2007

Computational Methods for the Objective Review of Forensic DNA Testing Results

Jason R. Gilder
Wright State University

Computational methods for the objective review of forensic DNA testing results

A dissertation submitted in partial fullfillment
of the requirements for the degree of
Doctor of Philosophy

By

JASON R. GILDER
M.S., Wright State University, 2003

2007
Wright State University

WRIGHT STATE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

May 21, 2007

I HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER MY SUPERVISION BY Jason R. Gilder ENTITLED Computational methods for the objective review of forensic DNA testing results BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Doctor of Philosophy.

Travis E. Doom, Ph.D.
Dissertation Director

Thomas Sudkamp, Ph.D.
Director, Computer Science and
Engineering Ph.D. Program

Joseph F. Thomas, Ph.D.
Dean, School of Graduate Studies

Committee on Final Examination

Travis E. Doom, Ph.D.

Dan Krane, Ph.D.

Michael Raymer, Ph.D.

Forouzan Golshani, Ph.D.

Nikolaos Bourbakis, Ph.D.

Laurence Mueller, Ph.D.

Allan Jamieson, Ph.D.

Abstract

Gilder, Jason R. Ph.D., Department of Computer Science and Engineering, Wright State University, 2007. Computational methods for the objective review of forensic DNA testing results.

Since the advent of criminal investigations, investigators have sought a "gold standard" for the evaluation of forensic evidence. Currently, deoxyribonucleic acid (DNA) technology is the most reliable method of identification. Short Tandem Repeat (STR) DNA genotyping has the potential for impressive match statistics, but the methodology not infallible. The condition of an evidentiary sample and potential issues with the handling and testing of a sample can lead to significant issues with the interpretation of DNA testing results. Forensic DNA interpretation standards are determined by laboratory validation studies that often involve small sample sizes.

This dissertation presents novel methodologies to address several open problems in forensic DNA analysis and demonstrates the improvement of the reported statistics over existent methodologies. Establishing a dynamically calculated RFU threshold specific to each analysis run improves the identification of signal from noise in DNA test data. Objectively identifying data consistent with degraded DNA sample input allows for a better understanding of the nature of an evidentiary sample and affects

the potential for identifying allelic dropout (missing data). The interpretation of mixtures of two or more individuals has been problematic and new mathematical frameworks are presented to assist in that interpretation. Assessing the weight of a DNA database match (a cold hit) relies on statistics that assume that all individuals in a database are unrelated – this dissertation explores the statistical consequences of related individuals being present in the database. Finally, this dissertation presents a statistical basis for determining if a DNA database search resulting in a very similar but nonetheless non-matching DNA profile indicates that a close relative of the source of the DNA in the database is likely to be the source of an evidentiary sample.

Table of Contents

List of Figures

List of Tables

Acknowledgments

Dan Krane for his strong guidance and support. He provides me with an endless supply of insight and new ideas.

Travis Doom for his continued assistance and support beginning with my undergraduate senior research project and continuing through my introduction to Bioinformatics.

Michael Raymer has been an invaluable guide through the world of Bioinformatics.

Simon Ford for his assistance and support, particularly in the area of computerized DNA analysis. His is continually helping me obtain new and useful information from DNA data files.

William Thompson for his assistance and support as well as many case studies that fuel my research.

Keith Inman for his valuable insight and even more valuable data.

Dedicated to my parents, Jim and Betsy: eternal idealists who have helped me strive for a better tomorrow.

Chapter 1

Introduction

Since the advent of criminal investigations, investigators have sought a "gold standard" for the evaluation of forensic evidence. For nearly a century, fingerprints were considered a unique identifier until recent failures in the technology resulted in increased scrutiny from the scientific community (Stacey, 2005). DNA has usurped the role of fingerprints as the most powerful means of identification currently available. While the foundation of forensic DNA testing technology is firmly rooted in science, the lack of an objective computational methodology for the interpretation of DNA testing results allows for the introduction of subjectivity.

The hallmark of the scientific method of understanding is reproducibility and objectivity. The means for collecting forensic DNA testing data is well-established and well-validated (Fregèau and Fourney, 1993; Edwards et al., 1992; Fregèau et al., 1999; Holt et al., 2002). However, there exist few statistically supportable objective standards for the evaluation and interpretation of DNA data. The principle objective of any forensic technique is to exclude individuals as possible contributors to evidence

samples and, only when that fails, to determine the chance that an individual is the actual source of materials that have been subjected to testing. In the absence of experimentally determined standards and related statistical approaches, analysts may either accidentally or deliberately introduce subjectivity into their interpretation of DNA testing results through a reliance simply upon their "training, experience and expertise." It then becomes possible to interpret the evidence as indicative of a scenario where an individual "matches" the evidence even when other interpretations are possible (and perhaps more statistically likely).

