relationships are needed.

Recently there have been reports of fetal microchimerism. This describes a situation where fetal cells become established in the mother and give rise to chimerism. Potentially a female could contain cell lines derived from a male child and hence this female could give rise to a weak Y chromosome typing result. This issue was investigated by Klintschar et al.,[468] who found that no interference was likely under normal template and amplification conditions.

There are a variety of commercially available multiplexes now produced for forensic Y chromosome applications. For example, ReliaGene Technologies Inc. produce the Y-plex^{TM}12 multiplex and Promega Corporation produce a PowerPlex Y kit. These kits allow forensic laboratories without the resources to develop their own multiplexes to undertake Y STR analysis. Specific sets of loci, often referred to as minimum haplotypes, have been chosen by groups in Europe[451] and the U.S. (Scientific Working Group on DNA Analysis Methods, SWGDAM) to encourage the standardization of population data. This objective was helped by the construction of a large database (see http://www.ystr.charite.de) that now includes American (see http://www.ystr.org/usa) and Asian (see http://www.ystr.org/asia) population data. Over 12,000 haplotypes are compiled on this database and are available online for researchers and caseworkers.[667] There is an ongoing search for suitable Y STRs with novel markers identified and validated regularly.[76,426,635,855] These STRs are occasionally duplicated, triplicated, or quadrupled,[55] giving rise to two, three, or four amplification products.

Y STR analysis is undertaken in the same way as autosomal STR analysis using the same equipment and methods. Homologous regions with varying degrees of conservation are present on the X and Y chromosomes. This means that supposedly Y chromosome specific primers may sometimes amplify similar products from the X chromosome. An example of such nonspecific amplification was described by the initial set of primers chosen for DYS391.[379] Since the Y chromosome does not recombine, it is treated as a haplotype in much the same way as mtDNA. Again the forensic loci will be linked to any deleterious or advantageous mutation. Even if forensic loci are selectively neutral, and there is little evidence either way, we must expect selection to be active on the whole haplotype including, by linkage, the forensic loci. For example, spermatogenic ability is thought to vary in differing Y lineages.[480]

Hallmark studies have included the investigation of whether Thomas Jefferson fathered a child by one of his slaves (the evidence supports that he

did)[317] and the finding that approximately 8% of all males in Asia may be descended from Genghis Khan.[445]

9.3.2 Mutation

Y chromosomes evolve along paternal lineages accumulating diversity only by mutational processes. The lineages are distributed in a clustered manner among human populations supposedly because of genetic and cultural factors.[440] The patrilineage of Y chromosome haplotypes reduces their diversity in the general population. To this effect is added a smaller effective population size and hence more rapid drift. The effective population size of the Y chromosome is one-fourth that of the autosomes and one-third that of the X chromosome,[d] but is the same as the mtDNA. The time to the most recent common ancestor for the Y chromosome is currently thought to be about 90,000 years, whereas it is about 240,000 years for mtDNA. Both of these estimates have considerable uncertainty. The shallower nature of the Y chromosome tree may be explained by men having a higher variance in reproductive success, by higher rates of migration, or by selection. There is some evidence in support of each of these.

This loss of diversity and a more structured population may be seen as disadvantages, given the high discriminating power that forensic scientists are accustomed to from multilocus autosomal profiles. However, this feature affords significant advantages as well, particularly for the study of population genetics and human evolution. The lower mutation rates for base substitutions make them extremely valuable in tree building. They tend to give a reliable coarse structure to the tree. However, a sufficient number are needed.[147,442,605] Microsatellites are more rapidly evolving and tend to be useful to investigate the fine structure of trees.

Some Y STRs originate from a duplicated tandem repeat array, additional alleles arising by insertion polymorphisms of the larger chromosomal region including the STR locus followed by a mutational change in the number of repeats within the STR locus. Locus multiplication (the multiplication of the amplicon including both forward and reverse primer binding sites) has been reported for several Y STR loci, including:

DYS19[445,452,666]
DYS385[142,451,695,708]
DYS390[76]
DYS435[445]

[d] Consider a male–female pair of individuals. They have four autosomes, three X and one Y chromosome between them. This suggests that the autosomes have an effective population size four times that of the Y. The X chromosome has an effective population size three times that of the Y.

Table 9.1 Mutation Rates for Some Y Chromosomal Loci (Number of Mutations in per meiosis or Mutation Rate)

Locus	Heyer et al. (1997)[403]	Kayser et al. (1997)[451]	Kayser et al. (2000)[454,c]	Dupuy et al. (2001)[238]
DYS19[a]	0 in 213	2 in 626	2 in 996	
DYS390[a]	0 in 213	1 in 94	4 in 466	0.013
DYS391[a]	0 in 213	0 in 41	2 in 415	0.007
DYS392[a]	1 in 213	0 in 42	0 in 415	
DYS393[a]	0 in 213	0 in 42	0 in 415	
DYS398[a]	2 in 213			
DYS385a/b[a]	1 in 213[e]	1 in 104	2 in 952	0.007
DYS389 I[a]		0 in 55	1 in 425	
DYS389 II[a]		1 in 53	2 in 425	0.013
DXYS156	0 in 213			
YCA Ia/b		0 in 72	0 in 150	
YCA IIa/b[b]		0 in 113	0 in 240	
YCA III		0 in 42		
DYS413a/b			1 in 100	
Total	4 in 213	2 in 626	14 in 4999	
MSY1	0.02 to 0.11 per generation[443]			
DYS464	2 in 7055[f]			
Average	0.0021			

[a] Indicates loci of the European "minimal" haplotype.

[b] Additional locus added to form the "extended" haplotype.

[c] Includes five mutations from Kayser et al.[451]

These can be explained as a result of insertion polymorphism of a larger chromosomal region (or replicative transposition)[716] followed by a mutational change in the number of repeats within the STR locus. This produces two, three, or four peaks of similar size if two, three, or four PCR products are produced. Using standard PCR methods, assignment of the alleles to loci cannot be done.[451]

Heyer et al.[403] determined the average mutation rate of Y specific microsatellites (see Table 9.1) as 2.1×10^{-3} based on the haplotypes of 36 males descended from 10 "founding fathers" over a total of 213 generations. Jobling et al.[444] reassessed microsatellite mutation rates by reanalysis of the deep-rooting pedigrees studied by Heyer et al.[403] using an additional locus, MSY1. This research supported some earlier assumptions of nonpaternity, thereby endorsing the previously reported mutation rate estimate. The mutation rate of MSY1 itself is estimated to be as high as 0.02 to 0.11 per generation.[443] Dupuy et al.[238] give an

[e] Heyer et al.[403] actually report 1 mutation for DYS385a and 0 for DYS385b.

[f] DYS464 is a multicopy STR showing four copies on the q arm of the Y chromosome.

average rate of 0.0042. Weber and Wong,[823] through analysis of 19 tetranucleotide autosomal STR loci, calculate an average mutation rate of 2.1×10^{-3}. Kayser et al.[451] analyzed 626 meioses at the DYS19 Y STR locus (446 confirmed father–son pairs from their study combined with data from 180 previously published meioses[690,776]), and calculated a mutation rate for this locus of 3.2×10^{-3}.

Kayser et al.[451] expected that the DYS19 data would provide an insight into the mutational processes for other Y STRs of similar structure. Analysis of 14 germline mutational events (including five from the earlier study) from 4999 meioses at 15 Y chromosome microsatellite loci justified this expectation.[454] The average tetranucleotide mutation rate across eight loci was calculated as 3.17×10^{-3} (95% CI 1.89–4.94×10^{-3}), the average dinucleotide mutation rate across six loci was calculated as 2.04×10^{-3} (95% CI 0.06–10.93×10^{-3}), and the overall average mutation rate across all 15 loci was calculated as 2.80×10^{-3} (95% CI 1.72–4.27×10^{-3}).

Bianchi et al.[63] extracted DNA from immortalized cell lines and profiled samples at seven Y-specific STRs. Two mutations were observed, but attributed to somatic events. The authors concluded that no germline mutations occurred in over 1743 meioses. Distinguishing somatic events from true germline mutations is a problematic feature of cell line research.[452]

Recent work by Zhivotovsky et al.[885] estimated an average mutation rate for a Y STR locus as 6.9×10^{-4} per 25 years by utilizing evolutionary studies. The difference between this estimate and that of the other published works is in accord with similar differences between direct observations of mutation rates in parent–child pairs and mutation rates developed from evolutionary studies.

Both the Heyer et al.[403] and Kayser et al.[451] studies support the stepwise mutation model[463] and argue against recombination-related mutations. Compound microsatellites were more likely to show mutations, with gains more likely than losses (observed ratio 10 gains:4 losses). Mutations tended to appear in the longest array of homogenous repeats.[454]

These features of Y STR mutation rate research correspond to autosomal findings. The four mutations reported by Heyer et al.[403] and the 14 reported by Kayser et al.[454] all occurred in uninterrupted arrays of $\geq$10 and $\geq$11 homogenous repeats, respectively. This is similar to the results of Brinkmann et al.,[92] where all 23 mutations observed in nine autosomal STR loci from over 11,000 meioses occurred in uninterrupted arrays of $\geq$10 homogenous repeats. Similarly, the higher mutation rate for tetrameric repeat regions compared to dimeric repeats[454] is also supported by autosomal findings.[823] Mutations at two loci (DYS390 and DYS389) were observed in one father–son pair, out of a total of 415 analyzed at the same nine loci. Such observations are also rare in autosomal microsatellite loci, but have occurred in paternity testing.[376]

These results imply that for di-, tri-, tetra-, and penta-nucleotide repeat loci the Y chromosome mutation rate is comparable, perhaps even slightly

higher, than that of the autosomes. When considered in conjunction with the reduced overall diversity on the Y chromosome, the high mutation rate implies that some identical Y microsatellite haplotypes may occur in the population through recurrent mutation events rather than shared patrilineage.

9.3.3 Databases of Y Haplotypes

Many of the considerations for mtDNA are equally applicable to Y chromosome databases. The considerable evidence for geographical substructure[822] at the Y chromosome means that certain sampling strategies will lead to overrepresentation of rare variants.[387] Of particular concern is the sampling of multiple populations and their assembly into global databases.

As with mtDNA, we can find no report in the literature yet of how to interpret Y chromosome haplotypes accounting for population subdivision. The logic of Balding and Nichols[36] leads to

$$LR \approx \frac{1}{\theta_Y + (1 - \theta_Y)\Pr(H)} \tag{9.2}$$

where $\Pr(H)$ is the probability of this haplotype, and θ_Y the estimate for the *Y* chromosome.

Again, use of this equation with large values for θ_Y would reduce the likelihood ratio substantially. As with mtDNA, we conclude from this that further investigation into how to compensate for population subdivision at the Y chromosome is warranted urgently. As many authors have commented, it is imperative that every effort should be made to use appropriate local databases and hence no correction or low values for θ_Y.

9.3.4 Y chromosome θ, R_{ST}, G_{ST}, or Φ_{ST} Values

A general shift to longer alleles has been observed in Asia with shortened alleles more prevalent in Africa and America.[451]

The presence of clines or discontinuities in the U.K.[822] and across Europe[380] is strong evidence for the presence of population subdivision. Mesa et al.[548] give G_{ST} values of 0.165 for the Y chromosome in five Columbian populations. They obtained higher G_{ST} values when they considered South America as a whole (0.287 for mtDNA; 0.299 for Y) or when they restricted themselves to Y chromosomes of confirmed Amerind origin (0.22). Zhivotovsky et al.[884] imply a high θ value for the Y chromosome in three Pakistani communities, but do not give a figure. Pérez-Lezaum et al.[614] give a comparison to autosomal markers and suggest that the readily observed increase in θ may be attributed to smaller effective population size.

Dupuy et al.[238] give data that suggest much lower substructuring within Norway with pairwise θ values of the order of 0.01 or less. Gusmão et al.[380]

and Martin et al.[536] give Φ_{ST} or R_{ST} values of less than 0.01 for samples of Iberian origin. They obtain much higher values when they compared to some other populations. De Knijff et al.[224] present distances between the Caucasian populations: Dutch, German, Swiss, and Italian. The maximum Φ_{ST} value observed was 0.0812 for Dutch and German.

9.3.5 Negative Correlation of Mitochondrial DNA and Y Chromosome Haplotypes

Using both mtDNA and Y haplotypes, Carvajal-Carmona et al.[150] estimate that 94% of the Y chromosome in the population of Antioquia (Columbia) is European, 5% African, and 1% Amerind. The mtDNA is estimated to be 90% Amerind. Similar results appear in Iceland, resulting from more ancient admixture. This may be a common occurrence in other admixed populations.[689] The prevalence and practical impact of such effects have not been considered in the literature.

9.3.6 Combining Y Chromosome Evidence with Autosomal or Mitochondrial DNA Evidence

As with the previous section on combining mtDNA frequency estimates with autosomal DNA frequency estimates, we can find nothing published on combining either of these with Y chromosome estimates. Bruce Walsh has brought some novel ideas to our attention.
If we write:

G_C: the autosomal genotype of the crime stain,
G_S: the autosomal genotype of the suspect,
M_C: the mitotype of the crime stain,
M_S: the mitotype of the crime stain,
Y_C: the Y chromosome haplotype of the crime stain,
Y_S: the Y chromosome haplotype of the suspect,

$$LR = \frac{\Pr(G_C, G_S, M_C, M_S, Y_C, Y_S|H_p)}{\Pr(G_C, G_S, M_C, M_S, Y_C, Y_S|H_d)}$$

$$= \frac{\Pr(G_C, M_C, Y_C|G_S, M_S, Y_S, H_p)\Pr(G_S, M_S, Y_S,|H_p)}{\Pr(G_C, M_C, Y_C,|G_S, M_S, Y_S, H_d)\Pr(G_S, M_S, Y_S|H_d)}$$

Assuming that $\Pr(G_S, M_S, Y_S|H_p) = \Pr(G_S, M_S, Y_S|H_d)$ and $\Pr(G_C, M_C, Y_C|G_S, M_S, Y_S, H_p) \approx 1$, we are led to

$$LR \approx \frac{1}{\Pr(G_C, M_C, Y_C|G_S, M_S, Y_S, H_d)}$$

This could be decomposed using the third law of probability in a number of different ways. It is unclear which of these will be the most easy to evaluate. For example, one of these decompositions is

$$LR \approx \frac{1}{\Pr(G_C, M_C, Y_C | G_S, M_S, Y_S, H_d)}$$

$$\approx \frac{1}{\Pr(Y_C | Y_S, G_C, G_S, M_C, M_S, H_d)\Pr(M_C | M_S, G_C, G_S, Y_S, H_d)\Pr(G_C | G_S, M_S, Y_S, H_d)}$$

One issue is that the information inherent in the Y chromosome haplotype and mitotype contains quite strong information as to the subpopulation of origin. In contrast, although the autosomal genotype also contains such information, it is possibly weaker, depending on how many loci have been typed. Terms such as $\Pr(Y_C | Y_S, G_C, G_S, M_C, M_S, H_d)$ represent the probability that the crime stain Y chromosomal haplotype will match given that the mitotypes and autosomal genotypes have already matched. This may be larger than the unconditional probability $\Pr(Y_C | Y_S, H_d)$. Walsh has suggested an approach that proceeds by adjusting the θ value upwards to account for the matches already observed. This is clearly promising and warrants urgent research. We eagerly await his publication.

Our suggestion until then would be to develop the nuclear DNA estimate using the subpopulation formulation of Balding and Nichols's and an appropriate nuclear DNA θ estimate.[36] The mtDNA estimate could be developed using

$$\theta_M + (1 - \theta_M)\frac{(count_M + 2)}{N_M + 3}$$

where θ_M refers to an appropriate mitochondrial θ estimate and $count_M$ and N_M are the count of matching mitotypes and the size of the mitochondrial DNA database. For the Y chromosomal data, we suggest

$$\theta_Y + (1 - \theta_Y)\frac{(count_Y + 2)}{N_Y + 3}$$

where θ_Y refers to an appropriate Y chromosome θ estimate and $count_Y$ and N_Y are the count of matching Y chromosome haplotypes and the size of the Y chromosomal database. However, it is likely that this will be an underestimate of the joint frequency of an autosomal, mitochondrial, and Y chromosomal match since this approach has not properly accounted for the expected raise in θ_M value from the autosomal DNA match and in the θ_Y value from the autosomal DNA and mtDNA match.

9.3.7 Y Chromosome Mixtures

In this section we consider the question of Y chromosome mixtures. This follows from an insightful inquiry and discussion with Oscar García of the Basque Country Forensic Genetics Laboratory.

The interpretation of mixtures of DNA from males using Y chromosome analysis differs from the interpretation of autosomal DNA mixtures in at least two respects. First, many Y chromosome loci are represented by a single allele at each locus. This means that the complications of heterozygote imbalance are eliminated for these loci. Some loci are duplicated, triplicated, or quadrupled such that two, three, or four amplicons are produced. For these loci, the term heterozygote imbalance is inappropriate; each of these "linked" alleles exists in a different sequence environment to others attributed to the same locus such that amplification efficiency may differ for each one and they may not appear "balanced."

Secondly, the Y chromosome is treated as a haplotype and this must be accounted for when considering Y chromosome mixtures.

Male/male mixtures have been successfully resolved at ratios of 10:1[445] and even 50:1 where no minor alleles were in stutter positions.[626] There are no reports of major and minor proportions changing significantly across loci and we have not observed this in our laboratory.

Like autosomal STRs, Y chromosome STR profiles from mixed samples can be affected by the presence of stutter peaks, and this needs to be taken into consideration when interpreting profiles from mixed samples. Stutter ratios for Y STR loci appear to be of a similar magnitude to those for autosomal STRs.[445,626,715] Similar rules need to be developed by each laboratory to determine how to interpret a peak in a stutter position.

Consider a simplified Y chromosome mixture with just three loci: 1, 2, and 3. We term the DNA type of this mixture, S. At each locus, there is a clear major contributor and a clear minor contributor. The two alleles at a locus are labeled 1 and 2. Hence the alleles at locus 1 are A_1^1 and A_2^1, where we label the A_1^1 allele to be from the major contributor. Clearly, then, this is a mixture of a major with haplotype $H_1 = A_1^1\, A_1^2\, A_1^3$ and a minor with haplotype $H_2 = A_2^1\, A_2^2\, A_2^3$. We assume a scenario such as a double rape where the two hypotheses may be:

H_p: The contributors are suspect 1, S_1 of haplotype H_1, and suspect 2, S_2 of haplotype H_2.
H_d: The contributors are two random men.

Under this scenario, assuming independence between random men,

$$\Pr(S|H_p) \approx \frac{1}{2}$$

$$\Pr(S|H_d) \approx \Pr(H_1)\Pr(H_2)$$

$$LR \approx \frac{1}{2\Pr(H_1)\Pr(H_2)}$$

or

$$LR \approx \frac{(1+\theta)(1+2\theta)}{2(\theta+(1-\theta)\Pr(H_1))(\theta+(1-\theta)\Pr(H_2))} \tag{9.3}$$

Next we consider the masking effect. This follows similar principles to mixtures of autosomal DNA. Consider a mixture where the major appears to be $H_1 = A_1^1 A_1^2 A_1^3$. There are minor peaks at locus 1 and 2, $A_2^1 A_2^2$. At locus 3, we see only the major peak.

In such a case, as long as we are confident that the minor allele has not dropped out, we can assume that it must be A_1^3 and that $H_2 = A_2^1 A_2^2 A_1^3$. The suspect has the haplotype H_2. This suggests

$$LR \approx \frac{1}{2\Pr(H_1)\Pr(H_2)}$$

or

$$\approx \frac{(1+\theta)(1+2\theta)}{2(\theta+(1-\theta)\Pr(H_1))(\theta+(1-\theta)\Pr(H_2))}$$

If we suspect that the minor allele may have dropped out, we need to consider all haplotypes that match at the remaining loci. We could write $H_2 = A_2^1 A_2^{2\star}$, where the $\star$ is meant to indicate that any allele may be present. To estimate this, we can simply count those haplotypes in the database that possess the pair of alleles $A_2^1 A_2^2$ and use the fact that the frequency of all alleles sum to 1.

Consider when a peak is present in a stutter position and we cannot conclude whether it is allelic or not. If the minor is sufficiently small that it may be the peak in the stutter position, then we need to consider two possibilities: the minor allele may be masked by the major peak or it could be the peak in the stutter position. Let the stutter peak be at position S. An allele at this position would be designated A_S^3 and let $H_S = A_2^1 A_2^2 A_S^3$. The suspect, again, has the haplotype $H_2 = A_2^1 A_2^2 A_1^3$. This would suggest

$$LR \approx \frac{1}{2\Pr(H_1)\{\Pr(H_2)+\Pr(H_S)\}}$$

or

$$\approx \frac{(1+\theta)(1+2\theta)}{2(\theta+(1-\theta)\Pr(H_1))(\theta+(1-\theta)[\Pr(H_2)+\Pr(H_S)])}$$

Last we come to the situation where peak area information suggests two approximately equal contributors or is unreliable for some reason. Hence the stain shows alleles $A_1^1 A_2^1$ at locus 1, $A_1^2 A_2^2$ at locus 2, and $A_1^3 A_2^3$ at locus 3. We cannot tell which alleles originated from one or the other of the contributors.

There are eight combinations of haplotypes possible for these three loci. These combinations come as four pairs. For example, we could have H_1 and H_8 or H_8 and H_1. Let

$$H_1 = A_1^1 A_1^2 A_1^3, \quad H_2 = A_1^1 A_1^2 A_2^3, \quad H_3 = A_1^1 A_2^2 A_1^3, \quad H_4 = A_1^1 A_2^2 A_2^3$$
$$H_8 = A_2^1 A_2^2 A_2^3, \quad H_7 = A_2^1 A_2^2 A_1^3, \quad H_6 = A_2^1 A_1^2 A_2^3, \quad H_5 = A_2^1 A_1^2 A_1^3$$

Recall that suspect 1 has haplotype H_1 and suspect 2 has haplotype H_2.

$$LR \approx \frac{1}{2\{\Pr(H_1)\Pr(H_8) + \Pr(H_2)\Pr(H_7) + \Pr(H_3)\Pr(H_6) + \Pr(H_4)\Pr(H_5)\}}$$

or

$$\approx \frac{(1+\theta)(1+2\theta)}{2(\theta-1)\left\{\begin{array}{c}\{\theta + (1-\theta)\Pr(H_1)\}\Pr(H_8) + \{\theta + (1-\theta)\Pr(H_2)\}\Pr(H_7) + \\ (1-\theta)\{\Pr(H_3)\Pr(H_6) + \Pr(H_4)\Pr(H_5)\}\end{array}\right\}}$$

For many loci, the enumeration of all haplotype combinations may become tedious and it may be better to write a small program.

9.4 Forensic X Chromosome Analysis

Most human females possess two X chromosomes that are present as a homologous pair. This is often written as "XX." It is thought that one of this pair is inactivated[528,750] and reduced to a Barr body. This explains why X chromosome monosomies, trisomies, and polysomies are not ubiquitously fatal. Complete and partial monosomies have been observed and are associated with Ullrich–Turner syndrome. The X chromosomes recombine in the female.

Normal males possess one X and one Y chromosome. This is often written as "XY." However, syndromes do exist whereby XY individuals present as the female phenotype. In such cases it is thought that the sex determining gene is either absent or inactive on the Y.

The genome database (www.gdb.org) lists 26 tri- and 90 tetranucleotide repeat polymorphisms on the X chromosome. 18 tetra-, 3 trinucleotide repeat loci and the VNTR locus DXS52 are in common forensic use.[750] X chromosome markers have advantages in deficient paternity cases, for example, when a biological sample is not available from the putative father and samples from paternal relatives are used instead. When females have the same father, they also share the same paternal X chromosome. This can be used to investigate,

for instance, half-sisters (if they have four alleles between them, then they are not half-siblings). If two close relatives are suspects in a paternity case, the X chromosome may have advantages over autosomal markers. Szibor et al.[750] consider a father–son pair. These two cannot have X chromosomes that are IBD. This may assist in determining the paternity of a female child.

X chromosome mutation rates would not be expected to be markedly different from autosomal rates; however, data are sparse (see Table 9.2 for some reported mutation rates).

Using both physical and genetic mapping methods, the relative locations of some X chromosome STRs of practical interest were investigated. This resulted in the map shown in Figure 9.2. Linkage disequilibrium was observed only for DXS101 and DXS7424 based on the investigation of 210 male DNA samples.[750]

We discuss the effect of linkage in paternity testing in Chapter 10, but in brief there are no implications in this context unless there are two meioses (for example, two children) or the phase of the mother is known from other data, such as the typing of the X chromosome of her father.

Table 9.2 X Chromosome STR Mutation Rates as Estimated from Tests of Trios[750]

STR	Mutations/Meioses	$\mu \times 10^{-3}$	95% CI
DXS6807	0/440	0.00	0.00–8.38
DXS9895	0/761	0.00	0.00–4.85
DXS8378	1/308	3.25	0.08–18.09
DXS9902	0/304	0.00	0.00–12.13
DXS7132	1/260	3.85	0.09–21.43
ARA[1]	4/562	4.92	1.01–14.37
DXS6800	0/440	0.00	0.00–8.38
DXS9898	0/754	0.00	0.00–4.89
DXS6789	0/752	0.00	0.00–4.91
DXS101	0/440	0.00	0.00–8.38
DXS7424	0/440	0.00	0.00–9.22
DXS7133	0/263	0.00	0.00–14.03
GATA172D04	0/370	0.00	0.00–9.97
HPRTB	3/610	4.92	1.01–14.37
DXS7423	2/234	8.55	1.03–30.87
DXS8377	5/760	6.58	2.13–15.35
CUMULATIVE	16/658	2.09	1.25–3.32

Reproduced from Szibor et al.[750] with kind permission of the authors and Springer-Verlag who retain ownership of the copyright.

[1]Dr Szibor has permitted us to see a forthcoming letter to the editor in which Szibor, Hering and Edlemann announce that they believe that HUMARA should not be further used for forensic purposes because of the principle that forensic DNA testing should never reveal personal risk factors or disease genes.

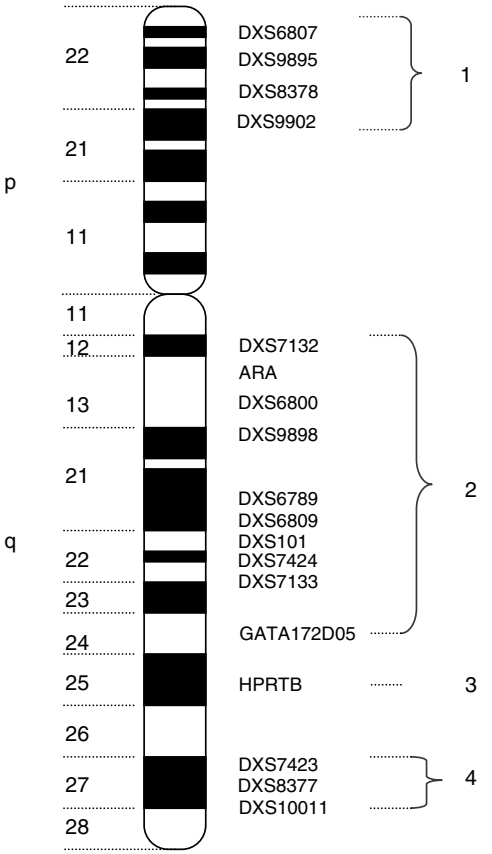


Figure 9.2 Diagrammatic map of the human X chromosome showing the approximate location of microsatellite loci used in forensic analyses. The numbers 1 to 4 represent four known linkage groups. This figure draws heavily from a more detailed description provided by Szibor et al.[750]

9.5 A Famous Case Example — The Romanovs

The Romanov dynasty ruled Russia from 1613 until 1918. The late 19th century had seen intermittent reforms in Russia, such as the abolition of serfdom, but at the close of the century substantial civil unrest existed. The 1905 war with Japan humbled Russia and led to renewed pressure for the reform of the Tsar's autocratic rule, the establishment of a representative assembly, and the granting of a constitution. In 1914 Russia declared war on Austria in support of its ally, Serbia, thus precipitating the First World War. A German declaration of war on Russia, also in accord with treaty obligations, followed and France honored its treaty with Russia by entering the war. Britain entered

the war a few days later on the side of its entente partners, France and Russia, after Germany infringement of Belgium neutrality.

Germany's plan was to strike at France first and then turn on the more slowly mobilizing Russia. However, a stalemate on the Western front saw Germany switch forces to the east where mobile warfare was still possible. A series of severe reversals resulted for Russia. By 1917 shortages of basic commodities precipitated a series of strikes and protests known as the February revolution. This resulted in the abdication of Tsar Nicholas II on March 2, 1917[g] in favor of his brother Michael, overlooking his hemophiliac son Alexei. Michael refused the throne and an interim democratic provisional government was formed. This government, loyal to the agreement not to make a separate peace, continued the war with Germany.

The Bolsheviks seized power seven months later between October 24 and 26, 1917 and shortly afterwards called for an armistice with Germany, effectively ending Russia's part in the First World War. However, Germany seized further substantial territory in a virtually unopposed offensive in February 1918. The Bolsheviks eventually signed a peace treaty at Germany's dictation in March 1918, the treaty of Brest-Litovsk. With this treaty, Russia surrendered Poland, the Ukraine, the Baltic provinces, Finland, and much of the Caucasus. Germany turned its forces west in an attempt to defeat Britain and France before the millions-strong American forces could be carried across the Atlantic.

The Russian royal family had been arrested by the Revolutionary Government of Russia after the abdication. They were held prisoner initially at Tsarskoe Selo, the royal residence near St. Petersburg, and then later transferred to Tobolsk in Siberia. Present in this internment were Tsar Nicholas II, Tsarina Alexandra, their son the Tsarevich Alexei, their four daughters, the Grand Duchesses Olga, Tatiana, Maria, and Anastasia, Demidova the maid, Aleksei Trupp the valet, Ivan Kharitonov the butler, and the family doctor Eugeny Botkin. The family was in good health except for the Tsarevich, who was hemophiliac and often sickly from bruising and subsequent blood poisoning, and the Tsarina, who had weak legs and could not stand for more than about five minutes.

During this imprisonment at Tobolsk, many of the family jewels were placed in the safety of the Tobolsk Monastery. In about 1933 they were found and confiscated. Present in this cache were 154 objects with a total value of 3,270,693 rubles. Included among them was a 100 carat diamond brooch, three hat pins (44 and 36 carats [*sic*]), a diamond crescent (70 carats,

[g] Possibly March 3. The instrument of abdication may have been signed after midnight but the Tsar was asked to "pre-date" it to 3:00 p.m. to avoid the implication that he had been forced to sign by Guchkov and Shulgin, the emissaries from Petrograd, who had arrived late in the afternoon.

reputedly a gift to the Tsar from the Turkish sultan), four diadems of the Tsarina, and other precious items. In early 1918 the family was moved to their final home at Ipatiev House in Ekaterinburg. By this time, Russia had sunk into a brutal civil war. In many places, White (Monarchist) forces were driving back the Red forces of the Bolshevik government. Ekaterinburg, the capital of the Red Urals, was threatened and expected to fall.

An order for the execution of the Royal family was signed by the Ural Soviet but had been initiated and confirmed by Moscow. The motive for the murders at this time, as opposed to some other time, may have been the approach of White Russian forces on Ekaterinburg.

Equally, the counter argument points out that the Tsarina and the four Grand Duchesses were German Princesses. It is therefore unlikely that the Bolsheviks would have wanted to affront the German forces with whom they had recently signed a peace treaty by killing German women and, worse, German aristocrats. Recall that the First World War had been started by the regicide of Franz Ferdinand, the nephew and heir to Emperor Franz Josef of Austria. In July 1918, Germany, although seriously pressed in the west, was still far from defeated.

One version of the deaths of the prisoners[631] is based, in part, on the "Yurovsky note." This is a typed note corrected in handwriting that has been identified as Yakov Yurovsky's, the head of the execution squad. At the end of this note is a handwritten description of the position of the grave. This note had allegedly been given to a historian in 1920 but Radzinsky[631] concluded that it had also been sent as an official report, hence the reason he (Radzinsky) could obtain it from the official files when they were declassified in 1989.

The Yurovsky note suggests that the Royal family and their four servants were awakened shortly after midnight on July 16, 1918.[h] They washed and dressed and were then led into the half-cellar of the Ipatiev house. They may have been lured to the cellar, unsuspectingly, by a ruse of danger to themselves from "unrest in the town." The room had been chosen because it had a plastered wooden partition that would not cause ricochets. In the next room was the detachment to carry out the execution. His father Nicholas carried the heir, Alexei. The Tsarina asked for chairs and two were brought. Alexei and the Tsarina Alexandra were seated. The rest of the prisoners were ordered to stand in two rows. In the front was the Royal family, at the back the servants and the doctor. In the room were the eleven intended victims, one dog, and twelve executioners.

When all was arranged, the detachment was called in. Yurovsky announced that "In view of the fact that your relatives are continuing their attack on Soviet Russia, the Ural Executive Committee has decided to execute

[h] Again, these events probably happened in the early hours of the 17th.

you." Nicholas purportedly asked him to repeat his words. The Tsar's last words were allegedly "You know not what you do."

Their Bolshevik murderers then shot the Royal family, their loyal servants, and the pet dog of one of the Grand Duchesses. The Tsar and Tsarina died immediately, as well as the servants Trupp and Kharitonov. Various reports have the four Grand Duchesses, the heir, Demidova the maid, and maybe Dr. Botkin surviving the first fusillade. The bullets "bounced off" the Grand Duchesses.

> The Tsar's two youngest daughters, pressed up against the wall, were squatting, covering their heads with their arms, and then two men fired at their heads .... Alexei was lying on the floor and they fired at him too. The lady-in-waiting [Demidova] was lying on the floor still alive. Then I ran into the execution room and shouted to stop the firing and finish off those still alive with bayonets. One of the comrades began plunging the bayonet of his American Winchester into her chest. The bayonet was like a dagger, but it was dull and would not penetrate. She grabbed the bayonet with both hands and began screaming. Later they got her with their rifle butts. (in the testimony of one of the executioners Kabanov quoted in Radzinsky)

The bodies were then carried out the front door to a waiting truck:

> When they laid one of the daughters on the stretcher, she cried out and covered her face with her arm. The others [the daughters] also turned out to be alive. We couldn't shoot anymore — with the open doors the shots could have been heard on the street .... Ermakov took my bayonet from me and started stabbing everyone dead who turned out to be alive. (in the testimony of Strekotin, another of the executioners quoted in Radzinsky)

> Yurovsky: "When they tried to stab one of the girls with a bayonet, the point just would not go through her corset."

Subsequently it was learnt that the Grand Duchesses had diamonds sewn into their corsets. This had deflected the bullets and blows.

The execution squad was under orders from the Ural Red Commissioner Philip Goloschokin to destroy or hide the bodies. The Russian Revolutionary Government's official announcement read that Nicholas had been preparing to escape and was accordingly shot. His family had been removed to a safe place.

Eight days later, on the 25th of July, the Bolsheviks surrendered Ekaterinburg to the Czech Legion and the Siberian White Army.

The official version of the events of the execution and burial follows largely from the account of Nikolai Sokolov, the official investigator appointed by the White Russians. The room in which the murders had occurred had been cleaned, but ample evidence remained of the violence. There were traces of washed blood, bullet holes fanned across two walls, and there were two bullet holes in the floor. The floor had dents possibly from bayonet blows. The bullets in the room were from a revolver, a Colt and a Mauser.

Traces of two bonfires were found, one by a nameless mine and another on a forest road. One of these fires contained charred human bones, a charred emerald cross, topaz beads, a child's military buckle, an eyeglass lens, buttons, hooks, and a large diamond (supposedly the 12 carat diamond given to the Tsarina by the Tsar). The buttons, hooks, and shoe buckles were similar to those from the Ipatiev house.

The nameless mine was full of water. Fresh branches and burned wood were floating on the surface and there was evidence that grenades had been thrown down the mine. Horse hooves and carts had trampled the area about the mine. When pumped dry, an amputated manicured finger with a long nail, Dr. Botkin's false teeth, his tie clasp, and the Tsarina's pearl earring were found. Also present was the body of a tiny dog, and photographs and icons attributable to the Royal family. No human bodies were found.

The truck used to transport the bodies was located. The back had been wiped but there were still visible traces of blood.

The Sokolov investigation concluded that the Tsar, his family, and four servants had been shot at Ipatiev house. The corpses were taken by truck to the unnamed mine. On July 18, a large quantity of gasoline and sulfuric acid was brought to the site. The bodies of the slain were chopped up with axes, doused with gasoline and sulfuric acid, and burned in bonfires. Sokolov never did find the bodies of the Tsar and his family.

Subsequent reconstruction of physical evidence and testimony suggests the following as a more plausible version of events. It is based, in part again, on the "Yurovsky note." The truck loaded with the bodies left Ekaterinburg. There was an incident where the truck overheated and the driver went for water. Having traveled about 5 versts (3.3 miles) from this point, the truck met a detachment of Red Guards who were expecting it. The bodies were transferred to carts. It was at this point that the "special corsets" containing diamonds were discovered on the Grand Duchesses. Yurovsky decided to strip the bodies at the burial site and had to threaten the crowd with a firing squad to avoid looting of the bodies immediately. Not all the bodies fitted on the carts, so the truck also continued.

The Red Guard apparently could not find the mine, the position of which they had known well the previous day. Yurovsky suspected that they simply

wanted to be alone with the "special corsets." Eventually they arrived at the first turn off the road to the mine, whereupon the truck finally broke down (or got stuck between trees, depending on the version). The bodies were moved to the mine on the remaining carts and stretchers.

By the mine the commandant ordered the bodies to be undressed and the clothes burnt. The corsets with the sewn-in diamonds were removed from the Grand Duchesses and the pearl belt made from several necklaces and sewn into linen was removed from the Tsarina. About half a pood (18 pounds) of diamonds were collected. Each of the girls was wearing a picture of Rasputin. The remaining clothing was burnt and the naked bodies were thrown into the mineshaft (by one account about $3\frac{1}{2}$ arshins deep (8 feet) with one arshin of water). In an attempt to collapse the mineshaft, Yurovsky tossed in some hand grenades.

However, an account of where and how the bodies of the Tsar and his family were hidden was circulating widely by the 17th. It was decided to recover the bodies and move them to another location. Time was short as the Whites were approaching.

By midnight on the 18th, the burial squad was back at the nameless mine. The mine was lit with torches and the bodies were hauled up with ropes. The bodies were again loaded onto carts, which proved unequal to the load and were unstable and falling apart. Another truck and two cars were summoned. The bodies were loaded onto the truck. Again, this truck became stuck in difficult ground.

Two of the bodies, described as Alexei and Demidova by Yurovsky, were burnt next to the truck. A common grave was dug for the remaining victims. It was $2\frac{1}{2}$ arshins deep and $3\frac{1}{2}$ arshins square. The bodies were placed into the hole and sulfuric acid was doused over their faces. Dirt and lime were scattered, boards placed on top, and then the truck was driven over the grave a few times. Yurovsky leaves a detailed account of the position of this grave.

In July 1991 President Boris Yeltsin authorized the exhumation of nine skeletons from a shallow grave approximately 20 miles from Ekaterinburg, in the Central Urals, Russia. This grave had been reported in a number of newspaper articles in 1989 by two amateur historians,[i] Gely Ryabov and Alexander Avdonin. Ryabov, Avdonin, and others may have located the site much earlier, possibly in 1979. They had retrieved three skulls only to "return" them to the ground fearfully a year later.[240,j]

[i] Gill et al.[351] use the term "amateur historian." Zhivotovsky[883] points out that Gely Ryabov was a militia Colonel and an operative of the Internal Minister of the U.S.S.R. Radzinsky[631] reports a conversation that suggests three geologists and a writer "found" the grave.

[j] Zhivotovsky introduces the possibility that the gravesite had been "recreated" in 1919 when the Reds recaptured Ekaterinburg. He also gives evidence that the grave was known and possibly opened in the years intervening between 1919 and 1979. The possibility is also raised that the "returned" skulls and bones were actually planted.

When examined by the Chief Forensic Medical Examiner of the Russian Federation, the grave was found to be a shallow pit one meter deep. There were signs that the grave was not completely undisturbed. A cable had been laid across one corner, suggesting that engineering work had been carried out at the site at some earlier date. Three of the skulls were in a box and some bones were in plastic bags. These may have been the skulls and bones "returned" by Ryabov and Avdonin. The skeletons showed signs of gunshot and bayonet marks, and destruction of the facial features. Many skeletons were represented by fewer than 50% of the bones. The gold, platinum, and porcelain dental work suggested that this group were aristocratic. Bullets in the grave were consistent with the revolvers used at the execution.[337] I have been unable to confirm whether the position of the grave corresponds exactly with Yurovsky's description.

Confirmation of the identity of the skeletons was attempted by the then novel DNA technology in a joint Russian/U.K. project. STR typing of the remains[351] at five loci together with amelogenin suggested that a family group was present, putatively assigned as the Tsar, Tsarina, and three of his four daughters. The mtDNA match of the three putative Grand Duchesses to the putative Tsarina supported the suggestion that this was a family group. For an interesting and comprehensive modern probabilistic reanalysis, see Egeland et al.[246] or Cowell and Mostad[202]. For an alternative view, see Zhivotovsky[883] or Knight et al.[470] This latter paper requires a number of unlikely events to have all occurred.

Hence putatively one daughter and the Tsarevich were missing. This would support some historical accounts that two of the bodies were either burned or buried separately. Notably, Yurovsky had stated that he had burnt Alexei and a female (he stated this to be Demidova) separately. Using photographic superimposition Russian scientists have concluded that the missing bodies are Alexei and Marie.[240] Dr. William Maples, using dental and bone specimens, believes them to be Alexei and Anastasia.[240,369]

The putative Dr. Botkin's genotype could also be compared to a living grandson.[k] Mrs. Schweitzer is also a descendant of Dr. Botkin, but I am unaware of any DNA comparison to her.[240]

The task of determining whether the family group was, indeed, the Russian Royal family was initially undertaken by comparison of the mtDNA from the Tsar and the Tsarina with living relatives (see Figures 9.3 and 9.4). The putative Tsarina matched exactly to Prince Phillip. The putative Tsar was matched to James Duke of Fife and Countess Xenia Cheremeteff-Sfiri at all except one nucleotide, that being at position 16,169. At this position, the Tsar was heteroplasmic for C and T in the ratio 72:28,

[k] There appears to be no published account that confirms this.

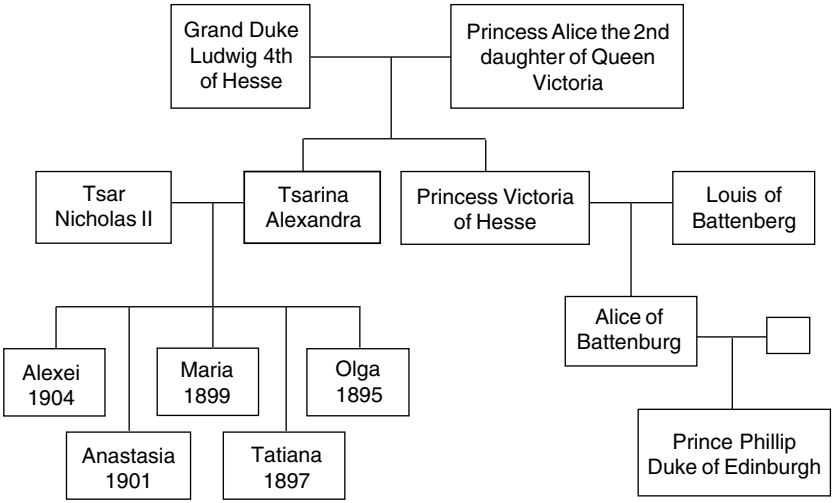


Figure 9.3 Pedigree including Tsarina Alexandra.

whereas James Duke of Fife and Countess Xenia Cheremeteff-Sfiri are both homoplasmic for T.[l]

Subsequently, the Tsar's brother Georgij, who had died of tuberculosis in 1899,[225,226,430] was exhumed from SS Peter and Paul Cathedral[m] in St. Petersburg. The mtDNA sequence of Georgij matched the putative Tsar at all positions, including the heteroplasmy at 16,169. At this position, Georgij showed a C to T ratio of 38:62.[n]

The 70-year-old controversy surrounding the identity of Anna Anderson Manahan, perhaps the most famous Anastasia claimant, has also been examined by forensic means.

In 1920 a woman, who had attempted suicide, was pulled from a canal in Berlin and taken to a clinic. Initially she called herself Tatiana, but later she began referring to herself as Anastasia. She bore numerous scars on her body, consistent with her story of being bayoneted before being left for dead in the cellar where the rest of her family were murdered.[90] She spoke Russian reluctantly, but German fluently. Russian was not the usual language in the Russian Royal household, French or German being preferred; however, her failure to speak Russian was unusual. It was explained by the trauma through which she had passed. She had a physical similarity to the Tsar's daughter and

[l] One alternative candidate is the Tsar's brother Grand Duke Mikhail who had been murdered in the nearby city of Perm and whose body had disappeared.
[m] Often referred to at the Fortress of SS Peter and Paul
[n] This examination also gives evidence on the progression of mtDNA from heteroplasmic to homoplasmic states.

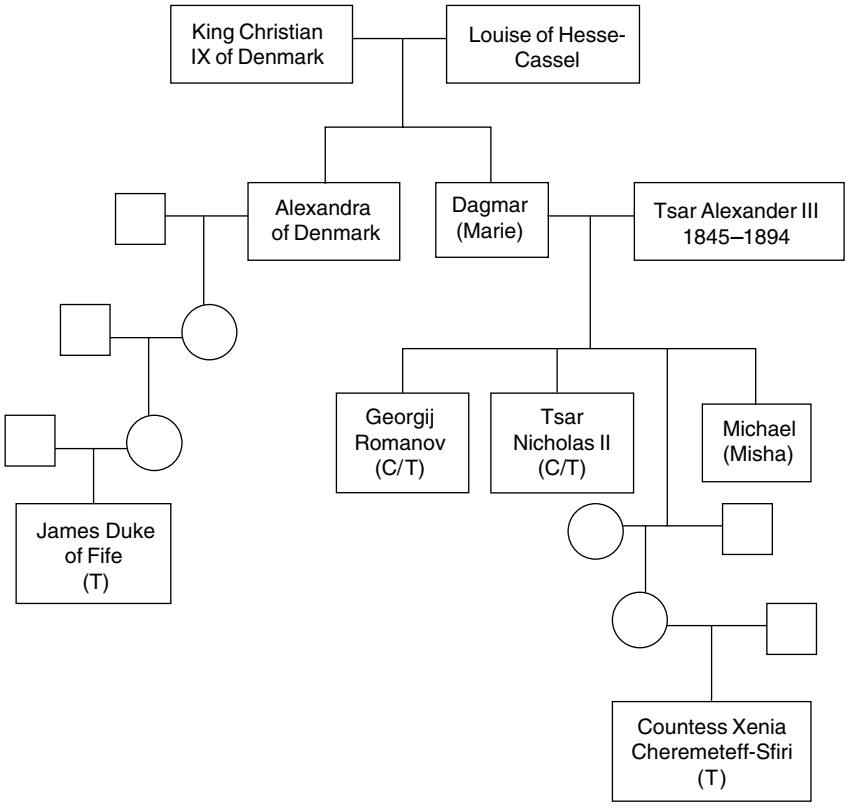


Figure 9.4 Pedigree including Tsar Nicholas II.

a trace of a birthmark where one had been removed from the young Anastasia. Her ear shape and handwriting were similar to those of the Grand Duchess. This woman, known as Anna Anderson, moved to the U.S. She later married Jack Manahan, becoming known as Anna Anderson Manahan. In 1970 she lost a court case aimed at getting her recognized as the Tsar's heiress. She died in 1984 and was cremated. Her ashes were buried in the crypt of the Romanovs' relatives, the Princes of Leuchtenberg.

A small bowel biopsy sample that had been removed as part of an examination for suspected cancer prior to her death was known to be held by a Charlottesville hospital. After some legal arguments, this sample was released for analysis.[196,240] In addition, six hairs said to have come from Anna Anderson had been kept by her husband as mementos in a book. These were found by an amateur historian, Susan Burkhart, who had been given access to Anna Anderson Manahan's estate.[240] DNA extracted from these samples was eventually analyzed by three differing laboratories. It was compared to

the skeletal remains of the Tsar and Tsarina and their living relatives at five STRs and by mtDNA typing. Four of the five STRs were inconsistent with Mrs. Manahan being a child of the Tsar and Tsarina. Six differences were found in the mtDNA between the small bowel sample and the hair samples (stated) to be from Mrs. Manahan and the Duke of Edinburgh[337,346] (see also Debenham[226] and Schweitzer[700]).

The evidence supports an alternative suggestion that Mrs. Manahan is Franzisca Schanzkowska (we follow Gill et al.[346]). Franzisca Schanzkowska was born about 1896 and lived in Pomerania in north Germany adjacent to the Polish border. During the First World War she worked in a Berlin munitions factory and was injured in an explosion. Subsequently, she was admitted to two different mental institutions, but disappeared in 1920 about the same time that Anna Anderson appeared and claimed to be the Grand Duchess Anastasia. The samples (stated to be) from Mrs. Manahan do match Carl Maucher, the great nephew of Schanzkowska and directly related through the maternal line. This sequence does not appear in a set of 300 searched by Gill et al.

On July 17, 1998, the 80th anniversary of their deaths, the Tsar, his wife, three daughters, and four of his loyal servants who would not leave him were reburied at the Cathedral of SS Peter and Paul in St. Petersburg.

Parentage Testing 10

JOHN BUCKLETON, TIM CLAYTON AND CHRISTOPHER TRIGGS

Contents

10.1 Introduction

Familial investigation features prominently in forensic science within both the criminal and civil jurisdictions.[558,772] This chapter reviews the application of autosomal STR evidence to parentage testing and discusses the issues associated with mutation, null alleles, and genetic anomalies. Use of single-locus probe data or multilocus probe data follows similar principles.[269] Application of both X- and Y-linked markers[750] to these purposes is also briefly reviewed. This chapter concentrates on pedigrees involving one or two alleged parents and one or more children. In a later chapter we will consider the application of familial testing to more complex pedigrees.

In a criminal context, such testing can be required following sexual assaults, for example, to identify the father of a child conceived as a result of an alleged assault. Likewise, familial testing can be used to confirm that two individuals are genetically related in order to support charges arising from entrance into a proscribed (incestuous) sexual relationship. In cases involving concealed births, abandoned children, or infanticide, it may be necessary to prove a genetic relationship to either ensure the rightful return of an infant or to support criminal charges.

During civil litigation familial investigation can be used to substantiate claims by an estranged partner for financial support and maintenance of a child. Similarly, in the field of wills and probate, disputes over inheritances can be informed by the application of genetic testing. Familial testing is also now being widely applied by governmental bodies to adjudicate in cases of immigration and naturalization. The identification of bodies for legal purposes can also be effected using familial testing.

As paternity and familial identification can provide evidence in either civil or criminal proceedings, a forensic scientist has a different responsibility in these two settings. In criminal cases, it is customary to concede reasonable doubt to the defendant. However, in civil cases it is not always obvious as to which way to concede any reasonable doubt. This may affect the method used to evaluate the evidence.

10.1.1 Testing Diverse Sample Types

Other than those samples taken from bodily remains, most samples for testing paternity will be pristine reference samples. These are presented as venous blood samples, samples of buccal mucosa (in the form of scrapes and oral rinses), or samples of plucked hair. Consequently, these types of samples present few difficulties during the extraction of DNA. However, a number of sample types that do pose significant technical challenges are sometimes encountered.

10.1.1.1 Termination Products

In many countries it is common for the scientist to be presented with samples from a pregnancy termination procedure. In the United Kingdom, terminations can be performed legally up to the 24th week of gestation. Clinically, those terminations performed up to about the 12th week of gestation use a technique known as vacuum aspiration. This results in the severe fragmentation of the fetus and subsequent mixing of those fragments with maternal tissues. Care must be taken to identify fetal structures among the largely maternal tissues in the termination products in these cases. The earlier in the gestation period the termination, the more technically difficult this is to achieve. Failure to identify sufficient fetal tissue will result in a solely maternal profile being obtained. Alternatively, if both maternal and fetal tissues are present, a mixture will be obtained in which the ratio of the mixture will be governed by the relative proportions of maternal and fetal cells (see Chapter 7). There is a maximum of three bands per locus due to the maternal relationship. In theory (if the maternal profile is known), such a mixture can be substantially resolved, using the principles outlined in Chapter 7, to yield the fetal profile.

10.1.1.2 Preserved Histology Samples

In some criminal cases, the complaints are retrospective and the only samples from a fetus may be in the form of archival histology samples. The most common form of tissue preservation used by histopathology laboratories is a procedure that involves a transient immersion of the tissue in formol saline followed by embedding in blocks of molten paraffin wax. This presents additional technical difficulties. First, sectioning from the block is required to produce thin slices with large surface areas to facilitate efficient removal of

the wax. Second, the wax in each slice must be dissolved using xylene. The xylene must then be removed by washing with ethanol. The tissue is then air dried to volatilize the ethanol remnants before digestion of the nascent tissue can begin. Third, the formalin used as preservative is known to have effects deleterious to the DNA. Depending upon the severity of the fixation treatment prior to embedding, the DNA can be rendered unamplifiable. Similarly, "wet" preserved samples stored in formalin solution have been demonstrated to be highly refractory to PCR analysis.

10.1.2 Principles of Mendelian inheritance

> The year 1900 marks the beginning of the modern period in the study of heredity. Despite the fact that there had been some development of the idea that a living organism is an aggregation of characters in (the) form (of) units of some description, there had been no attempts to ascertain by experiment, how such supposed units might behave in the offspring of a cross. In the year above mentioned the papers of Gregor Mendel came to light, being quoted almost simultaneously in the scientific correspondence of three European botanists, de Vries in Holland, Correns in Germany, and von Tschermak in Austria. Of Mendel's two papers, the important one in this connection, is entitled "Experiments in Plant Hybridisation," and was read at the meetings of the Natural History Society of Bruun in Bohemia at the sessions of February 8 and March 8, 1865. This paper had passed entirely unnoticed by the scientific circles of Europe, although it appeared in 1866 in the Transactions of the Society. From its publication until 1900, Mendel's paper appears to have been completely overlooked, except for the citation in Focke's "Pflanzenmischlinge." And the single citation of Hoffmann . . . .[650]

Two laws of heredity have been developed from Mendel's work. In modern times they are often phrased with the benefit of hindsight. We now know the chromosomal basis of inheritance associated with meiosis. However, at the time that Mendel wrote, none of this was known. An elegant phrasing of Mendel's laws without overreliance on modern terminology is given by Thompson.[763] We follow her treatment here:

The law of segregation. Each individual has two "factors" controlling a given characteristic, one being a copy of a corresponding factor in the father of the individual and one being a copy of the corresponding factor in the mother of the individual. Further, a copy of a randomly selected one of the two factors is copied to each child, independently for different children and independently of the factor contributed by the spouse.

The law of independent assortment. The factor copied from one pair is independent of the factor copied from another factor pair.

Modern molecular biology allows us to see the basis for these laws in the segregation of chromosomes and their recombination into a zygote. The human genome is diploid. It has a normal complement of 46 chromosomes arranged into 22 pairs of autosomes and a single pair of sex chromosomes (XY). The somatic cells divide mitotically to maintain their diploid status whereas the sex cells (gametes) are produced by meiotic divisions and are haploid. During meiosis, one of each of the pairs of the homologous chromosomes is randomly partitioned to the ovum or spermatozoon. In addition, there are recombination events that "shuffle" the genetic material further still. At fertilization, the union of an ovum and a single spermatozoon restores the diploid chromosomal constitution, and in doing so ensures that the embryo receives a random assortment of genes, half provided by one biological parent and the remaining half from the other biological parent (see Figure 10.1). Mendel's laws form the basis of familial testing.

An exception to the usual Mendelian inheritance pattern occurs for loci that are physically close on the same chromosome. If two loci are close enough, they tend to be inherited together. This phenomenon was discussed in Chapter 3 and is known as linkage.

Typically, the STR loci selected as forensic markers for familial analysis are situated on different chromosomes to ensure that they assort randomly and are inherited in a Mendelian fashion (the law of independent assortment). However, there are cases, for instance in the CODIS set, where two loci are

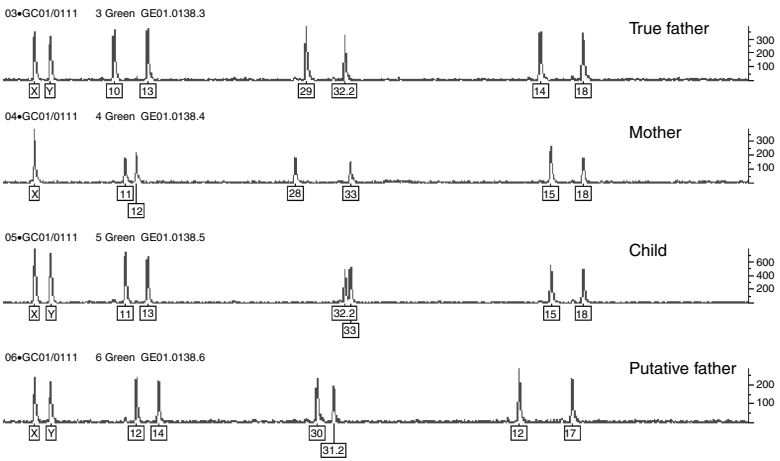


Figure 10.1 Profiles of mother, child, the true father, and a putative father at four autosomal STR loci.

resident on the same chromosome. As long as they are separated by sufficient genetic distance to guarantee an intervening recombination event, there are no potential consequences. As discussed in Chapter 1, there are some pairs of loci that are close enough that some linkage effects are expected. This situation will become more frequent as more loci are added to forensic sets, or when we move to SNPs. We include a small section later in this chapter on the consequences of linkage, which has not been extensively reported in the forensic literature.

