The application of statistical modelling to the interpretation of complex DNA profiles

Hannah Mary Kelly

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Forensic Science, The University of Auckland, 2013.

Abstract

In forensic DNA analysis a profile is typically produced from a biological sample collected from the scene of a crime and compared with the DNA of one or more persons of interest (POI). Single source pristine profiles are relatively easy to interpret and their analysis has achieved worldwide acceptance as a reliable scientific method. However, profiles from crime scenes are frequently compromised in quality, or quantity or both. Stochastic factors are often present in compromised profiles which complicate interpretation. Stochastic factors can include; heterozygous balance, allelic dropout, and increased stutter peaks and are characteristic of low template DNA (LtDNA) samples. Complicating interpretation even further is that in many cases, crime scene samples are composed from two or more people. The number of contributors can be unclear. The presence of three or more alleles at any locus signals the existence of more than one contributor, although it can be difficult to distinguish between the presence of a low level second contributor and stochastic effects.

This research investigates the behaviour of LtDNA profiles. The traditional guidelines, used in conventional DNA interpretation, are investigated with respect to their application to LtDNA profiles. Statistical models are created for; heterozygous balance, dropout, and stutter. These models use the explanatory variables identified in the data exploratory section of this research to describe the behaviour of each of the aforementioned stochastic effects. These models have been built with the aim that they will be implemented in a probabilistic approach to DNA interpretation.

This research also examines different approaches for the interpretation of forensic DNA profiles and the available methods for the calculation of the weight of the evidence of the profile. A two person LtDNA mixture is interpreted using different methods and the resulting statistics are presented. This is done with the aim of demonstrating how the same LtDNA evidence can be interpreted differently under current interpretation guidelines. It is important that that limits of current interpretation models are understood and are not extended beyond their means.

Acknowledgments

A PhD is so much more than a solo endeavour and there is no way that I ever could have dreamed of completing this without the assistance of a team of helpers.

To my supervisors John Buckleton and James Curran: thank you so much for your willingness to take me on.

John; thank you for offering me this topic, thank you for your ideas, and thank you for your enthusiasm and commitment to helping others to succeed. There wouldn't have been a PhD without you and I will be forever grateful for your help.

James; thank you for taking on a non-statistics student - it must at times have driven you crazy but thank you for putting up with my questions without any eye-rolling. Thank you for taking the time to explain things to me even though they must have seemed basic to you. Your patience, R skills and proof reading expertise have been extremely valuable.

Jo - thank you for all your help. Your experience and knowledge with DNA interpretation was an incredible asset to this research. Thank you so much for taking me under your wing and helping me in so many ways. I am incredibly lucky to have had you on my team.

ESR - thank you for having me. It has been amazing to have access to such good quality people and I have really enjoyed my time with you. Thank you for awarding me with the Institute of Environmental Science and Research Doctoral Scholarship, I wouldn't have been able to do this research without it.

The University of Auckland - thank you for awarding me with the University of Auckland Doctoral Scholarship. The financial assistance was much appreciated.

The Department of Statistics - thank you for housing me and for the use of the laptop. You are a great department and you made my life a lot easier.

Thank you to Mum and Dad for putting up with a perpetual student for a daughter. Your support has been boundless. Thanks for the care packages, thanks for listening and thanks for just being there when I needed. This is the end of the student road - for now anyway.

Lastly - thanks to my friends. Thanks for understanding I couldn’t always be around for your good times, but thanks for being there for my bad times nonetheless. Thanks for cooking for me, entertaining me, putting up with me, and for the inspirational quotes. You are all awesome.

Table of Contents

THE UNIVERSITY OF AUCKLAND
SCHOOL OF GRADUATE STUDIES

Co-Authorship Form

Graduate Centre
ClockTower – East Wing
22 Princes Street, Auckland
Phone: +64 9 373 7599 ext 81321
Fax: +64 9 373 7610
Email: postgraduate@auckland.ac.nz
www.postgrad.auckland.ac.nz

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Abstract.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 7: Modelling heterzygote balance in forensic DNA profiles

Nature of contribution by PhD candidate	Data analysis, statistical model construction, writing
Extent of contribution by PhD candidate (%)	60%

CO-AUTHORS

Name	Nature of Contribution
Jo-Anne Bright	Editing and data collation
John Buckleton	Editing
James Curran	Editing, help with statistical model construction

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- in cases where the PhD candidate was the lead author of the work that the candidate wrote the text.

Name	Signature	Date
Jo-Anne Bright		27.03.2013
John Buckleton		27.03.2013
James Curran		28.03.2013

THE UNIVERSITY OF AUCKLAND
SCHOOL OF GRADUATE STUDIES

Co-Authorship Form

Graduate Centre
ClockTower – East Wing
22 Princes Street, Auckland
Phone: +64 9 373 7599 ext 81321
Fax: +64 9 373 7610
Email: postgraduate@auckland.ac.nz
www.postgrad.auckland.ac.nz

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Abstract.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work. Chapter 8: Identifying and modelling the drivers of stutter in forensic DNA profiles	
Nature of contribution by PhD candidate	data analysis, statistical model contruction, writing
Extent of contribution by PhD candidate (%)	70%

CO-AUTHORS

Name	Nature of Contribution
Jo-Anne Bright	Editing and data analysis
John Buckleton	Editing
James Curran	Editing and help with statistical model construction

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- in cases where the PhD candidate was the lead author of the work that the candidate wrote the text.

Name	Signature	Date
Jo-Anne Bright		27.03.2013
John Buckleton		27.03.2013
James Curran		28.03.2012

THE UNIVERSITY OF AUCKLAND
SCHOOL OF GRADUATE STUDIES

Co-Authorship Form

Graduate Centre
ClockTower – East Wing
22 Princes Street, Auckland
Phone: +64 9 373 7599 ext 81321
Fax: +64 9 373 7610
Email: postgraduate@auckland.ac.nz
www.postgrad.auckland.ac.nz

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Abstract.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 4: The interpretation of low level mixtures

Nature of contribution by PhD candidate	writing, formulation of equations, interpretation
Extent of contribution by PhD candidate (%)	70%

CO-AUTHORS

Name	Nature of Contribution
Jo-Anne Bright	Editing and equation checking
John Buckleton	Editing and equation checking
James Curran	Editing

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- in cases where the PhD candidate was the lead author of the work that the candidate wrote the text.

Name	Signature	Date
Jo-Anne Bright		27.03.2013
John Buckleton		27.03.2013
James Curran		28.03.2013

THE UNIVERSITY OF AUCKLAND
SCHOOL OF GRADUATE STUDIES

Co-Authorship Form

Graduate Centre
ClockTower – East Wing
22 Princes Street, Auckland
Phone: +64 9 373 7599 ext 81321
Fax: +64 9 373 7610
Email: postgraduate@auckland.ac.nz
www.postgrad.auckland.ac.nz

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Abstract.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 3: A comparison of statistical models for the analysis of complex forensic DNA profiles

Nature of contribution by PhD candidate	writing
Extent of contribution by PhD candidate (%)	80%

CO-AUTHORS

Name	Nature of Contribution
Jo-Anne Bright	editing
John Buckleton	editing
James Curran	editing

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- in cases where the PhD candidate was the lead author of the work that the candidate wrote the text.

Name	Signature	Date
Jo-Anne Bright		27.03.2013
John Buckleton		27.03.2013
James Curran		28.03.2013

1: INTRODUCTION

In 1985 a ground breaking paper was published titled “Forensic application of DNA fingerprints” [1]. This paper delivered one of the single biggest advancements in forensic science in the 20th century. It detailed for the first time, a method for the identification of individuals using DNA profiling.

Any biological sample, including blood, semen, saliva, vaginal fluid and hair, is a potential source for DNA analysis. DNA is relatively stable and under optimal conditions does not change over time, allowing samples which were collected many years previously to be compared to samples collected recently [2].

Since 1985, DNA profiling has solved crimes, served as a legal base for opening old cases, freed wrongly convicted individuals, and is now considered the gold standard in the characterisation of forensic biological evidence. It has also undergone many stages of evolution.

1.1 The evolution of DNA profiling

The original method described by Gill et al., [1] was based upon the equally ground breaking paper “Individual-specific ‘fingerprints’ of human DNA” [3]. Jeffreys et at., [3] pioneered an approach which utilised tandem-repetitive DNA sequences or “minisatellites”. These minisatellites were made of relatively long repeating units of DNA, the number of repetitions differing between individuals. The minisatellites were flanked by a sequence of DNA containing a cleavage site that was recognized by a restriction enzyme and would result in the DNA being cut into fragments of varying lengths (a technique called restriction fragment length polymorphism or RFLP). When these fragments were analysed using gel electrophoresis, which separated them by size and then visualised using multilocus probes that hybridised many minisatellite loci at once, the result looked like a barcode – this gave rise to the term “DNA fingerprinting”.

Minisatellite analysis was usually restricted to violent and other serious crimes because it required a relatively large amount of good quality DNA, usually a visible blood or semen stain 25mm in diameter or larger. Additionally, the technique

contained a number of steps which made it time consuming. Interpretation of the resulting gels was also difficult [4].

The same RFLP technology was used in the next stage of DNA profiling except single locus probes were used in place of multilocus probes. This technique was called single-locus probes in New Zealand and the United Kingdom and variable number repeats (VNTR) in the United States of America. This may seem like a backwards step in the progress of DNA fingerprinting, but visualisation of a single locus at a time allowed analysts to determine if one or two alleles were present in the gel (i.e. if the contributor was a homozygote [5] or a heterozygote - assuming the second allele of a heterozygote had not run off the end of the gel). Therefore a statistic could be calculated for the evidential weight of a "match". However, the number of repeating units within the minisatellites was small compared to the minisatellite region plus the flanking sequence of DNA. The gel-based technology of the time, made it difficult to differentiate between alleles that differed by only one or two repeat units.

In the mid-1990's DNA profiling evolved to encompass the use of the polymerase chain reaction (PCR). In the words of Kary Mullis, the creator of PCR, "PCR makes life much easier for molecular biologists; it gives them as much of a particular DNA as they want" [6-8]. The inclusion of PCR in forensic DNA analysis meant that DNA profiles could now be generated from much smaller quantities of starting DNA template. The methods described previously required around 500 ng for a successful outcome whereas with the inclusion of PCR 1 ng or less could generate a profile.

The use of minisatellites was also phased out in favour of short tandem repeat (STR) loci. STR loci consist of repeated sequences of two to seven nucleotides that are generally less than 350 bp in total length. These are much smaller than minisatellites which ranged between one kb and 20 kb in size [9]. The ratio of the repeat unit to flanking sequence of DNA is smaller in STRs than minisatellites which means that the distance between alleles that differ by one repeat unit is improved. This made the differentiation of alleles easier when using resolution by polyacrylamide gel electrophorese (PAGE).

The first widely used STR based analysis technique was the "quadruplex" [10], a technique that analysed four STRs in one reaction, known as a multiplex [11]. The

discriminatory power of the quadruplex was not very strong so it was often coupled with single-locus probe profiling (VNTRs).

The quadruplex evolved with the addition of two more STRs and was renamed the second generation multiplex (SGM)[12]. The SGM also included a marker that targeted amelogenin, a sex specific gene so the sex of the contributor could now be determined. In 2000, the United Kingdom replaced SGM with SGM Plus which had an additional four loci, bringing the total to 10 loci plus amelogenin [13]. This meant the discriminatory power of the multiplex was much improved and the chance of an adventitious match was reduced. Automated detection of peaks was also introduced using capillary gel electrophoresis [14].

There is a wide choice of such loci for use in DNA analysis due to the abundance of STRs in the human genome. Selection for their use in forensic casework usually depends on their discriminating power, their structure and their ability to co-amplify with other STR loci in a single multiplex reaction. The multiplex currently in use in New Zealand is the AmpFℓSTR® Identifiler® PCR Amplification kit. This kit co-amplifies and simultaneously detects 15 internationally recognised STRs [15].

1.2 The interpretation of DNA

1.2.1 Generation of a statistic

As DNA profiling techniques have evolved so have the statistical methods used to evaluate the evidentiary value of the resulting DNA profile. The ultimate purpose of DNA profiling is to ascertain if two samples share a common source. This could be crime scene to person of interest or crime scene to crime scene. However, there are varying degrees of the strength of a match. For example, the number of loci included in the analysis will greatly affect the discriminatory power of the result. The more loci included in the analysis, the more discriminating the result will be. Therefore, without some measure of the strength of the evidence, the result of an examination can be misleading.

There are three general conclusions that can result from the analysis of a DNA profile;

1). An exclusion; The profiles are different and must have originated from different sources. No statistic is required to support the evidential value of an exclusion.

2). Inconclusive; It is not possible to determine from this analysis if the DNA samples share a common source. This may be due to a number of reasons including contamination of the sample(s), degradation of the sample(s), or inhibition of the amplification. No statistic is required to support the evidential value of an inconclusive result.

3). Inclusion; The profiles share the same alleles at the loci analysed and may have originated from the same source. A statistic is required to measure the strength of this result.

If two profiles share the same alleles this can arise because;

1) The two profiles share a common source, or
2) The two profiles coincidentally share the same alleles.

Therefore the "match statistic" or strength of the evidence takes into consideration the rarity of the alleles in a defined population. Initially, a standard way to measure the strength of an inclusion was to count the occurrence of the particular genotype in a random sample of the appropriate population. Then using classical statistical formulas, upper and lower confidence limits were placed on the estimate [16].

This method developed into using allelic frequencies instead of genotype frequencies. Allele frequencies refer to the rarity of the allele within a defined population and were calculated by the number of times the allele was observed in the appropriate database divided by the total number of alleles observed at that locus. The frequency of the genotype at each locus would then be calculated assuming Hardy Weinberg (HW) equilibrium [17].

HW equilibrium assumes random mating and can be stated symbolically. If we let p_i and p_j be the proportions of two alleles A_i and A_j within a population, then the HW expectations are [17]:

Homozygous genotypes: A_iA_i: = p_i^2

Heterozygous genotypes: $A_iA_j = 2p_ip_j$

The HW equilibrium is based on the laws of segregation and independent assortment [2]. These laws state that each gene pair behaves independently, and is therefore statistically independent. As a result, the frequency of each genotype at a locus can be multiplied across the profile to produce an overall estimate for frequency of the observed profile within a defined population. This is called the product rule [17].

A concern that arose regarding the use of the product rule questioned population sub-structure. Some people expressed serious concern that within defined census groups (such as Asian, Hispanic, North American Caucasian) there are subgroups and people tend to stay within their subgroups [18-21]. This means that allele frequencies aren't homogenized between the subgroups and the census group.

The need to account for possible population substructure lead to the 1992 recommendation by the US National Research Council (NRC) to apply the ceiling principle [16]. The ceiling principle used the allele frequency that was the largest, regardless of ethnicity. It also used upper confidence limits, subject to an overall minimum of 10% as an interim measure, which could be reduced to 5% after further research.The product rule was then used to calculate the frequency of the profile. The ceiling principle was deemed to be conservative because the largest allele frequency was used at each locus [16].

The recommendation of the use ceiling principle by the NCR generated wide criticism as the method was seen to be complicated and ad-hoc [22]. In 1994 Balding and Nichols proposed an alternative method called the "sampling formula", which in their words was "logically coherent and conservative while not understating the high discriminatory power of DNA profiling" [23].

The sampling formula of Balding and Nichols is an overcorrection for population substructure. It models the belief that if we have observed allele A_i in the population before then we are more likely to see it again. More specifically, if there are x copies

of allele A_i in a sample of *n* alleles, then the probability that the next allele will be of type A_i is:

$$\Pr\left(A_i \mid x \text{ copies of } A_i, n\right) = \frac{x\theta + (1-\theta)p_i}{1+(n-1)\theta}$$

where p_i is the probability of allele A_i, and θ is the coancestry coefficient (F_{ST}).

The use of the sampling formula was recommended in NCRII 1996 [17].

These statistics are generally presented using one of two types of methods; the frequentist approach or the logical approach.

1.2.2 Presenting a statistic

1.2.2.1 Frequentist approach

The frequentist approach uses probabilities or genotype frequencies to address the evidence (E) given a hypothesis H_1, $\Pr(E \mid H_1)$. The hypothesis might be as simple as saying that the observed profile is from someone unrelated to the person of interest.

Coincidence probability or random match probability (RMP) and the probability of exclusion are frequentist approaches that are used by the forensic science community.

The coincidence approach

The coincidence approach or RMP assesses if two profiles share the same alleles due to chance (coincidence). Allele frequencies are calculated using data that assumes HW equilibrium and no linkage disequilibrium (the non-random association of alleles at two or more loci that may or may not be on the same chromosome). The frequency of the genotype is then calculated using the product rule [24].

Probability of exclusion

The probability of exclusion or random man not excluded (RMNE) offers an approximation of the proportion of the population that has a genotype

composed of at least one allele not observed in the profile. For example, the person of interest (POI) is not excluded, what the probability that a random person would be excluded as the donor of this DNA? It is inferred that the POI is unlikely a random person $\left(\Pr\left(E \mid H_1\right)\right)$, therefore the evidence supports the null hypothesis $\left(\Pr\left(E \mid H_0\right)\right)$ that the POI is the donor of the DNA. The higher the exclusion probability the more support given to the null. For example, the exclusion probability: "Approximately 99.99999% of unrelated males would be excluded as the source of this DNA", implies strong support for the null hypothesis that the POI is the donor of the DNA [24].

Both the coincidence approach and the probability of exclusion appear easy to implement and are simple to explain to a jury. It has been argued that they are more conservative and more easily understood than the logical approach [25]. The main issue with these methods is that only part of the evidence is assessed (i.e. the crime scene DNA profile), the profile of the POI is not taken into account in the calculation. The danger in such an analysis is the inference that comes from thinking that because the probability of the DNA profile given the POI is not the source of the DNA $\left(\Pr\left(E \mid H_1\right)\right)$ is small; the probability of the profile given the POI is the source of the DNA $\left(\Pr\left(E \mid H_0\right)\right)$ is large.

1.2.2.2 The logical approach

The logical approach is better known as the likelihood ratio (*LR*). The classical *LR* approach consists of the comparison of the likelihood of obtaining the observed DNA profile given alternative competing hypotheses; typically termed the prosecution hypothesis and the defence hypothesis (Equation 1.1).

$$LR = \frac{\Pr(E \mid H_p)}{\Pr(E \mid H_d)} \qquad \text{Equation 1.1}$$

where E represents the evidence,

H_p is the prosecution's hypothesis, and

H_d is the defence's hypothesis.

In a single contributor simple DNA profile, the probability of observing the evidence given the hypothesis that the POI is the contributor is 1 under H_p. Under H_d it is assumed an unknown person, unrelated to the POI, is the true contributor. Therefore, the probability of observing the evidence, conditioned on an unknown contributor is modelled by the probability of observing the genotype in the target population; Equation 1.2; where $\Pr(E \mid H_p) = 1$ and f is the frequency of the genotype in a defined population

$$LR = \frac{1}{f} \qquad \text{Equation 1.2}$$

The *LR* is discussed further in Chapter 2.

1.2.2.3 Extensions to the logical approach

The classical *LR* defined under the logical approach works well for relatively simple analyses with clearly defined propositions under the prosecution and defence hypotheses. However, sometimes it is difficult to simplify a casework profile into two propositions. For example, if a profile is recovered from a crime scene (G_c), and the POI's profile (G_s) cannot be excluded from the crime scene profile, under the *LR* we could state the following propositions:

H_p: The DNA came from the POI.

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

However, note that the propositions are not exhaustive. What about males that are related to the POI? Genetics would suggest these people are the most likely to share a similar profile to the POI and are therefore the most important to consider under the H_d. The H_d can accommodate relatives if that is the hypothesis put forward by the person of interest. However, there are experts who state that it is not the responsibly of the POI or independent analysts (defence analysts) to put forward the H_d and that this hypothesis should include all situations from the outset.

What is required is a method that can handle many propositions. This can be implemented under the *LR* and derived from Bayes Theorem using Equation 1.3:

$$LR = \frac{\Pr(G_c \mid G_s, H_1, I)}{\sum_{i=2}^{N} \Pr(G_c \mid G_s, H_i, H_2, I) \Pr(H_i \mid H_2, I)} \qquad \text{Equation 1.3}$$

where for a population of size *N*, the POI is indexed as person 1 and the remaining members of the population as 2,…,*N*. The proposition that person *i* is the source of the DNA is H_i. Since the POI is indexed as person 1 the proposition that the POI is the source of the DNA is H_1. The propositions $H_2 \ldots H_N$ are the propositions where the source of the DNA is not the POI. *I* refers to the information at hand, i.e. the offender is European.

This method is also computationally intense and would require automation or software for its implementation. This research does not discuss this method in any more detail as this research requires the designation of H_d in situations to demonstrate the effect of different analysis methods on the resulting statistic.

However, this extension to the logical approach is not without merit and could be used in the development of interpretation methods for complex mixtures where the resolution of genotypes under the H_d is complex.

1.3 Problems with interpretation

The interpretation methods discussed in the previous section all function adequately when single source good quality DNA profiles are analysed. However, forensic DNA samples are often sub-optimal in quantity, or quality, and can be further complicated by factors such as the number of contributors to the profile. If the contributors to a profile are both present in similar quantities then, individual genotypes can often not be distinguished (complex mixtures). Additionally, if (a) contributor(s) is present at very low levels then it becomes difficult to distinguish between alleles of a minor profile and stochastic effects.

1.3.1 "Problem" profiles

The advent of increasingly more sensitive DNA analytical techniques has enabled scientists to generate profiles from samples that contain much smaller amounts of DNA. There is an approximate relationship between the resulting peak height in a profile and the amount of DNA template present in the original extract [27, 28]. If the quantity of DNA is low in the original extract then this often results in a reduction in peak heights in the resulting profile and problems with profile interpretation can arise [29].

Samples analysed with enhanced detection sensitivity techniques are often called low copy number (LCN) [30, 31] or low template DNA (LtDNA) samples [32, 33]. LCN analysis originated largely in Australia and generally refers to the use of increased PCR cycles to enhance the sensitivity of the DNA analysis [34]. However, some of the literature has adopted the term LCN to describe all samples that contain low levels of DNA.

There are various techniques used to improve the sensitivity of DNA profiling, other than increasing the PCR cycle number, and this has resulted in ambiguity regarding the use of the term LCN. Caddy et. al., [32] introduced the term LtDNA which refers specifically to samples that contain low levels of DNA, regardless of the subsequent enhancement technique used to address this issue.

Another complication in the classification of DNA samples as LtDNA is the lack of a clear definition that can be applied to delineate between LtDNA profiles and conventional profiles. Budowle et al., suggest in two separate papers that all profiles that give a quantification value of under either 100 pg or 200 pg should be defined as LtDNA [2, 35]. However, we believe that the application of an official threshold is unwarranted and will result in the misclassification of profiles. LtDNA profiles typically exhibit increased variability and increased stochastic effects. An increase in variability means that there can be problems with the reproducibility between profiles, while an increase in stochastic effects can make the interpretation of the profile difficult. However, profiles that contain over 200 pg of DNA can also exhibit these stochastic effects. Equally, profiles 199 pg and under may not exhibit any of these effects and can behave like conventional profiles. Additionally, if the resulting DNA

profile contains more than one contributor, then the quantification value does not take into account the relative contribution from each person. That is, although the overall quantification value may be well over the 200 pg threshold, the minor contributor may donate less than 200 pg of DNA [36-40].

We believe that the resulting electropherogram (epg) is the best indicator of the quantity of DNA present in the sample. In this research we regard any DNA profile whose epg exhibits stochastic effects, which makes subsequent interpretation of the profile difficult, as LtDNA.

1.3.2 Stochastic factors

Due to the kinetics of the PCR process a small number of DNA templates in any amplification will experience random or stochastic sampling effects. If there are only a limited number of starting templates, as in LtDNA samples, and these sampling effects occur in the early cycles of PCR amplification, the resulting DNA profile morphology can be compromised. Only those alleles that are amplified efficiently during the first few cycles will be able be detected, specifically, with a heterozygous locus, unequal sampling of alleles can result in a failure to detect one or both alleles.

Loss of a single allele is referred to as allele dropout while the loss of both alleles is referred to as locus dropout (Figure 1). Heterozygote balance refers to the ratio of peak heights (or areas) between the two alleles of a heterozygous locus and often used by analysts as a guideline when interpreting profiles. In LtDNA samples, heterozygote pairs are more prone to imbalance (Figure 1). Peaks from a heterozygote locus may be very imbalanced, with one of the alleles giving a much larger peak than the other. An analyst may then judge the difference between the peaks to be so great that they cannot be paired under the conventional heterozygote interpretation guidelines. Heterozygote balance is discussed further in Chapter 7 and dropout is discussed in Chapters 5 and 6.

Another complicating variable in DNA interpretation is the presence of stutter peaks. Here, stutter refers to the mis-copy of an allele resulting in an amplicon one repeat sequence shorter than the parent allele. Stutter is a by-product of the PCR process and also appears in full, single source DNA profiles. It has been reported within the

literature that increased variability in stutter percentages are present in LtDNA samples [41]. This can make it difficult to determine if the observed peak is in fact stutter. However, because stutter is a PCR by-product it is more likely that the resulting stutter peak height is dependent on the number of cycles used in the PCR process rather than the amount of DNA template present. Consequently it is LCN samples that are likely to have increased (mean) stutter peaks not LtDNA samples. This misconception is most likely a "hangover" from the confusion regarding LCN/LtDNA nomenclature. However, in mixed DNA profiles stutter peaks can be confused with a low level minor contributor making the interpretation of the profiles difficult (Figure 1). Stutter is discussed further in Chapters 8 and 9.

A third variable that can complicate interpretation is allele drop in (Figure 1). Drop in refers to extraneous alleles (unrelated to the crime scene) being amplified in the analysis. Drop in is related to the increase in sensitivity of the PCR process. This extra DNA is usually limited to one or two alleles and is attributed to "DNA falling from the ceiling" [42] (a random event) rather than laboratory or consumable based contamination. Drop in is usually limited to LCN samples and be calculated by observing the number of drop in events that occur within a specific lab. The calculation of the probability of drop in is not discussed within this research but the methods in which to include a probability of drop in within the calculation of a statistic for the weight of evidence is briefly touched upon in Chapter 2.

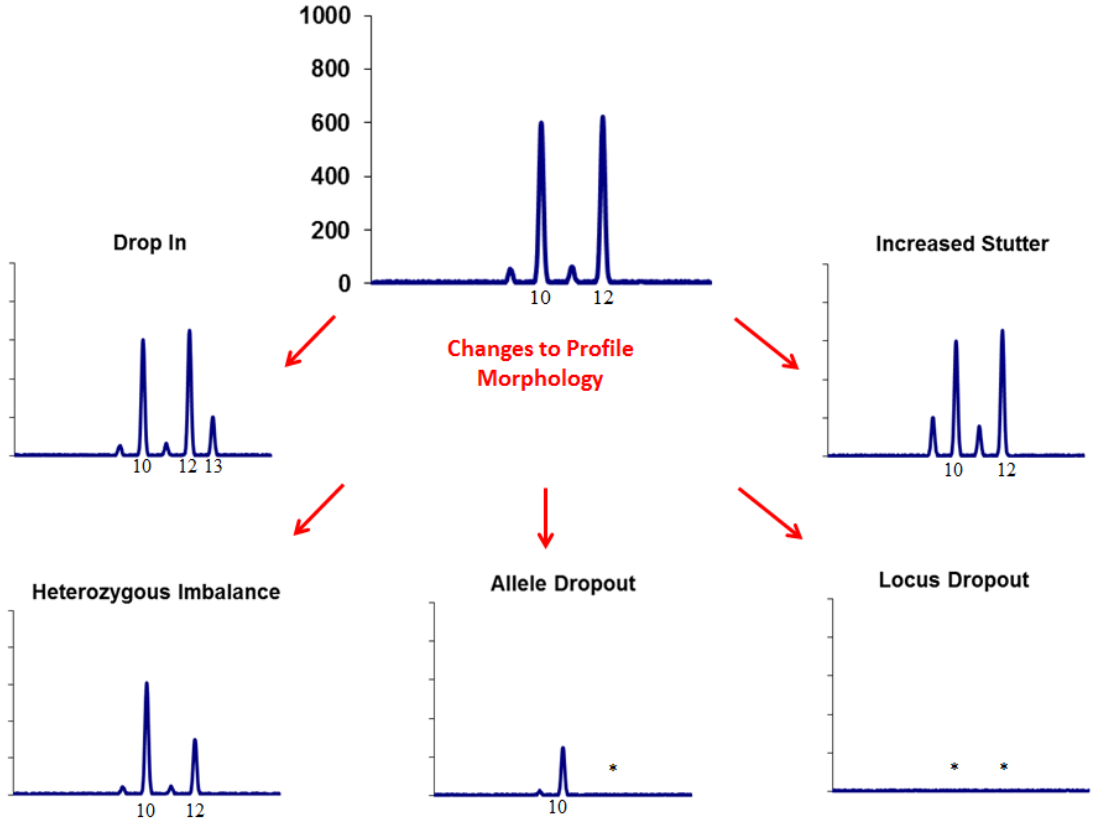


Figure 1: Examples of variations to a single source heterozygous locus (top centre) that complicate DNA interpretation.

These stochastic effects create uncertainty when designating peaks as alleles or pairing alleles as genotypes. Therefore they cannot be interpreted like a single source good quality profile where the results are more certain and follow the two step interpretation process (inclusion followed by the calculation of a statistic). Instead interpretation becomes more questionable and needs to be probability based. For example, one might ask, "what is the probability of seeing this peak if it is stutter?" Or, "what is the probability of this profile if the POI is the contributor?" As yet there is no consensus as to the best method to answer these questions. This has resulted in judicial issues regarding the acceptance of LCN and LtDNA evidence.

1.4 General acceptance of LtDNA and LCN evidence

LCN evidence was severely criticised in the court case *R vs Hoey* ([2007] NICC 49, 20 December, 2007). Part of the evidence in this case was DNA left on explosives connected to an attack in which 29 people were killed and over 200 injured. The judge

questioned the validity and the reliability of the LtDNA analysed using the LCN technique [[2007] NICC 49]. Although the Judge did not rule the evidence inadmissible, his comments were so damning that LCN work in the United Kingdom was shut down.

The comments most relevant to interpretation were;

- that there was no international agreement on the validation of LCN analysis,
- the Azores conference held in 2005 had ended with agreement only that more work was needed in that area [43], and
- that the lack of agreement in LCN analysis was in marked contrast to conventional DNA profiles for which there was internationally-agreed validation guidelines and definitions approved by the Scientific Working Group on DNA Analysis Methods (SWGDAM).

The UK's Forensic Science Regulator (Andrew Rennsion) commissioned an independent and objective view of the standards of the science used in the analysis of trace amounts of human DNA. That review conducted by Professor Brian Caddy (with the assistance of Dr Adrian Linarce and Dr Graham Taylor) was released on the 12th of April 2008 and concluded that the technique was "robust" and "fit for purpose" [32]. The review also enumerates 21 recommendations for specific improvements that should be undertaken to improve the methodology, including: the development of a consensus on the interpretation of test results and efforts to establish "best practices" for interpretation. LtDNA/LCN has since been reinstated in the United Kingdom, however no such consensus on the interpretation of results has been established.

Gilder et. al., (2008) have stated that they feel that the conclusions of the review are inconsistent with its recommendations in a number of respects [44]. For example, it is difficult to determine how a forensic technique could be deemed adequately validated for use in the courtroom when there is not yet an agreement on how its results should be interpreted. They state that this establishes grounds for concern about how earlier LCN cases have been interpreted. Additionally there are a number of recent articles that highlight interpretation inconsistencies for LtDNA profiles [45-48].

The lack of a consensus on the best practice for the interpretation of LtDNA profiles is problematic, especially as the SWGDAM 2010 guidelines state that “the laboratory must perform statistical analysis in support of any inclusion that is determined to be relevant in the context of a case, irrespective of the number of alleles detected” [49]. There are several types of statistical methods that an analyst can use to evaluate the significance of an inclusion. However not all of the methods can accommodate the stochastic effects associated with LtDNA. Regardless of this fact, these statistical methods are being applied, at times, incorrectly, and presented in court. Mixed DNA profiles that cannot be separated into individual contributors are also problematic to interpret and like LtDNA profiles, are subject to the opinions of the examiner, raising issue of bias in interpretation.

Geddes, L., highlights a LtDNA mixture case in which two probabilities for the same DNA profile, calculated using the same statistical method, were presented in court and differed by factor of approximately 2000 (1 in 47 vs. 1 in 95,000) [50]. When the profile was examined in a review of the evidence, a third analyst obtained a probability of 1 in 13 using a variation of the original statistical method. When a statistical method was used that attempted to take into account the stochastic effects, the resulting statistic indicated that the DNA profile was two times more likely if it came from the person of interest (POI) than someone unrelated to the POI else, evidence so weak it is virtually inconclusive. However, the lower two statistics were not presented in court and the defendant was convicted.

This is not an isolated incident. In 2005, Dr John Butler at NIST issued a range of forensic laboratories in the USA, the same mixture DNA and reference profiles and asked them to provide their conclusions about whether the profiles “matched” [46]. Results were reported using a variety of statistical methods but for those probabilities presented using the same statistical method, the final figures differed by 10 orders of magnitude.

Dror et al., (2011 [48]) carried out a similar experiment in which 17 expert DNA analysts, who were working in casework in an accredited government laboratory in North America, were presented with a mixed LtDNA profile and reference profiles from a sexual assault case. In the original case the analyst determined that the suspect “could not be excluded” from the LtDNA profile. Of the 17 experts asked to reanalyse

the case, one analyst concluded the suspect “could not be excluded”, four analysts concluded “inconclusive”, and 12 analysts concluded “exclude”. These analysts all work in the same laboratory and follow the same interpretation guidelines. This study highlights the subjectivity that can be present in the interpretation of complex or LtDNA profiles. It also emphasises the importance of establishing a consensus on the best practice for interpretation.

The increased variability of LtDNA profiles was recognised in the late 1990’s and an interpretation strategy called *the statistical model* was developed that significantly compensated for the stochastic effects present [51]. However, statisticians haven’t had the funding to adapt this theory to forensic use.

The aim of this research was to investigate the behaviour of LtDNA. Traditional guidelines used to interpret conventional DNA profiles were investigated with regards to their application to LtDNA profiles. The behaviour of DNA at low levels was investigated with respect to heterozygous balance (*h*), dropout (Pr(D)) and the stutter ratio (*SR*). The distributions of the resulting low level data under these guidelines were modelled using statistical methods. The resulting models can be applied across all DNA profiles amplified at 28 cycles, irrespective of the distinction between conventional and LtDNA profiles.

This research focused on identifying LtDNA and complex mixture profiles from casework data. These profiles were then utilised in investigating the behaviour of the distributions of the guidelines and in the model construction. We feel that this is an important distinction from theoretical studies or empirical studies that use pristine DNA, pristine DNA dilutions or pristine DNA mixtures. Although there is research that shows that there is no difference between pristine and casework heterozygote balance (*h*) data [52], we are aware that there is an observable degradation slope in casework data which effects higher molecular weight alleles more than lower weight molecular alleles [53]. This molecular weight based effect will not be so obvious in analyses that are locus/ratio based i.e. heterozygous balance (*h*) or the stutter ratio (*SR*). However, this effect may be more pronounced in analyses that compare loci across a profile. In particular, higher molecular weight alleles will be more prone to dropout than lower weight molecule alleles. In addition, by utilising casework data we

obtained a large representation of alleles that are present in the population, compared to pristine studies that often arise from only a few donors.

1.5 Thesis structure

The following chapters are made up of papers presented in their journal format, including references (in the reference style of the journal the paper was published in). The papers that have appeared are consistent with their published version except where noted in the preceding introduction of the chapter. As a result there is some repetition between chapters, primarily within the introductory sections, however each chapter can stand alone and be read independently of the other chapters.

The papers have been written with the aim of bridging communication between statisticians and the people we hope will use these methods – forensic biologists. Therefore, this thesis also contains additional explanatory chapters that detail why a particular method was chosen or that contain additional data analysis.

Chapter 2 provides a detailed discussion on the merits of the likelihood ratio and why the *LR* was chosen as the methodological framework to calculate the statistic. The merits of the likelihood ratio are discussed alongside the merits and short-comings of the other widely used framework in DNA interpretation, random man not excluded. RMNE is widely used and there has been apprehension to move towards a *LR* framework. This chapter aims to clarify why the *LR* is the logical (and only) method that can be used to interpret complex mixtures and LtDNA.

Chapter 3 discusses the various models that are used to evaluate the *LR*. The binary model, the semi-continuous model and the continuous model are compared with respect to their limitations and as to how far they can be extended to cope with LtDNA and complex mixtures.

Chapter 4 compares three models that are used when determining the genotypes possible under the *LR* hypotheses. Two of these models, termed the *F* and *Q* models take into account peak height, or quantitative information and are extensions of the binary model. The third model, termed the unconstrained combinatorial approach does not use peak heights and only makes use of the qualitative data in the profile. These three methods are used to interpret the same low level two person mixture and

a statistic is then calculated for each of them using the *LR*. This chapter highlights the shortcomings of models that do not make use of quantitative data. It also provides tables of formulae for many situations for one, two and three contributors using the *F*, *Q* and *UC* models.

Chapter 5 investigates the effect of the inclusion of a dropout probability in the generated statistic. The DNA Commission of the ISFG stressed the importance of considering the probability of allelic dropout (Pr(D)) in its recommendation on mixture interpretation, but how to best assess Pr(D) was not formalised. This chapter demonstrates how important the estimation of Pr(D) is in calculating the weight of evidence under the *LR*.

Chapter 6 calculates a probability of dropout using casework data and logistic regression. This builds on previous work by Tvedebrink et al. [54]. Peak height data is utilised as a proxy for DNA template and a degradation slope is included that is based on molecular weight. This model is compared to Tvedebrink's for transportability. i.e. how well do these respective models that are trained on one set of data, work on a second set of data. This was done in order to ensure that the model would be able to be utilised in casework.

Chapter 7 investigates heterozygote balance (*h*). We concentrated on building a model for the expected *h* given the observed average peak height for a locus. Data were collected that allowed this distribution to be modelled. The variance of *h* is shown to decrease at a rate inversely proportional to the average peak height at the locus $\left(\overline{\phi}\right)$. The variance of *h* is most extreme at LtDNA levels. The difference in the number of repeat units between the heterozygous alleles at a locus $\left(\delta\right)$ was also found to have an effect on *h*. Using statistical modelling techniques, a model was built that uses δ and $\overline{\phi}$ to give an expected estimate for *h*, as well as an interval that *h* would be expected to fall within (with reasonable probability).

Chapter 8 identifies and models the drivers of the stutter ratio (*SR*) in forensic DNA profiles. The *SR* describes the ratio of the stutter peak to its parent peak. The longest uninterrupted sequence (*LUS*) of repeat units in an allele was found to affect *SR*. A locus effect was also seen. A linear model was designed which describes the

behaviour of the expected *SR* with respect to locus and *LUS*. The expected (mean) *SR* was not affected by parent peak height and is more likely to be dependent on PCR cycle number.

Chapter 9 describes additional work investigating the drivers of stutter. In particular, what causes loci to stutter differently? D21S11 is examined in detail as it was noted that D21S11 does not follow the typical linear behaviour with respect to *LUS* as the other loci. In addition to investigating D21S11, this chapter compares The Next Generation Multiplex (NGM™ SElect Kit) *SR* data to AmpFℓSTR® Identifiler® *SR* data. Because NGM^{TM} SElect uses 30 cycles of PCR and AmpFℓSTR® Identifiler® uses 28, the observed *SR* for NGM^{TM} SElect data is higher. Using the difference in *SR* a "PCR efficiency" value was created which can be used to give an approximate probability of *SR* for D21S11 for any cycle number.

Chapter 10 is the discussion, conclusion and future areas of work identified in this research. Forensic DNA analysis is a huge asset to law enforcement, yet at the extremes of its applications there are limitations and the boundaries of these limits are now under critical review.

The introduction of statistics to forensic science has not always been an easy process for forensic biologists and reluctance to introduce seemingly more complicated methods is understandable. However how best to interpret complex and LtDNA profiles is a challenging and highly topical subject and requires discussion. It is important that the limits of current interpretation models are communicated and that models that can cope with the stochastic factors associated with LtDNA are introduced in a manner in which the forensic biology community will accept and adopt.

References

1. Gill, P., A. Jeffreys, and D. Werrett, Forensic application of DNA 'fingerprints'. Nature, 1985. 318 (12 December): p. 577 - 579.

2. Budowle, B., A.J. Eisenberg, and A. van Daal, Validity of Low Copy Number Typing and Applications to Forensic Science. Croatian Medical Journal, 2009(50): p. 207-217.

3. Jeffreys, A.J., A.C. Wilson, and S.L. Thein, Individual specific "fingerprints" of human DNA. Nature, 1985. 316: p. 75-79.

4. Donnelly, P., Match probability calculations for multi-locus DNA profiles. Genetica, 1995. 96(1-2): p. 55-67.

5. Buckleton, J., J.M. Curran, and C.J. Basten, Evaluating the statistical significance of single band profiles in VNTR analyses. Science & Justice, 2000. 40(1): p. 27-33.

6. Mullis, K. and F. Faloona, Specific synthesis of DNA in vitro via a polymerase catalysed chain reaction. Methods in Enzymology, 1987. 155: p. 335 - 350.

7. Mullis, K.B., et al., Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. Cold Spring Harbor Symposia on Quantitative Biology, 1986. 51: p. 263-273.

8. Mullis, K.B., The unusual origin of the polymerase chain reaction. Scientific American, 1990. 262(4): p. 64-65.

9. Smith, J.C., et al., Highly polymorphic minisatellite sequences: allele frequencies and mutation rates for five locus-specific probes in a Caucasian population. Journal of the Forensic Science Society, 1990. 30(1): p. 19-32.

10. Lygo, J.E., et al., The validation of short tandem repeat (STR) loci for use in forensic casework. International Journal of Legal Medicine, 1994. 107: p. 77-89.

11. University of Leicester, STR profiling in the UK http://www2.le.ac.uk/departments/emfpu/genetics/explained/str-uk.

12. 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. International Journal of Legal Medicine, 1996. 109(4): p. 195-204.

13. Jobling, M.A. and P. Gill, Encoded evidence: DNA in forensic analysis. Nature Reviews Genetics, 2004. 5(10): p. 739-751.

14. Cohen, A.S., D.R. Najarian, and B.L. Karger, Separation and analysis of DNA sequence reaction products by capillary gel electrophoresis. Journal of Chromatography A, 1990. 516(1): p. 49-60.

15. Bright, J.A., Validation of the AmpFℓSTR® Identifiler® PCR Amplification kit at the Forensic Biology Group of ESR, 2006.

16. 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, Washington, D.C.: National Academy Press.

17. NRC II, National Research Council Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence. 1996, Washington, D.C.: National Academy Press.

18. Lewontin, R.C. and D.L. Hartl, Population genetics in forensic DNA typing. Science, 1991. 254: p. 1745 - 1750.

19. Lewontin, R., Which population? American Journal of Human Genetics, 1993. 52: p. 205.

20. Lewontin, R., Forensic DNA typing dispute. Nature, 1994. 372: p. 398.

21. Hartl, D., Forensic DNA typing dispute. Nature, 1994. 372: p. 398-399.

22. Aldhous, P., Geneticists attack NRC report as scientifically flawed. Science, 1993. 259(5096): p. 755-756.

23. Balding, D.J. and R.A. Nichols, DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Science International, 1994. 64(2–3): p. 125-140.

24. Buckleton, J., A Framework for Interpreting Evidence, in Forensic DNA Evidence Interpretation. 2005, CRC Press. p. 27-65.

25. Thompson, W.C., DNA evidence in the O.J. Simpson trial. University of Colorado Law Review, 1996. 67(4): p. 827-857.

26. Perlin, M., et al., Validating Trueallele® DNA mixture interpretation Journal of Forensic Science, 2011. 56(6): p. 1430-1447.

27. Clayton, T., et al., Analysis and interpretation of mixed forensic stains using DNA STR profiling. Forensic Science International, 1998. 91: p. 55 - 70.

28. Gill, P., et al., Interpreting simple STR mixtures using allelic peak areas. Forensic Science International, 1998. 91: p. 41 - 53.

29. Whitaker, J.P., E.A. Cotton, and P. Gill, A comparison of the characteristics of profiles produced with the AmpFℓSTR® SGM Plus™ multiplex system for

both standard and low copy number (LCN) STR DNA analysis. Forensic Science International, 2001. 123: p. 215-223.

30. Gill, P., J. Whitaker, and J. Buckleton, Few cell DNA profiling, 2001: International, US, European and Australian.

31. Whitaker, J.P. and E.A. Cotton, Low copy number. First International Conference on Forensic Human Identification in the Millennium, 1999.

32. Caddy, B., G. Taylor, and A. Linacre, A review of the science of low template DNA analysis., 2008. http://www.bioforensics.com/articles/Caddy_Report.pdf

33. Regulator, T.F.S., Response to Professor Brian Caddy's Review of the Science of Low Template DNA Analysis, 2008. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/117557/response-caddy-dna-review.pdf

34. Findlay I, Frazier R, Taylor A, Urquhart A. Single cell DNA fingerprinting for forensic applications. Nature, 1997. 389: p 555-556.