While forensic DNA testing results are amenable to computational analysis, very few computational algorithms have been developed. Currently, standards for interpretation are often developed during a laboratory's validation process (DNA Advisory Board, 2000a; Federal Bureau of Investigation, 2005; Scientific Working Group on DNA Analysis Methods, 2000; Moretti et al., 2001a). Validation studies are typically performed when a laboratory is implementing a new technology, such as a new genetic analyzer or DNA testing kit. Validation usually entails testing a relatively small number of (usually pristine) samples and characterizing the resulting observations. The lab-specific standards that are created from these observations become the established practice for all future casework.

While validation is a valuable process, it does not guarantee adequate performance of a technology for ongoing casework. Testing a small number of samples may provide an inadequate bases for accurate classification standards. Since validation is typically only performed with the introduction of a new technology, tests are usually performed with new instruments, reagents, and freshly trained staff. The age of the equipment

and chemicals, and the level of analyst experience may result in testing results that are not adequately characterized by the parameters derived from the initial validation studies. Perhaps most importantly, current validation studies characterize only a subset of the issues encountered during routine casework.

Computational techniques and research studies address the issue of inadequate validation by exploring issues on a large and statistically relevant scale. Instead of using thresholds obtained with a small number of initial observations, statistical measures can be employed to give accuracy and, more importantly, a much-needed level of acceptable error to an ever-growing body of data. This document presents five novel computational methods to address such problems and reports the improvement over existent methodologies.

1.1 Summary overview of forensic DNA testing

The process of obtaining a DNA test result is one that involves many steps (see Chapter Two for a full summary of DNA theory and testing methodology). The most widely-used contemporary process of performing a DNA test can be summarized with the following steps:

1. Collect a DNA sample

2. Extract, purify, and amplify (replicate) the DNA sample

3. Examine a sample using a genetic analyzer (capturing the signal output digitally)

4. Utilize software to separate and label the multiplexed signal data

5. Manually interpret the results to generate a DNA profile

6. Compare DNA profiles of known individuals to tested samples to determine the presence of a DNA profile match

7. Calculate the appropriate statistics to determine the evidentiary weight of a DNA profile match

Steps one through four are well-established and fairly mechanical. The process becomes more difficult once the results must be manually interpreted in step five. The individual evaluating the results must determine what portion of the signal is the result of DNA products and what portion of the signal is the result of technical artifacts that must be disregarded. The decisions made at this point will directly affect the remaining two steps of the analysis process. The use of rigorous algorithmic standards can help identify spurious results and increase the confidence in the testing results.

DNA test results are typically generated using a genetic analyzer that operates through a process known as capillary electrophoresis (see Chapter 2). The results of a DNA test are often presented in the form of an electropherogram (see Figure 1.1). Each peak represents a particular DNA marker (called an allele). The height of each peak is measured in relative fluorescence units (RFUs) and is indicative of the approximate amount of DNA present in a sample. The x-axis of the electropherogram corresponds to the size of the DNA fragments detected. Fragments increase in size from left to right across an electropherogram.

Figure 1.1: **STR DNA electropherogram. Peaks are labeled with two boxes. The first contains the peak's allele call (number of repeats) and the second contains the peak's height (in RFUs). Peaks are broken into groups of loci (chromosomal locations), labeled in boxes above the peaks.**

Electropherograms are separated into groups of tested locations, or loci (singular: locus). Each locus is illustrated on the electropherogram as a grouping of peaks. A tested cell contains two copies of the human genome (one inherited from each parent), so a DNA test produces two markers for each tested locus. The markers can be the same between both copies (homozygote) or differ between each copy (heterozygote). The result is that for a single-source sample, such as a reference sample or positive control, the presence of two peaks at a locus indicate a heterozygote, while a single peak is consistent with being a homozygote. If three or more peaks are observed at a single locus, it is an indication that the sample is composed of a mixture of two or more contributors. Contamination and technical artifacts can also introduce additional peaks and make the interpretation process more difficult (see Section 2.5).

Sample comparison is usually carried out between the DNA profiles from a known reference sample and an item of evidence. If every allele found in the reference sample is present in an evidence sample, then that individual cannot be excluded from being

a contributor to that item of evidence. Barring testing issues, if any of the alleles found in a reference sample are not present in an evidentiary sample, that individual is said to be excluded as the source of the evidence. Situations involving mixtures of two or more individuals and low amounts of DNA (among others) can give rise to ambiguities due to the increased number of possible interpretations.