The Mendelian inheritance pattern is most easily represented in the format of a family tree (Figure 10.2). If one locus is considered and the two paternal alleles are represented as P_1, P_2 and the maternal alleles as M_1, M_2, then there are four combinations for the offspring: P_1M_1, P_1M_2, P_2M_1, and P_2M_2. By Mendel's first law, each combination is equally probable, occurring with probability $\frac{1}{4}$.

More recently, the nonautosomal DNA in the mitochondrion and the X and Y chromosomes have begun to play an important part in familial testing. For most of its length, the Y chromosome does not recombine and is patrilinearly inherited. Y STR markers therefore form a "haplotype" that, barring mutation, will be passed down the male line. This pattern of inheritance can be useful in certain cases where there is a shortage of relatives (so-called deficiency cases) or where there are large generational gaps.

The X chromosome can recombine in the female but not the male. As discussed in Chapter 9, individual loci of forensic interest may be linked. The major advantages of markers on the X chromosome again arise in deficient paternity cases, when a biological sample is not available from the putative father and samples from paternal relatives are used instead. When females have the same father, they also share the same paternal X chromosome.[750]

Mitochondrial DNA was also discussed in Chapter 9. It is assumed to be exclusively inherited matrilineally with no recombination in mammals. Its uses strongly mirror those of the Y chromosome for paternity cases, especially when there is a shortage of relatives.

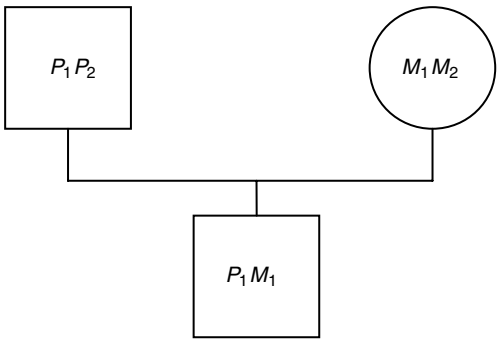


Figure 10.2 Pedigree for simple paternity trio.

10.2 Evaluation of Evidence

Three methods have been offered for the evaluation of parentage testing results. These are often termed the paternity index (*PI*), the probability of paternity, and an exclusion probability.[653,803] All these have strong parallels to concepts used elsewhere in this book. Every interpretation method assumes correct typing of the biological samples. However, Hallenberg and Morling[385] report typing error rates of 0.3 and 0.1% in surveys from 2000 and 2001, respectively.

Strong support is given for the paternity index approach by many authorities, including Evett and Weir[267] and the Paternity Testing Commission of the International Society of Forensic Genetics.[559]

All three methods use the frequency of certain alleles. These frequencies must be estimated from a relevant database as has been discussed in Chapter 5. A "rare" allele could be an allele that has not previously been reported or is not contained in the relevant database, or an allele with very few previous observations. Methods for dealing with "rare" alleles have been discussed in a general context in Chapter 6. Aside from not undertaking a calculation at all (5% of surveyed laboratories), three methods have been reported by Hallenberg and Morling[385] as being used to deal with rare alleles. For an allele previously unseen, these were:

- a minimum allele probability (50% of responding laboratories),
- $1/N$, where N is the number of alleles in the database (32%), and
- the alleles observed in the present case added to the database (9%).

For an allele previously seen once, these were:

- a minimum allele probability (50% of responding laboratories),
- $1/N$, where N is the number of alleles in the database (45%), and
- the alleles observed in the present case added to the database (5%).

All of these methods will be perfectly satisfactory in practice, but none have a firm statistical backing. There are more robust and sophisticated methods available (see Chapter 6), of which the Bayesian posterior method attributed to Painter (see Weir et al.[848] and Curran et al.[217]) is likely to be superior for rare alleles.

Before our detailed discussion of the paternity index, we will briefly consider the other two approaches.

10.2.1 Exclusion Probability

Consider the most common case of parentage testing where we have a mother (*M*), child (*C*), and a man alleged to be the father (*AF*). These three

persons have been typed and found to have the genotypes G_M, G_C, G_{AF}, respectively. The genotypes of the mother and the child define one (or in some cases one of two) paternal allele(s) at each locus.

An exclusion probability may be defined as "that fraction of men who do not possess the paternal allele or alleles." As such, it is strongly akin to the exclusion probability in mixtures evaluation.

If the possible paternal alleles at a locus are $A_1 \ldots A_n$ (often there is only one possible paternal allele), then the exclusion probability at this locus (PE_l) is

$$PE_l = \left(1 - \sum_{i=1}^{n} \Pr(A_i)\right)^2$$

assuming Hardy–Weinberg equilibrium. The exclusion probability across multiple loci (PE) is calculated as

$$PE = 1 - \prod_l (1 - PE_l)$$

For an extension to the consideration of relatives, see Fung et al.[325]

We have previously discussed Dr. Charles Brenner's[81] explanation of the shortcomings of the probability of exclusion. We follow his treatment again here.

Let us describe the evidence as:

1. Blood type of the mother.
2. Blood type of the child.
3. Blood type of the alleged father.

From this information, we can infer that:

4. The alleged father is not excluded.

Brenner points out that although statement 4 can be deduced from statements 1, 2, and 3, statement 3 cannot be deduced from 1, 2, and 4. Hence, the use of statement 4 represents a loss of information. The exclusion probability is a summary of the evidence in 1, 2, and 4.

10.2.2 Paternity Index

The paternity index is a specialist term used in paternity testing to describe the likelihood ratio. Its structure is exactly as described for the likelihood ratio in Chapter 2, but has been used in paternity testing for longer than in other areas of forensic biology.[255] Hallenberg and Morling[385] reported that 73% of respondents in the year 2000 and 78% in 2001 used the paternity

index or the probability of paternity to interpret parentage evidence. Consider the two hypotheses:

H_p: The alleged father is the true father.
H_d: The alleged father is not the true father.

Hypothesis H_p represents one side of the allegation. In many paternity cases, the action will be civil and it may not be appropriate to view this as the "prosecution" hypothesis. Fortunately the same letter can stand for "paternity." Hypothesis H_d represents the other side of the allegation; similarly it may not be appropriate to view this as the "defense" hypothesis.

If we consider some evidence, E, typically the genotypes of a child, the alleged father, and possibly the mother, then Bayes's theorem informs us that

$$\frac{\Pr(H_p|E)}{\Pr(H_d|E)} = \frac{\Pr(E|H_p)}{\Pr(E|H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)}$$

The likelihood ratio term

$$\frac{(\Pr(E|H_p)}{\Pr(E|H_d)}$$

is usually written as PI and is the central term calculated under this approach.

10.2.2.1 Use of the Product Rule in the Evaluation of the Paternity Index

In the evaluation of the likelihood ratio in previous chapters, we have discussed the small bias inherent in the use of the product rule when population substructure exists. The method of Balding and Nichols[36] can be used to evaluate likelihood ratios, or paternity indices, for paternity duos and trios when population substructure exists.

When the Balding and Nichols's correction is applied to a whole race or when conservatively large values of θ are used, this is thought to be an overcorrection that may err too much in one direction. This "conservative" behavior is considered desirable by some courts and scientists in criminal cases. However, this property of the subpopulation correction does not have such an obvious justification in civil cases. Of course, this is not to suggest that the population genetic fact of population subdivision has disappeared; however, in civil proceedings the product rule may have more desirable properties. NRC II discusses this matter briefly in the U.S. context in footnote 74.

If we assume independence, we can write $\Pr(A_p|G_M, G_{AF}, H_d)$ as $\Pr(A_p|H_d)$, which is the unconditional allele probability of the allele A_p. This assumption is similar to using the product rule. Without this assumption a slightly more complex approach is suggested, similar to the θ-based approaches discussed in Chapters 3, 5, and 7 on population genetic models and mixtures.

10.2.3 Probability of Paternity

Recall Bayes's theorem that states

$$\frac{\Pr(H_p|E)}{\Pr(H_d|E)} = PI \times \frac{\Pr(H_p)}{\Pr(H_d)}$$

We see that the paternity index relates the odds on paternity prior to considering the genetic evidence to those after considering that evidence. As with any Bayesian treatment, the posterior probability of paternity can be calculated from the paternity index and the prior odds. The prior odds relate to the probability of paternity based on the nongenetic evidence. This could include statements of the mother as to with whom she had intercourse, or evidence that may suggest that the alleged father was out of the country or in prison at the time of conception. Such evidence, if relevant and admissible, affects the prior odds.

However, it has become customary to set the prior odd to 1:1, that is, to assign prior probabilities of 50% to both H_p and H_d, when calculating the probability of paternity. This assumption is hard to justify at the fundamental level (Good[362] (pp. 68, 89–91); Robertson and Vignaux[659]) and must be seen simply as a pragmatic tool. It may be completely appropriate in many cases but equally may be totally inappropriate in others. It would seem wise, however, to make this assumption of equal prior odds explicit.

Utilizing this assumption we see that

$$\frac{\Pr(H_p|E)}{\Pr(H_d|E)} = PI$$

and hence that

$$\frac{\Pr(H_p|E)}{1 - \Pr(H_p|E)} = PI$$

yielding

$$\Pr(H_p|E) = \frac{PI}{1 + PI}$$

Given the assumption of prior odds of 1:1, we can simply tabulate the posterior probability of paternity given the paternity index (Table 10.1).

Hallenberg and Morling[385] report the requirements from a number of laboratories for issuing a report with a positive weight for paternity (see Table 10.2). There is no theoretical requirement for a lower limit for reporting, and a decision to have such a limit is based on pragmatic reasons such as avoiding court cases where the biological evidence is evaluated as less than the limit.

We (and others) cannot support the assumption of prior odds despite its extensive use and rather advocate use of the *PI* alone.[56,653] This stance is taken

Table 10.1 Posterior Probability of Paternity Given the Assumption of Equal Prior Odds (Rounded to Seven Significant Figures)

Paternity Index	Probability of Paternity
10	0.9090909
100	0.9900990
1000	0.9990010
10,000	0.9999000
100,000	0.9999900
1,000,000	0.9999990
10,000,000	0.9999999

by the Paternity Testing Commission of the International Society of Forensic Genetics:

> If the weight of the evidence is calculated, it shall be based on likelihood ratio principles. The paternity index, *PI*, is a likelihood ratio.[560]

It is also worthwhile noting that Birus et al.[69] noted two false inclusions with a *PI* over 10,000 while examining war victims in Croatia. In such cases, the assignment of prior odds of 1:1 may have led to a very erroneous view of the evidence.

Specific recommendations for the number and type of loci used were not made by the Paternity Testing Commission of the International Society of Forensic Genetics[557] and this choice is made by each laboratory.

10.2.4 Paternity Trios: Mother, Child, and Alleged Father

We begin by considering at least two hypotheses. In the most common case, these could be:

H_p: The alleged father is the true father (and the mother is the true mother).
H_d: A random person who is not related to the alleged father is the true father (and the mother is the true mother).

Table 10.2 Requirements for Issuing a Report with a Positive Weight for Paternity[385]

Paternity Index	Probability of Paternity	2000 (N=33)%	2001 (N=36)%
100–1000	99–99.9%	24	19
1000–10,000	99.9–99.99%	33	22
10,000–100,000	99.99–99.999%	21	25
>100,000	>99.999%	3	8
Fewer than a certain number of inconsistencies		9	11
No requirement		9	14

The assumption that the person labeled as the mother is the true mother of the child is usually unstated. Although these two hypotheses are the most commonly used, we note that they are not exhaustive as the random person may be a relative of the alleged father. This again suggests an alternative approach based on the general form of Bayes's theorem. Such an approach is not in use in any laboratory of which we are aware.

Typically then we require

$$PI = \frac{\Pr(G_C, G_M, G_{AF}|H_p)}{\Pr(G_C, G_M, G_{AF}|H_d)}$$

It is customary to decompose these probabilities using the third law of probability. Usually to evaluate the probabilities of the observing genotypes of individuals, they are conditioned on the genotypes of their ancestors. For example,

$$PI = \frac{\Pr(G_C, G_M, G_{AF}|H_p)}{\Pr(G_C, G_M, G_{AF}|H_d)} = \frac{\Pr(G_C|G_M, G_{AF}, H_p)\Pr(G_M, G_{AF}|H_p)}{\Pr(G_C|G_M, G_{AF}, H_d)\Pr(G_M, G_{AF}|H_d)}$$

where the genotype of the youngest person, the child, is conditioned on the parents, as opposed to

$$PI = \frac{\Pr(G_C, G_M, G_{AF}|H_p)}{\Pr(G_C, G_M, G_{AF}|H_d)} = \frac{\Pr(G_{AF}|G_M, G_C, H_p)\Pr(G_M, G_C|H_p)}{\Pr(G_{AF}|G_M, G_C, H_d)\Pr(G_M, G_C|H_d)}$$

Both decompositions are, of course, formally equivalent mathematically. However, the former is easier to evaluate. Thus, we will work with the former decomposition.

It is customary to assume that the joint probability of observing the genotypes of the putative parents does not depend on the particular hypothesis, i.e.,

$$\Pr(G_M, G_{AF}|H_p) = \Pr(G_M, G_{AF}|H_d) = \Pr(G_M, G_{AF})$$

This assumption essentially states that the joint probability of observing the genotypes of the mother and alleged father are not conditioned on whether the alleged father is the true father or not. This is only true in the absence of any conditioning on the genotypes of any other children or descendants. Given this assumption, the paternity index becomes

$$PI = \frac{\Pr(G_C|G_M, G_{AF}, H_p)}{\Pr(G_C|G_M, G_{AF}, H_d)}$$

Evaluation of the *PI* can proceed directly from this equation. The numerator can be evaluated using a Punnett square at each locus where both parents are present in the conditioning.

Assuming that the mother is the true mother, it is often possible to determine the maternal and paternal alleles, A_m, and A_p, unambiguously. This allows us to write

$$\begin{aligned}\Pr(G_C|G_M, G_{AF}, H_d) &= \Pr(A_p, A_m|G_M, G_{AF}, H_d)\\ &= \Pr(A_m|G_M, G_{AF}, A_p, H_d)\Pr(A_p|G_M, G_{AF}, H_d)\end{aligned}$$

Conventionally, using the further assumption that

$\Pr(A_m|G_M, G_{AF}, A_p, H_d) = \Pr(A_m|G_M)$ allows the probability in the denominator of *PI* to be written as

$$\Pr(G_C|G_M, G_{AF}, H_d) = \Pr(A_m|G_M)\Pr(A_p|G_M, G_{AF}, H_d)$$

Now $\Pr(A_m|G_M)$ is $\frac{1}{2}$ or 1, depending on whether the genotype G_M containing the maternal allele is heterozygous or homozygous. We denote this probability as the maternal Mendelian factor M_M. Evaluation of $\Pr(A_p|G_M, G_{AF}, H_d)$ is slightly more problematic.

As with previous chapters, we now turn to consideration of a series of examples and show in detail how to evaluate *PI* for paternity trios.

Example 10.1

	Genotype
Mother	*cd*
Child	*ac*
Alleged father	*ab*

Under H_p we assume that the alleged father is the true father, and may proceed by using a Punnett square:

		Genes from the Father	
		a	*b*
Genes from the Mother	*c*	*ac*	*bc*
	d	*ad*	*bd*

We see that the child's genotype is one of the four (equiprobable) outcomes and assign the probability $\Pr(G_C|G_M, G_{AF}, H_p) = \frac{1}{4}$.

The mother is heterozygous for the maternal allele ($A_m = c$) and can assign the value $M_M = \frac{1}{2}$ to the maternal Mendelian factor. The paternal allele is $A_p = a$. Under the hypothesis H_d, we assign the probability $\Pr(A_p|G_M, G_{AF}, H_d) = p_a$, the allele probability of the *a* allele in this population. Hence the paternity index is

$$PI = \frac{\frac{1}{4}}{\frac{1}{2} \times p_a} = \frac{1}{2p_a}$$

Example 10.2

	Genotype
Mother	*cc*
Child	*ac*
Alleged father	*ab*

Again under H_p we assume that the alleged father is the true father, and the Punnett square becomes

		Genes from the Father	
		a	*b*
Genes from the Mother	*c*	*ac*	*bc*
	c	*ac*	*bc*

We see that the child's genotype occurs in two of the four (equiprobable) outcomes and assign the probability $\Pr(G_C|G_M, G_{AF}, H_p) = \frac{1}{2}$.

The mother is homozygous for the maternal allele ($A_m = c$) and we can assign $M_M = 1$. The paternal allele $A_p = a$. As before, we assign the probability $\Pr(A_p|G_M, G_{AF}, H_d) = p_a$ under the hypothesis H_d. Hence

$$PI = \frac{\frac{1}{2}}{1 \times p_a} = \frac{1}{2p_a}$$

Example 10.3

	Genotype
Mother	*ab*
Child	*ab*
Alleged father	*bc*

Under H_p we assume that the alleged father is the true father, and proceed by a Punnett square:

		Genes from the Father	
		b	*c*
Genes from the Mother	*a*	*ab*	*ac*
	b	*bb*	*bc*

We see that the child's genotype occurs in one of the four (equiprobable) outcomes and assign the probability $\frac{1}{4}$ to this genotype.

This example was introduced because of a small complexity that occurs under H_d. This arises because either of the mother's alleles may be the maternal allele, making attribution of both the maternal and the paternal allele

ambiguous. Under H_d we can see that the mother may contribute the a allele ($A_m = a$) with probability $M_M = \frac{1}{2}$ or the b allele ($A_m = b$) with probability $M_M = \frac{1}{2}$. If the maternal allele is $A_m = a$, then the paternal allele A_p must be b. If the maternal allele is $A_m = b$, then the paternal allele must be a. The denominator is therefore the sum of two terms. Hence

$$PI = \frac{\frac{1}{4}}{\frac{1}{2}p_a + \frac{1}{2}p_b} = \frac{1}{2(p_a + p_b)}$$

There are 15 distinct combinations of maternal and paternal genotypes possible, but if we use the product rule to evaluate *PI* we find that *PI* takes only four possible forms, depending on whether the alleged father is a homozygote or a heterozygote and whether or not the child's paternal allele can be unambiguously identified.[499] Table 10.3 tabulates the possible combination of mother, child, and alleged father along with the *PI* formulae utilizing the product rule.

10.2.4.1 Distribution of PI

The considerations given above allow us to calculate *PI* after the genotypes have been observed. This is the usual situation. However, when evaluating new multiplexes or when advising customers, it may be interesting to consider the performance of a multiplex over a great many prospective cases.

Table 10.3 Form of *PI* for all Nonexcluded Combinations of Maternal and Paternal Genotypes (Lee et al.[499])

Genotype Mother	Genotype Child	Genotype Alleged Father	*PI* (Alleged Father is True Father)
aa	*aa*	*aa*	$\frac{1}{p_a}$
ab			
bb	*ab*		
bc			
aa	*aa*	*ab*	$\frac{1}{2p_a}$
ab			
ac			
bb	*ab*		
bc			
bc	*ac*		
cc			
cd			
ab	*ab*	*aa*	$\frac{1}{p_a + p_b}$
		ab	
		ac	$\frac{1}{2(p_a + p_b)}$

This may be achieved by considering the distribution of *PI* values produced by a particular multiplex.

Suppose that an STR locus has n alleles, a, b, c, …, with allele frequencies p_a, p_b, …, respectively. For paternity trios, *PI* can then take one of the values

$$\frac{1}{p_a}, \frac{1}{2p_a}, \frac{1}{p_a + p_b}, \frac{1}{2(p_a + p_b)}, \text{ or } 0$$

if the alleged father is excluded. There are a total of $n(n + 1)$ values for the *PI* if we consider only the nonexcluded alleged fathers. The distribution of the *PI* under H_p can be derived by enumerating all possible pairs of genotypes for a mother and alleged father (G_M, G_{AF}), and evaluating $\Pr[G_M, G_{AF}]$ using the product rule, enumerating all possible genotypes for a child G_C, and then evaluating $\Pr[\, G_C | G_M, G_{AF}]$ using Mendel's laws. The probabilities of all trios G_M, G_{AF}, and G_C which give rise to a particular value for *PI* can then be summed. The distribution of *PI* for a single locus is a discrete distribution with $n(n + 1)$ possible values (if all of the allele frequencies p_a, p_b, … differ).

The distribution of *PI* under H_d can similarly be derived by enumerating all possible pairs of genotypes for a mother and alleged father (G_M, G_{AF}), and evaluating $\Pr[G_M, G_{AF}]$ using the product rule. However, when enumerating all possible genotypes for a child G_C and evaluating $\Pr[G_C | G_M, G_{AF}]$, the genotype of the alleged father is ignored. Most trios represent an exclusion and a *PI* of zero may be immediately assigned.

As we type more loci, the number of possible values for the *PI* increases greatly and becomes problematic to enumerate. Accordingly, simulation may be preferable. As a demonstration, we simulate the distribution of the *PI* over the ten SGM$^+$ loci for trios where the parents' genotypes are drawn from a large randomly mating population with allele frequencies from the New Zealand Caucasian population. Although the *PI*, even over ten loci, can take only a finite number of discrete values, there are so many of these discrete values that we treat the distribution as essentially continuous.

Two distributions are necessary (see Figure 10.3): when H_p is true and when H_d is true. If we discount mutation, then when H_p is true the *PI* is never zero. When H_d is true, *PI* is most often zero. Hence the distribution under H_d is dominated by a spike at 0 (not shown). There is a small probability of obtaining nonzero values for the *PI* by chance alone, when the falsely accused alleged father is not excluded. These latter situations are termed nonexcluded nonpaternities. For the ten SGM$^+$ loci, the probability that an unrelated randomly chosen person will have a *PI* greater than 0 with respect to a child of two unrelated randomly chosen people is approximately 2.2×10^{-5}.

As shown, the distribution of *PI* when H_d is true is shifted toward lower values than when H_p is true, even if *PI* is positive, i.e., the putative father has not been excluded. The domain of *PI* is the same in each case.

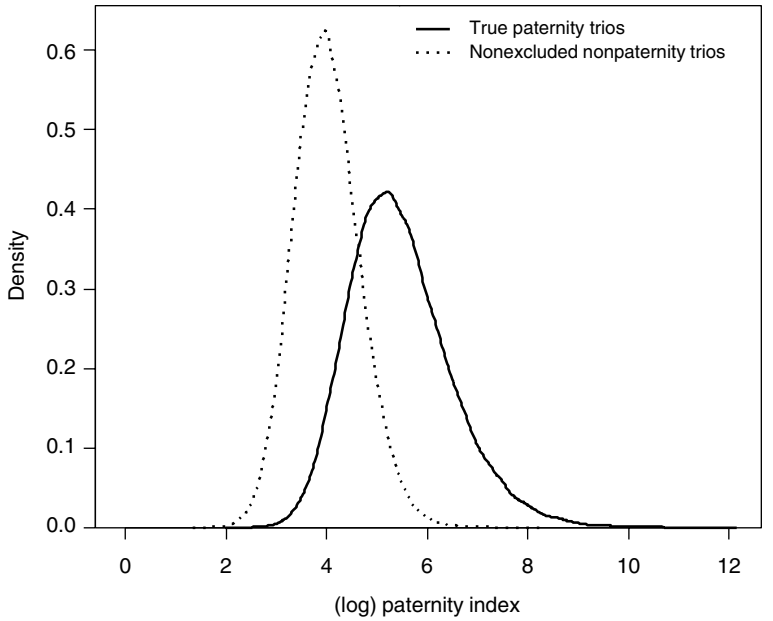


Figure 10.3 Distribution of the values of $\log_{10}(PI)$ for true paternity and nonexcluded nonpaternity trios.

Simulations such as that represented in Figure 10.3 give an indication of the expected performance of the multiplex in use in parentage testing casework. Consideration of additional loci, for example, using the 15-locus AmpF*l*STR® Identifiler™ (Applied Biosystems, Foster City, CA) system, moves the distribution of *PI* for both true paternity and nonexcluded nonpaternity trios to the right, that is, to larger values. For true paternity trios, the median value of *PI* increases from 2.2×10^5 to 5.5×10^7.

10.2.5 Paternity Duos: Child and Alleged Father

As usual, we begin by considering at least two hypotheses. In the most common case, these could be:

H_p: The alleged father is the true father (and the true mother's genotype is unknown).

H_d: A random person who is not related to the alleged father is the true father (and the true mother's genotype is unknown).

We require

$$PI = \frac{\Pr(G_C, G_{AF}|H_p)}{\Pr(G_C, G_{AF}|H_d)} = \frac{\Pr(G_C|G_{AF}, H_p)\Pr(G_{AF}|H_p)}{\Pr(G_C|G_{AF}, H_d)\Pr(G_{AF}|H_d)}$$

Next it is customary to assume $\Pr(G_{AF}|H_p) = \Pr(G_{AF}|H_d)$; hence

$$PI = \frac{\Pr(G_C|G_{AF}, H_p)}{\Pr(G_C|G_{AF}, H_d)}$$

This assumption is essentially stating that the genotype of the alleged father is unconditional on whether the alleged father is the true father or not. Evaluation of the *PI* proceeds directly from this equation.

If we assume allelic independence between people, we can write

$$PI = \frac{\Pr(G_C|G_{AF}, H_p)}{\Pr(G_C|G_{AF}, H_d)} = \frac{\Pr(G_C|G_{AF}, H_p)}{\Pr(G_C|H_d)}$$

Below we give an example using the assumption of independence, and a compilation of formulae appears in Table 10.4.

Example 10.4

	Genotype
Child	*ab*
Alleged father	*ac*

Under H_p we assume that the alleged father is the true father. As the genotype of the true mother is unknown, we cannot use a Punnett square to evaluate the numerator of the *PI*. It is easier to proceed by noting that the paternal allele under H_p is a and this will be passed $\frac{1}{2}$ of the time. Then the maternal allele (under H_p) must be b. We assign the probability $\Pr(b) = p_b$ to the event of this coming from a random person as before.

Under H_d and the assumption of independence, we see that we require the probability of the child's genotype but have neither the genotype of the true

Table 10.4 Form of the *PI* for all Nonexcluded Combinations of Paternal and Child Genotypes; we agree with Lee et al.[499]

Genotype Child	Genotype Alleged Father	*PI* (Alleged Father is True Father)
aa	*aa*	$\frac{1}{p_a}$
aa	*ab*	$\frac{1}{2p_a}$
ab	*aa*	
ab	*ab*	$\frac{p_a + p_b}{4p_a p_b}$
ab	*ac*	$\frac{1}{4p_a}$

mother nor the true father to help us determine this. We assign the probability of this event as $\Pr(G_c = ab|H_d) = 2p_a p_b$, which is the product rule assignment for this genotype. Hence,

$$PI = \frac{\frac{1}{2}p_b}{2p_a p_b} = \frac{1}{4p_a}$$

When the product rule is used to evaluate the case of a child–alleged father duo, the *PI* takes one of four possible forms (set out in Table 10.4).

10.2.6 Linked Loci

Consider a paternity trio with a pair of linked loci (A and B). We consider the situation where these loci have recombination fractions R_F and R_M for the female and male, respectively. Label the parental alleles as shown in Figure 10.4.

We consider the phase of M and F to be unknown. In other words, we do not know whether the A_3 and B_3 alleles in the mother are on the same chromosome or not. We assume that each phase is equiprobable. We ask the question: What is the chance that the mother will pass A_3, B_3? Label the two possible phases P_1 and P_2. In P_1, A_3 and B_3 are on the same chromosome, and in P_2 they are on different chromosomes.

To pass the A_3, B_3 set, we require that one of the following events occurs: (P_1 *and* no recombination *and* this chromosome (of the two) chosen) *or* (P_2 *and* recombination *and* this chromosome (of the two post recombination) chosen).

This suggests that $\frac{1}{2}(1 - R_F)\frac{1}{2} + \frac{1}{2}(R_F)\frac{1}{2} = \frac{1}{4}$, which is equal to the maternal Mendelian factor for two unlinked loci. The same result occurs for the male. This result holds true as long as each chromosome is only involved in one meiosis. The conclusion is that no correction for linked loci is necessary as long as we are only considering sets of single meioses (as in all trios and duos).[a]

For more complex pedigrees there is an effect, such as shown in Figure 10.5.

Consider the following hypotheses:

H_p: The alleged father (AF) is the father of C_2.
H_d: A random man is the father of C_2.

We again assume that the phases of the parents, M_1, M_2, and A_F, are unknown. The two mothers are each involved in only one meiosis and hence we can use the result, given above, that their Mendelian factor for two loci will be the same as the unlinked factor. However, the alleged father is involved in multiple meioses and hence there is an effect of linkage. Hence we will consider P_1^{AF} to be phase 1 for AF and P_2^{AF} to be phase 2 for AF. We arbitrarily assign P_1^{AF} to be the phase with A_1, B_1 on the same chromosome. Hence P_2^{AF} is the phase with A_1, B_2 on the same chromosome.

[a] Dr. Charles Brenner made us aware of this fact.

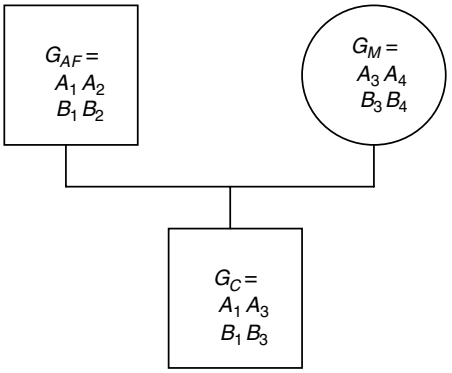


Figure 10.4 A paternity trio with two linked loci.

Consider

$$\Pr(C_1, C_2|M_1, M_2, AF, H_p)$$
$$= \Pr(C_1, C_2|M_1, M_2, AF, P_1^{AF}, H_p)\Pr(P_1^{AF}) + \Pr(C_1, C_2|M_1, M_2, AF, P_2^{AF}, H_p)\Pr(P_2^{AF})$$

Assuming, as before, that $P_1^{AF} = P_2^{AF} = \frac{1}{2}$,

$$\Pr(C_1, C_2|M_1, M_2, AF, H_p)$$
$$= \frac{\Pr(C_1, C_2|M_1, M_2, AF, P_1^{AF}, H_p) + \Pr(C_1, C_2|M_1, M_2, AF, P_2^{AF}, H_p)}{2}$$

Considering the recombination fraction, we obtain

$$\Pr(C_1, C_2|M_1, M_2, AF, H_p) = \frac{\frac{1}{4}\frac{(1-R_M)}{2}\frac{1}{4}\frac{(1-R_M)}{2} + \frac{1}{4}\frac{R_M}{2}\frac{1}{4}\frac{R_M}{2}}{2} = \frac{(1-R_M)^2 + R_M^2}{32}$$

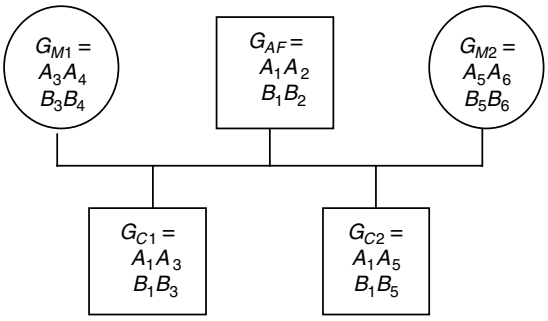


Figure 10.5 A pedigree involving multiple meioses of the same chromosome.

For unlinked loci we have

$$\Pr(C_1, C_2|M_1, M_2, AF, H_p) = \frac{1}{64}$$

which is equivalent to setting $R_M = \frac{1}{2}$ in the above equation.

As a numerical example, consider the pair of linked CODIS loci HUMCSF1PO and HUMD5S818 that are reported to be separated by 25 centiMorgans (cM).[30] Using Haldane's distance, this suggests a recombination fraction of 0.197.

$$\Pr(C_1, C_2|M_1, M_2, AF, H_p) = 0.021$$

as compared with 0.016 if we do not consider the linkage. We see that linkage at this distance (25 cM) has a moderate but not a severe effect.

10.2.7 Paternally Imprinted Alleles

Techniques are becoming available that allow the identification of the paternal allele.[572] This requires no modification to our general approach, but does cause some differences in detail. When the child's maternal allele can be determined unambiguously, the *PI* in this situation is unchanged. When the child shares the same genotype as the mother, the additional information provided by a knowledge of the paternal allele leads to simpler formulae for *PI*. We reproduce the table for the paternity trios using the product rule but accounting for knowledge of the paternal allele (Table 10.5).

10.3 Nonautosomal DNA Markers

10.3.1 Y Chromosome Analysis

The Y chromosome is inherited patrilinearly, and for most of its length does not recombine. Since there is no maternal allele, paternity duos and trios are treated in the same way. If the effect of mutation is ignored, even cases where the alleged father is untyped and a more distant paternal relative is used as the reference sample lead to the same formula. *PI* has a very simple form. Consider a mother, alleged father, child trio; an alleged father, child duo; or an alleged paternal relative, child duo. In each case, we assume an exact match of the haplotype of the Y chromosome between the alleged father or the paternal relative and the child under the hypothesis H_p.

The exclusion probability is $PE = (1 - f)$, where f is the frequency of the child's (and alleged father's) haplotype. For the paternity index, we need to assess

$$PI = \frac{\Pr(G_C|G_{AF}, H_p)}{\Pr(G_C|G_{AF}, H_d)} = \frac{1}{f}$$

This formula does not account for mutation.[669]

Table 10.5 Form of *PI* for all Nonexcluded Trios Accounting for Knowledge of the Paternal Allele

Genotype Mother	Genotype Child[a]	Genotype Alleged Father	*PI* (Alleged Father is True Father)
aa	***aa***	*aa*	$\frac{1}{p_a}$
ab			
bb	***ab***		
bc			
aa	***aa***	*ab*	$\frac{1}{2p_a}$
ab			
ac			
bb	***ab***		
bc			
bc	***ac***		
cc			
cd			
ab	***ab***	*aa*	$\frac{1}{p_a}$
		ab	$\frac{1}{2p_a}$
		ac	

[a]The paternal allele is marked in bold.

10.3.2 X Chromosome Analysis

The X chromosome does not recombine in the father. Again paternity duos, trios, and the use of more distant paternal relatives lead to the same formulae. Consider again a mother, alleged father, child trio; an alleged father, child duo; or an alleged paternal relative, child duo. In each case, we assume an exact match of the haplotype of the paternal X chromosome between the alleged father or the paternal relative and the child under the hypothesis H_p.

The exclusion probability is $PE = (1 - f)$, where f is the frequency of the child's (and alleged father's) haplotype.

As usual, we assess

$$PI = \frac{\Pr(G_C|G_{AF}, H_p)}{\Pr(G_C|G_{AF}, H_d)}$$

We use the same reasoning for the X and Y chromosomes so, ignoring the possibility of mutation, $PI = 1/f$, where f is the multilocus frequency of the child's X chromosomal type. We speculate that this frequency may be approximated as the product of single-locus frequencies since recombination does occur in the

female. Recall that linkage does not necessarily imply linkage disequilibrium (see Chapter 3). Szibor et al.[750] noted one instance in which they observed linkage disequilibrium among 16 loci in a set of 210 males. This result suggests that sufficient recombination occurs to bring the frequencies close to equilibrium expectations in this population. It is likely that the same modeling considerations given to autosomal DNA may apply to a large extent for the X chromosome. However, more and larger studies would be welcome.

10.3.2.1 Maternity Analysis for X Chromosomes

In rare circumstances, we may be required to consider a pedigree analysis where it is the mother whose parentage is under question. In such a case, the linkage of the loci on the X chromosome will need to be considered. However, since we cannot tell the phase of the mother, the finding given above applies in the case of a single meiosis.

The exclusion probability calculation is analogous to autosomal DNA. The calculation of the *PI* makes use of the unknown phase of the mother and hence applies a factor of $\frac{1}{2}$ per heterozygotic locus. Again ignoring the possibility of mutation, $PI = 1/2^N f$, where f is the multilocus genotype probability of the child's X chromosome and N is the number of heterozygotic X chromosome loci considered.

10.3.3 Mitochondrial DNA Analysis

For forensic purposes, we typically assume that mitochondrial DNA is matrilineally inherited without recombination. Therefore, it is of use in parentage testing only when the parentage of the mother is in question. The development and formulae are the same as for the Y chromosome and for paternity testing with the X chromosome. As in these cases, each of child, father, alleged mother trios; alleged mother, child duos; and duos incorporating a maternal relative yield the same result if mutation is ignored.

Following the Y chromosome analysis given above and assuming an exact match between the alleged mother and the child, the exclusion probability is $PE = (1 - f)$ and the paternity index is $PI = 1/f$, where f is the frequency of the child's (and alleged mother's) mtDNA haplotype.

10.4 Use of the Subpopulation Model

Using the subpopulation model of Balding and Nichols,[27,36,39,41,499] the evaluation of $\Pr(A_p|G_M, G_{AF}, H_p)$ for a paternity trio proceeds as we have outlined above. The probability of observing the paternal allele from an alternative donor in the same subpopulation may be affected by the genotypes of the mother and the alleged father. Hence in general,

$$\Pr(A_p|G_M, G_{AF}, H_d) \neq \Pr(A_p|H_d)$$

A relatively easy way to evaluate the *PI* in this situation is to write the term containing the paternal allele $\Pr(A_p|G_M, G_{AF}, H_d)$ as the probability of the paternal allele conditioned on the correct selection of conditioning alleles. If we consider that the mother and alleged father are in the same subpopulation as the true father, then we condition on four alleles, those of the mother and alleged father. Equally, if we consider the true father to belong to the same subpopulation as the alleged father but not that of the mother, then we condition on the two alleles of the alleged father. If we consider the true father to belong to the subpopulation of neither the mother nor the alleged father, then we condition on no alleles and the product rule is appropriate. Once again, we will illustrate the implementation of this approach through a series of worked examples.

Example 10.5. Consider the same trio of genotypes as in Example 10.1.

	Genotype
Mother	*cd*
Child	*ac*
Alleged father	*ab*

Under H_d we can see that the mother is heterozygotic for the maternal allele ($A_m = c$) and can assign the value $M_M = \frac{1}{2}$ to the mother's Mendelian factor. The paternal allele is $A_p = a$. Assuming that we are going to condition on the genotypes of both the mother and the alleged father, we write the probability $\Pr(A_p|G_M, G_{AF}, H_d)$ as $\Pr(a|abcd)$. The procedure for evaluation of this term follows the same approach as described in Chapter 7 on mixtures. Here we reproduce some of the "shortcut" rules for evaluating these terms.

Shortcut rules: Replace each of the paternal allele frequencies with one of these terms.

1st allele *a*	$(1-\theta)p_a$
2nd allele *a*	$\theta + (1-\theta)p_a$
3rd allele *a*	$2\theta + (1-\theta)p_a$
4th allele *a*	$3\theta + (1-\theta)p_a$
⋮	
*n*th allele	$(n-1)\theta + (1-\theta)p_a$

Divide by a correction term based on the number of alleles behind the conditioning bar.

2 behind	$(1+\theta)$
4 behind	$(1+3\theta)$
6 behind	$(1+5\theta)$
⋮	
M behind	$(1+(M-1)\theta)$

In Example 10.5 we condition on one copy of each of the four alleles, a, b, c, and d. Thus, the PI becomes

$$\begin{aligned} PI &= \frac{\frac{1}{4}}{\frac{1}{2} \times \Pr(a|abcd)} \\ &= \frac{1}{2\Pr(a|abcd)} \\ &= \frac{1+3\theta}{2(\theta+(1-\theta)p_a)} \end{aligned}$$

Example 10.6. Consider the following trio:

	Genotype
Mother	bc
Child	ab
Alleged father	aa

Under H_d we can see that the mother is heterozygotic for the maternal allele ($A_m = b$) and can assign the value $M_M = \frac{1}{2}$ to the mother's Mendelian factor. The paternal allele is $A_p = a$, and under H_d we condition on having observed two copies of the paternal allele from among the four alleles observed from this subpopulation. We write the conditional probability for this as $\Pr(a|aabc)$. Hence

$$\begin{aligned} PI &= \frac{\frac{1}{2}}{\frac{1}{2} \times \Pr(a|aabc)} \\ &= \frac{1}{\Pr(a|aabc)} \\ &= \frac{1+3\theta}{2\theta+(1-\theta)p_a} \end{aligned}$$

In Example 10.7 we consider the same set of trios as in Example 10.6, but in addition we assume that the mother comes from a different subpopulation than the alleged and true fathers.

Example 10.7. Consider the following trio of genotypes:

	Genotype
Mother	bc
Child	ab
Alleged father	aa

The maternal and paternal alleles and their Mendelian factors are as described above. Under H_d we condition on having observed two copies of

the paternal allele from among a total of only two alleles from the same subpopulation as opposed to four. We write the conditional probability for this as $\Pr(a|aa)$. Hence

$$PI = \frac{\frac{1}{2}}{\frac{1}{2} \times \Pr(a|aa)}$$
$$= \frac{1}{\Pr(a|aa)}$$
$$= \frac{1+\theta}{2\theta + (1-\theta)p_a}$$

Tables 10.6, 10.7, 10.8, and 10.9 give the form of the paternity index for paternity trios and duos under situations where the mother is a member of

Table 10.6 Form of *PI* for all Nonexcluded Trios of Maternal, Paternal, and Child Genotypes When Considering Subpopulations (Specifically, When the Mother, Alleged Father and True Father are Members of the Same Subpopulation)

Genotype Mother	Genotype Child	Genotype Alleged Father	Same Subpopulation
aa	*aa*	*aa*	$\frac{1+3\theta}{4\theta+(1-\theta)p_a}$
ab			$\frac{1+3\theta}{3\theta+(1-\theta)p_a}$
bb	*ab*		$\frac{1+3\theta}{2\theta+(1-\theta)p_a}$
bc			
aa	*aa*	*ab*	$\frac{1+3\theta}{2(3\theta+(1-\theta)p_a)}$
ab			$\frac{1+3\theta}{2(2\theta+(1-\theta)p_a)}$
ac			
bb	*ab*		$\frac{1+3\theta}{2(\theta+(1-\theta)p_a)}$
bc			
bc	*ac*		
cc			
cd			
ab	*ab*	*aa*	$\frac{1+3\theta}{4\theta+(1-\theta)(p_a+p_b)}$
		ab	
		ac	$\frac{1+3\theta}{2\left[3\theta+(1-\theta)(p_a+p_b)\right]}$

Table 10.7 Form of *PI* for all Nonexcluded Trios of Maternal, Paternal, and Child Genotypes When Considering Subpopulations (Specifically, When the Alleged Father and True Father are from the same subpopulation. The Mother is from a different subpopulation or race)

Genotype Mother	Genotype Child	Genotype Alleged Father	Different Subpopulation
aa	*aa*	*aa*	$\frac{1+\theta}{2\theta+(1-\theta)p_a}$
ab			
bb	*ab*		
bc			
aa	*aa*	*ab*	$\frac{1+\theta}{2(\theta+(1-\theta)p_a)}$
ab			
ac			
bb	*ab*		
bc			
cc	*ac*		
cd			
ab	*ab*	*aa*	$\frac{1+\theta}{2\theta+(1-\theta)(p_a+p_b)}$
		ab	
		ac	$\frac{1+\theta}{2\left[\theta+(1-\theta)(p_a+p_b)\right]}$

Table 10.8 Form of *PI* for all Nonexcluded Duos of Paternal and Child Genotypes When Considering Subpopulations (Specifically, When the Mother, Alleged Father and True Father are Members of the Same Subpopulation)

Genotype Child	Genotype Alleged Father	Same Subpopulation
aa	*aa*	$\frac{1+2\theta}{3\theta+(1-\theta)p_a}$
aa	*ab*	$\frac{1+2\theta}{2(2\theta+(1-\theta)p_a)}$
ab	*aa*	$\frac{1+2\theta}{2(2\theta+(1-\theta)p_a)}$
ab	*ab*	$\frac{(1+2\theta)(2\theta+(1-\theta)(p_a+p_b))}{4(\theta+(1-\theta)p_a)(\theta+(1-\theta)p_b)}$
ab	*ac*	$\frac{1+2\theta}{4(\theta+(1-\theta)p_a)}$

Table 10.9 Form of *PI* for all Nonexcluded Duos of Paternal and Child Genotypes When Considering Subpopulations (Specifically When the Alleged Father and True Father are Members of the same Subpopulation. The Mother is a different race with Allele Probabilities p_i'. If the Mother is the same race but a different Subpopulation then $p_i'=p_i$ with a considerable cancellation)

Genotype Child	Genotype Alleged Father	Different Subpopulation
aa	*aa*	$\frac{1+\theta}{2\theta+(1-\theta)p_a}$
aa	*ab*	$\frac{1+\theta}{2(\theta+(1-\theta)p_a)}$
ab	*aa*	$\frac{(1+\theta)p_b'}{\left[(2\theta+(1-\theta)p_a)p_b'+(1-\theta)p_b p_a'\right]}$
ab	*ab*	$\frac{(1+\theta)(p_a'+p_b')}{2\left[(\theta+(1-\theta)p_a)p_b'+p_a'(\theta+(1-\theta)p_b)\right]}$
ab	*ac*	$\frac{(1+\theta)p_b'}{2\left[(\theta+(1-\theta)p_a)p_b'+(1-\theta)p_b p_a'\right]}$

the same subpopulation as the alleged and true fathers and also for situations when she is in a different subpopulation or race.

Riancho and Zarrabeitia[643] have offered software that performs these calculations.

10.5 Relatedness in Paternity Cases

We envisage two potential situations where relatedness must be considered in the course of a parentage testing case:

1. A person is alleged to be the father of a child but his genotype is not available (due to death for instance), and the paternity analysis proceeds using genotypes of the mother, the child, and the alleged father's relatives.
2. A plausible alternative father is a relative of the accused man.

The structure for solving these problems is based on three-allele descent measures as demonstrated by Evett and Weir[267] and Weir[836] (see also Berry and Geisser,[59] Brenner,[86] or Morris et al.[562]). It is often possible to assume that neither the alleged father nor the mother are inbred, and this reduces the complexity of the problem considerably. In such cases it is possible to deal with a single factor, θ' (θ_{AT} in Weir). This factor θ' is very similar

to the factor θ discussed in Chapter 3 in that it is a two-allele measure giving the probability that the paternal allele and a random allele from the alleged father (or a genotyped person) are identical by descent (IBD).

Relationship	θ'
Siblings, parent/child	$\frac{1}{4}$
Uncle/nephew,grandparent/grandchild, half-siblings	$\frac{1}{8}$
Cousins	$\frac{1}{16}$

10.5.1 A Relative of the Accused is Suggested as the Alleged Father

Consider the situation where the accused man (previously called the "alleged father") puts forward the suggestion that his brother is the father of the child (hence, under H_d, the brother is the alleged father). Thus, the pair of hypotheses being considered are:

H_p: The accused man is the father of the child in question.
H_d: The brother of the accused man is the father of the child in question.

Let the genotype of the accused man be G_{AF} as before. Under H_d, we require the probability that the paternal allele A_p is a certain allele given the genotype, G_{AF}, of the brother of the donor of this allele. This requires us to consider IBD states between one allele from each of two brothers (or other relatives). Unlike the situation described in Chapter 4, the fact that two siblings may share two pairs of alleles does not cause extra complexity here.

Consider the situation where $G_M = cc$, $G_C = ac$, and $G_{AF} = ab$. We see that under both hypotheses, H_p and H_d, the paternal allele can be unambiguously identified as $A_p = a$. We require the probability that $A_p = a$ given the fact that $G_{AF} = ab$ and given that the donor of the paternal allele is the brother of the accused. There are three states that can occur:

1. A_p is IBD with the a allele of AF (with probability $\frac{1}{4}$).
2. A_p is IBD with the b allele of AF (probability $\frac{1}{4}$).
3. A_p is not IBD with either allele of AF (probability $\frac{1}{2}$).

Thus:

1. If A_p is IBD with the a allele of AF, then it should always be type a, $\Pr(A_p = a) = 1$.
2. If A_p is IBD with the b allele of AF, then it should never be type b, $\Pr(A_p = b) = 0$.

3. If A_p is not IBD with either allele of AF, then it is a random allele, $\Pr(A_p = a) = \Pr(a|abcc)$.[b]

Hence

$$\Pr(A_p = a|G_{AF} = ab \text{ and } G_M = cc) = \frac{1}{4} + \frac{\Pr(a\,|abcc)}{2}$$

If we are prepared to make the assumption of independence, this is $1/4 + p_a/2$, whereas under the assumption that the mother and the accused come from the same subpopulation

$$\Pr(A_p = a|G_{AF} = ab, G_M = cc) = \frac{1}{4} + \frac{1}{2}\frac{\theta + (1-\theta)p_a}{1+3\theta}$$

We can calculate the probability in the numerator of the PI using Mendelian factors $\Pr(G_C|G_{AF}, G_M, H_p) = \frac{1}{2}$.

For the denominator we need the probability $\Pr(G_C|G_{AF}, G_M, H_d)$. Since the maternal Mendelian factor is 1 (she is a homozygote), the denominator of the likelihood ratio is just $\Pr(A_p = a|G_{AF} = ab, G_M = cc)$, the probability we have just derived. Thus,

$$PI = \frac{2}{1+2p_a}$$

assuming independence and

$$PI = \frac{2(1+3\theta)}{1+2(\theta+(1-\theta)p_a)}$$

assuming substructure with inbreeding parameter θ (see Tables 10.10–10.13).

10.5.2 Deficient Paternity Analysis (The Alleged Father is Unavailable)

We envisage a situation where the alleged father is unavailable but where, for example, his brother is. In such a case, the two hypotheses may be of the form:

H_p: The person, X, is a sibling of the true father.
H_d: The person, X, is unrelated to the child.[c]

The evaluation of the probability in the denominator of the PI, $\Pr(G_C|G_X, G_M, H_d)$, is the same as for a probability trio illustrated earlier since, under H_d, person X is unrelated to the child.

[b] Note that we are conditioning on both the genotypes of the alleged father and the mother of the child. Other conditioning may be appropriate in certain cases.
[c] The conventional additional assumption is made that the genotype of the true mother is known and is common to both H_p and H_d, but not explicitly stated.

Table 10.10 Form of *PI* for all Nonexcluded Trios of Maternal, Paternal, and Child Genotypes When the Alternative Hypothesis is That the Brother of the Alleged Father is the True Father and When the Mother, Alleged Fathers and True Father are Members of the Same Subpopulation

Genotype Mother	Genotype Child	Genotype Alleged Father	Same Subpopulation
aa	*aa*	*aa*	$\frac{2(1+3\theta)}{1+7\theta+(1-\theta)p_a}$
ab			$\frac{2(1+3\theta)}{1+6\theta+(1-\theta)p_a}$
bb / *bc*	*ab*		$\frac{2(1+3\theta)}{1+5\theta+(1-\theta)p_a}$
aa	*aa*	*ab*	$\frac{2(1+3\theta)}{1+9\theta+2(1-\theta)p_a}$
ab / *ac*			$\frac{2(1+3\theta)}{1+7\theta+2(1-\theta)p_a}$
bb / *bc*	*ab*		$\frac{2(1+3\theta)}{1+5\theta+2(1-\theta)p_a}$
cc / *cd*	*ac*		
ab	*ab*	*aa* / *ab*	$\frac{2(1+3\theta)}{1+7\theta+(1-\theta)(p_a+p_b)}$
		ac	$\frac{2(1+3\theta)}{1+9\theta+2(1-\theta)(p_a+p_b)}$

In this case it is the numerator, $\Pr(G_C|G_X, G_M, H_p)$, that is not a product of simple Mendelian factors. Recall that the probability that a random allele from X will be IBD with an allele from AF is θ'. Under H_p, the paternal allele is a random choice from AF; hence, with probability θ' it will be IBD with a random allele from X. This gives a relatively straightforward structure for calculating the *PI* for any situation. Consider $G_M = cd$, $G_C = ac$, and $G_X = ab$. The maternal Mendelian factor $M_M = \frac{1}{2}$. Under H_p, the person X is a sibling of the father. There are three cases to consider:

1. A_p is IBD with the a allele of X (with probability $\frac{1}{4}$).
2. A_p is IBD with the b allele of X (probability $\frac{1}{4}$).
3. A_p is not IBD with either allele of X (probability $\frac{1}{2}$).

Table 10.11 Form of the *PI* for all Nonexcluded Trios of Maternal, Paternal, and Child Genotypes When the Alternative Hypothesis is That the Brother of the Alleged Father is the True Father and When the Alleged Father and True Father are Members of the same Subpopulation but the Mother is a Member of a different Subpopulation or Race

Genotype Mother	Genotype Child	Genotype Alleged Father	Different Subpopulation
aa	*aa*	*aa*	$\frac{2(1+\theta)}{1+3\theta+(1-\theta)p_a}$
ab			
bb	*ab*		
bc			
aa	*aa*	*ab*	$\frac{2(1+\theta)}{1+3\theta+2(1-\theta)p_a}$
ab			
ac			
bb	*ab*		
bc			
bc	*ac*		
cc			
cd			
ab	*ab*	*aa*	$\frac{2(1+\theta)}{1+3\theta+(1-\theta)(p_a+p_b)}$
		ab	
		ac	$\frac{2(1+\theta)}{1+3\theta+2(1-\theta)(p_a+p_b)}$

Table 10.12 Form of *PI* for all Nonexcluded Duos of Paternal and Child Genotypes When the Alternative Hypothesis is That the Brother of the Alleged Father is the True Father

Genotype Child	Genotype Alleged Father	Mother Same Subpopulation as Alleged and True Father
aa	*aa*	$\frac{2(1+2\theta)}{1+5\theta+(1-\theta)p_a}$
aa	*ab*	$\frac{2(1+2\theta)}{1+6\theta+2(1-\theta)p_a}$
ab	*aa*	
ab	*ab*	$\frac{2(1+2\theta)(2\theta+(1-\theta)(p_a+p_b))}{\Delta}$
		$\Delta=(1+4\theta+2(1-\theta)p_a)(\theta+(1-\theta)p_b)$ $+(\theta+(1-\theta)p_a)(1+4\theta+2(1-\theta)p_b)$
ab	*ac*	$\frac{2(1+2\theta)}{1+6\theta+4(1-\theta)p_a}$

Table 10.13 Form of *PI* for all Nonexcluded Duos of Paternal and Child Genotypes When the Alternative Hypothesis is That the Brother of the Alleged Father is the True Father

Genotype Child	Genotype Alleged Father	Mother from a Different Subpopulation or race from the Alleged and True Father. Allele probabilities p_i' in the race of the mother. If she is the same race but a different Subpopulation then a simplification occurs.
aa	*aa*	$\dfrac{2(1+\theta)}{1+3\theta+(1-\theta)p_a}$
aa	*ab*	$\dfrac{2(1+\theta)}{1+3\theta+2(1-\theta)p_a}$
ab	*aa*	$\dfrac{2(1+\theta)p_b'}{(1+3\theta)+(1-\theta)(p_a p_b'+p_a' p_b)}$
ab	*ab*	$\dfrac{2(1+\theta)(p_a'+p_b')}{\Delta}$ $\Delta=(1+3\theta+2(1-\theta)p_a)p_b' + p_a'(1+3\theta+2(1-\theta)p_b)$
ab	*ac*	$\dfrac{2(1+\theta)p_b'}{(1+3\theta+2(1-\theta)p_a)p_b'+2(1-\theta)p_b p_a'}$

1. If A_p is IBD with the *a* allele of *AF*, then it should be type *a* always, $\Pr(A_p = a) = 1$.
2. If A_p is IBD with the *b* allele of *AF*, then it should never be type *a*, $\Pr(A_p = a) = 0$.
3. If A_p is not IBD with either allele of *AF*, then it is a random allele, $\Pr(A_p = a) = \Pr(a|abcd)$[d]

Hence
$$\Pr(G_C|G_X, G_M, H_p) = M_M \times \left(\frac{1}{4} + \frac{\Pr(a|abcd)}{2}\right)$$
and hence
$$PI = \frac{M_M \times \left(\frac{1}{4} + \frac{\Pr(a|abcd)}{2}\right)}{M_M \times \Pr(a|abcd)} = \frac{1+2\Pr(a|abcd)}{4\Pr(a|abcd)}$$

Making the assumption of independence
$$PI = \frac{1+2p_a}{4p_a}$$

[d] Note that we are only conditioning on the genotypes of both the uncle and the mother. Other conditioning may be appropriate in certain cases.

and under the assumption that the mother, the alleged father, and the true father are from the same subpopulation,

$$PI = \frac{1 + 5\theta + 2(1 - \theta)p_a}{4[\theta + (1 - \theta)p_a]}$$

10.6 Multiple Children

The approach outlined above does not work with more than one child if both children are to be considered together.[e] This event rarely happens in criminal casework, but it does arise in immigration cases and provides an introduction to problems involving more complex pedigrees.

Consider a mother M, alleged father AF, and N children $C_1 \ldots C_N$. For the hypotheses:

H_p: The alleged father is the father of all the children (and the mother is the true mother), and

H_d: The alleged father is the father of none of the children, but they all have the same father (and the mother is the true mother)

$$PI = \frac{\Pr(C_1 \ldots C_N, M, AF|H_p)}{\Pr(C_1 \ldots C_N, M, AF|H_d)}$$

$$= \frac{\Pr(C_1 \ldots C_N|M, AF, H_p)\Pr(M, AF|H_p)}{\Pr(C_1 \ldots C_N|M, AF, H_d)\Pr(M, AF|H_d)}$$

Assuming, as previously, that $\Pr(M, AF|H_p) = \Pr(M, AF|H_d)$ and after enumeration of all possible genotypes for the true father as $R_1 \ldots R_M$, we obtain

$$PI = \frac{\Pr(C_1 \ldots C_N|M, AF, H_p)}{\sum_{i=1}^{M}\Pr(C_1 \ldots C_N|M, R_i, AF, H_d)\Pr(R_i|M, AF, H_d)} \qquad (10.1)$$

If we make the assumption of independence between people,

$$PI = \frac{\Pr(C_1 \ldots C_N|M, AF, H_p)}{\sum_{i=1}^{M}\Pr(C_1 \ldots C_N|M, R_i, H_d)\Pr(R_i|H_d)} \qquad (10.2)$$

[e] I have been asked why we would ever want to do this? How about the Tsar? Or see, for example, Macan et al.[529] who had a two child incest case.