35. Budowle, B. and A. van Daal, Reply to Comments by Buckleton and Gill on "Low copy number typing has yet to achieve 'general acceptance"' by Budowle, B., et al., 2009. Forensic Sci. Int.: Genet. Suppl. Series 2, 551–552. Forensic Science International: Genetics, 2011. 5(1): p. 12-14.

36. Buckleton, J., Validation issues around DNA typing of low level DNA. Forensic Science International: Genetics, 2009. 3(4): p. 225-260.

37. Budowle, B., R. Chakraborty, and A. van Daal, Author's response to Gill P, Buckleton J. Commentary on: Budowle B, Onorato AJ,

38. Buckleton, J. and P. Gill, Further Comment on "Low copy number typing has yet to achieve "general acceptance" " by Budowle, B., et al, 2009. Forensic Sci. Int. Genetics: Supplement Series 2, 551-552. Forensic Science International: Genetics, 2011. 5(1): p. 7-11.

39. Gill, P. and J. Buckleton, A universal strategy to interpret DNA profiles that does not require a definition of Low-Copy-Number. Forensic Science International: Genetics. 2010. 4(4): p. 221-227

40. Buckleton, J. and P. Gill, Further Comment on "Low copy number typing has yet to achieve "general acceptance" " by Budowle, B., et al, 2009. Forensic Sci. Int. Genetics: Supplement Series 2, 551-552. Forensic Science International: Genetics. 2011. 5(1): p. 7-11

41. nfstc. http://www.nfstc.org/pdi/Subject06/pdis06m0205.htm.

42. Buckleton, J. and P Gill, Low Copy Number, in Forensic DNA Interpretation. 2005, CRC Press. p 281.

43. Gill, P., et al., DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. . Forensic Science International., 2006. 160: p. 90-101.

44. Gilder, J., et al., Comments on the review of low copy number testing. International Journal of Legal Medicine, 2008. 123(6), p 535-536.

45. Geddes, L., What are the chances? The New Scientist, 2010. 207(2774): p. 8-10.

46. Geddes, L., C. King, and C. Stier, Between Prison and Freedom. The New Scientist, 2010. 207(2773): p. 8-11.

47. Butler, J. NIST Interlaboratory Mixture Interpretation Study 2005 (MIX05). 2005 [cited 2011 22 June]; Available from: http://www.cstl.nist.gov/biotech/strbase/interlab/MIX05.htm.

48. Dror, I.E. and G. Hampikian, Subjectivity and bias in forensic DNA mixture interpretation. Science & Justice : Journal of the Forensic Science Society, 2011. 51(4): p. 204-208.

49. (SWGDAM), SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. http://www.fbi.gov/about-us/lab/biometric-analysis/codis/swgdam-interpretation-guidelines [cited 2013].

50. Geddes, L., How DNA creates victims of chance. New Scientist, 2010(2774).

51. Gill, P., et al., An investigation of the rigor of interpretation rules for STR's derived from less than 100 pg of DNA. Forensic Science International, 2000. 112(1): p. 17-40.

52. Bright, J.A., et al., A comparison of stochastic variation in mixed and unmixed casework and synthetic samples. Forensic Science International: Genetics, 2012. 6(2): p. 180-184.

53. Nicklas, J.A., T. Noreault-Conti, and E. Buel, Development of a Real-Time Method to Detect DNA Degradation in Forensic Samples*. Journal of Forensic Sciences, 2012. 57(2): p. 466-471.

54. Tvedebrink, T., et al., Estimating the probability of allelic drop-out of STR alleles in forensic genetics. Forensic Science International, 2009. 3: p. 222-226.

2: USING THE LIKELIHOOD RATIO TO REPORT THE STATISTICAL WEIGHT OF DNA EVIDENCE

The introduction of DNA profiling constitutes one of the most important advancements in the criminal justice system. Forensic DNA analysis is a huge asset to law enforcement and is seen as a robust and objective forensic discipline. However, with improvements in technology, the limits of DNA profiling are now being tested. Increased sensitivity in analysis techniques means that a wider range of evidence types, previously thought to hold little evidential value because of the limited quantity of available DNA, can now be submitted for DNA analysis. Consequently, the resulting DNA profile is often compromised in quality. The increase in sensitivity of the analysis techniques also means that very low levels of DNA, not previously identified in routine analyses are now detected, resulting in an increase in the analysis of DNA mixtures.

The interpretation of DNA mixtures can be complicated. If the contributors to a profile are both present in similar quantities then individual genotypes cannot be distinguished. Additionally, if (a) contributor(s) is present at very low levels (LtDNA), then it becomes difficult to distinguish between alleles of a minor contributor and stochastic effects. These stochastic effects include; stutter, heterozygous balance, dropout, and drop in, and are the defining characteristics of LtDNA profiles.

Here, stutter refers to the mis-copy of an allele resulting in an amplicon one repeat sequence shorter than the parent allele. Stutter also appears in full, single source, DNA profiles, however in LtDNA samples the difference in height between the stutter peak and the parent peak can decrease, making it difficult to determine if the observed peak is in fact stutter. Additionally, if there are other smaller peaks present in the profile that could make up a minor component of a DNA mixture then the designation of stutter peaks becomes complicated.

Heterozygous balance refers to the ratio of peak heights (or areas) between the two alleles of a heterozygous locus. In LtDNA samples, heterozygous pairs are more prone to imbalance (Figure 2.1). An extreme manifestation of heterozygous imbalance is allele dropout.

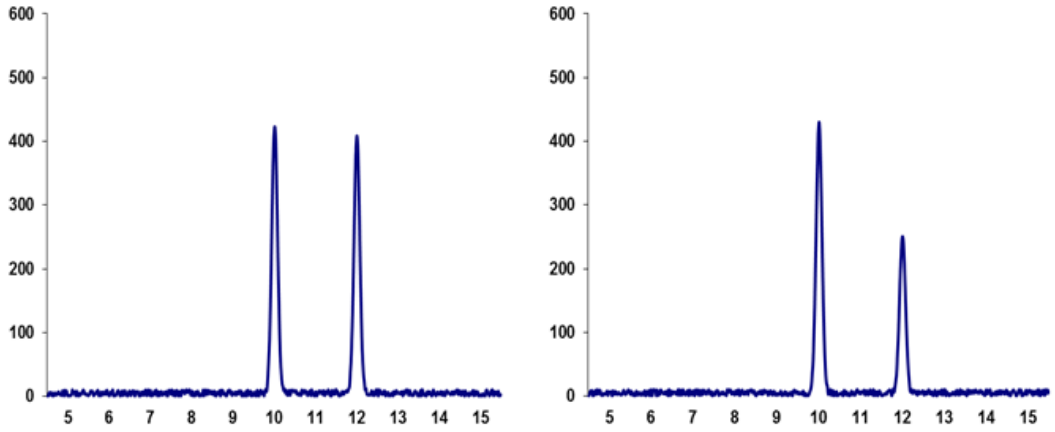


Figure 2.1: An electropherogram showing balanced heterozygous alleles on the left and imbalanced heterozygous alleles on the right. RFU are on the Y axis and allele designation is on the X axis.

Allele dropout is the failure of one of the alleles of a heterozygous pair to reach the detection threshold, either because it has not amplified or it is present in such low volumes it avoids detection. Therefore, the locus appears to be homozygous. Both alleles at a locus can also fail to be detected. This is termed extreme or locus dropout.

In addition to alleles failing to be amplified, extraneous alleles (unrelated to the crime scene sample) can be amplified, or “drop in”. This extra DNA is usually limited to one or two alleles and is attributed to “DNA falling from the ceiling” (a random event) rather than laboratory or consumable based contamination [1]. These stochastic effects create uncertainty in the interpretation of crime scene samples and therefore make it difficult to evaluate the weight of the evidence.

2.1 The interpretation of DNA

Traditionally the interpretation of DNA profiles has been thought of as a two-step process (Figure 2.2). The initial step involves the interpretation of the electropherogram (epg). Thresholds and guidelines are often used by forensic laboratories to aid in determining if peaks present in the epg can be designated as alleles, if alleles at a locus can pair (determining the genotype at a locus), and finally to aid in resolving the contributor(s) profile.

The second step involves comparing the resolved contributor profile to a reference profile. If the reference profile cannot be excluded from the contributor profile, then it is customary to provide some form of statistic to evaluate the weight of the evidence.

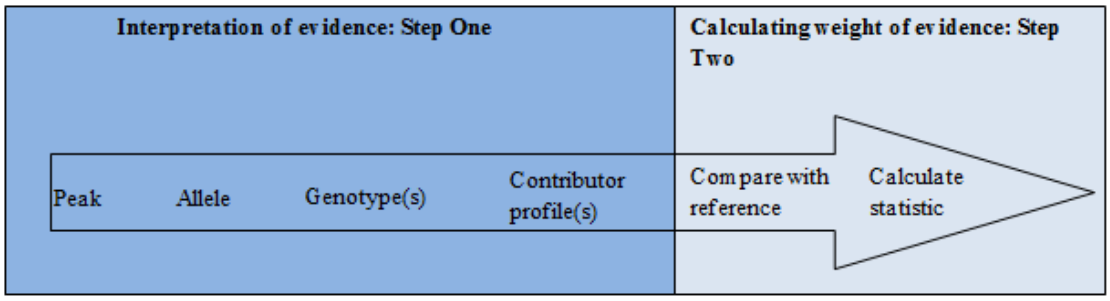


Figure 2.2: Schematic of the DNA interpretation process.

Two popular methods used to calculate the statistical weight for DNA profiles include the Random Man Not Excluded (RMNE) and the Likelihood Ratio (*LR*).

2.1.1 The Random Man Not Excluded

The RMNE in itself is as a two-step process. First, the suspect must be included as a contributor to the profile (a match) and then, given that a match is declared, the fraction of a defined population that would also match the contributor profile is calculated. RMNE supporters advocate its use because “it is more straightforward to implement and is easier to explain in court than the *LR*” [2]. However, “ease of explanation” is not an acceptable criterion [3].

The “match step” is a weakness of the RMNE. If there is uncertainty about the composition of a crime scene sample, then an “inclusion” can become blurry. If there

is the possibility of dropout and or drop in, then there is uncertainty that the observed profile faithfully reflects the true contributor.

Traditionally, problematic loci have been left out of the RMNE calculation. This is equivalent to treating them as a "one" (neutral) in the calculation. However this approach is biased as it does not account for the exclusionary potential of the locus. This uncertainty should be reflected in the statistic. The *LR* is able to incorporate this uncertainty.

2.1.2 The likelihood ratio

The *LR* does not require the "match or non-match" way of thinking, although it did use "match or non-match" for a long time. The need to designate an inclusion or exclusion is avoided because the conditioning that is used in the calculation negates the requirement for a two-step approach [4]. The classical *LR* approach consists of the comparison of the likelihood of obtaining the observed DNA profile given alternative competing hypotheses; typically termed the prosecution hypothesis and the defence hypothesis. Subjective bias can be avoided because utilising the *LR*, a statistical model can be employed that measures the strength of evidence that can favour the defence as well as the prosecution [4].

A number of *LR* based models have been described in the literature that can be used to interpret complex DNA profiles that (may) exhibit dropout and or drop in [4-13]. Because the *LR* requires a model to evaluate the probability of the evidence under each hypothesis it is the difference in these models that differentiates the various *LR* models. Although *LR* based methods have been in the literature for more than a decade, widespread laboratory acceptance of the *LR* has been slow. This is thought to be largely due to lack of understanding on the basis behind the *LR*, an inability to calculate it and the belief that the resulting statistic is difficult to explain in court. However, the *LR* has been acknowledged to be the most powerful and relevant statistic used to calculate the weight of the DNA evidence [14], and is recommended by the DNA commission of the International Society of Forensic Genetics (ISFG) in mixture interpretation [14].

In 2008, the United Kingdom's Forensic Science Regulator commissioned a review of LtDNA techniques. One of the conclusions of this review states that the development of a consensus on the interpretation of LtDNA profiles needs to be established [15]. As yet there is no consensus as to the best practice for the interpretation of complex mixtures or LtDNA profiles, however there are recommendations available [16]. There are a number of recent articles that highlight interpretation inconsistencies for these profiles [17-19].

There is difficulty in standardising interpretation due to different laboratories following different guidelines (Step one, Figure 2.2) and the lack of a consensus as to the most appropriate method to calculate the statistical weight of "incomplete" profiles that might match a reference profile (Step two, Figure 2.2). Regardless, the Scientific Working Group on DNA Analysis Methods (SWGDAM) 2010 guidelines state that "a laboratory must perform statistical analysis in support of any inclusion that is determined to be relevant in the context of a case, irrespective of the number of alleles detected" [20].

Given that the SWGDAM guidelines states that a statistic must be calculated for an "inclusion" despite there being the possibility of missing alleles from the contributor profile, and the review of LtDNA techniques calling for a consensus in interpretation, it is logical that an agreement on the most appropriate method to calculate an "inclusion" statistic is the sensible place to begin interpretation standardisation.

The recommendation of the use of the *LR* by the ISFG, and the evidence within the literature that indicates that the *LR* is the only method that can comprehensively evaluate the stochastic phenomena associated with complex profiles, suggests that the *LR* is the logical methodological framework that should be used in DNA interpretation. Therefore it is important that the capabilities and limits of the *LR* in regards to forensic DNA interpretation are explored and understood. The challenges in terms of statistical interpretation and in communicating the results to a criminal justice system must also be discussed.

2.1.2.1 Exploring the likelihood ratio

The *LR* is not exclusive to forensic DNA interpretation. It is used in many situations in statistics and is contained within Bayes' theorem (Equation 2.1) [6].

$$\frac{\Pr(H_1 \mid E, I)}{\Pr(H_2 \mid E, I)} = \frac{\Pr(E \mid H_1, I)}{\Pr(E \mid H_2, I)} \times \frac{\Pr(H_1 \mid I)}{\Pr(H_2 \mid I)} \qquad \text{Equation 2.1}$$

where:

H_1 is hypothesis one,

H_2 is an alternate hypothesis to H_1,

E represents the evidence, and

I represents background relevant information.

Bayes' theorem follows directly from the laws of probability and can be expressed in words as:

Posterior odds = likelihood ratio x prior odds.

In a forensic DNA context, the prior odds are the odds on the hypothesis before the DNA evidence. This is restricted to information relevant and admissible to the case. The posterior odds reflect the prior odds updated by the weight of the evidence (*LR*).

The role of the scientist is to present only the *LR*. In reporting the *LR* alone, the scientist is not offering an opinion on the hypothesis.

The typical *LR* approach for forensic DNA analysis consists of the comparison of the likelihood of obtaining the observed DNA profiles given the alternative competing hypotheses of the prosecution H_p and the defence H_d (Equation 2.2).

$$LR = \frac{\Pr(E \mid H_p, I)}{\Pr(E \mid H_d, I)}$$ Equation 2.2:

where:

under H_p, the POI contributed to the sample, and

under H_d, an unknown person unrelated to the POI contributed to the sample.

For single source profiles where there is sufficient DNA in the sample for the alleles to realistically reflect the true contributor and the observed profile "matches" that of the POI, then under H_p, the probability of observing the evidence given the hypothesis that the POI is the contributor is (usually) one. Under H_d it is assumed an unknown person is the contributor and as there is sufficient DNA to assume that the peaks present correctly represent the alleles of the true contributor, the probability of observing the evidence, conditioned on an unknown contributor can be modelled by the conditional probability of observing the genotype in the target population; Equation 2.3.

$$LR = \frac{1}{f}$$ Equation 2.3

where $\Pr(E \mid H_p) = 1$ and f is the probability of observing the genotype in a defined population

Generally f is a very small number which results in a large *LR*. This indicates that the *LR* supports the prosecution's hypothesis.

2.2 The interpretation of DNA when dropout it possible

Whenever dropout is a possibility there is also uncertainty that alleles present in the sample realistically reflect the true contributor. Therefore the numerator of the *LR* cannot be 1, and the denominator of the *LR* cannot be restricted to the frequency of the called profile within the population. Instead, this number should reflect the likelihood of other profiles being possible if alleles have dropped out. If the

possibility of drop out is not included in the calculation then the resulting probabilities will be too small, resulting in possible bias against the POI.

Supporters of the RMNE have extended the method to deal with problem loci by omitting the loci from the calculation completely. Ignoring a locus essentially gives it a value of "one". This ignores the exclusionary potential of the locus. For example, if in a LtDNA sample a peak is present at 299 RFU (with a homozygote threshold of 300 RFU) and there is no second peak present, then a heterozygote is not likely. The probability of the second allele dropping out is low and it is more than likely the contributor is homozygous at this locus. In a RMNE calculation this locus would be ignored and treated as "neutral" (as the allele is below threshold, therefore drop out is deemed possible [21]).

In reality the value of this locus cannot be one (neutral). The value of this locus must reflect the low probability that the partner allele has dropped and that the contributor is more than likely homozygous at this locus.

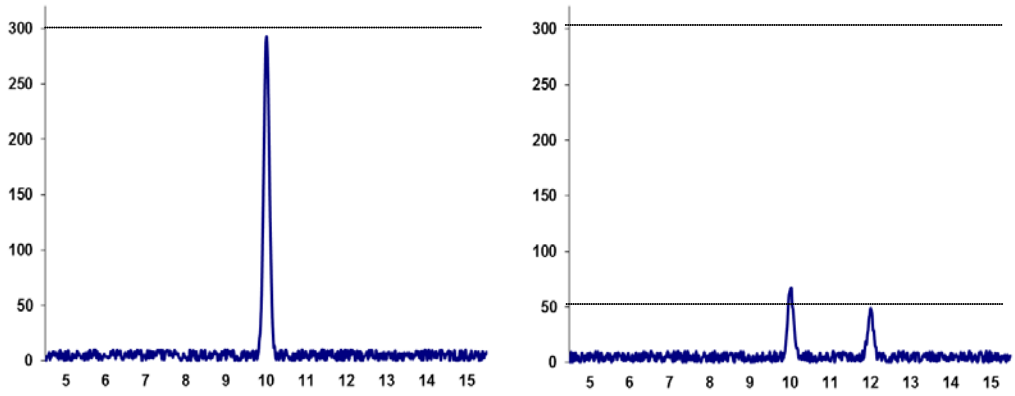


Figure 2.3: Two different epg scenarios seen alike under the RMNE

Van Nieuwerburgh et al. [2] describe a method they developed for the RMNE that allows for allelic dropout without omitting the locus. The Van Nieuwerburg method makes the assumption that a given number of alleles may have dropped out and then accepts any profile in the population that matches as "not excluded". The problem with this method is that dropout is assigned – i.e it is assumed to have occurred at the exact rate needed to include the POI. This again is raising similar issues where

possible exclusionary loci are designated incorrectly and the increases the likelihood ratio for innocent non-excluded persons. This is aided in part by the RMNE method not making use of the quantitative data– i.e. it does not see peak height. Alleles are considered present when above the stochastic threshold and absent when not. That the RMNE does not see peak heights makes the designation of possible "dropped out" alleles even more problematic. For example, the epg on the left in Figure 2.3 shows a single peak at a locus at 299 RFU. It is likely that the peak is a homozygote and the probability that a partner allele has dropped out is low. Assigning 'dropout' to this locus would include all heterozygous suspects bearing the observed peak. Most of these should be assigned a very low weight of evidence.

However, if the same peak was only just over the limit of detection and there was another smaller peak just below the limit of detection then it is likely that that locus is heterozygous, and the peak below threshold has dropped out (epg on the right, Figure 2.3). Although RMNE interprets these two situations in the same manner we can see that they are different which is why the probability of drop out should be factored into calculations at these loci.

We can assess the probability of the evidence given the hypothesis at each locus using the *LR* (Equation 2.4).

$$\Pr(E \mid H_G) \qquad \text{Equation 2.4}$$

where:

E is the evidence, or peak heights at the locus, and

H_G is the hypothesised genotype at the locus.

Buckleton and Gill [22] describe, in depth, a method that they developed to assess profiles using a locus by locus process. This method is described briefly here in order to demonstrate how the LR can be broken down. However other methods have since been published that extend this style of thinking using automated systems [9, 23-25]. These systems are likely to give a statistic that more accurately incorporates available

information. They are increasingly complex and difficult to explain. This is discussed further in Chapter 3.

It may seem strange to interpret the locus based on a hypothesis without ever trying to resolve the genotypes of the contributor(s) at the locus. However, this method is justified mathematically and represents a welcome and useful release from the need to make arbitrary decisions in pre-processing the profile. Consider a locus with a single peak at 70 RFU at position A. One must consider what is necessary, is to assess the probability of seeing this single A peak IF the contributor is an AA homozygote and to assess the probability of this single A peak IF the contributor is an AX heterozygote. Clearly if the peak is high, then the chance of observing a single A peak form an AX heterozygote is low and so forth. These assessments should be based on empirical data and are best described with the example below:

Consider the following scenario: Table 2.1; A LtDNA sample has been analysed and at a locus in replicate 1, a peak is present in position A. The peak is high and is probably an AA homozygote. However there is still a possibility (although low) that the partner allele has dropped out; AX. In order to explain an AA homozygote, both A alleles must not have dropped out $\left(\overline{D}_2\right)$. In order to explain an AX heterozygote, the A allele must not have dropped out but the partner allele has dropped out $\left(D\right)$. The probability of the genotype $\Pr\left(\mathrm{M}_j\right)$ is multiplied by the probability of the event (i.e. dropout or not dropout) and then, using the format of Table 2.1 the product column is summed. The $\Pr(E \mid H_d)$ is the summed product (all possible genotypes, regardless of their probability) and the $\Pr(E \mid H_p)$ is whichever cell in the product column represents their POI. The probability of dropout is central to this methodology and many papers have been published exploring this [5, 10, 26-29]. Chapters 5 and 6 discuss this further.

Table 2.1: The Buckleton and Gill model

$\mathbf{M_j}$	$\mathbf{PrM_j}$	$\mathbf{R_1}$ **AA**	**Product**
AA	$\Pr(AA \mid x)$	$\bar{D}_2$	$\bar{D}_2 P_{AA\mid x}$
AX	$2\Pr(AX \mid x)$	$\bar{D}D$	$2\bar{D}DP_{AX\mid x}$
			$\bar{D}_2 P_{AA\mid x} + 2\bar{D}DP_{AX\mid x}$

The beauty of Buckleton and Gill presenting their methodology in such a simple tabulated format is that it can be extended to cope with a number of other stochastic factors or other phenomena which are present in complex profiles. For example, if a second peak was present in the above example, then the probability of it having dropped in (C) could be considered along with the probability of it being a stutter peak (modelling stutter is discussed further in Chapter 8).

2.3 Conclusion

DNA profiling using STR variants is a well-established robust method. The ensuing interpretation of single source, simple profiles is usually routine. However, LtDNA casework poses additional challenges for interpretation in terms of heterozygous balance, allele dropout, stutter and drop in. The *LR* can be adapted to reflect the effects of these phenomena without compromising its statistical, or scientific, validity. The same cannot be said for the RMNE statistic. It is for this reason that we believe that the *LR* is the most sensible summary of the statistical weight of the evidence.

The *LR* is the methodological framework utilised in the rest of this research.

References

1. Buckleton, J. and P Gill, Low Copy Number, in Forensic DNA Interpretation. 2005, CRC Press. p 281.

2. Van Nieuwerburgh, F., et al., RMNE probability of forensic DNA profiles with allelic drop-out. Forensic Science International: Genetics Supplement Series, 2009. 2(1): p. 462-463.

3. Curran, J.M. Is forensic science the last bastion of resistance against statistics? Science and Justice 2013 53:p 251–252

4. Gill, P. and H. Haned, A new methodological framework to interpret complex DNA profiles using likelihood ratios. Forensic Science International: Genetics, 2013. 7(2): p. 251-263.

5. Tvedebrink, T., et al., Estimating the probability of allelic drop-out of STR alleles in forensic genetics. Forensic Science International, 2009. 3: p. 222-226.

6. Buckleton, J., A Framework for Interpreting Evidence, in Forensic DNA Evidence Interpretation. 2005, CRC Press. p. 27-65.

7. Cowell, R.G., S.L. Lauritzen, and J. Mortera, Probabilistic modelling for DNA mixture analysis. Forensic Science International: Genetics Supplement Series, 2008. 1(1): p. 640-642.

8. Kelly, H., et al., The interpretation of low level DNA mixtures. Forensic Science International: Genetics, 2012. 6(2): p. 191-197.

9. Gill, P., A. Kirkham, and J. Curran, LoComatioN: A software tool for the analysis of low copy number DNA profiles. Forensic Science International, 2007. 166(2-3): p. 128-138.

10. Gill, P. and J. Buckleton, A universal strategy to interpret DNA profiles that does not require a definition of low-copy-number. Forensic Science International: Genetics, 2010. 4(4): p. 221-227.

11. Lohmueller, K.E. and N. Rudin, Calculating the Weight of Evidence in Low-Template Forensic DNA Casework. Journal of Forensic Science,. 2013 58(1):p 243-249

12. Caragine, T., et al., Validation of Testing and Interpretation Protocols for Low Template DNA Samples Using AmpFℓSTR® Identifiler®. Croatian Medical Journal, 2009. 50: p. 250-267.

13. Mitchell, A.A., et al., Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in. Forensic Science International: Genetics, 2012. 6(6): p. 749-761.

14. Gill, P., et al., DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. Forensic Science International, 2006. 160(2-3): p. 90-101.

15. Caddy, B., G. Taylor, and A. Linacre, A review of the science of low template DNA analysis., 2008. http://www.bioforensics.com/articles/Caddy_Report.pdf

16. Gill, P., et al., DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods. Forensic Science International: Genetics, 2012. 6(6): p. 679-688.

17. Geddes, L., How DNA creates victims of chance. New Scientist, 2010(2774).

18. Butler, J. NIST Interlaboratory Mixture Interpretation Study 2005 (MIX05). 2005 [cited 2011 22 June];
Available from: http://www.cstl.nist.gov/biotech/strbase/interlab/MIX05.htm.

19. Dror, I.E. and G. Hampikian, Subjectivity and bias in forensic DNA mixture Interpretation. Science & Justice: Journal of the Forensic Science Society, 2011. 51(4): p. 204-208

20. (SWGDAM), SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. http://www.fbi.gov/about-us/lab/biometric-analysis/codis/swgdam-interpretation-guidelines [cited 2013].

21. Buckleton, J. and C.M. Triggs, Is the 2p rule always conservative? Forensic Science International, 2006. 159: p. 206-209.

22. Buckleton, J. and P. Gill, Low Copy Number, in Forensic DNA Evidence Interpretation. 2005, CRC Press. p. 275-297.

23. Bill, M., et al., PENDULUM--a guideline-based approach to the interpretation of STR mixtures. Forensic Science International, 2005. 148(2-3): p. 181-189.

24. Curran, J.M., A MCMC method for resolving two person mixtures. Science & Justice, 2008. 48(4): p. 168-177.

25. Perlin, M., et al., Validating Trueallele® DNA mixture interpretation Journal of Forensic Science, 2011. 56(6): p. 1430-1447.

26. Balding, D.J. and J. Buckleton, Interpreting low template DNA profiles. Forensic Science International: Genetics, 2009. 4(1): p. 1-10.

27. Bruce, B., J.E. Arthur, and D. Angela van, Low copy number typing has yet to achieve “general acceptance”. 2009. 2(1): p. 551-552.

28. Curran, J.M. and J. Buckleton, Inclusion Probabilities and Dropout. Journal of Forensic Sciences, 2010. 55(5): p. 1171-1173.

29. Puch-Solis, R., et al., Assigning weight of DNA evidence using a continuous model that takes into account stutter and dropout. Science & Justice, 2010. 50(1):p.46-47.

CHAPTER 3: A COMPARISON OF STATISTICAL MODELS FOR THE ANALYSIS OF COMPLEX FORENSIC DNA PROFILES

Hannah Kelly [a, b], Jo-Anne Bright [a], John Buckleton [a], James Curran [b]

a) ESR, PB 92021, Auckland, 1142, New Zealand,
b) Department of Statistics, University of Auckland, PB 92019, Auckland, 1142, New Zealand

This paper has been accepted by Science and Justice.

The DOI is: http://dx.doi.org/10.1016/j.scijus.2013.07.003

This paper is a summary of the current interpretation models available for the analysis of complex mixtures and low template DNA samples. In this paper "model" refers to the methods in which a genotype is determined and interpreted using the likelihood ratio. The models are broken down into three families; binary, semi-continuous, and continuous. The merits of each family are discussed alongside their limitations. This paper aims to educate forensic biologists on the models that are, or have been in use, and why those methods may be becoming outdated. Software is mentioned where appropriate so that analysts are aware of the limitations of any automated systems they may also be using. This paper aims to come across as a general discussion and the sometimes complicated mathematics of some of the models is intentionally avoided. The main objective of this work is to demonstrate to forensic biologists that there are limitations to the models that are currently used in complex mixtures and low template DNA sample interpretation and to show that the community will need to move towards the application of a continuous model.

This work was supported in part by grant 2011-DN-BX-K541 from the US National Institute of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of

Justice. The authors would like to thank Johanna Veth and Pauline Simon, for their contributions towards this paper.

3: A COMPARISON OF STATISTICAL MODELS FOR THE ANALYSIS OF COMPLEX FORENSIC DNA PROFILES

Keywords

Forensic DNA interpretation, binary model, drop model, continuous model

Abstract

Complex mixtures and LtDNA profiles are difficult to interpret. As yet there is no consensus within the forensic biology community as to how these profiles should be interpreted. This paper is a review of some of the current interpretation models, highlighting their weaknesses and strengths. It also discusses what a forensic biologist requires in an interpretation model and if this can be realistically executed under current justice systems.

3.1 Introduction

In forensic DNA analysis, a profile is typically produced from a biological sample collected from the scene of a crime and compared with the DNA profile of one or more persons of interest (POI). Traditional DNA analysis is sequential. Initially an electropherogram (epg) is produced. This raw output is processed by assigning peaks as allelic, stutter or artefactual. The deduced profile is then compared to the POI (if available), with the intention of producing either an inclusion, or an exclusion. If an inclusion is reached, then it is customary to provide a statistic to support the strength of the evidence. Analysis can involve either human or computerised processing, based on empirically devised guidelines, and can be complicated by factors such as the number of contributors to the profile, and the quality and quantity of the DNA.

Single source “pristine” profiles are relatively simple to interpret and their analysis has achieved worldwide acceptance as a reliable scientific method. However, profiles from crime scenes are frequently compromised in quality, or quantity, or both

(LtDNA). Stochastic factors are often present for such compromised profiles which complicates interpretation. These can include heterozygote imbalance, increased stutter peaks, allelic dropout, locus dropout, and drop in [1, 2]. Complicating interpretation even further is that in many cases, crime scene samples are composed from two or more people. Such profiles are referred to as mixtures.

The interpretation of mixtures can be difficult. The number of contributors is often unclear. The presence of three or more alleles at any locus signals the existence of more than one contributor, although it often is difficult to tell whether the sample originated from two, three, or even more individuals because the various contributors may share alleles. The number of contributors to the mixture is often assigned either by using the fewest number of individuals needed to explain the alleles [3-5], or by maximum likelihood methods [6]. In many cases there will be a major and a minor contributor present in the sample and the profiles can be resolved and interpreted as single source profiles. However, many profiles cannot be separated and are deemed "unresolvable". These complex mixtures are challenging to interpret and as yet, there is no consensus as to how such profiles should be dealt within the forensic biology community.

A 2010 article in New Scientist [7] highlights the disparity of practice in the interpretation of complex mixtures. In this article an epg from a previously analysed complex mixture was presented to 17 analysts in the same government laboratory for interpretation. Only one analyst agreed with the original finding, that the POI could not be excluded from the mixture. Four analysts deemed the evidence inconclusive, while the remaining 12 said that the POI could be excluded as having contributed to the mixture.

For a field which is widely regarded as objective, such a range of conclusions for the same evidence is worrying. Additionally, if the analyst is presented with the profile of a POI along with case circumstances strongly indicating that they are the offender, there is the perturbing issue of bias. If the accompanying statistic does not correctly represent the strength of the inclusion (or if no match statistic is provided) then there is the risk of the DNA evidence being misrepresented in court.

A 2005 study [8] highlights that not only are complex mixtures difficult to interpret, it can also be difficult to determine how many people have contributed to the mixture. The authors showed that more than 70% of four person mixtures could be wrongly interpreted as two or three person mixtures. In New Scientist [9] one of the authors from the 2005 study, Dan Krane states: “If you can’t determine how many contributors there were, it is ludicrous to suggest that you can tease apart who those contributors were or what their DNA profiles were”.

The following work is a review of some of the current interpretation models. We attempt here to highlight the weaknesses and strengths of these models. We also attempt to address the question of what a forensic biologist requires in a model and if this can be realistically implemented under current justice systems.

3.2 Calculation of a statistical weight

The DNA Commission of the International Society of Forensic Genetics (ISFG) recommends the use of the likelihood ratio (LR) in mixture interpretation [10]. The LR is accepted to be the most powerful and relevant statistic used to calculate the weight of the DNA evidence. It is the ratio of the probability of the evidence (E) given each of two competing hypotheses, H_1 and H_2, given all the available information, I. The available information, I, is taken to include the knowledge of the genotypes of the known contributors, K, the POI, S, and any other relevant and admissible evidence:

$$LR = \frac{\Pr(E \mid H_1, I)}{\Pr(E \mid H_2, I)}$$

The interpretation models discussed in this paper all utilise the likelihood ratio.

3.3 Interpretation models

3.3.1 The binary model

The binary model is probably better defined as a family of models rather than one specific model. The models in this family share the characteristic that they assign genotypes as possible or impossible given the data.

We define the genotype of the crime stain as G_c, and the genotypes of proposed donors as G_i for donor i. For an N donor mixture there are N proposed genotypes, G_i. We will denote the j^{th} combination of N genotypes, S_j. We can interpret the binary models as assigning a value of zero or one to $\Pr(G_C \mid S_j)$

The binary model assigns the values zero and one to the unknown probabilities, $\Pr(G_C \mid S_j)$, based on reasonable methods that approximate the relative values of $\Pr(G_C \mid S_j)$. In essence $\Pr(G_C \mid S_j)$ is assigned a value of zero if it is thought that this probability is very small relative to the other probabilities. $\Pr(G_C \mid S_j)$ is assigned a value of one if it is thought that this value is relatively large. As such, it is an approximation. Currently in most forensic biology laboratories this probability assignment is done manually and by the application of analysis thresholds and other rules based on empirical data.

Peak heights can vary between in epgs when replicates are run from the same sample. This variation between replicates from the same sample can be more dramatic if the sample is low template LtDNA. In LtDNA samples, some peaks at a locus may fail to reach the predetermined threshold to call a peak an allele in one replicate, but may exceed the threshold in a different replicate, therefore allowing it to be called. Since there is observable variation in replicates it is not possible that any crime scene profile (given a genotype set S_j) could occur with probability one, although zero is still possible. The reality is that all the probabilities, $\Pr(G_C \mid S_j)$, have some value in the interval $[0,1)$.

The most rudimentary implementation of the binary model treats alleles as present or absent and does not take into account peak height information [11-13]. We will term this the qualitative binary model.

Consider a set of allelic peaks $A_1 ..., A_M$. All sets of N genotypes that have these M alleles and no others are deemed included. Genotype sets are constrained by H_1 and H_2 (termed the allowed sets). The LR is assigned using the ratio of the sum of the probabilities of all allowed sets under H_1 and H_2

The computer programme POPSTATS, in common use in North America, implements this approach following the formulae of Weir et al. [12]. These formulae use the product rule and make no assessment of sampling uncertainty. This approach also appeared in the now obsolete DNAMIX I software [12]. It should be noted that this approach cannot be used if dropout is possible and if used may result in a seriously non-conservative assessment of the data. It is therefore not recommended for the interpretation of LtDNA or complex mixtures.

DNAMIX II extended this approach to include a subpopulation correction following NRC II recommendation 4.2 and implements the formulae of Curran et al. [13]. DNAMIX II makes no assessment of sampling uncertainty and, again, cannot be reliably used on profiles where dropout is possible.

DNAMIX III implements the formulae described in Curran et al. [13] and provides a limit on the confidence interval based on the work of Beecham and Weir [14]. The confidence interval itself is dependent on the extent of population substructure and the number of subpopulations. The software is not appropriate for profiles where dropout is possible.

Shortfalls in the qualitative binary approaches described above, such as the failure to take into account peak height and the inability to account for the possibility of dropout lead to the development of extensions which we will term the semi-quantitative binary model.

The semi-quantitative binary model declares some of the combinations that would have been allowed under the qualitative binary model as *possible* or *impossible* [5,

15]. Scientists use expert judgment together with a number of empirical guidelines to decide which genotype combinations at a locus can be excluded [5]. This assignment is often based on expert judgement or heuristics employing limits on variation in the mixture proportion (mx) and heterozygote balance (h).

The semi-quantitative model is mainly applied manually. However, GeneMapper® *ID-X* is a programme designed for the automated designation of forensic STR profiles [16]. It incorporates a mixture analysis tool that uses the number of peaks, peak height information, *mx* and interpretation guidelines to resolve two person mixed profiles in a semi-automated fashion based on Gill et al., [3].

Traditionally, the semi-quantitative binary model accounts for the possibility of drop-out by omitting the locus or using the $2p$ rule. The $2p$ rule assigns the probability $2\Pr(A_i)$ for the observation of a single allele, A_i, whose partner may have dropped out. The $2p$ rule had been assumed to be conservative in all circumstances, however this has proved a false assumption and is no longer recommended for use [10, 17].

One method to extend the binary model to profiles where dropout may have occurred (but alleles matching the POI are present within the profile) uses the '*F*' designation to denote an allele that may have dropped out or 'failed'. In this system the *F* designation represents any allele at the locus in question, including alleles already observed [18].

An alternate extension method uses a 'Q' designation in place of the *F*. A Q designation represents any allele at the locus except for those alleles already present. The formulae for the Q model can become very complex. As it is applied manually, this method is not readily extended to higher order mixtures (those containing more than two contributors) but there is the potential for automation of these extensions [19-21].

The UK Forensic Science Service (FSS) developed software, PENDULUM, that is automated and applies rules based on empirical data to assist in designating genotype sets as possible or not possible and uses the *F* designation [15]. However,

PENDULUM ends the process at these designations and does not proceed to calculate a LR, nor does it provide any other calculation of a statistical weight.

Binary models have served well for a number of years and in a great many cases.

The primary motivator for change is that the binary models described above cannot deal with a locus showing a non-concordance. This is a locus where at least one allele of the POI is not seen in the profile. In addition, none of the models can take into account multiple replicates. The challenges associated with the phenomena of dropout and drop in, in particular, have led to the evolution of a model which assesses the crime scene profile utilising primarily the concept of a probability of dropout.

3.3.2 The semi-continuous model

Figure 3.1 shows two examples of non-concordances when the POI is the genotype (7,9). Example A shows a large concordant 7 peak which is just under the homozygote threshold and no peak at the 9 allele position. Example B shows a small concordant 7 peak and a below threshold 9 peak. Previously both examples would have been treated using the $2p$ rule under the binary models. If we use subjectivity to assess the two examples we can see that they are both quite different. In reality there is considerable support for the genotype 7,7 in example A while in example B there is more support the genotype 7,9. Using the $2p$ rule in situations such as example A is non-conservative, which has led to the development of the Buckleton and Gill model [4, 22]

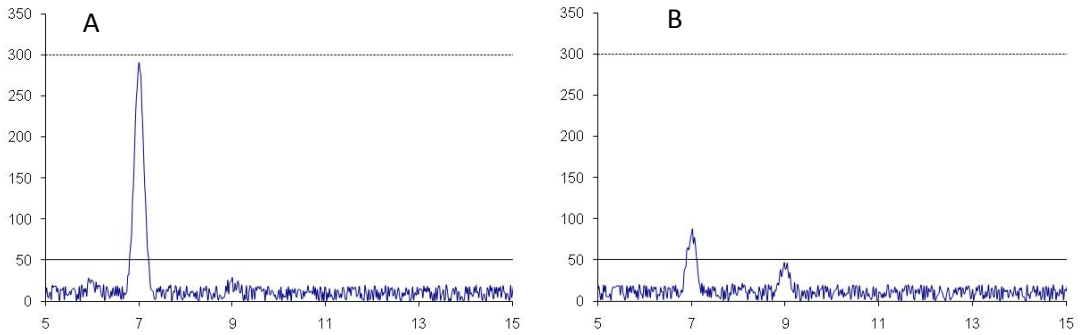


Figure 3.1. Two examples of non-concordance where POI = 7,9. A large concordant 7 allele with no 9 peak observed (non-tolerable non-concordance) and B small

concordant 7 allele with a non-concordant 9 peak visible sub-threshold (tolerable non-concordance). Stochastic threshold = 300 RFU, Limit of detection 50 RFU.

The Buckleton and Gill model assigns a probability to the event of an allele not appearing, $\Pr(D)$. This is usually shortened to D. (i.e. the probability that an allele would dropout) [19, 22, 23]. It can also factor in the presence of additional genetic material, referred to as drop in, $\Pr(C)$. In this model drop in is distinct from contamination. Drop in is not reproducible and is limited to only a few peaks per profile, whereas contamination refers to the presence of portions of reproducible extraneous DNA. This method also can cope with multiple replicates (for a more thorough discussion refer to Buckleton and Gill [24]).

The probability of dropout appears in both the numerator and denominator of the *LR*. There is no mathematical or logical reason why it should be the same in the numerator and denominator. There is also good reason to believe that it may be different at different loci within a profile and different between profiles.

The FSS implemented this approach in the software, LoComatioN [19]. However, the epg is still evaluated qualitatively first. The scientist must call peaks as alleles and assign stutter peaks. The assigned peaks are then entered into the computer program and the probabilities of the profile for all possible genotype sets are calculated. The software can calculate a likelihood ratio for a range of propositions manually entered into the program by the analyst. It enables a rapid evaluation of multiple propositions which would otherwise be laborious and error prone [19].

However, no peak height information is utilised when designating genotype sets. For example, in Figure 3.2, all of the genotype combinations would be given the same weight [20, 22, 25].

Tvedebrink et al., [26, 27] have suggested various improvements to the assignment of the probability of dropout. All of these methods use the profile itself to assess one or two covariates used to assign the probability of dropout. The treatment of the probability of dropout as a parameter assessed from the profile can be problematic as there is a recycling of the information. It would be better to treat the probability of dropout as a random variable and integrate it out [28]. This would require a sensible

distribution to describe the probability of dropout. Such a distribution would vary from case to case. As yet such concepts have been mentioned but not implemented.

The semi-continuous model is an improvement in the way complex mixtures and LtDNA profiles are interpreted. However it still does not make full use of the available information from the epg. Consider the epg shown in Figure 3.2. If we treat this as a two-person mixture, then six genotype combinations are deemed possible. These are:

Individual 1	**Individual 2**
7,9	11,13
7,11	9,13
7,13	9,11
9,11	7,13
9,13	7,11
11,13	7,9

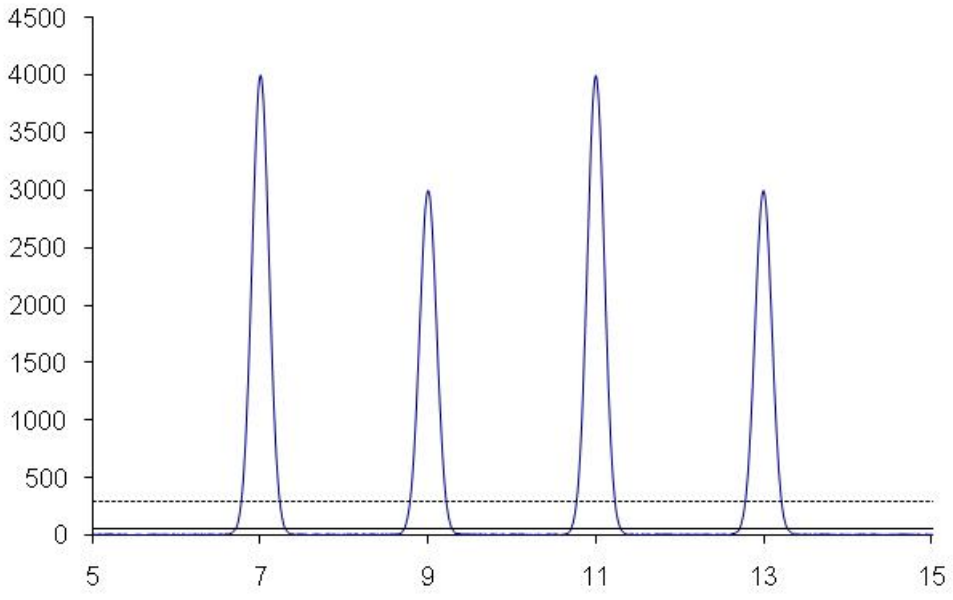


Figure 3.2 Artificial epg of four-peak locus for a two-person mixture

The combinations $7,11:9,13$ and $9,13:7,11$ are well supported by the peak heights. However, under the semi continuous model (and binary models) the profile is assigned the same probability for all of the genotype combinations listed. When this concept is extended to multiple loci only one combination will be the most supported.