In order to determine the weight associated with a DNA match, a statistical calculation must be performed to determine the chance of a coincidental match (see Chapter 2 for full details). The statistical question being asked is, what is the chance of randomly picking an unrelated person from a given population who cannot be excluded from the DNA profile observed on a given item of evidence? The random match probability (RMP) answers the question for single-source samples and the combined probability of inclusion (CPI) answers the question for mixed samples containing two or more contributors. The chance of a coincidental match for a single-source sample where genotype information has been determined across 13 STR loci is often less than one in a trillion (Butler, 2001).

Known suspects are not available for all criminal investigations. In these situations, databases of DNA profiles are often searched against the evidence profiles to attempt to find the source. In such a situation, a match between an evidentiary sample and a database entry is called a cold hit. DNA databases are usually comprised of individuals who have been convicted of a crime (or in some instances arrested or detained). Each state has its own DNA database and the FBI maintains a national database known as the COmbined DNA Index System (CODIS) (Federal Bureau of Investigation, 2007).

1.2 Overview

DNA tests can be performed with minute amounts of starting material. As a result, DNA samples contain low-level results. It can become difficult to determine the signal-to-noise threshold where genotype information can be reliably measured. Current validation techniques determine the sensitivity of an instrument and apply static and sometimes arbitrary signal-to-noise thresholds. Single static thresholds cannot adequately characterize all testing results as the level of noise in the system varies due to the age of the equipment, reagents, and other changing factors.

Chapter Three presents a method to establish a run-specific RFU threshold to determine the level of instrument noise in the system for a specific set of testing results. This method analyzes the signal data from control samples tested with every analysis run to determine a statistically significant threshold for signal detection. Run-specific RFU thresholds allow forensic scientists to glean as much useful information as possible from an electropherogram without a statistically significant risk of confusing instrument-related background noise as being signal derived through the DNA testing process.

Validation studies usually involve the analysis of reference samples that are of high quality and abundant quantity. Evidentiary samples are often collected in a condition that is far from ideal. Environmental effects can lead to the breakdown (degradation) of the DNA molecules. With degradation comes the increased risk of incomplete testing results (allelic dropout). The level of degradation in casework samples is currently identified by manual inspection. An analyst's "training, expertise, and experience"

determines how a sample is classified. Chapter Three proposes a method for the objective identification of potentially degraded samples. This approach provides a statistical metric for determining how similar or dissimilar a tested sample is from a population of established non-degraded samples.

Evidentiary samples often contain DNA from more than one individual. Validation usually does not extend to the interpretation of DNA mixtures, whose interpretation is notoriously problematic (Thompson et al., 2003a; Thompson et al., 2003b; Butler, 2001). As many individuals possess overlapping subsets of DNA markers, it is difficult to determine the number of contributors present in a mixture and determine which specific individuals gave rise to the observed profiles. Chapter Four presents an algorithm based on Boolean logic to create a mathematically provable framework for the resolution of DNA mixtures. The method attempts to accurately identify which DNA markers belong to which contributor. Even when a mixture cannot be fully resolved, additional knowledge that this algorithm produces can often be used to eliminate potential contributors and yield a partially (or possibly fully) resolved mixture.

The appropriate means of attaching a statistical weight to a DNA profile match that has resulted from a database search (a cold hit) has been the subject of significant debate within the scientific literature (National Research Council, 1992; National Research Council, 1996; DNA Advisory Board, 2000b; Balding and Donnelly, 1996). Most approaches are based upon the random match probability, which considers the pool of alternative suspects as being completely unrelated. When these methods are used with a cold hit, they assume independence of the underlying data. No studies

have been published that validate this assumption. Simulation studies allow for the creation of DNA databases that mirror the real world based on published allele frequency information. Virtual families can be created to introduce related (and thus dependent) individuals to the database population. Chapter Five presents a simulation study that explores the statistical effects of the presence of related individuals in a DNA database. This study provides insight as to the change in the weight assigned to assays of statistical identification necessary due to the makeup of a DNA database.

The issue of related individuals and database searches also comes into play with familial searches. In some circumstances a database search yielding a close, but imperfect match may lead law enforcement to investigate a relative of the partial database match. An objective statistical test should be used to determine when investigation of a relative is warranted. A likelihood ratio approach has been developed (Paoletti et al., 2006), but it has not been extensively validated or directly compared against existing methods of familial searches.

Chapter Five presents a comparison of a simulation of the CODIS software currently used by law enforcement and the likelihood ratio approach for several sets of individuals with varying degrees of relatedness. The CODIS software performs standard profile comparisons, while the likelihood ratio approach provides insight as to the most likely source of the evidence: a related or unrelated individual. Chapter Five expands the likelihood ratio method and provides additional insight into the most efficient means of performing a familial search in a variety of real-world situations, including the interpretation of two and three-person mixtures and in situations where the race of the perpetrator is unknown or misclassified.