In Example 10.8, we consider the situation with two children ($N = 2$).

Example 10.8

	Genotype
Mother	*cd*
Child 1	*ac*
Child 2	*ad*
Alleged father	*ab*

Given the genotypes of their parents, the genotypes of the children are independent. Hence

$$\Pr(C_1, C_2|M, AF, H_p) = \Pr(C_1|M, AF, H_p) \times \Pr(C_2|M, AF, H_p)$$
$$= \frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$$

This probability is the product of the Mendelian factors for each child. Under H_d we see that the true father must have the *a* allele. Therefore, we can restrict the possible candidates to those whose genotype, R_i, is either an *aa* homozygote or a heterozygote with a single copy of the *a* allele ($a\star$). Explicit use of this restriction is unnecessary, but simplifies the calculation by suppressing many terms that ultimately turn out to be zero. We can set out the calculation for the denominator, $\Pr(C_1, C_2|M, AF, H_d)$, in a tabular form:

R_i	$\Pr(R_i\|H_d)$	$\Pr(C_1, C_2\|M, R_i, H_d)$	$\Pr(C_1, C_2\|M, R_i, H_d) \times \Pr(R_i\|H_d)$
aa	p_a^2	$\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$	$\frac{p_a^2}{4}$
$a\star$	$2p_a(1 - p_a)$	$\frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$	$\frac{p_a(1 - p_a)}{8}$
Sum			$\frac{p_a(1 + p_a)}{8}$

In this table we make the assumption of independence (the product rule). The paternity index becomes

$$PI = \frac{1}{2p_a(1 + p_a)}$$

This expression may easily be extended to the subpopulation case by replacing the factors in the $\Pr(R_i|H_d)$ column with their subpopulation equivalents.

10.7 Mutation

A mutation is a change to the DNA sequence usually caused by an error during DNA replication at meiosis. Such a change will be transmitted to the next generation. Mutations can occur as either a single substitution in the nucleotide sequence (transition/transversion) or as deletions or insertions of tracts of DNA.

In 1912 Wilhelm Weinberg reported that children with short limbed dwarfism were often the last born children. Weinberg, correctly, and with astonishing insight, suggested that this was due to a mutation. In 1955 Penrose showed that the effect observed by Weinberg was due to paternal age, and not maternal age or birth order. These observations have led to the recognition of the phenomenon of the effect of paternal age on mutation rate. However, this may not be as simple as initially envisaged. Direct examinations of sperm suggest only a small increase in the number of mutant sperm, many fewer than expected from clinical data. Crow[210] reviews the current knowledge and suggests that there are three classes of gene mutation:

1. Nucleotide substitutions scattered along the gene usually with substantial sex and age effects.
2. Small insertions and deletions, mainly deletions, with no age effect and a slight maternal excess.
3. Hot spots occurring almost exclusively in males and rising steeply with age.

Crow also reviews the evidence for and against pre-mitotic selection (the preferential selection of mutant spermatogonia before the two cell divisions that give rise to sperm).

STR loci are particularly prone to mutation compared with coding sequences or noncoding and nonrepetitive DNA sequences. This is one of the reasons why STR loci are often very polymorphic (along with supposed selective neutrality). In essence, this is one of the properties that render such loci highly informative markers in forensic genetics. However, substitutions cannot be detected by most imaging systems as they produce the same sized allelic product as the unmutated sequence.

Observation of any putative paternal mutation will lower the value of the paternity index quite dramatically. Consequently, it would seem sensible to avoid a locus with high mutation rates. However, to be useful a locus should also be highly polymorphic. There is a general positive correlation between mutation rate and polymorphism, which makes the selection of microsatellite loci for investigating paternity difficult.

The relatively high mutation rates of STR loci stem from the tandem arrangement of their repeated sequences. STR loci are typically observed to

mutate in a predictable and characteristic way by the addition or loss of one or more units of the repeated sequence. At least two possible explanations have been advanced to account for this observation. By way of example, consider an idealized structure of the hypothetical STR locus in Figure 10.6.

There are two theories that suggest explanations as to how mutations at STR loci occur. The loop-out theory posits that during DNA replication *in vivo*, the DNA polymerase enzyme must traverse a stretch of DNA containing a tandem array of a repeated sequence. If the repeats have the same sequence motif, then there exists the possibility that either the template strand or the replicating strand can "loop out" one or more repeated units. The loop can be stabilized kinetically, provided that it is in phase and that there is sufficient sequence homology at either side. The effect of this loop is that the copied DNA strand will either have increased or decreased in length by one or more full repeat units, depending on in which strand the loop-out occurred. It seems highly likely that stuttering is the result of a similar process occurring *in vitro* with TAQ polymerase.[808]

The unequal cross-over theory posits that during a recombination event, there is an unequal exchange of DNA between the chromatids, resulting in a reciprocal change to both alleles.

As the typing of STR alleles is based on length polymorphism, both types of event will spawn a new allele. The theory behind this process has become known as the stepwise mutation model (SMM) and postulates that STR alleles mutate incrementally by integer expansions and contractions of the repeat motif.[876] A number of factors appear to affect the propensity for stepwise mutation (and stutter). A positive correlation between the mean number of uninterrupted repeats and mutation rate was reported by Brinkman et al.[92] The longer the run of repeats and the greater the "purity" of the repeat sequence, the greater the propensity to mutate or stutter.

Although the large majority of mutational events occurring at STR loci appear to be changes to the number of repeated units, it is also possible for there to be other types of mutational events. The possible types of mutation that could theoretically occur at our hypothetical STR locus are shown in Figure 10.7.

Only events resulting in a change in the length of the DNA between the primer binding sequences will be perceived as a mutation on typing. Insertions or deletion events not involving whole repeated units may produce noninteger

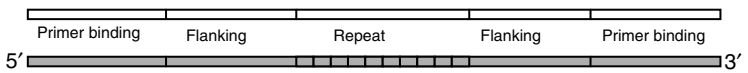
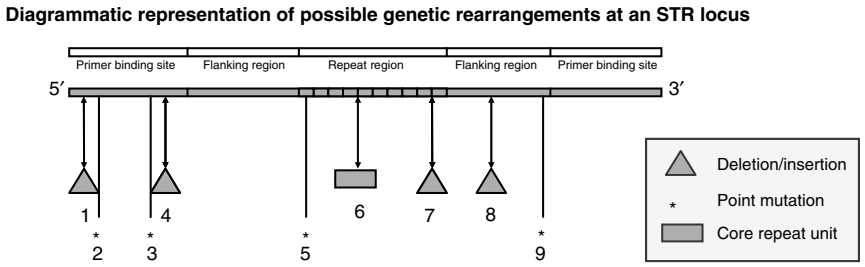


Figure 10.6 Idealized structure of an STR locus.



Type of rearrangement	Possible effect of rearrangement on STR profiling
1. Deletion/insertion@ 5'end of PBS	Reduced efficiency of primer binding
2. Point mutation@ 5'end of PBS	Reduced efficiency of primer binding
3. Point mutation@3'end of PBS	Abolition of primer binding
4. Deletion/insertion @ 3'end of PBS	Abolition of primer binding
5. Point mutation in RR	Sequence microvariant
6. Deletion/insertion of repeat units in RR	Increase/decrease in size of allele
7. Deletion/insertion in RR	Increase/decrease in size of allele
8. Deletion/insertion in FR	Increase/decrease in size of allele
9. Point mutation in FR	Sequence microvariant

Figure 10.7 Possible mutations at an STR locus.

changes to the allele length. The commonly occurring HUMTH01 9.3 allele appears to be the result of an ancestral mutation deleting a single nucleotide from one of the repeated units in the repeat region. Other loci (e.g., HUMD19S433, HUMD21S11, and HUMFIBRA/FGA) have a series of commonly encountered X.2 variants resulting from event(s) culminating in the loss or addition of two nucleotides in the flanking DNA. Changes to the nucleotide sequence of the flanking region and repeat region will result in sequence microheterogeneity, but will not be perceived as affecting the Mendelian pattern. This is not, however, the case for mutational events occurring in the primer binding sequences. Mutations in the primer binding sequences can produce silent (or null) alleles, depending on whether they occur toward the 5′ or 3′ ends of the primer binding sequence, shown in Figures 10.8 and 10.9.

The existence of polar mutation at minisatellites has been reported.[92,195,251,437,438,454,624,625,676,677] The general conclusion is that most mutations involve the preferential gain of one or a few repeat units at one end of the tandem array (dissenting data are given in Huang et al.[420] and Sajantila et al.[684]). However, whether stepwise mutations have a propensity toward expansion is still the subject of debate.[183,458,624,684]

This discussion may be settled by the suggestion that longer alleles preferentially contract whereas shorter alleles preferentially expand.[250,394,420,693,863] Huang et al. canvas the issue and provide further references to other papers discussing potential reasons as to why allele size does not increase endlessly, and produce very large alleles. Note also the dissenting data regarding the prevalence of single step mutations in this paper.

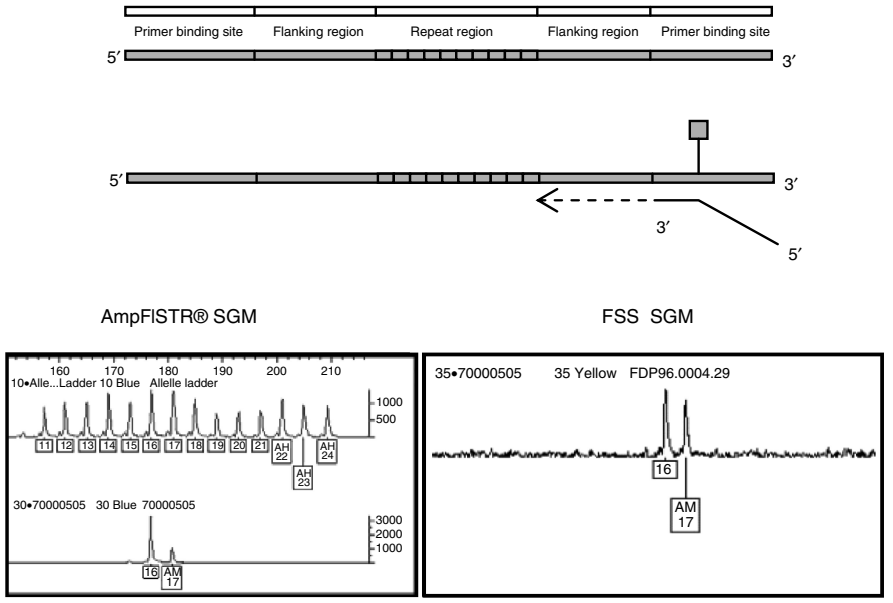


Figure 10.8 Primer binding site mutation affecting the amplification of the vWA 17 allele in the AmpF*l*STR® SGM$^+$ system (Applied Biosystems, Foster City, CA) whereas the allele has amplified efficiently using a different primer pair in the FSS SGM Multiplex.

Mechanisms for expansion, some including the effects of flanking DNA and induction of mutation by ionizing radiation, have been discussed.[393,433,437,438,824] One conclusion is that germline instability is controlled by elements outside the tandem repeat array.

The form of the mutations observed in a total of 400,000 meioses reported in several studies is summarized in Table 10.14.

An excess of paternal over maternal mutations has been reported.[92,823,876] Hallenberg and Morling[385] and Butler and Reeder [138] report detailed mutation rates summarized in Tables 10.15 and 10.16. The data in these two papers come from surveys of laboratories, and in some cases they may report the same data.

A statistical analysis of these data suggests a strong effect of sex with paternal rates higher than maternal rates ($p < 0.001$ for a generalized linear model two-way analysis of variance). The difference in mutation rates for male and female is plausibly explained by the observation that spermatozoa have undergone more cell divisions than ova during development. Crow[209] discussed a dissenting report and gives an important overview.

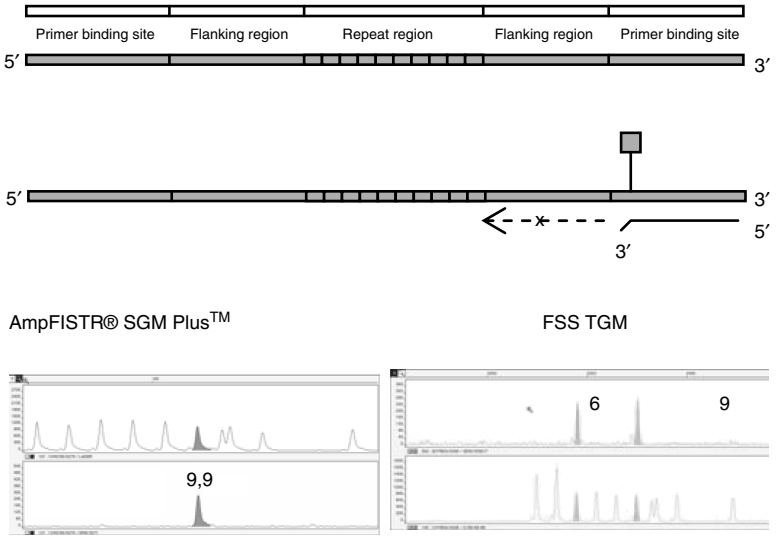


Figure 10.9 Due to a primer binding site mutation, the HUMTH01 locus amplified in the AmpF*l*STR® SGM^{+TM} appears as a 9,9 homozygote, whereas it is clearly a 6,9 heterozygote when analyzed under the FSS TGM (TGM stands for third-generation multiplex). The primer binding site mutation has caused the 6 allele to be silent under the SGM^{+} conditions.

Some loci have higher mutation rates than others ($p < 0.001$). Neither of these results is unexpected. However, the difference between male and female mutation rates was not consistent across all loci ($p = 0.007$, Figure 10.10). Where the sex effect is larger than expected (e.g., HUMFIBRA/FGA), the male rates are higher than the fitted line and the female rates are lower. Where the sex effect is smaller than expected (e.g., HUMD21S11), the male rates are lower than the fitted line and the female rates are higher.

One accrediting body, NATA,[574] states that loci with a mutation rate of greater than 0.25% cannot be used in parentage testing. This has caused some difficulty in court. This recommendation has been removed from the criminal testing guidelines but not the parentage testing guidelines.[610] We presume that the scientific motivation is to avoid apparent exclusions caused by mutation. However, if a likelihood ratio approach utilizing a mutation model is employed, this type of recommendation is not required.

The simplest possible mutation model is based on two assumptions:

1. At any locus, an allele has an equal probability of mutating to any other allele at that locus.
2. The rate of mutation from one allele to another is the same across all loci.

Table 10.14 Some Data on the Nature of the Mutation at STR Loci

Reference	Number of Meioses	Number of Mutations	Details
Clayton et al.[183]	50,274 SGM$^+$ loci	57	13 of 57 causing a null, 44 single step
Brinkman et al.[92]	10,844 9 autosomal loci	23	22 of 23 mutations were single step, 1 two step[f]
Weber and Wong[823]	~20,000 12 tetranucleotide microsatellite loci located on chromosome 19	35	13 of 13 established events were single step
Amorim et al.[16]	2899 12 autosomal loci	3	3 of 3 single step
Dauber et al.[218]	3830 10 autosomal loci	11	7 single step, 1 two step, 3 multistep
Sajantila et al.[684]	16,455	11	8 single step, 3 multistep
Xu et al.[876]	287,786 122 different autosomal tetranucleotide markers	236	Approximately 87% single step
Son et al.[721]	303 × 9 9 X chromosomal loci	29	11 minus 1 step, 3 minus 1 base, 6 plus 1 step, 2 plus one base, 4 other, 3 undetermined
Huang et al.[420]	362 autosomal dinucleotide loci	97	40:47 paternal: maternal of the 87 whose parental origin was unambiguous, 63% multistep

It is evident from empirical studies that such a model is a gross oversimplification. There is now compelling and cogent evidence that the probability of a mutation varies according to the sex of the individual and maybe the age, the locus being considered, which particular allele is being considered, the magnitude of change (number of steps), and the direction of that change.

A general mutation model relaxes these assumptions. At a given locus, we denote allele-specific mutation probabilities where the probability that allele X mutates to allele Y is denoted as $\mu_{X \to Y}$. Since paternal and maternal mutation rates differ, we will write these as $\mu^P_{X \to Y}$, $\mu^M_{X \to Y}$ respectively. Where we

[f] Eleven of 23 at the highly polymorphic ACTBP2 locus.

Table 10.15 Mutation Rates for VNTR Loci from Hallenberg and Marling[385] with permission from Elsevier

VNTR		Paternal		Maternal	
		Meioses (*N*)	Mutation Rate(%)	Meioses (*N*)	Mutation Rate (%)
D1S7	MS1	1157	4.24	1448	3.80
D1S80		1522	0.26	1507	0.13
D2S44	YNH24	9843	0.23	10,319	0.17
D4S139	Ph30	2245	1.25	2671	0.15
D5S110	Ms621	1833	1.71	2106	0.52
D7S21	MS31	10,704	1.47	11,293	0.06
D7S22	G3	2604	0.77	3071	0.10
D12S11	MS43a	10,270	0.10	10,780	0.01
D16S309	MS205	2452	0.82	2804	0.36

require the probability that allele X mutates to any other allele at the locus, we will write this as $\mu_{X \to *}$. Currently, the most prevalent data that exist are for locus mutation rates (μ).

Assume (probably falsely) that mutations occur independently from one another so that the probability of observing two mutations can be estimated by simple multiplication. Strictly, assuming the independence of mutational events may not be justified. For instance, a deficiency in a certain DNA repair enzyme may mean that multiple mutations occur in that individual with a greater frequency than might be expected. Or total mutation rates may be conditional on age, sex, or any other factor we have not yet discovered.

Dawid et al.[222] have recently developed a model based around these principles and comment, quite correctly, that "… the forensic treatment of sporadic parent/child inconsistency is sensitive to the mutation model assumed to underlie interallelic transitions." They compare four possible models:

1. Uniform: where every interallelic transition is equiprobable.
2. Proportional: where the probability of an interallelic transition is proportional to the frequency of the product allele.
3. Decreasing: which gives most probability to steps of the ± 1 repeat motif.
4. Stepwise stationary: again, this gives most probability to steps of the ± 1 repeat motif, but is modified by the frequency of the originating allele. This is the only model of the four that possesses stationarity; that is, that the allele probabilities should not change over time for large populations.

Empirical data (Table 10.14) suggest that small alleles (those at the low molecular weight end of the allelic range) are most likely to mutate to increased motif number. The most likely transition is the $+1$ motif.

Table 10.16 Mutation Rates for STR Loci from Hallenberg and Morling[385] (Upper Line with permission from Elsevier) and Butler and Reeder (Lower Line)[138]

STR	Paternal		Maternal	
	Meioses (*N*)	Mutation Rate(%)	Meioses (*N*)	Mutation Rate (%)
Amelogenenin	5753	0.05	5418	0.02
CSF1PO	4596	0.13	5038	0.02
	243,124	0.13	47,843	0.03
D2S1338	1755	0.23	2295	0.09
D3S1358	5762	0.21	6837	0.04
	8029	0.11	4889	0.00
D5S818	5113	0.16	5634	0.07
	130,833	0.15	60,907	0.04
D7S820	5031	0.10	5933	0.03
	131,880	0.15	50,827	0.03
D8S1179	2613	0.34	3295	0.03
	10,952	0.26	6672	0.07
D13S317	5545	0.20	5603	0.04
	69,598	0.15	59,500	0.06
D16S539	1544	0.19	2088	0.05
	48,760	0.08	42,648	0.03
D18S51	3346	0.30	3043	0.03
	9567	0.30	8827	0.09
D19S433	2582	0.12	2167	0.09
D21S11	4200	0.19	4029	0.15
	6754	0.18	6980	0.24
F13A01	723	0.14	760	0.00
FIBRA (FGA)	6926	0.46	7504	0.01
	189,973	0.29	8253	0.01
Penta D	233	0.00	325	0.00
Penta E	415	0.24	384	0.00
THO1	6242	0.02	7112	0.00
	74,426	0.02	42,100	0.01
TPOX	4579	0.04	5098	0.00
	45,374	0.02	28,766	0.01
vWA	17,717	0.30	19,116	0.02
	250,131	0.34	58,839	0.03

It appears, from limited data, that most multistep transitions are equiprobable. The bias to positive change in motif number declines as the alleles become larger, until the bias is toward a decrease in motif number for alleles at the high molecular weight end of the allelic range. It is conceivable that this model could also exhibit stationarity, although there is no absolutely compelling reason to expect stationarity in human populations.

Clayton et al. (unpublished results) suggest that mutations at differing loci may not be independent; this is to say that the rate of mutations at multiple

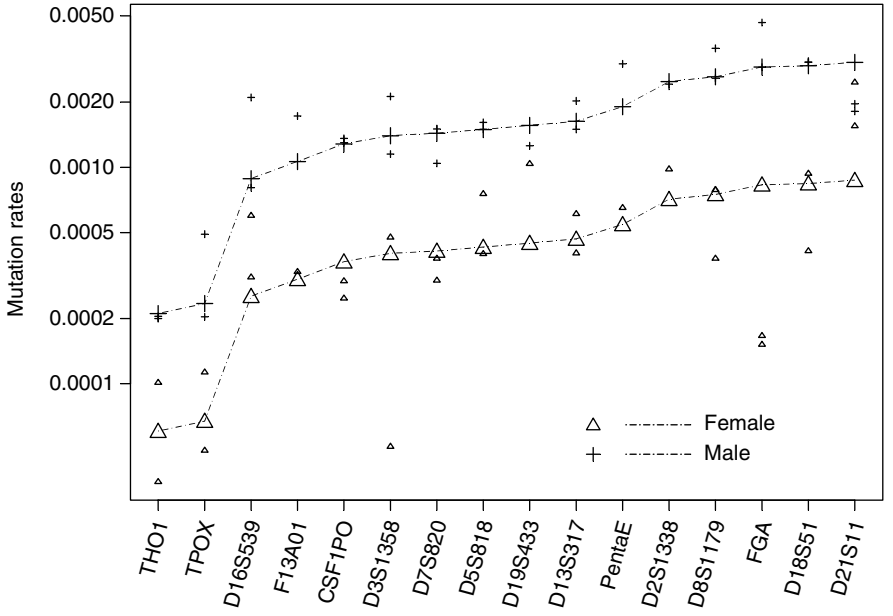


Figure 10.10 Maternal and paternal mutation rates. The smaller symbols are the observed data, and the larger symbols are the fitted values for the model of no sex × locus effect.

loci may be in excess of the simple product of single-locus mutation rates. Fortunately, this effect is likely to lead to conservative likelihood ratios if ignored, for trios exhibiting apparent multiple mutations.

The work of Clayton et al. strongly suggests the need for specific locus *and* allelic mutation rates preferably from empirical data or very robust models.

10.7.1 Paternity Trios with "Apparent Mutation"

10.7.1.1 Mother, Child, and Alleged Father — Father Does Not Possess the Paternal Allele

Hallenberg and Morling[385] report that the laboratories in their survey used four different formulae for calculating the *PI* when an apparent mutation had occurred. Unfortunately, the formulae in use are not given.

Consider a situation where the alleged father does not possess the paternal allele, A_p, such as $G_M = A_M A_M$, $G_{AF} = A_1 A_2$, $G_C = A_M A_p$. Under H_p, we hypothesize that the alleged father, *AF*, is the true father; hence one of his alleles must have mutated. If we further add the possibility that the maternal alleles may have mutated, we end up with a complex sum of terms.

$$\Pr(G_C|G_M, G_{AF}, H_p)$$
$$= \left(1-\mu^M_{A_M\to *}\right)\frac{\mu^P_{A_1\to A_p}+\mu^P_{A_2\to A_p}}{2}+\mu^M_{A_M\to A_p}\frac{\mu^P_{A_1\to A_M}+\mu^P_{A_2\to A_M}}{2}$$

However the second term

$$\mu^M_{A_M\to A_p}\frac{\mu^P_{A_1\to A_M}+\mu^P_{A_2\to A_M}}{2}$$

accounts for two mutational events and is expected to be small and may be ignored.

$$\Pr(G_C|G_M, G_{AF}, H_d) \approx \left(1-\mu^M_{A_M\to *}\right)p_{A_p}+\mu^M_{A_M\to A_p}p_{A_M}$$

and the numerical value of the second term $\mu^M_{A_M\to A_p}p_{A_M}$ is small compared to the first term. Ignoring these terms of small order, we have that

$$PI \approx \frac{\mu^P_{A_1\to A_p}+\mu^P_{A_2\to A_p}}{2p_{Ap}}$$

Recent debate among the members of the NIFS discussion group (National Institute of Forensic Sciences of Australia) has questioned whether the denominator can be treated in this simple way. We conclude that the short answer is yes. We consider the situation where mutation is considered in the denominator of the likelihood ratio. Hence an allele A_i may come from a random man possessing allele A_i, which has not mutated, or it may come from someone else whose allele has mutated to A_i. This suggests

$$\Pr(A_i^{N+1}) = \Pr(A_i^N)(1-\mu^p_{A_i\to *})+\sum_{j\neq i}\Pr(A_j^N)\mu^p_{A_j\to A_i}$$

where the Nth generation is the parental one and the $(N+1)$th the progeny. Under assumptions of stationarity $\Pr(A_i^{N+1}) = \Pr(A_i^N)$, but even without stationarity allele frequencies will be approximately stable across generations and $\Pr(A_i)^{N+1} \approx \Pr(A_i)^N$ except in rare and unlikely circumstances.

Dawid et al.[222] discuss a situation where it is unclear whether we have a maternal or a paternal mutation. We follow their treatment.

Consider $G_M = ab$, $G_C = ac$, $G_{AF} = aa$. It is unclear whether the paternal allele is a, and has been passed unmutated (and hence that the maternal allele must have mutated) or whether the maternal allele is a (and hence that the paternal allele has mutated). Again, ignoring those outcomes that posit

simultaneous paternal and maternal mutations, we have

$$PI \approx \frac{\dfrac{\mu^M_{a\to c} + \mu^M_{b\to c}}{2} + (1-\mu^P_{a\to *}) + \dfrac{\mu^P_{a\to c}}{2}\,(1-\mu^M_{a\to *} + \mu^M_{a\to b})}{p_c\,\dfrac{(1-\mu^M_{a\to *} + \mu^M_{b\to a})}{2} + \dfrac{\mu^M_{a\to c} + \mu^M_{b\to c}}{2}\,p_a}$$

$$\approx \frac{\mu^M_{a\to c} + \mu^M_{b\to c} + \mu^P_{a\to c}}{p_c}$$

for small mutation rates. This result differs from Dawid in only minor detail. Ayres[28] discusses the addition of the subpopulation correction.

10.7.1.2 Mother, Child, and Alleged Father — The Possibility of Silent (Null) Alleles

This issue arises when one alleged parent is an apparent homozygote, and the child is an apparent homozygote, and we have a system that has silent alleles, collectively represented by the symbol ϕ. We present below expressions for the paternity index accounting for the uncertainty induced by silent alleles. Typically such silent alleles are caused by a primer binding site mutation (see Chapter 1 for a brief discussion and some probabilities of observing silent alleles at different loci). However, as Clayton[181] has suggested, algebraic solutions are markedly inferior to a biological resolution of the uncertainty. The most obvious way to achieve this is to use a different primer to resolve the apparent silent allele.

Hallenberg and Morling[385] report that 23% of laboratories surveyed in 2000 and 64–82% in 2001 did not consider the possibility of silent alleles. If it was considered, then a large number of different formulae were used to calculate the *PI*.

Consider the situation $G_M = ab$, $G_C = b$, $G_{AF} = c$ and denote the allele probability of the silent allele as p_ϕ. Our treatment is completely general and could be applied to many other situations in this book where the use of a particular primer system suggests that a locus has silent alleles.

We have written the child's genotype as b rather than bb because, in any system that has silent alleles, it is difficult to be certain that the child has the homozygous genotype, bb, rather than the heterozygous $b\phi$ genotype. Similarly, the alleged father may be either cc or $c\phi$.

Assuming Hardy–Weinberg equilibrium, and in the absence of any other pedigree data, conditional probability arguments show that the apparent c homozygotes are composed of a fraction

$$\frac{p_c^2}{p_c^2 + 2p_c p_\phi} = \frac{p_c}{p_c + 2p_\phi}$$

who are actually *cc* homozygotes and a fraction

$$\frac{2p_c p_\phi}{p_c^2 + 2p_c p_\phi} = \frac{2p_\phi}{p_c + 2p_\phi}$$

who are $c\phi$ heterozygotes. Hence

$$\Pr(G_C|G_M, G_{AF}, H_p) = \frac{1}{2} \times \frac{1}{2} \times \frac{2p_c p_\phi}{p_c^2 + 2p_c p_\phi} = \frac{p_\phi}{2(p_c + 2p_\phi)}$$

if we consider a silent allele alone, and

$$\Pr(G_C|G_M, G_{AF}, H_p)$$

$$\approx \frac{1-\mu^M_{b\to *} + \mu^M_{a\to b}}{2} \times \left[\begin{array}{l} \dfrac{p_\phi}{2(p_c+2p_\phi)} \quad \left(1-\mu^P_{\phi\to *} + \mu^P_{\phi\to b} + \mu^P_{c\to b}\right) \\ \\ + \dfrac{p_c}{p_c+2p_\phi}\left(\mu^P_{c\to b} + \mu^P_{c\to *}\right) \end{array} \right]$$

$$+ \left(\frac{\mu^M_{a\to\phi} + \mu^M_{b\to\phi}}{2} \right) \left(\frac{p_\phi}{2(p_c+2p_\phi)} \left(\mu^P_{\phi\to b} + \mu^P_{c\to b}\right) + \frac{p_c}{p_c+2p_\phi}\, \mu^P_{c\to b} \right)$$

if we consider both silent alleles and mutation. Assuming small mutation rates,

$$\Pr(G_C|G_M, G_{AF}, H_p) \approx \frac{p_\phi + 2p_c\left(\mu^P_{c\to b} + \mu^P_{c\to *}\right)}{4(p_c + 2p_\phi)}$$

$$\Pr(G_C|G_M, G_{AF}, H_d) \approx \frac{1 - \mu^M_{b\to *} + \mu^M_{a\to b}}{2} (p_b + p_\phi) + \left(\frac{\mu^M_{a\to\phi} + \mu^M_{b\to\phi}}{2} \right) p_b$$

$$\approx \frac{\left(1 - \mu^M_{b\to *} + \mu^M_{a\to b}\right) p_\phi + \left(1 - \mu^M_{b\to *} + \mu^M_{a\to b} + \mu^M_{a\to\phi} + \mu^M_{b\to\phi}\right) p_b}{2}$$

$$\approx \frac{p_\phi + p_b}{2}$$

and hence

$$PI \approx \frac{p_\phi + 2p_c\left(\mu^P_{c\to b} + \mu^P_{c\to *}\right)}{2(p_c + 2p_\phi)(p_\phi + p_b)}$$

10.7.2 Mutation and Nonautosomal DNA

10.7.2.1 Y Chromosome

Consider a paternal relative, mother, child trio or paternal relative, child duo. We assume an exact match in haplotype of the Y chromosome between the paternal relative and the child. If we ignore mutation we obtain $PI \approx 1/f$, where f is the frequency of the child's (and alleged father's) haplotype.

Assume a mutation rate μ_l at each locus, l of the N loci typed on the Y chromosome. For the exact match described above,

$$PI = \frac{\prod_{l=1}^{N}(1-\mu_l)}{f} \approx \frac{(1-\overline{\mu})^N}{f}$$

where $\overline{\mu}$ is the average mutation rate across loci. We can find a series of closer approximations to the actual value of the PI if we have an estimate of the variance of the mutation rates about their mean,

$$s^2 = \frac{\sum_{i=1}^{N}(\overline{\mu}-\mu_i)^2}{N-1}$$

The first approximation is given by

$$PI = \frac{\prod_{l=1}^{N}(1-\mu_l)}{f} \approx \frac{(1-\overline{\mu})^N}{f}\left(1 - \frac{N-1}{2}\,\frac{s^2}{(1-\overline{\mu})^2}\right)$$

Assume that the paternal relative and the child differ at locus k and that m meioses have occurred between the child's Y chromosome and that of the alleged paternal relative. At locus k, the change is from allele x to y ($x \to y$).

$$PI \approx \frac{\prod_{l=1,\,l\neq k}^{N}(1-\mu_l)^m m\mu_k^{x\to y}(1-\mu_k)^{m-1}}{f}$$

Assuming that $\overline{\mu}$ is small, we obtain

$$PI \approx \frac{m\mu_k^{x\to y}}{f}$$

Assuming further that the $x \to y$ transition is a single step and occurs in approximately $\frac{1}{2}$ of all mutations at this locus, and also assuming a constant mutation rate across loci leads to the approximation

$$PI \approx \frac{m\overline{\mu}}{2f}$$

as given by Rolf et al.,[669] where f is now the frequency of the child's (not the paternal relative's) haplotype.

Some data on Y chromosome mutation rates using deep-rooted pedigrees have been published and are summarized in table 9.1.

These rates compare favorably with an average of 0.21% for Weber and Wong.[823] Kayser et al.[454] note that their data show a mutational bias toward increased length. The ratio of increase to decrease was 10:4 for the 14 mutations observed. Of these mutations, 13 were a single repeat change and one was a double repeat change.

10.7.2.2 X Chromosome Analysis

The interpretation of X chromosome evidence when used in paternity analysis follows the same principles as described for the Y chromosome. Assume a mutation rate μ_l at each locus, l, of the N loci typed on the X chromosome. For the exact match

$$PI = \frac{\prod_{l=1}^{N}(1-\mu_l)}{f} \approx \frac{(1-\overline{\mu})^N}{f}$$

where $\overline{\mu}$ is the average mutation rate across loci. This may be approximated more closely as in the previous section (see Chapter 9 for mutation rates).

Assume that the paternal relative and the child differ at locus k and that m meioses have occurred between the child's X chromosome and that of the alleged paternal relative. At locus k, the change is from allele x to y $(x \to y)$.

$$PI \approx \frac{\prod_{l=1,\, l\neq k}^{N}(1-\mu_l)^m m\mu_k^{x\to y}(1-\mu_k)^{m-1}}{f}$$

Assuming $\overline{\mu}$ is small we obtain $PI \approx m\mu_k^{x\to y}/f$. Further assuming that the $x \to y$ transition is a single step and occurs in approximately $\frac{1}{2}$ of all mutations at this locus and a constant mutation rate across loci suggests the approximation $PI \approx m\overline{\mu}/2f$, which is the X chromosome analog to that given by Rolf et al.[669] for the Y chromosome. f is now the frequency of the child's (not the paternal relative's) haplotype.

If the X chromosome is used in maternity analysis, the development is a composite of the recombination analysis given above and a mutation analysis. It requires no new principles. We make use of the result that linkage need not be considered if the phase of the mother is unknown and there is only one meiosis in the pedigree.

Assume that the mother and the child differ at locus k. At locus k, the mother has alleles a and b. The child has a maternal allele y not possessed by the mother (and not from the father). Hence we have observed, under H_p, a

mutation of $a \rightarrow y$ or $b \rightarrow y$. This leads to the expression

$$PI \approx \left(\frac{\mu_k^{a \rightarrow y} + \mu_k^{b \rightarrow y}}{2} \right) \frac{\prod_{l=1, l \neq k}^{N} (1 - \mu_l)}{f}$$

Assuming constant μ and that $\overline{\mu}$ is small, we obtain

$$PI \approx \left(\frac{\mu_k^{a \rightarrow y} + \mu_k^{b \rightarrow y}}{2} \right) \frac{1}{f}$$

10.7.2.3 Mitochondrial Maternity Analysis and Mutation

Consider a maternal relative, father,[g] child trio or a maternal relative child duo. We assume an exact match in haplotype of the mtDNA between the maternal relative and the child. If we ignore mutation, we obtain $PI \approx 1/f$, where f is the frequency of the child's (and alleged maternal relative's) haplotype.

Assume a mutation rate μ_l at each nucleotide position, l, of the N positions typed on the mtDNA. For the exact match described above,

$$PI = \frac{\prod_{l=1}^{N} (1 - \mu_l)}{f} \approx \frac{(1 - \overline{\mu})^N}{f}$$

where $\overline{\mu}$ is the average mutation rate across sites. This may also be approximated more closely as in the previous sections. Assume that the haplotypes differ at position k and that m meioses have occurred between the child's haplotype and that of the alleged maternal relative. At position k, the change is from nucleotide x to y ($x \rightarrow y$).

$$PI \approx \frac{\prod_{l=1, l \neq k}^{N} (1 - \mu_l)^m \, m \mu_k^{x \rightarrow y} (1 - \mu)^{m-1}}{f}$$

Assuming $\overline{\mu}$ is very small, we obtain

$$PI \approx \frac{m \mu_k^{x \rightarrow y}}{f}$$

f is now the frequency of the child's (not the maternal relative's) haplotype.

10.8 Inconsistencies in the Mendelian Pattern

If the pattern of alleles at an STR locus is inconsistent with Mendelian principles, there can be several possible explanations, including: (1) the pattern

[g] The father is irrelevant for mtDNA analysis.

in an individual may contain three bands, (2) the child may have a band not present in the genotype of either the mother or the alleged father.

In the latter situation, either a mutation has occurred, or a silent allele is present, or the individuals are not related as postulated. This situation is often referred to as an "exclusion." Although we will continue to use this familiar terminology in subsequent sections, as we will illustrate, it is an inappropriate expression.

10.8.1 Three-Banded Patterns

There are at least three potential causes of three-banded patterns: somatic mutation, trisomy, and translocation. These three phenomena affect familial testing in differing ways.

10.8.1.1 Somatic Mutation

Rolf et al.[668] report a case of somatic mutation involving germline cells in which the alleles segregated separately. This is straightforward to interpret in the unlikely circumstance that the fraction of germ cells containing each allele is known. For sperm, this should be estimable from relative peak areas of each allele in a sperm sample. Without such a sample, estimation of the fraction is difficult. There is also the possibility that an apparent heterozygote with genotype *ab*, say, is a mosaic of genotypes *aa* and *ab*.

Consider the trio of apparent genotypes:

	Genotype
Mother	*dd*
Child	*ad*
Alleged father	*abc*

Suppose that the fraction of sperm carrying allele *a* is q_a. Suppose further that the mother has genotype $G_M = dd$ (Mendelian factor $M_M = 1$) and the child has genotype $G_C = ad$. Then the paternal Mendelian factor and numerator of the *PI* is q_a. The denominator of the *PI* is not affected by the mosaicism in any significant way (it is in a minor way when we assume correlation of genotypes such as implied by use of the subpopulation model). Using the product rule, this is, as before, p_a. Thus if the alleged father is a mosaic for a standard paternity trio, the paternity index becomes $PI = q_a/p_a$.

If we assume that q_a cannot be estimated directly for this case, then it may be possible to infer its value from background studies on how often germline cells are affected by mosaicism when, say, buccal cells exhibit such mosaicism.

10.8.1.2 Trisomy and Translocation

Suppose that an alleged father shows the pattern *abc* with all alleles having the same peak area; then trisomy or translocation may be suspected. These are difficult to differentiate from the electropherogram, but trisomy can be determined by a number of other methods. Translocation can produce two copies of either the same or differing alleles on one chromosome. If the two copies are the same, they produce one band in the profile but of twice the expected area. When the translocated elements are strongly linked, they will be inherited as one unit, and can be treated as one rare allele (whether or not they are the same allele or they differ).

Suppose that we have the trio of apparent genotypes:

	Genotype
Mother	*dd*
Child	*abd*
Alleged father	*abc*

where we believe that the alleged father has a translocation from his genotype *and* that trisomy has been excluded (this cannot be inferred from observation of the genotype of the child in question, although the genotype of other acknowledged children of the alleged father can provide information as to his genotype). Then we do not know whether the alleles *a*, *b* are a pair on the same chromosome or whether the alleged father has some other pairing of alleles. If we assume that alleles *a*, *b* are on the same chromosome, then $\Pr(G_C|G_{AF}, G_M, H_p) = \frac{1}{2} \times M_M$ where M_M is the mother's Mendelian factor. This assumption maximizes the numerator of the paternity index. This is not appropriate unless the alleged father has other children whose genotypes have been examined or if STR examination of single sperm has been made. In the absence of such data, it may be reasonable, *a priori*, to assume that each of the three possible pairs of the alleles *a*, *b*, and *c* are equally likely. There are three combinations:

ab	*c*
ac	*b*
bc	*a*

and hence six possible sperm types. Accordingly, we write

$$\Pr(G_C|G_{AF}, G_M, H_p) = \frac{M_M}{6}$$

10.9 "Exclusions"

For the purposes of illustration, we consider the situation where STR tests have been employed at a number of loci to address the question of whether or not there is any evidence to support the proposition that the alleged father (*AF*) is the true biological father of a child (*C*). The alternative hypothesis is that some other unrelated man is *C*'s biological father. We could then pose the question:

> If *AF* is not *C*'s father, how many inconsistent loci would be required for a declaration of nonpaternity?

Many laboratories use a system where, if a predetermined number of inconsistencies are present, "exclusion" can be declared. Such a "numerical standard" would be dependent on the number of loci tested. For example, consider the evidential weight of two inconsistencies in a panel of six loci compared to that from two inconsistencies in a panel of 66 loci. Accordingly, for some laboratories this standard for exclusion is one, for others two, and for some three out of a standard panel of anywhere between 10 and 20 loci. Nutini et al.[586] report a double inconsistency in what they conclude is a true trio.

Not all mutations should carry equivalent evidential "weight". Certain types of mutations are inherently much less common than others.

We suggest that the most satisfactory approach is to use a likelihood ratio, or paternity index, combined with a mutation model. There is no necessity for a laboratory standard stipulating the number of inconsistent loci that must be present in a panel of STR loci before an exclusion is declared. It is desirable that such a mutation model be able to weigh changes according to the particular type of mutation involved. Bayesian methods offer a simple and logically coherent method of incorporating the effects of a mutation into the overall calculation of the paternity index.

If the paternity index is greater than one, even though one or more exclusions have been observed, there is still support for the proposition of parentage versus nonparentage notwithstanding the observed inconsistencies.

10.9.1 Mutation and Exclusion Probabilities[h]

The paternity index *PI* is a far superior measure of evidential weight than an exclusion probability, *PE*. However, it is possible to modify exclusion probabilities to account for mutations. There is an obvious interaction between what we

[h] This section was motivated by an inquiry by Tim Sliter.

define as an exclusion and the calculation of an exclusion probability. If we allow at most one apparent mutation in all L loci, then Chakraborty and Stivers[163] give

$$1 - \prod_{l=1}^{L} (1 - PE_l) \left\{ 1 + \sum_{j=1}^{L} \frac{PE_j}{(1 - PE_j)} \right\}$$

(we can confirm their derivation of this formula). These authors also give a recursive equation that allows the calculation allowing any number of mutations.

Let $Q_l = \sum_i \Pr(A_i)$, where the index i enumerates all those alleles (at locus l) allowed to mutate to the paternal allele. Suppose that we only allow alleles to mutate by ± 1 repeat; then $Q_l = \Pr(A^l_{p-1}) + \Pr(A^l_{p+1})$, where A^l_{p+1} and A^l_{p-1} are the alleles $+1$ and -1 repeats from the paternal allele (A^l_p), respectively. Then

$$PE = 1 - \prod_{l=1}^{L} (1 - PE_l) \left\{ 1 + \sum_{l=1}^{L} \frac{Q_l \, (2 - Q_l - 2\Pr(A^l_p))}{1 - PE_l} \right\}$$

If we allow the possibility that all alleles at a locus may mutate to the paternal allele, then $Q_l = 1 - \Pr(A^l_p)$ and we recover Chakraborty and Stivers's formula.

However, we are concerned that under some circumstances the exclusion probability calculated in this way may be misleading. The *PE* may give an indication of some substantial evidential value whereas the more reliable *PI* does not.

Disaster Victim Identification, Identification of Missing Persons, and Immigration Cases

11

JOHN BUCKLETON, CHRISTOPHER TRIGGS, AND TIM CLAYTON

Contents

11.1 Introduction

In earlier chapters we have considered the evaluation of evidence from single genetic profiles, from mixtures of profiles derived from two or more unrelated individuals, and in the previous chapter from three individuals, two of whom are the parents of the third. In this chapter we generalize the approach from a paternity trio to consider evidence from the profiles of a number of individuals who are members of a more complex pedigree. The types of situations the methods discussed in this chapter will allow us to consider are the identification of missing persons, the evaluation of the closeness of genetic relationships in immigration cases,[a] and the identification of victims of disasters. In each of these cases, we observe the profiles of several people and wish to use these to compare the likelihood of observing two or more postulated sets of relationships between the individuals.

The purpose of disaster victim identification may be to bring justice or closure to the families of the missing, and also to show respect to the victims themselves by treating them with individual dignity.[867] It is questionable whether scientific or historic curiosity is an adequate substitute for these motives and such endeavors must be dealt with very carefully. There are some mandated rights and responsibilities in this area. For example, the Geneva Convention defines a responsibility of states to help in the location of graves of persons who have died while in their detention.[289,332]

The identification of human remains has been undertaken without the aid of DNA typing for many years, and a wide range of approaches are still appropriate for differing circumstances.[203,330,717] However, in recent times DNA technology has been extensively and usefully applied to the identification problem. The identification of the remains may only be a part of the problem. It is important to remember that the position of the body or the burial may be a crime scene in itself and that much information may be obtained from a proper scene examination.[237] However, the identification of remains presents the forensic scientist with the dual problems of obtaining a profile from remains that are often in a state of advanced decomposition and

[a] We will often refer to the "body" in this chapter. For immigration cases, this is inappropriate and a term such as "applicant" should be substituted. The mathematics, however, are identical.

obtaining a reference profile for the missing person.[b] In this chapter we address these issues and discuss both the practical and interpretational difficulties surrounding them.

11.2 Mitochondrial or Nuclear DNA?

Before embarking on practical measures to identify human remains, it is usually necessary to consider several factors that will affect the decision as to whether to proceed with mtDNA typing (see Chapter 9) or nuclear DNA profiling.[629]

11.2.1 MtDNA

In certain circumstances, mtDNA offers a number of advantages over nuclear DNA.[327,364] High numbers of mtDNA molecules are present in every cell; therefore, there is a greater chance of some template surviving degradation. It is relatively easy to extract typeable quantities of material from hair shafts and other skeletal structures where the amount of nuclear DNA may be very limited indeed. Lastly, in cases where reference samples from immediate genetic relatives are scant (so-called genetic deficiency cases), mtDNA is useful as it can bridge the gaps in the pedigree. Thus, any relative on the maternal lineage can be used for comparison even though they may be separated by more than one generation. In this way, Gill et al.[351] were able to use surviving relatives of Tsar Nicholas II to identify his putative remains, as was discussed in Chapter 9.

However, the relative lack of discriminating power associated with mtDNA typing needs to be balanced against these advantages. This may be a major factor if multiple sets of remains are present. Similarly, if those remains were from individuals who were matrilineally related, one would be unable to distinguish between them.

11.2.2 Nuclear DNA

Consideration should also be given to which nuclear DNA typing system to use. If patrilineal relatives are available in genetic deficiency cases, a panel of Y chromosome markers may be appropriate. Of course, these do not have the advantage of multiple copy number. However, in the majority of cases, autosomal multiplex STR profiling will be the method of choice. The very high discriminating power afforded by STR loci combined with gender information has proven to be an excellent tool, especially when multiple sets of remains are present. In 1995 Clayton et al.[184] applied such technology to

[b] It is quite useful to use the term "missing person" to refer to the person from the pedigree. The body is not known to be from the pedigree; indeed that is the matter in question. Hence "missing person" can be used to refer to the identity of the body only under H_p but not under H_d.

identify multiple victims of the Waco incident. More recently, we have seen methods used in a large number of incidents.[44,83,88,143,364,375,496,593,799] This technology is now used routinely to identify human remains. It has been applied in massive programs such as those to identify victims of ethnic cleansing in the former Yugoslavia and victims of the September 11 attack on America, and includes compromised samples such as the successful typing of the remains of a neonate who had been kept in a vinyl bag for 15 years.[552] The value of a multidisciplinary approach is emphasized by an example from the World Trade Centre human identification project.[116] Due to co-mingling of soft tissues, DNA testing alone would have led to problems, and a partnership between forensic anthropologists and DNA specialists was advocated.

11.3 Human Remains — Obtaining a Profile from Bodily Remains

When human remains are found, they may be in a variety of conditions ranging from recently deceased to fully skeletonized. A body may be intact but, more often than not, the remains that are of forensic interest are fragmentary due to:

1. Physical forces (such as the violent impacts encountered in transport accidents, explosions, or mechanical wave action).
2. Scavenging and predation by land or marine organisms.
3. Dismemberment and/or disarticulation by an individual seeking to dispose of or conceal the body or its identity.

Moreover, the remains may have been exposed to fire (thermal insult) or to aggressive compounds such as lime or acid (chemical insult). The level of decomposition of the tissues in human remains can vary dramatically according to the time since death and the prevailing environmental conditions. Extreme cold can preserve remains as though they had been deep frozen while hot dry conditions can lead to complete desiccation and mummification. Conversely, warm damp conditions or water immersion can lead to rapid decomposition. By contrast, an aquatic but cold and/or anaerobic environment can sometimes serve to preserve tissue. A recently reported European collaborative exercise on artificially degraded DNA samples[696] gives insight into some likely effects of degradation and their prevalence. These authors were able to compare normal 28-cycle approaches with "enhanced" approaches. These enhanced approaches included elevated cycle number or increased template. They noted strong peak imbalance, particularly for the enhanced conditions, artifactual peaks, allelic dropout, and the reporting of some erroneous alleles. Many of these effects are similar to those encountered in LCN casework. The authors made recommendations for dealing with degraded samples.

For the purposes of this discussion, it is useful to categorize remains based upon the extent of decomposition. The approach to obtaining DNA from the remains can differ, depending on which category is being considered.

11.3.1 Category 1: Remains Displaying Relatively Few Signs of Decomposition

At the post-mortem examination, it will generally be possible to obtain blood samples from some areas of the body, and the internal organs and soft tissues will be largely unaffected. As the blood and soft tissues are rich sources of DNA, obtaining a DNA profile from remains in this category should present few, if any, technical difficulties.

11.3.2 Category 2: Remains Exhibiting Partial Decomposition

Generally at the post-mortem examination, no blood will be available and the superficial soft tissues may be exhibiting signs of putrefaction. Successful typing of the soft tissues and sera may still be possible, depending on the temperature in the area of death and the pathological conditions that the victim experienced.[753] Deep tissues such as psoas muscle or bone marrow will usually be in a reasonable condition. In this situation, targeting deeper tissues is often more successful than attempting to use more superficial tissues as a source of DNA. Contamination of the extract by decomposition products is known to inhibit subsequent PCR. Pusch et al.[629] review some potential inhibitors. These include Maillard reaction products, remains of porphyrins, degraded nucleic acids, soil components such as humic and fulvic acids, tannins and ferric ions, and in particular degraded human collagen type I. Normally, tissue samples will require additional treatment to release the DNA and purify it from the cell debris and other decomposition products. The most common protocol employed is to use a buffered solution containing Proteinase K, SDS, and DTT followed by an organic cleanup using phenol or phenol/chloroform. Using ethanol precipitation or microfiltration further purifies the DNA. Other protocols have been published, all of which are designed to maximize recovery and minimize the presence of inhibitory decomposition products.

DNA profiling may also reveal DNA degradation (exhibited as a gradual loss of signal from the high molecular weight loci first). This degradation may be extensive enough to prevent a full profile from being obtained.

11.3.3 Category 3: Remains in an Advanced State of Decomposition

Typically, in this situation, most of the soft tissues will have lost their integrity and some may have formed adipocere. Remnants of bone marrow may still be present. If the remains are in such an advanced state of decay, obtaining a

DNA profile from liquefied tissues or adipocere is seldom successful. The bone marrow can sometimes be better preserved and may provide sufficient DNA. However, if profiling from the putrefied soft tissues and marrow is unsuccessful, then recourse should be made to skeletal structures (See Category 4 below).

11.3.4 Category 4: Remains that are Fully Skeletonized (Including Mummified or Desiccated Remains)

At this stage of decomposition, generally only hardy structures such as bone, hair, nails, and teeth will be available. The skeletal structures, bone matrix, and tooth pulp are rich in mtDNA. They also contain low, but typeable, quantities of nuclear DNA. As the DNA is encased within a hardened calcified matrix, it requires special procedures to release and purify it (see below). Hair shafts are relatively rich sources of mtDNA but contain only traces of highly fragmented nuclear DNA. Attempts have been made to prepare STR profiles from hair shafts or other sources using short amplicon strategies.[51,139] This is a very logical approach but has not yet been implemented widely.

11.3.4.1 Extraction of DNA from Bone, Tooth, Hair, and Nail

Extraction of DNA from tooth,[533,718] hair,[406,791] nail,[752,790] and bone[301,412] poses a significant technical challenge.[282,427] The preferred starting material is either a molar tooth (preferably free from an amalgam filling) or approximately 1 g of compact (noncancellous) bone. First, the surface of the tooth or bone is cleaned to remove surface contamination. This is often done by sanding away the surface layer of bone or, in the case of a tooth, by transient acid immersion followed by washing. Next, the tooth or bone must be powdered. Two methods are generally employed. The first utilizes a specialized piece of bone-milling equipment in which a bone chip is enclosed in a vial with a heavy metal bar. The vial is then submerged in liquid nitrogen to render the bone brittle. An oscillating magnetic field is then applied. This causes the metal bar to vibrate violently, pulverizing the bone chip into a fine powder. The same effect can be achieved by grinding using a drill bit, but care must be taken to avoid generating high temperatures as a result of frictional forces. Once a fine powder has been produced, the DNA is released by enzymatic digestion using a buffered solution of Proteinase K, DTT, and Tween 60. The resulting extract is heavily contaminated with calcium salts leached from the bone that are then removed by mixing with a concentrated solution of EDTA. Finally, the extract is cleaned and concentrated by microfiltration. Nevertheless, the yields of nuclear DNA are typically low and, often, elevated cycle number PCR ($>$ 28 cycles) is often needed to generate typing results.

Nakanishi et al.[573] report the successful typing of nails after various environmental insults. Their findings confirm the usefulness of nail material as a source of DNA.

11.4 Comparisons

Having successfully obtained a profile from the unidentified remains, a comparison must then be carried out with the profile of the missing person. However, it is often the case that the missing person's DNA profile is not available. Fortunately, there are a number of ways in which information can be obtained.

11.4.1 Surrogate Samples

Increasingly, we are seeing the use of a "surrogate" DNA sample, that is, a sample thought to contain DNA from the missing person. All 155 victims of the Kaprun cable car disaster in 2000 were identified within 19 days based mainly on the use of such surrogate samples.[549] From an evidential perspective, perhaps the best type of surrogate sample one can obtain is an archival clinical sample. Often clinics, hospitals, and other medical institutions will retain labeled and indexed archival specimens that are traceable through patient records. The most common format is as a wax-embedded histology block. However, microscope slides, deep frozen tissue specimens, and neonatal blood cards have all been utilized in the past. Tissue supposedly well preserved in formalin solution has proven to be highly refractory to PCR analysis.

In those cases with an active investigation, law enforcement agencies may obtain personal effects that may carry biological material. Generally, these are articles of a personal or individual nature, such as a toothbrush, razor, underwear, or hairbrush. However, a variety of other samples have, in the past, proven useful; for example, saliva beneath stamps, bedding, handkerchief, and cigarette butts.

Many jurisdictions have now instituted national DNA intelligence databases. These repositories often contain the DNA STR profiles from hundreds of thousands of individuals. The U.K. National DNA Database currently holds in excess of two million records. If a missing person is previously on record, then it may be possible to obtain a copy of that person's profile to assist in the identification of human remains.

11.4.1.1 Twins

If a sample from a surviving monozygotic twin is available, then the deceased should share the same DNA profile. However, consideration should be given

as to how the surviving individual knows that he/she is monozygotic. Similarity of appearance may provide strong evidence of monozygosity, but this is not in itself conclusive and, generally, genetic tests are required to establish that the individuals are monozygotic as opposed to dizygotic.

11.4.2 Pedigree Analysis

In the remainder of this chapter, we will describe the alternative approach — the use of pedigree (kinship) analysis. We will concentrate on the calculation of likelihood ratios based on the probability of the observed genotypes if the postulated relationship is true or false. This requires access to appropriate databases. Chakraborty and Jin[162] and Ehm and Wagner[247] have proposed summary measures based on the number of shared and nonshared alleles. This approach was tested on 315 mother–child and 91 full-sibling pairs and by simulation.[623] Such measures may be of some use as a screening tool or as a filter for computer software, but are less powerful than likelihood ratio methods because they do not take account of the relative rareness of shared alleles. For example, the evidence for a relationship is stronger if the shared alleles are rare than if they are common.

There are many possible variations in individual family trees. However, general guidance on how to proceed has been published.[763] It is possible to formulate some simple guidelines in terms of the "appropriateness" of obtaining samples from certain relatives. This stems from a consideration of the amount of genetic information they would contribute to the inference process. As a practical solution, investigators could be instructed to obtain samples from any or all first-degree relatives. In some cases, this will lead to a redundancy of information. For instance, if both parents of the missing person are available, then the genotypes of his/her siblings are redundant (although those of his/her children are not). In certain instances, there may be sufficient information from the children of missing persons to completely determine their genotype. In this case, the information given by the genotypes of the missing person's parents is redundant. However, obtaining samples from all available first-degree relatives may be the most expedient course of action.