Addressing these shortcomings leads into the concept of the continuous model. This model seeks move away from very discreet all $\left(\Pr(G_C \mid S_j) = 1\right)$ or nothing $\left(\Pr(G_C \mid S_j) = 0\right)$ nature of the binary model by making better use of the available information.

3.3.3 The continuous model

We define a fully continuous model for DNA interpretation as one which assigns a value to the probability $\Pr(G_C \mid S_j)$ using some model for peak heights for all peaks in the profile. These models have the potential to handle any type of non-concordance and may assess any number of replicates without pre-processing and the consequential loss of information. Continuous models are likely to require models to describe the stochastic behaviour of peak heights and potentially stutter.

Many of the qualitative or subjective decisions that the scientist has traditionally handled such as the designation of peaks as alleles, the allocation of stutters and possible allelic combinations may be removed. Instead, the model takes the quantitative information from the epg such as peak heights, and uses this information to calculate the probability of the peak heights given all possible genotype combinations. Removing the subjectivity or qualitative analysis of the profile will ensure consistency in DNA interpretation and reporting across laboratories.

TrueAllele is an example of commercial software implementing a continuous model [29].

3.4 General acceptance of a universal DNA model

It is appropriate, when assessing the advantages and weaknesses of these models, to begin by discussing which aspects of an interpretation model are desirable and/or suitable in the forensic context. Accuracy, reliability and comprehensibility are definitely desirable aspects of a DNA interpretation model. None of these are easy to define in this context.

If we think of the product of an interpretation model as a likelihood ratio, then we may think of accuracy as closest to the true answer. The true answer in DNA interpretation is somewhat elusive and plausibly does not exist at all. For this paper we will think of accuracy as making the best use of all the available information in a logically robust manner.

We will use the word reliability in this context to refer to the chance of serious misapplication of the method, either to a situation for which it is unsuited or misapplication to a situation for which it is suited.

Comprehensibility may come in two forms. Is the method comprehensible to the forensic scientist? Is the method explainable to a court? There is therefore interplay between comprehensibility to the scientist and reliability (Figure 3.3). This point is possibly worth some expansion.

It is often assumed that complex and especially computerised methods are at most risk and this is plausible. However the risk exists for any method, computerised or otherwise, to be misunderstood.

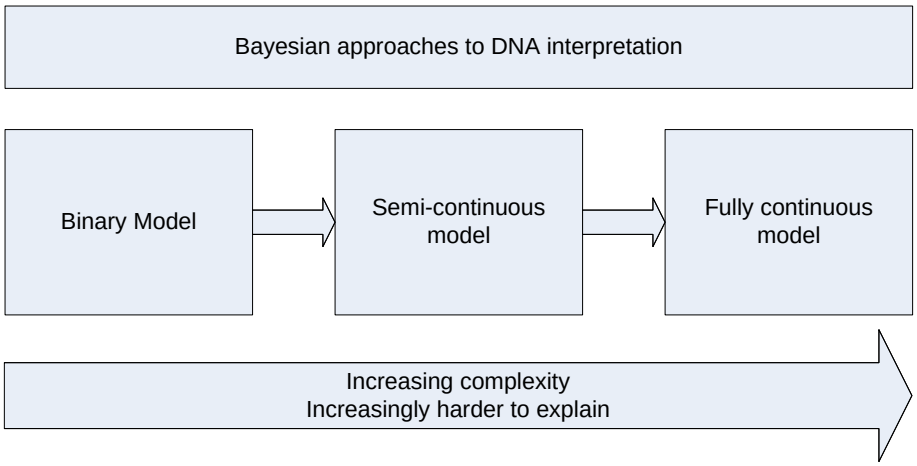


Figure 3.3 Summary of the relationship of the different models for forensic DNA interpretation

The simplest model, when assessed against these criteria, is the qualitative binary model. One could easily justify the argument that it is the most comprehensible and reliable. And yet there is evidence that this is not so. This method is not suitable for profiles where dropout is possible but it is often applied to such profiles. It may be that if interpretation is not given sufficient importance within an organisation, then adequate training and research resources may not be invested in it. Organisations giving low priority to interpretation may choose simple systems and also have low investment in interpretation training and research. The conclusion is that even the simplest method may descend into the category of misunderstood.

The semi-quantitative binary model, when applied manually as usually is the case, is the one that has the scientist most intimately involved in the interpretation. This places considerable training and research requirements on the organisation but in many ways this is a good thing. Parameters of importance for interpretation need to be assessed such as variability in heterozygote balance and stutter peak heights. Staff must be trained to a high degree of competence but, again, this is desirable both from a professional standards viewpoint, and from the ability of the scientist to represent the evidence in court. However the binary model, in any version, is incapable of handling non-concordances. This is the primary motivator for a move away from this method. There is also the difficulty in extending the model to multi-person mixtures.

The Buckleton and Gill model retains many of the best aspects of the semi-quantitative binary model but allows extension to profiles showing non-concordances. Software is required to extend to mixtures of three or more persons or to multiple replicates. Programs have been developed [19, 22].

Coming finally to the continuous model; this approach is undoubtedly the premier choice in terms of accuracy as defined here, if we can adequately model the behaviour of peak height with empirical observation and verify the mathematical logic of the development from these foundations. Such methods will need to be consigned to a computer. Training and research demands will be considerable to underpin the approach and to allow scientists to represent the evidence in court [30], as the continuous model is likely to be the least comprehensible of the three models [31, 32].

Additionally, computer software is only as reliable as the analyst that is using it. There is the risk that, with complicated automated programs, analysts will not understand the limitations and the program will be inadvertently used in situations where it is not appropriate to do so. However, properly developed and used, the continuous model will make the best use of the available information and give a considerable enhancement in objectivity [33, 34]. Replicates may be easily accommodated [35, 36]. The mathematics may be placed in the public domain by publication and hence it will be available for scrutiny by other qualified experts or subject to examination in court [37]. In many ways a well described mathematical process is more transparent than the often subjective decisions of experts.

3.5 Conclusion

DNA profiling is the stronghold in the characterisation of forensic biological evidence. The advent of increasingly more sensitive DNA analytical techniques has enabled scientists to generate profiles from samples that contain much lower amounts of DNA. This means that a wider range of evidence types can be analysed. However, the benefit of increased sensitivity, at times, means a reduction in profile quality and problems with profile interpretation due to the nature of the evidence types being sampled. Complex mixtures and LtDNA have stochastic factors present that complicate interpretation and current interpretation models are struggling.

Although extensions have been made to binary models we are being forced to move away from them, largely due to their inability to handle non-concordances but also by the difficulty in extending the semi-quantitative method to multi-person mixtures and the associated loss of information when expedients are used.

The options to move forward with are the semi-continuous Buckleton and Gill model and the continuous approach. Both of these are defensible scientifically. Of the two the continuous model makes best use of the available information. Since both are likely to be encapsulated into software the risk of them being misused must be ameliorated. This will be a challenge but perhaps a worthwhile one in terms of professionalism.

What must be decided is if we should move towards a model that is most likely to deliver us the more accurate answer, yet the mechanics are complex to explain to a jury and additionally raises the risk of our interpretation scientists becoming somewhat redundant, or if we should move towards a method where the scientist has a more hands on approach and the model is easier to explain to a jury but does not use all the information available.

Realistically, the model which makes best use of the available evidence has to be implemented. Therefore, we must advocate a move to a continuous method founded on sound biological models, which themselves are based on empirical data.

References

[1] B. Caddy, G. Taylor, A. Linacre, A review of the science of low template DNA analysis., (2008). http://www.bioforensics.com/articles/Caddy_Report.pdf

[2] The Forensic Science Regulator, Response to Professor Brian Caddy's Review of the Science of Low Template DNA Analysis, (2008). https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/117557/response-caddy-dna-review.pdf

[3] P. Gill, R.L. Sparkes, R. Pinchin, T. Clayton, J.P. Whitaker, J.S. Buckleton, Interpreting simple STR mixtures using allelic peak areas, Forensic Science International, 91 (1998) 41 - 53.

[4] J.S. Buckleton, C.M. Triggs, S.J. Walsh, Forensic DNA Evidence Interpretation, CRC Press, Boca Raton, Florida, 2004.

[5] T. Clayton, J.P. Whitaker, R.L. Sparkes, P. Gill, Analysis and interpretation of mixed forensic stains using DNA STR profiling, Forensic Science International, 91 (1998) 55 - 70.

[6] H. Haned, L. Pene, F. Sauvage, D. Pontier, The predictive value of the maximum likelihood estimator of the number of contributors to a DNA profile, Forensic Science International: Genetics, 5 (2011) 281-284.

[7] L. Geddes, Fallible DNA evidence can mean prison or freedom New Scientist, 2774 (2010).

[8] D.R. Paoletti, T.E. Doom, C.M. Krane, M.L. Raymer, D.E. Krane, Empirical analysis of the STR profiles resulting from conceptual mixtures, Journal of Forensic Sciences, 50 (2005) 1361-1366.

[9] L. Geddes, How DNA creates victims of chance, New Scientist, (2010).

[10] P. Gill, C.H. Brenner, J.S. Buckleton, A. Carracedo, M. Krawczak, W.R. Mayr, N. Morling, M. Prinz, P.M. Schneider, B.S. Weir, DNA commission of the

International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. , Forensic Science International., 160 (2006) 90-101.

[11] I.W. Evett, C. Buffery, G. Willott, D.A. Stoney, A guide to interpreting single locus profiles of DNA mixtures in forensic cases, Journal of the Forensic Science Society, 31 (1991) 41-47.

[12] B.S. Weir, C.M. Triggs, L. Starling, L.I. Stowell, K.A.J. Walsh, J.S. Buckleton, Interpreting DNA Mixtures, Journal of Forensic Sciences, 42 (1997) 213-222.

[13] J.M. Curran, C.M. Triggs, J.S. Buckleton, B.S. Weir, Interpreting DNA mixtures in structured populations, Journal of Forensic Sciences, 44 (1999) 987 - 995.

[14] G.W. Beecham, B.S. Weir, Confidence Interval of the Likelihood Ratio Associated with Mixed Stain DNA Evidence, Journal of Forensic Sciences, 56 (2011) S166-S171.

[15] M. Bill, P. Gill, J. Curran, T. Clayton, R. Pinchin, M. Healy, J. Buckleton, PENDULUM - A guideline based approach to the interpretation of STR mixtures, Forensic Science International, 148 (2005) 181-189.

[16] Applied Biosystems, GeneMapper® ID-X Software Mixture Analysis Tool. Quick Reference Guide, http://www3.appliedbiosystems.com/cms/groups/applied_markets_marketing/documents/generaldocuments/cms_087508.pdf (2008).

[17] J. Buckleton, C.M. Triggs, Is the 2p rule always conservative?, Forensic Science International, 159 (2006) 206-209.

[18] H. Kelly, J.-A. Bright, J. Curran, J. Buckleton, The interpretation of low level DNA mixtures, Forensic Science International: Genetics, 6 (2011) 191-197.

[19] P. Gill, A. Kirkham, J. Curran, LoComatioN: A software tool for the analysis of low copy number DNA profiles, Forensic Science International, 166 (2007) 128-138.

[20] H. Haned, Forensim: An open-source initiative for the evaluation of statistical methods in forensic genetics, Forensic Science International: Genetics, 5 (2011) 265-268.

[21] D. Balding, likeLTD (likelihoods for low-template DNA profiles), https://sites.google.com/site/baldingstatisticalgenetics/software/likeltd-r-forensic-dna-r-code.

[22] D.J. Balding, J. Buckleton, Interpreting low template DNA profiles, Forensic Science International: Genetics, 4 (2009) 1-10.

[23] P. Gill, J.P. Whitaker, C. Flaxman, N. Brown, J.S. Buckleton, An investigation of the rigor of interpretation rules for STR's derived from less that 100 pg of DNA, Forensic Science International, 112 (2000) 17-40.

[24] J.S. Buckleton, P. Gill, Low copy number, in: Forensic DNA Evidence Interpretation, CRC Press, Boca Raton, 2004, pp. 275-297.

[25] P. Gill, L. Gusmão, H. Haned, W.R. Mayr, N. Morling, W. Parson, L. Prieto, M. Prinz, H. Schneider, P.M. Schneider, B.S. Weir, DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods, Forensic Science International: Genetics, 6 (2012) 679-688.

[26] T. Tvedebrink, P.S. Eriksen, H.S. Mogensen, N. Morling, Estimating the probability of allelic drop-out of STR alleles in forensic genetics, Forensic Science International: Genetics, 3 (2009) 222-226.

[27] T. Tvedebrink, P.S. Eriksen, H.S. Mogensen, N. Morling, Statistical model for degraded DNA samples and adjusted probabilities for allelic drop-out, Forensic Science International: Genetics, 6 (2012) 97-101.

[28] D. Balding, Integrating D, Personal Communication, (2010).

[29] M.W. Perlin, M.M. Legler, C.E. Spencer, J.L. Smith, W.P. Allan, J.L. Belrose, B.W. Duceman, Validating TrueAllele® DNA mixture interpretation. , Journal of Forensic Sciences, 56 (2011) 1430-1447.

[30] M.W. Perlin, Easy reporting of hard DNA: Computer comfort in the courtroom., Forensic Magazine, 9 (2012) 32-37.

[31] Cybergenetics, DNA identification for scientists, http://www.cyben.com/information/courses/page.shtml (2012).

[32] M.W. Perlin, DNA identification science, in: C.H. Wecht (Ed.) Forensic Sciences, LexisNexis Matthew Bender, Albany, 2012.

[33] M.W. Perlin, A. Sinelnikov, An information gap in DNA evidence interpretation., PLoS ONE, 4 (2009) e8327.

[34] J. Curran, A MCMC method for resolving two person mixtures., Science & Justice, 48 (2008) 168-177.

[35] V.L. Pascali, S. Merigioli, Joint Bayesian analysis of forensic mixtures, Forensic Science International: Genetics, 6 (2012) 735-748.

[36] J. Ballantyne, E.K. Hanson, M.W. Perlin, DNA mixture genotyping by probabilistic computer interpretation of binomially-sampled laser captured cell populations: Combining quantitative data for greater identification information, Science & Justice, Article in press doi:10.1016/j.scijus.2012.04.004.

[37] M.W. Perlin, J. Galloway, Computer DNA evidence interpretation in the Real IRA Massereene terrorist attack., Evidence Technology Magazine, 10 (2012) 20-23.

CHAPTER 4: THE INTERPRETATION OF LOW LEVEL MIXTURES

Hannah Kelly [a, b], Jo-Anne Bright [a], James Curran [b], John Buckleton [a],

a) ESR, PB 92021, Auckland, 1142, New Zealand,
b) Department of Statistics, University of Auckland, PB 92019, Auckland, 1142, New Zealand

This paper appeared in Forensic Science International: Genetics, Volume 6, Issue 2, March 2012, pages 191-197.

The DOI is: http://dx.doi.org/10.1016/j.fsigen.2011.04.013

This paper compares three models that can be used to determine the possible genotype combinations present in a low level mixture. Two of these models use the quanititative data available from the profile but a third method, termed the unconstrained combinatorial approach, does not. The models that use the quantitative data (because they use quantitiative data) can accommodate the possible dropout of alleles unrelated to the person of interest. The unconstrained combinatorial approach cannot account for dropout. This paper takes a low level two person mixture and uses each of the three models to interpret the profile and compare with a POI. The resulting inclusion statistic is calculated using the likehood ratio. The resulting *LR*s differ in magnitude and demonstrate the variabilty in interpretation models. The aim of this paper is to show biologists that different *LR*s can result from the interpretation of the same profile and why it is important to understand the limitations of the model you are using.

This work was supported in part by grant 2011-DN-BX-K541 from the US National Institute of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice. The authors would like to thank Catherine McGovern, Richard Wivell, and two anonymous reviewers for their contributions towards this paper.

CHAPTER 4: AMENDMENTS

After publication the following errors were noticed in the manuscript:

Table 4.2: Person of interest profile (POI) and *LR* for each locus for the three different models;

- FGA contained an incorrect allele probability in the *LR* calculation.
- TH01 was missing a "2" from the calculation to account for the heterozygous locus.
- The above two errors lead to a miscalculation in the overall *LR* for each of the models.

These corrections have been updated in the text and are included below (Table A.1).

Table A.1: Amendments to Table 4.2: Person of interest profile (POI) and *LR* for each locus for the three different models

Locus	**POI**	$\boldsymbol{LR_{UC}}$	$\boldsymbol{LR_{Q}}$	$\boldsymbol{LR_{F}}$
TH01	7,7	1.02	3.46	3.46
FGA	21,25	5.98	5.98	5.98
LR		2.88×10^{06}	2.15×10^{05}	4.38×10^{03}

Table 4.3: *UC*, *F* and *Q* model *LR* calculations for locus D21S11, where POI = (28,30)

- The formula for the Pr(E|Hd) under the *F* model was missing an addition sign.
- The formula for the Pr(E|Hd) under the *Q* model was missing an addition sign.

These corrections have been updated in the text and are included below (Table A.2).

Table A.2: Amendments to Table 4.3: *UC*, *F* and *Q* model *LR* calculations for locus D21S11, where POI = (28,30)

F model	Allelic vector (28,30) Pr(E|Hd)= $\frac{12\left(\theta+(1-\theta)\,p_{28}\right)\left(\theta+(1-\theta)\,p_{30}\right)}{(1+\theta)(1+2\theta)}$
Q model	Allelic vector (28,30) Pr(E|Hd)= $\frac{2\left(\theta+(1-\theta)\,p_{28}\right)\left(\theta+(1-\theta)\,p_{30}\right)}{(1+\theta)(1+2\theta)}\times$ $\left[6-\frac{6\left(2\theta+(1-\theta)\,p_{28}\right)}{(1+3\theta)}-\frac{6\left(2\theta+(1-\theta)\,p_{30}\right)}{(1+3\theta)}+\frac{2\left(2\theta+(1-\theta)\,p_{28}\right)\left(3\theta+(1-\theta)\,p_{28}\right)}{(1+3\theta)(1+4\theta)}+\frac{2\left(2\theta+(1-\theta)\,p_{30}\right)\left(3\theta(1-\theta)\,p_{30}\right)}{(1+3\theta)(1+4\theta)}\right.$ $\left.+\frac{3\left(2\theta+(1-\theta)\,p_{28}\right)\left(2\theta+(1-\theta)\,p_{30}\right)}{(1+3\theta)(1+4\theta)}\right]$

4: THE INTERPRETATION OF LOW LEVEL MIXTURES

Abstract

The occurrence of mixed DNA profiles in forensic samples is not uncommon. Interpretation of these profiles however can be challenging. In this paper we compare three different models for interpreting mixed DNA profiles prior to the calculation of a statistical weight. Two of these models take into account the peak height of the alleles. The third method uses an unconstrained combinatorial approach. We compare the statistical weights calculated after applying the three different models to one low level two-person mixed DNA profile derived from a crime sample and provide tables of equations that can be applied to many different scenarios including single source and two and three person mixed DNA profiles.

Keywords

Forensic DNA interpretation, Low template DNA, mixed DNA profiles, Binary model

4.1 Introduction

In forensic DNA analysis, mixed DNA profiles arise from samples containing DNA from two or more people. In some circumstances alleles from a true contributor may not be visualised in the electropherogram (epg). This is thought to arise because of limitations in the quantity and or quality of the template DNA in the original sample and is called allele dropout. Epgs that exhibit exaggerated stochastic effects such as allele dropout are termed low template DNA (LtDNA).

In some cases DNA from one contributor to a mixed DNA profile may be present in a larger amount than DNA from another contributor. This component is sometimes referred to as originating from the major contributor and may be interpreted as a single source DNA profile. In other cases none of the contributors to the mixed DNA profile can be inferred.

Typically the hypotheses of interest will be whether or not the person of interest (POI), could be a contributor to the DNA profile. Although the genotype of this person is typically known it is desirable that all subjective judgements regarding the epg are made without knowledge of this genotype. This goes to issues of bias being raised by some scientists at court [1]. For example, the examiner should determine if dropout is possible at a locus before looking at a reference profile.

After these judgements are made the mixed profile is compared with the POI. If the POI cannot be excluded, then it is customary to provide some measure of the weight of the evidence. For mixed DNA samples this weight of evidence is typically supplied by calculating an exclusion probability (RMNE) or its complement the cumulative probability of inclusion (CPI), a random match probability (RMP), or a likelihood ratio (*LR*). Whenever dropout is a possibility a meaningful exclusion probability cannot be calculated for the full profile. RMNE may still have a meaning in this situation if, for example, two contributors represent the bulk of the DNA (the major contributors) and there is a third or additional trace contributor. Both the RMP and the *LR* approach may be extended to deal with situations where dropout is possible and there are no non-concordant alleles. A non-concordant allele is one present in the POI that is not visualised in the epg. The *LR* approach, but not the RMP, may be further extended to handle the situation where non-concordant alleles exist.

For mixed DNA profiles at low levels with exaggerated stochastic effects (LtDNA) the calculation of a *LR* may proceed by either a binary [2], a semi-continuous [3], or a fully continuous method [2, 4]. The binary method treats alleles as present or absent, the semi-continuous method assigns a probability to the events of dropout or non-dropout but still treats alleles as present or absent. Fully continuous methods deal with the probabilities of stochastic events (like dropout) based on the heights of the peaks visualised at a locus. These methods improve in power, flexibility and elegance in the order binary, semi-, and fully continuous. However the simplicity of binary methods retains much appeal and allows manual or semi-manual implementation. There is no modification of the binary method that can deal with a non-concordant allele in a comprehensive manner [5]. In this paper we explore methods to extend the

binary method to complex mixtures that have no non-concordant alleles when compared to the POI.

We term DNA mixtures *n* person mixtures where *n* stands for the number of contributors. It is inadvisable to push existing models too far with regard to the number of contributors. However there is considerable interest in interpreting at least 2 and 3 person mixtures. Computation, manual or otherwise, is simplified if a mixture can be resolved into profiles attributable to one or more contributors. Previous authors [6-10] have described methods to split a mixture, either fully or partially, into its component parts. The computation is simplified because the number of possibilities for the unknown contributors is reduced. Use of a mixture resolution method prior to evidence evaluation is termed the "constrained conditional method" [11] and the splitting termed "deconvolution". However in some complex mixtures such deconvolution is either not possible or is avoided simply for ease of implementation [11].

Methods have been offered that automate the application of the unconstrained combinatorial method in those cases where dropout is not possible [12, 13]. Neither of the original papers explicitly mentions the condition that dropout must not be a possibility for these methods to be applicable. Perhaps more regretfully the recently published ISFG recommendations [11] also fail to give an explicit warning. The use of the unconstrained combinatorial method to LtDNA mixtures without any allowance for dropout is a misapplication of the method. In this paper we have referred to it as the unconstrained combinatorial method (*UC*).

In this paper we describe two methods to extend the use of the binary method to complex mixtures where dropout is possible but where there are no non-concordant alleles. This type of modelling will not hold if there are any alleles present which are inconsistent with those of the individuals believed to be relevant to the hypothesis. We compare these methods with the *UC*.

The two methods that we introduce are termed the *F* and the *Q* methods where *F* is used to designate any allele and *Q* to designate any allele other than those already observed.

4.2 Methods

In order to understand the theory, the following statistical and biological principals are helpful.

4.2.1 Conversion of the observed peaks to alleles

It is often possible to infer the minimum number of copies of each type of each allele from the observed peaks at a locus. At a minimum there is one copy of each allele observed above the designated limit of detection but often it is possible to infer the existence of extra copies of these alleles. When converting peaks into alleles a number of contributing factors are taken into account including; the number of assumed contributors to a DNA profile, peak height ratio, average mixture proportion, and known conditioning profiles if available and applicable. The allelic vector may be different under the prosecution (H_p) and the defence (H_d) hypotheses since the known profiles differ by at least the POI.

Consider the example in Figure 4.1. If we assume this profile is a two person mixture, then the 14 peak exceeds the stochastic threshold (sometimes referred to as the homozygote threshold), T, where dropout of one allele at a homozygous locus is deemed possible. This indicates the presence of two 14 alleles. The allelic vector becomes $(13,14,14,15)$.

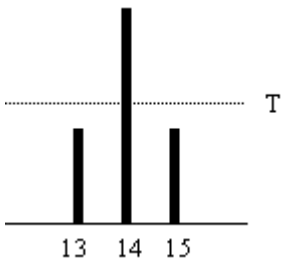


Figure 4.1 Example of a 3 peak locus, with one peak above and two below the stochastic threshold, *T*

4.2.2 Ambiguity in the allelic vector

In certain circumstances it may not be possible to infer a unique vector of alleles, but there might be a limited number of possibilities. Consider, for example, the scenario in Figure 4.2. This is an idealized example of a two person mixture with three peaks of equal height above *T*. We can infer that there is one more copy of the 13, 14, or 15 allele but we cannot tell which allele is duplicated. The possible allelic vectors are therefore $(13,13,14,15)$, $(13,14,14,15)$ and $(13,14,15,15)$.

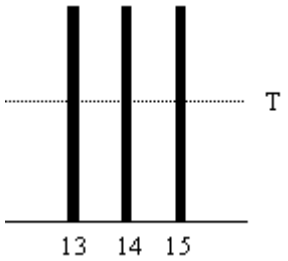


Figure 4.2 Example of a 3-peak locus, with all peaks above the stochastic threshold, *T*.

4.2.3 Permutations

Once the peaks at a locus have been converted to alleles the next step is to calculate the number of permutations. The number of permutations is the number of ways that the alleles can be arranged as pairs. For example; if at a given locus the alleles 13, 14, 15, and 16 are observed, then the possible allele combinations for a two person mixture are as follows:

Individual 1	Individual 2
13,14	15,16
13,15	14,16
13,16	14,15
14,15	13,16
14,16	13,15
15,16	13,14

Within each of the combinations above each set can be ordered two ways, for example 13,14 or 14,13 and 15,16 or 16,15. This factor of two for each combination, takes into account the allele being maternal or paternal in origin. For each combination there are 4 possible orders, so for 6 combinations the permutation set is 24.

In general the number of permutations at a locus exhibiting m_i copies of allele A_i is given by

$$\binom{n}{m_1, m_2, \ldots, m_l} = \frac{n!}{m_1!m_2!,\ldots,m_l!}$$

Where n is the total number of alleles at the locus and m is the number of times each allele is seen at the locus. The ! denotes a factorial. The factorial of a positive integer j, denoted by $j!$, is the product of all positive integers fewer than or equal to j. For example, $4! = 4 \times 3 \times 2 \times 1 = 24$.

For Figure 4.1, the number of permutations is $\frac{4!}{1!2!1!} = 12$.

4.2.4 Assigning the allelic designation for a locus

We assume that the observed peaks have been converted into alleles for a mixed profile with n contributors and the total number of alleles at that locus is $\leq 2n$. In some instances where dropout may have occurred a "placeholder" is required. For example, consider a locus where the alleles 13,14 and 15 are observed and the peak heights are below stochastic threshold, T indicating that dropout is possible. We might

assume that the profile is a two person mixture on the basis of the number of (above threshold) peaks seen at other loci. The proposed allelic vector then becomes $(13,14,15,F)$ where *F* is the placeholder indicating possible allelic dropout.

4.2.5 Assigning probability using the *F* model

An *F* designation represents all possible alleles at a locus, including those already observed, and is used to denote possible dropout. Figure 4.3 shows an example of a three peak locus in a two person mixed DNA profile

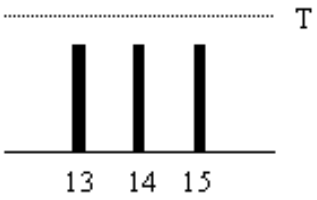


Figure 4.3 Example of a 3-peak locus, with all peaks below the stochastic threshold, T

Only one *F* allele is needed to explain the possible allelic vector at this locus where the fourth allele may or may not have dropped out. In this example *F* represents all possible alleles at that locus, including the alleles 13, 14, 15. The number of permutations is calculated using *F* as a placeholder and then *F* is dropped from the equation. The allelic vector becomes $(13,14,15,F)$. Therefore, for the example given in Figure 4.3 we seek to calculate $\Pr(13,14,15,F \mid X)$ where *X* is the set of "conditioning" alleles that consists of individuals relevant to the hypothesis such as the POI/suspect/complainant profile.

$$\begin{aligned}\Pr(13,14,15,F \mid X) &= \frac{4!}{1!1!1!1!}\Pr(13,14,15,F \mid X) \\ &= 24\Pr(13,14,15 \mid X)\end{aligned}$$

4.2.6 Assigning probability using the *Q* model

The *Q* model uses more information than the *F* model by calculating the probability of all possible allelic combinations for that locus. In the Figure 3 example above the possible allelic combinations are:

[13,13,14,15] [13,14,14,15] [13,14,15,15] [13,14,15,*Q*]

In this model, the *Q* designation represents all other possible alleles at that locus except for the alleles already observed. In this example the *Q* represents all other alleles possible at this locus except for (13,14,15).

The number of permutations for each allelic combination is worked out as for the *F* model. The number of permutations is calculated using *Q* as a placeholder. *Q* is then substituted out of the calculation. Because the probability of all alleles at a locus adds to one, the probability of the composite allele *Q* is one minus the observed alleles at that locus. In this example the probability of *Q* is $p_Q = 1 - p_{13} - p_{14} - p_{15}$, where p_i is the frequency of allele A_i in the subpopulation.

Q is best substituted out at the end stages of the calculation for ease. For example, the full calculation using the scenario above develops into:

$$\begin{aligned}
&\frac{4!}{2!1!1!}\Pr(13,13,14,15 \mid X) + \frac{4!}{1!2!1!}\Pr(13,14,14,15 \mid X) \quad + \\
&\frac{4!}{1!1!2!}\Pr(13,14,15,15 \mid X) + \frac{4!}{1!1!1!1!}\Pr(13,14,15,Q \mid X) \quad = \quad 12\Pr(13,14,15 \mid X) \\
&\qquad\qquad \times \begin{bmatrix} \Pr(13 \mid 13,14,15,X) \\ +\Pr(14 \mid 13,14,15,X) \\ +\Pr(15 \mid 13,14,15,X) \\ +2\Pr(Q \mid 13,14,15,X) \end{bmatrix}
\end{aligned}$$

Ignoring population substructure (which we do not recommend for criminal cases), this becomes

$$12 p_{13} p_{14} p_{15} \left(p_{13} + p_{14} + p_{15} + 2\left(1 - p_{13} + p_{14} + p_{15}\right)\right) = 12 p_{13} p_{14} p_{15} \left(2 - p_{13} - p_{14} - p_{15}\right)$$

4.2.7 Assigning probability using the *UC* method

The *UC* method does not take into account peak heights and therefore, does not convert peaks to alleles or account for the possibility of dropout.

It sums the probabilities of all sets of n contributors ($2n$ alleles) that explain the mixture but have no alleles outside the mixture. It is the requirement to "have no alleles outside the mixture" that leads to the discrepancy when dropout is possible.

For the example given in Figure 4.3, using the *UC* method, the probability (ignoring substructure) becomes:

$$12\Pr(13,14,15\mid X)\begin{bmatrix}\Pr(13\mid 13,14,15,X)\\ +\Pr(14\mid 13,14,15,X)\\ +\Pr(15\mid 13,14,15,X)\end{bmatrix}=12p_{13}p_{14}p_{15}\left(p_{13}+p_{14}+p_{15}\right)$$

This probability remains the same for all loci that have three peaks regardless of whether one of those peaks represents two alleles or if one or all of the peaks are below the stochastic threshold T.

4.2.8 Applying the sampling formula

Once the allele designations have been applied and the permutation multipliers have been calculated, the sampling formula of Balding and Nichols [14] is applied.

The sampling formula is a correction term for population substructure that models the belief that if we have observed allele A in the population before then we are more likely to see it again. More specifically, if there are x copies of allele A in a sample, (X), of n alleles then the probability that the next allele will be of type A is:

$$\Pr(A\mid X)=\frac{x\theta+(1-\theta)p_a}{1+(n-1)\theta}$$

where p_a is the probability of allele A, and θ is the coancestry coefficient (F_{ST}).

4.2.9 Comparison

In general, the probabilities produced by the *UC* model are smaller than those produced by the *Q* model, which in turn are smaller than those produced by the *F* model. Since the *LR* is the ratio of two of these probabilities it is not trivial to predict the effect on the *LR*. In our experience the *LR* for the *F* model is always lower than the *LR* for the *Q* model and we are unable to construct any artificial situation where the *LR* for the *Q* model is less than that for the *F* model.

There is a variety of methodologies and models for resolving mixed DNA profiles and calculating weights of evidence. It is expected that the models which use more of the information might be considered to give answers which are nearer to a fair and reasonable assessment of the evidence. Of the three models discussed here the *Q* model makes best use of the available information.

4.2.9.1 Example

In this example we compare the three methods, described in the previous section, for assigning an *LR* to an unresolvable two person mixed DNA profile. The profile epg is given in Figure 4.5 and the peaks and heights in Table 4.1. The person of interest profile is displayed in Table 4.2. The calculated *LR* for each of the three models is also provided for each locus and a combined total.

The calculations using each of the three methods with two loci, D21S11 and D3S1358, are provided in Tables 4.3 and 4.4 respectively. The *LR*s were calculated based on the allele frequencies for the New Zealand Caucasian population and a θ of 0.02 [15]. The stochastic threshold we have applied is 300 RFU.

In the supplementary material we provide the formulae for many situations for 1 to 3 contributors, respectively, using the *F*, *Q*, and *UC* models. In addition, we also provide calculations for two loci for a three person mixed DNA profile.

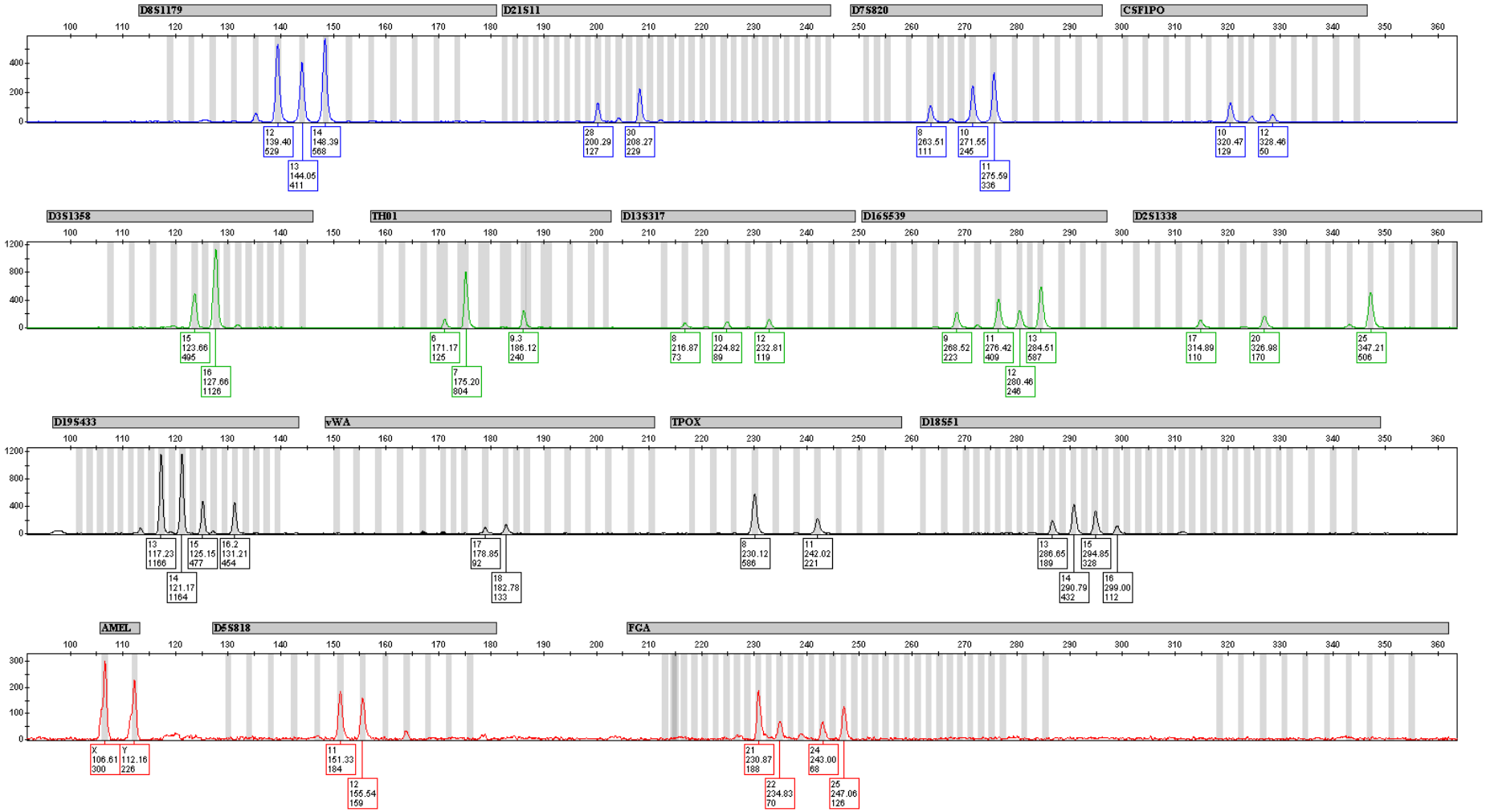


Figure 4.5 Epg of a two-person Identifiler® mixed DNA profile

Table 4.1 Peaks and heights from DNA profile, Figure 4.5, and allelic vector used for the *F* and *Q* models, taking into account peak heights

Marker	Peak	Height (RFU)	Allelic vector for *F* and *Q* models
D8S1179	12	529	(12, 13, 14)
	13	411	
	14	568	
D21S11	28	127	(28, 30)
	30	229	
D7S820	8	111	(8, 10, 11)
	10	245	
	11	336	
CSF1PO	10	129	(10, 12)
	12	50	
D3S1358	15	495	(15, 15, 16, 16)
	16	1126	
TH01	6	125	(6, 7, 7, 9.3)
	7	804	
	9.3	240	
D13S317	8	73	(8, 10, 12)
	10	89	
	12	119	
D16S539	9	223	(9, 11, 12, 13)
	11	409	
	12	246	
	13	587	
D2S1338	17	110	(17, 20, 25, 25)
	20	170	
	25	506	
D19S433	13	1166	(13, 14, 15, 16.2)
	14	1164	
	15	477	
	16.2	454	
vWA	17	92	(17, 18)
	18	133	
TPOX	8	586	(8, 8, 11)
	11	221	
D18S51	13	189	(13, 14, 15, 16)
	14	432	
	15	328	
	16	112	
D5S818	11	184	(11, 12)
	12	159	
FGA	21	188	(21, 22, 24, 25)
	22	70	
	24	68	
	25	126	

Table 4.2 Person of interest profile (POI) and *LR* for each locus for the three different models, from Figure 4.5

Marker	**POI**	LR_{UC}	LR_{Q}	LR_{F}
D8S1179	12,14	4.14	0.67	0.44
D21S11	28,30	5.74	3.07	1.86
D7S820	10,11	3.72	2.10	1.56
CSF1PO	10,10	2.23	0.96	0.56
D3S1358	15,15	2.61	1.95	1.95
TH01	7,7	1.02	3.46	3.46
D13S317	8,10	10.95	9.12	7.96
D16S539	11,13	1.52	1.52	1.52
D2S1338	25,25	2.41	7.77	7.77
D19S433	13,14	1.02	1.02	1.02
vWA	17,18	4.42	2.48	1.40
TPOX	8,8	1.06	0.98	0.56
D18S51	14,15	3.23	3.23	3.23
D5S818	11,12	2.10	1.51	0.63
FGA	21,25	5.98	5.98	5.98
		$2.88x10^{06}$	$2.15x10^{05}$	$4.38x10^{03}$

Table 4.3: *UC*, *F* and *Q* model *LR* calculations for locus D21S11, where POI = (28,30)

UC	Allelic vector (28,30) Pr(E|Hp) = 1 x 2Pr(28,30|28,30) + Pr(28,28|28,30) + Pr(30,30|28,30) $\left[\frac{2\left(\theta+(1-\theta)\,p_{28}\right)\left(\theta+(1-\theta)\,p_{30}\right)}{(1+\theta)(1+2\theta)}\right]+\left[\frac{\left(\theta+(1-\theta)\,p_{28}\right)\left(2\theta+(1-\theta)\,p_{28}\right)}{(1+\theta)(1+2\theta)}\right]+\left[\frac{\left(\theta+(1-\theta)\,p_{30}\right)\left(2\theta+(1-\theta)\,p_{30}\right)}{(1+\theta)(1+2\theta)}\right]$ Pr(E|Hd)= 2Pr(28,30|28,30)[2Pr(28,28|28,30,28,30) + 3Pr(28,30|28,30,28,30) + 2Pr(30,30)|28,30,28,30] $\frac{2\left(\theta+(1-\theta)\,p_{28}\right)\left(\theta+(1-\theta)\,p_{30}\right)}{(1+\theta)(1+2\theta)}\times\left[\frac{2\left(2\theta+(1-\theta)\,p_{28}\right)\left(3\theta+(1-\theta)\,p_{28}\right)}{(1+3\theta)(1+4\theta)}+\frac{3\left(2\theta+(1-\theta)\,p_{28}\right)\left(2\theta+(1-\theta)\,p_{30}\right)}{(1+3\theta)(1+4\theta)}+\frac{2\left(2\theta+(1-\theta)\,p_{30}\right)\left(3\theta+(1-\theta)\,p_{30}\right)}{(1+3\theta)(1+4\theta)}\right]$
F model	Allelic vector (28,30) Pr(E|Hp)= 1

	Pr(E|Hd)= 12Pr(28,30|28,30) $\frac{12\left(\theta+(1-\theta)\,p_{28}\right)\left(\theta+(1-\theta)\,p_{30}\right)}{(1+\theta)(1+2\theta)}$
Q model	Allelic vector (28,30) Pr(E|Hp)= 1 Pr(E|Hd)= $4\Pr(28,28,28,30\mid 28,30)+6\Pr(28,28,30,30\mid 28,30)+4\Pr(28,30,30,30\mid 28,30)+12\Pr(28,28,30,Q\mid 28,30)$ $+12\Pr(28,30,30,Q\mid 28,30)$ $+12\Pr(28,30,Q,Q\mid 28,30)$ $=2\Pr(28,30\mid 28,30)\times\begin{bmatrix}6-6\Pr(28\mid 28,28,30,30)-6\Pr(30\mid 28,28,30,30)+2\Pr(28,28\mid 28,28,30,30)\\+2\Pr(30,30\mid 28,28,30,30)\\+3\Pr(28,30\mid 28,28,30,30)\end{bmatrix}$

	$\frac{2\left(\theta+(1-\theta)\,p_{28}\right)\left(\theta+(1-\theta)\,p_{30}\right)}{(1+\theta)(1+2\theta)}\times$ $\left[6-\frac{6\left(2\theta+(1-\theta)\,p_{28}\right)}{(1+3\theta)}-\frac{6\left(2\theta+(1-\theta)\,p_{30}\right)}{(1+3\theta)}+\frac{2\left(2\theta+(1-\theta)\,p_{28}\right)\left(3\theta+(1-\theta)\,p_{28}\right)}{(1+3\theta)(1+4\theta)}+\frac{2\left(2\theta+(1-\theta)\,p_{30}\right)\left(3\theta(1-\theta)\,p_{30}\right)}{(1+3\theta)(1+4\theta)}\right.$ $\left.+\frac{3\left(2\theta+(1-\theta)\,p_{28}\right)\left(2\theta+(1-\theta)\,p_{30}\right)}{(1+3\theta)(1+4\theta)}\right]$

Table 4.4: *UC*, *F* and *Q* model *LR* calculations for locus D3S1358, where POI (15,15).

UC	Allelic vector (15,16) Pr(E|Hp)= Pr(16,16|15,15) + 2Pr(15,16|15,15) $\left[\frac{\left((1-\theta)\,p_{16}\right)\left(\theta+(1-\theta)\,p_{16}\right)}{(1+\theta)(1+2\theta)}+\frac{2\left(2\theta+(1-\theta)\,p_{15}\right)\left((1-\theta)\,p_{16}\right)}{(1+\theta)(1+2\theta)}\right]$ $\Pr(\text{E}|\text{Hd})= 2\Pr(15,16\,|\,15,15)\left[2\Pr(15,15\,|\,15,16,15,15)+3\Pr(15,16\,|\,15,16,15,15)+2\Pr(16,16)\,|\,15,16,15,15\right]$ $\frac{2\left(2\theta+(1-\theta)\,p_{15}\right)\left((1-\theta)\,p_{16}\right)}{(1+\theta)(1+2\theta)}\times\left[\begin{array}{l}\frac{2\left(3\theta+(1-\theta)\,p_{15}\right)\left(4\theta+(1-\theta)\,p_{15}\right)}{(1+3\theta)(1+4\theta)}+\frac{3\left(3\theta+(1-\theta)\,p_{15}\right)\left(\theta+(1-\theta)\,p_{16}\right)}{(1+3\theta)(1+4\theta)}\\ +\frac{2\left(\theta+(1-\theta)\,p_{16}\right)\left(2\theta+(1-\theta)\,p_{16}\right)}{(1+3\theta)(1+4\theta)}\end{array}\right]$
F model	Allelic vector (15,15,16,16) Pr(E|Hp)= Pr(16,16|15,15)

	$\frac{\left((1-\theta)\,p_{16}\right)\left(\theta+(1-\theta)\,p_{16}\right)}{(1+\theta)(1+2\theta)}$ Pr(E|Hd)= 6Pr(15,15,16,16|15,15) $\frac{6\left(2\theta+(1-\theta)\,p_{15}\right)\left(3\theta+(1-\theta)\,p_{15}\right)\left((1-\theta)\,p_{16}\right)\left(\theta+(1-\theta)\,p_{16}\right)}{(1+\theta)(1+2\theta)(1+3\theta)(1+4\theta)}$
Q model	Allelic vector (15,15,16,16) Pr(E|Hp)= Pr(16,16|15,15) $\frac{\left((1-\theta)\,p_{16}\right)\left(\theta+(1-\theta)\,p_{16}\right)}{(1+\theta)(1+2\theta)}$ Pr(E|Hd)= 6Pr(15,15,16,16|15,15) $\frac{6\left(2\theta+(1-\theta)\,p_{15}\right)\left(3\theta+(1-\theta)\,p_{15}\right)\left((1-\theta)\,p_{16}\right)\left(\theta+(1-\theta)\,p_{16}\right)}{(1+\theta)(1+2\theta)(1+3\theta)(1+4\theta)}$

4.3 Conclusions

In this work, we have extended the binary model using two different methods to deal with complex mixtures that have no non-concordant alleles. Both methods are able to take into account the possibility of dropout. We have compared the performance of these models with the misapplication of the unconstrained combinatorial method, *UC*. We stress that these methods will not hold if there are non-concordances.