11.4.2.1 General Principles

A pedigree specifies the relationships between a set of individuals. Those individuals at the top of the pedigree whose parents are not specified are termed the founders of the pedigree. Those individuals at the base of the pedigree who have no offspring are termed the final individuals. Pictorially, males are usually designated by squares and females by circles. In much of the work to come, the sex of a member of a pedigree is not relevant. In such cases, we may use a diamond as a general symbol.

We begin by considering unlinked loci. We imagine the situation where there is a body with genotype G_B. We have a pedigree of N people with genotypes $G_1 \ldots G_N$. We denote the pedigree information itself by I. Typically we seek to evaluate the evidence given two hypotheses:

H_p: The body is this person in this pedigree.
H_d: The body is not related to this pedigree.[c]

This suggests the evaluation of the likelihood ratio

$$LR = \frac{\Pr(G_B, G_1 \ldots G_N | H_p, I)}{\Pr(G_B, G_1 \ldots G_N | H_d, I)}$$

The joint probabilities in the numerator and denominator of the likelihood ratio are often evaluated by decomposition into a chain of conditional probabilities. For nonfounders, we condition on the genotypes of their parents. Sets of children can usefully be taken out together. Where the probability of a child is conditioned on both parents, conditioning on other members of the pedigree, either siblings or other ancestors, is not required. Where we have no genotype for a person in the pedigree, we consider all possible genotypes for that person and sum over their mutually exclusive probabilities. For unlinked loci, the equation

$$LR = \frac{\Pr(G_B, G_1 \ldots G_N | H_p, I)}{\Pr(G_B, G_1 \ldots G_N | H_d, I)}$$

can be approximated by

$$LR = \prod_{loci} LR_{locus}$$

These principles are difficult to follow in the abstract but become natural with practice.[187,198,407,431,512,534,571,761]

11.4.2.2 Parents

If samples are available from both parents of the deceased, then strong inferences can be drawn regarding his/her profile. Following the principles given in Chapter 10, it can be seen that, depending on the pattern of parental alleles, between one and four distinct combinations exist at each locus for their biological child.

We envisage a situation where we have a body with genotype G_B. We also assume that both the putative parents of the missing person are available: the mother (M) with genotype G_M and the father (F) with genotype G_F.

[c] Strictly speaking, the two hypotheses H_p, H_d specify two different pedigrees. We illustrate the difference in the discussion of Example 11.1.

Evaluating the evidence in such a scenario requires hypotheses of the following type:

H_p: The remains are from a biological child of M and F.
H_d: The remains are from an unknown person unrelated to M and F.

We evaluate the likelihood ratio

$$LR = \frac{\Pr(G_B, G_F, G_M|H_p, I)}{\Pr(G_B, G_F, G_M|H_d, I)}$$

First, using the third law of probability, we take out the youngest person, which is the body.

$$LR = \frac{\Pr(G_B|G_F, G_M, H_p, I)\Pr(G_F, G_M|H_p, I)}{\Pr(G_B|G_F, G_M, H_d, I)\Pr(G_F, G_M|H_d, I)}$$

We assume that the joint probabilities of observing the genotypes of the founders of the pedigree, that is, the mother and father, do not depend on the two hypotheses H_p and H_d,

$$\Pr(G_F, G_M|H_d, I) = \Pr(G_F, G_M,|H_p, I)$$

and hence

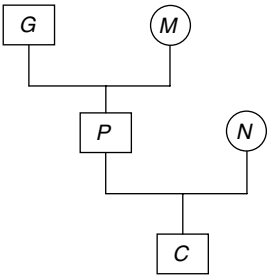
$$LR = \frac{\Pr(G_B|G_F, G_M, H_p, I)}{\Pr(G_B|G_F, G_M, H_d, I)}$$

The term $\Pr(G_B|G_F, G_M, H_p, I)$ can be assessed by applying Mendel's laws using a Punnett square. The term $\Pr(G_B|G_F, G_M, H_d, I)$ is the probability of observing the genotype of a person conditioned on those of two people to whom he/she is unrelated under H_d. This can be assessed either using the product rule or the subpopulation correction as desired.

It is worthwhile considering the usefulness of the subpopulation correction in this context. We believe that the product rule has a slight bias in favor of H_p whereas the subpopulation correction has a strong bias in favor of H_d. This is especially so if a conservative value is assigned for θ. If the intent of the case is to identify bodies to return the remains to the correct relatives, then there seems to be little value to the subpopulation correction. If the case is criminal, then there may be some value in using the subpopulation correction in order to give a conservative likelihood ratio to the court. If the subpopulation correction is applied, then we are assuming that the body is not part of this pedigree but is from the same subpopulation as the members of the pedigree. Birus et al.[69] give powerful evidence in support of this approach. Fung et al.[326] outline the general approach to incorporating the subpopulation correction.

Box 11.1 Evaluation of the Probability of Observing a Set of Genotypes Conditional on a Specified Pedigree

Suppose that we observe five genotypes: G_G, G_M, G_P, G_N, and G_C. We further believe that individual G and M are the parents of P, and that P and N are the parents of C, and denote this specification of relationships as H. We could also describe the relationships in a simple pedigree.



We notice that members of the pedigree fall into two classes: founders and nonfounders. The founders G, M, and N are defined to be unrelated,[d] and further we have not observed the genotypes of any of their parents. We can express the probability of observing the whole pedigree, $\Pr(G_P, G_G, G_M, G_N, G_C|H)$, as the product of the conditional probability of observing the nonfounders given the genotypes of the founders and the unconditional probability of observing the founders.

$$\Pr(G_P, G_G, G_M, G_N, G_C|H) = \Pr(G_P, G_C|G_G, G_M, G_N, H) \times \Pr(G_G, G_M, G_N|H)$$

We can further simplify $\Pr(G_P, G_C|G_G, G_M, G_N, H)$ into a product of the conditional probabilities of observing the genotype of an individual given the genotypes of his parents. For the simple pedigree above,

$$\Pr(G_P, G_G|G_M, G_N, G_C, H) = \Pr(G_C|G_P, G_N, H) \times \Pr(G_P|G_G, G_M|H)$$

If the genotypes of all of the founders of the pedigree are specified, evaluation of the probability of observing the pedigree is direct and follows Mendel's laws.

[d] Otherwise there would be more branches to the pedigree.

Table 11.1 gives the formulae for the likelihood ratio at a single locus. These formulae ignore mutation. For autosomal loci, there is no need to differentiate the mother from the father, so that we can call these persons parent 1 and 2. For the ten-locus SGM$^+$ system, this typically leads to values of the likelihood ratio of the order of 10^5 to 10^6.

If a sample from only one parent is available (because the identity of the other parent is unknown or because he/she is deceased), the likelihood ratio can still be calculated. Although inferences can still be drawn regarding the profile of the deceased, the numerical value of the likelihood ratio will be smaller and the "power" of the analysis is dramatically reduced.

The hypotheses are likely to be of the following form:

H_p: The remains are from a biological child of the parent.
H_d: The remains are from an unknown individual unrelated to the parent.

Table 11.1 Formulae for the Likelihood Ratio for Situations Where Genotypes of Both Parents (But No Children) of the Missing Person are Available as Reference Samples

Parent 1	Parent 2	Body	*LR* Assessed By the Product Rule	LR Assessed Using the Subpopulation Correction (both parents and the body are members of the same subpopulation)
aa	*aa*	*aa*	$\frac{1}{p_a^2}$	$\frac{(1+3\theta)(1+4\theta)}{(4\theta+(1-\theta)p_a)(5\theta+(1-\theta)p_a)}$
	ab	*aa*	$\frac{1}{2p_a^2}$	$\frac{(1+3\theta)(1+4\theta)}{2(3\theta+(1-\theta)p_a)(4\theta+(1-\theta)p_a)}$
		ab	$\frac{1}{4p_ap_b}$	$\frac{(1+3\theta)(1+4\theta)}{4(3\theta+(1-\theta)p_a)(\theta+(1-\theta)p_b)}$
	bb	*ab*	$\frac{1}{2p_ap_b}$	$\frac{(1+3\theta)(1+4\theta)}{2(2\theta+(1-\theta)p_a)(2\theta+(1-\theta)p_b)}$
	bc	*ab*	$\frac{1}{4p_ap_b}$	$\frac{(1+3\theta)(1+4\theta)}{4(2\theta+(1-\theta)p_a)(\theta+(1-\theta)p_b)}$
ab	*ab*	*aa*	$\frac{1}{4p_a^2}$	$\frac{(1+3\theta)(1+4\theta)}{4(2\theta+(1-\theta)p_a)(3\theta+(1-\theta)p_a)}$
		ab	$\frac{1}{4p_ap_b}$	$\frac{(1+3\theta)(1+4\theta)}{4(2\theta+(1-\theta)p_a)(2\theta+(1-\theta)p_b)}$
	ac	*ab*	$\frac{1}{8p_ap_b}$	$\frac{(1+3\theta)(1+4\theta)}{8(2\theta+(1-\theta)p_a)(\theta+(1-\theta)p_b)}$
		aa	$\frac{1}{4p_a^2}$	$\frac{(1+3\theta)(1+4\theta)}{4(2\theta+(1-\theta)p_a)(3\theta+(1-\theta)p_a)}$
	cd	*ac*	$\frac{1}{8p_ap_c}$	$\frac{(1+3\theta)(1+4\theta)}{8(\theta+(1-\theta)p_a)(\theta+(1-\theta)p_c)}$

Let G_B be the genotype of the body and G_P the genotype of the parent. We require

$$LR = \frac{\Pr(G_B|G_P, H_p, I)}{\Pr(G_B|G_P, H_d, I)}$$

The term $\Pr(G_B|G_P, H_p, I)$ can be assessed by simple Mendelian principles as before considering a missing parent. It is not necessary in this instance to enumerate all the possibilities for the missing parent as the expedient of a random allele as described in Section 10.2.5 of Chapter 10 can be used.[e]

The term $\Pr(G_B|G_P, H_d, I)$ is the probability of observing the genotype of a person conditioned on the genotype of another person to whom he/she is unrelated. This can be assessed using the product rule or the subpopulation correction as desired. Table 11.2 gives some likelihood ratios for a single parent situation ignoring mutation.

The values of the likelihood ratio obtained using the ten loci from the SGM$^+$ system are typically reduced to the order of 10^2 to 10^3.

If only a single parent is available, then the paucity of genetic information can often be supplemented by reference to other members of the pedigree if they are available.

11.4.2.3 Children

A sample from a child provides information regarding its biological parent. If the deceased had a number of children, then testing all of them will provide the most information regarding his/her genotype. Moreover, if a sample is available from the children's other parent, the process is aided as this often allows the determination of which allele was contributed by the deceased. Consider the situation where the male deceased has a wife (genotype *ab*) and two children (genotypes *aa* and *bc*). From this it can be deduced that at this locus the deceased's genotype is *ac*. Without the sample from the deceased's wife, there would have been two possible genotypes (*ab* or *ac*).

The magnitude of the likelihood ratio obtained from the analysis of children will be governed by the availability of a sample from the other parent, and also the number of children available for testing.

We consider the situation where there is one child and the partner of the missing person is available. The hypotheses are likely to be of the following form:

H_p: The remains are from the biological parent of the child.
H_d: The remains are from an unknown individual unrelated to the child.

[e] Recall that this expedient will not work if there are multiple children. In this formulation, the body is the only typed child of the parent and the missing parent. If, for example, siblings of the missing person are available, then the possible genotypes of the missing parent must be enumerated.

Table 11.2 Formulae for the Likelihood Ratio for Situations Where the Genotype of One Parent of the Missing Person is Available as a Reference Sample

Parent	Body	*LR* Assessed By the Product Rule	*LR* Assessed Using the Subpopulation Correction (mother, parent and body in the same subpopulation)
aa	aa	$\frac{1}{p_a}$	$\frac{1+2\theta}{3\theta+(1-\theta)p_a}$
	ab	$\frac{1}{2p_a}$	$\frac{1+2\theta}{2(2\theta+(1-\theta)p_a)}$
ab	aa	$\frac{1}{2p_a}$	$\frac{1+2\theta}{2(2\theta+(1-\theta)p_a)}$
	ab	$\frac{p_a+p_b}{4p_ap_b}$	$\frac{(1+2\theta)[2\theta+(1-\theta)(p_a+p_b)]}{4(\theta+(1-\theta)p_a)(\theta+(1-\theta)p_b)}$
	ac	$\frac{1}{4p_a}$	$\frac{1+2\theta}{4(\theta+(1-\theta)p_a)}$

Let G_B be the genotype of the body, G_C the genotype of the child, and G_P the genotype of the partner of the missing person. We require

$$LR = \frac{\Pr(G_B, G_P, G_C|H_p, I)}{\Pr(G_B, G_P, G_N|H_d, I)}$$

As usual, first we condition the genotype of the youngest person in the pedigree on the genotypes in the rest of the pedigree. In this example, the youngest person is the child.

$$LR = \frac{\Pr(G_B, G_P, G_C|H_p, I)}{\Pr(G_B, G_P, G_C|H_d, I)} = \frac{\Pr(G_C|G_B, G_P, H_p, I)\Pr(G_B, G_P,|H_p, I)}{\Pr(G_C|G_B, G_P, H_d, I)\Pr(G_B, G_P,|H_d, I)}$$

Again assuming that $\Pr(G_B, G_P \mid H_p, I) = \Pr(G_B, G_P \mid H_d, I)$, we see that

$$LR = \frac{\Pr(G_C|G_B, G_P, H_p, I)}{\Pr(G_C|G_B, G_P, H_d, I)}$$

The term $\Pr(G_C|G_B, G_P, H_p, I)$ can be assessed by simple Mendelian principles as before.

When we turn to the consideration of the pedigree under H_d, the term $\Pr(G_C|G_B, G_P, H_d, I)$ is the probability of observing the genotype of a child, given that of one parent (recall that B is not related to C and P under H_d), and should be assessed as previously for paternity duos. As usual, this can be assessed using the product rule or the subpopulation correction as desired. Some formulae for typical situations are given in Table 11.3. Table 11.4 gives the likelihood ratios when the partner of the deceased is not available.

Table 11.3 Formulae for the Likelihood Ratio for Situations Where the Genotypes of the Partner and One Child of the Missing Person are Available as Reference Samples

Body	Child	Partner of the Missing Person	*LR* Assessed By the Product Rule	*LR* Assessed Using the Subpopulation Correction (partner, child, and body in the same subpopulation)
aa	*aa*	*aa*	$\frac{1}{p_a}$	$\frac{1+3\theta}{4\theta+(1-\theta)p_a}$
		ab	$\frac{1}{p_a}$	$\frac{1+3\theta}{3\theta+(1-\theta)p_a}$
	ab	*ab*	$\frac{1}{p_a+p_b}$	$\frac{1+3\theta}{4\theta+(1-\theta)[p_a+p_b]}$
		bb or *bc*	$\frac{1}{p_a}$	$\frac{1+3\theta}{2\theta+(1-\theta)p_a}$
ab	*aa*	*aa*	$\frac{1}{2p_a}$	$\frac{1+3\theta}{2[3\theta+(1-\theta)p_a]}$
		ab or *ac*	$\frac{1}{2p_a}$	$\frac{1+3\theta}{2[2\theta+(1-\theta)p_a]}$
	ab	*ab*	$\frac{1}{p_a+p_b}$	$\frac{1+3\theta}{4\theta+(1-\theta)[p_a+p_b]}$
		bb or *bc*	$\frac{1}{2p_a}$	$\frac{1+3\theta}{2[\theta+(1-\theta)p_a]}$
	ac	*cc* or *cd*	$\frac{1}{2p_a}$	$\frac{1+3\theta}{2[\theta+(1-\theta)p_a]}$

It is worth noting a further feature of an analysis using children. Suppose that the genotype of the deceased's wife was *pr* and the four children at a locus were *pq*, *pq*, *qr*, and *qr*. On this occasion, the deceased can be inferred to carry the *q* allele. There are at least two possible explanations for this observation:

1. The deceased has genotype *qq* at this locus.
2. The deceased has genotype *qx* at the locus (where *x* is any allele other than *q*).

From this finding it can be seen that the first explanation, that the deceased is a *qq* homozygote, becomes more supported and the second explanation, that the deceased is a heterozygote, is less supported — since if the deceased was in fact *qx* then the probability of transmitting only allele *q* to each of his four children is $(\frac{1}{2})^4$, i.e., 1 in 16 compared with certain transmission for the homozygote. We see that one will never be certain that the deceased is a *qq*

Table 11.4 Formulae for the Likelihood Ratio for Situations Where the Genotype of One Child of the Missing Person is Available as a Reference Sample But the Partner is Unavailable

Body	Child	*LR* Assessed By the Product Rule	*LR* Assessed Using the Subpopulation Correction (body and child members of the same subpopulation)
aa	*aa*	$\frac{1}{p_a}$	$\frac{1+2\theta}{3\theta+(1-\theta)p_a}$
	ab	$\frac{1}{2p_a}$	$\frac{1+2\theta}{2[2\theta+(1-\theta)p_a]}$
ab	*aa*	$\frac{1}{2p_a}$	$\frac{1+2\theta}{2[2\theta+(1-\theta)p_a]}$
	ab	$\frac{p_a+p_b}{4p_ap_b}$	$\frac{(1+2\theta)\{2\theta+(1-\theta)[p_a+p_b]\}}{4(\theta+(1-\theta)p_a)(\theta+(1-\theta)p_b)}$
	ac	$\frac{1}{4p_a}$	$\frac{1+2\theta}{4[\theta+(1-\theta)p_a]}$

homozygote at this locus, although this may be overwhelmingly the most supported genotype.

11.4.2.4 Siblings

It is possible for two full siblings to have no alleles in common at a locus. Although this is unlikely to occur at all loci of a multiplex, a single surviving sibling offers little predictive information regarding the profile of the deceased. However, if two or more siblings are available, the information increases concomitantly. Only when all four distinct alleles are represented among the surviving siblings will it be possible to deduce the four parental alleles at a locus. It will still be impossible to determine which genotype is paternal and which is maternal. When multiple loci have all four alleles present, there will be ambiguity about the multilocus paternal and maternal genotypes. The ambiguity in the deceased is influenced by the ambiguity in the parental profiles and by Mendelian factors.

The magnitude of the *LR* obtained from the analysis of siblings will depend upon the number of siblings available for testing and on the rarity of shared alleles. In addition, it is possible that a match may exist between the deceased and a putative sibling, but that the calculated *LR* value across multiple loci is less than one. In other words, there is support for the alternative proposition (that this individual is not a sibling of the body) despite the apparent match. This tends to happen when the profile from

the body has several loci where it shares no alleles with the putative sibling. At such a locus, the value of the likelihood ratio is $\frac{1}{4}$. As a very rough rule of thumb, where the body shares one allele at a locus the *LR* is generally close to 1, and when both alleles are shared it will be greater than 1, although its magnitude depends on the rarity of the alleles concerned. (see Table 11.5)

11.4.2.5 Distribution of the LR

The formulae in Table 11.5 can also be used to assess the strength of the evidence as to whether two individuals with known genotypes are full siblings or not. For a particular multiplex we can assess the performance of the likelihood ratio by considering a large number of prospective cases. The values of the *LR* can be evaluated for pairs of individuals known to be full sibs and pairs known to be unrelated. As a demonstration, we simulate the distribution of *LR* over the 15 Identifiler loci for pairs of individuals known to be full sibs and known to be unrelated using allele frequencies from the New Zealand Caucasian population. We also simulate for pairs known to be half sibs, and pairs known to be cousins. The results for half sibs are the same as would be observed for uncle–nephew pairs or grandparent–grandchild pairs.

Table 11.5 Formulae for the Likelihood Ratio for Situations Where the Genotype of One Sibling of the Missing Person is Available as a Reference Sample

Sibling	Body	*LR*	
		Product Rule	Using Subpopulation Correction (sibling and body in the same subpopulation)
aa	*aa*	$\frac{(1+p_a)^2}{4p_a^2}$	$\frac{(1+\theta)(1+2\theta)+2(1+2\theta)(2\theta+(1-\theta)p_a)+(2\theta+(1-\theta)p_a)(3\theta+(1-\theta)p_a)}{4(2\theta+(1-\theta)p_a)(3\theta+(1-\theta)p_a)}$
	bb	$\frac{1}{4}$	$\frac{1}{4}$
	ab	$\frac{1+p_a}{4p_a}$	$\frac{1+4\theta+(1-\theta)p_a}{4(2\theta+(1-\theta)p_a)}$
	bc	$\frac{1}{4}$	$\frac{1}{4}$
ab	*aa*	$\frac{1+p_a}{4p_a}$	$\frac{1+4\theta+(1-\theta)p_a}{4(2\theta+(1-\theta)p_a)}$
	ab	$\frac{1+p_a+p_b+2p_ap_b}{8p_ap_b}$	$\frac{(1+\theta)(1+2\theta)+(1+2\theta)(2\theta+(1-\theta)(p_a+p_b))+2(\theta+(1-\theta)p_a)(\theta+(1-\theta)p_b)}{8(\theta+(1-\theta)p_a)(\theta+(1-\theta)p_b)}$
	ac	$\frac{1+2p_a}{8p_a}$	$\frac{1+4\theta+2(1-\theta)p_a}{8(\theta+(1-\theta)p_a)}$
	cc	$\frac{1}{4}$	$\frac{1}{4}$
	cd	$\frac{1}{4}$	$\frac{1}{4}$

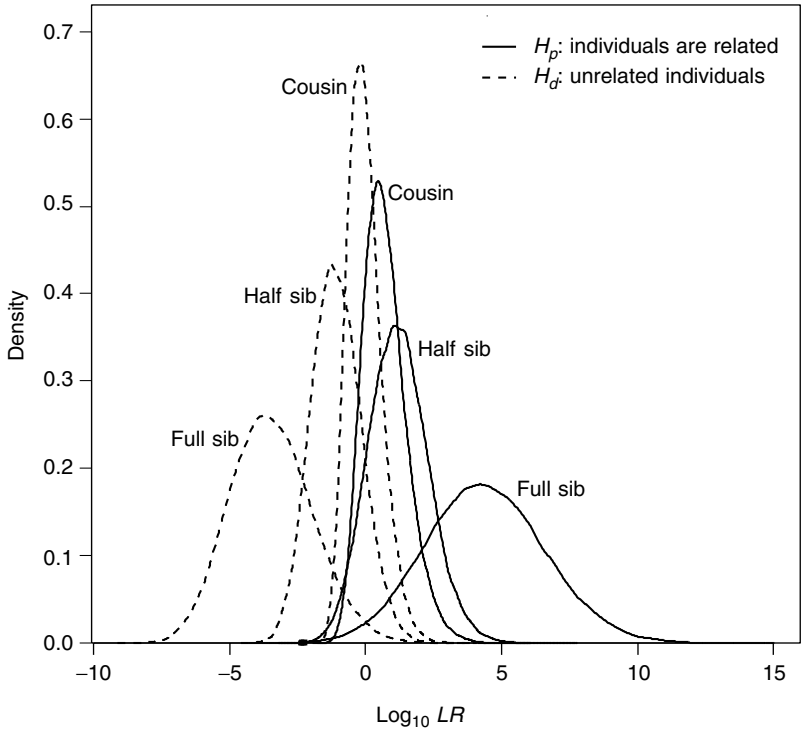


Figure 11.1 Distributions of the logarithm to the base 10 of the likelihood ratio when a pair of individuals are known to be first-degree relatives, and when they are known to be unrelated.

For each relationship, Figure 11.1 shows the distributions of the likelihood ratio under the following pair of hypotheses:

H_p: The pair of individuals have the first-degree relationship.
H_d: The pair of individuals are unrelated.

We see that the ability to discriminate a pair of full sibs from a pair of unrelated individuals is high, but weakens rapidly for more distant relatives. Recall that a pair of half sibs share an IBD allele with probability ½, but for cousins this probability is only ¼. For the 15 Identifiler loci, the modes of the distribution of the likelihood ratio are given in the table below.

	Cousins	Half Sibs	Full Sibs
H_p-related individuals	2.95	11.6	1.6×10^4
H_d-unrelated individuals	0.67	0.053	1.9×10^{-4}

Adding more loci will move the distributions of the *LR* under H_p to the right, and under H_d to the left. The separation will be faster for full sibs than for the more distant relationships. The operating characteristics of the likelihood ratio for testing other pairs of postulated relationships, for example, to assess the weight of evidence as to whether a pair of individuals are full sibs as opposed to half sibs, can be evaluated in a similar fashion.

11.4.2.6 Other Combinations of First-Degree Relatives

On many occasions, the pedigree contains combinations of relatives (e.g., a child and sibling). In these instances, the formulae for the likelihood ratio must be derived on a case-by-case and locus-by-locus basis, depending on the specific pattern of alleles.[763] We give examples of this derivation below. Software programs such as Charles Brenner's symbolic kinship program[86,88] have excellent algorithms and offer considerable savings of labor in this regard. Dawid et al.[223] describe the adaptation of general-purpose computer software for probabilistic expert systems such as Hugin,[422] GeNIe,[333] or XBAIES[875] to the pedigree problem.

Example 11.1 (*Missing body given putative parent and child*). Suppose that we have a simple three-generation pedigree as depicted on page 414. The person labelled *D* is missing and a body, *B*, has been found that may belong to this person.
For the hypotheses

H_p: The body is the missing person, and
H_d: The body is unrelated to this pedigree,

we require

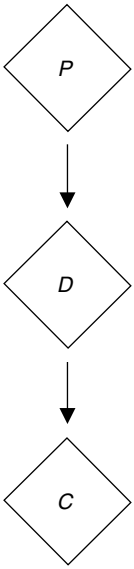
$$LR = \frac{\Pr(G_C, G_B, G_P|H_p, I)}{\Pr(G_C, G_B, G_P|H_d, I)}$$

Using the third law, we decompose this to arrange for children to be conditioned on parents (if any are present in the pedigree).

$$LR = \frac{\Pr(G_C|G_B, G_P, H_p, I)\Pr(G_B|G_P, H_p, I)\Pr(G_P|H_d, I)}{\Pr(G_C|G_B, G_P, H_d, I)\Pr(G_B|G_P, H_d, I)\Pr(G_P|H_d, I)}$$

Next we can simplify this if we make a series of assumptions. The assumptions are of two types. The first type is an assumption that the joint probabilities of observing the genotypes of the founders of the pedigree under the two hypotheses are the same.

The probability of observing the genotypes of the founder(s) of the pedigree does not depend on the specific hypothesis being considered,



$\Pr(G_P|H_p, I) = \Pr(G_P|H_d, I)$. The genotype of P does not depend on whether B is in the pedigree or not (recall that we do not know G_B at this point).

The second type of assumption follows from the application of Mendel's laws to the pedigree.

1. $\Pr(G_C|G_B, G_P, H_p, I) = \Pr(G_C|G_B, H_P, I)$ since the conditioning on P is redundant under the assumptions of H_p.
2. $\Pr(G_B|G_P, H_d, I) = \Pr(G_B|H_d, I)$ since B and P are unrelated under H_d.
3. $\Pr(G_C|G_B, G_P, H_d, I) = \Pr(G_C|G_P, H_d, I)$ since C and B are unrelated under H_d.

Note that if we use the subpopulation correction, we leave the general conditioning in place in steps 2 and 3 but treat the genotypes as coming from unrelated individuals. Hence

$$LR = \frac{\Pr(G_C|G_B, H_p, I)\Pr(G_B|G_P, H_p, I)}{\Pr(G_C|G_P, H_d, I)\Pr(G_B|H_d, I)}$$

Let $G_P = ab$, $G_B = bc$, and $G_C = cd$.

Now $\Pr(G_C|G_B, H_p, I) = p_d/2$, $\Pr(G_B|G_P, H_p, I) = p_c/2$, and $\Pr(G_B|H_d, I) = 2p_b p_c$.

To assess the term $\Pr(G_C|G_P, H_d, I)$ in general,[f] it is necessary to consider the genotype of the missing person, D. This is unknown so we denote by

[f] There are shortcuts at this point that work well for one child but are more difficult for multiple children. By using IBD states, we see that there is a $\frac{1}{2}$ chance that neither of the alleles in P is IBD, with the alleles in C. If neither are IBD, then a cd child will result with probability $2p_c p_d$. This yields directly $\frac{1}{2} \times 2p_c p_d = p_c p_d$.

G_{D_i}, $i = 1, \ldots, n$, the range of possible genotypes for this person. Suppose that the locus that is being considered has m alleles; then there are m possible homozygous genotypes and $\frac{1}{2}\, m(m-1)$ possible heterozygous genotypes to consider. However, unless the genotype G_{D_i} has (at least) one allele in common with that of P, then $\Pr(G_{D_i}|G_P, H_d, I) = 0$. Similarly, $\Pr(G_C|G_{D_i}, H_d, I) = 0$ unless the genotype G_{D_i} has (at least) one allele in common with that of C. Hence once the genotypes of P and C are known, the length, n, of the list of genotypes, $\{G_{D_i}\}$, that needs to be considered is greatly reduced. Thus,

$$\Pr(G_C|G_P, H_d, I) = \sum_{i=1}^{n} \Pr(G_C|G_{D_i}, G_P, H_d, I) \Pr(G_{D_i}|G_P, H_d, I)$$

Note that

$$\Pr(G_C|G_{D_i}, G_P, H_d, I) = \Pr(G_C|G_{D_i}, H_d, I)$$

since the genotype *of* G_C is independent of the genotype G_P conditional on H_d if the genotype of D, the child of P, and the parent of C is specified.

This term is amenable to evaluation using a tabular approach, with one row of the table below for each possible genotype G_{D_i}. Any possible genotype must contain one allele in common with G_P and one in common with G_C.

Recall $G_P = ab$, $G_B = bc$, and $G_C = cd$.

G_{D_i}	$\Pr(G_{D_i}\|G_P, H_P, I)$	$\Pr(G_C\|G_{D_i}, H_P, I)$	$\Pr(G_C\|G_{D_i}, G_P, H_P, I) \times \Pr(G_{D_i}\|G_P, H_P, I)$
ac	$\frac{p_c}{2}$	$\frac{p_d}{2}$	$\frac{p_c p_d}{4}$
ad	$\frac{p_d}{2}$	$\frac{p_c}{2}$	$\frac{p_c p_d}{4}$
bc	$\frac{p_c}{2}$	$\frac{p_d}{2}$	$\frac{p_c p_d}{4}$
bd	$\frac{p_d}{2}$	$\frac{p_c}{2}$	$\frac{p_c p_d}{4}$
$\sum_{i=1}^{n} \Pr(G_C\|G_{D_i}, H_d, I) \Pr(G_{D_i}\|G_P, H_d, I) =$			$p_c p_d$

Assembling the above terms, we find that $LR = 1/8 p_b p_c$.

We can write the likelihood ratio as the product of two terms:

$$LR = \frac{1}{\Pr(G_B|H_d, I)} \times \frac{\Pr(G_C|G_B, H_p, I)\Pr(G_B|G_P, H_p, I)}{\Pr(G_C|G_P, H_p, I)}$$

The first

$$\frac{1}{\Pr(G_B|H_d, I)} = \frac{1}{2p_b p_c}$$

is the likelihood ratio that would be obtained if we knew the genotype of person D. The second can be expressed as a fraction, since

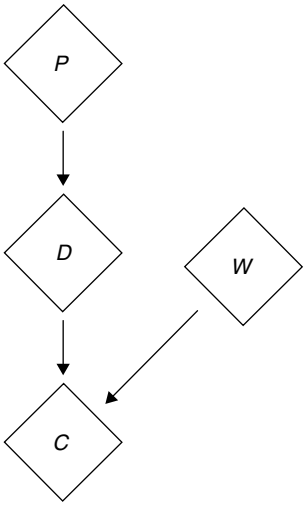
$$\frac{\Pr(G_C|G_B, H_p, I)\Pr(G_B|G_P, H_p, I)}{\Pr(G_C|G_P, H_d, I)} = \frac{\Pr(G_C|G_B, H_p, I)\Pr(G_B|G_P, H_p, I)}{\sum_{i=1}^{n}\Pr(G_C|G_{D_i}, H_d, I)\Pr(G_{D_i}|G_P, H_d, I)}$$

After we have enumerated the genotypes $D_1, \ldots, D_n$, we can drop the conditioning on the hypotheses H_p and H_d. We thus have

$$\frac{\Pr(G_C|G_B, H_p, I)\Pr(G_B|G_P, H_p, I)}{\Pr(G_C|G_P, H_d, I)} = \frac{\Pr(G_C|G_B, I)\Pr(G_B|G_P, I)}{\sum_{i=1}^{n}\Pr(G_C|G_{D_i}, I)\Pr(G_{D_i}|G_P, I)}$$

The genotype of the (nonexcluded) body B must be one of the genotypes in the list, D_k say, $G_B = G_{D_i}$. Hence the fraction is always less than 1 in value if there is more than one possible genotype for the missing body that is consistent with the pedigree. In this example, there are four possible genotypes and, coincidentally, LR is ¼ that for a simple identification.

Example 11.2 (*Missing body given putative parent, partner, and child*). We have the simple three-generation pedigree as depicted below. The person labelled D is missing and a body, B, has been found that may be this person. For the hypotheses



H_p: The body is the missing person, and
H_d: The body is unrelated to P, W, and C,

we require

$$LR = \frac{\Pr(G_C, G_B, G_W, G_P|H_p, I)}{\Pr(G_C, G_B, G_W, G_P|H_d, I)}$$

Using the third law, we decompose this to arrange for genotypes of children to be conditioned on the genotypes of their parents (if any are present in the pedigree); hence

$$LR = \frac{\Pr(G_C|G_B, G_W, G_P, H_p, I)\Pr(G_B|G_W, G_P, H_p, I)\Pr(G_W, G_P|H_p, I)}{\Pr(G_C|G_B, G_W, G_P, H_d, I)\Pr(G_B|G_W, G_P, H_d, I)\Pr(G_W, G_P|H_d, I)}$$

As in the previous example, we can simplify this by making the same four assumptions plus one further assumption:

1. The joint probability of observing the genotypes of the founders of the pedigree P and W is the same for both hypotheses.

The remaining assumptions are the implications of Mendel's laws for the probabilities of observing the genotypes G_P, G_B, G_W, and G_C under the two hypotheses H_p and H_d.

2. Under H_p the genotype of C is determined by those of B and W; hence
$$\Pr(G_C|G_B, G_W, G_P, H_p, I) = \Pr(G_C|G_B, G_W, H_p, I)$$
3. Under H_d the genotype of B is independent of those of P and W.
4. Under H_d the genotypes of B and C are unrelated and hence the probability of observing the genotype of B is determined by the genotypes of P and W.
5. In addition, we further assume that W and B are unrelated and so

$$\Pr(G_B|G_W, G_P, H_p, I) = \Pr(G_B|G_P, H_p, I)$$

Note that if we use the subpopulation correction, we leave the general conditioning in place in steps 2–5 but treat the genotypes as coming from unrelated individuals. Hence

$$LR = \frac{\Pr(G_C|G_B, G_W, H_p, I)\Pr(G_B,|G_P, H_p, I)}{\Pr(G_C|G_W, G_P, H_d, I)\Pr(G_B|H_d, I)}$$

Let $G_P = ab$, $G_B = bc$, $G_W = de$, and $G_C = cd$. Now $\Pr(G_C \mid G_B, G_W, H_p, I) = \frac{1}{4}$, $\Pr(G_B|G_P, H_p, I) = p_c/2$, and $\Pr(G_B|H_d, I) = 2p_b p_c$. The term $\Pr(G_C|G_W, G_P, H_d, I)$ is

assessed as before by considering the range of possible genotypes, G_{D_i}, for the missing person, D.

$$\Pr(G_C|G_W, G_P, H_d, I) = \sum_{i=1}^{n} \Pr(G_C|G_{D_i}, G_W, G_P, H_d, I)\Pr(G_{D_i}|G_W, G_P, H_d, I)$$

$$\Pr(G_{D_i}|G_W, G_P, H_d, I) = \Pr(G_{D_i}|G_P, H_d, I)$$

since D and W are unrelated and

$$\Pr(G_C|G_{D_i}, G_W, G_P, H_d, I) = \Pr(G_C|G_{D_i}, G_W, H_d, I)$$

since the genotype of the child is independent of that of P under H_p if the genotypes of the child's parents are specified. Hence we can write

$$\Pr(G_C|G_W, G_P, H_d, I) = \sum_{i=1}^{n} \Pr(G_C|G_{D_i}, G_W, H_d, I)\Pr(G_{D_i}|G_P, H_d, I)$$

The likelihood ratio can be written as

$$LR = \frac{1}{\Pr(G_B|H_d, I)} \times \frac{\Pr(G_C|G_B, G_W, H_p, I)\Pr(G_B|G_P, H_p, I)}{\Pr(G_C|G_P, G_W, H_p, I)}$$

$$= \frac{1}{\Pr(G_B|H_d, I)} \times \frac{\Pr(G_C|G_{D_k}, G_W, I)\Pr(G_{D_k}|G_P, G_W, I)}{\sum_{i=1}^{n} \Pr(G_C|G_{D_i}, G_W, I)\Pr(G_{D_i}|G_P, G_W, I)}$$

As in the previous example, we can enumerate the list of possible genotypes, D_i, and evaluate the sum in the denominator using the table below. The table includes rows for each of the same set of genotypes as in Example 11.1, but for two genotypes the term $\Pr(G_C|G_{D_i}, G_W, H_d, I) = 0$.

Recall $G_P = ab$, $G_B = bc$, $G_W = de$, and $G_C = cd$.

G_{D_i}	$\Pr(G_{D_i}\|G_P, H_d, I)$	$\Pr(G_C\|G_{D_i}, G_W, H_d, I)$	$\Pr(G_C\|G_{D_i}, G_W, H_d, I) \times \Pr(G_{D_i}\|G_P, H_d, I)$
ac	$\frac{p_c}{2}$	$\frac{1}{4}$	$\frac{p_c}{8}$
ad	$\frac{p_d}{2}$	0	0
bc	$\frac{p_c}{2}$	$\frac{1}{4}$	$\frac{p_c}{8}$
bd	$\frac{p_d}{2}$	0	0
$\sum_{i=1}^{n} \Pr(G_C\|G_{D_i}, G_W, H_d, I)\Pr(G_{D_i}\|G_P, H_d, I) =$			$\frac{p_c}{4}$

Assembling the above terms we find that $LR = 1/4p_b p_c$. Reducing the list of possible genotypes for the missing person, D, from four to two has coincidently increased the likelihood ratio by a factor of 2.

Example 11.3 (*Missing body given putative parent, partner, and two children*). We consider a simple three-generation pedigree as depicted below. The person labelled D is missing and a body, B, has been found that may be this person.

For the pair of hypotheses

H_p: The body is the missing person, and
H_d: The body is unrelated to this pedigree,

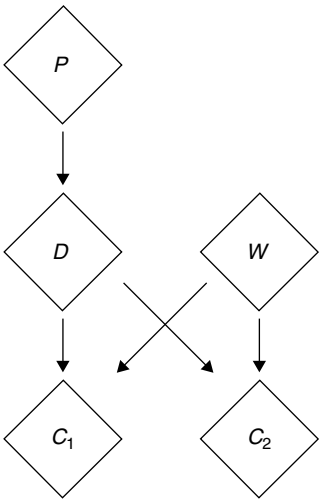
we require

$$LR = \frac{\Pr(G_{C1}, G_{C2}, G_B, G_W, G_P|H_p, I)}{\Pr(G_{C1}, G_{C2}, G_B, G_W, G_P|H_d, I)}$$

$$= \frac{\Pr(G_{C1}, G_{C2}|G_B, G_W, G_P, H_p, I)\Pr(G_B|G_W, G_P, H_p, I)\Pr(G_W, G_P|H_p, I)}{\Pr(G_{C1},G_{C2}|G_B, G_W, G_P, H_d, I)\Pr(G_B|G_W, G_P, H_d, I)\Pr(G_W, G_P|H_d, I)}$$

If we make the same five assumptions as in the previous example, we can write the likelihood ratio as

$$LR = \frac{1}{\Pr(G_B|H_d, I)} \times \frac{\Pr(G_{C1}, G_{C2}|G_B, G_W, H_p, I)\Pr(G_B|G_P, H_p, I)}{\Pr(G_{C1}, G_{C2},|G_W, G_P, H_d, I)}$$



Let $G_P = ab$, $G_B = bc$, $G_W = de$, $G_{C_1} = bd$, and $G_{C_2} = ce$.

Now $\Pr(G_{C_1}, G_{C_2}|G_B, G_W, H_p, I) = \frac{1}{16}$, $\Pr(G_B|G_P, H_P, I) = p_c/2$, and $\Pr(G_B|H_d, I) = 2p_b p_c$. The term $\Pr(G_{C_1}, G_{C_2}|G_W, G_P, H_d, I)$ is assessed as before by considering the range of possible genotypes, G_{D_i}, for the missing person, D, that are consistent with the genotypes of W, P, C_1, and C_2 and then evaluating the sum

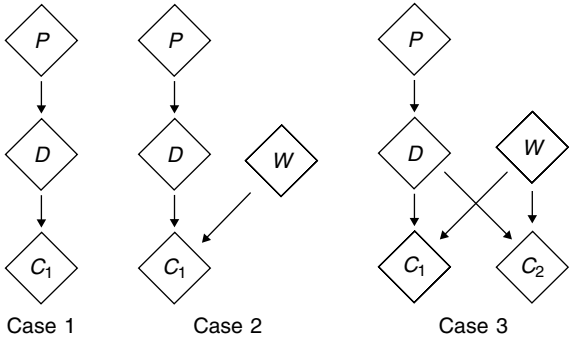
$$\Pr(G_{C_1}, G_{C_2}|G_W, G_P, H_d, I) = \sum_{i=1}^{n} \Pr(G_{C_1}, G_{C_2}|G_{D_i}, G_W, I) \Pr(G_{D_i}|G_P, I)$$

This pedigree differs from that in Example 11.2 because, under H_p, the genotypes of the two children, C_1 and C_2, together with the genotype of the partner suffice to determine the genotype of the missing person. Our table now has only a single row. Three of the previously considered genotypes lead to a value of 0 for $\Pr(G_{C_1}, G_{C_2},|G_{D_i}, G_W, I)$.

G_{D_i}	$\Pr(G_{D_i}\|G_P, I)$	$\Pr(G_{C_1}, G_{C_2}\|G_{D_i}, G_W, I)$	$\Pr(G_{C_1}, G_{C_2}\|G_{D_i}, G_W, I) \times \Pr(G_{D_i}\|G_P, I)$
bc	$\frac{p_c}{2}$	$\frac{1}{16}$	$\frac{p_c}{32}$
$\sum_{i=1}^{n} \Pr(G_{C_1}, G_{C_2}\|G_{D_i}, G_W, I)\Pr(G_{D_i}\|G_P, I) =$			$\frac{p_c}{32}$

The numerator and denominator of the second term in the expression for LR are the same and hence we find that $LR = 1/2p_b p_c$. Its value is now the same as that for a simple identification.

Example 11.4 (*Missing body given putative parent, partner, and two children*). We consider the same pedigree as in the previous three examples when the genotype of the recovered body, B, is homozygous.



Person	Case 1	Case 2	Case 3
P	ab	ab	ab
W	—	cd	cd
B	aa	aa	aa
C_1	ac	ac	ac
C_2	—	—	ad

Taking the same pair of hypotheses and making the same assumptions as in the three previous examples, we use the same method to evaluate the likelihood ratio. Under H_p we can enumerate the lists of possible genotypes, D_i, that lead to nonzero probabilities in each of the three cases.

D_i	Case 1	Case 2	Case 3
D_1	aa	aa	aa
D_2	ab	ab	ab
D_3	ac	ac	ac
D_4	bc	—	—
D_5	ax	ax	ax

The allele "x" is taken to be any allele other than a, b, or c. In each case, $\Pr(G_B|H_d, I) = p_a^2$. Using the tables below, we consider each case in turn.

Case 1: $P = ab, B = aa, C_1 = ac$

G_{D_i}	$\Pr(G_{D_i}\|G_P, I)$	$\Pr(G_{C_1}\|G_{D_i}, I)$	$\Pr(G_{C_1}\|G_{D_i}, I) \times \Pr(G_{D_i}\|G_P, I)$
$G_B = aa$	$\frac{p_a}{2}$	p_c	$\frac{p_a p_c}{2}$
ab	$\frac{p_a + p_b}{2}$	$\frac{p_c}{2}$	$\frac{(p_a + p_b)p_c}{4}$
ac	$\frac{p_c}{2}$	$\frac{p_a + p_c}{2}$	$\frac{(p_a + p_c)p_c}{4}$
bc	$\frac{p_c}{2}$	$\frac{p_a}{2}$	$\frac{p_a p_c}{4}$
ax	$\frac{1 - p_a - p_b - p_c}{2}$	$\frac{p_c}{2}$	$\frac{(1 - p_a - p_b - p_c)p_c}{4}$
$\sum_{i=1}^{n} \Pr(G_{C_i}\|G_{D_i}, I) \Pr(G_{D_i}\|G_P, I) =$			$\frac{(1 + 4p_a)p_c}{4}$

We find that

$$LR = \frac{2}{p_a(1+4p_a)} = \frac{1}{p_a^2} \times \frac{2p_a}{1+4p_a}$$

Its value is very much less than that for a simple identification if the *a* allele is rare.

Case 2: Inclusion of other parent of the child of the missing person. $P = ab$, $B = aa$, $W = cd$, $C_1 = ac$

G_{D_i}	$\Pr(G_{D_i}\|G_P, I)$	$\Pr(G_{C_1}\|G_{D_i}, I)$	$\Pr(G_{C_1}\|G_{D_i}, I) \times \Pr(G_{D_i}\|G_P, I)$
$G_B = aa$	$\frac{p_a}{2}$	$\frac{1}{2}$	$\frac{p_a}{4}$
ab	$\frac{p_a + p_b}{2}$	$\frac{1}{4}$	$\frac{p_a+p_b}{8}$
ac	$\frac{p_c}{2}$	$\frac{1}{4}$	$\frac{p_c}{8}$
ax	$\frac{1-p_a-p_b-p_c}{2}$	$\frac{1}{4}$	$\frac{1-p_a-p_b-p_c}{8}$
$\sum_{i=1}^{n} \Pr(G_{C_1}\|G_{D_i}, I) \Pr(G_{D_i}\|G_p, I) =$			$\frac{1+2p_a}{8}$

We find that

$$LR = \frac{2}{p_a(1+2p_a)} = \frac{1}{p_a^2} \times \frac{2p_a}{1+2p_a}$$

Its value is a little greater than that for Case 1, but is still less than that for a simple identification.

Case 3: Inclusion of a second child. $P = ab$, $B = aa$, $W = cd$, $C_1 = ac$, $C_2 = ad$

G_{D_i}	$\Pr(G_{D_i}\|G_P, I)$	$\Pr(G_{C_1}, G_{C_2}\|G_{D_i}, I)$	$\Pr(G_{C_1}, G_{C_2}\|G_{D_i}, I) \times \Pr(G_{D_i}\|G_P, I)$
$G_B = aa$	$\frac{p_a}{2}$	$\frac{1}{4}$	$\frac{p_a}{8}$
ab	$\frac{p_a + p_b}{2}$	$\frac{1}{16}$	$\frac{p_a + p_b}{32}$
ac	$\frac{p_c}{2}$	$\frac{1}{16}$	$\frac{p_c}{32}$
ax	$\frac{1-p_a-p_b-p_c}{2}$	$\frac{1}{16}$	$\frac{1-p_a-p_b-p_c}{32}$
$\sum_{i=1}^{n} \Pr(G_{C_1}, G_{C_2}\|G_{D_i}, I) \Pr(G_{D_i}\|G_p, I) =$			$\frac{1+4p_a}{32}$

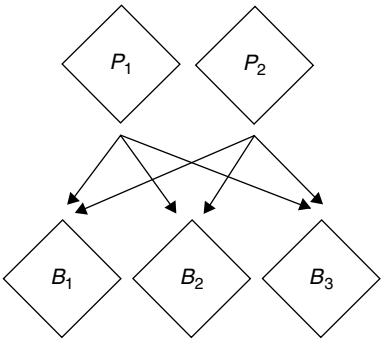
We find that

$$LR = \frac{4}{p_a(1+4p_a)} = \frac{1}{p_a^2} \times \frac{4p_a}{1+4p_a}$$

In the following table we compare the numerical value for the likelihood ratio in each of the four cases for a range of allele frequencies for the *a* allele.

Allele Frequency (p_a)	Simple Identification	Case 1	Case 2	Case 3
0.005	40,000	390	400	780
0.010	10,000	190	200	380
0.025	1600	73	76	150
0.050	400	33	36	67
0.100	100	14	17	29
0.250	16	4.0	5.3	8.0
0.350	8.2	2.4	3.4	4.8

Example 11.5



In this instance we consider three siblings whose parents have not been typed. The parents are drawn in the pedigree, but recall that their genotypes are not available. Sibling B_1 is missing and a body, *B*, has been found that may be this person.

For the hypotheses

H_p: The body is the missing person, and
H_d: The body is unrelated to this pedigree,

we require

$$LR = \frac{\Pr(G_{B_1}, G_{B_2}, G_{B_3} | H_p, I)}{\Pr(G_B, G_{B_2}, G_{B_3} | H_d, I)}$$

Assuming that B_1 and B_2, B_3 are independent if unrelated, this yields

$$LR = \frac{\Pr(G_{B_1}, G_{B_2}, G_{B_3}|H_p, I)}{\Pr(G_B|H_d, I)\Pr(G_{B_2}, G_{B_3}|H_d, I)}$$

Evaluation of the denominator is relatively straightforward. $\Pr(G_B|H_d, I)$ can be assessed by the product rule; $\Pr(G_{B_2}, G_{B_3}|H_d, I)$ is the joint probability for two siblings and is tabulated in Chapter 4. This leaves $\Pr(G_{B_1}, G_{B_2}, G_{B_3}|H_p, I)$. This is the joint probability for the genotypes of three siblings. Two approaches are available.

1. Complete enumeration of the possible genotypes for the parents. This approach is consistent with the approach used in Examples 11.1–11.4. If there are $1 \ldots n$ possibilities for the parents P_1 and P_2, then

$$\Pr(G_{B_1}, G_{B_2}, G_{B_3}, |H_p, I) = \sum_{i=1}^{n} \Big(\Pr(G_{B_1}, G_{B_2}, G_{B_3}, |G_{p_1}, G_{p_2}, H_p, I \times \Pr(G_{p_1}, G_{p_2}|H_p, I) \Big)$$

2. The use of six-allele descent measures, which are the logical extension of Weir's four-allele measures, as presented in Chapter 4.

Six-allele descent measures were discussed by Thompson[763] (p. 43) in the context of two siblings and an aunt or a niece or half sibling. We illustrate their use in this case. Let $G_{B_1} = ab$, $G_{B_2} = ab$, $G_{B_3} = cd$ (this is the easiest three-sibling calculation since all four distinct alleles are present).

Enumeration of the possibilities for the genotypes of parents leads to a calculation that can be set out in a table as in the examples above.

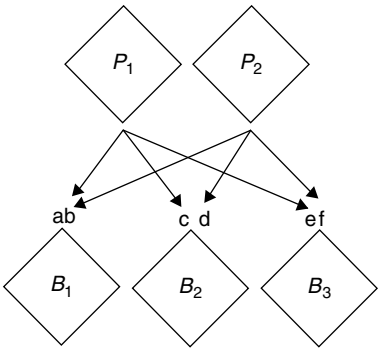
P_1	P_2	$\Pr(G_{P_1}, G_{P_2}\|H_P, I)$	$\Pr(G_{B_1}, G_{B_2}, G_{B_3}\|G_{P_1}, G_{P_2}, H_p, I)$
ac	bd	$4p_a p_b p_c p_d$	$\frac{1}{64}$
ad	bc	$4p_a p_b p_c p_d$	$\frac{1}{64}$
bd	ac	$4p_a p_b p_c p_d$	$\frac{1}{64}$
bc	ad	$4p_a p_b p_c p_d$	$\frac{1}{64}$
$\sum_{i=1}^{n} \Pr(G_{B_1}, G_{B_2}, G_{B_3}\|G_{P_1}, G_{P_2}, H_p, I) \Pr(G_{P_1}, G_{P_2}\|H_p, I) =$			$\frac{p_a p_b p_c p_d}{4}$

Now $\Pr(G_{B_1}|H_d, I) = 2p_a p_b$, and from Table 4.9 (Chapter 4) we see that

$$\Pr(G_{B_2}, G_{B_3}|H_d, I) = p_a p_b p_c p_d$$

Hence $LR = 1/8p_a p_b$.

In order to apply the six-allele[g] descent measures, we label the alleles in the three children a to f. We consider that the alleles a and c could be IBD since they have come from the same parent. This will occur $\frac{1}{2}$ of the time. Hence the alleles labeled a, c, and e will be IBD (a ≡ b ≡ c) $\frac{1}{4}$ of the time. Similarly, the alleles b, d, and f will be IBD $\frac{1}{4}$ of the time. Hence the event a ≡ b ≡ c and b ≡ d ≡ f will occur $\frac{1}{16}$ of the time.



IBD State		Pr(*IBD state*)	Pr $(G_{B_1}, G_{B_2}, G_{B_3}$|*IBD state*)
a≡b≡c	b≡d≡f	$\frac{1}{16}$	0
	b≡d	$\frac{1}{16}$	0
	b≡f	$\frac{1}{16}$	0
	d≡f	$\frac{1}{16}$	0
a≡c	b≡d≡f	$\frac{1}{16}$	0
	b≡d	$\frac{1}{16}$	$4p_a p_b p_c p_d$
	b≡f	$\frac{1}{16}$	0
	d≡f	$\frac{1}{16}$	0
a≡e	b≡d≡f	$\frac{1}{16}$	0

(*Continued*)

[g] The extension to *N* alleles is discussed by Thompson.[761,763]

Table (***Continued***)

	b≡d	$\frac{1}{16}$	0
	b≡f	$\frac{1}{16}$	0
	d≡f	$\frac{1}{16}$	0
c≡e	b≡d≡f	$\frac{1}{16}$	0
	b≡d	$\frac{1}{16}$	0
	b≡f	$\frac{1}{16}$	0
	d≡f	$\frac{1}{16}$	0
Taking the product across the rows and the sum down yields→			$\frac{p_a p_b p_c p_d}{4}$

This is the same result as achieved by enumerating the parental possibilities.

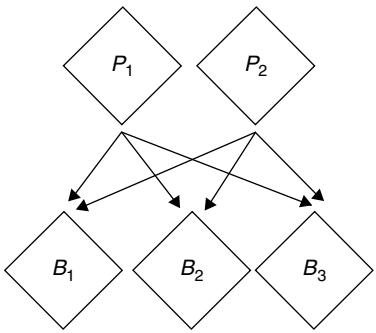
Example 11.6

In this instance, we consider the same pedigree of three siblings whose parents have not been typed as given in Example 11.5. For the hypotheses

H_p: The body is the missing person, and
H_d: The body is unrelated to this pedigree,

we evaluate

$$LR = \frac{\Pr(G_{B_1}, G_{B_2}, G_{B_3}|H_p, I)}{\Pr(G_B|H_d, I)\Pr(G_{B_2}, G_{B_3}|H_d, I)}$$



as usual; however, in this example we consider a more difficult arrangement of genotypes.

Let $G_B = ab$, $G_{B_2} = ab$, $G_{B_3} = ac$. Enumeration of the possibilities for the parents can be attempted by the tabular approach again. The symbol $\star$ is used to mean "not a."

P_1	P_2	$\Pr(G_{P_1}, G_{P_2}\|H_P, I)$	$\Pr(G_{B_1}, G_{B_2}, G_{B_3}\|G_{P_1}, G_{P_2}, H_p, I)$
aa	bc	$2p_a^2 p_b p_c$	$\frac{1}{8}$
ac	ab	$4p_a^2 p_b p_c$	$\frac{1}{64}$
$a^\star$	bc	$4p_a p_b\, p_c(1-p_a)$	$\frac{1}{64}$
bc	aa	$2p_a^2 p_b p_c$	$\frac{1}{8}$
ab	ac	$4p_a^2 p_b\, p_c$	$\frac{1}{64}$
bc	$a^\star$	$4p_a p_b\, p_c(1-p_a)$	$\frac{1}{64}$
$\sum_{i=1}^{n} \Pr(G_{B_1}, G_{B_2}, G_{B_3}\|G_{P_1}, G_{P_2}, H_p, I) \times \Pr(G_{P_1}, G_{P_2}\|H_P,\ I) =$			$\frac{p_a p_b p_c}{8}(1+4p_a)$

$$\Pr(G_B|H_d, I) = 2p_a 2p_b,\ \Pr(G_{B_2}, G_{B_3}|H_d, I) = \frac{p_a p_b p_c(1+2p_a)}{2}$$

from Table 4.11 (Chapter 4), so

$$LR = \frac{1+4p_a}{8p_a p_b(1+2p_a)}$$

Use of the descent measures leads to the table below.

IBD State		$\Pr(IBD\ state)$	$\Pr(G_{B_1}, G_{B_2}, G_{B_3}\|IBD\ state)$
a≡b≡c	b≡d≡f	$\frac{1}{16}$	0
	b≡d	$\frac{1}{16}$	$p_a p_b p_c$
	b≡f	$\frac{1}{16}$	0
	d≡f	$\frac{1}{16}$	0

(*Continued*)

Table (*Continued*)

a≡c	b≡d≡f	$\frac{1}{16}$	$p_a p_b p_c$
	b≡d	$\frac{1}{16}$	$4p_a^2 p_b p_c$
	b≡f	$\frac{1}{16}$	$p_a^2 p_b p_c$
	d≡f	$\frac{1}{16}$	$p_a^2 p_b p_c$
a≡e	b≡d≡f	$\frac{1}{16}$	0
	b≡d	$\frac{1}{16}$	$p_a^2 p_b p_c$
	b≡f	$\frac{1}{16}$	0
	d≡f	$\frac{1}{16}$	0
c≡e	b≡d≡f	$\frac{1}{16}$	0
	b≡d	$\frac{1}{16}$	$p_a^2 p_b p_c$
	b≡f	$\frac{1}{16}$	0
	d≡f	$\frac{1}{16}$	0
Taking the product across the rows and the sum down yields→			$\frac{p_a p_b p_c}{8}(1+4p_a)$

The value

$$\frac{p_a p_b p_c (1+4p_a)}{8}$$

is the same result as that achieved by enumerating the possible parental genotypes.

11.5 Complicating Factors

There are some factors that are liable to complicate a pedigree analysis. These include:

- Biological nonpaternity of one (or more) of the individuals in the pedigree.

- Germline mutation.
- Linkage.

Unfortunately, human relationships are such that biological nonpaternity is a real possibility, however convinced a party may be that a man is the true father of a certain child. For this reason, any assumptions made during analysis regarding paternity should be explicitly stated. In some situations, it may be possible to carry out comparisons based initially on maternal profiles.

The issue of germline mutation has been discussed previously in Chapter 10. The possibility of germline mutation must be borne in mind, especially when a large number of parent/child transmissions have been studied. It is likely that such an effect will be restricted to a single locus. The effects on the analysis may be negligible if the mutation occurs in a known sample.

The analysis given above was for unlinked loci. In such cases, the likelihood ratio may be calculated on a locus-by-locus basis and the results multiplied across all the loci considered,

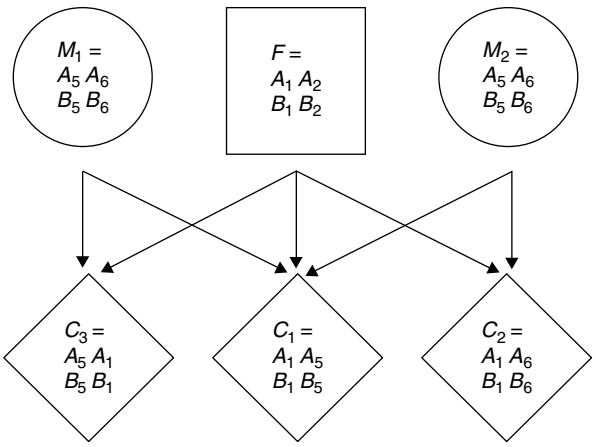
$$LR = \prod_{loci} LR_{locus}$$

For linked loci, such as the pair HUMD5S818 and HUMCSF1PO in the CODIS set, this equation applies only when there is no information about the phase of the people in dispute. Information about phase is present whenever a relevant individual in the supposed pedigree is involved in two or more meioses.[82]

Consider the following pedigree:

For the hypotheses

H_p: C_3 is the person indicated in this pedigree, and
H_d: C_3 is unrelated to this pedigree,



the likelihood ratio is

$$LR = \frac{(1 - R_M)^3 + R_M^3}{8\left\{(1 - R_M)^2 + R_M{}^2\right\} p_{A_1} p_{A_5} p_{B_1} p_{B_5}}$$

where the male and female recombination fractions are R_M and R_F, respectively. This compares with

$$LR = \frac{1}{16\, p_{A_1} p_{A_5} p_{B_1} p_{B_5}}$$

if the loci were unlinked. This is because the genotypes of the children C_1 and C_2 either singly or together give information about the phase of F. The practical implications are likely to be confined to only a fraction of the pedigrees, and unless there are multiple linked loci the implications will be small.

11.6 Mass Disasters

In a mass disaster situation, some of the difficulties of body identification by genetic means are compounded.

First, in many cases a cataclysmic physical event such as an impact or explosion will have caused severe fragmentation of the bodies. In such a situation, simply locating and recovering the post-mortem remains is a major organizational and technical challenge. A key part of the identification process will be to reunify fragmentary remains as well as identify the deceased.

Second, the remains may have been subjected to extreme thermal, chemical, or microbial insult. In turn, this will increase the difficulty of obtaining DNA in satisfactory concentrations and of sufficient quality.

Third, it may be the case that a number of individuals from the same family are thought to be victims. This creates the twofold problem of coping with the interrelatedness of some of the remains and obtaining sufficient genetic information via surviving relatives (see Clayton[180]).

Lastly, the sheer scale of the exercise may present difficulties. For instance, tracking and collating results from tens of thousands of samples, actioning rework, and comparing a large number of ante-mortem samples with a large number of post-mortem samples may be too great an endeavor for a single individual or group of individuals. Ballantyne[44] reports that Olaisen's group operated around the clock for three weeks on the 1996 Spitzbergen crash, which involved 141 victims. The situation following the terrorist attacks on the World Trade Centre on September 11, 2001 has precipitated the development of a specialist software package (MFISys) by a bioinformatics company (Genecode Forensics) to handle the vast quantities of data and information generated.

Brenner and Weir[83] and Vastag[799] identify three steps in the identification process.

Collapsing: This refers to the association of like profiles to condense the amount of data. These authors describe an approximate probabilistic approach to deal with the "collapse" of partial profiles. Olaisen et al.[593] and Ballantyne[44] report DNA typing of the 141 victims of the crash of a Tupolev 154 aircraft into the Opera mountain in Spitzbergen in 1996. The DNA results from 257 fragmented remains could be collapsed to 141 different genotypes equating to the number of victims. The 1277 tested remains from the crash of Swissair Flight 111 could be collapsed to the 228 genotypes expected from the 229 victims containing one pair of identical twins.[496] Goodwin et al.[364] describe the analysis of 187 tissue samples from 104 victims of Cebu Pacific Flight 387 in 1998. The tissue samples could be collapsed into 55 groups by mtDNA, subdivided into 80 groups when STR results were added, and further into 95 groups when post-mortem results were considered. The 14,249 typing results from the World Trade Centre tragedy were collapsed to at least 1487 distinct profiles. There are 2792 missing persons in this set.[83]

Screening: This refers to the comparison of every victim profile in the collapsed list with every missing person profile. Brenner and Weir point out that screening of direct comparisons between a victim and a sample from the missing person (from, say, a toothbrush) is straightforward, whereas screening against relatives is more difficult. The problem of false positives was highlighted since the large number of pairings will give a number of false indications of membership of a given pedigree.

Testing: This is the confirming calculation of likelihood ratios and was undertaken as described in the equations and tables above and in Brenner.[86]

11.6.1 Closed Set Matching

Consider a situation where there are N persons associated with a mass disaster. We denote the missing people as $M_1 \ldots M_N$. The bodies are denoted as $B_1 \ldots B_N$ with genotypes $G_1 \ldots G_N$. We have a list of the people who are missing. In addition, for each person a sample known to be from them (say a toothbrush, hair, a database sample, or some pedigree information) is available. We denote the known sample or pedigree information as $P_1 \ldots P_N$. This is termed "closed set" matching because the issue is to match a finite number of bodies or body parts to a finite set of missing people.[83,243,244] It uses the information more effectively than open set matching since both the consistency with a pedigree and the dissimilarity to other pedigrees are used.

The direct match comparisons between bodies, body parts, and complete profiles of missing people present few statistical problems. However, Leclair et al.[496] reported that in five of 47 instances in the Swissair Flight 111 crash investigation, the surrogate reference sample was incorrect, being that of

another family member. This does suggest the use of confirmatory pedigree analysis in at least some situations. Let us assume that L bodies can be assigned without error to missing people in this way. There are $N-L$ bodies unassigned and hence there are $(N-L)!$ possible assignments of the remaining bodies to pedigrees. To keep the subscripting simple, we arrange to label the directly assigned bodies as $B_1 \ldots B_L$ and the missing people as $M_1 \ldots M_L$. In principle, for small sets of missing people and bodies, the remaining assignments can be assessed by direct enumeration. We label each of the possible sets of assignments of the unassigned bodies to missing people as $A_{L+1} \ldots A_{(N-L)!}$ and evaluate the equation

$$\Pr(A_j | G_{L+1} \ldots G_N, P_{L+1} \ldots P_N)$$

$$= \frac{\Pr(G_{L+1} \ldots G_N | P_{L+1} \ldots P_N, A_j) \Pr(A_j)}{\sum_{i=L+1}^{(N-L)!} \Pr(G_{L+1} \ldots G_N | P_{L+1} \ldots P_N, A_i) \Pr(A_i)}$$

for each possible assignment, j, in this set. The prior probabilities $\Pr(A_j)$ can be assigned from a physical examination of the bodies, location or other information, or may be set to a flat prior for the remaining assignments (after the direct comparisons have been removed). Egeland et al.[244,246] give advice on how this may be undertaken in an elegant manner. They also suggest giving low priors to highly incestuous pedigrees, those involving what they term promiscuity, and to pedigrees that extend over multiple generations.

This process yields the posterior probability on each member of the possible set of assignments. However, we will probably be interested, not in the posterior probability for the entire set, which may be quite small, but the posterior probability for each assignment of a certain body to a certain missing person. This is obtained by summing the terms $\Pr(A_j | G_{L+1} \ldots G_N, P_{L+1} \ldots P_N)$ for those assignments A_j that contain this pairing.

The term $\Pr(G_{L+1} \ldots G_N | P_{L+1} \ldots P_N, A_j)$ may be approximated by assuming no interperson correlations within the set of missing people by

$$\prod_{i=L+1}^{N} \Pr(G_i | P_{L+1} \ldots P_N, A_j)$$

This is clearly incorrect, especially for related persons, but may be an adequate approximation. Any assignment, A_j, that produces $\Pr(G_i | P_{M+1} \ldots P_N, A_j) = 0$ for any genotype G_i can be tentatively eliminated.[88] Such an elimination is tentative as a mutation or a pedigree error (for example, if a person who

thinks they are a parent is in fact not) will produce a false exclusion. This provides a quick screen to eliminate potential assignments. However, it may be appropriate to "keep this combination alive" by assigning it a small number in case a mutation or a pedigree error may have occurred. Olaisen et al.[593] report four mutations in their analysis while Leclair et al.[496] report one. A superior approach[244] would include a full mutation model but would be computationally expensive.

Egeland et al.[246] demonstrate this approach on the nine bodies found in the grave in Ekaterinburg and thought to include most of the Russian royal family. In this set, there are 4536 possible family relationships. The number of permutations may be reduced by dividing the bodies into children (who cannot themselves have children in this example) and adults after age determination. However, the full-scale closed set approach given above may be impractical in many cases due to the size of the problem. A more typical approach is to eliminate possible assignments; for instance, body j cannot be from pedigrees x.... In addition, it may be possible to break the remaining assignments into male and female using amelogenin[246] or other methods. This may leave two smaller sets of missing persons, one of female and one of male, to be assigned to bodies also divided by sex. Brenner also describes an effective "lattice" approach to reducing the number of permutations that need be considered. The use of mitochondrial DNA and Y chromosome typing to simplify the number of comparisons has great promise. Cowell and Mostad[202] suggest a method for identifying small clusters of closely related people based on a likelihood ratio-based measure of distance. They demonstrate the effectiveness of their approach using real examples and simulations.

Next we assume that pedigree information is not available for some missing persons. For such persons, the term $\Pr(G_i|P_{L+1} \ldots P_N, A_j)$ can be set to the product rule estimate $\Pr(G_i)$ since it is an unconditional probability. Under such situations, there will be some sets of assignments with equivalent posteriors. These will be those A_j with alternate arrangements of the people with no pedigree information.

In many circumstances there will be fewer recovered bodies or parts than missing persons. After collapsing of the DNA profiles down to the subset of unique profiles, there will be even fewer profiles to assign to pedigrees. Hence, in a practical example, we may have fewer pedigrees than required and fewer unique DNA profiles than we have missing persons.

Let the number of unique DNA profiles be X, which will be fewer than N. Of these unique profiles, L are assigned by direct comparison to samples from the missing person (toothbrushes, etc.). This leaves $X - L$ unassigned bodies to assign to $N - L$ missing persons, of which Q do not have pedigree

information.[h] In such cases, $X - L < N - L$. Again, all that is required is to assess the equation

$$\Pr(A_j | G_{X-L+1} \ldots G_X, P_{L+1} \ldots P_{N-L-Q})$$

$$= \frac{\Pr(G_{X-L+1} \ldots G_X | P_{L+1} \ldots P_{N-L-Q}, A_j) \Pr(A_j)}{\sum_{i=L+1}^{N} \Pr(G_{X-L+1} \ldots G_X | P_{L+1} \ldots P_{N-L-Q}, A_i) \Pr(A_i)}$$

$$\approx \frac{\prod_{k=X-L+1}^{X} \Pr(G_k | P_{L+1} \ldots P_{N-L-Q}, A_j) \Pr(A_j)}{\sum_{i=L+1}^{(N-L)!} \prod_{k=X-L+1}^{X} \Pr(G_k | P_{L+1} \ldots P_{N-L-Q}, A_i) \Pr(A_i)}$$

Even with these expedients, this may be an insurmountable computational problem. The larger the number of bodies unrecovered and the larger the number of missing persons without pedigree or direct comparison information, the less the advantage of closed set matching. There will come a point where this approach is not worthwhile. Brenner and Weir[83] discuss a number of shortcuts.

11.6.2 The Waco Incident

In 1993 an offshoot of the Seventh Day Adventist Church, calling itself the Branch Dravidians, had established themselves in a ranch known as Mount Carmel, not far from the town of Waco in Texas. The Branch Dravidians believed, inter alia, that the second coming of Christ was imminent and that the end of the world was approaching. Subsequent official reports state that they had armed themselves with multiple firearms, including fully automatic and 50 caliber rifles, in anticipation of an apocalyptic showdown with government agents, who they likened to the Babylonians and Assyrians.[702]

Vernon Howell (also known as David Koresh) had been the leader of this group since 1987. Mr. Koresh and seven "zealots" had taken control of the Mount Carmel ranch from George Roden by armed assault in 1987. Mr. Roden was wounded in the affray. Subsequently, a jury did not convict any of the eight with regard to this incident. Mr. Roden later killed a man with an axe and was confined to a mental hospital.[425]

On February 28, 1993 at approximately 9:30 U.S. Central time, in an operation variously reported as being codenamed "Showtime"[372] or "Trojan Horse," two agents of the Bureau of Alcohol, Tobacco, and Firearms (BATF; the agents of this bureau are termed ATF agents) attempted to execute an arrest warrant for David Koresh for federal firearm and explosive violations.

[h] For example, no reference samples were available for 2 of the 141 victims in the 1996 Tupolev 154 crash.[44]

The accompanying search warrant authorized ATF agents to search the compound where Mr. Koresh and his followers lived. Helicopters were reported to have been used to distract the Dravidians on February 28.[527]

A firefight resulted. Who fired first has been the subject of much debate.[527] ATF agents reported that they had been met by a hail of gunfire. Grigg[372] reports that the Dravidians had been alerted, presumably because the press had been tipped off in advance and this information had circulated back to the Dravidians. This action was supposed to be part of a plan by the BATF to recover from bad publicity arising from a "60 Minutes" television program on sexual harassment within the Bureau and from the 1990 court proceedings filed in federal court by black agents claiming racial discrimination. The BATF denied alerting the press. Speculation has also arisen over the "Showtime" codename for the operation.[478] Dravidian survivors testified that Mr. Koresh opened the front door and urged the federal agents to hold fire.[372] According to the Dravidians, this action by Mr. Koresh was met by a hail of gunfire. Four ATF agents were killed and 16 wounded. Dravidian casualties were unknown at the time. Lynch[527] reports two dead and five wounded but five Dravidian graves were later found, possibly attributable to this initial exchange of fire.

Further speculation has arisen as to why the ATF agents had not arrested Mr. Koresh on one of his many visits to Waco or while jogging. Lynch reports that after the initial firing a cease-fire was arranged, under the terms of which the Dravidians would hold fire if the ATF agents left.[527] Subsequent retrospective legal analysis suggests that the arrest and search warrants may have been illegal as the paperwork for the firearms at the compound may have been intact. Grigg[372] and Kopel and Blackman[478] report that the Dravidians had a legitimate firearms and memorabilia business.

At this point the ATF requested trained negotiators from the FBI to attend. By afternoon, the FBI and Department of Treasury officials had deployed the elite Hostage Rescue Team. On March 1 at the request of Treasury department officials, the FBI took the lead in the incident.[702] During the next 51 days, 700 law enforcement officials participated. Between 250 and 300 FBI agents were present at any given time. Over the next period, negotiations or discussions occurred between the FBI and Mr. Koresh, Steve Schneider (Koresh's chief lieutenant), and about 54 other Dravidians. Mr. Koresh was reported to have said that he would come out when God told him to. Between February 28 and March 23, 35 Dravidians left (14 adults and 21 children). These people were possibly "expelled" by Mr. Koresh as being weak or troublemakers. These escapees were immediately arrested in an action presumably designed to encourage others to surrender.

The official report states that the FBI did not fire during the entire standoff. Pressure on the compound was slowly increased by tightening the

perimeter, cutting off electricity, and by playing loud music and annoying noises.

Finally, a plan to insert tear gas over a 48-hour period was approved by the Attorney General. On April 19 at 5:59, the FBI telephoned the compound to tell Mr. Koresh that gas would be introduced but that the FBI were not assaulting. At 6:02 an FBI M1 heavy tank with attached boom began inserting CS[i] gas. The Dravidians are reported to have fired on the vehicle that was designed to withstand the 50 caliber armour piercing fire. At 6:47, 389 Ferret rounds were fired into the building. These rounds also carried gas.

At 12:07 the Dravidians are alleged to have started fires at three or more locations. By 12:25 FBI agents heard systematic gunfire. Grigg[372] reports that Forward Looking Infra Red Film taken by helicopters of the final fire showed that gunshots had been directed at the Mount Carmel complex by the FBI. Lynch reports that the FBI stopped firetrucks approaching the fire, possibly because of fear for the safety of the firefighters from either gunfire or the burning ammunition. There were nine survivors of the fire. Seven escaped by themselves. A further two were aided by the FBI.[527] The fire, fueled by ordinance, and with the rescue services kept away, burnt until the fuel load was exhausted.[184]

The bodies of 75 people were eventually recovered (50 adults and 25 children under 15). At least 17 of these had died from gunshot injuries and one from stabbing. Five additional bodies, presumably those of the Dravidians who had died on February 28, were also recovered. Injuries to some of these five suggested that they were shot from inside the compound, and one set of injuries suggested suicide.[702] This finding was disputed by Warlow.[818] From within the compound 305 firearms were recovered. Approximately 1.9 million "cooked-off" (i.e., burnt in the fire) cases were present. Warlow gives a full inventory of the weaponry and reports his findings on the 23 British dead.

The time between death and recovery of the bodies was approximately six days. Insect activity, mainly due to flies, had been intense. To keep the flies at bay, the bunker, a central feature of the compound where many bodies lay, had been fumigated nightly. One group of deceased were so co-mingled that they were removed together in three sections. Three anthropologists required three days to separate them. One body, presumably from the initial encounter, had been buried outside in a sleeping bag. This burial had been videotaped. Four further bodies had been buried inside in the underground bunker.[789]

The bodies were fragmented, charred, partially incinerated, or putrefying. DNA analysis was undertaken by the FSS and the U.S. Armed Forces DNA

[i] σ-Chlorobenzilidinemalenonitrile. A powerful irritant and lachrymatory agent.

Identification Laboratory (AFDIL) in Washington, D.C. A U.K. developed quadruplex of STR loci and sometimes the results for the amelogenin locus were used by each laboratory. In this way, 61 bodies were identified, 40 of which could not be identified by any other means. Full profiles were obtained for 50 bodies, 6 bodies gave partial profiles, and 5 gave no result. The partial profiles showed the loss of the high molecular weight loci. Amelogenin was used in 23 instances.[180,184] In all, 26 positive identifications were made. Other identifications were limited by a shortage of relatives.

Lynch reports the trial in 1994 of 11 survivors on charges of conspiracy to murder ATF agents. All 11 were acquitted of these charges. Seven were convicted of lesser charges. Four were completely acquitted.[527] On August 25, 1999 the FBI issued a statement that pyrotechnic devices may have been used in the early morning of April 19. Tapes later confirmed the use of pyrotechnic rounds. Heyman[404] discusses improvements that have been made to Federal (U.S.) procedure.

DNA Intelligence Databases

12

SIMON WALSH AND JOHN BUCKLETON

Contents

The recent growth of DNA intelligence databases has increased the involvement of forensic science in law enforcement. Beyond their primary function of linking crimes and suspects, DNA intelligence databases also provide a repository of information on crimes and criminals. Analysis of these data can increase our understanding of the context within which forensic science operates as a tool of the criminal justice system (CJS). In addition, DNA databases require us to consider interpretation issues specific to their utilization.

12.1 A Brief History

The ability of DNA to incriminate or exonerate has been extended during the 1990s by the advent of DNA intelligence databases. The first jurisdiction to pass legislation allowing the collection and storage of DNA samples from convicted offenders was the Commonwealth of Virginia, U.S., in 1989.[24] During these early years, the databases were often comprised of VNTR profiles

and were sometimes constructed on an ad hoc basis.[189,563] By 1995 the U.K. had a legally established national DNA profile database based on a platform of STR technology.[851–853] This model was followed shortly afterwards by New Zealand.[390,851] National DNA databases were introduced into Holland and Austria in 1997; Germany (1998); Finland and Norway (1999);[537] Switzerland, Belgium, Denmark, France, South Africa, and Hong Kong (2001), Botswana (2002), Peru, Venezuela, Thailand, Jordan, Qatar, Syria, Tunisia, Croatia, Cyprus, Czech Republic, Italy, Cyprus, Greece, Iceland, the Ukraine, and many others.[408] A parallel process has occurred in Canada[320,804] and the U.S.,[418] where standardization was based on 13 STR loci known as the Combined DNA Index System (CODIS). Australian states and territories are operating DNA databases that are planned to be combined under the federally administered CrimTrac program.[553] Databases are planned in countries including Lesotho, Mauritius, Tanzania, Zimbabwe, Argentina, Bahamas, Chile, Columbia, Cuba, Uruguay, Malaysia, Macao, Bahrain, Lebanon, Libya, and Oman. Local, as opposed to national, databases exist in China, India, Uzbekistan, and many other countries.[408]

The first execution of an individual identified through a DNA database search occurred in Virginia in April 2002 when James Earl Patterson was put to death by lethal injection.[19] While imprisoned for another matter, Patterson's DNA sample "hit" samples associated with the 1987 rape and murder of 56-year-old Joyce Alridge in 1999. Following the DNA match, Patterson confessed and pled guilty to the allegations. Virginia was also the first state to execute a person convicted on the basis of DNA evidence when Timothy W. Spencer was sent to the electric chair in 1993.[191]

12.2 Functional Aspects

Typically, DNA intelligence databases consist of two separate collections of profiles: a database of the profiles of individuals who have either volunteered or been compelled to submit samples, and a database of profiles obtained from samples from crime scenes, or exhibits associated with an alleged offense (Figure 12.1). Often the operator of a database has a program that can compare:

- Crime samples to other crime samples
- Crime samples to individuals
- Individuals to individuals.

Corresponding profiles revealed through the above comparisons could each be termed a "hit," but each has a very different meaning. Obviously,

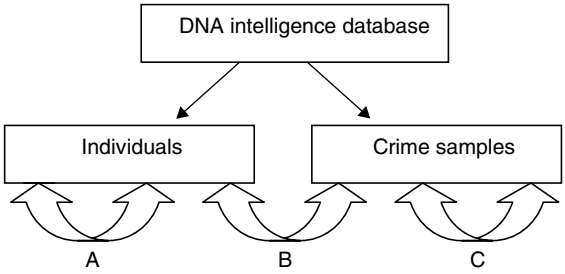


Figure 12.1 A diagrammatic representation of DNA database structure and match process, where *A* = individual to individual, *B* = crime to individual, and *C* = crime to crime. Reproduced from Walsh et al.[811] with the kind permission of the editor of *Science and Justice*.

crime-to-crime hits may suggest that the same person was at both scenes. Crime-to-individual hits may suggest that this particular individual was at this particular scene and may lead investigators to others who were involved.

Blakey[a,73,74] observed that: on average, every identification[b] leads to 1.4 detections, with over 61% of all identifications contributing some form of intelligence.

Individual-to-individual hits may contain information regarding the discriminating power of our systems and have implications for the reliability of our estimation procedure. In many cases, databases are not "clean." By this we mean that the same individual may be on the database more than once either under the same name or under aliases. This makes aspects of individual-to-individual comparisons difficult.

Blakey notes that there is an emerging national (U.K.) issue with multiple false identities appearing on the National DNA database (NDNAD). This indicates a need for more scrupulous attention to detail during custody handling.[74] The same issue is occurring in New Zealand who now have 50 unresolved duplicates in their database of 39,000. These are matching profiles that cannot as yet be ascribed to the same or different people. Investigation of these 50 duplicates is continuing but their resolution is a laborious process, even in a relatively small jurisdiction with a single police force.

Often programs that compare profiles do not require a perfect match. For instance, the Forensic Evidence Database (FoRED) software that manages entries and profile matching in New Zealand can allow for "no mismatch" results across all loci, or alternatively the operator may choose a fixed number of mismatching alleles. If very loose parameters are chosen for a match, then many individuals may match a given sample. These matches need to be

[a] The Blakey quotations are reproduced with the kind permission of Her Majesty's Stationary Office.
[b] Blakey uses the term identification synonymously with hit. Detection is used to mean that a person is charged.

assessed separately from the program results. The U.K. equivalent is the "near miss" program written by Richard Pinchin and Steve Knight. This process has caused considerable discussion among civil libertarian groups, but we can see little risk as long as these "near misses" are subsequently examined in detail during the confirmation process. Champod and Ribaux[172] elaborate on this and other aspects of the use of forensic identification databases.

12.2.1 Administration

As the technology that forms the basis for DNA intelligence databases is specialized, the operational components have remained the responsibility of forensic biology laboratories. In general, the database and its products are the property of law enforcement agencies, with the analytical and matching processes administered on their behalf by forensic institutions. All aspects of the process, whether handled by police or scientists, are subject to governing legislation. Often this legislation contains clauses that facilitate external review of operations by delegated parliamentary authorities. This is highly desirable.

From a forensic scientist's perspective, the legal basis for the administration of DNA databases represents an additional level of legislative governance over their work. Almost ubiquitously, DNA laws contain sections that prescribe the appropriate conditions under which a DNA sample can be collected, analyzed and stored, and the criminal sanctions that are enforceable for individuals in breach of these requirements. Although it is not possible to itemize all the various offense categories here, they generally include intentionally or recklessly supplying forensic material for analysis, improperly accessing or disseminating information stored on the DNA database, and matching profiles on the database unlawfully. Penalties can include fines and/or prison sentences. Some legal commentators feel that they are too lenient and do not provide sufficient deterrence to prevent a rogue scientist acting to pervert the course of justice through inappropriate use of a DNA database,[542] but we differ. The U.K. is an example where the custodial and contributor functions of the database are separated. The custodian is appointed by the Association of Chief Police Officers (ACPO), and is, at the moment, the FSS. Contributors include the FSS and also other government and private laboratories, all of whom may contribute to the national database. The custodial function of the FSS is administratively separate to the contributory laboratories. The custodian has a responsibility for monitoring the quality of contributing laboratories and may close them if required. The growth of the U.K. NDNAD has been significant, and it is projected to reach a target of 5 million samples by 2004. The system operates by analysis of buccal (mouth) scrapes or hair roots taken from any individual suspected or convicted of a recordable

offense. These are known as criminal justice samples. Results are stored on computer in the form of a digital code that is based on the nomenclature of each STR, and the database is known as an intelligence database. Early samples were genotyped using the SGM system and have six loci recorded. Current samples are typed using SGM^+ and have ten autosomal STR loci and amelogenin recorded. Older SGM profiles are not necessarily upgraded to SGM^+ for database purposes. During normal casework, operational laboratories carry out analysis of crime material such as semen or bloodstains. The DNA profiles derived from these samples are compared against the criminal justice samples in the existing database. If a match is found, then police are informed to enable further investigations to be carried out.

The FBI database, named CODIS, links all 50 states in the U.S. with the capability to search criminal DNA profiles. The structure of CODIS reflects the sociopolitical organization of the U.S. in that it has local (LDIS), state (SDIS), and national (NDIS) levels of operation. A laboratory can load and compare samples within their own database at the local level. Forensic DNA records originating at the local level can be uploaded to the state database that is housed at a central laboratory. This laboratory manages the collation of information at the state level and uploads to the national database that is maintained by the FBI. In order for a state to have DNA profiles included on the NDIS, it must first sign a memorandum of understanding whereby the state agrees to adhere to the FBI-issued quality assurance standards. The complexities of the U.S. model are paralleled in other jurisdictions, such as Australia, where state or provincial laws that govern database operations must be combined and compared at a national level. In the U.S., only complete convicted offender profiles are able to be loaded to the NDIS (13 STR loci or 4 RFLP loci). For casework samples, ten of the STR loci are required or a minimum of three RFLP loci.[137]

The demand for DNA testing has surpassed the capabilities of forensic laboratories in many parts of the U.S. Casework backlogs across the nation grew to such a degree that President George W. Bush sponsored "The President's Initiative to Advocate Justice through DNA Technology" and pledged US$1 billion to resource the eradication of the sample backlog. As a result of the vast increases in the volume of DNA database cases, private genetic testing laboratories are commonly contracted to analyze convicted offender samples on behalf of the central forensic agency. Forensic samples (usually exhibits relating to the alleged offense) tend to remain the responsibility of the crime lab. Some examples of this model exist in regions of the U.S., such as:

Virginia	
Forensic samples	Virginia Division of Forensic Science
Felon samples	Bode Technology Group

Indiana	
Samples	Indiana State Police Crime Laboratory
Felon samples	GenLex Corporation

Automated laboratory techniques can now facilitate the high-throughput analysis of thousands of forensic samples in a matter of days or weeks, allowing DNA databases to grow at a considerable rate.[321,339,798,850]

12.2.2 Performance Management

Aside from their core operational function, DNA databases can be seen as an amalgamation of DNA profile data from a large volume of crimes. In addition, for a proportion (usually in excess of 30–40% of the total number), there is information relating crimes to individuals or other crimes on the basis of a common DNA profile. Analysis of these data can provide information on features of the CJS such as crime distribution, offender demographics, DNA analysis methods, and police submission and management strategies. This information can in turn be used to better incorporate forensic evidence into investigative and intelligence frameworks or toward achieving important societal outcomes from the CJS. There is growing evidence that there is a substantial gap between the potential social good that could arise from a DNA database and the benefit actually realized. However, the full potential benefit is difficult to assess, as many jurisdictions do not monitor the performance of their databases beyond reporting a one-dimensional index relating to the proportion of hits. This is a major omission. Databases represent a significant financial investment and an intrusion on personal liberty; hence it is vital that their performance be monitored and optimized. The act of monitoring performance is likely to lead to improvements.[22]

In the U.K., performance monitoring was undertaken and reported in 2000[73] and 2002.[74] The author of these works, Inspector Blakey, highlights the role of the Scientific Support Manager within each force and recommends that these persons be placed high in the chain of command. In 2000, 60% of these were civilian employees and 40% police officers. Several factors were identified in these reports that required attention: In about 10% of all cases of identification, there was no evidence of any subsequent inquiry. A fraction, in the case of one force 10%, of all identifications did not relate to any recorded crime (presumably the sample had been submitted without being associated with the relevant crime record).

Poor file management: The original inspection endorsed the view, then widely held by the service (the English and Welsh police forces), that many DNA and fingerprint identifications did not result in detections. There was, then, little understanding of what action was taken in respect of these lost identifications and there was the presumption of the existence of a "black

hole" into which they disappeared. This assessment has determined that the black hole is now somewhat gray.

Blakey's last comment suggests that the act of investigating these lost identifications is likely to start processes by which this loss is reduced. Several factors were advanced as contributing to the underutilization of the DNA database resource:

Lack of good feedback information: Many forces have a great deal of difficulty in managing the process of turning identifications into detections, and this is rooted in a paucity of quality performance information. It is worth highlighting that intelligence, in this context, extends beyond that of crime and criminal. Only one of the forces assessed had established a process for feeding good practice back to Scene of Crime Officers (SOCOs), particularly information about unusual locations for retrieval of DNA stains and current crime patterns.

Timeliness: Timeliness is a matter of concern and there are significant delays in most of the forces assessed in commencing an investigation following receipt of the identification. About 16% of the cases examined in 2000 related to cases that had been finalized prior to the identification. About 51% of the delay time was spent within the force and about 49% within analytical process at the Forensic Science Service.

Interview technique: Blakey states that there are grounds to suspect that inept interview techniques and the inappropriate disclosure of evidence to defense representatives prior to interviews may afford suspects sufficient information with which to formulate spurious explanations for the presence of marks and stains and thus avoid prosecution.

Blakey argues convincingly that the answer to many of these problems lies in leadership from senior police, and in coordination and integration of forensic intelligence.

The lack of full engagement from chief officers was reflected in many of the responses from those forces assessed in this revisit, in that only three of the ten enjoyed the active participation of an ACPO ranking officer in championing the scientific support function. The remainder had devolved accountability to their heads of CID, and in many cases *de facto* accountability rests with Basic Command Unit commanders. Her Majesty's Inspector understood the pressures on chief officers, but remained convinced of the value of a forensic coordination group established within each force and championed by a chief officer if maximum results were to be obtained.

There is evidence that forces have recognized the need for a strategic approach to developing the full potential of a scientific support to the investigation of volume crime.

The assessment revealed several Basic Command Units that had successfully incorporated a contribution from Scientific Support Staff in their tasking

and coordination process. One of the main benefits of this approach was that the parameters for screening could be adjusted within a routine assessment of volume crime and aligned with other problem-solving tactics.

All of the forces assessed had made encouraging improvements in their use of scientific support functions in support of intelligence processes. The rollout of the National Intelligence Model has been a significant catalyst. As yet however, only two of the forces assessed had developed the means to target scene attendance by SOCOs as a result of their tasking and coordination process.

It is clear from the responses that producing fully effective crime scene attendance and screening policies continue to present difficulty.

In the U.K., various forces reported attending between 52 and 90% of all recorded burglaries of dwellings. No relationship to the number of DNA samples collected was noted. This suggests that those forces targeting the use of SOCOs are performing no better than those who do not target. About 2–4% of burglaries yielded DNA. This seems quite low and suggests that research into how many burglary scenes do, indeed, contain typeable evidence would be profitable. A linear relationship exists both in the U.K. and New Zealand between the number of samples submitted and the number of identifications obtained.

Blakey[73] reports that scientific support unit costs in the U.K. varied between 1.6 and 2.8% of total force spending. Costs per identification ranged between £443 and £13,114. Cost per detection ranged between £788 and £2342. He also noted that the number of individuals sampled fell below that permitted by the legislation by a factor of about 33%.

Another attempt to monitor DNA database performance more comprehensively and increase its overall informativeness was undertaken in New Zealand.[811] An additional database component was designed to extract intelligence information from case data held on the central system, particularly that relating "matches" or "hits." The result is an increased level of reporting capability whereby the New Zealand police can obtain reports on request detailing submission, analysis, or match trends from their station, district, or region. Data are represented against national figures and can also incorporate the geographic distribution of database links.

12.3 Legislation

The development of DNA intelligence databases has almost always been preceded by the alteration of existing legislation, or the creation of entirely new laws that codify, and often extend, the rights of the police to obtain and store DNA samples. In general, such legislation covers the grounds under which DNA samples can be obtained from suspects and convicted offenders, and

provides for the creation and administration of a DNA database. Many differences exist between jurisdictions. Usually these surround the focus and extent of post-conviction sampling and the grounds for requesting a DNA sample from a suspect. In general however, a clear trend has been shown for governments to embrace the concept of DNA databases and enact legislation accordingly.

For example, the U.K. has an aggressive version of DNA laws in comparison with other international jurisdictions. Any person convicted of, charged with, or suspected to have had involvement in the pursuance of a "recordable offense," may be required to provide a DNA sample.[22,414] The right to order a sample can rest with a police officer with a minimal rank of Superintendent. The most recent iteration (Criminal Justice and Police Act 2001 (U.K.)) has been referred to as a "compendious 'catch-all' criminal justice package."[820] In a highly significant case,[23] a suspect was identified following a DNA database match to a reference profile that should have been destroyed several months earlier (according to the legislative model of the time, the Police and Criminal Evidence Act 1984 or "PACE Act").[c] The House of Lords overruled the trial judge's decision to exclude the evidence. The legislative response was to remove the requirement to destroy samples following acquittal and in its place insert a rule that allows samples retained in such circumstances to be used for "purposes related to the prevention and detection of crime, the investigation of any offense or the conduct of any prosecution" (s.82). This amendment allows any sample that should have been destroyed, but has not been, to be used in the investigation of subsequent crimes. Wasik[820] expresses alarm not only at the legal decision to allow evidence that appeared to have been improperly obtained, but also at the willingness of governments to extend police powers, despite obvious breaches of legally enshrined individual liberties. "In the case where the police have acted in breach of PACE by retaining DNA evidence which quite clearly should have been destroyed, the legislative response has been to change the law with retrospective effect so as to legitimise what was improperly done."[820]

The U.S. provides another interesting example of the willingness to clear a legislative path for the use of DNA evidence and the construction of DNA databases. As discussed earlier in this chapter, states of the U.S. began to consider laws covering the compulsory acquisition of DNA samples from convicted offenders (usually sex offenders) as early as 1988. By 1994, 29 states had passed some form of DNA legislation, and by 1998 all states had done so, the last being Rhode Island (a summary of the current legislative status in the U.S. is provided in Table 12.1).[449] In 1994 Congress introduced the DNA Identification Act, specifically authorizing the FBI to create an index of DNA

[c] In particular s.64, which stated that *"any sample (intimate or non-intimate) must be destroyed if a suspect is cleared of an offense or not prosecuted."*

Table 12.1 Qualifying Offenses for DNA Database Laws in States of the U.S.

Type of Offense	No. of States (2000)	No. of States (2003)
Sexual offenses	50	50
Murder	37	50
All violent offenses	—	47
Assault/battery	28	—
Felony attempts	25	—
Juveniles	24	32
Robbery	19	—
Drug offenses	—	35
Burglary	18	44
All felonies	7	30
Some misdemeanors	—	23
Arrestees/suspects	—	4

Data adapted from Auslinkas et al.[24] and the DNA Resource website.[233] Although it is not possible to directly cross-compare all categories (a lack of data is signified by "—"), there is a general trend to increase the scope of DNA laws in the U.S. This is most notable with regard to burglary, juveniles, felony offenses, and arrestees.

profiles collected from all persons convicted of crimes, evidence recovered from crime scenes, and missing persons.

An interesting feature of the process in the U.S. (and undoubtedly elsewhere) is that in almost all cases no infrastructure existed to process, let alone record and match, DNA database submissions. For example, in 1999 the Governor of New York State (NYS) unveiled a plan to test all persons convicted of a felony, or attempted felony. The plan would increase the size of the NYS offender database by over 50,000 samples per year. By this time, 6000 offender samples had been received since collection began in 1994. It was then revealed, however, that of the 6000 samples collected over the preceding five-year period, only 1500 had been analyzed and none had actually been "matched" due to a lack of appropriate guidelines covering the database operations.

In the same year, Louisiana became the first U.S. state to legislate for the collection of DNA samples from all arrestees — the most sweeping legislative powers of any jurisdiction in the U.S. (and perhaps beyond). This was despite the fact that, at the time, no DNA database had actually been implemented in Louisiana at all. It existed only on paper. In a short time across the U.S., backlogs of convicted samples and evidence from crime scenes have developed, an issue that remains a focus of political, fiscal, and resource commitment.[760] It is clear from these examples how enamored legislators have become with the power of DNA testing. So much so, that sweeping laws have been widely introduced into jurisdictions that have yet to experience for themselves the pros and cons of operating a DNA database, and have inadequate fiscal or technological resources to operate one.

DNA databases suffered a rare legal defeat in the U.S. lower court (in Boston) when the Massachusetts law requiring a DNA sample from violent and sexual offenders was invalidated. In the case of *Landry v Attorney General,*[487] the judge agreed with the inmates' claim that the 1997 law amounted to an unreasonable search under both the state and federal constitutions. The ruling was eventually overturned in the U.S. Supreme Court, which ruled: "The state has established an indisputable interest in preserving a permanent record of convicted persons for resolving past and future crimes … and now will use DNA identification for these purposes."[449] Other courts have uniformly rejected challenges on such grounds.

In Australia a confounding issue has been dealing with interjurisdictional legislative variation (Table 12.2). As one of the fundamental aims of DNA-based laws was to codify the establishment of a NDNAD, divergences have slowed the pace of implementation of a unified national system. Obviously, key provisions vary to suit the particular concerns of the jurisdictional law enforcement agencies; however, it is unclear what in particular motivates the introduction of diverse models and what bearing the divergences will have on the effectiveness of the DNA database.

Internationally, there are several countries in which a regime of court authorized compulsory sampling exist. In Canada, the Criminal Code (DNA Search Warrant, 1995) allows for the taking of bodily samples for DNA typing from suspects in serious, violent offenses.[804] In the Netherlands, legislation allows the investigating judge to force a nonconsenting defendant to provide a sample for DNA testing, following an application by the Crown, provided the DNA analysis will assist in proving the case.[469]

Table 12.2 Various Legislative Models among Australian States and Territories are an Example of the Difficulty in Harmonizing DNA Laws across Jurisdictions

Jurisdiction	Legislation
Australian Capital Territory	*Crimes (Forensic Procedures) Act 2000*
Australia (Commonwealth)	*Crimes Amendment (Forensic Procedures) Act 1998*
New South Wales	*Crimes (Forensic Procedures) Act 2000*
Queensland	*Police Powers and Responsibilities Act 1997*
Victoria	*Crimes (Amendment) Act 1997*
South Australia	*Criminal Law (Forensic Procedures) Act 1998*
Western Australia	*Criminal Investigation (Identifying People) Act 2002*
Northern Territory	*Police Administration Amendment Act (No. 2) 1998* *Juvenile Justice Amendment Act (No. 3) 1998* *Prisons (Correctional Services) Act (No. 2) 1998*
Tasmania	*Crimes (Forensic Procedures) Act (2000) (Tas)*

Due to the concern about "function creep,"[d] some jurisdictions require the compulsory destruction of samples as a component of their DNA laws. This occurs in New Zealand, where under the *Criminal Investigations (Blood Samples) Act 1995* only the computerized record of the DNA profile is retained. The original biological sample and all other products of DNA analysis are destroyed within three months of receipt at the forensic laboratory. Similarly, forensic agencies in Germany, The Netherlands, Norway, and Belgium must destroy a sample once the DNA profile has been obtained, precluding reanalysis for the purpose of confirmation or updating DNA profile data.[537]

Some jurisdictions (such as the state of New South Wales in Australia) have not implemented this provision, and their administrators argue pragmatically that should the need arise to alter the technological platform upon which the database operates, this could only be possible if the samples were available for retesting.

12.4 Aspects of Forensic Significance

Historically, police and forensic scientists have had a focus on clearing one crime at a time. Forensic scientists receive cases, process them, and determine whether they have evidence supporting the prosecution or defense for a specific suspect and a specific crime. This has been tempered with occasional cases where forensic science can produce valuable intelligence before a suspect is located. This paradigm is changing. Utilization of DNA technology and DNA databasing is a means by which forensic scientists can expect to produce intelligence information in crimes for which there is no suspect.

Under a general concept referred to as intelligence-led policing, law enforcement agencies have recently moved to utilize academic and scientific expertise to assist in forming strategies to prevent, reduce, detect, or prosecute crime. This approach often leads to a greater utilization of forensic science, and there are examples where "forensic intelligence" has been successfully adopted into operational policing initiatives.[357,644] DNA intelligence databases have the potential for strategic use by police in a proactive as well as reactive context. We will discuss some examples of this in more detail, and highlight the need for the ongoing refinement of this approach.

Blakey[74] argues strongly for the positive aspects of database use and the use of forensic science *per se*:

> The contribution to the detection of both major and volume crime which comes from forensic science and fingerprints is clear

[d] This is the notion that the assembly of thousands (or millions) of DNA samples on criminal databases will present too great a temptation for governments who may sometime in the future stand to profit, or otherwise benefit, through widespread or targeted genetic research such as the search for genetic links to criminal behavior.

> and substantial. The use of both DNA and fingerprints has been developed and improved over the past few years and continues to provide new opportunities. If these opportunities are to be fully grasped the police processes and management need to be developed to keep up with the science.
>
> The FSS Pathfinder project focused on the impact of increased forensic activity during crime scene attendance, in particular using Low Copy Number (LCN) DNA, footwear and toolmarks and improving the capacity to link forensic intelligence. The project was particularly important for the empirical testing of a hypothetical model of the impact of forensic science on crime detection, prosecution and resultant reduction.
>
> Assessment indicated not only the frequency of availability of forensic material but also its typical value to detections. Where cases were charged and referred to the CPS for prosecution, the evaluation tracked the progress of the cases and assessed the contribution of the forensic evidence. The availability of forensic evidence appears to result in a high proportion of guilty pleas.[74]

While most forensic scientists and the police would like to share Blakey's view that the improved scientific means for identification is a source of assistance and support for crime victims by the inducement of a rapid guilty plea, the relationship between DNA evidence and outcomes from the CJS (such as an increased number of guilty pleas) is inconclusive. In particular, it has been suggested in the legal literature that justifications for DNA's use on the basis of its potential to solve serious crimes such as sexual assault and rape are misleading. The vast majority of these cases are characterized by the victim's reluctance to report the offense, a time delay between the offense and its reporting, and the existence or nonexistence of consent as the pivotal issue. DNA evidence is of little use resolving such issues. In fact, it has been proposed that the presence of DNA evidence actually increases the likelihood that a consent defense will be run at trial, subsequently disadvantaging the Crown case and increasing the plight of the complainant.[703] Recent criminological research from Australia differs from some of Blakey's findings.[93] This research indicated that while the existence of DNA evidence is a strong predictor that a case will be prosecuted, it did not show any significant effect in producing guilty pleas. This outcome led Briody[93] to conclude that DNA typing "places an increased financial burden on the taxpayer through increasing the number of court cases and, it might be argued, through its association with a larger prison population." Findlay and Grix[299] share this view

that in sexual assault cases in which DNA evidence is to be presented, the defense will focus on consent, placing more pressure, rather than less, on the victim.

In the earliest piece of DNA-linked criminological research of which we are aware, Purcell et al.[628] reviewed 55 cases between 1987 and 1991. Independent variables relating to the accused and the event characteristics were considered in combination with the presence or absence of DNA evidence. Given the timing of this research, the authors make the interesting observation that at the beginning of the study "there was concern that the case outcome would not be a variable at all. That is, given the scientific weight accorded DNA evidence, it was possible that the state would win all of the cases, resulting in a unimodal distribution, with all cases in the win column." In summary, the key findings of Purcell et al. are as follows:

1. The characteristics of the accused (age and employment status) rather than the characteristics of the crimes (black defendant on white complainant, stranger versus nonstranger) appear to have influenced the prosecutor's evaluations of the likelihood of a guilty verdict without DNA evidence.
2. A single variable made a statistically significant contribution, and cases involving older defendants were viewed as requiring DNA evidence.
3. Older defendants were less likely to be convicted than younger ones (although the contribution of this variable was just outside the 0.05 level of significance).
4. Case outcome appears to be independent of the selected crime characteristics in this study.

The presence of DNA experts at the trial was, by itself, a very good predictor of case outcome. In cases in which DNA experts testified, defendants were significantly more likely to be convicted than in those cases in which no such witnesses appeared.

The presence of DNA experts at trial explained 7% of the variance (in Model 3 — related to sentencing trends). The authors explain that this may be an example of the old adage — "you take up some of my time and I'll take up some of yours" — implying that the DNA evidence that was available before trial did not induce the defendant to plead guilty.

Through their findings, Purcell et al. conclude that "case outcome (i.e., conviction versus any other outcome) is to be understood largely in terms of only one of the variables included; DNA experts at trial. That is, in cases in which a DNA expert gave testimony, all other selected variables being equal, the accused was found guilty." This suggestion, that the use of DNA evidence has an impact on the case outcome, and possibly the sanctions imposed,

warrants further attention and provides some of the first empirical evidence of DNA's impact on CJS outcomes.

As previously mentioned, a clear consequence of DNA database operations has been an increase in the volume of casework. Unsolved crimes for the specific purpose of DNA Databank comparison now comprise over 40% of all forensic cases submitted in New Zealand. This has begun a change in thinking within forensic science. Historically, each case was processed as a unit or perhaps as a small series. This is evolving to a situation where all cases and individual samples are potentially linked. For some time, the ability to link cases and identify a crime series has been an effective strategy in law enforcement.[368] Incorporating DNA databases into the existing array of investigative data enhances the ability to link cases committed by the same individual or organization, adding also a highly discriminating mechanism for identification. Achieving greater integration of forensic and investigative techniques is an important developmental aim that should enhance the overall effectiveness of law enforcement.

In the 1980s, DNA profiling was primarily used to solve serious crimes. It now contributes to the investigation of a broad spectrum of crimes, including property offenses such as burglary. In New Zealand, property offenses collectively make up over 80% of overall database case submissions whereas violent offenses, including sexual assault, abduction, aggravated robbery, and homicide, comprise 12% of overall submissions. The variation in jurisdictional legislative and operational frameworks can alter the case submission profile; for example, certain countries in Europe have primarily focused on the investigation of serious, violent and sexual offenses.[694]

Related experiences involve the solution of crimes other than the one under investigation. There are instances in both the U.K. and New Zealand where the investigation of a major crime using DNA technology has led to the solution of a number of different and often completely unassociated crimes.

> Although DNA material recovered during the investigations into the murders of Michelle Bettles and Lorraine Turner has not helped to pinpoint their killers, officers say that it is helping to push up detection levels in offenses such as vehicle crime and burglary.[45]

Recidivism trends indicate an association between the number of prior arrests or appearances and the likelihood of re-offending[879] as well as a crossover among repeat offenders between petty crime and serious crime.[494] For example, 81% of burglars convicted in New Zealand in 1995 were re-convicted of an offense within two years.[725] Over 56% were re-convicted for another property offense and 24% were re-convicted for a violent offense.

Over 82% of violent offenders convicted in 1995 had prior convictions, 56% having prior convictions for property offenses. These figures should suggest the potential for crime clearance utilizing DNA databasing; however, research that integrates the function of DNA intelligence databases with criminological outcomes has been limited. The New Zealand database has about a 53% crime-to-person hit rate and a 28% crime-to-crime hit rate, which are above U.K. figures.

Through technological improvements, a range of evidence types can now be investigated for database purposes (many of these have been discussed in earlier chapters). Table 12.3 gives some approximate laboratory and database "success" rates for various evidence types analyzed at the New Zealand DNA database between 1996 and 2001.[811] We note but cannot explain the low database success rate for cases in which semen was the primary evidence type.

Assessment of the effectiveness of DNA intelligence databases is generally limited to primary indexes such as "match rates" or "hit rates," as described here. In the U.S., the CODIS system measures effectiveness by an index defined as the number of cases/investigations that CODIS assisted through a "hit," where a "hit" is defined as a match produced by CODIS that otherwise would not have been developed.[521]

Walsh et al.[811] present data that suggest a correlation between submissions and hit rates (Figure 12.2). This is not surprising, but it does support the hypothesis that the police are continuing to submit a good fraction of samples from criminally active people. An obvious leveling of the hit rate versus submissions would suggest that the police were no longer submitting samples from criminally active persons. A continued correlation would support the continued submission of samples or even a submission at increased rates. Further exploration of this relationship could augment sampling and submission strategies for the more efficient utilization of DNA database

Table 12.3 Summary of Analytical Success and Database Match Rates for Each Major Evidence Type

Sample Type	Success Rate (%)	Crime-to-Person Match Rate (%)	Crime-to-Crime Match Rate (%)
Blood	73	40	40
Semen	72	27	11
Cig. butt	50	38	50
Bottle swab	42	33	48
Saliva	41	36	51
"Trace"	23	36	44
Hair	16	27	31

Reproduced from Walsh et al.[811] with the kind permission of the editor of *Science and Justice*.

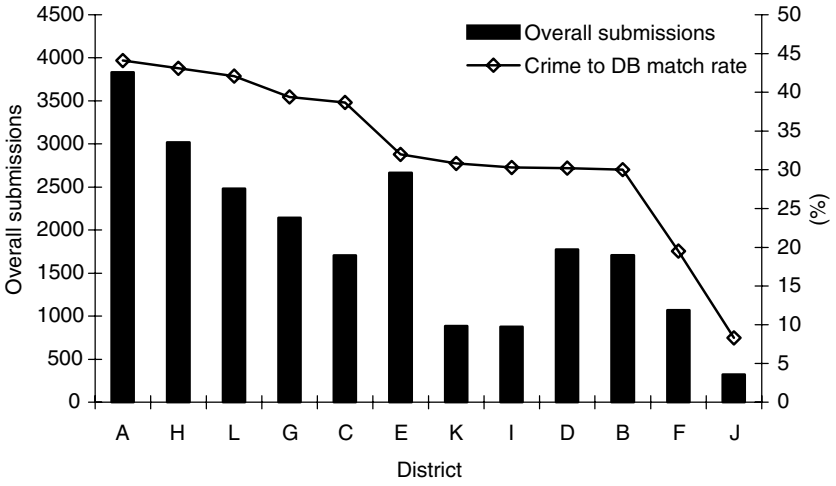


Figure 12.2 Submission rates and hit rates for varying districts in New Zealand. Reproduced from Walsh et al.[811] with the kind permission of the editor of *Science and Justice*.

resources. Examining data following the initial years of operation of the U.K. database, Werrett and Sparkes[852] observed "an apparent straight line correlation between the number of stains from scenes of crime submitted and the number of matches reported back." They also point out how the primary link provided via database matches often allows the police to resolve numerous other offenses which have been linked via alternative forms of intelligence. A vital correlation observed by Werrett and Sparkes exists between timely result reporting from the database and the ability of the police to successfully utilize the match result to obtain a detection or case resolution. Rapid reporting is an essential component for operational intelligence. This research illustrates the crime-solving potential of DNA intelligence databases, particularly when the results are promptly available and combined with the suite of preexisting intelligence tools available to law enforcement agencies.

Figure 12.3 shows the origin of the individual sample that hit against a crime sample from New Zealand's two largest cities. This demonstrates how the submission of samples from individuals resident in a particular locality can be important to the crime clearance in a different locality. These data also attest to the mobility of offenders and the capacity of DNA intelligence databases to transcend jurisdictional boundaries and contribute to investigations at a national level. Geographical representation of crime data has become commonplace,[633,634] and similar representations of cases linked via DNA evidence may augment the understanding of aspects of crime and offender behavior.

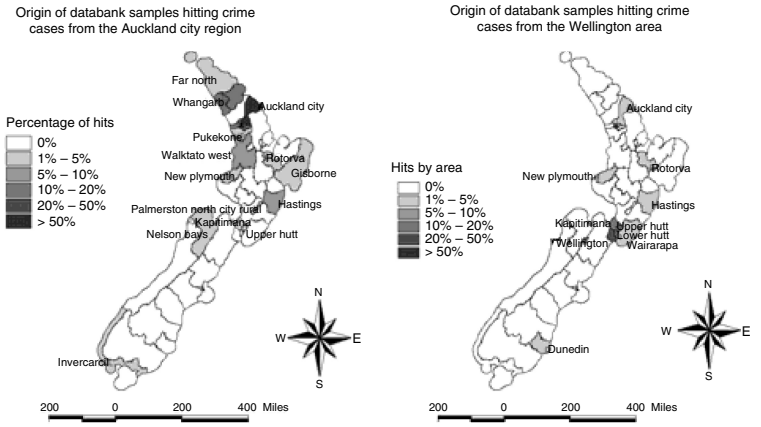


Figure 12.3 Geographical distribution of crime-to-person matches from two New Zealand Police Districts. Reproduced from Walsh et al.[811] with the kind permission of the editor of *Science and Justice*.

A convincing demonstration of the potential utility of forensic science is given by Operation VENDAS, undertaken in New South Wales, Australia.[724] In this operation the police optimized SOCO attendance at break and enter and motor vehicle scenes, raising scene attendance from 30 to 73% and 50 to 95% in two of the three districts trialled (we do not have data for the third). Also trialled was the use of digital photography to assist rapid intelligence from fingerprints and a seven-day turn-around time for DNA. The use made of the forensic intelligence was also optimized through access to investigative and other forms of operational support. This operation produced the results given in Table 12.4. Confirmatory publication of the results of this operation is still awaited.

It would have been interesting to see whether or not there was also an effect on violent and sex crime. Crossover rates among offenders would suggest that there should have been a reduction in these crimes as well, if not contemporaneously, then perhaps over subsequent months or years.

Balls[45] reports the results of increased scenes of crime effort in Norfolk, U.K. The number of DNA samples rose from 1098 to 2142 in a year, with a comparable increase in the number of DNA hits. As well as increasing DNA recovery, the increased attendance by SOCOs had also led to an 18% increase

Table 12.4 Reductions in Break and Enter and Motor Vehicle Crime Achieved during Operation VENDAS

	State Average	Brisbane Water	Lake Illawarra	Miranda
B&E	14.6%	41.9%	9.1%	30.9%
Motor	25.7%	42.6%	43.6%	30.4%

in fingerprint submissions and a 7% increase in identifications without an increase in the staffing level at the fingerprint bureau.

In the U.K., forensic evidence such as fingerprints and DNA are integrated into the successful Force Linked Intelligence System (FLINTS), which was developed by the West-Midlands police.[309] This integrated intelligence model has received broad support, in particular for its spatial representation of crime links and networks.