The *UC* method is not appropriate. It does not use all the information available and may be extremely misleading. In addition as the formulae used do not deal with dropout the effect on the *LR* is unpredictable. There is a considerable risk that the *LR* will be significantly non-conservative. We are unable to recommend the *UC* method whenever dropout is possible for this reason. These limitations in the *UC* method are known [12, 13] and analysts should be aware of them when using this method for the interpretation of LtDNA profiles.

The *F* model does not use all of the available information as efficiently as the *Q* model. However, the formulae are easier to apply and in our experience the *F* model always produces a lower *LR* than the *Q* model. It is possible to produce *LR*s using the *F* model that are so much lower than those produced by the *Q* model that we feel that in these situations, the *F* model assigns the wrong weight to the evidence. For example, the *F* model may produce a *LR* less than 1 in some situations where such a *LR* is not reasonable when compared with the *Q* model.

The *Q* model is more complex than the *F* model and for analysts may appear somewhat more daunting. However, the formulae for both models can be easily implemented in a computer programme which, once validated, lifts considerable burden from the analyst

All things considered, we support the use of the *Q* model as it makes best use of the available evidence. However, the *F* model is still an acceptable approximation in some situations and is clearly easier to implement and less error prone.

It is important to highlight that none of these methods make full use of the available information. They are effectively methods to extend the working life of the binary model but better models are now becoming available. More intelligent models that

can accommodate an assessment of the probability of dropout and drop in offer a way forward that makes better use of the available data. These could include semi-continuous models like LoComatioN [16] or fully-continuous models like TrueAllele [4]. We are also aware that the Forensic Science Service has a fully continuous model in development (DNA Insight), but this is not currently commercially available [17]. These models, in our opinion are better suited for the interpretation of LtDNA profiles than the binary model.

Appendix A: Supplementary material

Table 4.5: Formulae for many situations for 1 contributor using the *F*, *Q* and *UC* models. We use X to represent the alleles from known individuals who may be assumed to be from the same subpopulation as the POI.

Peaks	Alleles	Model	Formula
2	2 (A,B)	*UC*, *F* & *Q*	$2\Pr(A,B \mid X)$
1	ALL	*UC*	$\Pr(A,A \mid X)$
	2 (A,A)	*F* & *Q*	$\Pr(A,A \mid X)$
	1 (A)	*F*	$2\Pr(A \mid X)$
		Q	$\Pr(A \mid X)\left(2-\Pr(A \mid X)\right)$

Table 4.6: Formulae for many situations for 2 contributors using the *F*, *Q* and *UC* models

Peaks	Alleles	Model	Formula
4	4 (A,B,C,D)	*UC*, *F* & *Q*	$24\Pr(A,B,C,D \mid X)$
3	ALL	*UC*	$12\Pr(A,B,C \mid X)\left(\Pr(A \mid A,B,C,X)+\Pr(B \mid A,B,C,X)+\Pr(C \mid A,B,C,X)\right)$
	4 (A,B,C,D)	*F* & *Q*	$12\Pr(A,A,B,C \mid X)$
	3 (A,B,C)	*F*	$24\Pr(A,B,C \mid X)$
		Q	$12\Pr(A,B,C \mid X)\left(2-\Pr(A \mid A,B,C,X)-\Pr(B \mid A,B,C,X)-\Pr(C \mid A,B,C,X)\right)$
2	ALL	*UC*	$2\Pr(A,B \mid X)\left(2\Pr(A,A \mid A,B,X)+3\Pr(A,B \mid A,B,X)+2\Pr(B,B \mid A,B,X)\right)$
	4 (A,A,B,B)	*F* & *Q*	$6\Pr(A,A,B,B \mid X)$
	4 (A,A,A,B)	*F* & *Q*	$4\Pr(A,A,A,B \mid X)$
	3 (A,A,B)	*F*	$12\Pr(A,A,B \mid X)$
		Q	$2\Pr(A,A,B \mid X)\left(6-4\Pr(A \mid A,A,B,X)-3\Pr(B \mid A,A,B,X)\right)$
	2 (A,B)	*F*	$12\Pr(A,B \mid X)$
		Q	$2\Pr(A,B \mid X)\begin{pmatrix}6-6\Pr(A \mid A,B,X)-6\Pr(B \mid A,B,X)+2\Pr(A,A \mid A,B,X)+2\Pr(B,B \mid A,B,X)\\ +3\Pr(A,B \mid A,B,X)\end{pmatrix}$
1	ALL	*UC*	$\Pr(A,A,A,A \mid X)$
	4	*F* & *Q*	$\Pr(A,A,A,A \mid X)$
	3	*F*	$4\Pr(A,A,A \mid X)$

		Q	$\Pr(A,A,A\mid X)\left(4-3\Pr(A\mid A,A,X)\right)$
	2	F	$6\Pr(A,A\mid X)$
		Q	$\Pr(A,A\mid X)\left(6-8\Pr(A\mid A,A,X)+3\Pr(A,A\mid A,A,X)\right)$
	1	F	$4\Pr(A\mid X)$
		Q	$\Pr(A\mid X)\left(4-6\Pr(A\mid A,X)+4\Pr(A,A\mid A,X)-\Pr(A,A,A\mid A,X)\right)$

Table 4.7: Formulae for many situations for 3 contributors using the *F*, *Q* and *UC* models

Peaks	Alleles	Model	Formula
6	6 (A,B,C,D,E,F)	*UC*	$720\Pr(A,B,C,D,E,F \mid X)$
		F & *Q*	$720\Pr(A,B,C,D,E,F \mid X)$
5	ALL	*UC*	$360\Pr(A,B,C,D,E \mid X)\begin{pmatrix}\Pr(A \mid A,B,C,D,E,X)+\Pr(B \mid A,B,C,D,E,X)+\Pr(C \mid A,B,C,D,E,X)\\+\Pr(D \mid A,B,C,D,E,X)+\Pr(E \mid A,B,C,D,E,X)\end{pmatrix}$
	6 (A,A,B,C,D,E)	*F* & *Q*	$360\Pr(A,A,B,C,D,E \mid X)$
	5 (A,B,C,D,E)	*F*	$720\Pr(A,B,C,D,E \mid X)$
		Q	$360\Pr(A,B,C,D,E \mid X)\begin{pmatrix}2-\Pr(A \mid A,B,C,D,E,X)-\Pr(B \mid A,B,C,D,E,X)-\Pr(C \mid A,B,C,D,E,X)\\-\Pr(D \mid A,B,C,D,E,X)-\Pr(E \mid A,B,C,D,E,X)\end{pmatrix}$
4	ALL	*UC*	$60\Pr(A,B,C,D \mid X)\begin{pmatrix}2\Pr(A,A \mid A,B,C,D,X)+3\Pr(A,B \mid A,B,C,D,X)+3\Pr(A,C \mid A,B,C,D,X)\\+3\Pr(A,D \mid A,B,C,D,X)+2\Pr(B,B \mid A,B,C,D,X)+3\Pr(B,C \mid A,B,C,D,X)\\+3\Pr(B,D \mid A,B,C,D,X)+2\Pr(C,C \mid A,B,C,D,X)+3\Pr(C,D \mid A,B,C,D,X)\\+2\Pr(D,D \mid A,B,C,D,X)\end{pmatrix}$
	6 (A,A,B,B,C,D)	*F* & *Q*	$180\Pr(A,A,B,B,C,D \mid X)$
	6 (A,A,A,B,C,D)	*F* & *Q*	$120\Pr(A,A,A,B,C,D \mid X)$
	5 (A,A,B,C,D)	*F*	$360\Pr(A,A,B,C,D \mid X)$

		Q	$60\Pr(A,A,B,C,D\mid X)\begin{pmatrix}6-4\Pr(A\mid A,A,B,C,D,X)-3\Pr(B\mid A,A,B,C,D,X)-3\Pr(C\mid A,A,B,C,D,X)\\-3\Pr(D\mid A,A,B,C,D,X)\end{pmatrix}$
	4 (A,B,C,D)	*F*	$360\Pr(A,B,C,D\mid X)$
		Q	$60\Pr(A,B,C,D\mid X)\begin{pmatrix}6-6\Pr(A\mid A,B,C,D,X)-6\Pr(B\mid A,B,C,D,X)-6\Pr(C\mid A,B,C,D,X)\\-6\Pr(D\mid A,B,C,D,X)+2\Pr(A,A\mid A,B,C,D,X)+3\Pr(A,B\mid A,B,C,D,X)\\+3\Pr(A,C\mid A,B,C,D,X)+3\Pr(A,D\mid A,B,C,D,X)+2\Pr(B,B\mid A,B,C,D,X)\\+3\Pr(B,C\mid A,B,C,D,X)+3\Pr(B,D\mid A,B,C,D,X)+2\Pr(C,C\mid A,B,C,D,X)\\+3\Pr(C,D\mid A,B,C,D,X)+2\Pr(D,D\mid A,B,C,D,X)\end{pmatrix}$
3	ALL	*UC*	$30\Pr(A,B,C\mid X)\begin{pmatrix}\Pr(A,A,A\mid A,B,C,X)+2\Pr(A,A,B\mid A,B,C,X)+2\Pr(A,A,C\mid A,B,C,X)\\+2\Pr(A,B,B\mid A,B,C,X)+3\Pr(A,B,C\mid A,B,C,X)+2\Pr(A,C,C\mid A,B,C,X)\\+\Pr(B,B,B\mid A,B,C,X)+2\Pr(B,B,C\mid A,B,C,X)+2\Pr(B,C,C\mid A,B,C,X)\\+\Pr(C,C,C\mid A,B,C,X)\end{pmatrix}$
	6 (A,A,B,B,C,C)	*F* & *Q*	$90\Pr(A,A,B,B,C,C\mid X)$
	5 (A,A,B,B,C)	*F*	$180\Pr(A,A,B,B,C\mid X)$
		Q	$30\Pr(A,A,B,B,C\mid X)\left(6-4\Pr(A\mid A,A,B,B,C,X)-4\Pr(B\mid A,A,B,B,C,X)-3\Pr(C\mid A,A,B,B,C,X)\right)$
	4 (A,A,B,C)	*F*	$180\Pr(A,A,B,C\mid X)$

		Q	$30\Pr(A,A,B,C\mid X)\begin{pmatrix}6-8\Pr(A\mid A,A,B,C,X)-6\Pr(B\mid A,A,B,C,X)-6\Pr(C\mid A,A,B,C,X)\\+3\Pr(A,A\mid A,A,B,C,X)+4\Pr(A,B\mid A,A,B,C,X)+4\Pr(A,C\mid A,A,B,C,X)\\+2\Pr(B,B\mid A,A,B,C,X)+3\Pr(B,C\mid A,A,B,C,X)+2\Pr(C,C\mid A,A,B,C,X)\end{pmatrix}$
	3 (A,B,C)	*F*	$120\Pr(A,B,C\mid X)$
		Q	$30\Pr(A,B,C\mid X)\begin{pmatrix}4-6\Pr(A\mid A,B,C,X)-6\Pr(B\mid A,B,C,X)-6\Pr(C\mid A,B,C,X)+\\4\Pr(A,A\mid A,B,C,X)+4\Pr(B,B\mid A,B,C,X)+4\Pr(C,C\mid A,B,C,X)\\+6\Pr(A,B\mid A,B,C,X)+6\Pr(A,C\mid A,B,C,X)\\+6\Pr(B,C\mid A,B,C,X)-\Pr(A,A,A\mid A,B,C,X)-2\Pr(A,A,B\mid A,B,C,X)\\-2\Pr(A,A,C\mid A,B,C,X)-2\Pr(A,B,B\mid A,B,C,X)\\-2\Pr(A,C,C\mid A,B,C,X)-3\Pr(A,B,C\mid A,B,C,X)-2\Pr(B,B,C\mid A,B,C,X)\\-\Pr(B,B,B\mid A,B,C,X)-2\Pr(B,C,C\mid A,B,C,X)-\Pr(C,C,C\mid A,B,C,X)\end{pmatrix}$
2	ALL	*UC*	$\Pr(A,B\mid X)\begin{pmatrix}6\Pr(A,A,A,A\mid A,B,X)+15\Pr(A,A,A,B\mid A,B,X)+20\Pr(A,A,B,B\mid A,B,X)\\+15\Pr(A,B,B,B\mid A,B,X)+6\Pr(B,B,B,B\mid A,B,X)\end{pmatrix}$
	6 (A,A,A,B,B,B)	*F* & *Q*	$20\Pr(A,A,A,B,B,B\mid X)$
	5 (A,A,A,B,B)	*F*	$60\Pr(A,A,A,B,B\mid X)$
		Q	$5\Pr(A,A,A,B,B\mid X)\left(12-9\Pr(A\mid A,A,A,B,B,X)-8\Pr(B\mid A,A,A,B,B,X)\right)$
	4 (A,A,B,B)	*F*	$90\Pr(A,A,B,B\mid X)$
		Q	$5\Pr(A,A,B,B\mid X)\begin{pmatrix}18-24\Pr(A\mid A,A,B,B,X)-24\Pr(B\mid A,A,B,B,X)+9\Pr(A,A\mid A,A,B,B,X)\\+16\Pr(A,B\mid A,A,B,B,X)+9\Pr(B,B\mid A,A,B,B,X)\end{pmatrix}$
	3 (A,A,B)	*F*	$60\Pr(A,A,B\mid X)$

		Q	$\Pr(A,A,B\mid X)\begin{pmatrix}60-120\Pr(A\mid A,A,B,X)-90\Pr(B\mid A,A,B,X)+90\Pr(A,A\mid A,A,B,X)\\+120\Pr(A,B\mid A,A,B,X)+60\Pr(B,B\mid A,A,B,X)-24\Pr(A,A,A\mid A,A,B,X)\\-45\Pr(A,A,B\mid A,A,B,X)-40\Pr(A,B,B\mid A,A,B,X)-15\Pr(B,B,B\mid A,A,B,X)\end{pmatrix}$
	2 (A,B)	*F*	$30\Pr(A,B\mid X)$
		Q	$\Pr(A,B\mid X)\begin{pmatrix}30-60\Pr(A\mid A,B,X)-60\Pr(B\mid A,B,X)+90\Pr(A,B\mid A,B,X)+60\Pr(A,A\mid A,B,X)\\+60\Pr(B,B\mid A,B,X)-60\Pr(A,A,B\mid A,B,X)-60\Pr(A,B,B\mid A,B,X)\\-30\Pr(A,A,A\mid A,B,X)-30\Pr(B,B,B\mid A,B,X)+6\Pr(A,A,A,A\mid A,B,X)\\+6\Pr(B,B,B,B\mid A,B,X)+15\Pr(A,A,A,B\mid A,B,X)+15\Pr(A,B,B,B\mid A,B,X)\\+20\Pr(A,A,B,B\mid A,B,X)\end{pmatrix}$
1	ALL	*UC*	$\Pr(A,A,A,A,A,A\mid X)$
	6	*F* & *Q*	$\Pr(A,A,A,A,A,A\mid X)$
	5	*F*	$6\Pr(A,A,A,A,A\mid X)$
		Q	$\Pr(A,A,A,A,A\mid X)\left(6-5\Pr(A\mid A,A,A,A,A,X)\right)$
	4	*F*	$15\Pr(A,A,A,A\mid X)$
		Q	$\Pr(A,A,A,A\mid X)\left(15-24\Pr(A\mid A,A,A,A,X)+10\Pr(A,A\mid A,A,A,A,X)\right)$
	3	*F*	$20\Pr(A,A,A\mid X)$
		Q	$\Pr(A,A,A\mid X)\left(20-45\Pr(A\mid A,A,A,X)+36\Pr(A,A\mid A,A,A,X)-10\Pr(A,A,A\mid A,A,A,X)\right)$
	2	*F*	$15\Pr(A,A\mid X)$
		Q	$\Pr(A,A\mid X)\begin{pmatrix}15-40\Pr(A\mid A,A,X)+45\Pr(A,A\mid A,A,X)-24\Pr(A,A,A\mid A,A,X)\\+5\Pr(A,A,A,A\mid A,A,X)\end{pmatrix}$
	1	*F*	$6\Pr(A\mid X)$

		Q	$\Pr(A \mid X)\begin{pmatrix} 6-15\Pr(A \mid A,X)+20\Pr(A,A \mid A,X)-15\Pr(A,A,A \mid A,X)+6\Pr(A,A,A,A \mid A,X) \\ -\Pr(A,A,A,A,A \mid A,X) \end{pmatrix}$

Figure 4.6 Epg showing TPOX and D18S51 loci of a three-person Identifiler® mixed DNA profile

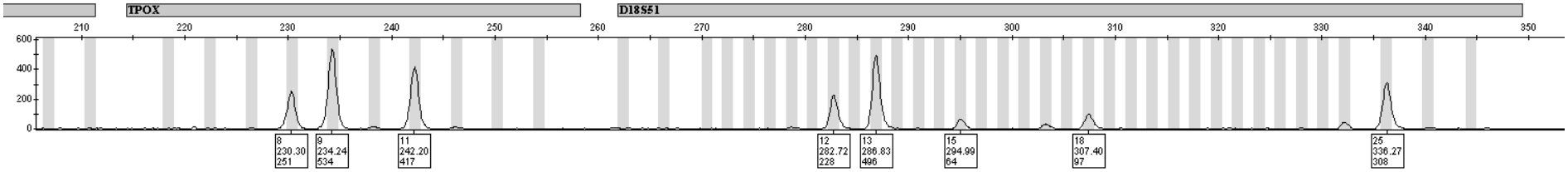


Table 4.8 Peaks and heights from DNA profile, Figure 4.6, POI profile and *LR* for the three different models

Marker	Peak	Height (RFU)	Allelic vector for F and Q models	POI profile	LR_{UC}	LR_{Q}	LR_{F}
TPOX	8 9 11	251 534 417	$(8,9,11)$	9,11	2.38	2.22	1.00
D18S51	12 13 15 18 25	228 496 64 97 308	$(12,13,15,18,25)$	13,25	16.50	15.84	14.24

References

1. D.E. Krane, S. Ford, J.R. Gilder, K. Inman, A. Jamieson, R. Koppl, et al., Sequential Unmasking: A Means of Minimizing Observer Effects in Forensic DNA Interpretation, Journal of Forensic Sciences, 53(4) (2008) 1006-1007.

2. J.S. Buckleton, C.M. Triggs, S.J. Walsh. DNA Evidence. Boca Raton, Florida: CRC Press, 2004.

3. P. Gill, J.P. Whitaker, C. Flaxman, N. Brown, J.S. Buckleton, An investigation of the rigor of interpretation rules for STR's derived from less that 100 pg of DNA, Forensic Science International, 112(1) (2000) 17-40.

4. M.W. Perlin, M.M. Legler, C.E. Spencer, J.L. Smith, W.P. Allan, J.L. Belrose, et al. Validating TrueAllele® DNA mixture interpretation. 2011 [updated 2011; cited 2011 14 January 2011]; Available from: http://www.cybgen.com/documents/publications/PLSSABD2010.pdf.

5. J. Buckleton, C.M. Triggs, Is the 2p rule always conservative?, Forensic Science International, 159 (2006) 206-209.

6. M. Bill, P. Gill, J. Curran, T. Clayton, R. Pinchin, M. Healy, et al., PENDULUM - A guideline based approach to the interpretation of STR mixtures, For Sci Int, 148 (2005) 181-189.

7. P. Gill, R.L. Sparkes, R. Pinchin, T. Clayton, J.P. Whitaker, J.S. Buckleton, Interpreting simple STR mixtures using allelic peak areas, Forensic Science International, 91 (1998) 41 - 53.

8. T. Clayton, J.P. Whitaker, R.L. Sparkes, P. Gill, Analysis and interpretation of mixed forensic stains using DNA STR profiling, Forensic Science International, 91 (1998) 55 - 70.

9. M.W. Perlin, B. Szabady, Linear mixture analysis: A mathematical approach to resolving mixed DNA samples., Journal of Forensic Sciences 46(6) (2001) 1372-1377.

10. M.W. Perlin, G. Lancia, S.-K. Ng, Toward fully automated genotyping: genotyping microsatellite markers by deconvolution, American Journal of Human Genetics, 57 (1995) 1199-1210.

11. P. Gill, C.H. Brenner, J.S. Buckleton, A. Carracedo, M. Krawczak, W.R. Mayr, et al., DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. , Forensic Science International, 160 (2006) 90-101.

12. B.S. Weir, C.M. Triggs, L. Starling, L.I. Stowell, K.A.J. Walsh, J.S. Buckleton, Interpreting DNA Mixtures, Journal of Forensic Sciences, 42(2) (1997) 213-222.

13. J.M. Curran, C.M. Triggs, J.S. Buckleton, B.S. Weir, Interpreting DNA mixtures in structured populations, Journal of Forensic Sciences, 44(5) (1999) 987 - 995.

14. D.J. Balding, R.A. Nichols, DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands, Forensic Science International, 64 (1994) 125-140.

15. J.-A. Bright, J.S. Buckleton, C.E. McGovern, Allele frequencies for the four major sub-populations of New Zealand for the 15 Identifiler® loci, Forensic Science International: Genetics, 4(2) e65-e66.

16. P. Gill, A. Kirkham, J. Curran, LoComatioN: A software tool for the analysis of low copy number DNA profiles, Forensic Science International, 166(2-3) (2007) 128-138.

17. J. Curran, Personal Communication (2012)

5: THE INFLUENCE OF THE PROBABILITY OF DROPOUT ON THE WEIGHT OF THE EVIDENCE

The challenge of how to best account for the possibility that alleles have dropped out is central to the implementation of extended interpretation methods to cope with LtDNA and complex mixtures.

The DNA commission of the International Society of Forensic Genetics (ISFG) stressed the importance of considering allelic dropout in the recommendation on mixture interpretation [1]. However, the question of how to assess the probability of dropout (Pr(D)) was not formalised. The estimation of Pr(D) is important because it influences the estimation of the weight of evidence in the calculation of the likelihood ratio (*LR*).

5.1 A case example: Garside and Bates

A 2006 case (Bates, R. v [2006] EWCA Crim 1395) [2] highlights the importance of the development of a model that can incorporate the probability of dropout. On the 2nd of October 2001 Marilyn Garside was stabbed and killed as she answered the door of her elderly mother's house in Rose Lane, Romford, UK. It was the prosecution's case that the victim's husband, James Garside, had hired Richard Bates to murder her. Samples were taken from seven locations in the vicinity of the homicide and analysed for the presence of DNA. The results from one of the crime scene samples are presented in the form of a table; Table 5.1 includes the profile of the deceased and the person of interest (Bates) [3].

The evidential item labelled, "SJP/22 Area 4 Chrome Handle", produced an electropherogram (epg) that indicated a clear major contributor that corresponded with the deceased and a partial minor low level contributor that had eight alleles that differed from the deceased. These eight alleles are present in the POI's profile however the POI has alleles at D2S1338 and D18S51 that are not present in the crime scene profile. There is the possibility that because the minor contributor is only present at a low level, that these alleles are not represented in the epg because they have dropped out.

Table 5.1: The partial minor component from the crime scene sample "SJP/22 Area 4 Chrome Handle" with reference profiles from the deceased and POI.

Locus	Deceased	POI	Minor component
D3S1358	16,16	13,16	13
vWA	15,17	16,16	16
D16S539	11,12	11,12	—
D2S1338	20,20	**19**,22	22
D8S1179	12,13	8,13	8
D21S11	30,32.2	30,31.2	31.2
D18S51	14,14	**12,15**	—
D19S433	12,14	12,15	15
TH01	9.3,9.3	7,7	7
FGA	23,25	21,21	21

When calculating the probability match for the POI the expert called on by the prosecution assumed dropout of the "missing" alleles to have occurred. At D18S51, where both alleles were assumed to have dropped out, the analyst employed the widespread practice of assigning a value of one to the locus. This is largely believed to be treating the locus as neutral (as a likelihood ratio of one supports neither the prosecution nor defence. This is discussed further in Chapter 2).

Only one allele corresponding with the POI was detected at the locus D2S1338. The prosecution's expert used the "2*p* rule" at this locus. Based on these assumptions the resulting probability match reported by the prosecution was 1 in 610,000.

The defence employed a second expert to interpret the profile who stated that it is not correct to assign a value of one to loci that are "missing" data due to the potential exclusionary potential of the unreported alleles and that using the *"2p rule"* may not be conservative. However, at the time it was not possible to calculate the effect of the "missing" loci on the profile so a "true" match probability could not be obtained, although the expert claimed it was likely to be lower than 1 in 610,000.

When presented with the differing opinions the judge in the case stated:

"*What are the consequences of the impossibility of assigning a statistical weight to the voids? The alternatives are to exclude the evidence entirely or to admit it subject to an appropriate warning to the jury of the limitations of the evidence, and particularly highlighting the fact that although what was found was consistent with Bates' DNA profile, the voids at D2S1338 and D18S51 in particular may have contained an allele or alleles, the presence of which would have been wholly exculpatory. In arriving at the correct conclusion it is important to remember that scientific evidence frequently only provides a partial answer to a case, or to an issue in a case. However, the test of admissibility is not whether the answer is complete, but whether science can properly and fairly contribute to the matter in question. . . . . ".* [2]

And the Court ruled:

"*We can see no reason why partial profile DNA evidence should not be admissible provided that the jury are made aware of its inherent limitations and are given a sufficient explanation to enable them to evaluate it*." [2]

In allowing the admission of partial profiles as evidence (therefore assuming dropout to have occurred) the Court has made it imperative that these profiles are correctly represented in their evidential value. Therefore it is important that a statistic is calculated that is more likely to represent a value closer to the true likelihood of the profile. This calculation must take into account the probability that "missing" alleles have dropped out rather than just assuming dropout to have occurred.

5.1.1 Reanalysing the evidence using a dropout proxy

If we reanalyse the two questionable loci; D18S51 and D2S1338, using a proxy for the probability of dropout, the Buckleton and Gill model (Chapter 2) and the *Q* model (Chapter 4) (peak height information was not readily available from the SJP/22-4 profile), then we can determine the effect that the "missing" alleles may have on the overall statistic.

At D2S1338 the alleles 20, 20 and 22 are called. The deceased has the genotype [20,20]. The genotype of the minor contributor might be [22,22], [22,20] or [22,Q] where *Q* represents all of the other alleles possible at D2S1883 not already called at the locus .

Recall the POI is 19,22. The evidence given the hypothesis of the prosecution $\Pr\left(E \mid H_p\right)$ becomes:

$$\Pr\left(E \mid H_p\right) = D\bar{D}$$

where D stands for the probability of dropout and $\bar{D}$ stands for the probability of an allele not dropping out (i.e. the allele has been seen).

The evidence given the defence's hypothesis $\Pr\left(E \mid H_d\right)$ must include all of the possible genotypes that explain the 22 allele (the allele not attributed to the deceased) as well as a correction from Hardy-Weinberg equilibrium. The alleles from the known contributor that may mask the presence of the minor contributor's alleles (in this case the 20 allele) are not assigned a value for $\bar{D}$ as they have already been seen in the profile ($\bar{D} = 1$):

$$\Pr\left(E \mid H_d\right) = 2\bar{D}P_{20,22|19,22} + \bar{D}_2 P_{22,22|19,22} + 2\bar{D}DP_{Q,22|19,22}$$

The overall likelihood ratio (*LR*), with the Balding and Nichols correction [4] applied in the end stages (with the approximation $D_2 \approx D^2$), becomes:

$$LR_D \approx \frac{D\bar{D}}{2\bar{D}P_{20,22|19,22} + \bar{D}_2 P_{22,22|19,22} + 2\bar{D}DP_{Q,22|19,22}}$$

$$LR_D \approx \frac{D\bar{D}}{2\bar{D}P_{20,22|19,22} + \bar{D}(1+D)P_{22,22|19,22} + 2\bar{D}DP_{Q,22|19,22}}$$

$$LR_D \approx \frac{D}{2P_{20,22|19,22} + (1+D)P_{22,22|19,22} + 2DP_{Q,22|19,22}}$$

$$LR_D \approx \frac{D}{2P_{20,22|19,22} + (1+D)P_{22,22|19,22} + 2D(1-P_{22|19,22} - P_{20|19,22})P_{22|19,22}}$$

$$LR_D \approx \frac{D}{P_{22|19,22}\left[2P_{20|19,22,22} + (1+D)P_{22|19,22,22} + 2D(1-P_{22|19,22} - P_{20|19,22})\right]}$$

$$LR_D \approx \frac{D}{\left(\theta + (1-\theta)P_{22}\right)\left[2\left(1-\theta\right)P_{20} + (1+D)\left(2\theta + \left(1-\theta\right)P_{22}\right) + 2D\left(1-\theta\left(1-P_{22} - P_{20}\right)\right)\right]}$$

where P_i is the allele frequency of the ith allele within the defined population and θ is the coancestry coefficient, or F_{ST}. In this work we use allele frequencies from the Australian Caucasian subpopulation data [5] and a value of θ=0.02.

In contrast, when the *2p* rule is used, the *LR* is:

$$LR_D \approx \frac{1}{2(\theta + (1-\theta))P_{22}}$$

If we try a range of values for D and compare the resulting *LR*s to the 2*p* rule, then we can examine the effect of D. We can see from Figure 5.1 that when the probability of dropout is very low the 2*p* rule is not conservative – that is, it is overestimating the value of the evidence at that locus. As the probability of dropout nears one then the two *LR*s start to converge. When we plot the ratio of the two *LR*s (Figure 5.2) we can see that the 2*p* rule overstates the *LR* fairly dramatically at low level, but then flattens out as the probability of dropout increases.

Using a proxy of $D = 0.5$ (a rough estimation of what D could be given the epg and the hypothesis that the POI is the minor) the *LR* using the dropout model is 7.7 and the *LR* using the 2*p* rule is 9.2. Consequently both methods indicate that the evidence given the H_p is more likely than the evidence given the H_d however, the 2*p* rule favours the H_p 1.2 times more than the model utilising a value for dropout.

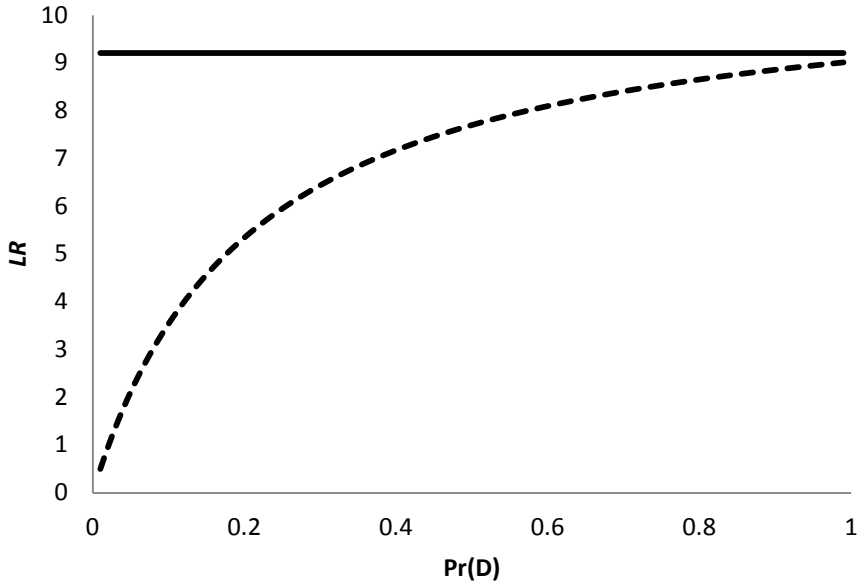


Figure 5.1: A comparison of the *LR*s calculated for D2S1338 using a dropout proxy. The solid line is the *LR* using the *2p* rule and dashed line is the *LR* using the Buckleton and Gill model.

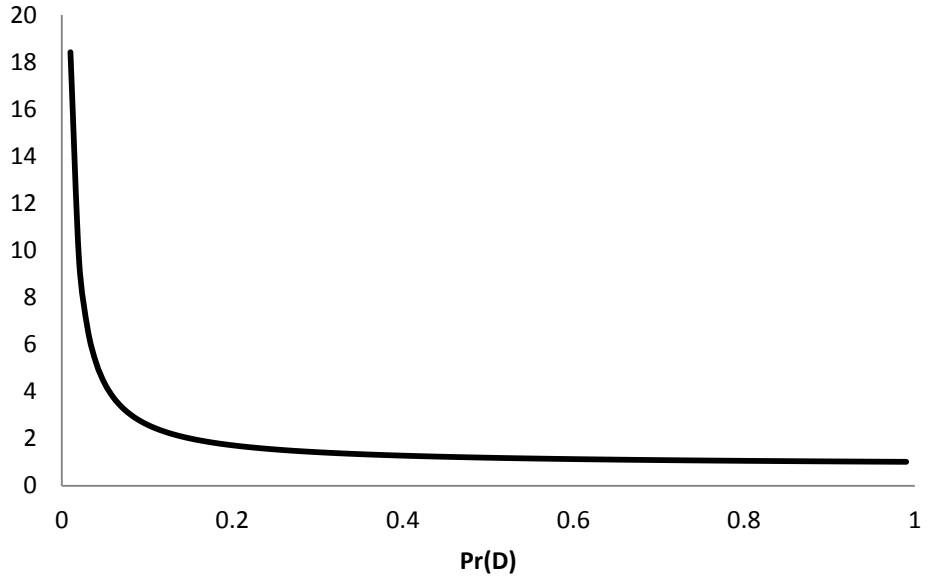


Figure 5.2: The ratio of the *2p* rule *LR* to the Buckleton and Gill *LR*

The analyst called the alleles [14,14] at the locus D18S51. The deceased has the genotype [14,14] and the POI has the genotype [12,15] at this locus. If the POI is the true minor contributor, both alleles must have dropped out: $\Pr\left(E \mid H_p\right) = D^2$.

The probability of the evidence given the defence's hypothesis is: $\Pr\left(E \mid H_d\right) = P_{14,14|12,15} + 2DP_{14,Q|12,15} + D^2 P_{Q,Q|12,15}$

And the resulting *LR* becomes:

$$LR \approx \frac{D^2}{P_{14,14|12,15} + 2DP_{14,Q|12,15} + D^2 P_{Q,Q|12,15}}$$

$$LR \approx \frac{D^2}{\left[(1-\theta)\left(\theta + (1-\theta)\right)P_{14}^2 + (1-\theta)DP_Q\left\{2P_{14|} + DP_Q\right\}\right]}$$

We can compare the *LR*s for a range of values for D to the assigned *LR* of "1" used in the case. Looking at Figure 5.3 we can see that assigning a value of one to the locus favours the

prosecution. The true value of the locus could favour the defence if the probability of dropout is low. If we take the proxy we used for dropout in the earlier example (0.5), then the Buckleton and Gill model returns a *LR* of 0.08 compared to the assigned *LR* of 1. This means that in this example, assigning a 1 at this locus overstates the *LR* by a factor of 12.5, a non-insignificant difference in favour of the prosecution.

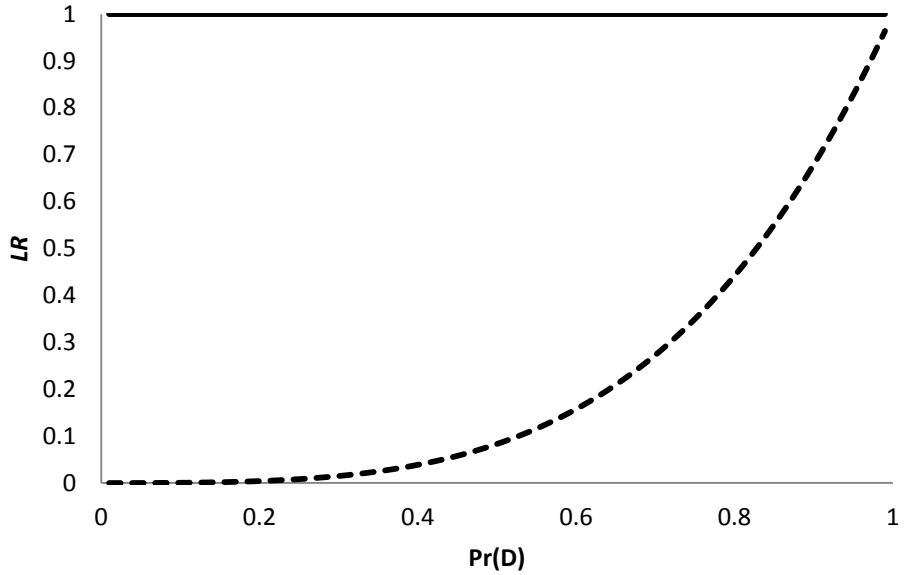


Figure 5.3: A comparison of the *LR* calculated for D18S51 using a dropout proxy and the Buckleton and Gill model (dashed line) and the *LR* of 1 (solid line) used by the prosecution.

5.2 Discussion

The differences will be compounded in cases where multiple loci are assessed in this manner. As the results are multiplied across loci the differences will be much larger in magnitude. Although assigning a one to a "missing" locus appears on the outset to be neutral, in reality, missing information should down weight the overall statistic. The value of this weight must depend on the quality of the profile being analysed. If analysts are assuming alleles are "missing", then this is already an assumption in favour of the prosecution. By assuming there are alleles missing we are including the POI at the locus in question when in reality perhaps the POI's genotype at this locus is not missing and it's actually an exclusion.

The analysis of these two loci has demonstrated that neither method employed by the prosecution in this case was conservative, and that the methods break down as the probability of dropout decreases. Loci should be assessed using a probability of dropout in cases where

dropout is possible, rather than ignoring the possibility it has occurred. This will ensure that the *LR* presented to the court is closer to the "true" answer.

At the time of the Garside and Bates trial, and until recently, there was no satisfactory model for dropout. However, now several models of the probability of dropout have been published [6-12].

Tvedebrink et al., introduced the concept of modelling the probability of dropout using logistic regression [6-8]. Logistic regression is a standard way to estimate the probabilities for a random variable with two possible outcomes (i.e. dropout verses no dropout), when it is thought that the probability changes with respect to one or more explanatory variables. The logistic model in Tvedebrink's original paper is straightforward as there is only one explanatory variable: $\hat{H}$, which is the average allele peak height across the DNA profile with an individual locus effect included in the intercept [7].

In casework there is often an observable decrease in allelic peak height as the molecular weight of the alleles increases. This has been termed the "ski slope" or degradation slope [13, 14]. Degradation is important to consider when modelling dropout, as higher weight alleles have been observed to dropout out more frequently than low weight alleles (Figure 5.4). If an average peak height is used across a profile, then it is likely that the probability of dropout will be overestimated for low molecular weight alleles and underestimated for high molecular weight alleles.

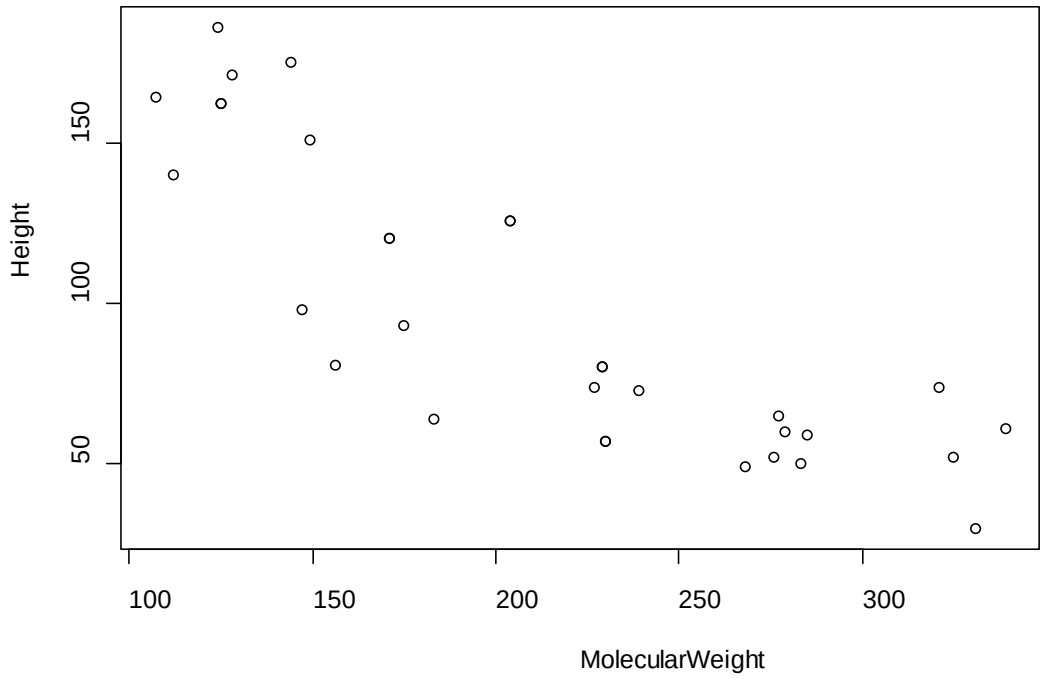


Figure 5.4: A plot showing the degradation slope present in a casework profile when the peak heights of the alleles are plotted by molecular weight

In the context of a DNA profile, the best explanatory variable to predict the probability of dropout would be the true, but unknown, template available at each locus for amplification, rather than a proxy across the profile. Tvedebrink et al., built on their original model in a recent publication [8] by adjusting the proxy for DNA quantity to correct for degradation. In this updated work, the proxy for the available DNA template is an exponential function of molecular weight [15] . However, the group also includes an individual locus effect. The inclusion of a locus effect may be justified as experience in casework has suggested that loci may dropout with differing probabilities.

However, we were concerned with the transportability of the 2012 Tvedebrink model [8]. If the probability of dropout was calculated including a locus effect using a training set of data, then how transportable would those locus coefficients be on different profiles? Theoretically the locus effects could include batch-to-batch variability of PCR reagents, the vulnerability of each primer to inhibitors known to be prevalent in casework type samples, or environmental

conditions and/or contaminants. Factors influencing the locus coefficients in the training data may differ from the casework profile that is being interpreted.

The following chapter outlines research into the transportability of the original Tvedebrink model [7], the 2012 Tvedebrink model [8] and a third model. The third model is a modified version of Tvedebrink's 2012 model with the individual locus coefficients removed. The following work also compares the three models on their efficacy using the data they were trained on.

References

1. Gill, P., et al., DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. Forensic Science International, 2006. 160(2-3): p. 90-101.

2. http://judgmental.org.uk/judgments/EWCA-06 EWCA Crim 1395.html.

3. Balding, D. J. at al., Interpreting low template DNA profiles. Forensic Science International: Genetics, 2009. 4(1): p. 1-10.

4. Balding, D.J. and R.A. Nichols, DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Science International, 1994. 64(2–3): p. 125-140.

5. Bright, J., Personal Communication, (2012).

6. Tvedebrink, T., et al., Allelic drop-out probabilities estimated by logistic regression—Further considerations and practical implementation. Forensic Science International: Genetics, 2012. 6(2): p. 263-267.

7. Tvedebrink, T., et al., Estimating the probability of allelic drop-out of STR alleles in forensic genetics. Forensic Science International, 2009. 3: p. 222-226.

8. Tvedebrink, T., et al., Statistical model for degraded DNA samples and adjusted probabilities for allelic drop-out. Forensic Science International: Genetics, 2012. 6(1): p. 97-101.

9. Cowell, R.G., S.L. Lauritzen, and J. Mortera, Probabilistic expert systems for handling artifacts in complex DNA mixtures. Forensic science international. Genetics, . 2011. 5(3): p. 202-209.

10. Curran, J.M. and J. Buckleton, Inclusion Probabilities and Dropout. Journal of Forensic Sciences, 2010. 55(5): p. 1171-1173.

11. Miller, C.R., P. Joyce, and L.P. Waits, Assessing Allelic Dropout and Genotype Reliability Using Maximum Likelihood. Genetics Society of America, 2002(160): p. 357-366.