In another comprehensive and successful example of integration of forensic case information (of the kind envisaged by Blakey), law enforcement and forensic personnel from the French-speaking regions of Switzerland have merged the individualizing potential of forensic evidence types, including but not limited to DNA, with traditional sources of investigative intelligence such as modus operandi and spatiotemporal crime characteristics. At the heart of this approach is a sophisticated computerized platform utilizing case-based reasoning models of artificial intelligence.[644,645] This is based on a framework that fully integrates forensic case data around two central components:

1. A structured memory representing what is known about the criminality under consideration at a given point in time (outstanding crimes, active series, linked cases, and so on).
2. An organized repertoire of inference structures that reveal how to combine the use of different types of data in the course of the analysis, aiming to mirror the decision-making of an experienced investigator.[e]

In many respects, the success of the Swiss approach lies in its dispensation of what could be termed "traditional attitudes" toward forensic databases, and its acute awareness of what constitutes meaningful intelligence. Ribaux et al.[646] define intelligence (in general terms) as "the timely, accurate and usable product of logically processed information … (where) … in the context of the criminal justice system the information pertains to crime and the context in which it occurs." And forensic intelligence as "the accurate, timely and useful product of logically processing forensic case data."[646] Through their ability to provide technical information capable of directing police investigations (such as by identifying a suspect or a crime-to-crime link), DNA databases can be seen as a source of intelligence. In this way, DNA databases operate in a similar manner to automated fingerprint index systems (or AFIS). This proactive role is somewhat novel for forensic DNA evidence, which is usually employed to corroborate or refute a previously held belief or version of events. In some respects, the remnants of this retrospective, identification-driven use

[e] We recommend anyone interested in the intelligence-based use of forensic case data to review this novel and successful approach.

of DNA evidence provide a conceptual barrier to its exploitation under a truly intelligence-based model. These include:

1. A lack of timeliness (such as that identified above through the work of Blakey and in the examples relating to the present U.S. backlogs).
2. The lack of a framework to integrate DNA links with other forensic or investigative information. This can be accentuated by attempts to replicate the identification-based approach of DNA and AFIS databases with other forms of forensic evidence. This has proven to be complex and computationally laborious due to the fragmentary and visual nature of other trace evidence types.
3. An underutilization of the value of scene-to-scene DNA links due to a myopic emphasis on identifications. Scene linking is a fundamental goal of crime analysis and DNA databases can complement and reinforce these efforts. Some advantages and impediments to this union include:

    a. An increase in the number of "detections" related to each DNA match
    b. Recognition of serial crimes and crime "hot spots"
    c. A tendency for the "aura of certainty" associated with DNA matches to supplant rather than complement other investigative strategies
    d. A failure to systematically exploit information associated with DNA-based cases and links, such as characteristics of the crime and the evidence, or the offender profile.

4. Although there is much interesting research and development occurring, the forensic community is yet to determine the most effective manner in which to coordinate and exploit the intelligence potential of DNA (and other) evidence. In most countries, DNA databases are still in their infancy.

12.5 Social and Ethical Considerations

There are worrying aspects to the present and future uses of forensic DNA profiling for some members of the international criminal justice community.[465,636] The formidable expansion in the use of DNA, catalyzed largely through the introduction of forensic DNA databases, has increased the relevance of sociolegal and ethical perspectives in strategies for applying forensic DNA techniques. Our field has evolved into a far-reaching public tool, and the forensic scientist's responsibilities now extend beyond the traditional boundaries of the laboratory door and the courthouse steps.

Considerable debate has accompanied the use and development of forensic DNA profiling.[813] The context of the debate has changed from a focus on

the technology itself, to one that encompasses broader issues related to the most appropriate means for its use. Most debate in this area has emanated from the legal community (a review of the relevant features of this discussion is forthcoming[810]).

The contribution of legal commentators covers ideological issues, issues of identity, and operational issues. DNA databases and DNA-based legislation are contemporary operational issues and have been the subject of substantial review. Specifically, contentious issues include:

- Construction of DNA-based legislation.
- Hurried passage of legislation and the lack of public, legal, and political scrutiny.
- Lack of transparency associated with crucial policy decisions that affect the operational reach of the DNA database (such as sampling arrestees, suspects, or offenders).
- Interjurisdictional variation and how to address incongruence.
- Appropriateness of increased police powers.
- Striking the appropriate balance between the crime investigation needs of the state and the privacy rights of its citizens.
- Perceived encroachment into previously sacred territory of criminal law and an associated diminution of rights, including the right to silence and the right against self-incrimination (for example, Gans[328] describes the strategy of "DNA request surveillance"). Concern has also been expressed over the storage of human genetic information and the potential for future misuse.[25]
- Justification for DNA databases and associated legislative change.
- Links between legislative changes and increasingly punitive attitudes toward criminal justice.
- Lack of a convincing empirical basis to harness the investigative value of DNA profiling (and a perception that the strategy is simply one of "more is better").
- Lack of empirical justification for DNA databases or an assessment of their utility (other than poorly defined notions such as general deterrence).

Findlay and Grix[299] point out that the use of DNA dragnets further erodes civil rights. If we recall Professor Sir Alec Jeffreys's early DNA case involving the murders of two 15-year-old girls, Lynda Mann and Dawn Ashworth,[816] the perpetrator was apprehended because of his attempt to avoid giving a sample. Noncompliance in a voluntary dragnet is taken as a cause for suspicion rather than the exercise of a legitimate choice. "There has been enough challenge to the reality of informed consent within forensic procedures without the added strain concerned with the actuality of violation in mass-testing

situations." Pro use advocates would state that this erosion of rights was justified by results. Anti use advocates point to the emerging fiction regarding the concept of voluntary consent.

Despite these concerns, the technology underlying the use of forensic DNA profiling continues to develop at a substantial rate. Since its inception, the technology has gone through discernible phases of refinement and standardization (the settlement on STR methodology), sophistication and expansion (more automated techniques and DNA databasing), and more recently enhancement of capability (through miniaturization, portability, and determination of features such as hair color). Many of the changes in the landscape of forensic DNA profiling have a bearing on its use in criminal investigations, and its potential to impact outcomes of the CJS. For example, a U.K. Government White Paper recently called for every baby's DNA to be stored for future health care (not forensic) purposes.[192] Likewise in Iceland and Estonia, the entire community's genetic material has been licensed to private genetic research companies.[252,318] Whether such a resource would be used for forensic purposes in the future is an issue that would warrant careful debate. To balance the ongoing and increasing use of such technology, with its associated cost and privacy, rights, and misuse concerns, the benefits must be clear and must substantially outweigh the risks.

While some of these issues may seem nonforensic in origin, it is impossible for us to isolate discussion that is relevant only to a single discipline and/or irrelevant to another. A forensic scientist now must be aware of their legislative obligations under DNA laws and the sociolegal and ethical controversy associated with the regime under which they work. The forensic community must continue to refine not only the technology but, most crucially, the manner in which it is applied, so that the most effective and judicious use of DNA databases is ensured. We must be receptive, rather than reactionary, to social, ethical, and cultural concerns.

12.6 Interpretation Issues Associated with DNA Databases

The advent of DNA intelligence databases has not only altered operational features of the role of forensic DNA evidence, but also the context within which this evidence is interpreted. Interpretation models for forensic DNA evidence have evolved considerably. As a recent adaptation, it is understandable that testimony relating to DNA databases is now being scrutinized. Issues associated with this area are discussed here.

12.6.1 Adventitious Matches

In this section we attempt to show how to estimate the chance of adventitious matches in a database of size N. This is a very difficult task, largely

because of the structure of databases and the fact that the match probabilities between two people, or between people and crime samples are not constant. This matter has been tackled previously by Weir.[843] Initially we assume that databases have no structure, that is, they have no pairs of brothers or other relatives, no ethnic structure, and that the match probability for any two people taken from the database or from the contributors to the crime database is the same (P_m). This model is unrealistically simple, but it is the model most often considered. A more realistic model will be discussed later but is difficult to evaluate.

Let the database of individuals be of size N. Let there be C stains on the crime database. Then there are $N(N-1)/2$ pairs of people on the database, $C(C-1)/2$ pairs of crime stains, and NC person-to-crime comparisons. It is educational to see how quickly these numbers become large. Consider a database of size 27,000, with 2000 crime stain profiles. This results in 364 million comparisons between people, 2 million comparisons between crimes, and 54 million comparisons between crimes and people. Initially we assume a constant match probability (P_m) between people and treat the number of matches as binomially distributed (Weir has previously pointed out that the trials are not independent, nor is the match probability constant as required for the binomial to apply).

As an example, let P_m be 1 in 50 million. Then we expect the following.

Number of Matches	Probability of This Number of Matches between People and Crimes	Probability of This Number of Matches between People	Probability of This Number of Matches between Crimes
0	0.34	0.00	0.96
1	0.37	0.00	0.04
2	0.20	0.02	0.00
3	0.07	0.04	0.00
4	0.02	0.08	0.00
5	0.00	0.12	0.00
6	0.00	0.14	0.00
7	0.00	0.15	0.00
8	0.00	0.13	0.00
9	0.00	0.11	0.00
10	0.00	0.08	0.00
11	0.00	0.05	0.00

If the match probability is reduced, then these probabilities become weighted toward the top of the table. Below we give the same table by assuming a match probability of 1 in a billion.

Number of Matches	Probability of This Number of Matches between People and Crimes	Probability of This Number of Matches between People	Probability of This Number of Matches between Crimes
0	0.95	0.69	1.00
1	0.05	0.25	0.00
2	0.00	0.05	0.00
3	0.00	0.01	0.00
4	0.00	0.00	0.00
5	0.00	0.00	0.00
6	0.00	0.00	0.00
7	0.00	0.00	0.00
8	0.00	0.00	0.00
9	0.00	0.00	0.00
10	0.00	0.00	0.00
11	0.00	0.00	0.00

If we abandon this simple model, as we must, the key factors to consider are:

1. Many crime profiles are partial, that is they do not have all loci scored. This is often due to technical issues such as limited sample or degradation. Hence they will have a higher P_m than full profiles. In such cases the match probability of, say, 1 in a billion may not be reasonable. Some individuals will be scored at a limited set of loci. In the New Zealand and U.K. cases, this will most likely have occurred because they were typed with SGM (six loci) rather than SGM$^+$ (ten loci).
2. There will be related people on the database. For these people the match probability will be larger than the match probability between unrelated people. These related people will increase the number of person-to-person matches but will not increase the number of adventitious hits.
3. Crime samples may be from people related to those on the database. This will increase the number of adventitious hits.

For these reasons, it is impractical to treat match probability as known and constant. The more reasonable alternative is to consider the match probability as a variable between pairs of people and between people and crimes.

Weir,[843] as usual, gives a superior formulation allowing varying match probabilities and accounting for subpopulation effects:

$$P_M = \sum_i \Pr(A_iA_i|A_iA_i)\ \Pr(A_iA_i) + \sum_{i\neq j} \Pr(A_iA_j|A_iA_j)\ \Pr(A_iA_j)$$

$$= \frac{\left\{\begin{array}{c} 6\theta^2 + \theta^2(1-\theta)\left(2 + 9\sum_i p_i^2\right) + 2\theta(1-\theta)^2\left(2\sum_i p_i^2 + \sum_i p_i^3\right) \\ +(1-\theta)^3\left[2\left(\sum_i p_i^2\right)^2 - \sum_i p_i^4\right] \end{array}\right\}}{(1+\theta)(1+2\theta)}$$

Triggs and Buckleton give an approximation to a model allowing varying match probabilities for classes of relationship (Box 12.1).

To our knowledge, this model has not yet been implemented and would require some detailed knowledge of relatedness within a database or between offenders and persons on the database to implement. Without such a model, we can still get the flavor of the results. Matches between people on the database are expected to occur. The number of these will be increased by the relatedness of people on the databases. Adventitious matches between crime samples and people are also expected. This will most often be between people with partial profiles on the database or to those crime samples for which there is a partial profile. Post match confirmation will be beneficial in reducing this only when additional loci can be compared. People on the database may be at risk of matching to crime samples left by relatives.

Box 12.1 Triggs and Buckleton

Person-to-person comparisons: Consider partitioning the pairs of people on the database (or the crime person pairs) into, say, those pairs that are brothers, cousins, unrelated Caucasians, unrelated Caucasian Maori pairs, etc. Let the *i*th partition contain N_i pairs. The match probability for this degree of relatedness is P_i (again assumed constant and known). The distribution of matches for each partition X_i is modeled as having a binomial distribution $X_i \sim B(N_i, P_i)$. This does not take proper account of Professor Weir's correct objection that the pairs are not independent trials. If we model using a Poisson distribution, then the distribution of matches for each partition is $X_i \sim Poisson\ (N_iP_i)$ and the total number of adventitious matches X is $X \sim Poisson\ (\sum_i N_iP_i)$. This model still assumes that the match probability is constant and known within partitions and that the pairs are independent, and hence it is still an incomplete treatment. The New Zealand database is dynamic in that profiles may be added or removed. Occasionally a snapshot or cut is taken of the New Zealand database and cleaned. This process identifies all matches, with all matches subsequently examined. This is not trivial, and in our latest cut we still have 50 unresolved pairs. For instance, persons with the same surname at the same address may be one person using aliases or may be brothers. It would be unwise to assume that they were the same person simply from this data. After examining these pairs, we are usually left with some matches. In one cut on the database, we had 10,907 SGM six-locus profiles giving about 59 million pairs of people. This produced ten matches.[641] Of these, eight were twins or brothers, one was a case of subterfuge where person *A* had persuaded person *B* to give a sample as person *A* as well as himself (person *B*),

(*continued*)

Box 12.1 (*continued*)

and one was a match between apparently unrelated people. The match probability for unrelated people at the SGM loci is about 1 in 50 million.

Crime-to-person comparisons: Next consider a crime stain profile. We assume that this has been left by a person who is not on the database. If the true donor is on the database, our problem is very much alleviated as we expect to get a true match. We may also get one or more adventitious matches, but the plurality of these will alert us to our difficulty. Therefore, suppose that the true donor is not on the database. However, the true donor may be related to people on the database. Again we can follow the partitioning approach by considering *NS* pairs as comprising partitions that can be treated as brothers, cousins, etc. The result is the same as above.

12.6.2 Assessing the Strength of the Evidence from a Match Derived from the Intelligence Database

The strength of the DNA evidence resulting from an intelligence database match is often presented without any mention that the hit was obtained from a database. It is usually not in the suspect's interest to let a jury know that he/she has a sample on the DNA database. The question of whether searching an intelligence database for a match affects the strength of the evidence has been discussed extensively and forcefully in the literature. The issue also affects any search among suspects whether they are on a database or not. Unfortunately, there is much confused writing and it would be very difficult for a court to make a reasoned decision based on a simple assessment of literature recommendations.

The first National Research Council (NRC) report[584] suggested that those loci used in the matching process should not be used in assessing the weight of evidence. The second NRC report[585] (pp. 133–135) recommended that an adjustment be applied by multiplying the match probability (P_m) by the number of people on the database. Using an example of an intelligence database of 1000 and a multiplex with a P_m of 10^{-6}, this would result in a P_m of 0.001. To date, Aitken,[8] Balding and Donnelly,[37,40] Balding et al.,[42] Berry,[57] Dawid,[219] Dawid and Mortera,[221] Donnelly and Friedman,[236] Evett,[262] Evett and Weir,[267] Evett et al.,[273] and Finkelstein and Levin[300] have all suggested that the evidence is slightly stronger after a database search and hence no correction is required. Meester and Sjerps[543] suggest that a flat prior be assumed and that the posterior probability be reported, but essentially support the "unadjusted" *LR*. Taroni et al.[757] take a Bayes net approach and reach the same conclusion: "the result of the database search has the character of an additional piece of information."

The two NRC reports,[584,585] Devlin,[229] Lempert,[502] and Stockmarr[728–730] have suggested that the match probability be multiplied by the number of people on the database (termed the *NP* approach). Lempert[502] suggests multiplying by the size of the suspect population and not the database. Morton[565] suggests confirmatory markers or the use of the *NP* approach. The National Commission on the Future of DNA Evidence appears undecided.[641] It is unclear what Goldman and Long[359] desired, but they did suggest that "the estimated match probability must be adjusted to take into account multiple testing."

Here is one of our favorite pieces:

> Because the probative effect of a match does not depend on what occasioned the testing, we conclude that the adjustment recommended by the committee (NRC II) should not be made. The use of Bonferroni's adjustment (multiplying by the size of the database) may be seen as a frequentist attempt to give a result consistent with Bayesian analysis without a clear-cut Bayesian formulation. (Finkelstein and Levin.[300])

Gornik et al.[367] give an interesting practical demonstration of the problem with regard to the identification of war victims using reverse paternity.

Again, the only scientific safeguard is confirmation at additional loci. We must recommend continued scientific investigation to foster an understanding and assessment of these risks. It seems likely that the public, at least in New Zealand, is not fully informed of these risks nor has informed public debate occurred.

What would a biologist or court make of this? The mathematical arguments given by either side appear impressive; however, we believe that the weight of logical argument is on the "no correction" side.

We attempt here to formulate a summary of the most well-supported approach based on quality of the mathematical argument. We use terms familiar from Chapters 3, 4, and 5. For a population of size N, we index the suspect as person 1 and the remaining members of the population as $2 \ldots N$. The first $1 \ldots M$ of these are on the database. As before, we will call the hypothesis that person i is the source of the DNA: H_i. Since the suspect is indexed person 1, the hypothesis that the suspect is, in fact, the source of the DNA is H_1. The remaining hypotheses, $H_2, \ldots, H_N$, are those hypotheses where the true offender is one of the $N-1$ "other" people. Before we examine the evidence, each person has some probability of being the offender, $\Pr(H_i) = \Pi_i$. The suspect is on the database and has been genotyped. As before, we call his genotype G_s. The stain from the scene has been typed and found to have genetic profile G_c. The search of the database reveals that $G_c = G_s$ and no other profiles on the database match (this latter requirement is unnecessary

and a generalization is easy). We know that the $2 \ldots M$ people on the database other than the suspect do not match. The remaining $M + 1, \ldots, N$ members of the population have genotypes $G_{M+1} \ldots G_N$, which are unknown. We require the probability $\Pr(H_1|G_c, G_1 \ldots G_M)$.

$$\Pr(H_1|G_c, G_1 \ldots G_M) = \frac{\Pr(G_c|H_1, G_1 \ldots G_M)\Pr(H_1|G_1 \ldots G_M)}{\sum_{i=1}^{N} \Pr(G_c|H_i, G_1 \ldots G_M)\ \Pr(H_i \mid G_1 \ldots G_M)}$$

$$= \frac{\Pr(G_c|H_1, G_1 \ldots G_M)\Pr(H_1|G_1 \ldots G_M)}{\left(\begin{array}{l} \Pr(G_c|H_1, G_1 \ldots G_M)\Pr(H_1|G_1 \ldots G_M) \\ + \sum_{i=2}^{M} \Pr(G_c|H_i, G_1 \ldots G_M)\Pr(H_i|G_1 \ldots G_M) \\ + \sum_{i=M+1}^{N} \Pr(G_c|H_i, G_1 \ldots G_M)\Pr(H_i|G_1 \ldots G_M) \end{array}\right)}$$

In the above equation, we have split the denominator into the suspect, the other people on the database, and those other people not on the database. Further, we will assume $\Pr(G_c|H_1, G_1 \ldots G_M) = 1$. This assumption, in words, is: The stain should match the suspect if it did indeed come from him.

$$\Pr(H_1|G_c, G_1 \ldots G_M) = \frac{\Pr(H_1|G_1 \ldots G_M)}{\left(\begin{array}{l} \Pr(H_1|G_1 \ldots G_M) \\ + \sum_{i=2}^{M} \Pr(G_c|H_i, G_1 \ldots G_M)\Pr(H_i|G_1 \ldots G_M) \\ + \sum_{i=M+1}^{N} \Pr(G_c|H_i, G_1 \ldots G_M)\Pr(H_i|G_1 \ldots G_M) \end{array}\right)}$$

Since the $2 \ldots M$ "other" people on the database do not match, $\Pr(G_c|H_i, G_1 \ldots G_M) = 0$ for $i = 2$ to M.

$$\Pr(H_1|G_c, G_1 \ldots G_M) = \frac{\Pr(H_1|G_1 \ldots G_M)}{\Pr(H_i|G_1 \ldots G_M) + \sum_{i=M+1}^{N} \Pr(G_c|H_i, G_1 \ldots G_M)\Pr(H_i|G_1 \ldots G_M)}$$

Under most conditions, we assume that the nonmatching people do not change our view of the match probability. This is not strictly true as the finding of only one or few matching people may reinforce our view that the profile is rare.[40] However, making this assumption and assuming that the genotypes do not affect the prior,

$$\Pr(H_1|G_c, G_1) = \frac{\Pr(H_1)}{\Pr(H_1) + \sum_{i=M+1}^{N} \Pr(G_c|H_i, G_1)\Pr(H_i)} \tag{12.1}$$

We compare this with the equivalent probability for the same match without a database search:

$$\Pr(H_1|G_c, G_1) = \frac{\Pr(H_1)}{\Pr(H_1) + \sum_{i=2}^{M} \Pr(G_c|H_i, G_1)\Pr(H_i) + \sum_{i=M+1}^{N} \Pr(G_c|H_i, G_1)\Pr(H_i)} \quad (12.2)$$

Under any set of priors, Equation (12.1) gives a larger value for the posterior probability than Equation (12.2) since the denominator is larger in (12.2). Hence, reporting the approach leading to (12.2) is always conservative. In other words, the "no correction" approach is always conservative.

Up to this point we have made no really contentious assumptions (although we accept Balding and Donnelly's argument that the search also informs the match probability). If we now assume that every person has the same prior chance of being the true offender (this assumption is quite contentious) and we write $\Pr(G_c|H_i, G_1)$ as P_m in Equation (12.1), we obtain

$$\Pr(H_1|G_c, G_1) = \frac{1}{1 + (N - M - 1)P_m} \quad \text{(database search)} \quad (12.3)$$

as compared with

$$\Pr(H_1|G_c, G_1) = \frac{1}{1 + (N - 1)P_m} \quad \text{(no database search)} \quad (12.4)$$

Again, Equation (12.3) gives a larger posterior than (12.4). Conclusion: The evidence is stronger after a database search and reporting the standard *LR* is always conservative.

This approach easily accommodates varying priors. Consider that people on the database are x times more or less likely to be the offender in this case than people not on the database. Inserting into Equation (12.1) gives

$$\Pr(H_1|G_c, G_1) = \frac{1}{1 + (N + M - 1)P_m/x} \quad (12.5)$$

We supplement this mathematical argument with some fables given by Buckleton and Curran[100] that may be more use in court than reasoned mathematical arguments.

Fable 12.1. Three people enter an empty train carriage, which is then locked. When the carriage is unlocked by the police, it is discovered that one of these passengers has been stabbed to death. The remaining two passengers immediately become suspects. Not surprisingly, each states that the other did it. What are the police to do? On examining the suspects, they find that one

is a tetraplegic and could never hold a knife, let alone stab a person with it. Therefore the police charge the other passenger. Moral: Excluding alternative suspects increases the probability of guilt for those suspects remaining.

Fable 12.2. Consider a case where we have typed many loci of a stain at the scene of a crime. The estimated match probability is 1 in 4 billion. We search a database of every person in the world and find that one (and only one) person matches. Clearly we believe that this is the true perpetrator. Why? Because he matches the crime stain *and* because no one else does (remember there could have been two or more matches as the match probability is not a count in the population). Moral: Excluding other persons increases our belief that the profile is rare.

Fable 12.3. Consider a case where we have investigated two crimes that are identical. A stain has been left at the scene and a partial driver's license giving an address and surname has also been left behind. In the one case we type brother *A* (of five from this address with this surname) first. He matches. We stop. In the second case we find a stain at the scene and type brothers *B–E* first. They are excluded. Then we type brother *A*. He matches. Stockmarr's analysis would have us downweight the evidence in this second scenario since we have searched among five persons. But we have eliminated the persons most likely to also match (his brothers) who are also the other primary suspects due to the partial driver's license evidence. Surely this does not downweight the evidence but increases it? Moral: Eliminating genetically close relatives is sound practice and increases the weight of evidence.

It seems plausible that if the suspect has been identified by a database search, the prior odds are lower than if he has been identified by some other means (such as an eyewitness identification). We see therefore that the possible effect on the posterior odds of a database search is not via a lowering of the likelihood ratio but by plausibly lower prior odds (although this is not necessarily always the case).

It is important that we:

1. Continue to encourage the development of evidence that a certain profile is, indeed, rare.
2. Continue to eliminate other likely suspects.
3. Encourage the investigation either genetically or otherwise of brothers who may be potential alternative suspects.

For an elegant analysis of the same issue in fingerprints, see Champod and Ribaux.[172] We would like to draw attention to their illuminating scenarios on pp. 475 and 476.

It is important that an incorrect analysis of database searches is not allowed to damage interpretation in this area.

12.7 Summary

Undeniably, the introduction and expansion of DNA intelligence databases has modified the landscape of forensic science significantly. Through the use of DNA intelligence databases, forensic science has become more important as a potential source of law enforcement intelligence, more widely applied in crime investigation, and more able to contribute proactively to strategies of crime reduction and prevention.

When coupled with developments such as portable, miniaturized, diagnostic laboratory tools, this impact is destined to increase even further. As the focus of law enforcement becomes increasingly trans-national, there is the potential for DNA intelligence databases to contribute to criminal investigation at a global level.

As with all areas of forensic science, we must continue to challenge and refine our understanding of this technology.

Glossary

STRBase[138,680] also provides a glossary.

2p rule A genotype probability assignment often associated with the *F* designation. It is used when the genotype may contain allele *p* and any other allele.

Admixture We are using this term to describe the population genetic event when two populations are mixing.

Adventitious matches A match to the profile of a person who is not the true donor of that profile.

Allele Technically, this refers to the different forms of a gene. However, in forensic DNA profiling it is misused to refer to the different forms of the intron, which, technically, is not a gene.

Allelic dropout The condition where an allele cannot be visualized.

Allelic ladder A standardization tool used to assist in sizing alleles. It consists of most of the common alleles at a locus.

ASCLD American Society of Crime Laboratory Directors. The parent body of ASCLD/LAB, the laboratory accreditation board.

Autosomes Any pair of chromosomes other than the XY pair.

Balding and Nichols formula A correction for the subpopulation effect. See Equations (3.4).

Bases The primary units of DNA. There are four bases that can be assembled into the DNA molecule, abbreviated as A, T, C, and G.

Bayes' theorem A mathematical theorem developed by the Reverend Bayes, stating that the posterior odds are equal to the prior odds multiplied by the likelihood ratio; see Equation (2.1 and 2.2).

Billion 1,000,000,000.

Biparental inheritance Inheritance from both parents.

Bp Base pairs, typically used as a unit of measurement.

CE Capillary electrophoresis.

Chimerism The situation where an individual shows differing genotypes in differing cells.

Chloroplast An organelle in plants thought to have descended from symbiotic blue-green algae.

Chromosome A physical structure of the nucleus that contains the DNA sequence. From the Latin for a colored body from their affinity to take up dye.

ChrX An abbreviation for the X chromsome.

ChrY An abbreviation for the Y chromsome.

CODIS Combined DNA index system.

Concatemers When linear DNA replicates, there is a gap left at the 5′ end of the new strand that the polymerase cannot extend across because of the absence of a primer terminus. One of the ways that viruses deal with this is for polymerase and ligase action to join the incomplete ends of the replicated linear strands. This aggregate of double-stranded linear DNA is called a concatemer. This aggregate is then cut by a nuclease into genome-sized bundles.

Control region of the mitochondrial DNA A noncoding region of the mtDNA.

D-loop A name for the control region of mtDNA that arose due to structures that are visible during replication.

Diploid Describes an organism or cell with two copies of each chromosome.

DNA fingerprinting Coined by Sir Alec Jeffreys to describe his initial multilocus probes that visualized the fragments produced by the enzymatic digestion of DNA. These probes produced a bar code-type output, which was described as the DNA fingerprint. The analogy to fingerprinting was found to be seriously unhelpful, and Evett and Buckleton advocated a change to DNA profiling that has been largely accepted.

DNA profiling The process of revealing parts of the DNA code. It can describe the visualization of anything from one to many loci. It is now universally accepted as the more useful term among forensic scientists.

Dropin The phenomenon where an allele has appeared as a contaminant in a profile.

Dropout The phenomenon when an allele has failed to amplify.

E The evidence.

EDNAP The European DNA profiling group.

Electrophoresis A technique used in DNA technology to separate fragments by size; it also has much wider applications in other areas of science.

ENFSI European Network of Forensic Science Institutes.

EPG Electropherogram.

Eukaryote An organism of one or many cells that contain a nucleus and separate organelles. Examples include all plants and animals. Contrast with prokaryote, which includes all bacteria.

Exon Portions of coding DNA that eventually give rise to protein structures.

F The within-person inbreeding coefficient.

***F* designation** A designation used to describe a genotype where an allele may have dropped out. For instance 16, *F* implies that the genotype may be 16 and anything else.

Final individuals in a pedigree Individuals in the pedigree without offspring.

Founders of a pedigree Those individuals at the top of a pedigree whose parents are unspecified.

F_{ST} The between-person inbreeding coefficient (used synonymously with θ in this book).

Gamete The reproductive cells: an egg or a sperm. These are haploid.

G_c Genotype of the crime stain.

Gene A functional sequence of DNA.

Gene diversity A measure of the diversity of a locus in a population, similar, but not identical, to heterozygosity.

General form of Bayes's theorem See Equation (2.4).

Genome The entire haploid DNA complement of an individual.

Genotype The genetic makeup of an organism as differentiated from its appearance. The genotype of an individual may be described at any number of loci.

Gonosomal This refers to the XY chromosome pair.

Good shedder A term used to describe a person who readily deposits DNA onto touched objects.

G_s Genotype of the suspect.

G_{ST} Nei's coefficient of gene variation (G_{ST}). Following Crow[208] (pp. 62–66), let p_{us} be the frequency of the uth allele in the sth subpopulation. Define gene diversity in the sth subpopulation $D_s = 1 - \sum_u p_{us}^2$. Crow comments that D_s is the probability of drawing two different alleles from within the sth subpopulation. If there are n subpopulations, $\bar{p}_i$ is the mean allele frequency across the subpopulations, $\bar{D}_s$ is the mean of the D_s values, and $D_T = 1 - \sum_u \bar{P}_u^2$, then $G_{ST} = 1 - \bar{D}_s / D_T$.

Haploid An organism or cell with a single copy of each chromosome.

Hardy–Weinberg equilibrium An assumption of independence at one locus. See Equation (3.1).

Hb Heterozygote balance.

H_d Defense hypothesis.

Heteroplasmy This term is used in mitochondrial work to describe the situation where an individual shows two mitotypes.

Heterozygosity A measure of the diversity of a locus in a population.

Heterozygote The genotype at this locus has two different alleles.

Heterozygote balance The area (or height) difference between the two peaks of a heterozygote.

Heterozygote imbalance The area (or height) difference between the two peaks of a heterozygote.

Homopolymeric Used to describe a sequence where one base is repeated.

Homozygote The genotype at this locus has two copies of the same allele.

$\mathbf{H_0}$ The null hypothesis.

H_p Prosecution hypothesis.

HV1, HV I Hypervariable region 1 (of the mitochondrial DNA).

HV2, HV II Hypervariable region 2 (of the mitochondrial DNA).

HV3, HV III Hypervariable region 3 (of the mitochondrial DNA).

Hypervariable A region of DNA that has many different variants among individuals.

I The background (nonscientific) evidence in a case.

Interspecific crosses Crosses between species.

Intron A portion of noncoding DNA usually between exons.

IUB International Union of Biochemistry and Molecular Biology.

IUPAC International Union of Pure and Applied Chemistry.

LCN Low copy number. This refers to the situation where there are very few DNA molecules available in the sample.

Linkage equilibrium An assumption of independence between loci. See Equation (3.2).

Loci/locus A position on the genome (loci is the plural).

Locus dropout The situation where the entire genotype cannot be visualized at this locus.

Logical approach The term preferred in this book to the more commonly used "Bayesian approach."

Madonna plot A stylized rendition of an EPG.

Matrilineal inheritance Inheritance solely from the mother.

Mendelian inheritance Inheritance that follows Mendel's two laws.

Mitochondria An organelle in eukaryotes associated with the production of ATP.

Mitochondrial DNA The DNA present as small circular molecules in the mitochondria.

Mitotype The genotype of the mitochondrial DNA.

Mosaicity A situation where an individual shows more than one genotype either in one tissue or in differing tissues. This is usually assumed to have been caused by somatic mutation. It may be relatively common. In fact, all individuals may be mosaic to some extent.

MSY The male specific region of the Y chromosome. This is the region that does not recombine with the X chromosome.

MtDNA Mitochondrial DNA.

Multilocus probes Used to describe two probes 33.15 and 33.6, which visualized many loci forming a "bar code" type of image.

Nested PCR An amplification procedure utilizing two amplifications. An aliquot from the first amplification is used as a template for the second, which may use differing primers and conditions.

ng nanogram, 10^{-9} grams.

Nonconsensus A repeat unit that differs in structure from the main repeating sequence.

NRC National Research Council.

NRC I National Research Council, Committee on DNA Technology in Forensic Science et al. (1992). *DNA Technology in Forensic Science*. Washington, DC, National Academy Press.

NRC II National Research Council and C.o.D.F. Science (1996). *The Evaluation of Forensic DNA Evidence*. Washington, DC, National Academy Press.

NRY The nonrecombining region of the Y chromosome.

Null allele The name implies an "absent" allele. However, in forensic work it refers to an allele that cannot be visualized. In multilocus or single-locus work, this is usually caused by an allele that is outside the range that can be seen on the gel. In PCR work, this is usually caused by a primer binding site mutation. Hence, these are strictly not null alleles and the term "silent" allele would seem more appropriate.

Paternal inheritance Inheritance from the father.

Paternity index A term used in paternity testing for the likelihood ratio.

PCR Polymerase chain reaction. An *in vitro* reaction that amplifies the DNA.

pg picogram, 10^{-12} grams.

Polymorphic This is used to describe a locus that has many different alleles.

Polyploid Any organism with many copies of each chromosome.

Poor shedder A term used to describe a person who deposits very little DNA onto touched objects.

Posterior odds Usually referring the odds after considering the evidence.

Primer A section of DNA that binds to the template DNA and is used to initiate the PCR reaction.

Prior odds Usually referring to the odds before considering the evidence.

Probability of paternity A term used in paternity testing for the posterior probability of paternity given prior odds of 1.

Probe A short fragment of DNA usually labeled with a radioactive or chemical tag. This allows visualization of DNA fragments after electrophoresis by binding to complementary sequences.

Product rule A model for estimating profile probabilities that assumes independence. See Equations (3.1) and (3.2).

Pseudoautosomal A section of the Y chromosome that recombines with the X chromosome.

Punnett square A method for assigning the probabilities of children conditional on their parents' genotypes.

***r* or R_c** The recombination fraction.

Recommendation 4.1 A recommendation of NRC II. See Equation (3.3).

Repeat unit A short sequence that is repeated several times at a locus.

RFLP Restriction fragment length polymorphism, a technique in which DNA is cut into fragments of differing lengths. These fragments include relatively long intronic DNA sequences and smaller flanking regions of coding DNA.

RFU Relative fluorescence units. A unit used in measuring peak height.

SGM Second-generation multiplex. A six locus multiplex, plus amelogenin.

SGM$^+$ Second-generation multiplex plus. A ten locus multiplex, plus amelogenin.

Silent allele An allele that cannot be visualized in the system used. It may be possible to visualize such an allele using different primers.

Single-locus probes This utilizes RFLP technology; however, the probes used to visualize the product were altered to visualize one locus at a time.

Singleplex A PCR system that amplifies only one locus.

SLP Single-locus probes.

Somatic cells All the cells of the body other than those associated with cellular ancestry.

Somatic mutation A mutation after gamete fusion usually leading to the situation where the individual has different genotypes in different cells.

Spurious alleles Alleles assumed to have arrived in a profile via contamination.

STR Short tandem repeat.

Stutter ratio The ratio of the area of the stutter peak to the parent allelic peak.

Stuttering A miscopying of the DNA template by the PCR reaction. Usually a stutter produces a product shorter by one repeat unit.

Subpopulation formulae A correction for the subpopulation effect. See Equations (3.4).

θ The between-person inbreeding coefficient (used synonymously with F_{ST} in this book).

Transition A mutation from one purine (A or G) to another or from one pyrimidine (C or T) to another.

Translocation A copy of the gene has been inserted somewhere else on the genome.

Transversion A mutation from a purine (A or G) to a pyrimidine (C or T) or vice versa.

Trisomy The situation where an individual has three copies of a chromosome rather than the usual pair.

Uniparental inheritance Inheritance from only one parent.

VNTR Variable number of tandem repeats. Utilized as the U.S. equivalent of SLP technology, although the name actually refers to the underlying DNA variability and hence would apply to multilocus work as well.

References

1. Abaz, J. et al., Comparison of variables affecting the recovery of DNA from common drinking containers. *Forensic Sci. Int.*, 2002; 126: 233–240.
2. Abshire, R., WACO revisited. A look back for answers. *Law Enf. Tech.*, 1997; 24: 28–32.
3. Adcock, G.J. et al., Mitochondrial DNA sequences in ancient Australians: implications for modern human origins. *Proc. Natl. Acad. Sci. USA*, 2001; 98: 537–542.
4. Ago, K. et al., Polymerase chain reaction amplification of samples from Japanese monkeys using primers for human short tandem repeat loci. *Leg. Med.*, 2003; 5: 204–206.
5. Aitken, C. and Taroni, F., Interpretation of scientific evidence. *Sci. Justice*, 1996; 36: 290–292.
6. Aitken, C.G.G., The use of statistics in forensic science. *J. Forensic Sci. Soc.*, 1987; 27: 113–115.
7. Aitken, C.G.G., Statements of probability. *J. Forensic Sci. Soc.*, 1988; 28: 329–330.
8. Aitken, C.G.G., *Statistics and the Evaluation of Evidence for Forensic Scientists*. A book in the series: *Statistics in Practice*, Barnett, V., Series Editor. 1995; John Wiley and Sons Ltd.: Chichester.
9. Aitken, C.G.G. and Gammerman, A.J., Probabilistic reasoning in evidential assessment. *J. Forensic Sci. Soc.*, 1989; 29: 303–316.
10. Aitken, C.G.G. and Stoney, D.A., *The Use of Statistics in Forensic Science*. 1991; Ellis Horwood Ltd.: Chichester.
11. Aitken, C.G.G. and Taroni, F., Interpretation of scientific evidence (Part II). *Sci. Justice*, 1997; 37: 65.
12. Aitken, C.G.G. and Taroni, F., A verbal scale for the interpretation of evidence. *Sci. Justice*, 1998; 38: 279–281.
13. Aloni, Y. and Attardi, G., Symmetrical *in vivo* transcription of mitochondrial DNA in HeLa cells. *Proc. Natl. Acad. Sci. USA*, 1971; 68: 1757–1761.
14. Alves, C. et al., VWA STR genotyping: further inconsistencies between Perkin-Elmer and Promega kits. *Int. J. Leg. Med.*, 2001; 115: 97–99.
15. Alves, C. et al., Contribution for an African autosomic STR database (AmpF*l*STR Identifiler and Powerplex 16 system) and a report of genotypic variations. *Forensic Sci. Int.*, 2004; 139: 201–205.
16. Amorim, A. et al., Somatic and germinal mutations in STRs. In *Progress in Forensic Genetics*, Vol. 8, Sensabaugh, G.F., Lincoln, P.J., and Olaisen, B., Eds. 2000; Elsevier Science: Amsterdam, p. 37–39.

17. Anderson, S. et al., Sequence and organization of the human mitochondrial genome. *Nature*, 1981; 290: 457–463.

18. Andrews, R.M. et al., Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genet.*, 1999; 23: 147.

19. Anon., DNA cold search execution. *Law & Order*, 2002; 50: 6.

20. Anslinger, K. et al., Identification of the skeletal remains of Martin Bormann by mtDNA analysis. *Int. J. Leg. Med.*, 2001; 114: 194–196.

21. Associated Press, U.S. Widens Probe of FBI's DNA Analysis. *USA Today*, 2003; Available from http://www.usatoday.com/news/nation/2003-04-27-fbi-dna-lab_x.htm (accessed 17 March 2004).

22. Asplen, C. and Lane, S.A., Considerations in the development of DNA databases. *Interfaces*, 2003; 36: 1–2.

23. *Attorney General's Reference (No. 3 of 1999)* [2000] 2 Cr. App. R. 416 and [2001] 2 Cr. App. R. 475 (HL).

24. Aulinskas, T.H. et al., No-suspect forensic casework: a one in five hit rate! In *Eleventh International Symposium on Human Identification*. 2000; Biloxi, Mississippi.

25. Australian Law Reform Commission. Protection of Human Genetic Information (Discussion Paper 66). 2002; NSW, Australia.

26. Awadalla, P. et al., Linkage disequilibrium and recombination in hominoid mitochondrial DNA. *Science*, 1999; 286: 2524–2525.

27. Ayres, K.L., Relatedness testing in subdivided populations. *Forensic Sci. Int.*, 2000; 114: 107–115.

28. Ayres, K.L., Paternal exclusion in the presence of substructure. *Forensic Sci. Int.*, 2002; 129: 142–144.

29. Ayres, K.L. et al., Implications for DNA identification arising from an analysis of Australian forensic databases. *Forensic Sci. Int.*, 2002; 129: 90–98.

30. Bacher, J.W. et al., Chromosome localization of CODIS loci and new pentanucleotide repeat loci. In *Proceedings of the Tenth International Symposium on Human Identification*. 2000; Biloxi, Missouri.

31. Bai, U. et al., Mitochondrial DNA deletions associated with aging and possibly presbycusis: a human archival temporal bone study. *Am. J. Ontol.*, 1997; 18: 449–453.

32. Baker, H., Cautions in viewing objects. The Microscope Made Easy. 1742, chap. 15. Science Heritage Ltd, Lincolnwood, Illinois. ISBN 094009533.

33. Balding, D.J., Estimating products in forensic identification using DNA profiles. *J. Am. Stat. Assoc.*, 1995; 90: 839–844.

34. Balding, D.J., When can a DNA profile be regarded as unique? *Sci. Justice*, 1999; 39: 257–260.

35. Balding, D.J. and Donnelly, P., The prosecutor's fallacy and DNA evidence. *Crim. Law Rev.*, 1994; October: 711–721.

36. Balding, D.J. and Nichols, R.A., DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Sci. Int*., 1994; 64: 125–140.

37. Balding, D.J. and Donnelly, P., Inference in forensic identification. *J. R. Stat. Soc. A*, 1995; 158: 21–53.

38. Balding, D.J. and Donnelly, P., Inferring identity from DNA profile evidence. *Proc. Natl. Acad. Sci. USA*, 1995; 92: 11741–11745.

39. Balding, D.J. and Nichols, R.A., A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica*, 1995; 96: 3–12.

40. Balding, D.J. and Donnelly, P., Evaluating DNA profile evidence when the suspect is identified through a database search. *J. Forensic Sci*., 1996; 41: 603–607.

41. Balding, D.J. and Nichols, R.A., Significant genetic correlations among Caucasians at forensic DNA loci. *Heredity*, 1997; 78: 583–589.

42. Balding, D.J. et al., Some causes for concern about DNA profiles — comment. *Stat. Sci*., 1994; 9: 248–251.

43. Balding, D.J. et al., Population genetics of STR loci in Caucasians. *Int. J. Leg. Med*., 1996; 108: 300–305.

44. Ballantyne, J., Mass disaster genetics. *Nature Genet*., 1997; 15: 329–331.

45. Balls, R., Forensic examinations helping police. Available at http://www.edp24.co.uk/content/News (accessed 8 September 2003).

46. Balogh, M.K. et al., STR genotyping and mtDNA sequencing of latent fingerprint on paper. *Forensic Sci. Int*., 2003; 137: 188–195.

47. Bär, W., DNA profiling: evaluation of the evidentiary value. *Leg. Med*., 2003; 5: 41–44.

48. Bär, W. et al., DNA recommendations. Further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. International Society for Forensic Haemogenetics [published erratum appears in *Int. J. Leg. Med*., 1999; 112: 145–150]. *Int. J. Leg. Med*., 1997; 110: 175–176.

49. Bär, W. et al., Guidelines for mitochondrial DNA typing. *Vox Sanguinis*, 2000; 79: 121–125.

50. Bär, W. et al., DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. *Forensic Sci. Int*., 2000; 110: 79–85.

51. Barbaro, A. et al., DNA typing from hair shaft. In *Progress in Forensic Genetics*, Vol. 8, Sensabaugh, G.F., Lincoln, P.J., Olaisen, B. et al., Eds. 2000; Elsevier Science: Amsterdam, p. 523–525.

52. Basten, C.J., North Carolina State University Summer Institute in Statistical Genetics Course Notes, 2003.

53. Bayes, T., An essay towards solving a problem in the doctrine of chances. *Phil. Trans*., 1763; 53: 370–418.

54. Bennett, J.H., On the theory of random mating. *Ann. Eugen*., 1954; 18: 311–317.

55. Berger, B. et al., Molecular characterization and Austrian Caucasian population data of the multi copy Y-chromosome STR DYS464. *Forensic Sci. Int.*, 2003; 137: 221–230.

56. Berry, D.A., Probability of paternity. In *The Use of Statistics in Forensic Science*, Aitken, C.G.G. and Stoney, D.A., Eds. 1991; Ellis Horwood: New York, p. 150–156.

57. Berry, D.A., Comment on Roeder, K (1994). DNA fingerprinting: a review of the controversy. *Stat. Sci.*, 1994; 9: 252–255.

58. Berry, D.A., DNA, statistics, and the Simpson case. *Chance: New Directions Stat. Comput.*, 1994; 7: 9–12.

59. Berry, D.A. and Geisser, S., Inference in cases of disputed paternity. In *Statistics and the Law*, DeGroot, M.H. et al., Eds. 1986; John Wiley and Sons: New York, p. 353–382.

60. Bertoni, B. et al., Admixture in Hispanics: distribution of ancestral population contributions in the continental United States. *Hum. Biol.*, 2003; 75: 1–11.

61. Betz, A. et al., DYS STR analysis with epithelial cells in a rape case. *Forensic Sci. Int.*, 2001; 118: 126–130.

62. Beyleveld, D., Ethics in the use of statistics in genetics. In *Handbook in Statistical Genetics*, Balding, D.J., Bishop, M., and Cannings, C., Eds. 2001; John Wiley and Sons, Ltd.: Chichester, p. 697–720.

63. Bianchi, N.O. et al., Characterisation of ancestral and derived Y-chromosome haplotypes of New World native populations. *Am. J. Hum. Genet.*, 1998; 63: 1862–1871.

64. Bini, C. et al., Different informativeness of the three hypervariable mitochondrial DNA regions in the population of Bologna (Italy). *Forensic Sci. Int.*, 2003; 135: 48–52.

65. Birky, W.C.J., Transmission genetics of mitochondria and chloroplasts. *Annu. Rev. Genet.*, 1978; 12: 471–512.

66. Birky, W.C.J., Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc. Natl. Acad. Sci. USA*, 1995; 92: 11331–11338.

67. Birky, W.C.J., The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annu. Rev. Genet.*, 2001; 35: 125–148.

68. Birky, W.C.J. et al., Mitochondrial transmission genetics: replication, recombination, and segregation of mitochondrial DNA and its inheritance in crosses. In *Mitochondrial Genes*, Attardi, G., Borst, P., and Slonimski, P.P., Eds. 1982; Cold Spring Harbor Laboratory Press: Plainview, NY, p. 333–348.

69. Birus, I. et al., How high should paternity index be for reliable identification of war victims by DNA typing? *Croat. Med. J.*, 2003; 44: 322–326.

70. Bittles, A.H., The role and significance of consanguinity as a demographic variable. *Popul. Devel. Rev.*, 1994; 20: 561–584.

71. Bittles, A.H. and Neel, J.V., The costs of human inbreeding and their implications for variations at the DNA level. *Nature Genet.*, 1994; 8: 117–121.

72. Bittles, A.H. et al., Reproductive behaviour and health in consanguineous marriages. *Science*, 1991; 252: 789–794.

73. Blakey, D., *Under the Microscope: Thematic Inspection Report on Scientific and Technical Support*. 2000; Ditching: Her Majesty's Inspectorate of Constabulary.

74. Blakey, D., *Under the Microscope Refocused. A Revisit to the Thematic Inspection Report on Scientific and Technical Support*. 2002; Ditching: Her Majesty's Inspectorate of Constabulary. p. 1–18.

75. Bobrowski, A. et al., Non-homogeneous infinite sites model under demographic change: mathematical description and asymptotic behaviour of pairwise distributions. *Math. Biosci.*, 2002; 175: 83–115.

76. Bosch, E. et al., High resolution Y chromsome typing: 19 STRs amplified in three multiplex reactions. *Forensic Sci. Int.*, 2002; 125: 42–51.

77. Botstein, D. et al., Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.*, 1980; 32: 314–331.

78. Bourn, J., Improving Service Delivery. The Forensic Science Service. Report of the Comptroller and Auditor General. HC 523 session 2002–2003, 2003.

79. Bowcock, A.M. et al., High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, 1994; 368: 455–457.

80. Brandstatter, A. and Parson, W., Mitochondrial DNA heteroplasmy or artefacts — a matter of the amplification strategy? *Int. J. Leg. Med.*, 2003; 117: 180–184.

81. Brenner, C., What's wrong with the "exclusion probability?" 1997; Available from http://www.dna-view.com/exclusn.htm (accessed 17 March 2004).

82. Brenner, C., personal communication, 2003.

83. Brenner, C. and Weir, B.S., Issues and strategies in the DNA identification of World Trade Center victims. *Theor. Popul. Biol.*, 2003; 63: 173–178.

84. Brenner, C. and Morris, J.W., Paternity Index calculations in single locus hypervariable DNA probes: validation and other studies. Unpublished.

85. Brenner, C. et al., Likelihood ratios for mixed stains when the number of donors cannot be agreed. *Int. J. Leg. Med.*, 1996; 109: 218–219.

86. Brenner, C.H., Symbolic kinship program. *Genetics*, 1997; 145: 535–542.

87. Brenner, C.H., DNA frequency uncertainty — why bother? 1997; Available from http://www.dna-view.com/noconfid.htm (accessed 17 March 2004).

88. Brenner, C.H., Kinship analysis when there are many possibilities. In *Progress in Forensic Genetics*, Vol. 7, Sensabaugh, G.F. et al., Eds. 1999; Elsevier Science: Amsterdam, p. 94–96.

89. Bright, J. and Petricevic, S.F., Recovery of trace DNA and its application to DNA profiling of shoe insoles. *Forensic Sci. Int.*, 2004; 145: 7–12.

90. Bright, J.-A., Advantages, disadvantages and the use of mitochondrial DNA in forensic science. 1998; Unpublished Research Essay, University of Auckland.

91. Brinkmann, B., Overview of PCR-based systems in identity testing. *Methods Mol. Biol.*, 1998; 98: 105–119.

92. Brinkmann, B. et al., Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am. J. Hum. Genet.*, 1998; 62: 1408–1415.

93. Briody, M., The effects of DNA evidence on sexual offence cases in court. *Curr. Iss. Crim. Justice*, 2002; 14: 159–181.

94. Broeders, A.P.A., Some observations on the use of probability scales in forensic identification. *Forensic Linguistics*, 1999; 6: 228–241.

95. Brookfield, J., DNA profiling on trial — correspondence. *Nature*, 1994; 369: 351.

96. Brown, A.H.D. and Feldman, M.W., Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics*, 1980; 96: 523–536.

97. Brownlie, A.R., Does justice require less precision than chemistry? *Sci. Justice*, 1997; 37: 73–74.

98. Buckleton, J., personal communication, 1996.

99. Buckleton, J., Lecture Notes: STR Interpretation Workshop. In *First International Conference on Forensic Human Identification in the Millennium*. 1999; London.

100. Buckleton, J. and Curran, J., personal communication, 2001.

101. Buckleton, J. et al., Interpreting evidence — evaluating forensic science in the courtroom. *Nature*, 1995; 377: 300.

102. Buckleton, J. et al., Evaluating the statistical significance of single band profiles in VNTR analyses. *Sci. Justice*, 2000; 40: 27–33.

103. Buckleton, J. et al., The fallacy of independence testing and the use of the product rule? *Sci. Justice*, 2001; 41: 81–84.

104. Buckleton, J.S., What the 90's can teach us about good forensic science. In *First International Conference on Forensic Human Identification in the Millennium*. 1999; London.

105. Buckleton, J.S., Bayes beats frequentist (letter). *The Forensic Panel*, 1999; 3: 5.

106. Buckleton, J.S. and Gill, P.D., personal communication, 1988.

107. Buckleton, J.S. and Gill, P.D., P17961 International Patent "Preferential Degradation." 2003.

108. Buckleton, J.S. and Pinchin, R.A., P17995 International Patent "Intelligence Tools." 2003.

109. Buckleton, J.S. and Triggs, C.M., The effective of linkage for relatives, in preparation.

110. Buckleton, J.S. et al., A continuous model for interpreting the positions of bands in DNA locus-specific work. *J. Forensic Sci. Soc.*, 1991; 31: 353–363.

111. Buckleton, J.S. et al., Who is "random man?" *J. Forensic Sci. Soc.*, 1991; 31: 463–468.

112. Buckleton, J.S. et al., Setting bounds for the likelihood ratio when multiple hypotheses are postulated. *Sci. Justice*, 1997; 38: 23–26.

113. Buckleton, J.S. et al., Detection of deviations from genetic equilibrium — a commentary on Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, Caucasians, Hispanics, Bahamians, Jamaicans and Trinidadians. *J. Forensic Sci.*, 1999; 44: 1277–1286; *J. Forensic Sci.*, 2001; 46: 198–202.

114. Buckleton, J.S. et al., A stratified approach to the compilation of blood group frequency surveys. *J. Forensic Sci. Soc.*, 1987; 27: 103–122.

115. Buckleton, J.S. et al., P17960 International Patent "Statistical Simulations in Mixtures." 2003.

116. Budimlija, Z.M. et al., World Trade Center human identification project: experiences with individual body identification cases. *Croat. Med. J.*, 2003; 44: 259–263.

117. Budowle, B., The effects of inbreeding on DNA profile frequency estimates using PCR-based loci. In *Human Identification: The Use of DNA Markers*, Weir, B.S., Ed. 1995; Kluwer Academic Publishers: London, p. 21–25.

118. Budowle, B., STR allele concordance between different primer sets — a brief summary. *Profiles in DNA*, 2000; 3: 10.

119. Budowle, B., Statistics and Mixture Interpretation Workshop. In *Twelfth International Symposium on Human Identification*. 2001; Biloxi, Mississippi, U.S.A.

120. Budowle, B., Population studies on 17 STR loci routinely used in forensic cases. *Int. Congress Series*, 2003; 1239: 71–74.

121. Budowle, B. and Chakraborty, R., Author's response to Buckleton et al. Detection of deviations from genetic equilibrium — a commentary on Budowle B et al. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, US Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. *J. Forensic Sci.*, 2001; 46: 201–202.

122. Budowle, B. et al., A reassessment of frequency estimates of PVUII-generated VNTR profiles in a Finnish, an Italian, and a general Caucasian database — no evidence of ethnic subgroups affecting forensic estimates. *Am. J. Hum. Genet.*, 1994; 55: 533–539.

123. Budowle, B. et al., Estimating minimum allele frequencies for DNA profile frequency estimates for PCR-based loci. *Int. J. Leg. Med.*, 1996; 108: 173–176.

124. Budowle, B. et al., Interpretation guidelines for mitochondrial DNA sequencing. In *Tenth International Symposium on Human Identification*, Promega Corporation. 1999; Madison, WI.

125. Budowle, B. et al., Critique of interpretation of high levels of heteroplasmy in the human mitochondrial region I from hair. *Forensic Sci. Int.*, 2002; 126: 30–33.

126. Budowle, B. et al., Characterization of heteroplasmy and hypervariable sites in HV1: critique of D'Eustachio's interpretations. *Forensic Sci. Int.*, 2002; 130: 68–70.

127. Budowle, B. et al., The assessment of frequency estimates of Hae III-generated VNTR profiles in various reference databases. *J. Forensic Sci.*, 1994; 39: 319–352.

128. Budowle, B. et al., The assessment of frequency estimates of Hinf I-generated VNTR profiles in various ethnic databases. *J. Forensic Sci.*, 1994; 39: 988–1008.

129. Budowle, B. et al., Source attribution of a forensic DNA profile. *Forensic Sci. Comm.*, 2000; 2; Available from http://www.fbi.gov/hq/lab/fsc/current/backissu.htm (accessed 17 March 2004).

130. Budowle, B. et al., *DNA Typing Protocols: Molecular Biology and Forensic Analysis*. BioTechniques. 2000; Eaton Publishing: Natick, MA.

131. Budowle, B. et al., Reply to Weir. *Forensic Sci. Comm.*, 2001; 3; Available from http://www.fbi.gov/hq/lab/fsc/current/backissu.htm (accessed 17 March 2004).

132. Budowle, B. et al., Low copy number — consideration and caution. In *Twelfth International Symposium on Human Identification*. 2001; Biloxi, Missouri.

133. Budowle, B. et al., Population data on the thirteen CODIS core short tandem repeat loci in African Americans, US Caucasians, Hispanics, Bahamanians, Jamaicans and Trinidadians. *J. Forensic Sci.*, 1999; 44: 1277–1286.

134. Budowle, B. et al., Population studies on three Native Alaskan population groups using STR loci. *Forensic Sci. Int.*, 2002; 129: 51–57.