12. Puch-Solis, R., et al., Assigning weight of DNA evidence using a continuous model that takes into account stutter and dropout. Science & Justice, 2010. 50(1): p. 46-47.

13. Nicklas, J.A., T. Noreault-Conti, and E. Buel, Development of a Real-Time Method to Detect DNA Degradation in Forensic Samples. Journal of Forensic Sciences, 2012. 57(2): p. 466-471.

14. Chung, D.T., et al., A study on the effects of degradation and template concentration on the amplification efficiency of the STR Miniplex primer sets. Journal of Forensic Sciences, 2004. 49(4): p. 733-740.

15. Bright, J.A., et al., Degradation in forensic DNA profiles. Australian Journal of Forensic Sciences, 2013. IN PRESS http://dx.doi.org/10.1080/00450618.2013.772235.

16. Tvedebrink, T., et al., Estimating the probability of allelic drop-out of STR alleles in forensic genetics. Forensic Science International: Genetics, 2009. 3(4): p. 222-226.1.

CHAPTER 6: MODELLING THE PROBABILITY OF DROPOUT

A version of this work has been submitted for publication to Forensic Science International: Genetics under the title:

Utilising allelic dropout probabilities estimated by logistic regression in casework.

The authors on the paper are:

John Buckleton [1], Hannah Kelly [1, 2], Jo-Anne Bright [1], Duncan Taylor [3], Torben Tvedebrink [4] and James Curran [2]

1) ESR, PB 92021, Auckland, 1142, New Zealand,
2) Department of Statistics, University of Auckland, PB 92019, Auckland, 1142, New Zealand
3) Forensic Science SA, 21 Divett Place, Adelaide, SA 5000, Australia
4) Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg East, Denmark

This paper has been rewritten as a chapter to appear in this thesis. There are sections that overlap with the submitted paper such as the results tables and methods section in particular.

This research expands on the paper that has been submitted predominantly due to the difference in the intended target audience. Sections in this chapter have been broadened in an effort to make the results of this research more accessible for case work forensic biologists.

This research investigates dropout in forensic DNA samples. Dropout has been highly topical within the literature and is a phenomenon associated with low template DNA. Models have been proposed that calculate the probability of dropout using logistic regression. Two of these models are compared with a third model that the authors have built. The comparison tests the

three models on their efficacy using the data they are trained on then tests their ability to be used on data outside of their training set.

The aim of this work is to determine if a dropout model is trained using a specific set of data, can it be applied to data outside this set? For example, are there variables that are specific to the training profiles that will not translate to other profiles? This work shows that although it has been shown that locus affects the probability of dropout, this is likely to be profile specific. These locus effects are likely to be related to the amplification conditions, the internal multiplex variability, and other events that change between profiles.

This work was supported in part by grant 2011-DN-BX-K541 from the US National Institute of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice. The authors would like to thank Johanna Veth and Sue Vintiner for their helpful contributions towards this paper.

6: MODELLING THE PROBABILITY OF DROPOUT

In forensic DNA analysis, if a sample is low level or degraded some alleles may fail to be visualized in the resulting electropherogram (epg). This can be because;

1) they were not present in the aliquot amplified,

2) they have not amplified at all, or

3) they have amplified but the peak is below some threshold set for assignment of peaks.

This phenomenon is termed dropout. The DNA commission of the International Society of Forensic Genetics (ISFG) stressed the importance of considering allelic dropout in the recommendation on mixture interpretation [1]. However, the question of how to assess the probability of dropout (Pr(D)) was not formalised.

A model for the probability of dropout has been published by Tvedebrink et. al., in 2009 [2]. It models the probability of dropout using logistic regression and has an explanatory variable that uses an average of peak heights across a profile, $\left(\hat{H}\right)$. The model also includes a locus effect. This model allows a different intercept per locus $\left(\beta_{0,\ell}\right)$ with a constant slope defined by average peak height $\left(\beta_1 \times \hat{H}\right)$ and introduces the concept that the probability of dropout differs at each locus.

Our initial exploratory work confirmed the locus effect identified by Tvedebrink et al. [2] (Figure 6.1). This dataset was comprised of approximately 100 casework profiles that were identified as exhibiting possible dropout. Case files were examined and data collated for profiles suggesting a single contributor where the circumstances allowed a reasonable inference about the source.

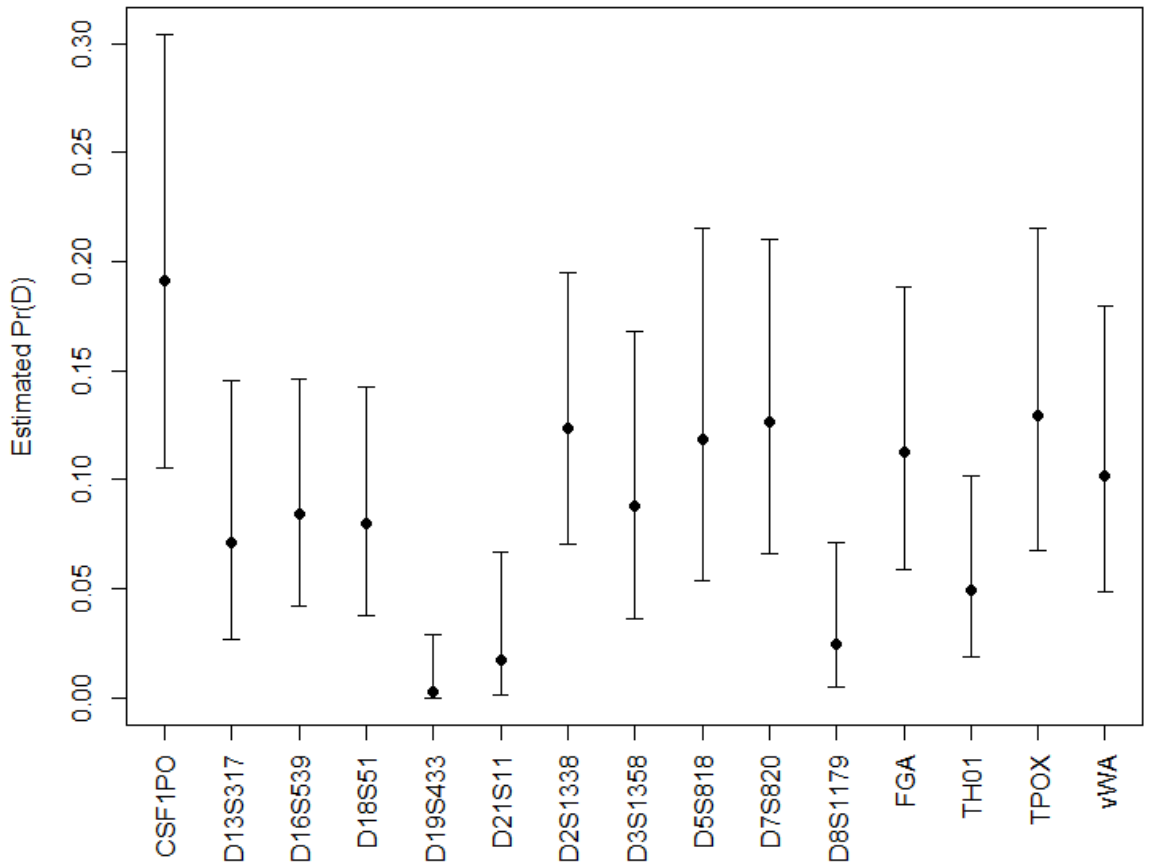


Figure 6.1: Estimated Pr(D) per locus, averaged across ~100 LtDNA casework profiles. The intervals were generated using the posterior distribution of dropout rate (assuming a Beta (0.5, 0.5) prior).

A typical casework electropherogram (epg) has an observable decrease in allelic peak height as the molecular weight (mwt) of the alleles increases. This has been termed the “ski slope” or degradation slope [3, 4]. Degradation is important to consider when modelling dropout, as higher weight alleles have been observed to dropout out more frequently than lower weight alleles. This observation implies that the amount of available DNA template pre-amplification may not be constant across loci. Therefore it is not unexpected that loci dropout with differing probabilities.

Using the explanatory variable $\left(\hat{H}\right)$, the same amount of available DNA template pre-amplification, is assumed per locus. Locus specific effects are contained within the intercept coefficients $\left(\beta_{0,\ell}\right)$ (including the observed degradation slope).

A better method to model the probability of dropout is to account for the degradation slope in the explanatory variable. This provides an estimate for the true, but unknown, template available at each locus.

6.1 Modelling degradation

The simplest model for degradation is linear [5]. Under this model, the expected peak height declines constantly with respect to molecular weight. The linear model has empirical support in that one can take a copy of an epg and draw a downward sloping straight line across the apex of heterozygous peaks from the lowest molecular weight locus to the highest molecular weight locus [6].

However, the suggestion of an exponential relationship for degradation has been made by Tvedebrink et al., [7]. This relationship can be justified theoretically as, if the degradation of the DNA strand was random with respect to location, then we would anticipate that the expected height of peak *a*, E_a would be exponentially related to molecular weight and to whether the peak was heterozygous or homozygous (Figure 6.2). Let X_a be the count of allele *a*. $X_a = 1$ for a heterozygous locus and $X_a = 2$ for a homozygous locus. The expected height, E_a, of peak *a* is therefore modelled as:

$$\mathrm{E}_a = \hat{H}_a X_a \qquad \hat{H}_a = \alpha_0 \, \mathrm{e}^{\alpha_1 \mathrm{w}_a}$$

where w_a is the molecular weight of allele *a* and where $\hat{H}_a$ subsumes the concepts of template number and degradation. Thus, $\hat{H}_a$ is a proxy for the available DNA template at allelic position *a*.

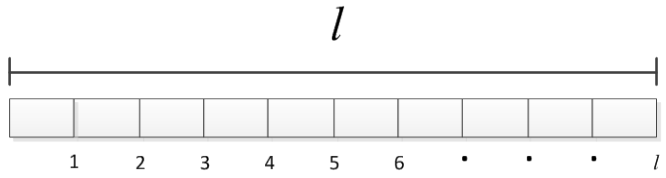


Figure 6.2: The expected height (E_a) of allele *a*, is exponentially related to the molecular weight of *a* (w_a) if the breakdown of DNA is random. If the probability of a break is *p*, at any of the locations 1…*l* then the chance of the full fragment being amplified is $(1-p)^l$.

6.2 Modelling dropout

The updated Tvedebrink et. al., [7] probability of dropout model, includes the proxy for available DNA template per locus $\hat{H}_a$ using the exponential function of molecular weight. However, this updated model also retains the locus effect in the intercept. Thus the parameters of the model are $\beta_{0,l}$ and $\beta_1 \times \hat{H}_a$.

It is interesting that Tvedebrink et. al., [7] include a locus effect despite accounting for the differing amount of template available at each locus. If template number is removed from locus effect, then it is reasonable to assume that any remaining locus effect must be amplification specific. That is, batch to batch variability of multimix, the vulnerability of each primer to inhibitors, or environmental conditions or contaminants, however, experience establishes that a locus effect is observed.

Consequently, the question arises; do locus effects developed for one set of "training" data translate to a future set of data? If multiplex master mix batches, or even samples, differ in locus amplification efficiency, then the transportability of the model to future profiles may be an issue. Accordingly it may be advantageous to consider a model that incorporates the concept of degradation but does not include a locus effect.

This work investigates the transportability of the 2009 Tvedebrink model (T_1) [2], the 2012 Tvedebrink model (T_2) [7] and a new model (T_3). T_3 is a revised version of Tvedebrink's 2012 model with the individual locus coefficients removed. The following work also compares the three models on their efficacy using the data they were trained on.

6.3 Method

T_1 uses a function of $\ln \hat{H}$ as the explanatory variable and T_2 and T_3 use $\ln \hat{H}_a$. Using this approach we model the probabilities of dropout of a single allele D_a and of a homozygous locus, D_{2a}, as:

$$D_a = \frac{e^{(\beta_{0,\ell})+(\beta_1 \times \ln \hat{H})}}{1+e^{(\beta_{0,\ell})+(\beta_1 \times \ln \hat{H})}} \quad D_{2a} = \frac{e^{(\beta_{0,\ell})+(\beta_1 \times \ln 2\hat{H})}}{1+e^{(\beta_{0,\ell})+(\beta_1 \times \ln 2\hat{H})}} \quad \text{(for locus ℓ) .....} T_1 \text{ model}$$

$$D_a = \frac{e^{(\beta_{0,\ell})+(\beta_1 \times \ln \hat{H}_a)}}{1+e^{(\beta_{0,\ell})+(\beta_1 \times \ln \hat{H}_a)}} \quad D_{2a} = \frac{e^{(\beta_{0,\ell})+(\beta_1 \times \ln 2\hat{H}_a)}}{1+e^{(\beta_{0,\ell})+(\beta_1 \times \ln 2\hat{H}_a)}} \quad \text{(for locus ℓ) .....} T_2 \text{ model}$$

$$D_a = \frac{e^{\beta_0+\beta_1 \ln \hat{H}_a}}{1+e^{\beta_0+\beta_1 \ln \hat{H}_a}} \quad D_{2a} = \frac{e^{\beta_0+\beta_1 \ln 2\hat{H}_a}}{1+e^{\beta_0+\beta_1 \ln 2\hat{H}_a}} \quad \text{.....} T_3 \text{ model}$$

β_0 and β_1 are developed from empirical data by logistic regression. In T_1 and T_2 the values, each locus has a different β_0. In T_3 one β_0 and one β_1 are applied to all loci.

The three models described, (T_1, T_2 and T_3) were applied to the datasets outlined below. $\hat{H}$ and $\hat{H}_a$ values were obtained using least squares fitting. For the methods T_1 and T_2 the β_0 values were constrained to within a factor of two of the average to avoid them moving to unreasonable values. For the various datasets, one was chosen to train the models, which was then tested against the others. The threshold for dropout was set at 50 RFU. Peaks below this were deemed to have dropped out (1) and peaks above this were deemed present, or not dropped out (0).

6.3.1 The datasets

Case files were examined and data collated for profiles suggesting a single contributor where the circumstances allowed a reasonable inference about the source. The case file dates varied from November 2009 to May 2012. The DNA samples had been extracted using DNA IQ™ (Promega Corporation) method for saliva, bloodstains and trace samples. All samples were

quantified using Applied Biosystems Quantifiler™ human DNA detection system (Life Technologies, Carlsbad CA) and 1.5 ng of DNA was targeted for Applied Biosystems Identifiler® (Life Technologies, Carlsbad CA) amplification on a 9700 thermal cycler (Applied Biosystems) with a silver block. Amplified DNA was analysed using a 3130xl capillary electrophoresis instrument and DNA profile data was analysed using GeneMapper™ ID software (Applied Biosystems).

Datasets I_1 and I_2 were made up of 92 and 74 casework samples respectively, and was one data set split in approximate time order. Dataset I_3 was made up of 47 casework samples that were purposely collected to emphasise low template DNA.

Single source profiles of known origin from blood and semen stains were collated. The DNA samples had been extracted using DNA IQ™ (Promega Corporation) and quantified using Applied Biosystems Quantifiler™ human DNA detection system (Life Technologies, Carlsbad CA) and 1.5 ng of DNA was targeted for Applied Biosystems Identifiler® (Life Technologies, Carlsbad CA) amplification on a 9700 thermal cycler (Applied Biosystems) with a silver block. Amplified DNA was analysed using a 3130xl capillary electrophoresis instrument and DNA profile data was analysed using GeneMapper™ ID software (Applied Biosystems). This data was classified as pristine and used to create dataset I_4.

Dataset I_4 was made up of 118 pristine samples.

Buccal swabs collected from 10 volunteers were extracted using DNA IQ™ (Promega) as per the manufacturer's directions. Extracted DNA was quantified twice using Applied Biosystems Quantifiler™ human DNA detection system (Life Technologies, Carlsbad CA) and an average taken. Varying quantities of DNA (1 ng, 500 pg, 250 pg, 100 pg, 75 pg, 50 pg, 10 pg, 5 pg and 1 pg) were amplified using Promega's PowerPlex® 21 System in 12.5µL reactions on a 9700 thermal cycler (Applied Biosystems) with a silver block. Amplified DNA was analysed using a 3130xl capillary electrophoresis instrument and DNA profile data was analysed using GeneMapper™ ID software (Applied Biosystems).

Dataset P was created from the PowerPlex® data. Dataset P was made up of 70 samples and was one dataset run on one plate and split multiple times into equal halves using random numbers.

6.4 Results

The log likelihoods of the data for different combinations of model and training sets $I_{1...}$ I_4 are given in Table 6.1. The log likelihood values were obtained by assigning;

$$Y_i = \begin{cases} 1 \\ 0 \end{cases}$$

where 1 refers to dropout, and

0 refers to no dropout.

The distribution of the data is approximated as:

$$Y_i \approx \text{Bernoulli}(\theta_i).$$

The $\log(\text{odds})$ for logistic regression $=\beta_0 + \beta_1 x$, which here, is equal to the density:

$$\text{logit}(\theta_i) = f\left(\hat{H}, \ell\right)$$

Where :

$$f\left(\hat{H}, \ell\right) = \begin{cases} \beta_{0,\ell} + \beta_1 \times \hat{H} & \text{for } T_1 \\ \beta_{0,\ell} + \beta_1 \times \hat{H}_a & \text{for } T_2 \\ \beta_0 + \beta_1 \times \hat{H}_a & \text{for } T_3 \end{cases}$$

The final scores are calculated by

$$\begin{cases} \dfrac{\log(\theta_i)}{\log(1-\theta_i)} & Y_i = 1 \\ & Y_i = 0 \end{cases}$$

Sets I_1 and I_2 are an approximately chronological split of one large casework set. To test if the effect seen was based on the chronological split these were subdivided into two sets of size 107 and 106 using random numbers. I_1 and I_2 are relabelled $I_{1'}$ and $I_{2'}$. The results are

also given in Table 6.1. Set P was pristine DNA. Five different 35:35 splits of the same data were trialled. The data are presented in Table 6.2.

Set used for training	Data set	Method applied		
		T_1	T_2	T_3
Train	I_1	-0.51	**-0.50**	-0.64
	I_2	-1.26	-1.19	**-1.09**
	I_3	-3.43	-3.39	**-3.36**
	I_4	-1.25	-1.26	**-1.02**
Train	I_1,	-0.48	**-0.47**	-0.61
	I_2,	-0.94	-0.93	**-0.72**
	I_3	-3.49	**-3.35**	-3.48
	I_4	-1.30	-1.29	**-1.03**
	I_1	-0.72	-0.75	**-0.64**
Train	I_2	-0.99	**-0.88**	-1.08
	I_3	**-3.23**	**-3.23**	-3.34
	I_4	-1.31	-1.33	**-1.05**
	I_1,	-0.72	-0.71	**-0.65**
Train	I_2,	-0.61	**-0.60**	-0.68
	I_3	-3.47	**-3.35**	-3.36
	I_4	-1.23	-1.21	**-1.04**
	I_1	-0.74	-0.72	**-0.68**
	I_1,	-0.67	**-0.64**	-0.68
	I_2	-1.23	-1.20	**-1.14**
	I_2,	-0.83	-0.81	**-0.71**
Train	I_3	-2.59	**-2.53**	-3.26
	I_4	-1.24	-1.20	**-1.14**
	I_1	-0.77	-0.77	**-0.66**
	I_1,	-0.73	-0.72	**-0.64**
	I_2	-1.46	-1.35	**-1.16**
	I_2,	-0.95	-0.88	**-0.74**
	I_3	-3.74	**-3.59**	-3.64
Train	I_4	-0.85	**-0.84**	-0.99

Table 6.1. Previous page: Log likelihoods per profile for different combinations of model and training sets I_1..I_4 and $I_{1'}$ and $I_{2'}$. The highest value in each row is marked in bold.

Table 6.2: Log likelihoods per profile for different splits of model and training sets for the PowerPlex® 21 set. The highest value in each row is marked in bold.

Trained on half of set P and tested on the other half.	Method applied		
	T_1	T_2	T_3
Train	**-1.6**	-1.7	-2.2
	-3.2	-3.3	-3.8
Train	**-1.7**	-1.8	-2.8
	-3.3	-3.4	**-3.2**
Train	**-2.2**	-2.3	-2.9
	-2.6	-2.7	-3.1
Train	**-1.6**	-1.7	-2.4
	-3.5	-3.6	-3.6
Train	**-2.6**	-2.7	-3.6
	-2.1	**-2.1**	-2.4

6.5 Discussion

The bigger, or higher, log likelihood is the more favoured scenario according to the data. In almost all tests performed on the Identifiler® sets method T_2 produced the highest log likelihood in the training set and T_3 produced the highest log likelihood in the test sets. We interpret this as meaning that a locus effect does exist but it is profile specific. That is, the locus effect that the model is trained on is not transportable. This is not an unexpected result. If template number is removed from locus effect, then it is reasonable to assume that any remaining locus effect must be amplification or profile specific. Namely, any preferential amplification/failure of amplification of loci will be due to inhibition, environmental conditions and contaminants – all factors which will be specific to the sample being analysed at the time.

For the PowerPlex® 21 data T_1 regularly gave the highest log likelihoods in the test sets. We interpret this as meaning that pristine source data does not show the expected degradation effect and therefore is not suitable to train these logistic systems. However, it would be interesting in future work to remove the locus effect in the T_1 model to determine if it improves the transportability of the model.

Of the three methods studied T_3 trained on casework data is narrowly the best for immediate use in casework due to its portability since it produced the largest log likelihoods more often. It is also worthwhile to emphasise that this research used casework profiles. This is important as this is the type of data that is most susceptible to locus effects outside of the amount of template available for amplification. The T_1 and T_2 models were developed using pristine DNA dilutions, therefore locus effects were likely to be minimal and transportable within the authors' data set.

We conclude that further development is required in the application of locus specific effects and it is likely that these locus effects will vary from profile to profile or between amplifications.

References

1. Gill, P., et al., DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. Forensic Science International, 2006. 160(2-3): p. 90-101.

2. Tvedebrink, T., et al., Estimating the probability of allelic drop-out of STR alleles in forensic genetics. Forensic Science International, 2009. 3: p. 222-226.

3. Nicklas, J.A., T. Noreault-Conti, and E. Buel, Development of a Real-Time Method to Detect DNA Degradation in Forensic Samples. Journal of Forensic Sciences, 2012. 57(2): p. 466-471.

4. Chung, D.T., et al., A study on the effects of degradation and template concentration on the amplification efficiency of the STR Miniplex primer sets. Journal of Forensic Sciences, 2004. 49(4): p. 733-740.

5. Bright, J.-A., et al., Developing allelic and stutter peak height models for a continuous method of DNA interpretation. Forensic Science International: Genetics, 2013. 7(2): p. 296-304.

6. Bright, J.A., et al., Degradation in forensic DNA profiles. Australian Journal of Forensic Sciences, 2013. IN PRESS.http://dx.doi.org/10.1080/00450618.2013.772235

7.. Tvedebrink, T., et al., Statistical model for degraded DNA samples and adjusted probabilities for allelic drop-out. Forensic Science International: Genetics, 2012. 6(1): p. 97-101.

CHAPTER 7: MODELLING HETEROZYGOTE BALANCE IN FORENSIC DNA PROFILES

Hannah Kelly [a,b], Jo-Anne Bright [a], James M. Curran [b], John Buckleton [a]

a) ESR, PB 92021, Auckland, 1142, New Zealand
b) Department of Statistics, University of Auckland, PB 92019, Auckland, 1142, New Zealand

This paper appeared in Forensic Science International: Genetics, Volume 6, Issue 6, December 2012, Pages 729-734. This was a focus issue of Forensic Science International: Genetics, titled: Analysis and biostatistical interpretation of complex and low template DNA samples.

The DOI is: http://dx.doi.org/10.1016/j.fsigen.2012.08.002

The aim of this paper is to model the distribution of heterozygous balance in forensic DNA case work samples so that predictions can be made regarding the expected heterozygous balance of questioned genotypes. Case work samples are utilised in order to witness the stochastic effects, degradation, and other variables that might affect the resulting peak height of an allele after amplification. Using case work data also results in a larger representation of alleles than dilution or pristine studies that usually only create profiles from a limited number of donors.

In particular, this paper investigates the behaviour of heterozygous balance at low template levels. The resulting model that is constructed uses the difference in allele designation as an explanatory variable for the expected (mean) heterozygous balance, and the variance of heterozygous balance is shown to decrease at a rate inversely proportional to the average peak height at the locus. That is, the variance of heterozygous balance is more extreme at low template levels.

Guidelines exist for the interpretation of conventional DNA profiles using heterozygous balance, and this work investigates the application of these guidelines to low template work. This work shows that the upper boundary of the current guidelines (1.67) is “safe” when the average peak height of a profile is above 174 RFU and the lower boundary (0.60) is “safe” when the average peak height is above 434 RFU. Above 1000 RFU, the 95% credible intervals developed in this research become much narrower than the traditional guidelines

which could be useful in the interpretation of complex mixtures as fewer genotype combinations would be considered probable.

This work was conducted with the aim of producing a model that could be incorporated into software that utilises a continuous model for the interpretation of forensic DNA profiles. Alternatively, we hope this model will be used as an aide by the scientist if using the binary model. For example; the scientist can calculate the expected heterozygous balance and credible interval and compare these values to the observed values at the locus. The scientist can use this comparison to determine how probable the questioned genotype is.

The BUGS model code is included in the supplementary material, following the references, at the end of this chapter. This should allow the reader to apply the methodology to their own data set. The code for running JAGS is not included as it is specific to the data set.

This work was supported in part by grant 2011-DN-BX-K541 from the US National Institute of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice. The authors declare no conflict of interest. The authors would like to thank Catherine McGovern, Sue Vintiner and two anonymous reviewers for the helpful contributions towards this paper.

7: MODELLING HETEROZYGOTE BALANCE IN FORENSIC DNA PROFILES

Abstract

In this paper we investigate the relationship between heterozygous balance (h) and average peak height ($\bar{\phi}$) in a set of Identifiler® data. The mean of heterozygote balance is unaffected by average peak height but the variance about this mean is much lower at high average peak heights.

Keywords Heterozygous balance; mixed DNA profiles; LtDNA; DNA thresholds; DNA interpretation.

7.1 Introduction

Modern forensic DNA analysis is dominated by the use of the polymerase chain reaction (PCR) to amplify short tandem repeat (STR) loci [1]. Typically DNA, from samples associated with a crime, is extracted from body fluids and/or cellular material present on a wide range of surfaces. A sample of this extract is taken and amplified using the PCR process. The reagents used are tagged with a fluorescent dye that becomes incorporated into the amplified product. The resulting post-amplification product can be visualised using capillary electrophoresis which results in an electropherogram (epg). Alleles appear as peaks in the epg. There is an approximate relationship between the height of the resulting peaks and the amount of DNA template in the extract.

The relationship between height and starting template is stochastic because the sampling, extraction, and PCR stages all introduce variability into the process. The factors affecting this variability are thought to be both random and systematic. One manifestation of this variability is the relative height of the two peaks of a heterozygote; termed heterozygote balance (h).

Heterozygote balance refers to the ratio of peak heights (or areas) between the two alleles of a heterozygote. In this paper we will use peak heights exclusively. h can be regarded as a summary of relative variation of peak heights. There are two well-known definitions of

heterozygous balance and both are typically referred to as hb. In this paper we will use the definition

$$h = \frac{\phi_{HMW}}{\phi_{LMW}} \tag{1}$$

where HMW and LMW refer to the higher and lower molecular weight allele respectively, and where ϕ is peak height. We will discuss the alternative definition in Section 7.5.1.

If multiple epgs were to be formed from the same extract, then the heterozygote balance of two alleles at the same locus would vary between epgs. Therefore, in a more formal statistical framework, we would call h a random variable, and describe its pattern of behaviour with a probability density function. Many probability distributions are described, or parameterized, by their mean, and sometimes their variance. The normal distribution is an extremely common example. Therefore, if we can use data to describe the mean and the variance in h, then we may be able to model h using a distribution that exhibits the observed behaviour.

It is important to understand the variability in h because h is used heavily in both the interpretation of mixed DNA samples and in the interpretation of low template DNA samples (LtDNA). For example, h may be used to classify combinations of alleles (or genotypes) as possible or impossible when considering a mixture. This use of h is the defining feature of the binary model [2,3]. There are acknowledged shortcomings of the binary model (see [4] for an early review) and newer models are able to utilise the peak height information without this possible/impossible dichotomy. Such models are inherently superior [5], however they, even more so, rely on an understanding of the variability inherent in peak height information. Even single source DNA profiling benefits from an understanding of the variability in h.

At LtDNA levels one peak of a heterozygote may be so imbalanced that it either does not exceed the threshold set for the declaration of an allele or it is simply not detected at all. This allele is said to have dropped out. The situation where the person of interest has an allele not present in the resulting epg is termed a non-concordance [6] and is difficult to interpret at LtDNA levels using the binary model [7]. The more elegant solution lies in models that treat peak heights and allele designations as probabilistic [5], [8].

Previous work has identified several factors that appear to influence observed heterozygote balance. The mean of h is affected by the difference in the number of repeat sequences (δ) between the alleles at a heterozygote locus. Alleles with a larger number of repeat sequences are thought to produce relatively smaller height peaks [4] both because they stutter more but also because they amplify less [9]. The variability of h as a function of the average peak height ($\overline{\phi}$) of the two peaks has also been investigated [10-12]. These works found that variability was much greater at lower peak heights.

There have been efforts made to statistically model the DNA extraction and amplification processes and to observe the effect on heterozygote balance [13-16]. Comparisons of the predictions made by these models with empirical data are encouraging and suggest that at least the largest factors affecting the distribution of h have been identified. In this paper we undertake an investigation in an attempt to refine the model.

7.2 Data preparation

Ninety five ($n = 95$) single source Identifiler® DNA profiles were taken from casework samples of varying profile quality across a range of sample types including; bloodstains, saliva stains and cigarette butts. Samples were extracted using an organic [17] or DNA IQ [18] extraction method depending on the sample type.

All samples were quantified prior to amplification using Quantifiler (Applied Biosystems) on a 7500 (Applied Biosystems) according to the manufacturer's instructions [19]. Amplification was performed with Identifiler® (Applied Biosystems) in a 9700 thermal cycler with silver block. Amplified products were separated on 3130xl Genetic Analysers (Applied Biosystems) and the analysis of DNA profiling data was undertaken using GeneMapper ID version 3.2 (Applied Biosystems) using the panels and bins provided by Applied Biosystems. A peak detection threshold of 25 RFU was applied. All stutter alleles that were one repeat unit shorter than the parent allele were retained. Stutter, also known as reverse or back stutter, is defined as a minor product that is one repeat sequence shorter than the main peak and is thought to be caused by slipped strand mis-pairing during PCR [20]. Stutter peaks were included for this analysis as they are thought to provide a more accurate approximation to DNA quantity at a locus when their height is combined with that of the parent allele. Other types of stutter including forward or over stutter, and double back stutter were ignored if present.

The difference, δ, in the number of repeat units between alleles at each heterozygous locus was calculated by subtracting the allelic designation of the smaller molecular weight allele from the allelic designation of the larger molecular weight allele. For example; if at locus vWA we were to observe an allelic vector of $\{15,17\}$, then $\delta = 2$. All loci where $\delta = 1$ were removed from this data set because the stutter from the higher molecular weight allele cannot be separated from the lower molecular weight allele. This is important because of the way we calculate average peak height. Amelogenin was also removed from the data set because it is not an STR. In total, 644 heterozygous loci were retained in the data set.

The average peak height ($\overline{\phi}$) at each locus was calculated to include stutter peaks:

$$\overline{\phi} = \frac{\left(\left(\phi_A + \phi_S\right)_{HMW} + \left(\phi_A + \phi_S\right)_{LMW}\right)}{2} \tag{2}$$

where ϕ_A is the height of the allelic peak and ϕ_S is the height of its respective stutter peak.

The heterozygote balance for each heterozygous locus was calculated using Equation (1) and statistical analyses were carried out in R [21]. Further information on some of the statistical analyses used in this paper can be found in Curran [22].

7.3 Exploratory data analysis and modelling

We attempt here to model both the expectation (mean) and variance of $\log_e(h)$ using a linear model with candidate explanatory variables $\overline{\phi}$, δ and a locus effect. Many ratios are easier to model if logarithms are taken. In this work we experimented with both h and $\log_e(h)$, and found that the model was improved when using $\log_e(h)$. We have therefore persisted with the latter.

Previous work [4] has shown that there is a relationship between the difference in repeat units, δ, and $\log_e(h)$. Buckleton et al. [4] and others [23] have reported that, for heterozygous STR loci, there is a small, but significant, tendency for the smaller allele to amplify more efficiently than the larger allele. Exploratory data analysis on the data used in this paper replicates this finding, with h decreasing on average 3% for each unit increase in

δ. In addition, we found no evidence of a difference in h for different loci, nor was there an effect (on the mean) due to $\overline{\phi}$.

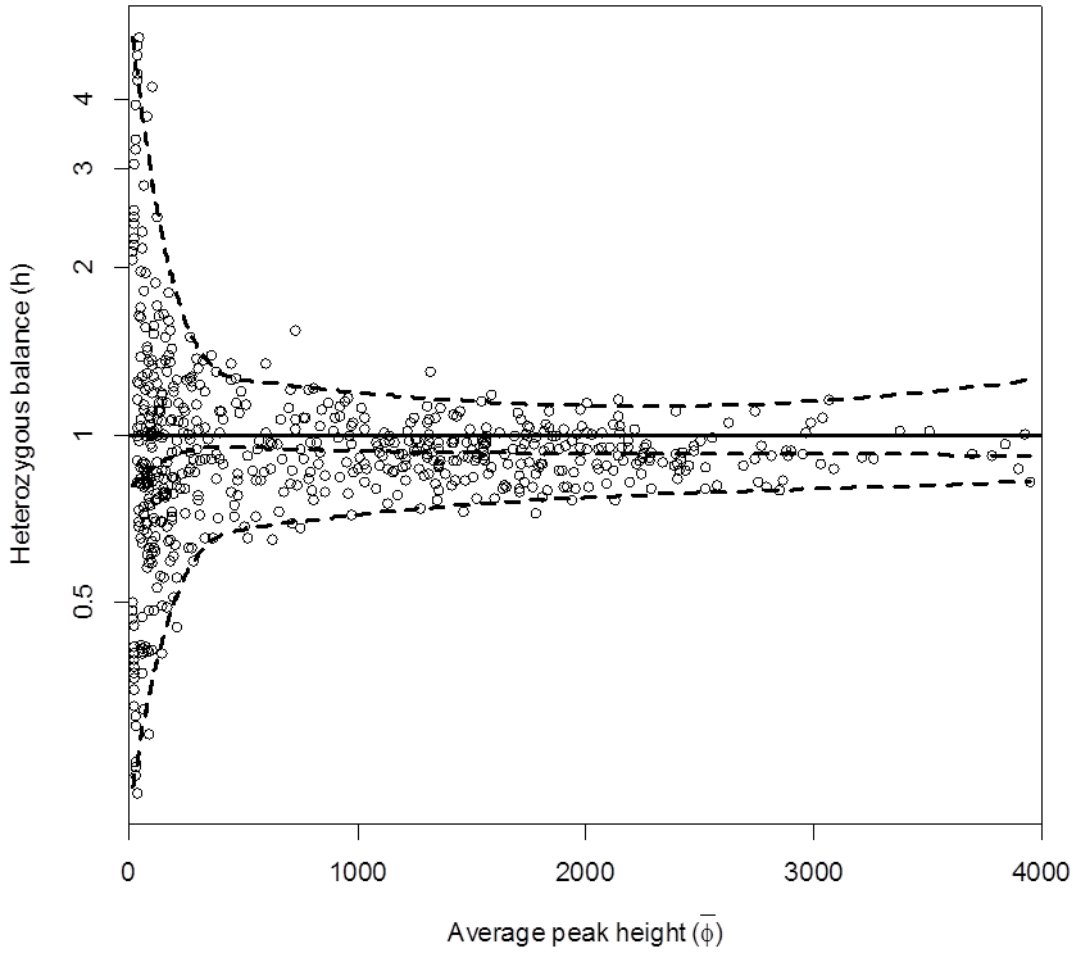


Figure 7.1: Heterozygous balance versus average peak height

The scatter plot in Figure 7.1 shows the relationship between $\overline{\phi}$ and $\log_e(h)$. In an idealized situation two heterozygous alleles amplified from the same amount of template DNA should have the same height in the epg. This idealized relationship is represented by the horizontal line at $h=1$. In reality, there are a number of factors which may result in differences. Previous work [24] has suggested that there is a relationship between average peak height, $\overline{\phi}$, and heterozygous balance. Analysis of our own data shows this effect is an artefact of the inferior definition of h employed in [24] and not a true representation of the underlying

behaviour. The black dashed lines in Figure 7.1 are quantile regression [25,26] lines for the 0.025, 0.500, and 0.975 quantiles of h, with B-splines applied to $\bar{\phi}$. We use this technique to robustly model the mean/median relationship between $\bar{\phi}$ and $\log_e(h)$ and, more importantly the variance relationship between $\bar{\phi}$ and $\log_e(h)$. We note that the central (median) line is slightly lower than the line $h = 1$ which indicates that, in our data set, the higher molecular weight allele is smaller on average than the lower molecular weight allele. The median line is also relatively constant over the range of $\bar{\phi}$ which we take this as evidence that $\bar{\phi}$ has relatively little influence on h. The curved dashed lines are can be thought of as roughly representing a 95% prediction interval for h given a certain value of $\bar{\phi}$. These curves guide us in our choice of model for the variability of h given $\bar{\phi}$ Figure 7.1 shows that variability (represented by the distance between the two curved dashed lines) decreases as $\bar{\phi}$ increases. On this basis we postulate that the observed variation in h is inversely proportional to $\bar{\phi}$.

7.4 Model construction

The data analysis in Section 7.3 enables the construction of a Bayesian model which explicitly includes the variance relationship and allows intuitive inferences about the model parameters. Our model is similar to that proposed by Tvedebrink et al. [27], but we take a Bayesian approach to model fitting rather than using maximum likelihood estimation. We briefly compare these two approaches, which yield very similar results, in the appendix to this paper.

We divide the model formulation process up into three distinct stages. These are 1) modelling the distribution for h, 2) modelling the mean relationship for h with respect to δ, and possibly $\bar{\phi}$, and 3) modelling the variance relationship for h.

We propose that the logarithm of h is normally distributed or alternatively, that h is lognormally distributed, for a given mean and variance. Whilst this is unlikely to be exactly true we consider it to be a fair approximation as Figure 7.1 suggests that the logarithm of h is approximately symmetric around the mean. Appealing to symmetry is not a compelling argument, as there are many families of symmetric distributions we may choose from. However, the normal and the lognormal families have wide acceptance in biology and

science in general. Furthermore, we will constrain the mean value of $\log_e(h)$ to pass through the origin when $\delta = 0$. This reflects the idea that when $\delta = 0$ we are dealing with a homozygote rather than a heterozygote, and so the concept of heterozygous balance becomes redundant. We express this as:

$$\log_e(h_i) \sim N\left(\mu_i, \sigma_i^2\right) \tag{3}$$

i.e. the distribution of $\log_e(h_i)$ is Normal with a mean of μ_i and a variance of σ_i^2. The parameters μ_i and σ_i^2 depend on δ and $\overline{\phi}$. These dependencies are modelled by

$$\mu_i = \beta\delta_i \tag{4}$$

and

$$\sigma_i^2 = \frac{\sigma^2}{\overline{\phi}_i} \tag{5}$$

respectively. Equation (4) is the mean model, and describes the dependency of the mean value of h on the difference between the numbers of the repeat sequences (δ_i) as a simple linear dependency where the intercept is zero, and β is the slope on δ_i.

Equation (5) models the observed relationship that the variability in h decreases as $\overline{\phi}$ increases. Again the inverse proportionality model is unlikely to be exactly true but it appears to fit the data well.

The package rjags was used to fit our model in R [28]. JAGS is variant of BUGS [29], a statistical package that allows the user to fit Bayesian models using Markov chain Monte Carlo (MCMC) techniques.

The result of the MCMC sampling scheme produces (correlated) samples from the posterior distribution of the parameters. Summary statistics on each of the samples of the parameters are given in Table 7.1.

Table 7.1: Posterior summary statistics from the Bayesian model.

Statistic	Mean	Median	Std. Dev.	95% HPD
β	-0.025	-0.025	0.002	(-0.021, -0.029)
σ_i^2	19.268	19.218	1.092	(17.281, 21.505)

We can use the mean model with the posterior parameter means for crude mean prediction. If $\log_e h_i \sim N\left(\mu_i, \sigma_i^2\right)$, then

$$E\left[h_i\right] = \exp\left(\mu_i + \frac{\sigma_i^2}{2}\right)$$

and

$$Var\left[h_i\right] = \left(\exp\left(\sigma_i^2\right) - 1\right)\exp\left(2\mu_i + \sigma_i^2\right)$$

An approximate credible interval for this estimate may be obtained from:

$$\exp\left(\mu_i \pm z_\alpha^* \sigma_i\right)$$

where z_α^* is the $(1-\alpha/2)$ quantile of the standard normal distribution. Using our posterior means as plug-in estimates of the parameters in this expression we can obtain a 95% credible interval from:

$$\exp\left(-0.025 \times \delta \pm 1.96\sqrt{\frac{19.27}{\bar{\phi}}}\right)$$

For example; if we have two alleles present at a locus, say $\{15,17\}$ then $\delta = 2$. Furthermore, if the $\bar{\phi}$ at the locus was 400 RFU , then the expected h is:

$$\exp\left(-0.025 \times 2 + 0.5 \times \frac{19.27}{400}\right) = 0.974$$

An approximate 95% credible interval for this value is:

$$\exp\left(-0.025\times 2 \pm 1.96\sqrt{\frac{19.27}{400}}\right) = (0.62,1.46)$$

We read this as h is within the bounds $(0.62,1.46)$ with probability 0.95. Note that this is a credible interval for a new h value. A credible interval for an average value of h will be much narrower. We can compare the observed value of h to the credible interval for a given $\bar{\phi}$ and δ to determine if our data falls within the 95% credible intervals. Note that these values fall within the recommended guidelines of $(0.60,1.67)$ suggested by Buckleton et al. [4]. Equivalently, if one has access to a function that calculates the quantiles of a lognormal distributed random variable (such as the qlnorm function in R), then a credible interval can be obtained by getting the 0.025 and 0.975 quantiles from a lognormal distribution with mean -0.025×2 and standard deviation $\sqrt{19.27/400}$. The results from this function are the same as our approximate result to four decimal places of accuracy.

7.4.1 Comparison to the conventional thresholds

Analysis was undertaken to determine where the conventional thresholds and the 95% credible intervals derived in this work intersect. It should be noted that the traditional bounds were developed for SGM+ , whereas this work uses Identifiler® . In the current model the traditional lower boundary of 0.60 is `safe' when the average peak height is above 434 RFU. The upper boundary of 1.67 is `safe' when the average peak height is above 174 RFU. Figure 7.2 shows the 95% credible intervals and the conventional h thresholds with respect to $\bar{\phi}$. The graph shows that the distribution of h is not even, and widens at low $\bar{\phi}$. The conventional thresholds can be supported when they are outside the 95% credible intervals but not once they intersect. The graph also shows that above approximately 1,000 RFU, the 95% credible intervals become much narrower. These narrower guidelines could be useful in the interpretation of complex mixtures at higher levels as fewer genotype combinations would be considered possible.

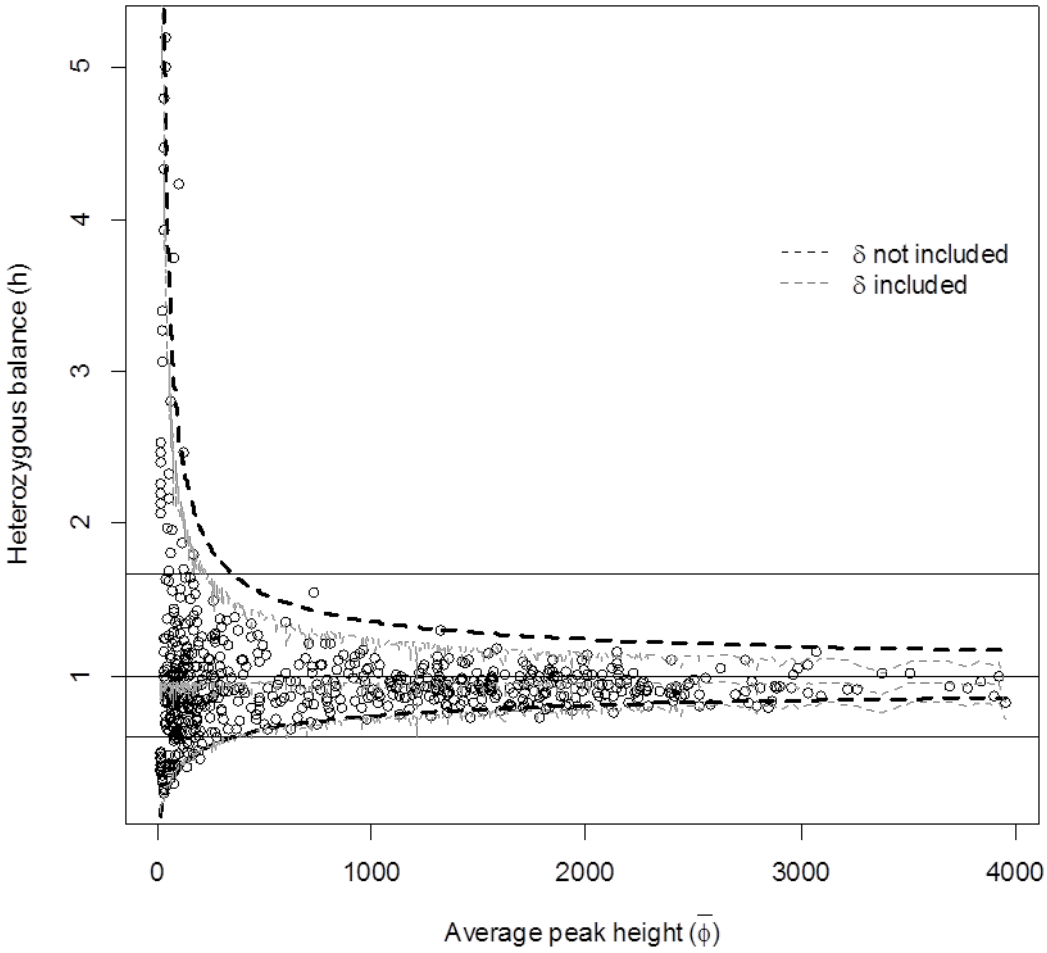


Figure 7.2: Fitted values from the lognormal model with and without the effect of δ. The horizontal lines represent h thresholds (0.60, 1.67). The dashed lines indicate the 95% credible intervals determined in this work.