135. Budowle, B. et al., Fixed bin analysis for statistical evaluation of continuous distributions of allelic data from VNTR loci. *Am. J. Hum. Genet.*, 1991; 48: 841–855.

136. Budowle, B. et al., STR primer concordance study. *Forensic Sci. Int.*, 2001; 124: 47–54.

137. Butler, J.M., *Forensic DNA Typing: Biology and Technology Behind STR Markers*. 2001; Academic Press: San Diego, CA.

138. Butler, J.M. and Reeder, D.J., STRBase: Short Tandem Repeat DNA Internet DataBase (http://www.cstl.nist.gov/biotech/strbase).

139. Butler, J.M. et al., The development of reduced size STR amplicons as tools for analysis of degraded DNA. *J. Forensic Sci.*, 2003; 48: 1054–1064.

140. Butler, J.M. et al., A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. *Forensic Sci. Int.*, 2002; 129: 10–24.

141. Caesar, G.J., *Commentaries on the Gallic War*. 1918 Edition, Dewey, F.H., Ed. Translation Publishing Co., New York.

142. Caglia, A. et al., Increased forensic efficiency of a STR based Y specific haplotype by addition of the highly polymorphic DYS385 locus. *Int. J. Leg. Med.*, 1998; 111: 142–146.

143. Calacal, G.C. et al., Identification of two fire victims by comparative nuclear DNA typing of skeletal remains and stored umbilical tissues. *Am. J. Forensic Med. Pathol.*, 2003; 24: 148–152.

144. Calloway, C.D. et al., The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. *Am. J. Hum. Genet.*, 2000; 66: 1384–1397.

145. Cantatore, P. and Attardi, G., Mapping of nascent light and heavy strand transcripts on the physical map of HeLa cell mitochondrial DNA. *Nucl. Acids Res.*, 1980; 8: 2605–2625.

146. Capelli, C. et al., "Ancient" protocols for the crime scene? Similarities and differences between forensic genetics and ancient DNA analysis. *Forensic Sci. Int.*, 2003; 131: 59–64.

147. Capelli, C. et al., A Y chromosome census of the British Isles. *Curr. Biol.*, 2003; 13: 979–984.

148. Carracedo, A. et al., Focusing the debate on forensic genetics. *Sci. Justice*, 1996; 36: 204–206.

149. Carroll, S.B., Genetics and the making of *Homo sapiens*. *Nature*, 2003; 422: 849–857.

150. Carvajal-Carmona, L.G. et al., Strong Amerind/White sex bias and a possible sephardic contribution among the founders of a population in Northwest Colombia. *Am. J. Hum. Genet.*, 2000; 67: 1287–1295.

151. Castle, W.E., The law of Heredity of Galton and Mendel and some laws governing race improvement by selection. *Proc. Natl. Acad. Sci. USA*, 1903; 39: 233–242.

152. Cavalli-Sforza, L.L. et al., *The History and Geography of Human Genes*. 1994; Princeton University Press: Princeton, NJ.

153. Centre for Molecular Medicine, Emory University, Atlanta, GA. MITOMAP: Human Mitochondrial Genome Database l. 1998; Available at http://www.gen.emory.edu/mitomap.htm (accessed 17 March 2004).

154. Cerda-Flores, R.M. et al., Maximum likelihood estimates of admixture in Northeastern Mexico using 13 short tandem repeat loci. *Am. J. Hum. Biol.*, 2002; 14: 429–439.

155. Cerda-Flores, R.M. et al., Genetic admixture in three Mexican Mestizo populations based on D1S80 and HLA-DQA1 loci. *Am. J. Hum. Biol.*, 2002; 14: 257–263.

156. Cerri, N. et al., Mixed stains from sexual assault cases: autosomal or Y-chromosome short tandem repeats? *Croat. Med. J.*, 2003; 44: 289–292.

157. Chakraborty, R., Detection of nonrandom association of alleles from the distribution of the number of heterozygous loci in a sample. *Genetics*, 1984; 108: 719–731.

158. Chakraborty, R., Sample size requirements for addressing the population genetic issues of forensic use of DNA typing. *Hum. Biol.*, 1992; 64: 141–160.

159. Chakraborty, R., DNA profiling on trial — correspondence. *Nature*, 1994; 369: 351.

160. Chakraborty, R. and Smouse, P.E., Recombination of haplotypes leads to biased estimates of admixture proportions in human populations. *Proc. Natl. Acad. Sci. USA*, 1988; 85: 3071–3074.

161. Chakraborty, R. and Kidd, K.K., The utility of DNA typing in forensic work. *Science*, 1991; 254: 1735–1739.

162. Chakraborty, R. and Jin, L., Determination of relatedness between individuals using DNA fingerprinting. *Hum. Biol.*, 1993; 65: 875–895.

163. Chakraborty, R. and Stivers, D.N., Paternity exclusion by DNA markers: effects of paternal mutations. *J. Forensic Sci.*, 1996; 41: 671–677.

164. Chakraborty, R. et al., Population amalgamation and genetic variation: observations on artificially agglomerated tribal populations of central and South America. *Am. J. Hum. Genet.*, 1988; 43: 709–725.

165. Chakraborty, R. et al., Evaluation of standard errors and confidence intervals of estimated multilocus genotype probabilities and their implications in DNA. *Am. J. Hum. Genet.*, 1993; 52: 60–70.

166. Chakraborty, R. et al., Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proc. Natl. Acad. Sci. USA*, 1997; 94: 1041–1046.

167. Chambers, G.K. et al., Forensic DNA profiling: the importance of giving accurate answers to the right questions. *Crim. Law Forum*, 1997; 8: 445–459.

168. Champod, C., Identification/individualisation. In *Encyclopedia of Forensic Sciences*, Seigal, J. et al., Eds. 2000; Academic Press: New York, p. 1077–1084.

169. Champod, C., personal communication, 2003.

170. Champod, C. and Taroni, F., Bayesian framework for the evaluation of fibre transfer evidence. *Sci. Justice*, 1997; 37: 75–83.

171. Champod, C. and Taroni, F., Interpretation of fibres evidence — the Bayesian approach. In *Forensic Examination of Fibres*, Robertson, J., Ed. 1999; Taylor and Francis, Ltd.: London, p. 379–398.

172. Champod, C. and Ribaux, O., Forensic identification databases in criminal investigations and trials. In *Harmonisation in Forensic Expertise*, Sprangers, W.J.J.M. and Nijboer, J.F., Eds. 2000; Thela Thesis: Amsterdam, p. 463–485.

173. Champod, C. and Evett, I.W., A probabilistic approach to fingerprint evidence. *J. Forensic Ident.*, 2001; 51: 101–122.

174. Champod, C. et al., Comments on the scale of conclusions proposed by the ad hoc Committee of the ENFSI Marks Working Group. *Inform. Bull. Shoeprint/Toolmark Examiners*, 2000; 6: 11–18.

175. Chang, D.D. and Clayton, D.A., Precise identification of individual promotors for transcription of each strand of human mitochondrial DNA. *Cell*, 1984; 36: 635–643.

176. Chang, Y.M. et al., Higher failures of Amelogenin sex test in an Indian population group. *J. Forensic Sci.*, 2003; 48: 1309–1313.

177. Chapman, R., personal communication, 2001.

178. Chow, S.L., Precis of statistical significance: rationale, validity, and utility. *Behav. Brain Sci.*, 1998; 21: 169–239.

179. Clarke, G.W., Commentary on: "How the probability of a false positive affects the value of DNA evidence." *J. Forensic Sci.*, 2003; 48: 1199.

180. Clayton, T., The WACO disaster, USA 1993. *Contact*, 1997; 25: 14–16.

181. Clayton, T., personal communication, 2003.

182. Clayton, T.M. et al., Analysis and interpretation of mixed forensic stains using DNA STR profiling. *Forensic Sci. Int.*, 1998; 91: 55–70.

183. Clayton, T. et al., The incidence of somatic and germline mutations in human autosomal tetrameric microsatellite loci, personal communication, 2001.

184. Clayton, T.M. et al., Identification of bodies from the scene of a mass disaster using DNA amplification of short tandem repeat (STR) loci. *Forensic Sci. Int.*, 1995; 76: 7–15.

185. Clayton, T.M. et al., Primer binding site mutations affecting the typing of STR loci contained within the AMPF*l*STR® SGM Plus™ kit. *Forensic Sci. Int.*, 2004; 139: 255–259.

186. Clisson, I. et al., Genetic analysis of human remains from a double inhumation in a frozen kurgan in Kazakhstan (Berel site, early 3rd Century BC). *Int. J. Leg. Med.*, 2002; 116: 304–308.

187. Cockerham, C.C., Higher order probability functions of alleles by descent. *Genetics*, 1971; 69: 235–247.

188. Cockerham, C.C. and Weir, B.S., Sib mating with two linked loci. *Genetics*, 1968; 60: 629–640.

189. Coffman, D., Florida Department of Law Enforcement's convicted offender DNA database: transition from RFLP to STR. In *Ninth International Symposium on Human Identification*. 1998; Promega Corporation, Madison, WI.

190. Coleman, D., A note on the frequency of consanguineous marriages in Reading, England in 1972/1973. *Hum. Hered.*, 1980; 30: 278–285.

191. *Commonwealth of Virginia v Spencer*. 384 SE 2d 253. 1989; W VA.

192. Connor, S., Every Baby's DNA may be Stored for Future Health Care. *The Independent*, 25 June 2003.

193. Cook, R. et al., A hierarchy of propositions: deciding which level to address in casework. *Sci. Justice*, 1998; 38: 231–240.

194. Cook, R. et al., A model for case assessment and interpretation. *Sci. Justice*, 1998; 38: 151–156.

195. Cooper, G. et al., Markov chain Monte Carlo analysis of human Y chromosome microsatellite provides evidence of biased mutation. *Proc. Natl. Acad. Sci. USA*, 1999; 96: 11916–11921.

196. Correction (Editorial). *Nature Genet.*, 1994; 8: 309.

197. Cotterill, J., "If it doesn't fit, you must acquit": metaphor and the O.J. Simpson criminal trial. *Forensic Linguistics*, 1998; 5: 141–158.

198. Cotterman, C.W., *A Calculus for Statistical Genetics*. 1940; Ph.D. thesis. Columbus, OH: Ohio State University.

199. Cotton, E.A. et al., Validation of the AmpF*l*STR SGM plus system for use in forensic casework. *Forensic Sci. Int.*, 2000; 112: 151–161.

200. Cotton, R.W. and Word, C.J., Commentary on: "How the probability of a false positive affects the value of DNA evidence." *J. Forensic Sci.*, 2003; 48: 1200.

201. Court of Appeal, Regina v Doheny and Adams. 1996; Royal Courts of Justice, The Strand, London.

202. Cowell, R.G. and Mostad, P.F., A clustering algorithm using DNA marker information for sub-pedigree reconstruction. *J. Forensic Sci.*, 2003; 48: 1239–1248.

203. Cox, M., A multidisciplinary approach to the investigation of crimes against humanity, war crimes and genocide: the Inforce Foundation. *Sci. Justice*, 2003; 43: 225–227.

204. Coyle, H.M. et al., An overview of DNA methods for the identification and individualization of marijuana. *Croat. Med. J.*, 2003; 44: 315–321.

205. Crews, S. et al., Nucleotide sequence of a region of human mitochondrial DNA containing the precisely identified origin of replication. *Nature*, 1979; 277: 192–198.

206. Crispino, F., Comment on JFI 51(3). *J. Forensic Ident.*, 2001; 51: 449–456.

207. Crouse, C.A. et al., Analysis and interpretation of short tandem repeat microvariants and three-banded allele patterns using multiple detection systems [published erratum appears in *J. Forensic Sci.*, 1999; 44]. *J. Forensic Sci.*, 1999; 44: 87–94.

208. Crow, J.F., *Basic Concepts in Population, Quantitative, and Evolutionary Genetics*. 1986; W.H. Freeman and Company: New York.

209. Crow, J.F., A new study challenges the current belief of a high human male:female mutation ratio. *Trends Genet.*, 2000; 16: 525–526.

210. Crow, J.F., There's something curious about paternal-age effects. *Science*, 2003; 301: 606–607.

211. Crow, J.F. and Kimura, M., *An Introduction to Population Genetics Theory*. 1970; Harper and Row: London.

212. Cummins, J.M. et al., Fate of microinjected spermatid mitochondria in mouse oocyte and embryo. *Zygote*, 1998; 6: 213–222.

213. Curran, J.M., SPURS II (Statistical Probability Using Re-Sampling), software.

214. Curran, J.M. et al., *Forensic Interpretation of Glass Evidence*. 2000; CRC Press LLC: Boca Raton, FL.

215. Curran, J.M. et al., What is the magnitude of the subpopulation effect? *Forensic Sci. Int.*, 2003; 135: 1–8.

216. Curran, J.M. et al., Interpreting DNA mixtures in structured populations. *J. Forensic Sci.*, 1999; 44: 987–995.

217. Curran, J. et al., Assessing uncertainty in DNA evidence caused by sampling effects. *Sci. Justice*, 2002; 42: 29–37.

218. Dauber, E.M. et al., Mutational events at human micro- and minisatellite loci: mutation rates and new STR-alleles. In *Progress in Forensic Genetics*, Vol. 8, Sensabaugh, G.F., Lincoln, P.J., and Olaisen, B., Eds. 2000; Elsevier Science: Amsterdam, p. 21–23.

219. Dawid, A.P., Comment on Stockmarr's "Likelihood ratios for evaluating DNA evidence when the suspect is found through a database search." *Biometrics*, 2001; 57: 976–980.

220. Dawid, A.P., Which likelihood ratio? *Law Prob. Risk*, 2004; 3(1): 65–71.

221. Dawid, A.P. and Mortera, J., Coherent analysis of forensic identification evidence. *J. R. Stat. Soc. B*, 1996; 52: 425–443.

222. Dawid, A.P. et al., Non-fatherhood or mutation? A probabilistic approach to parental exclusion in paternity testing. *Forensic Sci. Int.*, 2001; 124: 55–61.

223. Dawid, A.P. et al., Probabilistic expert systems for forensic inference from genetic markers. *Scand. J. Stat.*, 2002; 29: 577–595.

224. de Knijff, P. et al., Chromosome Y microsatellites: population genetic and evolutionary aspects. *Int. J. Leg. Med.*, 1997; 110: 134–140.

225. Debenham, P.G., Genetics leaves no bones unturned. *Nature Genet.*, 1994; 6: 113–114.

226. Debenham, P.G., Heteroplasmy and the Tsar. *Nature*, 1996; 380: 484–485.

227. Deka, R. et al., Linkage disequilibrium in four human populations on a 100 kb DNA fragment on chromosome 21: implications for mapping complex traits. *Hum. Biol. Assoc. Abstr.*, p. 13.

228. D'Eustachio, P., High levels of mitochondrial DNA heteroplasmy in human hairs by Budowle et al. *Forensic Sci. Int.*, 2002; 130: 63–67.

229. Devlin, B., The evidentiary value of a DNA database search. *Biometrics*, 2000; 56: 1276.

230. Devlin, B. et al., NRC report on DNA typing. *Science*, 1993; 260: 1057–1058.

231. Diamond, J., *The Rise and Fall of the Third Chimpanzee*. 1991; Vintage: London.

232. Dimsoski, P., Development of a 17-plex microsatellite polymerase chain reaction kit for genotyping horses. *Croat. Med. J.*, 2003; 44: 332–335.

233. DNA Resource website. Available at http://www.dnaresource.com/bill_tracking_list.htm (accessed 6 February 2004).

234. Doda, J.N. et al., Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. *Proc. Natl. Acad. Sci. USA*, 1981; 78: 6116–6120.

235. Donnelly, P., Match probability calculations for multi-locus DNA profiles. In *Human Identification: The Use of DNA Markers*, Weir, B.S., Ed. 1995; Kluwer Academic Publishers: London.

236. Donnelly, P. and Friedman, R.D., DNA database searches and the legal consumption of scientific evidence. *Mich. Law Rev.*, 1999; 97: 931–984.

237. Duhig, C., Non-forensic remains: the use of forensic archaeology, anthropology and burial taphonomy. *Sci. Justice*, 2003; 43: 211–214.

238. Dupuy, B.M. et al., Y-chromosome variation in a Norwegian population sample. *Forensic Sci. Int.*, 2001; 117: 163–173.

239. Eddy, S.R. and MacKay, D.J.C., Is the Pope the Pope? *Nature*, 1996; 382: 490.

240. Editorial, Anastasia and the tools of justice [published erratum appears in *Nature Genet.* 1994; 8: 309]. *Nature Genet.*, 1994; 8: 205–206.

241. Edwards, A. et al., DNA typing and genetic mapping at five trimeric and tetrameric tandem repeats. *Am. J. Hum. Genet.*, 1991; 49: 746–756.

242. Edwards, J.H., More on DNA typing dispute. *Nature*, 1995; 373: 98.

243. Egeland, T. and Mostad, P.F., Statistical genetics and genetical statistics: a forensic perspective. *Scand. J. Stat.*, 2002; 29: 297–307.

244. Egeland, T. et al., A computerized method for calculating the probability of pedigrees from genetic data. *Sci. Justice*, 1997; 37: 269–274.

245. Egeland, T. et al., Estimating the number of contributors to a DNA profile. *Int. J. Leg. Med.*, 2003; 117: 271–275.

246. Egeland, T. et al., Beyond traditional paternity and identification cases. Selecting the most probable pedigree. *Forensic Sci. Int.*, 2000; 110: 47–59.

247. Ehm, M.G. and Wagner, M., A test statistic to detect errors in sib-pair relationships. *Am. J. Hum. Genet.*, 1998; 62: 181–188.

248. Elder, J.K. and Southern, E.M., Measurement of DNA length by gel electrophoresis II: comparison of methods for relating mobility to fragment length. *Anal. Biochem.*, 1983; 128: 227–231.

249. Elder, J.K. et al., Measurement of DNA length by gel electrophoresis. I. Improved accuracy of mobility measurements using a digital microdensitometer and computer processing. *Anal. Biochem.*, 1983; 128: 223–226.

250. Ellegren, H., Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet.*, 2000; 16: 551–558.

251. Ellegren, H., Heterogeneous mutation processes in human microsatellite DNA sequences. *Nature Genet.*, 2000; 24: 400–402.

252. Enserink, M., Opponents criticize Iceland's database. *Science*, 1998; 282: 859.

253. Eriksen, B. and Svensmark, O., The effect of sample size on the estimation of the frequency of DNA-profiles in RFLP-analysis. *Forensic Sci. Int.*, 1994; 65: 195–205.

254. Erratum, *Int. J. Leg., Med.*, 1999; 112: 145–150.

255. Essen-Moller, E., Die Beweiskraft der Ahnlichkeit im Vatershaftsnachweis. Theoretische Grundlagen. *Mitt. Anthropol. Gesellschaft.*, 1938; 68: 9–53.

256. Evett, I.W., personal communication, 2001.

257. Evett, I.W., What is the probability that this blood came from that person? A meaningful question. *J. Forensic Sci. Soc.*, 1983; 23: 35–39.

258. Evett, I.W., A discussion of the deficiencies of the coincidence method for evaluating evidential value and a look towards the future. In *Tenth International Association of Forensic Sciences Triennial Meeting*. 1984; Oxford, U.K.

259. Evett, I.W., On meaningful questions: a two-trace transfer problem. *J. Forensic Sci.* Soc., 1987; 27: 375–381.

260. Evett, I.W., Evaluating DNA profiles in a case where the defence is "it was my brother." *J. Forensic Sci. Soc.*, 1992; 32: 5–14.

261. Evett, I.W., Avoiding the transposed conditional. *Sci. Justice*, 1995; 35: 127–131.

262. Evett, I.W., A Discussion of Recommendation 5.1 of the Second NRC Report on DNA Profiling. FSS Report No. TN 821, 1996.

263. Evett, I.W., Verbal conventions for handwriting opinions. *J. Forensic Sci.*, 2000; 45: 508–509.

264. Evett, I.W. and Buckleton, J.S., Some aspects of the Bayesian approach to evidence evaluation. *J. Forensic Sci. Soc.*, 1989; 29: 317–324.

265. Evett, I.W. and Weir, B.S., Flawed reasoning in court. *Chance: New Directions Stat. Comput.*, 1991; 4: 19–21.

266. Evett, I.W. and Buckleton, J.S., Statistical analysis of STR data. In *Advances in Forensic Haemogenetics*, Bär, W., Ed. 1996; Springer-Verlag: Berlin, p. 79–86.

267. Evett, I.W. and Weir, B.S., *Interpreting DNA Evidence — Statistical Genetics for Forensic Scientists.* 1998; Sinauer Associates, Inc.: Sunderland.

268. Evett, I.W. et al., Probabilistic analysis of DNA profiles. *J. Forensic Sci. Soc.*, 1989; 29: 191–196.

269. Evett, I.W. et al., Paternity calculations from DNA multilocus profiles. *J. Forensic Sci. Soc.*, 1989; 29: 249–254.

270. Evett, I.W. et al., An efficient statistical procedure for interpreting DNA single locus profiling data in crime cases. *J. Forensic Sci. Soc.*, 1992; 32: 307–326.

271. Evett, I.W. et al., Taking account of peak areas when interpreting mixed DNA profiles. *J. Forensic Sci.*, 1998; 43: 62–69.

272. Evett, I.W. et al., More on the hierarchy of propositions: exploring the distinction between explanations and propositions. *Sci. Justice*, 2000; 40: 3–10.

273. Evett, I.W. et al., Letter to the Editor. *Biometrics*, 2000; 56: 1274–1277 (with correspondence from Stockmarr and Delvin).

274. Evett, I.W. et al., Use of Monte Carlo simulation to calculate likelihood ratios for mixed DNA profiles. Unpublished results, 2001.

275. Evett, I.W. et al., DNA fingerprinting on trial. *Nature*, 1989; 340: 435.

276. Evett, I.W. et al., Bayesian analysis of single locus DNA profiles. In *International Symposium on Human Identification*. 1990; Promega Corporation, Madison, WI.

277. Evett, I.W. et al., A guide to interpreting single locus profiles of DNA mixtures in forensic cases. *J. Forensic Sci. Soc.*, 1991; 31: 41–47.

278. Evett, I.W. et al., The evidential value of DNA profiles. *J. Forensic Sci. Soc.*, 1993; 33: 243–244.

279. Evett, I.W. et al., Statistical analysis of a large file of STR profiles of British Caucasians to support forensic casework. *Int. J. Leg. Med.*, 1996; 109: 173–177.

280. Evett, I.W. et al., DNA profiling: a discussion of issues relating to the reporting of very small match probabilities. *Crim. Law Rev.*, 2000; May: 341–355.

281. Evett, I.W. et al., The impact of the principles of evidence interpretation on the structure and content of statements. *Sci. Justice*, 2000; 40: 233–239.

282. Evison, M.P. et al., Extraction of single-copy nuclear DNA from forensic specimens with a variety of postmortem histories. *J. Forensic Sci.*, 1997; 46: 1032–1038.

283. Excoffier, L., Analysis of population subdivision. In *Handbook of Statistical Genetics*, Cannings, C., Ed. 2001; John Wiley and Sons: Chichester.

284. Excoffier, L. and Yang, Z., Substitution rate variation among sites in the mitochondrial DNA hypervariable region I of humans and chimpanzees. *Mol. Biol. Evol.*, 1999; 16: 1357–1368.

285. Excoffier, L. and Hamilton, G., Comment on "Genetic structure of human populations." *Science*, 2003; 300: 1877.

286. Eyre-Walker, A. et al., How clonal are human mitochondria? *Proc. R. Soc. Lond. B*, 1999; 266: 477–483.

287. Fairley, W.B. and Mosteller, F., A conversation about Collins. *Univ. Chicago Law Rev.*, 1974; 41: 242–253.

288. Fenton, N.E. and Neil, M., The jury observation fallacy and the use of Bayesian networks to present probabilistic legal arguments. *Math. Today*, 2000; 36: 180–187.

289. Ferllini, R., The development of human rights investigations since 1945. *Sci. Justice*, 2003; 43: 219–224.

290. Fernandes, A. and Brehm, A., Y-chromosome STR haplotypes in the Açores Archipelago (Portugal). *Forensic Sci. Int.*, 2003; 135: 239–242.

291. Fienberg, S.E. et al., Understanding and evaluating statistical evidence in litigation. *Jurimetrics*, 1995; 36: 1–32.

292. Findlay, I., Single cell PCR. In *Methods in Molecular Medicine*, Lo, Y.M.D., Ed. 1998; Humana Press: Totowa, NJ, p. 233–263.

293. Findlay, I. and Quirke, P., Fluorescent polymerase chain reaction: part 1. A new method allowing genetic diagnosis and DNA fingerprinting of single cells. *Hum. Reprod. Update*, 1996; 2: 137–152.

294. Findlay, I. et al., DNA fingerprinting from single cells. *Nature*, 1997; 389: 555–556.

295. Findlay, I. et al., High throughput genetic diagnosis of single and small numbers of cells. *Today's Life Science*, 2001; Sept/Oct: 40–46.

296. Findlay, I. et al., Genetic diagnosis of single cells: multiplex PCR using seven fluorescent primers to diagnose sex and DNA fingerprint. In *Biotechnology European Symposium*. 1994; Miami.

297. Findlay, I. et al., Allelic dropout and preferential amplification in single cells and human blastomeres: implications for preimplantation diagnosis of sex and cystic fibrosis. *Mol. Hum. Reprod.*, 1995; 10: 1609–1618.

298. Findlay, I. et al., Simultaneous DNA "fingerprinting," diagnosis of sex and single-gene defect status from single cells. *Mol. Hum. Reprod.*, 1995; 10: 1005–1013.

299. Findlay, M. and Grix, J., Challenging forensic evidence? Observations on the use of DNA in certain criminal trials. *Curr. Iss. Crim. Justice*, 2003; 14: 269–282.

300. Finkelstein, M.O. and Levin, B., *Statistics for Lawyers*. 2001; Springer-Verlag: New York.

301. Fisher, D.L. et al., Extraction, evaluation, and amplification of DNA from decalcified and undecalcified United States Civil War bone. *J. Forensic Sci.*, 1993; 38: 60–68.

302. Fisher, R.A., *Theory of Inbreeding*. 1949; Oliver and Boyd: Edinburgh.

303. Fisher, R.A., *Statistical Methods for Research Workers*. 1950; Oliver and Boyd: London.

304. Fisher, R.A., Standard calculations for evaluating a blood-group system. *Heredity*, 1951; 5: 95–102.

305. Fisher, R.A., On a significance test in Pearson's Biometrika tables (no. 11). *J. R. Stat. Soc. B*, 1956; 18: 56–60.

306. Fisher, R.A., *Statistical Methods and Scientific Inference*. 1973; Collins Macmillan: London.

307. Fisher, R.P. and Clayton, D.A., A transcription factor required for promotor recognition by human mitochondrial RNA polymerase. *J. Biol. Chem.*, 1985; 260: 11330–11338.

308. Fisher, R.P. et al., Promotor selection in human mitochondria involves binding of a transcription factor to orientation-dependent upstream regulatory elements. *Cell*, 1987; 50: 247–258.

309. FLINTS — Linking police work with science and technology. Available at http://www.west-midlands.police.uk.flints/ (accessed 24 June 2003).

310. Foreman, L.A. and Lambert, J.A., Genetic differentiation within and between four UK ethnic groups. *Forensic Sci. Int.*, 2000; 114: 7–20.

311. Foreman, L.A. and Evett, I.W., Statistical analysis to support forensic interpretation of a new ten-locus STR profiling system. *Int. J. Leg. Med.*, 2001; 114: 147–155.

312. Foreman, L.A. et al., Bayesian analysis of DNA profiling data in forensic identification applications. *J. R. Stat. Soc. A*, 1997; 160: 429–469.

313. Foreman, L.A. et al., Rationale behind QMS guidelines (FSS-TS-404) for match probability calculations. FSS Report No. TN 823, 1997.

314. Foreman, L.A. et al., Regional genetic variation in Caucasians. *Forensic Sci. Int.*, 1998; 95: 27–37.

315. Foreman, L.A. et al., Bayesian validation of a quadruplex STR profiling system for identification purposes. *J. Forensic Sci.*, 1999; 44: 478–486.

316. Forrest, A.R.W., Sally Clark — a lesson for us all. *Sci. Justice*, 2003; 43: 63–64.

317. Foster, E.A. et al., Jefferson fathered slave's last child. *Nature*, 1998; 396: 27–29.

318. Frank, L., Estonia prepares for national DNA database. *Science*, 2000; 290: 31.

319. Freckelton, I., Expert evidence. In *Expert Evidence*, Selby, H., Ed. 2001; Law Book Company: North Ryde, NSW, p. 3600–3864.

320. Frégeau, C.J. et al., National casework and national DNA database: the Royal Canadian Mounted Police perspective. *Prog. Forensic Genet.*, 1998; 7: 541–543.

321. Frégeau, C.J. et al., The National DNA Data Bank of Canada — a laboratory bench retrospective on the first year of operation. *Int. Congress Series*, 2003; 1239: 621–625.

322. Frégeau, C.J. et al., AmpF*l*STR Profiler Plus short tandem repeat analysis of casework samples, mixture samples, and nonhuman DNA samples amplified under reduced PCR volume conditions (25 ml). *J. Forensic Sci.*, 2003; 48: 1014–1034.

323. Fung, W.K., Are convenience DNA samples significantly different? *Forensic Sci. Int.*, 1996; 82: 233–241.

324. Fung, W.K., 10% or 5% match window in DNA profiling. *Forensic Sci. Int.*, 1996; 78: 118.

325. Fung, W.K. et al., Power of exclusion revisited: probability of excluding relatives of the true father from paternity. *Int. J. Leg. Med.*, 2002; 116: 64–67.

326. Fung, W.K. et al., Testing for kinship in a subdivided population. *Forensic Sci. Int.*, 2003; 135: 105–109.

327. Gabriel, M.N. et al., Identification of human remains by immobilized sequence-specific oligonucleotide probe analysis of mtDNA hypervariable regions I and II. *Croat. Med. J.*, 2003; 44: 293–298.

328. Gans, J., Something to hide: DNA, surveillance and self-incrimination. *Curr. Iss. Crim. Justice*, 2001; 13: 168–184.

329. Gans, J. and Urbas, G., DNA identification in the criminal justice system. *Trends Iss. Crime Crim. Justice*, 2002; 226: 1–6.

330. Ganswindt, M. et al., Bone finds: a challenge to forensic science. *Leg. Med.*, 2003; 5: 382–385.

331. Geiringer, H., Further remarks on linkage theory in Mendelian Heredity. *Ann. Math. Stat.*, 1945; 16: 390–393.

332. Geneva Convention. 1949; Articles 130 and 136.

333. GeNIe & S.M.I.L.E.'s website, Decision Systems Laboratory, University Of Pittsburgh. Available at http://www2.sis.pitt.edu/~genie (accessed 6 February 2004).

334. George, K.H., DNA profiling: what should the jury be told? *J. NIH Res.*, 1996; 8: 24–26.

335. Gibbons, A., Calibrating the mitochondrial clock. *Science*, 1998; 279: 28–29.

336. Gigerenzer, G., *Reckoning with Risk*. 2002; Penguin Books: London.

337. Gill, P., The Romanovs. *Medico-Leg. J.*, 1997; 65: 122–132.

338. Gill, P., Application of low copy number DNA profiling. *Croat. Med. J.*, 2001; 42: 229–232.

339. Gill, P., Role of short tandem repeat DNA in forensic casework in the UK — past, present and future perspectives. *BioTechniques*, 2002; 32: 366–385.

340. Gill, P. and Buckleton, J., Floating bin analysis and the "Gillogram." Unpublished, 1988.

341. Gill, P. et al., Forensic application of DNA "fingerprints." *Nature*, 1985; 318: 577–579.

342. Gill, P. et al., The analysis of hypervariable DNA profiles: problems associated with the objective determination of the probability of a match. *Hum. Genet.*, 1990; 85: 75–79.

343. Gill, P. et al., Development of guidelines to designate alleles using an STR multiplex system. *Forensic Sci. Int.*, 1997; 89: 185–197.

344. Gill, P. et al., Interpretation of simple mixtures when artefacts such as stutters are present — with special reference to multiplex STRs used by the Forensic Science Service. *Forensic Sci. Int.*, 1998; 95: 213–224.

345. Gill, P. et al., Few cell DNA profiling. International Patent Number WO 01/79541 A2 International, U.S., European and Australian Patent.

346. Gill, P. et al., Establishing the identity of Anna Anderson Manahan. *Nature Genet.*, 1995; 9: 9–10.

347. Gill, P. et al., Report of the European Network of Forensic Science Institutes (ENSFI): formulation and testing of principles to evaluate STR multiplexes. *Forensic Sci. Int.*, 2000; 108: 1–29.

348. Gill, P. et al., An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. *Forensic Sci. Int.*, 2000; 112: 17–40.

349. Gill, P. et al., A new method of STR interpretation using inferential logic — development of a criminal intelligence database. *Int. J. Leg. Med.*, 1996; 109: 14–22.

350. Gill, P. et al., Interpreting simple STR mixtures using allele peak areas. *Forensic Sci. Int.*, 1998; 91: 41–53.

351. Gill, P. et al., Identification of the remains of the Romanov family by DNA analysis. *Nature Genet.*, 1994; 6: 130–135.

352. Gill, P. et al., Report of the European DNA profiling group (EDNAP): an investigation of the complex STR loci D21S11 and HUMFIBRA (FGA). *Forensic Sci. Int.*, 1997; 18: 25–33.

353. Gill, P. et al., Considerations from the European DNA profiling group (EDNAP) concerning STR nomenclature. *Forensic Sci. Int.*, 1997; 87: 185–192.

354. Gill, P.D. and Evett, I.W., Population genetics of short tandem repeat (STR) loci. *Genetics*, 1995; 96: 69–87.

355. Gill, P.D. et al., A comparison of adjustment methods to test the robustness of an STR DNA database comprised of 24 European populations. *Forensic Sci. Int.*, 2003; 131: 184–196.

356. Gilmore, S. et al., Short tandem repeat (STR) DNA markers are hypervariable and informative in *Cannabis sativa*: implications for forensic investigators. *Forensic Sci. Int.*, 2003; 131: 65–74.

357. Girod, A., Computerized classification of the shoeprints of burglars' soles. *Forensic Sci. Int.*, 1996; 82: 59–65.

358. Golding, G.B., Estimates of DNA and protein sequence divergence: an examination of some assumptions. *Mol. Biol. Evol.*, 1983; 1: 125–142.

359. Goldman, D. and Long, J., More on DNA typing dispute. Correspondence. *Nature*, 1995; 373: 99.

360. Good, I.J., When batterer becomes murderer. *Nature*, 1995; 375: 541.

361. Good, I.J., When batterer becomes murderer. *Nature*, 1996; 381: 481.

362. Good, P.I., *Applying Statistics in the Courtroom*. 2001; Chapman & Hall/CRC: London.

363. Goodwin, W. et al., Mitochondrial DNA analysis of individuals from a mass grave. In *Progress in Forensic Genetics*, Vol. 7, Olaisen, B., Brinkmann, B., and Lincoln, P.J., Eds. 1998; Elsevier Science BV: The Netherlands, p. 80–83.

364. Goodwin, W. et al., The use of mitochondrial DNA and short tandem repeat typing in the identification of air crash victims. *Electrophoresis*, 1999; 20: 1707–1711.

365. Goos, L.M. et al., The influence of probabilistic statements on the evaluation of a DNA match. *J. Can. Soc. Forensic Sci*., 2002; 35: 77–90.

366. Gordon, D. et al., Significant evidence for linkage disequilibrium over a 5-cM region among Afrikaners. *Genomics*, 2000; 66: 87–92.

367. Gornik, I. et al., The identification of war victims by reverse paternity is associated with significant risk of false inclusion. *Int. J. Leg. Med*., 2002; 116: 255–257.

368. Gotleib, S. et al., *Crime Analysis — From First Report to Final Arrest*. 1998; Alpha Publishing: Montclair, CA.

369. Goza, W.M., William R Maples, forensic historian: four men, four centuries, four countries. *J. Forensic Sci*., 1999; 44: 692–694.

370. Graham, J. et al., Conditional genotypic probabilities for microsatellite loci. *Genetics*, 2000; 155: 1973–1980.

371. Greenhalgh, M.J. and Buffery, C.M., Estimation of genetic distance between subpopulations using DNA profiling data from crime investigations. In *International Society of Forensic Haemogenetics Meeting*. 1995; Venice.

372. Grigg, W.N., Waco Revisited. *The New American*, 2003; p. 25–27.

373. Grzybowski, T. et al., High levels of mitochondrial DNA heteroplasmy in single hair roots: reanalysis and revision. *Electrophoresis*, 2003; 24: 1159–1165.

374. Grzybowski, T., Extremely high levels of human mitochondrial DNA heteroplasmy in single hair roots. *Electrophoresis*, 2000; 21: 548–553.

375. Guimaraes, M.A., The challenge of identifying deceased individuals in Brazil: from dictatorship to DNA analysis. *Sci. Justice*, 2003; 43: 215–217.

376. Gunn, P.R. et al., DNA analysis in disputed parentage: the occurrence of two apparently false exclusions of paternity, both at short tandem repeat (STR) loci, in the one child. *Electrophoresis*, 1997; 18: 1650–1652.

377. Guo, S.W. and Thompson, E.A., Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics*, 1992; 48: 361–372.

378. Gusmão, L. et al., Point mutations in the flanking regions of the Y-chromosome specific STR's DYS391, DYS437, and DYS438. *Int. J. Leg. Med.*, 2002; 116: 322–326.

379. Gusmão, L. et al., Alternative primers for DYS391 typing: advantages of their application to forensic genetics. *Forensic Sci. Int.*, 2000; 112: 49–57.

380. Gusmão, L. et al., Grouping of Y-STR haplotypes discloses European geographic clines. *Forensic Sci. Int*., 2003; 134: 172–179.

381. Gyllensten, U. et al., Paternal inheritance of mitochondrial DNA in mice. *Nature*, 1991; 352: 255–257.

382. Hadkiss, C., personal communication, 2001.

383. Haldane, J.B.S., A combination of linkage values, and the calculation of distances between loci of linked factors. *J. Genet.*, 1919; 8: 299–309.

384. Haldane, J.B.S., Some theoretical results of continued brother–sister mating. *J. Genet.*, 1937; 34: 265–274.

385. Hallenberg, C. and Morling, N., A report on the 2000 and 2001 paternity testing workshops of the English speaking working group of the International Society for Forensic Genetics. *Forensic Sci. Int.*, 2002; 129: 43–50.

386. Halverson, J., Questgen Forensics. 2004; Available at http://www.questgen.biz (accessed 25 March 2004).

387. Hammer, H.J. et al., Human population structure and its effect on sampling Y chromosome sequence variation. *Genetics*, 2003; 164: 1495–1509.

388. Hammer, M.F., A recent common ancestry for human Y chromosomes. *Nature*, 1995; 378: 376–378.

389. Harbison, S.A. and Buckleton, J., Applications and extensions of sub-population theory: a caseworkers guide. *Sci. Justice*, 1998; 38: 249–254.

390. Harbison, S.A. et al., The New Zealand DNA databank: its development and significance as a crime solving tool. *Sci. Justice*, 2001; 41: 33–37.

391. Harbison, S.A., Petricevic, S.F., and Vintiner, S.K., The persistence of DNA under fingernails following submersion in water. *Int. Congress Series*, 2003; 1239: 809–813.

392. Hardy, G.H., Mendelian proportions in a mixed population. *Science*, 1908; 28: 49–50.

393. Harr, B., Mismatch repair-driven mutational bias in *D. melanogaster*. *Mol. Cell*, 2002; 10: 199–205.

394. Harr, B. and Schlotterer, C., Long microsatellite alleles in *Drosophila melonogaster* have a downward bias and short persistence times, which cause their genomewide underrepresentation. *Genetics*, 2000; 155: 1213–1220.

395. Hartl, D., Forensic DNA typing dispute. *Nature*, 1994; 372: 398–399.

396. Hartmann, J.M. et al., The effect of ethnic and racial population substructuring on the estimation of multi-locus fixed-bin VNTR RFLP genotype probabilities. *J. Forensic Sci.*, 1997; 42: 232–240.

397. Hasegawa, M. and Horai, S., Time of the deepest root for polymorphism in human mitochondrial DNA. *J. Mol. Evol.*, 1991; 32: 37–42.

398. Hasegawa, M. et al., Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.*, 1985; 32: 37–42.

399. Hazlitt, G., DNA statistical evidence: unravelling the strands. *Judicial Officers' Bull.*, 2002; 14: 66–69.

400. Helgason, A. et al., mtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic Ancestry. *Am. J. Hum. Genet.*, 2001; 68: 723–737.

401. Henderson, J.P., The use of DNA statistics in criminal trials. *Forensic Sci. Int.*, 2002; 128: 183–186.

402. Henke, J. et al., Application of Y-chromosomal STR haplotypes to forensic genetics. *Croat. Med. J.*, 2001; 42: 292–297.

403. Heyer, E. et al., Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Hum. Mol. Genet.*, 1997; 6: 799–803.

404. Heymann, P.B., *Lessons of WACO: Proposed Changes in Federal Law Enforcement.* 1993; U.S. Department of Justice, NIJ.

405. Hibbert, D.B., Scientist *vs* the law. *Accred. Qual. Assur.*, 2003; 8: 179–183.

406. Higuchi, R. et al., DNA typing from single hairs. *Nature*, 1988; 332: 543–546.

407. Hilden, J., GENEX — an algebraic approach to pedigree probability calculus. *Clin. Genet.*, 1970; 1: 319–348.

408. Hitchin, S. and Schuller, W., *Global DNA Database Inquiry*, *2002*. 2002; Interpol DNA Unit.

409. Hixson, J.E. and Clayton, D.A., Initiation of transcription from each of the two human mitochondrial promotors requires unique nucleotides at the transcriptional start sites. *Proc. Natl. Acad. Sci. USA*, 1985; 82: 2660–2664.

410. Holden, C., DNA fingerprinting comes of age. *Science*, 1997; 278: 1407.

411. Holland, M.M. and Parsons, T.J., Mitochondrial DNA sequence analysis — validation and use for forensic casework. *Forensic Sci. Rev.*, 1999; 11: 21–50.

412. Holland, M.M. et al., Development of a quality, high throughput DNA analysis procedure for skeletal samples to assist with the identification of victims from the World Trade Center Attacks. *Croat. Med. J.*, 2003; 44: 264–272.

413. Honda, K. et al., Male DNA typing from 25-year-old vaginal swabs using Y-chromosome STR polymorphisms in a retrial request case. *J. Forensic Sci.*, 1999; 44: 868–872.

414. Hopwood, A., personal communication, 2003.

415. Horgan, J., High profile. *Sci. Am.*, 1994; October: 18–19.

416. Horstman, M., Maori men and women from different homelands. *News in Science*, 27 March 2003; Available at http://www.abc.net.au/science/news/stories/s817069.htm (accessed 17 March 2004).

417. Houshmand, M. et al., Is paternal mitochondrial DNA transferred to the offspring following intracytoplasmic sperm injection? *J. Assist. Reprod. Genet.*, 1997; 14: 223–227.

418. Hoyle, R., The FBI's national DNA database. *Biotechnology*, 1998; 16: 987.

419. Hsieh, H.-M. et al., A highly polymorphic STR locus in *Cannabis sativa*. *Forensic Sci. Int.*, 2003; 131: 53–58.

420. Huang, Q.-Y. et al., Mutation patterns at dinucleotide microsatellite loci in humans. *Am. J. Hum. Genet.*, 2002; 70: 625–634.

421. Hudson, R.R., Linkage disequilibrium and recombination. In *Handbook of Statistical Genetics*, Cannings, C., Ed. 2001; John Wiley and Sons, Ltd: Chichester.

422. Hugin Expert Software. Available at http://www.hugin.com (accessed 6 February 2004).

423. Hühne, J. et al., Heteroplasmic substitutions in the mitochondrial DNA control region in mother and child samples. *Int. J. Leg. Med.*, 1999; 112: 27–30.

424. Hühne, J. et al., Mitochondrial DNA in human hair shafts — evidence of intra-individual differences? *Int. J. Leg. Med.*, 1999; 112: 172–175.

425. Hustmyre, C., Waco ten years later. *Law & Order*, 2003; 72–76.

426. Iida, R. et al., Characterization and haplotype analysis of the polymorphic Y-STRs, DYS443, DYS444 and DYS445 in a Japanese population. *Int. J. Leg. Med.*, 2002; 116: 191–194.

427. Imaizumi, K. et al., The effects of decalcification and glass beads treatment on DNA typings from bone samples. Report of the National Research Institute of Police Science. *Res. Forensic Sci.*, 1998; 51: 19–28.

428. Imaizumi, Y., A recent survey of consanguineous marriages in Japan. *Clin. Genet.*, 1986; 30: 230–233.

429. Inman, K. and Rudin, N., *Principles and Practice of Criminalistics. The Profession of Forensic Science*. 2001; CRC Press: Boca Raton, FL.

430. Ivanov, P.L. et al., Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgji Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nature Genet.*, 1996; 12: 417–420.

431. Jacquard, A., Genetics information given by a relative. *Biometrics*, 1972; 28: 1101–1114.

432. Jarjoura, D. et al., Likelihood ratios for deoxyribonucleic acid (DNA) typing in criminal cases. *J. Forensic Sci.*, 1994; 39: 64–73.

433. Jeffreys, A.J., Spontaneous and induced minisattelite instability in the human genome. Bayer Lecture presented at the Spring Meeting of the Medical Research Society held at the Royal College of Physicians, 1997.

434. Jeffreys, A.J. et al., Hypervariable minisatellite regions in human DNA. *Nature*, 1985; 314: 67–72.

435. Jeffreys, A.J. et al., Individual specific "fingerprints" of human DNA. *Nature*, 1985; 316: 75–79.

436. Jeffreys, A.J. et al., Amplification of human minisatellites by polymerase chain reaction: towards DNA fingerprinting of single cells. *Nucl. Acids Res.*, 1988; 10: 10953–10971.

437. Jeffreys, A.J. et al., Complex gene conversion events in germline mutation at human minisatellites. *Nature Genet.*, 1994; 6: 136–145.

438. Jeffreys, A.J. et al., Mutation processes at human minisatellites. *Electrophoresis*, 1995; 16: 1577–1585.

439. Jiricny, J., Colon cancer and DNA repair: have mismatches met their match? *Trends Genet.*, 1994; 10: 164–168.

440. Jobling, M., In the name of the father: surnames and genetics. *Trends Genet.*, 2001; 17: 353–357.

441. Jobling, M.A. The Y chromosome as a forensic tool: progress and prospects for the new millennium. In *First International Conference on Forensic Human Identification in the Millennium.* 1999; London.

442. Jobling, M.A. et al., The Y chromosome in forensic analysis and paternity testing. *Int. J. Leg. Med.*, 1997; 110: 118–124.

443. Jobling, M.A. et al., Hypervariable digital DNA codes for human paternal lineages: MVR:PCR at the Y-specific minisatellite, MSY1 (DYF155S1). *Hum. Mol. Genet.*, 1998; 7: 643–653.

444. Jobling, M.A. et al., Y-chromosome-specific microsatellite mutation rates reexamined using a minisatellite, MSY1. *Hum. Mol. Genet.*, 1999; 8: 2117–2120.

445. Johnson, C.L. et al., Validation and uses of a Y-chromosome STR 10-plex for forensic and paternity laboratories. *J. Forensic Sci.*, 2003; 48: 1260–1268.

446. Johnston, M. and Stormo, G.D., Heirlooms in the attic. *Science*, 2003; 302: 997–999.

447. Jones, D.A., Blood samples: probability of discrimination. *J. Forensic Sci.* Soc., 1972; 12: 355–359.

448. Jorde, L.B. et al., Microsatellite diversity and the demographic history of modern humans. *Proc. Natl. Acad. Sci. USA*, 1997; 94: 3100–3103.

449. Jost, K., DNA databases. *CQ Researcher*, 1999; 9: 449–472.

450. Kawano, S. et al., A genetic system controlling mitochondrial fusion in the slime mould, *Physarum polycephalum. Genetics*, 1993; 133(2): 213–224.

451. Kayser, M. et al., Evaluation of Y chromosomal STRs: a multicentre study. *Int. J. Leg. Med.*, 1997; 110: 125–133.

452. Kayser, M. and Sajantila, A., Mutations at Y-STR loci: implications for paternity testing and forensic analysis. *Forensic Sci. Int.*, 2001; 118: 116–121.

453. Kayser, M. et al., Independent histories of human Y chromosome from Melanesia and Australia. *Am. J. Hum. Genet.*, 2001; 68: 173–190.

454. Kayser, M. et al., Characteristics and frequency of germline mutations at microsatellite loci from human Y chromosome, as revealed by direct observation in father/son pairs. *Am. J. Hum. Genet.*, 2000; 66: 1580–1588.

455. Kelly, J.F. and Wearne, P.K., *Tainting Evidence.* 1998; The Free Press: New York.

456. Kidd, J.K. et al., Studies of three Amerind populations using nuclear DNA polymorphisms. *Hum. Biol.*, 1991; 63: 775–783.

457. Kiernan, V., Tough tests to cross-examine DNA data. *New Sci.*, 1995; June: 8.

458. Kimpton, C. et al., Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. *Int. J. Leg. Med.*, 1994; 106: 302–311.

459. Kimpton, C.P. et al., A further tetranucleotide repeat polymorphism in the vWF gene. *Hum. Mol. Genet.*, 1992; 1: 287.

460. Kimpton, C.P. et al., Automated DNA profiling employing multiplex amplification of short tandem repeat loci. *PCR Methods Appl.*, 1993; 3: 13–22.

461. Kimpton, C.P. et al., Report on the second EDNAP collaborative STR exercise. European DNA Profiling Group. *Forensic Sci. Int*., 1995; 71: 137–152.

462. Kimura, M. and Ohta, T., Mutation and evolution at the molecular level. *Genetics*, 1973; 73(Suppl.): 19–35.

463. Kimura, M. and Ohta, T., Stepwise mutational model (SMM) and distribution of allelic frequencies in a finite population. *Proc. Natl. Acad. Sci. USA*, 1978; 75: 2868–2872.

464. King, J., DOJ aware of problems in FBI's DNA lab. 2003; Available at http://www/nacdl.org/MEDIA/proooo96.htm (accessed 29 April 2003).

465. Kirby, M., DNA evidence: proceed with care. *Aust. J. Forensic Sci*., 2001; 33: 9–13.

466. Kirkham, A., personal communication, 2001.

467. Klintschar, M. and Wiegand, P., Polymerase slippage in relation to the uniformity of tetrameric repeat stretches. *Forensic Sci. Int.*, 2003; 135: 163–166.

468. Klintschar, M. et al., Persisting fetal microchimerism does not interfere with forensic Y-chromosome typing. *Forensic Sci. Int.*, 2004; 139: 151–154.

469. Kloosterman, A. and Janssen, H., Forensic DNA testing and its legislation in the Netherlands. *Forensic Sci. Int.*, 1997; 88: 55.

470. Knight, A. et al., Molecular, forensic and haplotype inconsistencies regarding the identity of the Ekaterinberg remains. *Ann. Hum. Biol.*, 2004; preview article: 1–8.

471. Kocher, T.D. and Wilson, A.C., Sequence evolution of mitochondrial DNA in humans and chimpanzees: control region and protein coding region. In *Evolution of Life: Fossils, Molecules and Culture*, Honjo, T., Ed. 1991; Springer-Verlag: Tokyo, p. 391–413.

472. Koehler, J.J., Error and exaggeration in the presentation of DNA evidence at trial. *Jurimetrics*, 1993; 34: 21–39.

473. Koehler, J.J., DNA matches and statistics: important questions, surprising answers. *Judicature*, 1993; 76: 222–229.

474. Koehler, J.J., The base rate fallacy reconsidered: descriptive, normative, and methodological challenges. *Behav. Brain Sci.*, 1996; 19: 1–53.

475. Koehler, J.J., On conveying the probative value of DNA evidence: frequencies, likelihood ratios and error rates. *Univ. Colorado Law Rev*., 1996; 67: 859–886.

476. Koehler, J.J., When are people persuaded by DNA match statistics? *Law Hum. Behav.*, 2001; 25: 493–513.

477. Koehler, J.J. et al., The random match probability in DNA evidence: irrelevant and prejudicial? *Jurimetrics*, 1995; 35: 201–219.

478. Kopel, D.B. and Blackman, P.H., The unwarranted warrant: the WACO search warrant and the decline of the fourth amendment. *J. Public Law Policy*, 1996; 18: 1–9.

479. Krane, D. et al., Genetic differences at four DNA typing loci in Finnish, Italian, and mixed Caucasian populations. *Proc. Natl. Acad. Sci. USA*, 1992; 89: 10583–10587.

480. Kuroki, H. et al., Spermatogenic ability is different among males in different Y chromosome lineage. *J. Hum. Genet.*, 1999; 44: 289–292.

481. LaBerge, G.S. et al., Forensic utility of mitochondrial DNA analysis based on denaturing high-performance liquid chromatography. *Croat. Med. J.*, 2003; 44: 281–288.

482. Ladd, C. et al., A systematic analysis of secondary DNA transfer. *J. Forensic Sci.*, 1999; 44: 1270–1272.

483. Lagerstrom-Fermer, M. et al., Heteroplasmy of the human mtDNA control region remains constant during life. *Am. J. Hum. Genet.*, 2001; 68: 1299–1301.

484. Lander, E., DNA fingerprinting on trial. *Nature*, 1989; 339: 501–505.

485. Lander, E., Invited editorial: research on DNA typing catches up with courtroom application. *Am. J. Hum. Genet.*, 1991; 48: 819–823.

486. Lander, E. and Budowle, B., DNA fingerprinting dispute laid to rest. *Nature*, 1994; 371: 735–738.

487. *Landry v Attorney General* (No. 98462, 1998 WL 1181741).

488. Laurie, C. and Weir, B.S., Dependency effects in multi-locus match probabilities. *Theor. Popul. Biol.*, 2003; 63: 207–219.

489. Lauritzen, S.L. and Mortera, J., Bounding the number of contributors to mixed DNA stains. *Forensic Sci. Int.*, 2002; 130: 125–126.

490. Lavine, M. and Schervish, M.J., Bayes factors: what they are and what they are not. *Am. Stat.*, 1999; 53: 119–122.

491. Law, B., Statistical Interpretation of DNA Evidence. Ph.D. thesis, Department of Statistics. 2002; University of Auckland: Auckland, N.Z.

492. Law, B. et al., Effects of population structure and admixture on exact tests for association between loci. *Genetics*, 2003; 164: 381–387.

493. Lawton, M.E. et al., An international survey of the reporting of hypothetical cases. *J. Forensic Sci. Soc.*, 1988; 28: 243–252.

494. Leary, D. and Pease, K., DNA and the active criminal population. *Crime Prevention Community Safety: Int. J.*, 2003; 5: 7–12.

495. Lebel, R.R., Consanguinity studies in Wisconsin. I: Secular trends in consanguineous marriages, 1843–1981. *Am. J. Med. Genet.*, 1983; 15: 543–560.

496. Leclair, B. et al., Enhanced kinship analysis and STR-based DNA typing for human identification in mass disasters. In *Progress in Forensic Genetics*, Vol. 8, Sensabaugh, G.F., Lincoln, P.J., and Olaisen, B., Eds. 2000; Elsevier Science: Amsterdam, p. 91–93.

497. Leclair, B. et al., STR DNA typing: increased sensitivity and efficient sample consumption using reduced PCR reaction volumes. *J. Forensic Sci.*, 2003; 48: 1001–1013.

498. Lee, H.C. and Labriola, J., *Famous Crimes Revisited*. 2001; Strong Books: Southington.

499. Lee, H.-S. et al., Motherless case in paternity testing. *Forensic Sci. Int.*, 2000; 114: 57–65.

500. Lee, P.M., *Bayesian Statistics — An Introduction*, 2nd ed. 1997; John Wiley and Sons, Inc.: New York.

501. Lempert, R., Comment: theory and practice in DNA fingerprinting. *Stat. Sci.*, 1994; 9: 255–258.

502. Lempert, R., After the DNA wars: skirmishing with NRC II. *Jurimetrics*, 1997; 37: 439–468.

503. Lempert, R.O., Some caveats concerning DNA as criminal identification evidence: with thanks to the Reverend Bayes. *Cardozo Law Rev.*, 1991; 13: 303–341.

504. Lempert, R.O., DNA, science and the law: two cheers for the ceiling principle. *Jurimetrics*, 1993; 34: 41–57.

505. Lempert, R.O., The suspect population and DNA identification. *Jurimetrics*, 1993; 34: 1–7.

506. Leung, W.-C., The prosecutor's fallacy — a pitfall in interpreting probabilities in forensic evidence. *Med. Sci. Law*, 2002. 42: 44–50.

507. Levy, H., *And the Blood Cried Out*. 1996; Basic Books: New York.

508. Lewontin, R., Which population? *Am. J. Hum. Genet.*, 1993; 52: 205.

509. Lewontin, R., Forensic DNA typing dispute. *Nature*, 1994; 372: 398.

510. Lewontin, R.C., Comment: the use of DNA profiles in forensic contexts. *Stat. Sci.*, 1994; 9: 259–262.

511. Lewontin, R.C. and Hartl, D.L., Population genetics in forensic DNA typing. *Science*, 1991; 254: 1745–1750.

512. Li, C.C. and Sacks, L., The derivation of joint distribution and correlation between relatives by the use of stochastic measures. *Biometrics*, 1954; 10: 347–360.

513. Li, C.C. et al., Similarity of DNA fingerprints due to chance and relatedness. *Hum. Hered.*, 1993; 43: 45–52.

514. Li, W.H. and Nei, M., Stable linkage disequilibrium without epistasis in subdivided populations. *Theor. Popul. Biol.*, 1974; 6: 173–183.

515. Lidor, D., Be all that DNA can be. 2002; Available from http://www.wired.com/news/print/0,1294,54631,00.html (accessed 17 March 2004).

516. Lindley, D.V., Probabilities and the law. In *Utility, Probability and Human Decision Making*, Vlek, C., Ed. 1975; Reidel Publishing Company: Dordrecht, Holland, p. 223–232.

517. Lindley, D.V., A problem in forensic science. *Biometrika*, 1977; 64: 207–213.

518. Lindley, D.V., Probability and the law. *The Statistician*, 1977; 26: 203–220.

519. Lindley, D.V., Probability. In *The Use of Statistics in Forensic Science*, Stoney, D.A., Ed. 1991; Ellis Horwood: New York, p. 27–50.

520. Liu, B. et al., Mismatch repair gene defects in sporadic colorectal cancers with microsatellite instability. *Nature Genet.*, 1995; 9: 48–53.

521. Loftus, P., The FBI's combined DNA index system helps corrections and law enforcement professionals monitor inmates and solve crimes. *Corrections Today*, 1999; July: 68–71.

522. Lonjou, C. et al., Linkage disequilibrium in human populations. *Proc. Natl. Acad. Sci. USA*, 2003; 100: 6069–6074.

523. Lowe, A. et al., The propensity of individuals to deposit DNA and secondary transfer of low level DNA from individuals to inert surfaces. *Forensic Sci. Int.*, 2002; 129: 25–34.

524. Lowe, A. et al., Use of low copy number DNA in forensic inference. *Int. Congress Series*, 2003; 1239: 799–801.

525. Lutz, S. et al., Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. *Int. J. Leg. Med.*, 1998; 111: 67–77.

526. Lynch, M., God's signature: DNA profiling, the new gold standard in forensic science. *Endeavour*, 2003; 27: 93–97.

527. Lynch, T., No confidence. An unofficial account of the Waco incident. *Policy Analysis*, 2001; 395: 1–18.

528. Lyon, M.F., Gene action in the X-chromosome of the mouse (*Mus musculus L.*). *Nature*, 1961; 190: 372–373.

529. Macan, M. et al., Paternity testing in a case of brother–sister incest. *Croat. Med. J.*, 2003; 44: 347–349.

530. Macauley, V. et al., The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am. J. Hum. Genet.*, 1999; 64: 232–249.

531. Maiste, P.J. and Weir, B.S., A comparison of tests for independence in the FBI RFLP databases. *Genetica*, 1995; 96: 125–138.

532. Makalowski, W., Not junk after all. *Science*, 2003; 300: 1246–1247.

533. Malaver, P.C. and Yunis, J.J., Different dental tissues as a source of DNA for human identification in forensic cases. *Croat. Med. J.*, 2003; 44: 306–309.

534. Malecot, G., *Les mathematiques de l'heredite*. 1948; Masson et Cie: Paris.

535. Martin, B. et al., PENDULUM — a guideline based approach to the interpretation of STR mixtures, *Forensic Sci. Int.* (accepted).

536. Martin, P. et al., A Spanish population study of 17 Y-chromosome STR loci. *Forensic Sci. Int.*, 2004; 139: 231–235.

537. Martin, P.D. et al., A brief history of the formation of DNA databases in forensic science within Europe. *Forensic Sci. Int.*, 2001; 119: 225–231.

538. Mattick, J.S., Introns: evolution and function. *Curr. Opin. Genet. Dev.*, 1994; 4: 823–831.

539. Mattick, J.S., Non-coding RNA's: the architects of eukaryotic complexity. *Eur. Mol. Biol. Org. Rep.*, 2001; 2: 986–991.

540. McKasson, S., I think therefore I probably am. *J. Forensic Ident.*, 2001; 51: 217–221.

541. McKie, I.A.J., There's none ever fear'd that the truth should be heard but they whom the truth would indite. *Sci. Justice*, 2003; 43: 161–165.

542. Meagher, D., The quiet revolution? A brief history and analysis of the growth of forensic police powers in Victoria. *Crim. Law J.* 2000; 24: 76–88.

543. Meester, R. and Sjerps, M., The evidential value in the DNA database search controversy and the two-stain problem. *Biometrics*, 2003; 59: 727–732.

544. Melton, T. and Stoneking, M., Extent of heterogeneity in mitochondrial DNA of ethnic Asian populations. *J. Forensic Sci.*, 1996; 41: 591–602.

545. Melton, T. et al., Extent of heterogeneity in mitochondrial DNA of European populations. *J. Forensic Sci.*, 1997; 42: 437–446.

546. Mendel, G., Experiments in plant hybridisation. In *English Translation and Commentary by R.A. Fisher*, Bennett, J.H., Ed. 1866 (1965); Oliver and Boyd: Edinburgh.

547. Menotti-Raymond, M. et al., Quantitive polymerase chain reaction-based assay for estimating DNA yield extracted from domestic cat specimens. *Croat. Med. J.*, 2003; 44: 327–331.

548. Mesa, N.R. et al., Autosomal, mtDNA, and Y chromosome diversity in Amerinds: pre and post-Columbian patterns of gene flow in South America. *Am. J. Hum. Genet.*, 2000; 67: 1277–1286.

549. Meyer, H.J., The Kaprun cable car disaster — aspects of forensic organisation following a mass fatality with 155 victims. *Forensic Sci. Int.*, 2003; 138: 1–7.

550. Meyer, S. et al., Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics*, 1999; 152: 1103–1110.

551. Mills, K.A. et al., Tetranucleotide repeat polymorphism at the human alpha fibrinogen locus (FGA). *Hum. Mol. Genet.*, 1992; 1: 779.

552. Minaguchi, K. et al., DNA analysis of neonatal human remains wrapped and kept in a vinyl bag for 15 years. *Leg., Med.*, 2003; 5: 183–186.

553. Mobbs, J., The National Criminal Investigation DNA Database — an investigative tool for the future. In *The Use of DNA in the Criminal Justice System Public Seminar*. April 2001; Institute of Criminology, Sydney University Law School, Australia.

554. Monahan, S.L. et al., The use of likelihood ratios in reporting difficult forensic cases. In *Advances in Forensic Haemogenetics*, Bär, W., Ed. 1996; Springer-Verlag: Berlin, 87–89.

555. Morgan, T.H., Random segregation versus coupling in Mendelian inheritance. *Science*, 1911; 34: 384.

556. Morgan, T.H., *The Theory of Genes*. 1928; Yale University Press: New Haven.

557. Morling, N., International recommendations for paternity testing standards — Guest Editorial. *Int. J. Leg. Med.*, 2003; 117: 1.

558. Morling, N. and Carracedo, A., International recommendations for paternity testing standards. *Forensic Sci. Int.*, 2002; 129: 147.

559. Morling, N. et al., Paternity Testing Commission of the International Society of Forensic Genetics: recommendations on genetic investigations in paternity cases. *Forensic Sci. Int.*, 2002; 129: 148–157.

560. Morling, N. et al., Paternity Testing Commission of the International Society of Forensic Genetics Recommendations on genetic investigations in paternity cases. *Int. J. Leg. Med*., 2003; 117: 56–61.

561. Morris, J., *The Age of Arthur*. 1993; Butler and Tanner Ltd.: Frome and London.

562. Morris, J.W. et al., The avuncular index and the incest case. *Advances in Forensic Haemogenetics*. 1988; Springer-Verlag: Berlin.

563. Morrison, R., Databasing efforts in Alabama — past and present. In *Tenth International Symposium on Human Identification*. 1999; Promega Corporation, Madison, WI.

564. Mortera, J. et al., Probabilistic expert system for DNA mixture profiling. *Theor. Popul. Biol.*, 2003; 63: 191–205.

565. Morton, N.E., The forensic DNA endgame. *Jurimetrics*, 1997; 37: 477–494.

566. Mourant, A. *Blood Relations: Blood Groups and Anthropology.* 1983; Oxford University Press: New York.

567. Mourant, A.E., Blood groups and diseases. *Haematologia*, 1974; 8: 183–194.

568. Mullis, K.B. et al., Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol*., 1986; 51: 263–273.

569. Murphy, W. et al., Evidence for complete symmetrical transcription *in vivo* of mitochondrial DNA in HeLa cells. *J. Mol. Biol.*, 1975; 99: 809–814.

570. Murray-McIntosh, R.P. et al., Testing migration patterns and estimating founding population size in Polynesia by using human mtDNA sequences. *Proc. Natl. Acad. Sci. USA*, 1998; 95: 9047–9052.

571. Nadot, R. and Vayssex, G., Appartment et identite. Algorithme du calcul des coefficients d'identite. *Biometrics*, 1973; 29: 347–359.

572. Naito, E. et al., Novel paternity testing by distinguishing parental alleles at a VNTR locus in the differentially methylated region upstream of the human H19 gene. *J. Forensic Sci.*, 2003; 48: 1275–1279.

573. Nakanishi, A. et al., Effects of environmental conditions to which nails are exposed on DNA analysis of them. *Leg. Med.*, 2003; 5: 194–197.

574. National Association of Testing Authorities (NATA) Australia, ISO/IEC 17025 Application Document: Supplementary Requirements for Accreditation in the Field of Forensic Science. 2000.

575. Nei, M., *Molecular Evolutionary Genetics*. 1987; Columbia University Press: New York.

576. Nei, M. and Li, W.H., Linkage disequilibrium in subdivided populations. *Genetics*, 1973; 75: 213–219.

577. Neuhauser, C., Mathematical models in population genetics. In *Handbook of Statistical Genetics*, Cannings, C., Ed. 2001; John Wiley and Sons, Ltd: Chichester.

578. Newton, J. and Weistein, H., Simpson: Ito to release juror dismissal transcripts in July. *Los Angeles Times*, 24 June 1995; p. A1 and A24.

579. Neyman, J. and Pearson, E., On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. R. Soc. A*, 1933; 231: 289–337.

580. Nicholson, G. et al., Assessing population differentiation and isolation from single nucleotide polymorphism data. *J. R. Stat. Soc. B*, 2002; 64: 1–21.

581. Nickerson, R.S., Null hypothesis testing: a review of an old and continuing controversy. *Psych. Methods*, 2000; 5: 241–301.

582. Nijenhuis, L.E., A critical evaluation of various methods of approaching probability of paternity. In *Inclusion Probabilities in Parentage Testing*. Walker, R.H., Ed. 1983; American Association of Blood Banks: Arlington, VA, p. 103–112.

583. Nowak, R., Forensic DNA goes to court with O.J. *Science*, 1994; 265: 1352–1354.

584. NRC I — National Research Council Committee on DNA Technology in Forensic Science, Board on Biology, Commission on Life Sciences, DNA Technology in Forensic Science. 1992; National Academy Press: Washington, DC.

585. NRC II — National Research Council Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence. 1996; National Academy Press: Washington, DC.

586. Nutini, A.L. et al., Double incompatibility at human alpha fibrinogen and penta E loci in paternity testing. *Croat. Med. J.*, 2003; 44: 342–346.

587. Ohno, K. et al., Identification of a possible control element, Mt5, in the major noncoding region of mitochondrial DNA by intraspecific nucleotide conservation. *Biochem. Int.*, 1991; 24: 263–272.

588. Ohno, S., *Sex Chromosomes and the Sex Linked Genes*. 1967; Springer: Berlin.

589. Ohno, Y. et al., A simple method for calculating the probability of excluding paternity with any number of codominant alleles. *Forensic Sci. Int.*, 1982; 19: 93–98.

590. Ohta, T., Linkage disequilibrium due to random genetic drift in finite subdivided populations. *Proc. Natl. Acad. Sci. USA*, 1982; 79: 1940–1944.

591. Ohta, T. and Kimura, M., Linkage disequilibrium due to random genetic drift. *Genet. Res.*, 1969; 13: 47–55.

592. Ojala, D. et al., tRNA punctuation model of RNA processing in human mitochondrial DNA. *Nature*, 1981; 99: 470–474.

593. Olaisen, B. et al., Identification by DNA analysis of the victims of the August 1996 Spitsbergen civil aircraft disaster. *Nature Genet.*, 1997; 15: 402–405.

594. Olaisen, B. et al., DNA recommendations 1997 of the International Society for Forensic Genetics. *Vox Sang*, 1998; 74: 61–63.

595. Oota, H. et al., Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. *Nature Genet.*, 2001; 29: 20–21.

596. Orkin, S.T. et al., Linkage of b-thalassaemia mutations and b-globin gene polymorphisms with DNA polymorphisms in human b-globin gene cluster. *Nature*, 1982; 296: 627–631.

597. Ott, J., *Analysis of Human Genetic Linkage*, 3rd ed. 1985; The John Hopkins University Press: London.

598. Ottenheiner, M., Lewis Henry Morgan and the prohibition of cousin marriage in the United States. *J. Family History*, 1990; 15: 325–334.

599. Overall, A.D.J., The Geographic Scale of Human Genetic Differentiation at Short Tandem Repeat Loci. 1998; Ph.D. thesis, The University of London.

600. Overall, A.D.J. et al., Genetic differentiation within and between four UK ethnic groups. *Forensic Sci. Int.*, 2000; 114: 7–20.

601. Padar, Z. et al., Canine STR analyses in forensic practice. Observation of a possible mutation in a dog hair. *Int. J. Leg. Med.*, 2002; 116: 286–288.

602. Pai, C.-Y. et al., Allelic alterations at the STR markers in the buccal tissue cells of oral cancer patients and the oral epithelial cells of healthy betel quid chewers: an evaluation of forensic applicability. *Forensic Sci. Int.*, 2002; 129: 158–167.

603. Painter, I.S. and Weir, B.S., personal communication.

604. Panckhurst, P., All Swahili to Court. *New Zealand Herald*, 1994; p. 2

605. Paracchini, S. et al., Heirarchical high-throughput SNP genotyping of the human Y chromosome using MALDI-TOF mass spectrometry. *Nucl. Acids Res.*, 2002; 30: e27.

606. Parra, E.J. et al., Estimating African American proportions by use of population-specific alleles. *Am. J. Hum. Genet.*, 1998; 63: 1839–1851.

607. Parson, W. et al., The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organisation, results and perspectives. *Forensic Sci. Int.*, 2004; 139: 215–226.

608. Parsons, T.J. and Irwin, J.A., Questioning evidence for recombination in human mitochondrial DNA. *Science*, 2000; 288: 1931.

609. Parsons, T.J. et al., A high observed substitution rate in the human mitochondrial DNA control region. *Nature Genet.*, 1997; 15: 363–368.

610. Pearman, C., personal communication, 2002.

611. Pearson, K., On a generalized theory of alternative inheritance, with special references to Mendel's laws. *Phil. Trans. R. Soc. A*, 1904; 203: 53–86.

612. Pearson, K., *The History of Statistics in the 17th and 18th Centuries Against the Changing Background of Intellectual, Scientific and Religious Thought*. 1978; Charles Griffin and Company, Ltd.: London.

613. Pepinski, W. et al., Allele distribution of 15 STR loci in a population sample of Byelorussian minority residing in northeastern Poland. *Forensic Sci. Int.*, 2004; 139: 265–267.

614. Pérez-Lezaum, A. et al., Population genetics of Y-chromosome short tandem repeats in humans. *J. Mol. Evol.*, 1997; 45: 265–270.

615. Perlin, M.W. and Szabady, B., Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. *J. Forensic Sci.* 2001; 46: 1372–1378.

616. Pfannkuch, M. et al., Probability with less pain. *Teaching Statistics*, 2003; 24: 24–32.

617. Pfeiffer, H. et al., The results of a mtDNA study of 1200 inhabitants of a German village in comparison to other Caucasian databases and its relevance for forensic work. *Int. J. Leg. Med*., 2001; 114: 169–172.

618. Phillips, L.J., Jowitt, J., and Keene, J., *Regina v. Doheny and Adams. Crim. Appeal Rep*., 1997; 3: 369–389.

619. Pinchin, R.A. and Buckleton, J., PENDULUM. 1995; Birmingham.

620. Pizzamiglio, M. et al., DNA typing on latex gloves. In *Progress in Forensic Genetics*, Sensabaugh, G.F. et al., Eds. 2000; Elesvier: Amsterdam, p. 504–507.

621. Polymeropolous, M.H. et al., Tetranucleotide repeat polymorphism at the human tyrosine hydrolase gene (TH). *Nucl. Acids Res.*, 1991; 19: 3753.

622. Popovic, T. and Glass, M. Laboratory aspects of bioterrorism-related Anthrax — from identification to molecular subtyping to microbial forensics. *Croat. Med. J.*, 2003; 44: 336–341.

623. Presciuttini, S. et al., Allele sharing in first-degree and unrelated pairs of individuals in the Ge.F.I. AmpF*l*STR® Profiler Plus™ database. *Forensic Sci. Int.*, 2003; 131: 85–89.

624. Primmer, C. et al., Directional evolution in germline microsatellite mutations. *Nature Genet.*, 1996; 13: 391–393.

625. Primmer, C.R. et al., Unravelling the processes of microsatellite evolution through analysis of germline mutations in barn swallows *Hirundo rustica*. *Mol. Biol. Evol.*, 1998; 15: 1047–1054.

626. Prinz, M. et al., Validation and casework application of a Y chromosome specific STR multiplex. *Forensic Sci. Int.*, 2001; 120: 177–188.

627. Puers, C. et al., Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUMTH01 [AATG]n and reassignment of alleles on population analysis by using a locus specific ladder. *Am. J. Hum. Genet.*, 1993; 53: 953–958.

628. Purcell, N., Winfree, L.T., and Mays, G.L., DNA (deoxyribonucleic acid) evidence and criminal trials: an exploratory survey of factors associated with the use of "genetic fingerprinting" in felony prosecutions. *J. Crim. Justice*, 1994; 22: 145–157.

629. Pusch, C.M. et al., Molecular phylogenetics employing modern and ancient DNA. *J. Appl. Genet.*, 2003; 44: 269–290.

630. Quintana-Murci, L. et al., The human Y chromosome: function, evolution and disease. *Forensic Sci. Int.*, 2001; 118: 169–181.

631. Radzinsky, E., *The Last Tsar. The Life and Death of Nicholas II*. Translated from the Russian by Marian Schwartz. 1992; Anchor Books, Doubleday: London.

632. Rasmussen, E.M. et al., Sequencing strategy of mitochondrial HV1 and HV2 DNA with length heteroplasmy. *Forensic Sci. Int.*, 2002; 129: 209–213.

633. Ratcliffe, J., Implementing and integrating crime mapping into a police intelligence environment. *Int. J. Police Sci. Mgmt.*, 2000; 2: 313–323.

634. Ratcliffe, J.H., Intelligence-led policing and the problems of turning rhetoric into practice. *Policing Soc.*, 2002; 12: 53–66.

635. Redd, A.J. et al., Forensic value of 14 novel STRs on the human Y chromosome. *Forensic Sci. Int.*, 2002; 130: 97–111.

636. Redmayne, M., The DNA database: civil liberty and evidentiary issues. *Crim. Law Rev.*, 1998; 437–454.

637. *Regina v Galli*. 2001; New South Wales Court of Criminal Appeal, 12 December 2001.

638. *Regina v GK*. 2001; New South Wales Court of Criminal Appeal, 16 October 2001.

639. *Regina v Karger*. 2001; Supreme Court of South Australia 64, 29 March 2001.

640. Repping, S. et al., Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. *Nature Genet.*, 2003; 35: 247–251.

641. Research and Development Working Group, *The Future of Forensic DNA Testing: Predictions of the Research and Development Working Group*. 2000; United States Department of Justice.

642. Riancho, J.A. and Zarrabeitia, M.T., The prosecutor's and defendants's Bayesian nomograms. *Int. J. Leg. Med.*, 2002; 116: 312–313.

643. Riancho, J.A. and Zarrabeitia, M.T., A Windows-based software for common paternity and sibling analyses. *Forensic Sci. Int.*, 2003; 135: 232–234.

644. Ribaux, O. and Margot, P.A., Inference structures for crime analysis and intelligence: the example of burglary using forensic science data. *Forensic Sci. Int.*, 1999; 100: 193–210.

645. Ribaux, O. and Margot, P., Case based reasoning in criminal intelligence using forensic case data. *Sci. Justice*, 2003; 43: 135–143.

646. Ribaux, O. et al., Forensic intelligence and crime analysis. *Law Prob. Risk*, 2003; 2: 47–60.

647. Richards, M.B. et al., Tracing European founder lineages in the near eastern mtDNA pool. *Am. J. Hum. Genet.*, 2000; 67: 1251–1276.

648. Risinger, D.M. et al., The Daubert/Kuhmo implications of observer effects in forensic science: hidden problems of expectation and suggestion. *Calif. Law Rev.*, 2002; 90: 1–56.

649. Roberts, H., Interpretation of DNA evidence in courts of law — a survey of the issues. *Aust. J. Forensic Sci.*, 1998; 30: 29–40.

650. Roberts, H.F. et al., *Experiments in Plant Hybridisation*. 1965; Oliver and Boyd: London.

651. Robertson, B. and Vignaux, G.A., Inferring beyond reasonable doubt. *Oxford J. Leg. Stud.*, 1991; 11: 431–438.

652. Robertson, B. and Vignaux, G.A., Extending the conversation about Bayes. *Cardozo Law Rev.*, 1991; 13: 629–645.

653. Robertson, B. and Vignaux, G.A., Unhelpful evidence in paternity cases. *NZLJ*, 1992; (9): 315–317.

654. Robertson, B. and Vignaux, G.A., Expert evidence: law, practice and probability. *Oxf. J. Leg., Stud.*, 1992; 12: 392–403.

655. Robertson, B. and Vignaux, G.A., Probability — the logic of the law. *Oxf. J. Leg. Stud.*, 1993; 13: p. 457–478.

656. Robertson, B. and Vignaux, G.A., Taking fact analysis seriously. *Michigan Law Rev.*, 1993; 91: 1442–1464.

657. Robertson, B. and Vignaux, G.A., Crime investigation and the criminal trial. *J. Forensic Sci.* Soc., 1994; 34: 270.

658. Robertson, B. and Vignaux, G.A., DNA evidence: wrong answers or wrong questions? In *Human Identification: The Use of DNA Markers*, Weir, B.S., Ed. 1995; Kluwer Academic Publishers: Dordrecht, p. 145–152.

659. Robertson, B. and Vignaux, G.A., *Interpreting Evidence — Evaluating Forensic Science in the Courtroom*. 1995; John Wiley and Sons, Inc.: Chichester.

660. Robertson, B. and Vignaux, G.A., DNA on appeal — II. *NZLJ*, 1997: 247–250.

661. Robertson, B. and Vignaux, G.A., DNA on appeal. *NZLJ*, 1997; 210–212.

662. Robertson, B. and Vignaux, G.A., Bayes' theorem in the court of appeal. *The Criminal Lawyer*, 1998; January: 4–5.

663. Robertson, B. and Vignaux, G.A., Explaining evidence logically. *New Law J.*, 1998; 148: 159–162.

664. Robertson, B. and Vignaux, G.A., Don't teach statistics to lawyers! In *Fifth International Conference on Teaching of Statistics*. 1998; Singapore.

665. Roeder, K., DNA fingerprinting: a review of the controversy. *Stat. Sci.*, 1994; 9: 222–278.

666. Roewer, L. et al., Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. *Hum. Genet.*, 1992; 89: 389–394.

667. Roewer, L. et al., Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. *Forensic Sci. Int.*, 2001; 118: 106–113.

668. Rolf, B. et al., Somatic mutations at STR loci — a reason for three allele patterns and mosaicism. *Forensic Sci. Int.*, 2002; 126: 200–202.

669. Rolf, B. et al., Paternity testing using Y-STR haplotypes: assigning a probability for paternity in cases of mutations. *Int. J. Leg. Med.*, 2001; 115: 12–15.

670. Romualdi, C. et al., Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Res.*, 2002; 12: 602–612.

671. Rosenberg, N.A. et al., Genetic structure of human populations. *Science*, 2002; 298: 2381–2385.

672. Rosenberg, N.A. et al., Response to comment on "Genetic structure of human populations." *Science*, 2003; 300: 1877c.

673. Ross, A.M. and Harding, H.W.J., DNA typing and forensic science. *Forensic Sci. Int.*, 1989; 41: 197–203.

674. Rowlands, R.T. and Turner, G., Recombination between the extranuclear genes conferring oligomycin resistance and cold sensitivity in *Aspergillus nidulans*. *Mol. Genet.*, 1974; 133: 151–161.

675. Rozen, S. et al., Abundant gene conversion between arms of palindromes in human and ape Y-chromosomes. *Nature*, 2003; 423: 873–876.

676. Rubinsztein, D.C. et al., Microsatellite evolution — evidence for directionality and variation in rate between species. *Nature Genet.*, 1995; 10: 337–343.

677. Rubinsztein, D.C. et al., Mutational bias provides a model for the evolution of Huntington's disease and predicts a general increase in disease prevalence. *Nature Genet.*, 1994; 7: 525–530.

678. Rudin, N. and Inman, K., *An Introduction to Forensic DNA Analysis*. 1997; CRC Press LLC: Boca Raton, FL.

679. Rudin, N. and Inman, K., *An Introduction to Forensic DNA Analysis*, 2nd ed. 2002; CRC Press LLC: Boca Raton, FL.

680. Ruitberg, C.M., Reeder, D.J., and Butler, J.M., STRBase: a short tandem repeat DNA database for the human identity testing community. *Nucl. Acids Res.*, 2001; 29: 320–322.

681. Rutty, G.N., An investigation into the transference and survivability of human DNA following simulated manual strangulation with consideration of the problem of third party contamination. *Int. J. Leg. Med.*, 2002; 116: 170–173.

682. Rutty, G.N. et al., DNA contamination of mortuary instruments and work surfaces: a significant problem in forensic practice? *Int. J. Leg. Med.*, 2000; 114: 56–60.

683. Rutty, G.N. et al., The effectiveness of protective clothing in the reduction of potential DNA contamination of the scene of crime. *Int. J. Leg. Med.*, 2003; 117: 170–174.

684. Sajantila, A. et al., Experimentally observed germline mutations at human micro- and minisatellite loci. *Eur. J. Hum. Genet.*, 1999; 7: 263–266.

685. Saks, M.J. et al., Context effects in forensic science: a review and application of the science of science to crime laboratory practice in the United States. *Sci. Justice*, 2003; 43: 77–90.

686. Salmon, C., Commentary — DNA evidence. *Crim. Law Rev.*, 1997: 669–673.

687. Sanger, F. et al., DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, 1977; 74: 5463–5467.

688. Sans, M., Admixture studies in Latin America: from the 20th to the 21st century. *Hum. Biol.*, 2000; 72: 155–177.

689. Sans, M. et al., Unequal contributions of male and female gene pools from parental populations in the African descendants of the city of Melo, Uruguay. *Am. J. Phys. Anthropol.*, 2002; 118: 33–44.

690. Santos, F.R. et al., Genetic and population study of a Y-linked tetranucleotide repeat DNA polymorphism with a simple non-isotopic technique. *Hum. Genet.*, 1993; 90: 655–656.

691. Sawyer, S. et al., Fingerprinting loci do show population differences: comments on Budowle et al., *Am. J. Hum. Genet.*, 1996; 59: 272–274.

692. Saxena, R. et al., Four DAZ genes in two clusters found in AZFc region of human Y chromosome. *Genomics*, 2000; 67: 256–267.

693. Schlötterer, C. et al., High mutation rates of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. *Mol. Biol. Evol.*, 1998; 15: 1269–1274.

694. Schneider, P.M. and Martin, P.D., Criminal DNA databases: the European situation. *Forensic Sci. Int.*, 2001; 119: 232–238.

695. Schneider, P.M. et al., Tandem repeat structure of the duplicated Y chromosomal STR locus DYS385 and frequency studies in the German and three Asian populations. *Forensic Sci. Int.*, 1998; 97: 61–70.

696. Schneider, P.M. et al., STR analysis of artificially degraded DNA — results of a collaborative European exercise. *Forensic Sci. Int.*, 2004; 139: 123–134.

697. Schneider, S. et al., Arlequin: a software for population genetic data analysis. 1997; Genetics and Biometry Laboratory, Geneva, Switzerland.

698. Schultz, M.M. and Reichert, W., A strategy for STR-analysis of cryptic epithelial cells on several textiles in practical casework. In *Progress in Forensic Genetics*, Vol. 8, Sensabaugh, G.F., Lincoln, P.J., and Olaisen, B., Eds. 2000; Elsevier: Amsterdam, p. 514–516.

699. Schwartz, M. and Vissing, J., Paternal inheritance of mitochondrial DNA. *N. Engl. J. Med.*, 2002; 347: 576–580.

700. Schweitzer, R.R., Anastasia and Anna Anderson. *Nature Genet.*, 1995; 9: 345.

701. Scranage, J. and Pinchin, R.A., DNASYS. 1995; Birmingham.

702. Scruggs, R. et al., Report to the Deputy Attorney General on the Events at Waco, Texas. 28 February to 19 April 1993; Available at http://www.usdoj.gov/05publications/waco/wacotocpg.htm (accessed 17 March 2004).

703. Scutt, J., Beware of new technologies. *Leg. Service Bull.*, 1990; 15: 9–12.

704. Segal, M.R. et al., Comparing DNA fingerprints of infectious organisms. *Stat. Sci.*., 2000; 15: 27–45.

705. Seielstad, M.T. et al., Genetic evidence for a higher female migration rate in humans. *Nature Genet.*, 1998; 20: 278–280.

706. Seton, C., Life for sex killer who sent decoy to take genetic test. *The Times*, 1988; London.

707. Sharma, V. and Litt, M., Tetranucleotide repeat polymorphism at the D21S11 locus. *Hum. Mol. Genet.*, 1992; 1: 67.

708. Shepherd, C. et al., Y STR haplotype data for New Zealand population groups using the Y-Plex™ 6 kit. *Forensic Sci. Int.*, 2004; 145(1): 69–72.

709. Shinozaki, N. and Aoki, H., Inbreeding in Japan: results of a nationwide study. *Jpn. J. Hum. Genet.*, 1975; 20: 91–107.

710. Shitara, H. et al., Maternal inheritance of mouse mtDNA in interspecific hybrids: segregation of the leaked paternal mtDNA followed by prevention of subsequent paternal leakage. *Genetics*, 1998; 148: 851–857.

711. Shitara, H. et al., Selective and continuous elimination of mitochondria microinjected into mouse eggs from spermatids, but not from liver cells, occurs throughout embryogenesis. *Genetics*, 2000; 156: 1277–1284.

712. Shoemaker, J. et al., A Bayesian characterization of Hardy–Weinberg disequilibrium. *Biometrics*, 1998; 25: 235–254.

713. Sibille, I. et al., Y-STR DNA amplification as biological evidence in sexually assaulted female victims with no cytological detection of spermatozoa. *Forensic Sci. Int.*, 2002; 125: 212–216.

714. Siguroardottir, S. et al., The mutation rate in the human mtDNA control region. *Am. J. Hum. Genet.*, 2000; 66: 1599–1609.

715. Sinha, S.K. et al., Development and validation of the Y-Plex™ 5, a Y chromosome STR genotyping system for forensic casework. *J. Forensic Sci.*, 2003; 48: 985–1000.

716. Skaletsky, H. et al., The male specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, 2003; 423: 825–837.

717. Skinner, M. et al., Guidelines for international bio-archaeology monitors of mass grave exhumations. *Forensic Sci. Int.*, 2003; 134: 81–92.

718. Smith, B.C. et al., A systematic approach to the sampling of dental DNA. *J. Forensic Sci.*, 1993; 38: 1194–1209.

719. Smith, C.A.B., A note on testing the Hardy–Weinberg Law. *Ann. Hum. Genet.*, 1970; 33: 377–383.

720. Snabes, M.C. et al., Preimplantation single-cell analysis of multiple genetic loci by whole genome amplification. *Proc. Natl. Acad. Sci. USA*, 1994; 91: 6181–6185.

721. Son, J.Y. et al., Polymorphism of nine X chromosomal STR loci in Koreans. *Int. J. Leg. Med.*, 2002; 116: 317–321.

722. Sparkes, R. et al., The validation of a 7-locus multiplex STR test for use in forensic casework (II). Mixtures, ageing, degradation and species studies. *Int. J. Leg. Med.*, 1996; 109: 195–204.

723. Sparkes, R. et al., The validation of a 7-locus multiplex STR test for use in forensic casework (I). Mixtures, ageing, degradation and species studies. *Int. J. Leg. Med.*, 1996; 109: 186–194.

724. Spence, L. et al., Forensic science targeting volume crime recidivists. *Forensic Bull.*, 2003; June: 26–28.

725. Spier, P., Recidivism Patterns for People Convicted in 1995. Research Report of the New Zealand Ministry of Justice: Wellington, 2001.

726. St John, J. et al., Failure of elimination of paternal mitochondrial DNA in abnormal embryos. *Lancet*, 2000; 355: 200.

727. Sterne, J.A.C., Teaching hypothesis tests — time for a significant change? *Stat. Med.*, 2002; 21: 985–994.

728. Stockmarr, A., Likelihood ratios for evaluating DNA evidence when the suspect is found through a database search. *Biometrics*, 1999; 55: 671–677.

729. Stockmarr, A., The choice of hypotheses in the evaluation of DNA profile evidence. In *Statistical Science in the Courtroom*, Gastwirth, J.L., Ed. 2000; Springer-Verlag: New York, p. 143–159.

730. Stockmarr, A., Author's reply. *Biometrics*, 2001; 57: 978–980.

731. Stone, A.C. et al., Mitochondrial DNA analysis of the presumptive remains of Jesse James. *J. Forensic Sci.*, 2001; 46: 173–176.

732. Stoneking, M., Hypervariable sites in the mtDNA control region are mutational hotspots. *Am. J. Hum. Genet.*, 2000; 67: 1029–1032.

733. Stoneking, M. et al., Establishing the identity of Anna Anderson Manahan. *Nature Genet.*, 1995; 9: 9–10.

734. Stoney, D.A., Letters to the Editor — source individuality versus expressed individuality — discussion of "probability analysis and the evidential value of bolt arrangements." *J. Forensic Sci.*, 1989; 34: 1295–1296.

735. Stoney, D.A., What made us ever think we could individualize using statistics. *J. Forensic Sci. Soc.*, 1991; 31: 197–199.

736. Stoney, D.A., Relaxation of the assumption of relevance and an application to one-trace and two-trace problems. *J. Forensic Sci. Soc.*, 1994; 34: 17–21.

737. Strom, C.M., More on DNA typing dispute. Correspondence. *Nature*, 1995; 373: 98–99.

738. Strom, C.M. and Rechitsky, S., Use of nested PCR to identify charred human remains and minute amounts of blood. *J. Forensic Sci.*, 1998; 43: 696–700.

739. Strom, C.M. et al., Reliability of polymerase chain reaction (PCR) analysis of single cells for preimplantation genetic analysis. *J. Assist. Reprod. Genet.*, 1994; 11: 55–62.

740. Sturtevant, A.H., The linear arrangement of six sex-linked factors in *Drosophilia* as shown by their mode of association. *J. Exp. Zool.*, 1913; 14: 43–59.

741. Sullivan, K.M. et al., A rapid and quantitative DNA sex test fluorescence-based PCR analysis of X–Y homologous gene amelogenin. *BioTechniques*, 1994; 15: 636–641.

742. Sullivan, K.M. et al., A single difference in mtDNA control region sequence observed between hair shaft and reference samples from a single donor. In *Proceedings from the Seventh International Symposium on Human Identification*, Promega Corporation. 1996; Scottsdale, AZ.

743. Sun, G. et al., Global genetic variation at nine short tandem repeat loci and implications on (*sic*) forensic genetics. *Eur. J. Hum. Genet.*, 2003; 11: 39–49.

744. Sutherland, B. et al., Commentary on: Wickenheiser RA. Trace DNA: a review, discussion of theory, and application of the transfer of trace quantities of DNA through skin contact. *J. Forensic Sci.* 2002; 47: 442–450; *J. Forensic Sci.*, 2003; 48: 467.

745. Suzuki, H. et al., Existence of common homologous elements in the transcriptional regulatory regions of human nuclear genes and mitochondrial gene for the oxidative phosphorylation system. *J. Biol. Chem.*, 1991; 266: 2333–2338.

746. Suzuki, H. et al., Common protein-binding sites in the 5′-flanking regions of human genes for cytochrome c1 and ubiquinone-binding protein. *J. Biol. Chem.*, 1990; 265: 8159–8163.

747. Sykes, B., *The Seven Daughters of Eve*. 2001; W.W. Norton and Company: London.

748. Sykes, B. et al., The origins of the Polynesians: an interpretation from mitochondrial lineage analysis. *Am. J. Hum. Genet.*, 1995; 57: 1463–1475.

749. Sykes, B.C., *Adam's Curse*. 2003; Bantam Press: London.

750. Szibor, R. et al., Use of X-linked markers for forensic purposes. *Int. J. Leg. Med.*, 2003; 117: 67–74.

751. Taberlet, P. et al., Reliable genotyping of samples with very low DNA quantities using PCR. *Nucl. Acids Res.*, 1996; 24: 3189–3194.

752. Tahir, M.A. and Watson, N., Typing using DNA HLA-DQa alleles extracted from human nail material using polymerase chain reaction. *J. Forensic Sci.*, 1995; 40: 634–636.

753. Takayama, T. et al., Quality and quantity of DNA in cadavers' serum. *Leg. Med.*, 2003; 5: 180–182.

754. Taroni, F. and Aitken, C.G.G., Probabilistic reasoning and the law. Part 1: assessment of probabilities and explanation of the value of DNA evidence. *Sci. Justice*, 1998; 38: 165–177.

755. Taroni, F. and Aitken, C.G.G., Probabilistic reasoning and the law. Part 2: assessment of probabilities and explanation of the value of trace evidence other than DNA. *Sci. Justice*, 1998; 38: 179–188.

756. Taroni, F. et al., De Finetti's subjectivism, the assessment of probabilities and the evaluation of evidence: a commentary for forensic scientists. *Sci. Justice*, 2001; 3: 145–150.

757. Taroni, F. et al., A general approach to Bayesian networks for the interpretation of evidence. *Forensic Sci. Int.*, 2004; 139: 5–16.

758. Tereba, A., Tools for analysis of population statistics. *Profiles in DNA*, 1999; 2: 14–16.

759. Thagard, P., Why wasn't O.J. convicted? Emotional coherence in legal inference. *Cognition Emotion*, 2003; 17: 361–383.

760. The President's initiative to advocate justice through DNA technology. Available at http://www.USDOJ.GOV (accessed 30 April 2003).

761. Thompson, E.A., Gene identities and multiple relationships. *Biometrics*, 1974; 30: 667–680.

762. Thompson, E.A., Two-locus and three-locus gene identity by descent in pedigrees. *J. Math. Appl. Med. Biol.*, 1988; 5: 261–279.

763. Thompson, E.A., *Statistical Inference from Genetic Data on Pedigrees*. 2000; Institute of Mathematical Statistics: Beachwood, OH.

764. Thompson, W.C., Are juries competent to evaluate statistical evidence? *Law Contemp. Prob.*, 1989; 52: 9–41.

765. Thompson, W.C., Comment on Roeder, K (1994). DNA fingerprinting: a review of the controversy. *Stat. Sci.*, 1994; 9: 263–266.

766. Thompson, W.C., Subjective interpretation, laboratory error and the value of forensic DNA evidence: three case studies. In *Human Identification: The Use of DNA Markers*, Weir, B.S., Ed. 1995; Kluwer Academic Publishers: Dordrecht, p. 153–168.

767. Thompson, W.C., DNA evidence in the O.J. Simpson trial. *Univ. Colorado Law Rev.*, 1996; 67: 827–857.

768. Thompson, W.C., Accepting lower standards: the National Reseach Council's second report on forensic DNA evidence. *Jurimetrics*, 1997; 37: 405–424.

769. Thompson, W.C. and Schumann, E.L., Interpretation of statistical evidence in criminal trials — the prosecutor's fallacy and the defence attorney's fallacy. *Law Hum. Behav.*, 1987; 11: 167–187.

770. Thompson, W.C. et al., How the probability of a false positive affects the value of DNA evidence. *J. Forensic Sci.*, 2003; 48: 1–8.

771. Thompson, W.C. et al., Authors' response. *J. Forensic Sci.*, 2003; 48: 1201.

772. Thomson, J.A. et al., Analysis of disputed single-parent/child and sibling relationships using 16 STR loci. *Int. J. Leg. Med.*, 2001; 115: 128–134.

773. Tilford, C. et al., A physical map of the human Y chromosome. *Nature*, 2001; 409: 943–945.

774. Tippett, C.F. et al., Paint flakes from sources other than vehicles. *J. Forensic Sci. Soc.*, 1965; 8: 61–65.

775. Torres, Y. et al., DNA mixtures in forensic casework: a 4-year retrospective study. *Forensic Sci. Int.*, 2003; 134: 180–186.

776. Trabetti, E. et al., Application of a Y-linked tetranucleotide repeat polymorphism to paternity cases. In *Advances in Forensic Haemogenetics*, Vol. 5, Rossi, U., Ed. 1994; Springer: Berlin.

777. Triggs, C. and Buckleton, J., Unpublished results, 1996.

778. Triggs, C. and Patel, J., personal communication, 2001.

779. Triggs, C. et al., Assessing probabilities considering eyewitness evidence. *Sci. Justice*, 1995; 35: 263–266.

780. Triggs, C.M., personal communication, 2003.

781. Triggs, C.M. and Buckleton, J.S., Logical implications of applying the principles of population genetics to the interpretation of DNA profiling evidence. *Forensic Sci. Int.*, 2002; 128: 108–114.

782. Triggs, C.M. and Buckleton, J., The two-trace transfer problem revisited. *Sci. Justice*, 2003; 43: 127–134.

783. Triggs, C.M. and Buckleton, J.S., Comment on "Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence," Meester, R. and Sjerps, M. *Law Prob. & Risk* 2004; 3: 73–82.

784. Triggs, C.M. et al., The calculation of DNA match probabilities in mixed race populations. *Sci. Justice*, 2000; 40: 33–38.

785. Tully, G. and Lareu, M.V., Controversies over heteroplasmy. *Electrophoresis*, 2001; 22: 180–182.

786. Tully, G. et al., Considerations of the European DNA Profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. *Forensic Sci. Int.*, 2001; 124: 83–91.

787. Tully, L.A. et al., A sensitive denaturing gradient-gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mitochondrial DNA control region. *Am. J. Hum. Genet.*, 2000; 67: 432–443.

788. Tyler-Smith, C. and McVean, G., The comings and goings of a Y polymorphism. *Nature Genet.*, 2003; 35: 201–202.

789. Ubelaker, D.H. et al., The role of forensic anthropology in the recovery and analysis of Branch Dravidian compound victims: recovery procedures and characteristics of the victims. *J. Forensic Sci.*, 1995; 40: 335–340.

790. Uchida, S. et al., Utility of fingernail DNA for evaluation of chimerism after bone marrow transplantation and for diagnostic testing for transfusion-associated graft-versus-host disease. *Blood*, 1996; 87: 4015–4016.

791. Uchihi, R. et al., Deoxyribonucleic acid (DNA) typing of human leukocyte antigen (HLA)-DQA1 from single hairs in Japanese. *J. Forensic Sci.*, 1992; 37: 853–859.

792. Urquhart, A. et al., Variation in short tandem repeat sequences — a survey of twelve microsatellite loci for use as forensic identification markers. *Int. J. Leg. Med.*, 1994; 107: 13–20.

793. Urquhart, A.J. et al., Sequence variability of the tetranucleotide repeat of the human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2) locus. *Hum. Genet.*, 1993; 92: 637–638.

794. van Hoofstst, D.E.O. et al., DNA typing of fingerprints and skin debris: sensitivity of capillary electrophoresis in forensic applications using multiplex PCR. In *Second European Symposium of Human Identification*, Promega Corporation. 1998; Innsbruck, Austria.

795. Van Oorschot, R.A. and Jones, M., DNA fingerprints from fingerprints. *Nature*, 1997; 387: 767.

796. Van Oorschot, R.A. and Jones, M., Retrieval of DNA from touched objects. In *The Fourteenth International Symposium for Forensic Sciences*. 1998; Adelaide, Australia.

797. van Rentergeum, P. et al., Use of latent fingerprints as a source of DNA for genetic identification. In *Progress in Forensic Genetics*, Vol. 8, Sensabaugh, G.F. et al., Eds. 2000; Elsevier: Amsterdam, p. 501–503.

798. Varlaro, J. and Duceman, B., Dealing with increasing casework demands for DNA analysis. *Profiles in DNA*, 2002; 5: 3–6.

799. Vastag, B., Out of tragedy, identification innovation. *JAMA*, 2002; 288: 1221–1223.

800. Veth, J., The Characterisation of Artefacts found in DNA Profiles. MSc thesis, Forensic Science, University of Auckland, New Zealand, 2002.

801. Wahlund, S., Zuzammensetzung von populationen und korrelationserscheinungen vom standpunkt der vererbungslehre aus betrechtet. *Hereditas*, 1928; 11: 65–106.

802. Wakeley, J., Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. *J. Mol. Evol.*, 1993; 37: 613–623.

803. Walker, R.H. et al., Eds. *Inclusion Probabilities in Parentage Testing*. 1983; American Association of Blood Banks: Arlington, VA.

804. Walsh, J.J., Canada's proposed forensic DNA evidence bank. *Can. Soc. Forensic Sci. J.*, 1998; 31: 113–125.

805. Walsh, K. and Buckleton, J.S., A discussion of the law of mutual independence and its application to blood-group frequency data. *J. Forensic Sci. Soc.*, 1988; 28: 95–98.

806. Walsh, K.A.J. and Buckleton, J.S., Calculating the frequency of occurrence of a blood type for a "random man." *J. Forensic Sci. Soc.*, 1991; 31: 49–58.

807. Walsh, K.A.J. et al., Assessing prior probabilities considering geography. *J. Forensic Sci. Soc.*, 1994; 34: 47–51.

808. Walsh, P.S. et al., Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. *Nucl. Acids Res.*, 1996; 24: 2807–2812.

809. Walsh, S., Commentary on: Kline MC, Jenkins B, Rogers, S. *J. Forensic Sci.*, 1998; 43: 1103.

810. Walsh, S.J., Legal perceptions of forensic DNA profiling. Part I: A review of the legal literature. *Forensic Sci. Int.,* 2004 (in press).

811. Walsh, S.J. et al., The collation of forensic DNA case data into a multi-dimensional intelligence database. *Sci. Justice*, 2002; 42: 205–214.

812. Walsh, S.J. et al., Genetic diversity among the people of Aotearoa, New Zealand. 2002; U.K. Forensic Science Service Report RR 823, January 2002.

813. Walsh, S.J. et al., DNA profiling and criminal justice — a contribution to a changing debate. *Aust. J. Forensic Sci.*, 36: 34–43.

814. Walsh, S.J. et al., Evidence in support of self-declaration as a sampling method for the formation of sub-population databases. *J. Forensic Sci.*, 2003; 48: 1091–1093.

815. Walsh, S.J. et al., Characterisation of variant alleles at the HumD21S11 locus implies unique Australasian genotypes and re-classification of nomenclature guidelines. *Forensic Sci. Int.*, 2003; 135: 35–41.

816. Wambaugh, J., *The Blooding*. 1989; William Morrow and Company: New York.

817. Wan, Q.-H. and Fang, S.-G., Application of species-specific polymerase chain reaction in the forensic identification of tiger species. *Forensic Sci. Int.*, 2003; 131: 75–78.

818. Warlow, T.A., Ballistics examination of British citizens from the Waco siege. *Sci. Justice*, 1998; 38: 255–259.

819. Warne, D.C. et al., Tetranucleotide repeat polymorphism at the human beta actin replated pseudogene 2 (ACTBP2) detected using the polymerase chain reaction. *Nucl. Acids Res.*, 1991; 19: 6980.

820. Wasik, M., Legislating in the shadow of the Human Rights Act: the Criminal Justice and Police Act 2001. *Crim. Law J.*, 2001; 931–947.

821. Watanabe, Y. et al., DNA typing from cigarette butts. *Leg. Med.*, 2003; 5: 177–179.

822. Weale, M.E. et al., Y chromosome evidence for Anglo-Saxon mass migration. *Mol. Biol. Evol.*, 2002; 19: 1008–1021.

823. Weber, J.L. and Wong, C., Mutation of human short tandem repeats. *Hum. Mol. Genet.*, 1993; 2: 1123–1128.

824. Weetman, D. et al., Reconstruction of microsatellite mutation history reveals a strong and consistent deletion bias in invasive clonal snails, *Potamopyrgus antipodarum*. *Genetics*, 2002; 162: 813–822.

825. Wei, C.-C. et al., The spectrum of microsatellite loci on chromosomes 7 and 8 in Taiwan aboriginal populations: a comparative population genetic study. *Hum. Genet.*, 1999; 104: 333–340.

826. Weinberg, W., Uber den Nackweis der Vererbung beim Menschen. Jahresh. Verein f. Varerl *Naturk. Wurtemb.*, 1908; 64: 368–382.

827. Weir, B.S., personal communication, 2001.

828. Weir, B.S., Independence tests for VNTR alleles defined as quantile bins. *Am. J. Hum. Genet.*, 1993; 53: 1107–1113.

829. Weir, B.S., Population genetics in the forensic DNA debate. *Proc. Natl. Acad. Sci. USA*, 1992; 89: 11654–11659.

830. Weir, B.S., The status of DNA fingerprinting. *The World and I*, 1993; 214–219.

831. Weir, B.S., Forensic population genetics and the National Research Council (NRC). *Am. J. Hum. Genet.*, 1993; 52: 437.

832. Weir, B.S., The effects of inbreeding on forensic calculations. *Annu. Rev. Genet.*, 1994; 28: 597–621.

833. Weir, B.S., DNA profiling on trial — correspondence. *Nature*, 1994; 369: 351.

834. Weir, B.S., Comment in Roeder, K (1994). DNA fingerprinting: a review of the controversy. *Stat. Sci.*, 1994; 9: 266–267.

835. Weir, B.S., DNA statistics in the Simpson matter. *Nature Genet.*, 1995; 11: 365–368.

836. Weir, B.S., *Genetic Data Analysis*, Vol. II. 1996; Sinauer: Sunderland, MA.

837. Weir, B.S., Life & hard times of a statistician. *Stats*, 1996; 16: 15–17.

838. Weir, B.S. Court experiences in the USA: People v. Simpson. In *First International Conference on Forensic Human Identification in the Millennium*. 1999; London.

839. Weir, B.S., What is the structure of human populations? In *Evolutionary Biology*, Clegg, M.T. et al., Eds. 2000; Kluwer Academic: London.

840. Weir, B.S., DNA match and profile probabilities: comment on Budowle et al. (2000) and Fung and Hu (2000). *Forensic Sci. Comm.*, 2001; 3: Available at http://www.fbi.gov/programs/lab/fsc/current/weir.htm.

841. Weir, B.S., Interpreting DNA course notes. In NCSU Summer School in Statistical Genetics. 2001; Raleigh, North Carolina.

842. Weir, B.S., Forensics. In *Handbook of Statistical Genetics*, Cannings, C., Ed. 2001; John Wiley and Sons, Ltd.: Chichester.

843. Weir, B.S., Matching and partially-matching DNA profiles. 2003; Unpublished.

844. Weir, B.S., Summer School in Statistical Genetics course notes. 2003; Unpublished.

845. Weir, B.S., personal communication, 2003.

846. Weir, B.S. and Cockerham, C.C., Estimating F-statistics for the analysis of population structure. *Evolution*, 1984; 38: 1358–1370.

847. Weir, B.S. and Evett, I.W., Whose DNA? *Am. J. Hum. Genet.*, 1992; 50: 869.

848. Weir, B.S. et al., Assessing sampling error in DNA evidence. Personal communication, 2000.

849. Weir, B.S. et al., Interpreting DNA mixtures. *J. Forensic Sci.*, 1997; 42: 213–222.

850. Weitz, C. and van Niekerk, J., Complete robotic automation of a forensic DNA typing facility. *Insights*, 2002; 13: 11.

851. Werrett, D.J., The national DNA database. *Forensic Sci. Int.*, 1997; 88: 33–42.

852. Werrett, D.J. and Sparkes, R.L., 300 matches per week — the effectiveness and future development of DNA intelligence databases — Parts 1 and 2. In *Ninth International Symposium on Human Identification*. 1998; Promega Corporation, Madison, WI.

853. Werrett, D.J. and Sullivan, K., The National DNA Database of England and Wales and the strategic use of DNA profiling. In *The First International Conference on Forensic Human Identification in the Millennium*. 1999; London.

854. Whitaker, J.P. et al., A comparison of the characteristics of profiles produced with the AMPF*l*STR SGM Plus multiplex system for both standard and low copy number (LCN) STR DNA analysis. *Forensic Sci. Int.*, 2001; 123: 215–223.

855. White, P.S. et al., New, male-specific microsatellite markers from the human Y-chromosome. *Genomics*, 1999; 57: 433–437.

856. Whittaker, M., personal communication, 1996.

857. Whittle, M.R. et al., Updated Brazilian genetic data, together with mutation rates on 19 STR loci, including D10S1237. *Forensic Sci. Int.*, 2004; 139: 207–210.

858. Wickenhieser, R.A., Trace DNA: a review, discussion of theory, and application of the transfer of trace quantities of DNA through skin contact. *J. Forensic Sci.*, 2002; 47: 442–450.

859. Wiegand, P. and Kleiber, M., DNA typing of epithelial cells after strangulation. *Int. J. Leg. Med.*, 1997; 110: 181–183.

860. Wiegand, P. and Kleiber, M., DNA typing of epithelial cells. In *Progress in Genetics*, Olaisen, B. et al., Eds. 1998; Elsevier: Amsterdam, p. 165–167.

861. Wiegand, P. et al., STR typing of biological stains on strangulation tools. In *Progress in Forensic Genetics*, Vol. 8, Sensabaugh, G.F., Lincoln, P.J., and Olaisen, B., Eds. 2000; Elsevier: Amsterdam, p. 508–510.

862. Wiegand, T. et al., DNA typing of debris from fingernails. *Int. J. Leg. Med.*, 1993; 106: 81–83.

863. Wierdl, M. et al., Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics*, 1997; 146: 769–779.

864. Wiersema, S., Is the Bayesian approach for you? Presented at the *European Meeting for Shoeprint/Toolmark Examiners*. Berlin, 15–18 May 2001.

865. Wild, C.J. and Seber, G.A.F., *Chance Encounters*. 2000; John Wiley and Sons: New York.

866. Wilkinson-Herbots, H.M. et al., Site 73 in hypervariable region II of the human mitochondrial genome and the origin of European populations. *Ann. Hum. Genet.*, 1996; 60: 499–508.

867. Williams, E.D. and Crews, J.D., From dust to dust: ethical and practical issues involved in the location, exhumation, and identification of bodies from mass graves. *Croat. Med. J.*, 2003; 44: 251–258.

868. Wilson, M.R. et al., A family exhibiting heteroplasmy in the human mitochondrial DNA control region reveals both somatic mosaicism and pronounced segregation of mitotypes. *Hum. Genet.*, 1997; 100: 167–171.

869. Wolanska-Nowak, P., Population differentiation and the value of DNA evidence. *Z Zagadnien Nauk Sadowych*, 1999; 39: 103–121.

870. Wolanska-Nowak, P. et al., STR data for SGM Plus and penta E and D loci in a population sample from south Poland. *Forensic Sci. Int.*, 2002; 127: 237–239.

871. Wolfe, K., In *Gene Structure in Eukaryotic Microbes*, Kinghorn, J.R., Ed. 1987; IRL: Oxford.

872. Wright, S., Systems of mating II. *Genetics*, 1921; 6: 124–143.

873. Wright, S., The genetical structure of populations. *Ann. Eugen.*, 1951; 15: 159–171.

874. Wurmb-Schwark, N.V. et al., Mitochondrial mutagenesis in the brain in forensic and pathological research. *Leg. Med.*, 2003; 5: 1–6.

875. XBAIES download page. Avaliable at http://www.staff.city.ac.uk/~rgc/webpages/xbpage.html (accessed 23 February 2004).

876. Xu, X. et al., The direction of microsatellite mutations is dependent upon allele length. *Nature Genet.*, 2000; 24: 396–399.

877. Yoneda, M. et al., Marked replicative advantage of human mtDNA carrying a point mutation that causes the MELAS encephalomyopathy. *Proc. Natl. Acad. Sci. USA*, 1992; 89: 11164–11168.

878. Yule, G.U., Mendel's laws and their probable relation to inter-racial heredity. *New Phytol.*, 1902; 1: 192–207, 222–238.

879. Zamble, F.E. and Quinsey, G., *The Criminal Recidivism Process*. 1997; Chicago University Press: Chicago.

880. Zaykin, D. et al., Exact tests for association between alleles at arbitrary numbers of loci. *Genetica*, 1995; 96: 169–178.

881. Zaykin, D. et al., Truncated product method for combining *P*-values. *Genet. Epidemiol.*, 2002; 22: 170–85.

882. Zhang, K. et al., Randomly distributed crossovers may generate block-like patterns of linkage disequilibrium: an act of genetic drift. *Hum. Genet.*, 2003; 113: 51–59.

883. Zhivotovsky, L.A., Recognition of the remains of Tsar Nicholas II and his family: a case of premature identification. *Ann. Hum. Biol.*, 1999; 26: 569–577.

884. Zhivotovsky, L.A. et al., The forensic DNA implications of genetic differentiation between endogamous communities. *Forensic Sci. Int.*, 2001; 119: 269–272.

885. Zhivotovsky, L.A. et al., The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am. J. Hum. Genet.*, 2004; 74: 50–61.

886. Zietkiewicz, E. et al., Nuclear DNA diversity in worldwide distributed human populations. *Genetica*, 1997; 205: 161–171.