7.4.2 Model assessment

It is useful to provide some assessment of how well the lognormal distribution models the observed data. In the preceding sections we have modelled the conditional distribution of heterozygous balance given average peak height and a difference in repeat units. A standard tool for evaluating model fit is a quantile-quantile (Q-Q) plot of the residuals. We have a minor difficulty in that, unlike a standard regression model, the variance is not constant with respect to the explanatory variables. However, we do have an estimate of the variance of each

observation, and so we can standardize the residuals so that they have mean zero and unit variance. That is,

$$r_i = \frac{\log_e(h_i) - \hat{\beta}\delta_i}{\hat{\sigma} / \sqrt{\bar{\phi}_i}}$$

We can plot the residuals r_i against the quantiles of the standard normal distribution to get a normal Q-Q plot. This is shown in Figure 7.3.

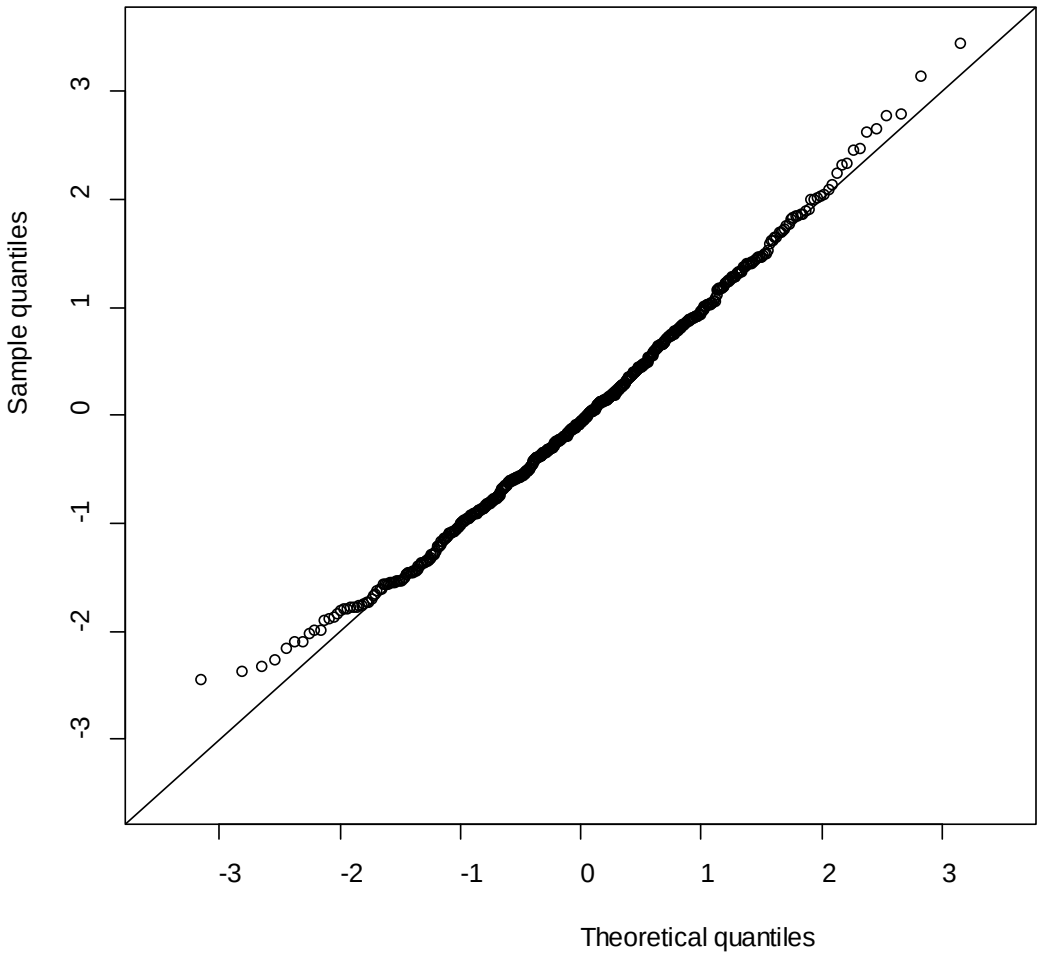


Figure 7.3: Normal Q-Q plot using the conditional distribution of h given $\bar{\phi}$

There is slight upward curvature in the Q-Q plot, which is indicative of right skew in the residuals. That is, the lognormal distribution is not capturing all of the skewness of the heterozygous balance statistic. Although not presented here, our work with the gamma model seems to remove this shortcoming. However, the trade-off in complexity between the gamma and lognormal model and a small amount of right skewness leads us to fall more on the side of the lognormal model.

7.5 Discussion

The ``fuzzy'' nature of the grey lines in Figure 7.2 is due to the inclusion of δ in the model which have been set to the observed δ values. It is apparent that, although statistically significant, δ has little impact on the credible intervals. The black dashed lines represent the 95% credible intervals for h respect to $\overline{\phi}$ ignoring δ.

It is important to note that the numerical results here are ``tuned'' for the particular data set used. That is, these results may not be transferable from lab to lab and may lose their value as machines age. However, the models do transfer and can be easily fitted to new data.

7.5.1 Comparison of methods used to calculate heterozygous balance

As mentioned above there are two main methods that are used within the literature to calculate heterozygous balance. The calculation:

$$h^{(2)} = \frac{\phi_{smaller}}{\phi_{larger}} \tag{6}$$

where $\phi_{smaller}$ and ϕ_{larger} represent heights of the alleles with the smaller and larger peak heights respectively, is common [24].

We suggest, however, that calculating heterozygous balance by this method results in a misleading representation of the data and a loss of information content. We attempt to explain our reasoning below.

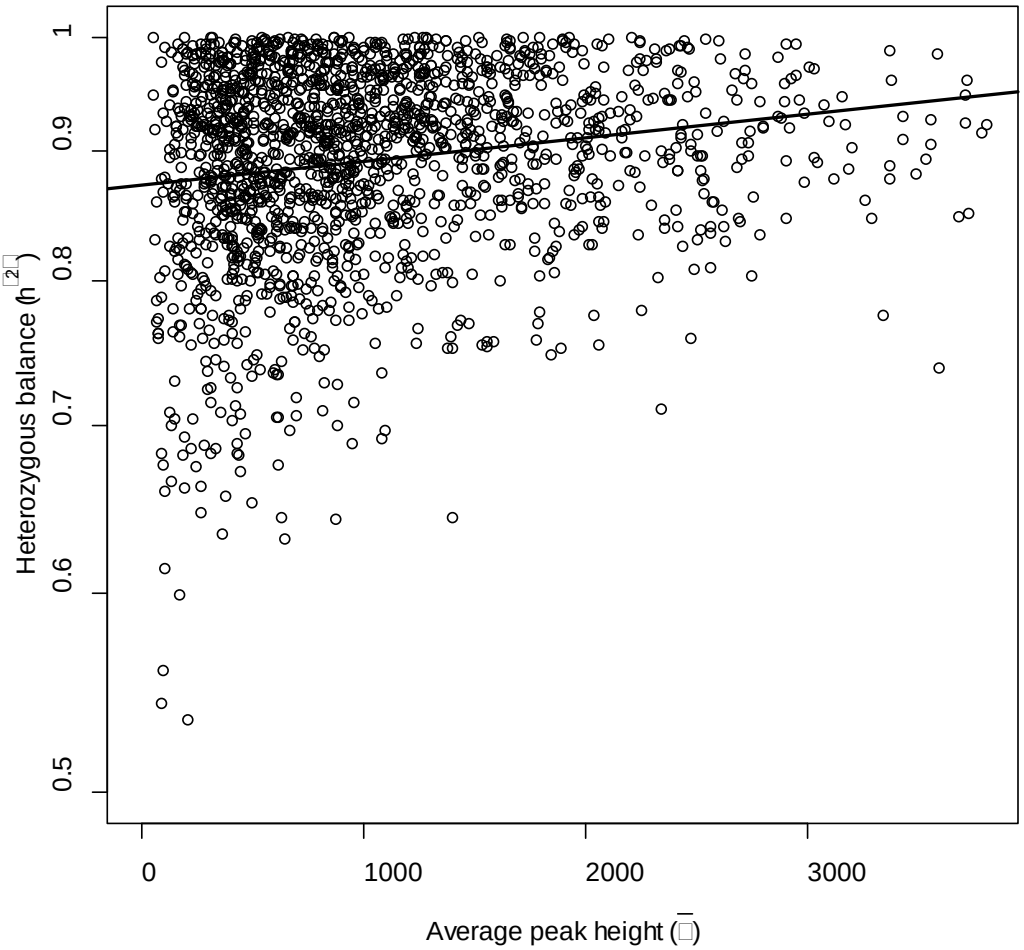


Figure 7.4: $h^{(2)}$ versus $\overline{\phi}$. The scale is logarithmic.

Figure 7.4 shows the behaviour of heterozygous balance with respect to average peak height when Equation (6) is used to compute heterozygous balance. Below approximately 1,000 RFU there is a decrease in peak height that results in a decrease in the mean of $h^{(2)}$. Simple linear regression shows that average peak height is a statistically significant predictor of $h^{(2)}$. However, as we have shown, this association of $\overline{\phi}$ to the mean is induced entirely by the statistic used to compute heterozygous balance and the increased variability of heterozygous balance (regardless of definition) at lower levels of template DNA.

We suggest that when calculating heterozygous balance Equation (1) provides a fairer representation of the data. In the discussion that follows we will refer to this definition as $h^{(1)}$

The definition of $h^{(1)}$ is fixed. That is, it does not change between two different epgs with the alleles present because the lower molecular weight allele will always be so regardless of the height of the peak, and similarly for the higher molecular weight allele. It is because of this fact that $h^{(1)}$ supplies us with more information than $h^{(2)}$. For example, a resulting ratio of less than 1 obtained using $h^{(1)}$ would indicate that the lower molecular weight allele was been preferentially amplified. This cannot be determined if we have used $h^{(2)}$ unless we use additional information.

7.6 Conclusion

The ultimate goal of this research was to model the joint distribution of h and $\bar{\phi}$, as it is a convenient and common way of looking at the distribution of peak heights.

To simplify the modelling we do the usual factoring, so that.

$$f(h,\bar{\phi}) = f(h \mid \bar{\phi}) f(\bar{\phi})$$

In this paper we have concentrated on building a model for $f(h \mid \bar{\phi})$. Data was collected that allowed us to model this distribution, and through data analysis we have identified what we regard as a satisfactory model. The difference in the number of repeat sequences (δ) between alleles has been identified as having a significant effect on the mean of h. The variance of h has been shown to decrease at a rate inversely proportional to the average peak height at the locus. Thus, the variance of h is most extreme at LtDNA levels.

We have used Bayesian modelling techniques to build a model that uses δ and $\bar{\phi}$ to give an expected estimate for h as well as a credible interval that h would be expected to fall within. Although δ was shown to be statistically significant, the effect of δ on the mean was relatively small and δ can, for practical purposes, be dropped from the model. Our model appears to describe the observed data well. This model differs to some of the findings within the literature [11], [24].

We have explicitly modelled the ratio of two peak height measurements in this work. This type of modelling may not address the root sources of variation. A better result might be achieved by modelling the variation of the peak heights directly with a constant component

and a component that is proportional to the amount of template DNA. Such models are currently being considered by the authors and a number of other researchers.

We have demonstrated that defining heterozygous balance in terms of the molecular weights of the alleles preserves information. We have also shown that when heterozygous balance is defined in this manner, there is no dependency between the average value of h and average peak height. For these reasons, we recommend this calculation over the definition given in Equation (6).

We hope that the model in this work may be used as an aide by the scientist when using the binary model. For example; the scientist may use this model to calculate both an expected value and credible interval for h and compare this to the observed h. The scientist can then use this comparison to determine if the questioned genotype is probable.

However, this model is a step away from the threshold based criterion and is a step toward a more probabilistic approach. We intend this model to be used in a semi-continuous or continuous model for the interpretation of DNA. The threshold based binary model fails when there are non-concordances present in a profile (dropout) and we must move toward models that can deal with this possibility [5]. Dropout is an extreme version of h and the proposed model can aid the scientist in determining the probability that a partner allele may have dropped out of the profile.

Appendix - Comparison with Maximum Likelihood Estimators

The maximum likelihood estimators for our model are reasonably easily derived. They are

$$\hat{\beta}_{mle} = \frac{\bar{\phi}_\bullet \delta_\bullet y_\bullet}{\bar{\phi}_\bullet \delta_\bullet^2} \tag{7}$$

and

$$\hat{\sigma}^2_{mle} = \frac{1}{n}\sum_{i=1}^{n} \bar{\phi}_i \left(y_i - \hat{\beta}\delta_i \right)^2 \tag{8}$$

where $y_i = \log_e(h_i)$ and $\bullet$ notation means the summation over all observations, e.g.

$$y_\bullet = \sum_{i=1}^{n} y_i$$

Equation (8) reduces to

$$\hat{\sigma}^2_{mle} = \frac{1}{n}\sum_{i=1}^{n} \bar{\phi}_i y_i^2 \tag{9}$$

when $\beta = 0$. In this Appendix we compare the maximum likelihood estimates to the estimates we found using MCMC. We used the following priors for β and σ^2 in our Bayesian analysis:

$$\beta \sim N(0, 10^6)$$

$$\frac{1}{\sigma^2} \sim \Gamma(10^{-3}, 10^{-3})$$

where the parameterization for the the Gamma distribution is in terms of shape and rate. Applying (7) and (8) to our data yields the maximum likelihood estimates in Table 7.2.

Table 7.2: Comparison of ML estimates and the estimated posterior means of β and σ^2.

Parameter	MLE	MCMC
β	-0.02513	-0.02516
σ^2	19.186	19.269

When $\beta = 0$, the posterior mean is $\hat{\sigma}^2_{mcmc} = 24.183$ and the ML estimate (from (9)) is $\hat{\sigma}^2_{mle} = 24.101$.

Table 7.3: Comparison of ML estimates and Bayesian estimates with respect to coverage (%).

δ incl.	MLE	MCMC
No	94.41	94.41
Yes	92.97	95.53

The coverage shows slight differences between the two methods when the effect of δ is included in the model. This is because the posterior distribution of the variance, σ^2 is right-skewed, and hence the mean will be slightly larger than the median and the mode where the maximum likelihood estimate is positioned.

References

[1] A. Edwards, H. A. Hammond, L. Jin, C. T. Caskey, and R. Chakraborty. DNA typing and genetic mapping at five trimeric and tetrameric tandem repeats. American Journal of Human Genetics, 49:746--756, 1991.

[2] T. Clayton, J. P. Whitaker, R. L. Sparkes, and P. Gill. Analysis and interpretation of mixed forensic stains using DNA STR profiling. Forensic Science International, 91:55 -- 70, 1998.

[3] P. Gill, R. L. Sparkes, R. Pinchin, T. Clayton, J. P. Whitaker, and J. S. Buckleton. Interpreting simple STR mixtures using allelic peak areas. Forensic Science International, 91:41--53, 1998.

[4] J. S. Buckleton, C. M. Triggs, and S. J. Walsh. Forensic DNA Evidence Interpretation. CRC Press, Boca Raton, FL, 2004.

[5] M. W. Perlin, M. M. Legler, C. E. Spencer, J. L. Smith, W. P. Allan, J. L. Belrose, and B. W. Duceman. Validating TrueAllele DNA mixture interpretation. Journal of Forensic Sciences, 56(6):1430--1447, 2011.

[6] D. J. Balding and J. Buckleton. Interpreting low template DNA profiles. Forensic Science International: Genetics, 4(1):1--10, 2009.

[7] J. Buckleton and C. M. Triggs. Is the 2p rule always conservative? Forensic Science International, 159:206--209, 2006.

[8] P. Gill, J. P. Whitaker, C. Flaxman, N. Brown, and J. S. Buckleton. An investigation of the rigor of interpretation rules for STRs derived from less that 100 pg of DNA. Forensic Science International, 112(1):17--40, 2000.

[9] C. Brookes, J. Bright, S. Harbison, and J. Buckleton. Characterising stutter in forensic STR multiplexes. Forensic Science International: Genetics, 6(1):58--63, 2012.

[10] J. Bright, J. Turkington, and J. Buckleton. Examination of the variability in mixed DNA profile parameters for the Identifiler® multiplex. Forensic Science International: Genetics, 4(2):111--114, 2009.

[11] J. R. Gilder, K. Inman, W. Shields, and D. E. Krane. Magnitude-dependent variation in peak height balance at heterozygous STR loci. International Journal of Legal Medicine, 125:87--94, 2011.

[12] J. Bright, E. Huizing, L. Melia, and J. Buckleton. Determination of the variables affecting mixed MiniFiler™ DNA profiles. Forensic Science International: Genetics, 5(5):381--385, 2011.

[13] P. Gill, J. M. Curran, and K. Elliot. A graphical simulation model of the entire DNA process associated with the analysis of short tandem repeat loci. Nucleic Acids Research, 33(2):632--643, 2005.

[14] H. Haned. Forensim: An open-source initiative for the evaluation of statistical methods in forensic genetics. Forensic Science International: Genetics, 5(4):265--268, 2011.

[15] H. Haned, T. Egeland, D. Pontier, L. Péne, and P. Gill. Estimating drop-out probabilities in forensic DNA samples: A simulation approach to evaluate different models. Forensic Science International: Genetics, 5(5):525--531, 2011.

[16] J. Weusten and J. Herbergs. A stochastic model of the processes in PCR based amplification of STR DNA in forensic applications. Forensic Science International: Genetics, 6(1):17--25, 2012.

[17] J. Bright and S. F. Petricevic. Recovery of trace DNA and its application to DNA profiling of shoe insoles. Forensic Science International, 145(1):7--12, 2004.

[18] Promega Corporation. DNA IQ system - small sample casework protocol, 2002.

[19] R. L. Green, I. C. Roinestad, C. Boland, and L. K. Hennessy. Developmental validation of the Quantifiler real time PCR kits for the quantification of human nuclear DNA samples. Journal of Forensic Sciences, 50:1--17, 2005.

[20] P. S. Walsh, N. J. Fildes, and R. Reynolds. Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. Nucleic Acids Research, 24:2807--2812, 1996.

[21] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.

[22] J. M. Curran. Introduction to data analysis with R for forensic scientists. CRC Press, Boca Raton, FL, 2010.

[23] J. Veth. The charactarization of artefacts found in DNA profiles. Master of Science thesis, University of Auckland, 2002.

[24] A. Debernardi, E. Suzanne, A. Formant, L. Péne, A. B. Dufour, and J. R. Lobry. One year variability of peak heights, heterozygous balance and inter-locus balance for the DNA positive control of AmpFℓSTR® Identifiler® STR kit. Forensic science international. Genetics, 5(1):43--49, 2011.

[25] R. Koenker. Quantile regression, volume 38. Cambridge Univ Pr, 2005.

[26] R. Koenker. quantreg: Quantile Regression, 2012. R package version 4.79.

[27] T. Tvedebrink, H. Smidt Mogensen, M. C. Stene, and N. Morling. Performance of two 17 locus forensic identification STR kits -- Applied Biosystems' AmpFℓSTR® NGMSElect and Promega's PowerPlex ESI17 kits. Forensic Science International: Genetics, 2012.

[28] M. Plummer. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In Proceedings of the 3 International Workshop on Distributed Statistical Computing, 2003.

[29] D. J. Lunn, D. Spiegelhalter, A. Thomas, and N. Best. The BUGS project: Evolution, critique and future directions (with discussion). Statistics in Medicine, 28:3049--3082, 2009

[30] R. G. Cowell, S. L. Lauritzen, and J. Mortera. Probabilistic modelling for DNA mixture analysis. Forensic Science International: Genetics Supplement Series, 1(1):640--2, 2008.

Supplementary material (BUGS model specification)

```
model{
    for(i in 1:N){
        hb[i] ~ dlnorm(mu[i], tau[i])
        mu[i] <- b0 + b1*delta[i]

        tau[i] <- aph[i]*invsigsq
        sigma[i] <- 1/sqrt(tau[i])
    }
    b0 ~ dnorm(0, 0.000001)
    b1 ~ dnorm(0, 0.000001)
    invsigsq ~ dgamma(0.001, 0.001)
    sigmasq <- 1/invsigsq
}
```

CHAPTER 8: IDENTIFYING AND MODELLING THE DRIVERS OF STUTTER IN FORENSIC DNA PROFILES

Hannah Kelly [1, 2], Jo-Anne Bright [1], John Buckleton [1], James Curran [2]

1) ESR, PB 92021, Auckland, 1142, New Zealand,
2) Department of Statistics, University of Auckland, PB 92019, Auckland, 1142, New Zealand

This paper has been accepted for publication by the Australian Journal of Forensic Sciences andwas published online on the 5th of July 2013.

The DOI is: http://dx.doi.org/10.1080/00450618.2013.808697

This paper investigates what causes alleles to stutter. The purpose of this paper is to identify the drivers of stutter and use them to build a model in which the expected stutter ratio (height of the stutter peak divided by the height of the parent peak) can be predicted for a questioned peak. It was found that the stutter ratio is affected by the longest uninterrupted repeat sequence (*LUS*) present in the parent allele. Locus also causes the stutter ratio to differ within a profile. A linear model is constructed which describes the behaviour of the expected stutter ratio with respect to locus and *LUS*.

The data used to identify the explanatory variables and to build the model are based on case work samples. Case work samples are utilised in order to observe a variety of alleles at each locus. It is important to the author that the resulting model is not built based only on the behaviour of the stutter ratio of a few alleles and that the research utilises a sample representing the alleles present within the New Zealand population.

This research is undertaken with the intention that the resulting model be implemented in a semi-continuous or continuous modelling system for the analysis of forensic DNA profiles.

The BUGS model code is included in the supplementary material, following the references, at the end of this chapter. This should allow the reader to apply the methodology to their own data set. The code for running JAGS is not included as it is specific to the data set.

This work was supported in part by grant 2011-DN-BX-K541 from the US National Institute of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice. The authors would like to thank Catherine McGovern for her contributions towards this paper.

8: IDENTIFYING AND MODELLING THE DRIVERS OF STUTTER IN FORENSIC DNA PROFILES

Abstract

This paper investigates the variables that could effect the stutter ratio (SR). Bayesian modelling techniques are used to model the distribution of the SR using the parameters identified; locus and the longest uninterrupted sequence (*LUS*). The final model gives an expected estimate for the SR as well as the distribution about this estimate.

Keywords: stutter; stutter ratio; mixed DNA profiles; LtDNA; DNA thresholds.

8.1 Introduction

The forensic analysis of DNA is often undertaken by the polymerase chain reaction (PCR) amplification of short tandem repeat (STR) loci followed by electrophoretic seperation. The AmpFℓSTR® Identifiler® PCR Amplification kit, currently in widespread use, has been developed to co-amplify and simultaneously detect 15 internationally recognised tetranucleotide STRs [1]. While the multiplexing of STR loci allows for greater discrimination between forensic samples, some compromise has to be made with respect to the amplification conditions of each locus. One consequence of both the PCR process, and the compromises inherent in multiplexing, is that the *Taq* enzyme responsible for DNA replication can miscopy. The most prevalent miscopy results in a loss of one complete repeat sequence. This is referred to as back stuttering or N-4 stuttering and is the subject of this research. For simplicity we will refer to this as stutter although strictly stutter would include *N+4* and *N-8* variants.

The size of a stutter peak is often characterised by stutter ratio, *SR*. In the following sections we define the stutter ratio as:

$$SR = \frac{\phi_S}{\phi_A} \tag{1}$$

where ϕ_S refers to the height or area (hereafter height) of the stutter peak, and ϕ_A refers to the height of the parent allele. Typically a threshold, T, is set for SR. If a peak is above this threshold, then it is designated as allelic. If it is below this threshold, then it could be stutter.

There is a known relationship between stutter and the number of repeat units in an allele whereby stutter increases as repeat unit number increases [2-4]. However if the allele contains several repeat sequences interrupted with a conserved, or non-consensus, segment, then the relationship appears to be dominated by the longest uninterrupted repeat sequence (LUS) [2, 3, 5]. Brookes et al. [6] defined the longest uninterrupted sequence as the longest stretch of basic repeat motifs within an allele.

It is known that the expected stutter ratio for allele *i*, $E[SR_i]$, varies between and within loci [6] and with cycle number [7]. There is variation around this expectation (var*SR*).

There is a belief that $E[SR_i]$ is greater for samples at low template levels. However, any effect is unlikely to appear in the expected stutter ratio $E[SR_i]$ since additional template is likely to simply act to make both parent and stutter peak larger preserving the ratio. But an effect would be expected and has been predicted in var*SR* [7]. The belief that $E[SR_i]$ is greater at low template levels is more likely due to low template DNA (LtDNA) being analysed using an increase in PCR cycle number (LCN).

In a more formal statistical framework we call SR a random variable and we describe its behaviour with a probability density function. In this paper we attempt to model the behaviour of SR using a family of well-known probability density functions. A standard approach to modelling data is to choose a parametric family of distributions which might describe the observed behaviour, and then to "fit" the distribution using data to estimate the parameters that control the family behaviour. Many families of probability distributions are parameterised by their mean, and occasionally by their variance. Therefore if we can use data to estimate the mean and variance of SR, then we may be able to model SR using a family of distributions that are representative the observed behaviour.

The interest in understanding the causes of stutter and the variability in stutter product is not merely an academic pursuit. There is considerable application of such understanding in both the interpretation of mixed source samples, and even in determining if a sample is mixed

source at all. Modern probabilistic methods for the interpretation of mixed or LtDNA profiles use probability models that could be informed by empirical studies of stutter. Despite considerable work in the field, we are not aware of any comprehensive evaluation of the drivers behind SR. Therefore, in this work, we use standard linear regression as a screening tool to determine the relationship (if any) between several possible explanatory variables and SR. Because the true template level is not known for any given sample we use the parent peak height ϕ_i where i is the allele designation, as a proxy for template. These variables investigated are; *LUS*, A-T content, locus, ϕ_i , and the number of repeat units. Since the goal of this work is to produce models for casework use it is necessary to constrain the possible explanatory variables to those that will be available in casework. Bayesian modelling techniques may then be used to estimate the model parameters of the distribution of SR using those variables deemed to be statistically significant. The final model allows us to determine an expected value for SR for allele, i, $E[SR_i]$, and to make probabilistic statements about the likely range of values one would expect to see.

8.2 Data preparation

A criticism of variability studies is that data sets are constructed using pristine DNA [8]. The criticism is that this practice does not adequately mimic casework conditions. This led Bright et. al., to investigate the difference between pristine and casework samples [9]. Although that study found little difference between pristine and casework samples it is still likely that the best surrogate for casework data is indeed casework data as long as the uncertainty in the source of the sample is accepted.

Of course the true source of a casework sample is never known. However in some circumstances a reasonable assessment may be made. If the sample is apparently single source and "matches" a reference sample then there is some reason to suppose that this is the source. Such an assumption is, of course, completely inappropriate in casework but may be permissible in research.

Profiles that had been assigned as single source in casework, of varying quality and from a variety of sample types including saliva stains and cigarette butts were examined for the presence of stutter peaks. The DNA from these samples had been previously extracted using an organic [10], or DNA IQ^{TM} extraction method [11], depending on sample type.

All samples were quantified prior to amplification using Applied Biosystems Quantifiler® Human Quantification Kits on a Applied Biosystems 7500 according to the manufacturer's instructions [12]. A target of 1ng was amplified with Applied Biosystems Identifiler® Kits according to the manufacturer's instructions in a 9700 thermal cycler with silver block. Amplified products were separated on Applied Biosystems 3130xl Genetic Analysers and analysis of DNA profiles undertaken using Applied Biosystems GeneMapper® ID version 3.2.1 using the panels and bins provided.

A peak detection threshold of 30 RFU was used in the data preparation in order to incorporate peaks at a low level. This was to maximise the detection of stutter peaks. Consider an allelic peak of height 500 RFU. It is likely that it could produce stutter peaks below 50 RFU which would not be detected. This would lead to a considerable missing data problem with all small stutters reported simply as "not detected." Examination of our baseline noise suggests that measurements down to 30 RFU and even lower could be sustained, at least for research. Even then considerable numbers of allelic peaks had undetected stutters. These were inserted at 15 RFU since "undetected" means anywhere between 0 and 30 RFU.

Heterozygous alleles one repeat unit length apart were removed from the data set to avoid confusion surrounding the true height of the smaller allele with the additional *N-4* stutter peak of the larger allele. 7,771 possible stutter peaks were identified from the re-analysed Identifiler® DNA profiles from a large spread of alleles.

The A-T content was determined from STR sequences taken from STRBase, the STR DNA Internet Database [13]. All STRs within the Identifiler® multiplex have a 75% A-T content except for D19S433 and D2S1338 which are 50%.

The longest uninterrupted sequence (*LUS*) was determined from STR sequences posted on STRBase. Three categories of STRs have been identified based on their repeat structure; simple, compound and complex. Simple repeats contain core sequences identical in sequence and length. Compound repeats contain two or more adjacent simple repeats. Complex repeats contain several repeat blocks of variable length with variable intervening sequences. In a minority of cases there were several sequences for the same allele designation. In such cases the average value of *LUS* has been used. This ambiguity is likely to inflate the variance for certain alleles. It would be advantageous to know the true sequence of every allele in order to obtain the true *LUS* but this knowledge is currently not available in routine casework. When

setting up to model casework the greater uncertainty in sequence must be accepted and is correctly reflected in a greater variation in *SR*.

The number of repeat units was taken as the allele designation. In simple STRs, the repeat units and the LUS are equivalent, but for compound and complex STRS the repeat unit length and LUS can differ. For example, at locus D19S433 allele 13 is a compound STR which has a repeat unit length of 13 but a LUS of 11. LUS is always smaller than repeat unit length if not equivalent.

The SR_i for each allele was calculated using Equation (1) and statistical analyses were carried out using the statistical package R [14]. Further information on some of the statistical analyses used in this paper can be found in Curran [15].

8.3 Exploratory data analysis and modelling

We are interested in modelling both the mean $E[SR_i]$ and the variance (varSR_i) of stutter ratio.

The candidate explanatory variables; LUS, A-T content, ϕ_i, locus and number of repeat units were investigated for their effect on $E[SR_i]$ and varSR_i using multiple regression. The main effects of each variable, as well as a term for the interaction between locus and LUS were included in the initial model. The interaction between locus and LUS was included in order to model the observed phenomenon where stutter appears to behave differently at different loci with respect to LUS.

Exploratory analysis suggests that the predominant effects on *SR* are *LUS* (Figure 8.1), locus and a *LUS*-locus interaction.

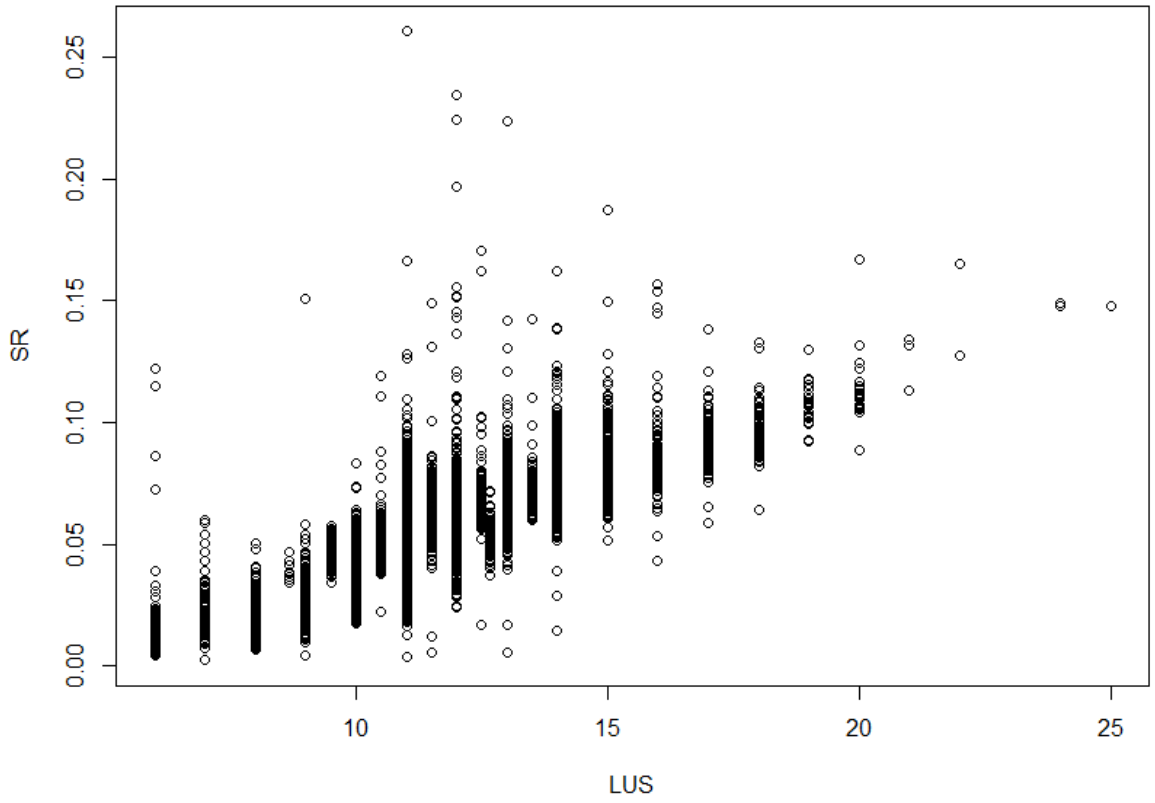


Figure 8.1: A scatter plot showing the relationship between LUS and stutter

We found an unexpected effect of parent peak height, ϕ_i, on $E[SR_i]$ but the coefficient is small 1.3x10^{-3} suggesting that omitting this explanatory variable will not significantly change the prediction. Note that the coefficient is positive. This is in opposition to any expectation that LtDNA samples stutter more.

The number of repeat units added very little extra to the model when LUS was included. Although this term was statistically significant ($P = 8.3\times10^{-15} \approx 0$), the coefficient was also small (2.4×10^{-3}), which, when combined with the relatively small repeat unit numbers (6 – 34.5), means there would be little effect on the $E[SR_i]$. Therefore number of repeat units was not included in the model.

The two loci with low A-T base content do not have exceptional β_{l0} coefficients. They are the 3rd and 7th largest coefficients in the multiplex. This would suggest that we have no evidence in this dataset for an effect of A-T content. We also note that A-T content is

completely predicted by locus and hence, if locus is present in the model AT content is redundant.

To examine those factors affecting varSR_i the residuals of the fitted model were regressed against the height of the parent peaks. There is a significant effect of height on the variance of stutter ($p = 7.16e-07$). However, again the coefficient is small ($2.78e-08$) which means that even when the height of the parent peak is in the 1000's there will be minimal effect on varSR_i.

8.4 Model construction

The data analysis in the preceding section was used to select a linear model which describes the behaviour of the expected value (or mean) of SR_i with respect to the locus, l, and LUS. This model is given in Equation (2).

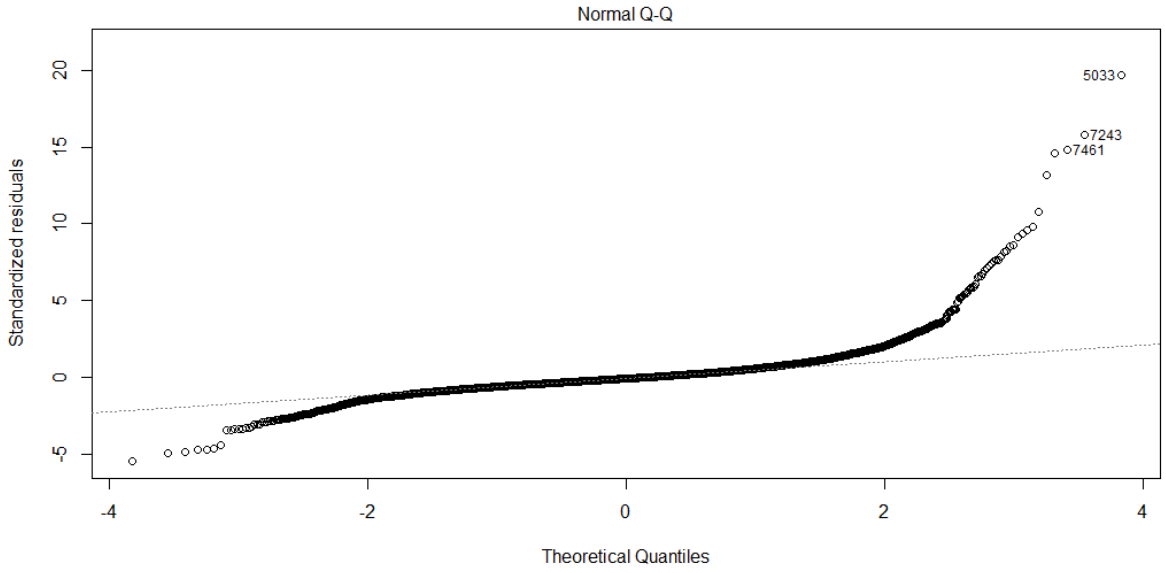
$$E[SR_i] = \beta_{l0} + \beta_{l1} LUS_i \quad (2)$$

Traditional regression models assume that the data is normally distributed with a mean described by a linear function and with constant scatter (variance) about the mean. This assumption is often checked by plotting the ordered residuals against the quantiles of a standard normal distribution. This plot is called a normal quantile-quantile (Q-Q) plot of the residuals. If the normality assumption holds, then the points on a normal Q-Q plot will follow a straight line whose slope provides an estimate of the standard deviation about the line. Strong curvature in one direction, or the other, is evidence of skewness. Sigmoidal shapes usually mean that the data has shorter or longer tails than a normal distribution. This behaviour is common if the mechanism generating it, is more variable than can be described by a normal distribution - a phenomenon often referred to as *heavy tails*. A normal Q-Q Plot of the residuals from this model is given Figure 8.2 A. The plot shows a large deviation from a straight line in the upper end of the plot, indicating that the tail on the right hand side of the distribution is longer than expected. We attempted to remedy this by taking the logarithms of the data (Equation (3)), however this increased the length of the tails on both sides of the distribution. Figure 8.2 shows the normal Q-Q plots of the residuals from both the normal (A) and log normal (B) models. As the tails of the distribution contain data of interest (peaks that could be deemed either allelic or stutter) we changed the model from normal, and log normal, to gamma. The gamma family of distributions are useful for modelling data which can take

values from zero to infinity, have a single mode, and which exhibit a degree of skewness to the right. The log normal family of distributions also has these properties, but the gamma family has a heavier right tail.

$$E\left[\log(SR_i)\right] = \beta_{l0} + \beta_{l1} LUS_i \tag{3}$$

A



B

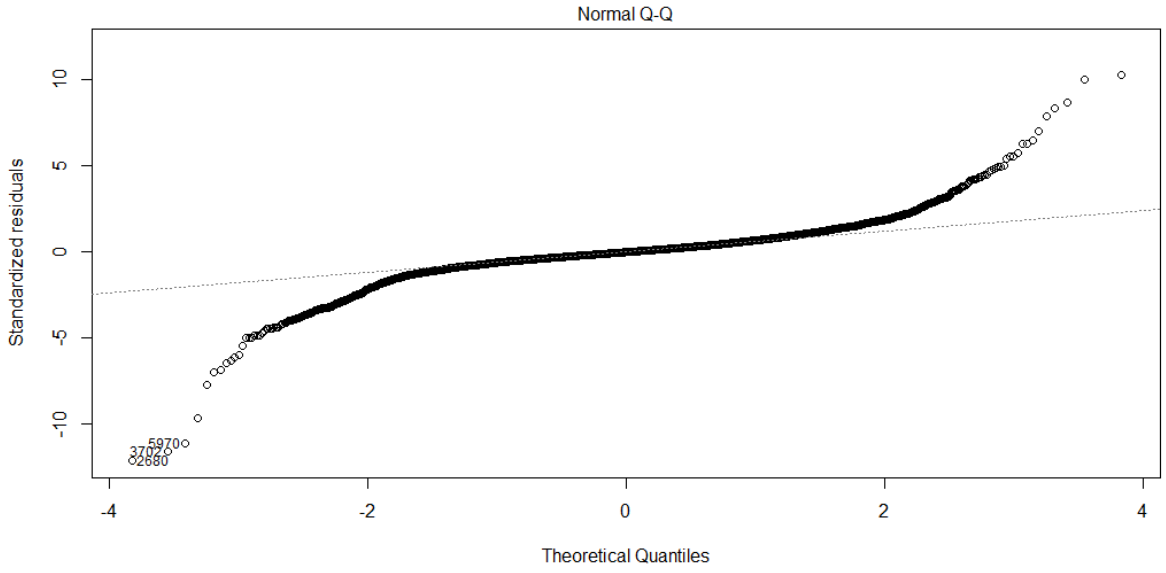


Figure 8.2: A: normal Q-Q plot showing the fit of the residuals from the model to the normal distribution and B: the residuals from the logarithms of the data (log normal)

The gamma family of distributions is a family of continuous probability density functions described by two parameters; shape α and rate K where $\alpha, \mathrm{K} > 0$. Sometimes scale, the reciprocal of rate, is used instead. The mean and variance of a gamma random variable X with shape α and rate K are given by

$$E[X] = \mu_X = \frac{\alpha}{\mathrm{K}} \text{ and } Var[X] = \sigma_X^2 = \frac{\alpha}{\mathrm{K}^2}$$

Therefore, if we have estimates $\hat{\mu}_X$ of μ_X, and $\hat{\sigma}_X^2$ of σ_X^2, then we can obtain estimates of α and K from

$$\hat{\alpha} = \left(\frac{\hat{\mu}_X}{\hat{\sigma}_X}\right)^2 \text{ and } \hat{\mathrm{K}} = \frac{\hat{\mu}_X}{\hat{\sigma}_X^2}$$

This mapping allows us to use the traditional specification of a mean model and a variance model whilst still using a gamma family of distributions which are not usually described by a mean and variance.

Therefore, in this paper, we model SR_i as a gamma distributed random variable, where the expected (mean) behaviour of SR_i is a strictly positive linear function of locus and LUS, and the variance is constant. We express this statistically as

$$SR_i \sim \Gamma\left(\alpha_i, \mathrm{K}_i\right)$$

$$E[SR_i] = \max\left(0, \beta_{l0} + \beta_{l1} LUS\right)$$

$$Var[SR_i] = \sigma^2$$

The parameters of this model are β_{l0}, β_{l1} and σ^2. The details regarding the fitting of this model are contained in the appendix.

8.5 Model assessment

A plot of the quantiles of SR from the observed data against the quantiles simulated from the model showed that there was a longer observed tail in the observed data than the gamma model predicts (data not shown).

We selected a cut-off point by eye ($SR \approx 0.12$) where the observed values started to visibly depart from the straight line. Using this cut-off we identified 97 observations that, according to the model, could be described as having unusually large stutters. The sample numbers from these observations were used to retrieve the original epgs. We randomly selected 16 of these epgs for investigation. Thirteen samples exhibited pull up, were degraded samples or low level mixtures. The remaining three had no detectable cause and presumably were simply large stutters. The two low level mixtures that were identified both had more than one "stutter" that did not fit the model. This suggests that such an approach as outlined here may assist in the detection of very low trace contributors. Recall that all of these 16 samples had been passed by human operator as single source and all the peaks assigned as stutter. These peaks were not even noticed as suspiciously large without the application of this model.

8.6 Using the model

This model returns the probability density of an observed stutter ratio for a given value of LUS for the parent allele, and the locus l. This information can be used to assess the "plausibility" of observing a particular stutter ratio. This could in turn be used in a continuous interpretation model or to decide whether a peak is stutter or allelic.

For example, consider a peak at vWA of height 1400 RFU and a peak in the stutter position of 70 RFU, so that the stutter ratio is 0.05. The parent peak has a LUS of 11. The predicted mean and variance can be calculated using the coefficients obtained from the gamma model (Appendix 1). In this example the predicted mean is 0.045. A 95% credible interval for the stutter ratio is $(0.0279, 0.0671)$. Our observed ratio lies very close to the middle of this interval (0.0475).

8.7 Discussion

This research had two primary aims; to determine the main variables affecting stutter ratio, and to then use the variables identified to build a model that could probabilistically evaluate observed stutter ratios. Data was collected that allowed us to model the distribution of SR, and through data analysis we have identified what we regard as a satisfactory model.

Both LUS and locus were found have a significant effect on the mean of SR. Stutter was found to be more likely as the uninterrupted sequence in the allele increased in length. That is, as LUS increased SR increased. There was also significant differences in the mean of SR between loci, even for the same value of LUS. This observed locus effect could be due to a number of factors and could be an area of future research for those interested in creating commercial STR kits.

The number of repeat units in the parent allele had a significant but small additional effect on the mean of SR in the regression analysis if LUS was accounted for. Given that the relationship appears to be dominated by locus and LUS, the number of repeat units was not included in the model.

Bayesian modelling techniques were used to build a model that used the variables; locus and LUS (with an interaction factor), to give an expected estimate for stutter ratio as well as predicted upper and lower quantiles within which stutter ratio would be expected to fall. The normal, log normal and gamma distributions were trialled with respect to modelling the stutter ratio. It was decided that due to the right handed skew of the data, the gamma distribution provided the best fit.

This model is a step towards a more probabilistic approach to identifying stutter. We intend this model to be used in a semi-continuous or continuous modelling system for the interpretation of STR DNA profile data. Threshold based interpretation is becoming outdated with the analysis of more complex mixtures and LtDNA samples whose peaks may not reach conventional thresholds. There is a need to move towards an interpretation model that can deal with these types of samples.

Appendix 1

Model fitting details

We need to specify priors in order to estimate these parameters in the Bayesian framework. The priors reflect our knowledge about these parameters before we observe the data. We will take a position of no prior information and therefore choose prior distributions that are very diffuse, or uninformative. By this we mean that we are stating that the true values could be "almost anything". We will also choose conjugate priors for mathematical and computational simplicity. We do not believe that such choices, although subjective, will unduly influence the posterior distributions. Therefore, for completeness, our priors are

$$\beta_{l0}, \beta_{l1} \sim N\left(0, 10^6\right), m = 1, \ldots, 15$$

$$\frac{1}{\sigma^2} \sim \Gamma\left(10^{-3}, 10^{-3}\right)$$

The package BRugs, an interface to the OpenBUGS programme, was used to fit our model in R [16]. BRugs is a statistical package that allows the user to fit Bayesian models using Markov chain Monte Carlo (MCMC) techniques.

The MCMC sampling scheme produces correlated samples from the posterior distributions of the parameters. The samples are also correlated with the initial choices for the parameter values, but this reduces over time. There is correlation between successive values. Therefore, we use a "burn-in" period to make sure that the samples have moved a sufficient distance from the initial parameter choices. This involves discarding the first n samples. In this paper we discarded the first $n = 10,000$ samples. The correlation between successive samples can be reduced by "thinning the chain". This process involves selecting every k^{th} sampled value. We took every 50^{th} sample from a run of 500,000 to produce a sample of size 10,000 from the posterior distribution of each parameter. We chose a relatively big number for the thinning process as the successive observations were highly correlated. Taking every 50^{th} sample corrected this. The high degree of correlation was induced by the requirement that the mean be positive. That is, $E[SR_i] = \max\left(0, \beta_{l0} + \beta_{l1} LUS\right)$. This meant that many of the

proposals were not accepted, and hence the chains had a tendency to remain in the same place for a long time.

Parameter estimates and summary statistics

Bayesian estimation procedures yield samples from the posterior distribution(s) of the parameter(s) in the model given the data. We list here the posterior mean, median, and the 2.5^{th} and 97.5^{th} percentiles which provide a 95% credible interval for each parameter.

Table 8.1: Summary statistics from the posterior distributions of β_{l0} by locus

Locus	2.50%	Mean	Median	97.50%
CSF1PO	-0.045	-0.037	-0.037	-0.029
D13S317	-0.050	-0.047	-0.047	-0.043
D16S539	-0.050	-0.044	-0.044	-0.038
D18S51	-0.043	-0.038	-0.038	-0.032
D19S433	-0.044	-0.037	-0.037	-0.030
D21S11	0.004	0.012	0.012	0.020
D2S1338	-0.019	-0.011	-0.012	-0.004
D3S1358	-0.053	-0.043	-0.043	-0.033
D5S818	-0.045	-0.038	-0.038	-0.030
D7S820	-0.046	-0.041	-0.041	-0.035
D8S1179	-0.015	-0.007	-0.007	0.001
FGA	-0.040	-0.034	-0.034	-0.028
TH01	-0.023	-0.018	-0.018	-0.014
TPOX	-0.021	-0.016	-0.016	-0.011
vWA	-0.117	-0.108	-0.108	-0.098

Table 8.2: Summary statistics from the posterior distributions of β_{l1} by locus

Locus	2.50%	Mean	Median	97.50%
CSF1PO	0.0064	0.0071	0.0071	0.0078
D13S317	0.0078	0.0082	0.0082	0.0085
D16S539	0.0078	0.0083	0.0083	0.0089
D18S51	0.0073	0.0076	0.0076	0.0080
D19S433	0.0086	0.0092	0.0092	0.0098
D21S11	0.0031	0.0038	0.0038	0.0046
D2S1338	0.0059	0.0065	0.0065	0.0071
D3S1358	0.0081	0.0089	0.0089	0.0097
D5S818	0.0069	0.0076	0.0076	0.0082
D7S820	0.0074	0.0079	0.0080	0.0085
D8S1179	0.0046	0.0053	0.0053	0.0060
FGA	0.0068	0.0072	0.0072	0.0075
TH01	0.0051	0.0058	0.0058	0.0065
TPOX	0.0039	0.0044	0.0044	0.0049
vWA	0.0131	0.0139	0.0139	0.0147

Table 8. 3: Summary statistics from the posterior distribution of σ

	2.50%	Mean	Median	97.50%
σ	0.0098	0.0101	0.0100	0.0102

References

1. Collins, P., et al., Validation of the AmpFℓSTR® Identifiler® PCR Amplification kit, in 11th International Symposium on Human Identification 2000, Promega Corporation.

2. Walsh, P.S., N.J. Fildes, and R. Reynolds, Sequence analysis and characterisation of stutter products at the tetranucleotide repeat locus vWA. Nucleic Acids Research, 1996. 24(14): p. 2807-2812.

3. Klintschar, M. and P. Wiegand, Polymerase slippage in relation to the uniformity of tetrameric repeat stretches. Forensic Science International, 2003. 135(2): p. 163-166.

4. Applied Biosystems., AmpFℓSTR® SGM Plus PCR Amplification Kit: User's Manual, 2005.

5. Lazaruk, K., et al., Sequence variation in humans and other primates at six short tandem repeat loci used in forensic identity testing. Forensic Science International, 2001. 119(1): p. 1-10.

6. Brookes, C., et al., Characterising stutter in forensic STR multiplexes. Forensic Science International: Genetics, 2012. 6(1): p. 58-63.

7. Weusten, J. and J. Herbergs, A stochastic model of the processes in PCR based amplification of STR DNA in forensic applications. Forensic Science International: Genetics, 2012. 6: p. 17-25.

8. Budowle, B., A.J. Eisenberg, and A. van_Daal, Validity of Low Copy Number Typing and Applications to Forensic Science. Croatian Medical Journal, 2009. 50: p. 207-217.

9. Bright, J.A., et al., A comparison of stochastic variation in mixed and unmixed casework and synthetic samples. Forensic Science International: Genetics, 2012. 6(2): p. 180-184.

10. Bright, J.A. and S.F. Petricevic, Recovery of trace DNA and its application to DNA profiling of shoe insoles. Forensic Science International, 2004. 145(1): p. 7-12.

11. Promega Corporation, DNA IQTM System - Small Sample Casework Protocol, 2002: Madison WI, USA.

12. Green, R.L., et al., Developmental validation of the QuantifilerTM real time PCR kits for the quantification of human nuclear DNA samples. Journal of Forensic Sciences, 2005. 50: p. 1-17.

13. Ruitberg, C.M., D.J. Reeder, and J.M. Butler, STRBase: a short tandem repeat DNA database for the human identity testing community. Nucleic Acids Research, 2001. 29(1): p. 320-322.

14. R Team: A language and environment for statistical computing. in R Foundation for tatistical Computing. 2012. Vienna.

15. Curran, J.M., Introduction to Data Analysis with R for Forensic Scientists 2011, Boca Raton, FL,: CRC Press.

16. Thomas, A., et al., Making BUGS open. R News, 2006. 6(1): p. 12-17.

Supplementary material (BUGS model specification)

```
model{
    for(i in 1:N){
        stutter[i] ~ dgamma(shape[i], rate[i])
        log.mu[i] <- b0[marker[i]] + b1*LUS[i]

        mu[i] <- exp(log.mu[i])

        shape[i] <- mu[i]*mu[i]/(sigma*sigma)
        scale[i] <- sigma*sigma/mu[i]
        rate[i] <- 1/scale[i]
    }

    for(j in 1:15){
     b0[j]~dnorm(0, 0.000001)
    }

    b1~dnorm(0, 0.000001)
    tau ~ dgamma(0.001,0.001)
    sigma <- 1/sqrt(tau)

    ## pred code
    pred.log.mu <- b[15] + b1*11
    pred.mu <- exp(pred.log.mu)
    pred.shape <- pred.mu*pred.mu/(sigma*sigma)
    pred.scale <- sigma*sigma/pred.mu

    post.mean <- pred.shape*pred.scale
    post.var <- post.mean*pred.scale
}
```

9: INVESTIGATING THE LOCUS EFFECT ON THE STUTTER RATIO

The investigation into the drivers of stutter identified that both the locus and allele *LUS* had a significant effect on the expected (mean) stutter ratio (*SR*) (Chapter 8). In this chapter we aim to identify the variables that contribute to the locus effect. Figure 9.1 shows the *SR* per locus. We can see that TH01 (μ = 0.022) has the lowest mean *SR* and D18S51 has the highest mean *SR* (μ = 0.080).

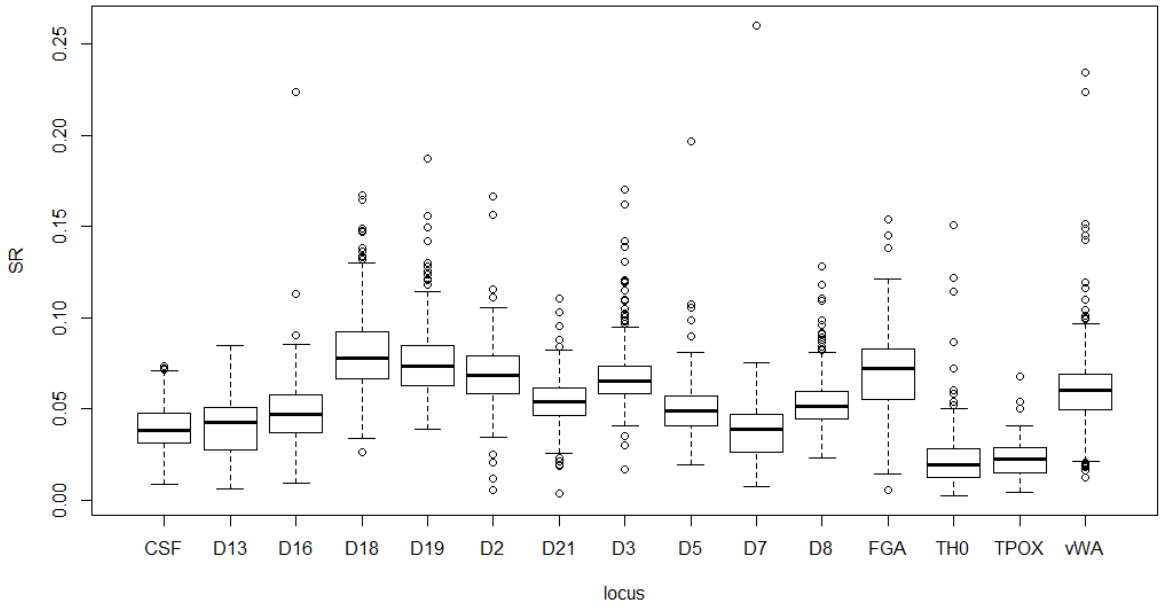


Figure 9.1: The *SR* per locus

A pairwise comparison of the mean *SR* values is shown in Table 9.1. We use this to identify multiple loci that stutter similarly.

Table 9.1: Loci that cannot be differentiated on mean *SR*

Locus 1	**Locus 2**	**P Value**
CSF1PO	D13S317	1.000
CSF1PO	D7S820	0.991
D13S317	D7S820	0.996
D16S539	D5S818	0.949
D21S11	D8S1179	0.998
D2S1338	D3S1358	0.947
D2S1338	FGA	0.996
FGA	D3S1358	0.230
TH01	TPOX	1.000

The examination of loci that had a similar mean *SR* indicated that the class of STR might affect stutter. STRs can be broken down into three main groups;

1) simple repeats contain core sequences identical in sequence and length,
2) compound repeats contain two or more adjacent simple repeats, and
3) complex repeats may contain several repeat blocks of variable repeat unit length with variable intervening sequences.

Table 9.2 lists the STR class, the STR sequence, the allele range present in the dataset, and the average *SR* for each locus in the data used in this research.

Table 9.2: STR type, sequence and allele range by locus.

Locus	STR class	Sequence	Allele Range in Dataset	SR
D16S539	simple	[GATA]	8-17	0.048
D18S51	simple	[AGAA]	10-25	0.080
D19S433	compound	[AAGG]/[TAGG]	11-17.5	0.076
D21S11	complex	[TCTA]/[TCTG]	26-34.5	0.054
D2S1338	compound	[TGCC]/[TTCC]	16-26	0.069
D8S1179	compound	[TCTA]/[TCTG]	8-17	0.053
FGA	complex	[CTTT]/[TTCC]	18-29	0.070
TH01	simple	[TCAT]	6-10	0.022
vWA	compound	[TCTA]/[TCTG]	14-20	0.060
D3S1358	compound	[TCTA]/[TCTG]	11-20	0.068
CSF1PO	simple	[AGAT]	8-15	0.040
D13S317	simple	[TATC]	6-16	0.040
D5S818	simple	[AGAT]	9-15	0.049
D7S820	simple	[GATA]	7-15	0.039
TPOX	simple	[AATG]	8-12	0.022

We used one way ANOVA to show that different STR classes have different mean stutter ratios (Figure 9.2). Simple repeats had the lowest average stutter ratio (0.044), followed by complex repeats (0.062). Compound repeats had the highest average stutter ratio (0.0650).

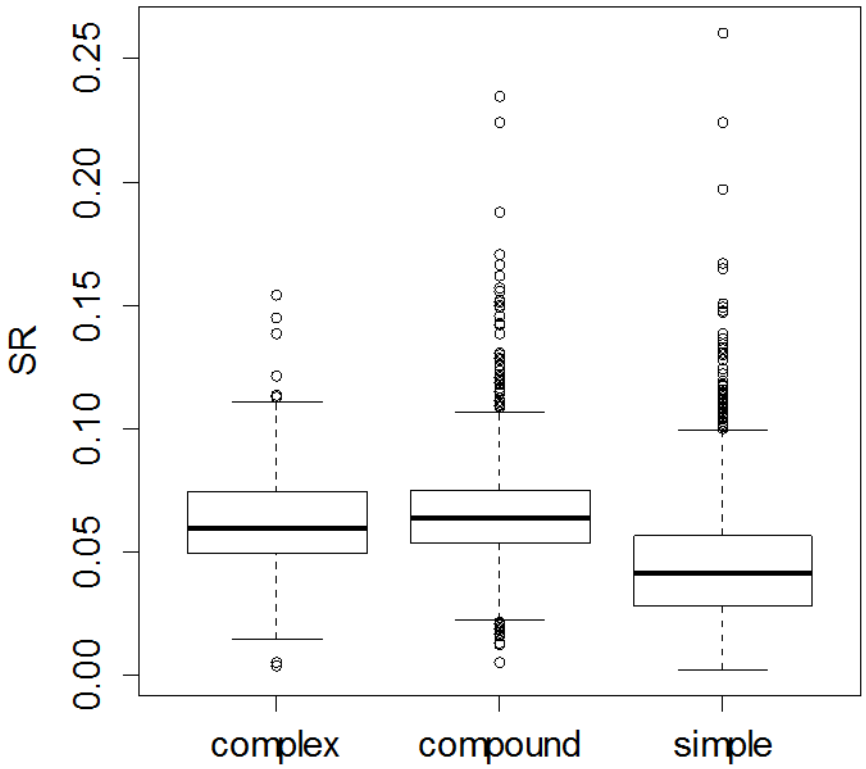


Figure 9.2: *SR* by STR class

SR was regressed against *LUS* and STR class with a *LUS*-STR class interaction factor using linear regression. The resulting adjusted R^2 value was 0.73 compared to the adjusted R^2 value of 0.78 of the *LUS*-locus model in Chapter 8 (and compared to an adjusted R^2 value of 0.65 for *SR* vs. *LUS* only). This indicates that STR class does contribute to the locus effect however; there is still some unexplained explanatory variable beyond STR class.

Past research suggests that the higher the GC content of a locus the more stable the locus is due to the additional hydrogen bonds (GC base pairs have three hydrogen bonds while AT base pairs have two hydrogen bonds) [1]. Within the AmpFℓSTR® Identifiler® multiplex, there are two loci with an AT content of 50% (D19S433 and D2S1338). The remaining 13 loci have an AT base pair content of 75%. Therefore the expected *SR* should be lower in loci with a higher GC content. However, in this research the two loci that contain the highest percentage GC base pairs have two of the highest *SR*s.

The adjusted R^2 value slightly increased to 0.74, when the *SR* was regressed against *LUS* and STR class with a *LUS*-STR class interaction factor and AT base pair content. This indicates that AT base pair content is not contributing significantly to the model if *LUS* and STR class are included. When the STR class is dropped from the model the adjusted R^2 value becomes 0.71, indicating that STR class is the better variable to retain in the model.

The repeat number, or allele designation, was also investigated with regard to its effect on the *SR*. However, because the length of the STR was already somewhat accounted for by the inclusion of the variable *LUS*, a new variable "Allele-*LUS*" was introduced into the dataset. This variable was created by taking the value of the allele designation and subtracting the *LUS* value. This value represents the residual repeat units left in the allele once the longest uninterrupted repeating sequence is removed. The resulting adjusted R^2 was 0.72, when the *SR* was regressed against *LUS* and Allele-*LUS*. When the *SR* was regressed versus *LUS* and STR class with a *LUS*-STR class interaction factor, AT base pair content and Allele-*LUS* the resulting adjusted R^2 was 0.74. This indicates that Allele-*LUS* was not contributing anything significant towards the locus effect.

The highest adjusted R^2 value that could be obtained when removing locus effect and adding variables thought to constitute locus effect was 0.74. This was the model that included STR class and AT base pair content. This indicates that although STR class and AT base pair content contribute to locus effect there is still some unexplained variation unaccounted for that is contained within locus effect.

As with any multiplex, a compromise has to be made within the PCR conditions for each locus. Due to this compromise it is likely that there will be slight behavioural differences between loci. These may manifest themselves in differing amplification efficiency of loci. So whilst endeavouring to understand the differences in the *SR* between loci is interesting from an academic and manufacturers' perspective, from a DNA analyst's perspective it is prudent to retain the original model which accounts for the loci effect as a whole.

9.1 D21S11

While investigating variables thought to contribute to the locus effect, it was discovered that the locus D21S11 did not exhibit the linear relationship between *SR* and *LUS* evident within the other loci.

Figure 9.3 shows the data for D21S11, a complex repeat (Table 9.2). This plot shows that there appears to be some *LUS* sequences that are stuttering less than expected (Table 9.3). For example, the expected *SR* for alleles with a *LUS* of 12.66 was 0.062, but the observed *SR* was 0.054. The sequences of the alleles present in the data set used (AmpFℓSTR® Identifiler®) were recorded using the STRbase D21S11 fact sheet [2], along with their reported *LUS* (Table 9.4).

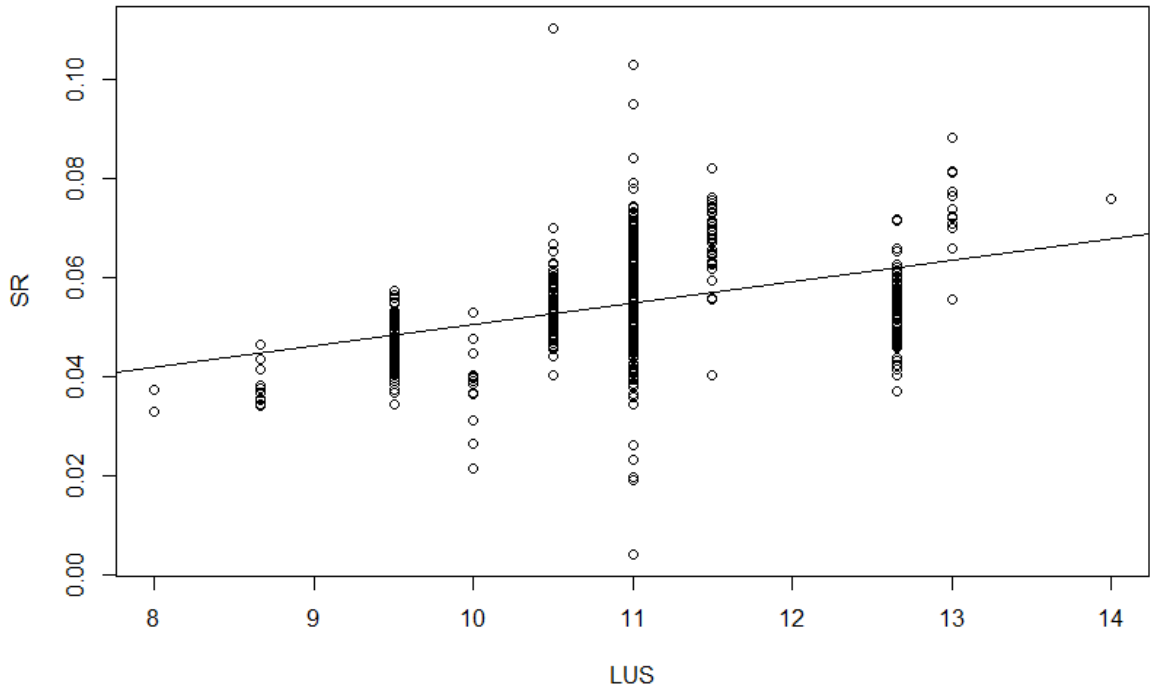


Figure 9.3: *SR* vs. *LUS* for D21S11

Table 9.3: Expected and Observed *SR* Values for D21S11 using *LUS*

LUS	Expected SR Mean	Observed SR Mean	Observed SR Median
8	0.042	0.035	0.035
8.66	0.045	0.038	0.037
9.5	0.048	0.046	0.046
10	0.051	0.038	0.040
10.5	0.053	0.054	0.054
11	0.055	0.057	0.058
11.5	0.057	0.067	0.068
12.66	0.062	0.054	0.054
13	0.063	0.074	0.073
14	0.068	0.076	0.076

It was noted that alleles 30 and 31.2 had the same recorded value of *LUS* (11) and therefore, theoretically, should stutter similarly. Figure 9.4 shows the *SR* of these two alleles. These two alleles stutter quite differently and the distinctive behaviour of the two alleles explains the large variability of the *SR* at *LUS* = 11 in Figure 9.3. The only notable difference between these two alleles is a TATCTA addition to the end of the 31.2 allele (Table 9.4). It is surprising to find that the longer allele has a lower *SR* as traditionally it has been thought that longer alleles had a higher *SR* [3].

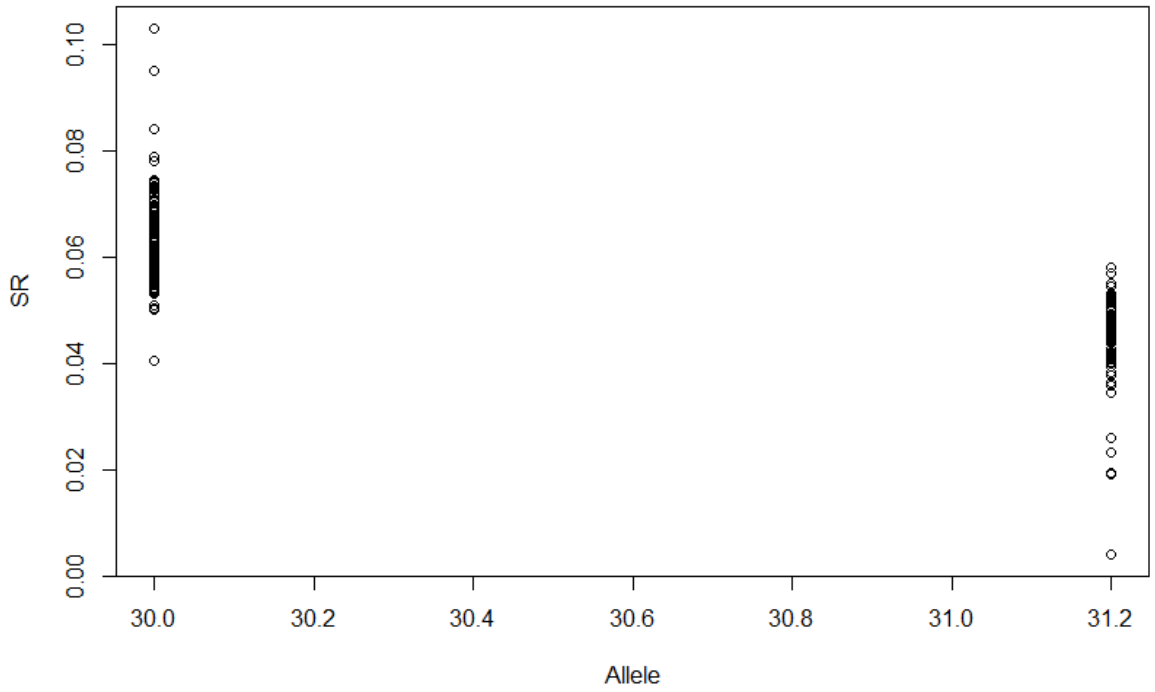


Figure 9.4: The *SR* of alleles 30 and 31.2 for D21S11 *LUS* 11

Allele – *LUS* was regressed against *SR* for the D21S11 data to see if there was any relationship present (Figure 9.5, the 0.5 designations represent 2 bp of a 4 bp repeat). No obvious relationship, that could be explained biologically, was evident between *SR* and the number of residual repeat units.

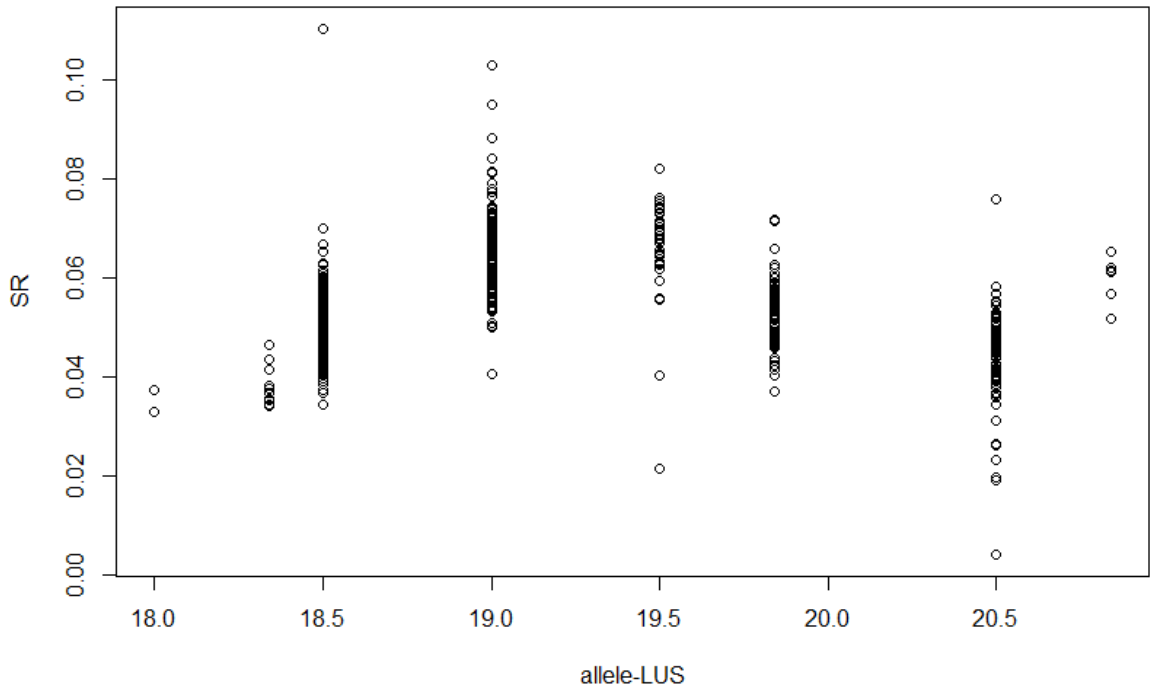


Figure 9.5: *SR* vs. Allele-*LUS* for D21S11

There are two observable relationships when the *SR* is regressed against allele designation (Figure 9.6). The alleles that have a 0.2 designation (i.e. the alleles with the TATCTA addition; 29.2, 30.2, 31.2, 32.2, 33.2, 34.2) stutter linearly with respect to each other, but they stutter less than they should compared to the alleles without the TATCTA addition. There appears to be a decrease in the observed *SR*, associated with the TATCTA addition, which is difficult to explain using our knowledge in this area. We speculate that these alleles have been mis-sequenced and their *LUS* is actually shorter than what they have been designated. It would be worthwhile to sequence these alleles in future work.

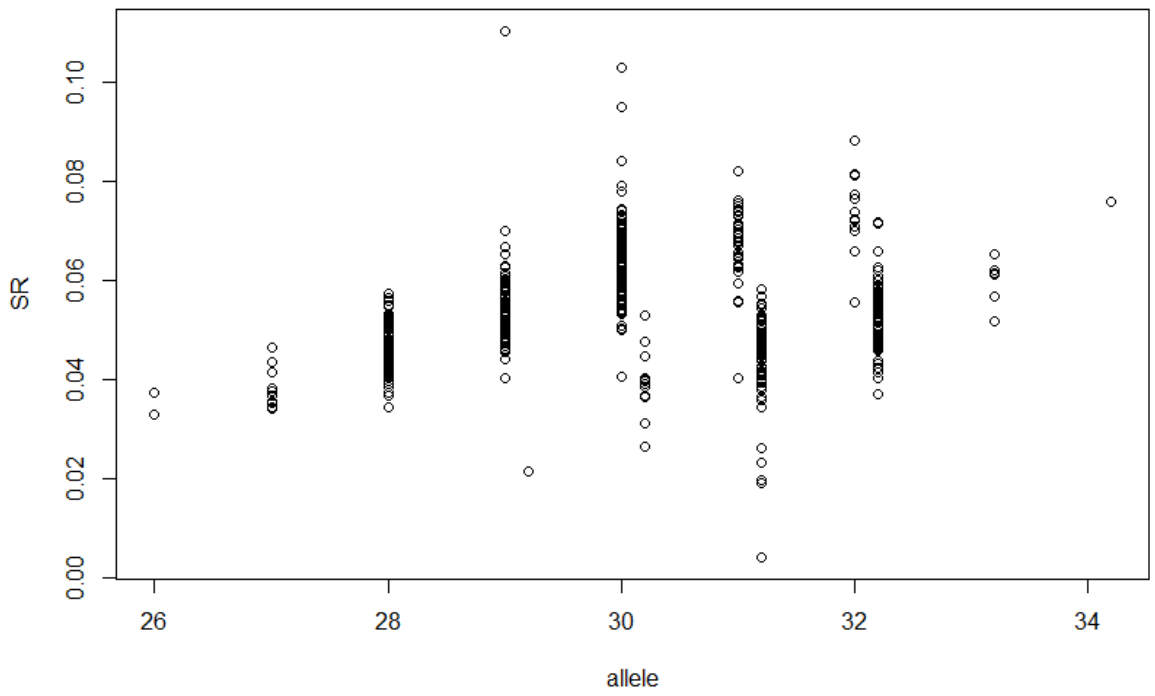


Figure 9.6: *SR* vs. allele for D21S11

9.2 D21S11: NGM™ SElect & AmpFℓSTR® MiniFiler™

Additional single source profiles of varying quality, and from a variety of sample types, were identified from casework. The DNA from these samples had been previously extracted either using an organic [4], or DNA IQ extraction method [5], depending on the sample type.

All samples were quantified prior to amplification using Applied Biosystems Quantifiler® Human Quantification Kits on a Applied Biosystems 7500 according to the manufacturer's instructions [6]. A target of 1 ng was amplified using either NGM™ SElect or AmpFℓSTR® MiniFiler™ kits according to the manufacturer's instructions in a 9700 thermal cycler with silver block. Amplified products were separated on Applied Biosystems 3130xl Genetic Analysers and analysis of DNA profiles undertaken using Applied Biosystems GeneMapper ID version 3.2.1 using the panels and bins provided.

A peak detection threshold of 50 RFU was used. 629 possible stutter peaks were identified in the AmpFℓSTR® MiniFiler™ dataset and 4646 possible stutter peaks were identified in the NGM™ SElect data set.

SR was plotted against *LUS* at locus D21S11 for each of the datasets (Figure 9.7 and Figure 9.8) to determine if the *SR* behaved as expected with respect to *LUS* using these kits. When looking at Figure 9.7, the AmpFℓSTR® MiniFiler™ data set, it is difficult to determine if the *SR* is behaving as expected. The circled points appear to be off-trend but there is a lack of data at this point to support this hypothesis.

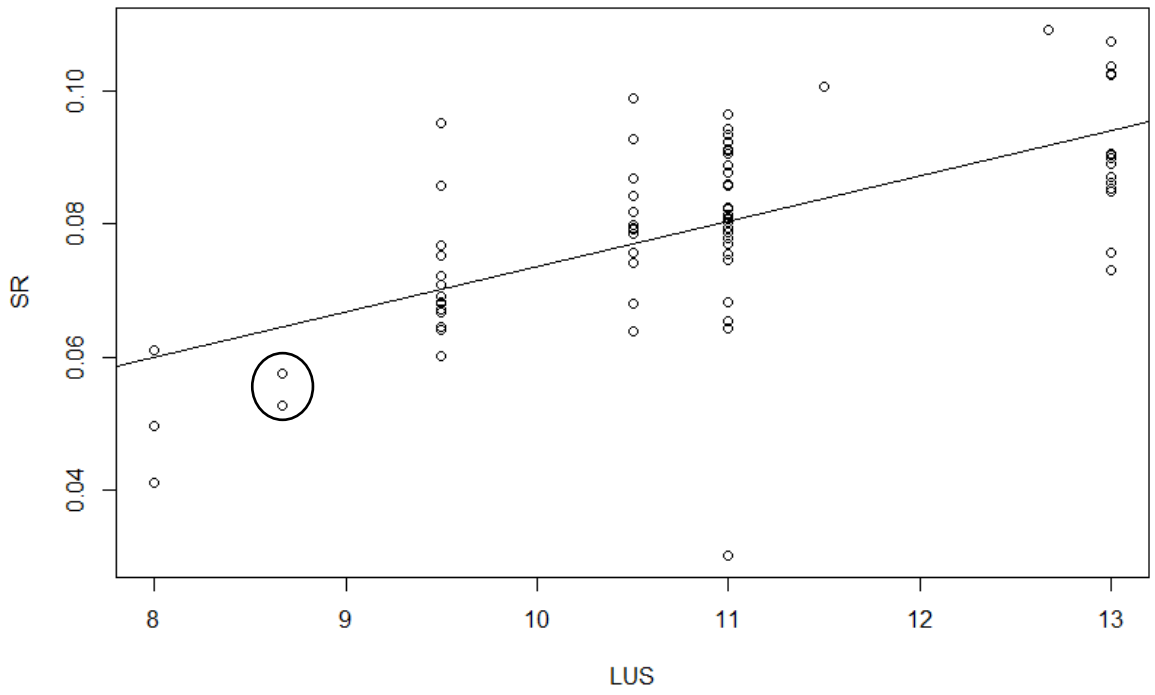


Figure 9.7: *SR* vs. *LUS* for D21S11 using AmpFℓSTR® MiniFiler™ data

Figure 9.8 shows that the alleles with a *LUS* of 11.5 in the NGM™ SElect dataset have a higher *SR* then expected with respect to the alleles with a *LUS* of 11. This reiterates what was seen in the AmpFℓSTR® Identifiler® data set. The peculiarities at the other *LUS* values as seen in the AmpFℓSTR® Identifiler® kit are not as obvious in the NGM™ SElect data set as at there is a lack of data at some *LUS* values (10, 13).

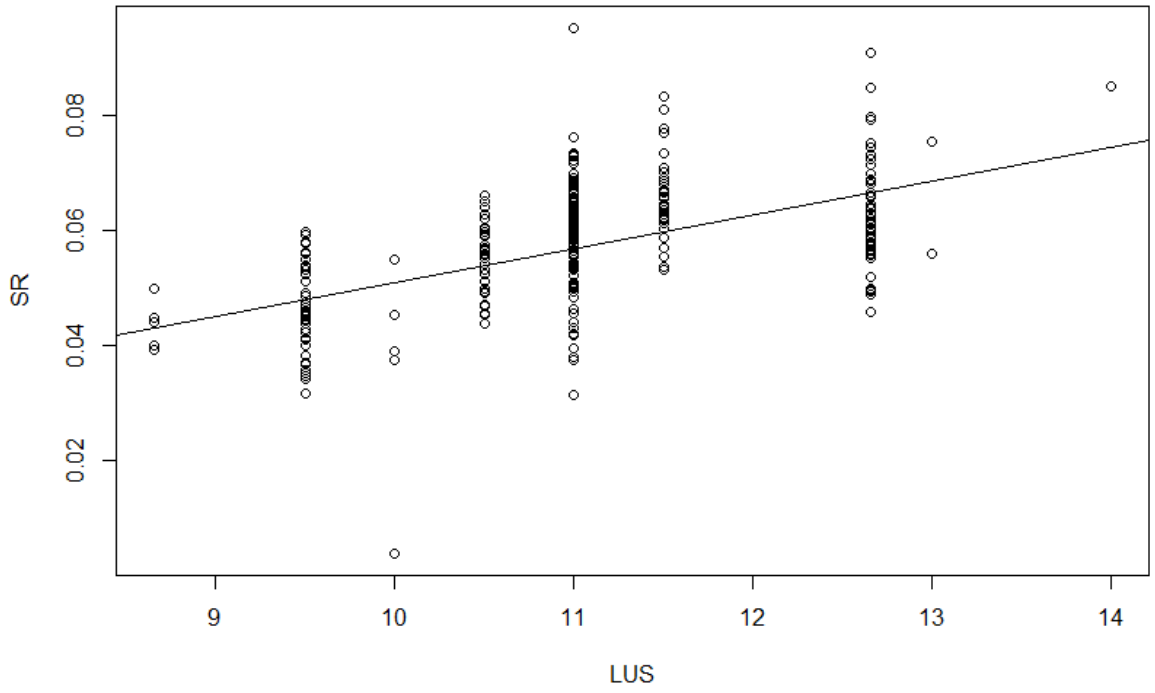


Figure 9.8: *SR* vs. *LUS* for D21S11 using NGM™ data

Plotting the D21S11 NGM™ SElect *SR* data by allele (Figure 9.9), allows the alleles with a *LUS* of 11 to be separated (30 and 31.2) and the difference in the observed *SR* between these two alleles with the same *LUS* is more obvious. These two alleles with a shared *LUS* in both the AmpFℓSTR® Identifiler® and NGM™ SElect data sets are evidence that this locus warrants further investigation. These alleles were also investigated using the AmpFℓSTR® MiniFiler™ kit. Allele 30 had an observed *SR* of 0.083 and allele 31.2 had an observed SR of 0.077 which suggests that AmpFℓSTR® MiniFiler™ data may have similar behaviour to AmpFℓSTR® Identifiler® and NGM™ SElect data. However, the lack of AmpFℓSTR® MiniFiler™ data made this assumption hard to prove.

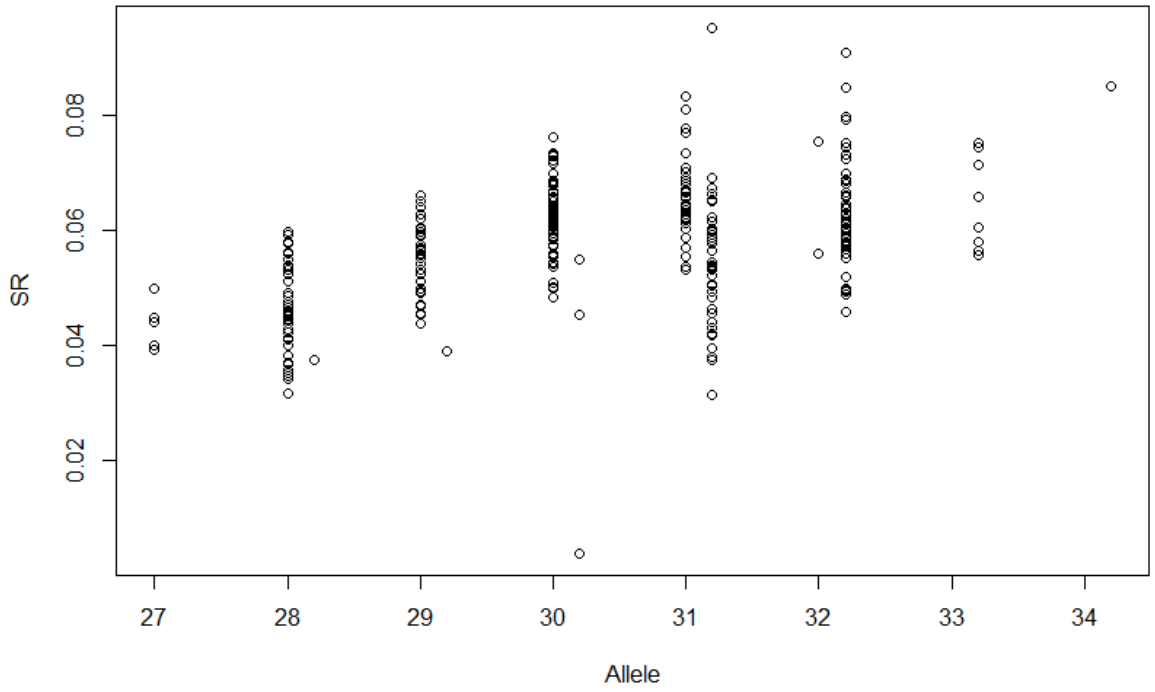


Figure 9.9: *SR* vs. Allele for D21S11 using NGM^{TM} SElect data

9.3 *SR* by PCR cycle number

Stutter peaks accrue with every PCR cycle. Given our belief that stutter has some molecular basis, as yet not fully understood, it is reasonable to suppose that there is some constant probability, *x*, that a given molecule will stutter at each cycle. This suggests that there is a probability of approximately *1-x* that the allele will copy correctly.

Let allelic template at cycle *t* be T_t, then the expected amount of template is:

$$E[T_t] = T_{t-1}(2-x)$$

and the expected stutter template (S_t) at cycle *t* is:

$$E[S_t] = S_{t-1}(2-x) + xT_{t-1}$$

Then the expected stutter ratio, (*SR*) is equal to:

$$E[SR_t] = \frac{S_t}{T_t} = \frac{S_{t-1}(2-x) + xT_{t-1}}{T_{t-1}(2-x)}$$

$$= SR_{t-1} + \frac{x}{(2-x)}$$

At $t = 0$ $SR = 0$, hence the expected SR is:

$$E[SR_t] = \frac{tx}{(2-x)}$$

This suggests that the SR increases linearly with cycle number t. If we know SR_t and t, then this equation may be solved for x, giving us a more fundamental parameter that is invariant with respect to cycle number.

Using the observed SR and LUS at D21S11 for the AmpFℓSTR® Identifiler® data (28 cycles) and the NGM™ SElect data (30 cycles), the approximate probability of stutter for each cycle of PCR could be ascertained using numerical optimisation techniques to calculate the minimum sum of the observed minus expected SR values. $1 - \mathrm{PCR}_{eff}$, where PCR_{eff} is the PCR efficiency, gives an approximate value for the probability of stutter (Table 9.4).

Table 9.4: Calculation of the expected *SR* per PCR cycle.

LUS	**Cycle number**	**Observed *SR***	**Expected *SR* per cycle**
12.66	28	0.0539	0.004044
	30	0.0633	

The probability of stutter at each PCR cycle can be calculated for each locus and each LUS value. This is novel as it has been observed that stutter varies depending on PCR cycle number, with SR increasing in samples that have been amplified using additional PCR cycles.

Assigning stutter in low template samples that have been amplified using LCN techniques (increasing PCR cycles from 28 up to 34) is problematic and has not yet been addressed in this research. However, using this method the expected SR can be calculated for a given locus and LUS at any cycle number. For example, at D21S11, LUS 12.66; the expected SR at 28 cycles is 0.0567. At 34 cycles this value increases to 0.0688.

This is a different and more fundamental way of thinking about stutter in forensic DNA profiles. However, this research only looked at one locus and two *LUS* values. A much larger investigation would need to be undertaken to take this research any further.

9.4 Conclusion

The main aim of this chapter was to identify variables that caused loci to stutter differently. The class of STR present at the locus was found to be significant. However, there was still some unexplained explanatory value to locus beyond STR class. The variables, AT base pair content and Allele-*LUS* were investigated but found to contribute little to the observed locus effect. It is likely that the locus effect includes a PCR effect and therefore can never be fully accounted for. It is prudent to retain the original model that accounts for the loci effect as a whole.

During the investigation into locus effect it was noted that D21S11, unlike the other loci, did not exhibit a linear relationship between *SR* and *LUS*. This observation warranted further investigation. It was discovered that there were two distinct linear relationships when the *SR* was plotted against allele (Figure 9.6). The alleles that have a 0.2 designation stutter linearly with regards to each other, but they stutter less than they should in regards to the alleles without the 0.2 designation. Some alleles within the NGM^{TM} SElect dataset also did not stutter as would be expected with regards to *LUS*. This is possibly due to the alleles having been labelled with an incorrect *LUS*/allele designation. The classification of D21S11 deserves further investigation.

Table 9.4: Alleles present in the D21S11 data set

Allele	Repeat structure	LUS	Average LUS
26	[TCTA]4 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]8	8	8
28	[TCTA]4 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]10	10	9.5
28'	[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]9	9	
30	[TCTA]6 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]11	11	11
30'	[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]11	11	
30"	[TCTA]4 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]12	12	
30"'	[TCTA]6 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]10	10	
30.2	[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]10 TA TCTA	10	10
30.2'	[TCTA]5 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]11 TA TCTA	11	
32	[TCTA]6 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]13	13	13
32'	[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]13	13	
32.2	[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]12 TA TCTA	12	12.66
32.2'	[TCTA]4 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]13 TA TCTA	13	
32.2"	[TCTA]5 [TCTG]6 [TCTA]2 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]13 TA TCTA	13	
34.2	[TCTA]4 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]14 TA TCTA	14	14
27	[TCTA]4 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]9	9	8.66
27'	[TCTA]6 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]8	8	
27"	[TCTA]5 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]9	9	
29	[TCTA]4 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]11	11	10.5

29'	[TCTA]6 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]10	10	
29.2	[TCTA]5 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]10 TA TCTA	10	10
31	[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]12	12	11.5
31'	[TCTA]6 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]12	12	
31"	[TCTA]6 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]11	11	
31'"	[TCTA]7 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]11	11	
31.2	[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]11 TA TCTA	11	11
33.2	[TCTA]5 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]13 TA TCTA	13	12.66
33.2'	[TCTA]6 [TCTG]5 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]13 TA TCTA	13	
33.2"	[TCTA]6 [TCTG]6 [TCTA]3 TA [TCTA]3 TCA [TCTA]2 TCCA TA [TCTA]12 TA TCTA	12	

References

1. Brookes, C.L., To Stutter or not to Stutter, Master of Science Thesis, 2006, University of Auckland.

2. STRbase. (http://www.cstl.nist.gov/strbase/str_D21S11.htm).

3. Walsh, P.S., N.J. Fildes, and R. Reynolds, Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. Nucleic Acids Research, 1996. 24(14): p. 2807-2812.

4. Bright, J.-A. and S.F. Petricevic, Recovery of trace DNA and its application to DNA profiling of shoe insoles. Forensic Science International, 2004. 145(1): p. 7-12.

5. Promega, DNA IQTM System - Small Sample Casework Protocol, 2002: Madison WI, USA.

6. Green RL, R.I., Boland C, Hennessy LK, Developmental validation of the quantifiler real-time PCR kits for the quantification of human nuclear DNA samples. Journal of Forensic Science, 2005. 50(4): p. 809-25.

10: DISCUSSION, CONCLUSION AND FUTURE RESEARCH

There can be little doubt that the advent of DNA typing constitutes one of the most important advances in forensic science of the 20th century. More than two decades have passed since Sir Alec Jeffrey's ground-breaking publication [1], and DNA typing is now routinely undertaken in forensic investigations worldwide.

The last two decades have seen DNA typing evolve from determining if a person of interest is excluded as the source of a biological sample to calculating the strength of evidence that the person of interest is the contributor of the biological sample, with the inclusion of statistical genetics. This is evident in the National Research Council's publication (NRC II) [2] in which it is said that the "ultimate purpose of DNA typing is to test the hypothesis that a particular person is the source of an item of biological evidence".

10.1 The interpretation of low template DNA

With the increasing sensitivity of the analysis techniques utilised in DNA typing, it is appropriate to discuss how far DNA interpretation can be pushed. The stochastic effects inherent in low template DNA profiles can complicate the interpretation of DNA typing, but the preceding chapters of this thesis have demonstrated that these factors can be modelled, and the probability of these events occurring, given the data, can be assessed.

Dropout, perhaps one of the most concerning manifestations of the analysis of smaller quantities of DNA was investigated in Chapters 5 and 6. Chapter 5 demonstrated how important the inclusion of a probability of dropout was on the calculation of the weight of evidence. Although assigning a "one" to a locus with potential dropout appears on the outset to be a neutral statistic in favour of neither the prosecution nor the defence, in reality any "missing" evidence should down weight the overall statistic. The probability that the data is "missing" should be assessed using the information available from the profile. The relative peak heights of the present data can indicate to the analyst the probability that the questioned allele has dropped out.

Chapter 6 investigated the application of a probability of dropout model. Models suggested by Tvedebrink et al. [3, 4] were compared to a third model that used the peak heights of the peaks present in a profile, conditioned on molecular weight, as an explanatory variable in the calculation of the probability of dropout using logistic regression. This chapter showed that the probability of dropout could be adequately modelled and implemented in a calculation of the evidential weight under the likelihood ratio method.

Heterozygous balance (*h*) was discussed in Chapter 7. It was found that the average (*h*) did not change significantly in regards to peak height. This means that the average (*h*) does not deteriorate at low template DNA levels as previously thought [5]. However, the variance of (*h*) was shown to decrease at a rate inversely proportional to the average peak height at a locus. The variance of (*h*) was most extreme at low template levels. Using statistical methods a model was built that could be used to give an estimate for expected (*h*) as well as an interval that (*h*) would be expected to fall within, with reasonable probability. However, the resulting model was complicated and it is expected that future work will not utilise (*h*) and will instead focus on the variability of peak heights [6].

Chapters 8 and 9 investigated stutter. Chapter 8 identified and modelled the drivers of the stutter ratio (*SR*). It was found that the longest uninterrupted sequence (*LUS*) and the locus had an effect on (*SR*). A linear model was designed which used *LUS* and locus to describe the behaviour of the expected *SR*. The *SR* was not affected by peak height and therefore did not behave differently at low template levels than in “conventional” profiles. The model that was built can be applied across all DNA profiles regardless of their classification as low template or conventional.

The *SR* was shown to be dependent on PCR cycle number. Profiles that were amplified using increased cycle number (such as in LCN analysis) showed increased stutter. Chapter 9 investigated calculating a PCR efficiency value, which could be used to give an approximate probability for the expected SR given any cycle number. This study only used the data from one locus (D21S11) and one *LUS* (12.66) and would need to be extended to be of any use in case work.

Chapter 9 also investigated the irregularity of D21S11. D21S11 did not exhibit the linear relationship between *SR* and *LUS* evident within the other loci. It is the opinion of the author

that D21S11 may have been mis-sequenced and the designated *LUS* is longer than it should be. It would be worthwhile to sequence these alleles in future work.

Chapter 2 demonstrated that the probabilistic assessment of these stochastic effects can only be undertaken using a method that can deal with the probability or likelihood of the event in a sensible and logical manner. The only method that can cope with an assessment of this type is the likelihood ratio. The conditioning step used in the likelihood ratio removes the need for the analyst to designate an inclusion or exclusion step early on in the interpretation and the likelihood of stochastic events can be accommodated within the formulation of the hypotheses. The formulation of the hypotheses and the manner in which they are then assessed also safeguards the evidence and the analyst from bias, as the probability of the evidence given an unlikely hypothesis will not be supported by the *LR*.

This is contrary to the belief of some authors who argue that bias is inherent in the interpretation of low template DNA. This argument stems from the belief that evidential profiles should be resolved in full before they are compared to a person of interest's profile to avoid the issue of reverse conditioning.

However, the issue of reverse conditioning is handled appropriately within the *LR* based method. For example, when assessing if a questioned peak could be stutter or allelic, the question becomes; "what is the probability of seeing this peak (at this height) if it is stutter?" If the evidence (epg) does not support the hypothesis that the peak is stutter then the resulting probability will reflect this. The evidence cannot be manipulated to "fit" the hypothesis if undertaken correctly.

The probabilistic models built within this research can be implemented under extensions to the binary models as discussed in Chapter 3. This is possibly the easiest implementation for forensic biologists to understand and requires a much more "hands on" approach than automated systems. However, the downfall of having a hands on approach is that the method becomes error prone. Some of the calculations are laborious and computationally intense as evident within Chapter 4 and in particular the *Q* model.

Binary models have served well for a number of years and the primary push to move away from this method of interpretation is that these models cannot deal with non-concordances. That is, these models cannot deal with a locus where at least one allele of the POI is not seen

in the profile. (Although Chapter 4 dealt with dropout under the binary model, these dropout events were not alleles relating to the POI and instead related to the hypothesis that alleles from a different contributor had dropped out).

This has resulted in the development of the semi-continuous model. Using the methodology of Buckleton and Gill [7], the probability of dropout (and other stochastic events such as drop in) can be included in the *LR* calculation. However, the semi-continuous model still does not make full use of all of the information available from the evidence. For example, Chapter 3 refers to an example containing an epg that comprises four peaks from a two person mixture. Using both the binary and semi-continuous models the combinations of alleles at this locus are deemed either possible (1) or not possible (0). In reality, given the peak heights of each allele there will be a genotype that is most supported by the data. This has led to the development of the continuous model.

The continuous model assigns a value to the probability of the profile genotype given the hypothesised genotype. These models have the potential to handle any type of non-concordance and may assess any number of replicates without pre-processing and the consequential loss of information. Many of the sometimes subjective decisions made by the analyst – such as the designation of peaks as alleles, the designation of stutter peaks and the possible genotype combinations may be removed and instead, assessed by the continuous model which makes use of the quantitative information present in the profile to assess the probability these events. It is the author's opinion that the forensic biology community should move towards utilising continuous models in their assessment of all DNA profiles.

Some of the issues associated with the interpretation of low template DNA have been identified and discussed prior to this work. However there has been apprehension within the forensic community to move towards interpretation models that can cope with the stochastic effects present in low template DNA. This is possibly due to the "fear of the unknown" in that there is belief that these models operate beyond the comprehension of biologists. That is a fear that is common with the automation or introduction of software within any domain. However, in this situation I believe it is unfounded.

The stochastic effects that are modelled in this research, although a defining characteristic of low template samples, are also present in conventional profiles and many, if not all, experienced DNA analysts will be familiar with them. The degradation slope discussed in

Chapter 5 is evidence that high molecular weight alleles are already often analysed at low template levels. Heterozygous balance and stutter are guidelines already utilised in the analysis of conventional profiles and the expectation of these values was shown to be invariant depending on peak height.

That leaves only the comprehension of the probabilistic models themselves. The author of this research is neither a statistician nor a biologist and the purpose of this research was to present these statistical models in a manner in which would be received by the forensic biology community.

The introduction of statistics to forensic science has not always been an easy process and the reluctance to introduce seemingly more complex statistical methods is understandable. However, with the increase in sensitivity of DNA profiling techniques, low template samples are becoming more common. How best to interpret these samples is a challenging and highly topical subject.

There is no doubt that the expertise of a DNA analyst is highly valued and there are many occasions where an expert can look at a profile and, using their experience and knowledge, gained through many thousands of hours of analysis, interpret the profile by “eye”. However, there are now frequent challenges to DNA interpretation guidelines – in that no specific guidelines exist. There are recommendations, but as of yet there is no worldwide standard that details how one must interpret a profile. It is time that this is addressed. The author is not recommending that all aspects of a DNA analysis be standardised, but rather that the assumptions of the analysis be clearly stated and that the analysis is repeatable. If one decides to use a continuous model as recommended by the author, then the assumptions and/or hypothesis of the model must be defined.

It is not realistic to expect the entire forensic community to implement the same model and thus variations in reported evidential weights are to be expected. However, it is not acceptable for the general conclusion (i.e. exclusion, inconclusive, inclusion) to differ between analysts as discussed in Chapters 1 and 3. The factors contributing to the interpretation problems discussed in these chapters have been addressed in this research and the author stresses that the first step in regulating DNA interpretation is for the forensic biology community to implement the *LR* as the standard statistical calculation for DNA

interpretation. The stochastic effects modelled in this research cannot be implemented under the other widely used statistical method; random man not excluded (RMNE).

If the move towards implementing a continuous model is undertaken, then this model must be validated and understood by the user as to the extent of its capabilities. Continuous models should not be shown to be better by obtaining the highest *LR*. They should be proven better by their ability to best represent the evidence, and there will be situations in which these models will not be able to be utilised. These situations must be understood by the user.

For example, the author foresees situations in which some complex mixtures will be too difficult to analyse using clearly defined propositions, under the defence hypotheses, in particular. If sensible propositions cannot be formulated under the hypotheses then it follows that a *LR* cannot be calculated.

10.2 The interpretation of complex mixtures

The interpretation of complex mixtures is difficult. The logical approach (*LR*) involves assumptions about the number and the identity of the contributors to the mixture. Consequently a number of calculations are sometimes required and the deconvolution of the mixture can be problematic.

The difficulty in the deconvolution of mixtures has led to some laboratories not presenting a match statistic in court, and simply “not excluding” a person of interest from the mixture if their alleles are present. However, several courts consider the statistical calculation the most important piece of the DNA evidence and will not allow DNA evidence that does not have an accompanying statistic [8, 9].

For example; in the court of appeal in the People v. Coy II [10], the appellant contested the analysis of a mixed DNA profile. Under examination was the evidence of a forensic biologist who testified at length as to the meaning of the DNA evidence. The analyst denied that any of the test results enabled her to testify positively that the DNA belonged to either the person of interest or the victim. However, she concluded that on the basis of the test results, neither the POI nor the victim could “be excluded as a possible contributor” to the mixed DNA sample. The analyst then went on to say that “once we determine that two samples could have come

from the same source then we could calculate a statistical estimate to give a likelihood of how common or how rare it is to find that set of characteristics in another individual,"

However, she performed no statistical interpretation of the results regarding the mixed DNA sample because "our laboratory policy is we do not calculate statistical estimates for mixed samples". In the appeal the judge referred to the following statement by the Committee on DNA Technology in Forensic Science:

"To say that two patterns match, without providing any scientifically valid estimate of the frequency with which such matches might occur by chance, is meaningless".

And the following from US v. Yee, [11]:

"Without the probability assessment, the jury does not know what to make of the fact that the patterns match: the jury does not know whether the patterns are as common as pictures with two eyes or as unique as the Mona Lisa".

Subsequently, the judge in the People v. Coy II ruled that the DNA evidence was inadmissible without a statistical weight to indicate the strength of the "match".

A high degree of variability currently exists with the calculation of statistics in mixture interpretation. Different laboratories follow different mixture interpretation guidelines and furthermore, different analysts following the same guidelines often interpret those guidelines differently. There is no current concordance on the "right" method for mixture interpretation.

A 2005 NIST study [12] presented participating laboratories with mixtures representing four different case scenarios. NIST requested that the results be reported as though they were from a real case including attaching a statistical value and a copy of their laboratory mixture interpretation guidelines.

Results differed in ~10 orders of magnitude for a statistic calculated from the same profile. In addition the following quotes were also obtained:

"Our laboratory does not "pull out" any profile from a mixture for interpretation or statistical purposes", and;

"We currently do not calculate and report statistics on mixture samples".

The Scientific Working Group on DNA Analysis Methods (SWGDAM) Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories 2010 [13], states that:

"3.5. When major or minor contributors cannot be distinguished because of similarity in signal intensities, the sample is considered to be an indistinguishable mixture. The classification as indistinguishable may be limited to some, not all, of the loci for which DNA typing results are obtained and does not imply that the profile is uninterpretable. Individuals may still be included or excluded as possible contributors to an indistinguishable mixture".

and:

"4.1. The laboratory must perform statistical analysis in support of any inclusion that is determined to be relevant in the context of a case, irrespective of the number of alleles detected and the quantitative value of the statistical analysis".

These guidelines (as opposed to standards) and the aforementioned court cases are clear indicators that the forensic biology community needs to introduce, and/or standardise mixture interpretation. This research has briefly touched upon mixtures in Chapter 4, but as these models were implemented using the logical approach (*LR*), the propositions under the defence hypothesis still need to be defined. In the opinion of the author, in the interpretation of unresolvable/complex mixtures, defining the propositions under the defence hypothesis becomes difficult and the number of possible genotype combinations becomes large quickly (even using a constrained approach).

Chapter 8 investigated the stutter ratio (*SR*) which is likely the most questioned variable within complex or LtDNA mixtures, for example, "is this peak stutter, or is there a second low level contributor present?" Within this chapter, the explanatory variables; the longest uninterrupted repeat sequence (*LUS*) and the locus, were shown to affect the *SR*. The distribution of the mean *SR* was able to be modelled using the explanatory variables and the probability of observing a peak (at height *x*, given the parent peak height *y*) could be predicted. i.e. The probability of observing peak *x* given it is stutter could be calculated. In assessing the functionality of the model, an interesting set of data were observed. These data deviated from the assumed distribution of the model, and could be described as unusually large stutters (according to the model). The sample numbers from these observations were

used to retrieve the original epgs. A random sample of the observations was investigated and within these observations low level mixtures were identified. These epgs had previously been passed by a human operator as single source and all of the peaks were assigned as stutter (although for this research the limit of detection was lowered to 30 RFU, and the original profiles were analysed at 50 RFU). These peaks were not noticed as suspiciously large without the application of this model. The author recommends the use of this model (or a similar model) to aid in the designation of peaks as either stutter or allelic.

In this thesis we describe mixtures that

- show evidence as having come from three or more contributors
- have severe heterozygous imbalance, or excessive dropout
- are difficult to formulate hypotheses for

as complex mixtures. Cases that exhibit any of the features listed above require specialist interpretation by senior case working scientists. Continuous models, such as those described in this thesis, and the software that implements them, are extremely important tools in the expert examiners' tool box. However, there will always be cases which are beyond interpretation. The reasons may be articulated quite easily in the *LR* framework. For example, take a gang rape scenario with an unspecified number of assailants. The victim asserts she thinks she was raped by four men, but it could be as many as eight. How should one formulate hypotheses in cases such as this where there are multiple accused, some of who have plead guilty, some of whom are being tried separately, and there is uncertainty about the exact number of contributors? Detractors of the *LR* see this as a weakness of the *LR* method itself. However, we believe it is a strength because it gives a logical set of reasons why interpretation was not carried out. The fact that a method, such as RMNE, can calculate a statistic regardless of circumstance merely highlights the assertion of many commentators that the RMNE does not address the questions of interest to the courts.

However, there is the possibility of future probabilistic models being built that can interpret such complex mixtures. These models will assess the likelihood of the evidence given the hypothesis that the POI is a contributor to the DNA mixture. They will not require an input regarding the number of contributors and will use the quantitative data available directly from

the epg. Therefore there will be no thresholds aside from a limit of detection and analysts will not be designating peaks as alleles or calling genotype combinations. Such a model will be exhaustive and computationally intense. Every possible combination of peaks will be assessed probabilistically and given a weight. For example, the probability of a peak if it is stutter will be weighted along with the probability of the same peak if it is allelic. Possible genotype combinations will also be weighted as to their likelihood of occurrence and those most likely will be given higher weights. This method of interpretation differs from the step by step resolution depicted in Figure 2.2.

Significant laboratory work and experimentation will need to be done to prove the necessary input for such a model. Significant statistical modelling work will need to be done to build the models which evaluate the evidence realistically and fairly. And lastly significant training will need to be put in to ensure that the model is understood by the forensic biology community and does not fall victim to the “black box” syndrome.

References

1. Jeffreys, A.J., V. Wilson, and S.L. Thein, Individual-specific /`fingerprints/' of human DNA. Nature, 1985. 316(6023): p. 76-79.

2. NCR II, National Research Council Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence1996, Washington, D.C.: National Academy Press.

3. Tvedebrink, T., et al., Estimating the probability of allelic drop-out of STR alleles in forensic genetics. Forensic Science International, 2009. 3: p. 222-226.

4. Tvedebrink, T., et al., Allelic drop-out probabilities estimated by logistic regression—Further considerations and practical implementation. Forensic Science International: Genetics, 2012. 6(2): p. 263-267.

5. Debernardi, A., et al., One year variability of peak heights, heterozygous balance and inter-locus balance for the DNA positive control of AmpFℓSTR® Identifiler® STR kit. Forensic science international. Genetics, 2011 5(1): p. 43-49.

6. Bright, J.-A., et al., Developing allelic and stutter peak height models for a continuous method of DNA interpretation. Forensic Science International: Genetics, 2013. 7(2): p. 296-304.

7. Buckleton, J. and P. Gill, Low Copy Number, in Forensic DNA Evidence Interpretation2005, CRC Press. p. 275-297.

8. United States v. Porter, 1992, D.C.App.

9. People v. Barney, 1992, 8 Cal.App.4th 798.

10. People v Coy II, http://caselaw.findlaw.com/mi-court-of-appeals/1030370.html.

11. US v. Yee, 1991, N.D.Ohio.

12. Butler, J. NIST Interlaboratory Mixture Interpretation Study 2005 (MIX05). 2005 [cited 2011 22 June]; Available from: http://www.cstl.nist.gov/biotech/strbase/interlab/MIX05.htm.

13. (SWGDAM),. SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. http://www.fbi.gov/about-us/lab/biometric-analysis/codis/swgdam-interpretation-guidelines [cited 2013].

Bibliography

(1991). United States v. Yee. Ohio.

(1992). DNA Technology in Forensic Science, The National Academies Press.

(1992). People v. Barney. Cal.App.

(1992). United States v. Porter. D.C.App.

(1996). The Evaluation of Forensic DNA Evidence, The National Academies Press.

(2007). People v. Coy II. Michigan. http://caselaw.findlaw.com/mi-court-of-appeals/1030370.html.

(SWGDAM), "SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories." http://www.fbi.gov/about-us/lab/biometric-analysis/codis/swgdam-interpretation-guidelines, 2013.

Aldhous, P. (1993). "Geneticists attack NRC report as scientifically flawed." Science 259(5096): 755-756.

Applied Biosystems (2005). AmpFℓSTR® SGM Plus PCR Amplification Kit: User's Manual.

Balding, D. J. "likeLTD (likelihoods for low-template DNA profiles), https://sites.google.com/site/baldingstatisticalgenetics/software/likeltd-r-forensic-dna-r-code."

Balding, D. J. and J. Buckleton (2009). "Interpreting low template DNA profiles." Forensic Science International: Genetics 4(1): 1-10.

Balding, D. J. and R. A. Nichols (1994). "DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands." Forensic Science International 64(2–3): 125-140.

Ballantyne, J., E. K. Hanson and M. W. Perlin (2012). "DNA mixture genotyping by probabilistic computer interpretation of binomially-sampled laser captured cell populations: Combining quantitative data for greater identification information." Science & Justice : Journal of the Forensic Science Society.

Beecham, G. W. and B. S. Weir (2011). "Confidence interval of the likelihood ratio associated with mixed stain DNA evidence." Journal of Forensic Sciences 56 Supplement 1: S166-171.

Bill, M., P. Gill, J. Curran, T. Clayton, R. Pinchin, M. Healy and J. Buckleton (2005). "PENDULUM--a guideline-based approach to the interpretation of STR mixtures." Forensic Science International 148(2-3): 181-189.

Bright, J.A., K. McManus, S. Harbison, P. Gill and J. Buckleton (2012). "A comparison of stochastic variation in mixed and unmixed casework and synthetic samples." Forensic Science International: Genetics 6(2): 180-184.

Bright, J.A. and S. F. Petricevic (2004). "Recovery of trace DNA and its application to DNA profiling of shoe insoles." Forensic Science International 145(1): 7-12.

Bright, J.A., D. Taylor, J. M. Curran and J. S. Buckleton (2013). "Developing allelic and stutter peak height models for a continuous method of DNA interpretation." Forensic Science International: Genetics 7(2): 296-304.

Bright, J.A., J. Turkington and J. Buckleton (2010). "Examination of the variability in mixed DNA profile parameters for the Identifiler® multiplex." Forensic Science International: Genetics 4(2): 111-114.

Bright, J. A. (2006). Validation of the AmpFℓSTR® Identifiler® PCR Amplification kit at the Forensic Biology Group of ESR.

Bright, J. A., J. S. Buckleton and C. E. McGovern (2010). "Allele frequencies for the four major sub-populations of New Zealand for the 15 Identifiler® loci." Forensic Science International: Genetics 4(2): e65-66.

Bright, J. A., E. Huizing, L. Melia and J. Buckleton (2011). "Determination of the variables affecting mixed MiniFiler™ DNA profiles." Forensic Science International: Genetics 5(5): 381-385.

Bright, J. A., D. Taylor, J. M. Curran and J. Buckleton (2013). "Degradation in forensic DNA profiles." Australian Journal of Forensic Sciences IN PRESS. http://dx.doi.org/10.1080/00450618.2013.772235

Brookes, C., J. A. Bright, S. Harbison and J. Buckleton (2012). "Characterising stutter in forensic STR multiplexes." Forensic Science International: Genetics 6(1): 58-63.

Brookes, C. L. (2006). To Stutter or not to Stutter. MSc, University of Auckland.

Buckleton, J. (2005). A Framework for Interpreting Evidence. Forensic DNA Evidence Interpretation, CRC Press: 27-65.

Buckleton, J. (2009). "Validation issues around DNA typing of low level DNA." Forensic Science International: Genetics 3(4): 255-260.

Buckleton, J. and T. Clayton (2004). Mixtures. Forensic DNA Evidence Interpretation, CRC Press.

Buckleton, J., J. M. Curran and C. J. Basten (2000). "Evaluating the statistical significance of single band profiles in VNTR analyses." Science and Justice 40(1): 27-33.

Buckleton, J. and P. Gill (2005). Low Copy Number. Forensic DNA Evidence Interpretation, CRC Press: 275-297.

Buckleton, J. and P. Gill (2011). "Further Comment on “Low copy number typing has yet to achieve “general acceptance” by Budowle, B., et al, 2009. Forensic Sci. Int. Genetics: Supplement Series 2, 551–552." Forensic Science International: Genetics 5(1): 7-11.

Buckleton, J. and C. Triggs (2006). "Is the 2p rule always conservative?" Forensic Science International 159(2–3): 206-209.

Budowle B, Chakraborty R and v. D. A. (2010). "Authors' response to commentary by Gill and. Buckleton." Journal of Forensic Science 55(1): 269-272.

Budowle, B., A. J. Eisenberg and A. van Daal (2009). "Low copy number typing has yet to achieve "general acceptance"." Forensic Science International: Genetics Supplement Series 2(1): 551-552.

Budowle, B., A. J. Eisenberg and A. van Daal (2009). "Validity of Low Copy Number Typing and Applications to Forensic Science." Croatian Medical Journal(50): 207-217.

Budowle, B. and A. van Daal (2011). "Reply to Comments by Buckleton and Gill on “Low copy number typing has yet to achieve ‘general acceptance”’ by Budowle, B., et al., 2009. Forensic Sci. Int.: Genet. Suppl. Series 2, 551–552." Forensic Science International: Genetics 5(1): 12-14.

Butler, J. (2005). "NIST Interlaboratory Mixture Interpretation Study 2005 (MIX05)." Retrieved 22 June, 2011, from http://www.cstl.nist.gov/biotech/strbase/interlab/MIX05.htm.

Caddy, B., G. Taylor and A. Linacre (2008). A review of the science of low template DNA analysis. http://www.bioforensics.com/articles/Caddy_Report.pdf

Callaghan, T. F., A. D. Manna, A. M. Gross, R. A. Guerrieri, J. C. Luttman and D. L. McClure (2010). "Mixture interpretation: defining the relevant features or guidelines for the assessment of mixed DNA profiles in forensic casework." Journal of Forensic Science 55(1): 265-268.

Caragine, T., J. Mikulasovich, E. B. Tamariz, J. Sebestyen, H. Baum and M. Prinz (2009). "Validation of Testing and Interpretation Protocols for Low Template DNA Samples Using AmpFℓSTR® Identifiler®." Croatian Medical Journal 50: 250-267.

Chung, D. T., J. Drabek, K. L. Opel, J. M. Butler and B. R. McCord (2004). "A study on the effects of degradation and template concentration on the amplification efficiency of the STR Miniplex primer sets." Journal of Forensic Sciences 49(4): 733-740.

Clayton, T. M., J. P. Whitaker, R. Sparkes and P. Gill (1998). "Analysis and interpretation of mixed forensic stains using DNA STR profiling." Forensic Science International 91(1): 55-70.

Cohen, A. S., D. R. Najarian and B. L. Karger (1990). "Separation and analysis of DNA sequence reaction products by capillary gel electrophoresis." Journal of Chromatography A 516(1): 49-60.

Cowell, R. G., S. L. Lauritzen and J. Mortera (2008). "Probabilistic modelling for DNA mixture analysis." Forensic Science International: Genetics Supplement Series 1(1): 640-642.

Cowell, R. G., S. L. Lauritzen and J. Mortera (2011). "Probabilistic expert systems for handling artifacts in complex DNA mixtures." Forensic Science International. Genetics. 5(3): 202-209

Curran, J. M. (2008). "A MCMC method for resolving two person mixtures." Science & Justice 48(4): 168-177.

Curran, J. M. (2010). Introduction to data analysis with R for forensic scientists. Boca Raton, FL, CRC Press.

Curran, J.M.(2013) "Is forensic science the last bastion of resistance against statistics?" Science & Justice 53: 251–252.

Curran, J. M. and J. Buckleton (2010). "Inclusion Probabilities and Dropout." Journal of Forensic Sciences 55(5): 1171-1173.

Curran, J. M., C. M. Triggs, J. Buckleton and B. S. Weir (1999). "Interpreting DNA mixtures in structured populations." Journal of Forensic Sciences 44(5): 987-995.

Debernardi, A., E. Suzanne, A. Formant, L. Pène, A. B. Dufour and J. R. Lobry (2011). "One year variability of peak heights, heterozygous balance and inter-locus balance for the DNA positive control of AmpFℓSTR© Identifiler© STR kit." Forensic science international:. Genetics. 5(1):43-49

Donnelly, P. (1995). "Match probability calculations for multi-locus DNA profiles." Genetica 96(1-2): 55-67.

Dror, I. E. and G. Hampikian (2011). "Subjectivity and bias in forensic DNA mixture interpretation." Science & Justice : Journal of the Forensic Science Society 51(4): 204-208.

Edwards, A., A. Civitello, H. A. Hammond and C. T. Caskey (1991). "DNA typing and genetic mapping with trimeric and tetrameric tandem repeats." American Journal of Human Genetics 49(4): 746-756.

Evett, I. W., C. Buffery, G. Willott and D. Stoney (1991). "A guide to interpreting single locus profiles of DNA mixtures in forensic cases." Journal of the Forensic Science Society 31(1): 41-47.

Findlay I, Frazier R, Taylor A, Urquhart A. (1997) "Single cell DNA fingerprinting for forensic applications." Nature 389:555-556.

Geddes, L. (2010). "Fallible DNA evidence can mean prison or freedom." New Scientist(2773).

Geddes, L. (2010). "How DNA creates victims of chance." New Scientist(2774).

Gilder, J., R. Koppl, I. Kornfield, D. Krane, L. Mueller and W. Thompson (2008). "Comments on the review of low copy number testing." International Journal of Legal Medicine. 123(6), p 535-536

Gilder, J. R., K. Inman, W. Shields and D. E. Krane (2011). "Magnitude-dependent variation in peak height balance at heterozygous STR loci." International Journal of Legal Medicine 125(1): 87-94.

Gill, P., C. H. Brenner, J. S. Buckleton, A. Carracedo, M. Krawczak, W. R. Mayr, N. Morling, M. Prinz, P. M. Schneider and B. S. Weir (2006). "DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures." Forensic Science International 160(2-3): 90-101.

Gill, P. and J. Buckleton (2004). Low Copy Number. Forensic DNA Evidence Interpretation, CRC Press.

Gill, P. and J. Buckleton (2010). "A universal strategy to interpret DNA profiles that does not require a definition of low-copy-number." Forensic Science International: Genetics 4(4): 221-227.

Gill, P., J. Curran and K. Elliot (2005). "A graphical simulation model of the entire DNA process associated with the analysis of short tandem repeat loci." Nucleic Acids Research 33(2): 632-643.

Gill, P., L. Gusmão, H. Haned, W. R. Mayr, N. Morling, W. Parson, L. Prieto, M. Prinz, H. Schneider, P. M. Schneider and B. S. Weir (2012). "DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods." Forensic Science International: Genetics 6(6): 679-688.

Gill, P. and H. Haned (2013). "A new methodological framework to interpret complex DNA profiles using likelihood ratios." Forensic Science International: Genetics 7(2): 251-263.

Gill, P., A. J. Jeffreys and D. J. Werrett (1985). "Forensic application of DNA `fingerprints'." Nature 318(6046): 577-579.

Gill, P., A. Kirkham and J. Curran (2007). "LoComatioN: A software tool for the analysis of low copy number DNA profiles." Forensic Science International 166(2-3): 128-138.

Gill, P., B. Sparkes and J. S. Buckleton (1998). "Interpretation of simple mixtures of when artefacts such as stutters are present - with special reference to multiplex STRs used by the Forensic Science Service." Forensic Science International 95(3): 213-224.

Gill, P., R. Sparkes, R. Pinchin, T. Clayton, J. Whitaker and J. Buckleton (1998). "Interpreting simple STR mixtures using allele peak areas." Forensic Science International 91(1): 41-53.

Gill, P., J. Whitaker, C. Flaxman, B. Brown and J. Buckleton (2000). "An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA." Forensic Science International(112): 17-40.

Green RL, R. I., Boland C, Hennessy LK (2005). "Developmental validation of the quantifiler real-time PCR kits for the quantification of human nuclear DNA samples." Journal of Forensic Sciences 50(4): 809-825.

Haned, H. (2010). "Forensim: An open-source initiative for the evaluation of statistical methods in forensic genetics." Forensic Science International. Genetics.

Haned, H., T. Egeland, D. Pontier, L. Pene and P. Gill (2011). "Estimating drop-out probabilities in forensic DNA samples: a simulation approach to evaluate different models." Forensic Science International: Genetics 5(5): 525-531.

Haned, H., L. Pène, F. Sauvage and D. Pontier (2011). "The predictive value of the maximum likelihood estimator of the number of contributors to a DNA mixture." Forensic Science International: Genetics 5(4): 281-284.

Hartl, D. L. (1994). "Forensic DNA typing dispute." Nature 372(6505): 398-399.

http://judgmental.org.uk/judgments/EWCA-Crim/2006/[2006]_EWCA_Crim_1395.html.

Jeffreys, A. J., V. Wilson and S. L. Thein (1985). "Individual-specific `fingerprints' of human DNA." Nature 316(6023): 76-79.

Jobling, M. A. and P. Gill (2004). "Encoded evidence: DNA in forensic analysis." Nature Reviews: Genetics 5(10): 739-751.

Kelly, H., J.A. Bright, J. M. Curran and J. Buckleton (2012). "Modelling heterozygote balance in forensic DNA profiles." Forensic Science International: Genetics 6(6): 729-734.

Kelly, H., J. A. Bright, J. M. Curran and J. Buckleton (2012). "The interpretation of low level DNA mixtures." Forensic Science International: Genetics 6(2): 191-197.

Klintschar, M. and P. Wiegand (2003). "Polymerase slippage in relation to the uniformity of tetrameric repeat stretches." Forensic Science International 135(2): 163-166.

Koenker, R. (2005). Quantile Regression, Cambridge University Press.

Koenker, R. (2012). Quantreg: Quantile Regression. http://www.r-project.org.

Krane, D. E., S. Ford, J. R. Gilder, K. Inman, A. Jamieson, R. Koppl, I. L. Kornfield, D. Michael Risinger, N. Rudin, M. S. Taylor and W. C. Thompson (2008). "Sequential Unmasking: A Means of Minimizing Observer Effects in Forensic DNA Interpretation." Journal of Forensic Sciences 53(4): 1006-1007.

Lazaruk, K., J. Wallin, C. Holt, T. Nguyen and P. S. Walsh (2001). "Sequence variation in humans and other primates at six short tandem repeat loci used in forensic identity testing." Forensic Science International 119(1): 1-10.

Lewontin, R. C. (1993). "Which population? ." American Journal of Human Genetics 52: 205.

Lewontin, R. C. (1994). "Forensic DNA typing dispute." Nature 372(6505): 398-398.

Lewontin, R. C. and D. L. Hartl (1991). "Population genetics in forensic DNA typing." Science (New York, N.Y.) 254(5039): 1745-1750.

Lohmueller, K. E. and N Rudin (2013). "Calculating the Weight of Evidence in Low Template Forensic DNA Casework." Journal of Forensic Science. 58(1): 243-249.

Lunn, D., D. Spiegelhalter, A. Thomas and N. Best (2009). "The BUGS project: Evolution, critique and future directions." Stat Med 28(25): 3049-3067.

Lygo, J. E., P. E. Johnson, D. J. Holdaway, S. Woodroffe, C. P. Kimpton, P. Gill, J. P. Whitaker and T. M. Clayton (1994). "The validation of short tandem repeat (STR) loci for use in forensic casework." International Journal of Legal Medicine 107(2): 77-89.

Miller, C. R., P. Joyce and L. P. Waits (2002). "Assessing Allelic Dropout and Genotype Reliability Using Maximum Likelihood." Genetics Society of America(160): 357-366.

Mitchell, A. A., J. Tamariz, K. O'Connell, N. Ducasse, Z. Budimlija, M. Prinz and T. Caragine (2012). "Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in." Forensic Science International: Genetics 6(6): 749-761.

Mullis, K. B. (1990). "The unusual origin of the polymerase chain reaction." Scientific American 262(4): 64-65.

Mullis, K. B. and F. Faloona (1987). "Specific synthesis of DNA in vitro via a polymerase catalysed chain reaction." Methods in Enzymology 155: 335-350.

Mullis, K. B., F. Faloona, S. Scharf, R. Saiki, G. Horn and H. Erlich (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harbour Symposia on Quantitative Biology. 51: 263-273.

nfstc. "http://www.nfstc.org/pdi/Subject06/pdi_s06_m02_05.htm."

Nicklas, J. A., T. Noreault-Conti and E. Buel (2012). "Development of a Real-Time Method to Detect DNA Degradation in Forensic Samples." Journal of Forensic Sciences 57(2): 466-471.

Paoletti, D. R., T. E. Doom, C. M. Krane, M. L. Raymer and D. E. Krane (2005). "Empirical analysis of the STR profiles resulting from conceptual mixtures." Journal of Forensic Sciences 50(6): 1361-1366.

Pascali, V. L. and S. Merigioli (2012). "Joint Bayesian analysis of forensic mixtures." Forensic Science International: Genetics 6(6): 735-748.

Perlin, M. (2012). "Cybergenetics DNA identification for scientists https://www.cyben.com."

Perlin, M. (2012). Easy reporting of hard DNA: computer comfort in the court room. Forensic Magazine. 9: 32-37.

Perlin, M. and J. Galloway (2012). Computer DNA evidence interpretation in the Real IRA Massereene terrorist attack. Evidence Technology Magazine. 10: 20-23.

Perlin, M., M. Legler, C. Spencer, J. Smith, W. Allan, J. Belrose and B. Duceman (2011). "Validating Trueallele® DNA mixture interpretation " Journal of Forensic Science 56(6): 1430-1447.

Perlin, M. W., G. Lancia and S. K. Ng (1995). "Toward fully automated genotyping: genotyping microsatellite markers by deconvolution." American Journal of Human Genetics 57(5): 1199-1210.

Perlin, M. W. and A. Sinelnikov (2009). "An information gap in DNA evidence interpretation." PLoS One 4(12): e8327.

Perlin, M. W. and B. Szabady (2001). "Linear mixture analysis: a mathematical approach to resolving mixed DNA samples." Journal of Forensic Sciences 46(6): 1372-1378.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. http://mcmc-jags.sourceforge.net.

Promega (2002). DNA IQTM System - Small Sample Casework Protocol. Madison WI, USA.

Puch-Solis, R., T. Clayton, J. Whitaker and S. Pope (2010). "Assigning weight of DNA evidence using a continuous model that takes into account stutter and dropout." Science & Justice 50(1): 46-47.

R Core Team (2012). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Rennison, A. (2008). Response to Professor Brian Caddy's Review on the Science of Low Template DNA Analysis. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/117557/response-caddy-dna-review.pdf

Ruitberg, C. M., D. J. Reeder and J. M. Butler (2001). "STRBase: a short tandem repeat DNA database for the human identity testing community." Nucleic Acids Research 29(1): 320-322.

Smith, J. C., R. Anwar, J. Riley, D. Jenner, A. F. Markham and A. J. Jeffreys (1990). "Highly polymorphic minisatellite sequences: allele frequencies and mutation rates for five locus-specific probes in a Caucasian population." Journal of the Forensic Science Society 30(1): 19-32.

Sparkes, R., C. Kimpton, S. Watson, N. Oldroyd, L. Barnett, J. Arnold, C. Thompson, R. Hale, J. Chapman, A. Urquhart, P. Gill and T. Clayton (1996). "The validation of a 7-locus multiplex STIR test for use in forensic casework." International Journal of Legal Medicine 109(4): 186-194.

STRbase. "(http://www.cstl.nist.gov/strbase/str_D21S11.htm)."

Thomas, A., B. O'Hara, U. Ligges and S. Sturtz (2006). "Making BUGS Open." R News 6(1): 12-17.

Thompson, W. (1996). "DNA evidence in the O.J. Simpson Trial." University of Colorado Law Review 67: 827-857.

Tvedebrink, T., P. S. Eriksen, M. Asplund, H. S. Mogensen and N. Morling (2012). "Allelic drop-out probabilities estimated by logistic regression—Further considerations and practical implementation." Forensic Science International: Genetics 6(2): 263-267.

Tvedebrink, T., P. S. Eriksen, H. S. Mogensen and N. Morling (2009). "Estimating the probability of allelic drop-out of STR alleles in forensic genetics." Forensic Science International 3: 222-226.

Tvedebrink, T., P. S. Eriksen, H. S. Mogensen and N. Morling (2010). "Evaluating the weight of evidence by using quantitative short tandem repeat data in DNA mixtures." Journal of the Royal Statistical Society: Series C (Applied Statistics) 59(5): 855-874.

Tvedebrink, T., P. S. Eriksen, H. S. Mogensen and N. Morling (2012). "Statistical model for degraded DNA samples and adjusted probabilities for allelic drop-out." Forensic Science International: Genetics 6(1): 97-101.

Tvedebrink, T., H. S. Mogensen, M. C. Stene and N. Morling (2012). "Performance of two 17 locus forensic identification STR kits-Applied Biosystems's AmpFℓSTR® NGMSElect and Promega's PowerPlex(R) ESI17 kits." Forensic Science International: Genetics 6(5): 523-531.

University of Leicester, "STR profiling in the UK " http://www2.le.ac.uk/departments/emfpu/genetics/explained/str-uk.

Van Nieuwerburgh, F., E. Goetghebeur, M. Vandewoestyne and D. Deforce (2009). "RMNE probability of forensic DNA profiles with allelic drop-out." Forensic Science International: Genetics Supplement Series 2(1): 462-463.

Veth, J. (2002). The characterisation of artefacts found in DNA profiles. Master of Science, The University of Auckland.

Walsh, P. S., N. J. Fildes and R. Reynolds (1996). "Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA." Nucleic Acids Research 24(14): 2807-2812.

Weir, B., C. Triggs, L. Starling, L. Stowell, K. Walsh and J. Buckleton (1997). "Interpreting DNA mixtures." Journal of Forensic Sciences 42: 213-222.

Weusten, J. and J. Herbergs (2012). "A stochastic model of the processes in PCR based amplification of STR DNA in forensic applications." Forensic Science International: Genetics 6(1): 17-25.

Whitaker, J. and E. A. Cotton (1999). Low copy number. First International Conference on Forensic Human Identification in the Millennium.

Whitaker, J. P., E. A. Cotton and P. Gill (2001). "A comparison of the characteristics of profiles produced with the AmpFℓSTR® SGM Plus™ multiplex system for both standard and low copy number (LCN) STR DNA analysis." Forensic Science International 123(2-3): 215-223.