The methods presented herein can be readily adopted by laboratory analysts and DNA experts to better evaluate forensic DNA testing results. Each method also lends itself to automation, so that it can be directly incorporated into a laboratory protocol without creating a significant burden. These tools may also be used as the groundwork for further studies of additional issues in forensic DNA analysis and interpretation to make those processes more objective as well (Chapter Six).

Chapter 2

Background

2.1 Basic DNA theory

DNA contains the blueprint of the human body (Campbell, 1996). Almost all human cells contain DNA. Most human DNA is wrapped into 23 pairs of chromosomes within each cell. Half of each chromosome pair is inherited from the individual's mother and half comes from the individual's father. Chromosomes each contain many functional components called genes. When it is possible to distinguish between two or more variants of a single gene, those variants are called alleles. At the finest level of scale, chromosomes and genes can be described as a sequence of nucleotides (named for the nitrogenous bases they contain). Nucleotides are present in four forms: adenine (A), guanine (G), cytosine (C), and thymine (T). In its typically double stranded form, an "A" on one DNA strand always pairs with the "T" on another, and a "C" will always pair with a "G". This association is called a "base pair" (Voet et al., 1999). At this level, the familiar double helix can be visualized (see Figure 2.1).

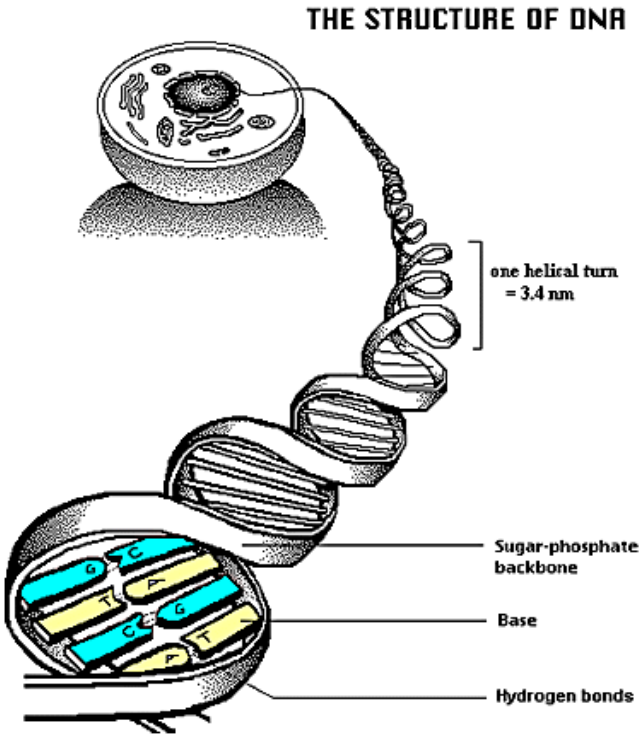

Figure 2.1: **DNA being unraveled from the nucleus of a cell (The National Health Museum, 2007). The sugar-phosphate backbone creates the familiar double helix and is made up of linkages between the phosphate of one nucleotide and the sugar of the adjoining nucleotide. The interior of the double helix is made up of pairs of nitrogenous bases, held together by hydrogen bonds. The pairing of bases is specific: adenine (A) can only pair with thymine (T) and cytocine (C) can only pair with guanine (G). DNA is grouped into 23 pairs of separate chromosomes and resides in the nucleus of all nucleated cells (Campbell, 1996).**

The human genome contains approximately 3.12 billion base pairs and is virtually identical between all humans (Dennis et al., 2002). A distinction is often made between meaningful DNA (e.g. genes or the regions containing the coding information for proteins and other gene products) and non-coding, or "junk" DNA that holds no known function. Even distantly related humans are approximately 99.5% identical at this level of their DNA (Dennis et al., 2002). Differences are found predominantly in "junk" DNA. Unlike coding regions, changes (mutations) to non-coding regions are less likely to affect the survival of the organism and are therefore more likely to be tolerated.

"Junk" DNA is not entirely useless. Non-coding regions hold the information used for current forensic DNA testing: Short Tandem Repeats (STRs). STRs are short repeating segments of DNA that often differ between people (Fregèau and Fourney, 1993). STR sequences used for forensic analyses are made up of four base pair increments (tetranucleotide repeats). STR alleles are named for the number of observed STR repeats at a given locus. An allele is a term for the genetic form present and a locus (pl: loci) is a chromosomal location. For example, if the sequence "GATA" is observed ten consecutive times at a particular locus, that person is said to have a *10* allele at that locus. DNA testing kits analyze a set of specific loci. For example, Applied Biosystems' Profiler Plus® and COfiler® kits (currently the most widely-used testing kits) analyze a combined 13 loci plus the sex-determining locus Amelogenin (see Figure 2.2).

Since the human genome contains two copies of every chromosome, two alleles are represented for every locus. If the alleles are different, the individual is said to

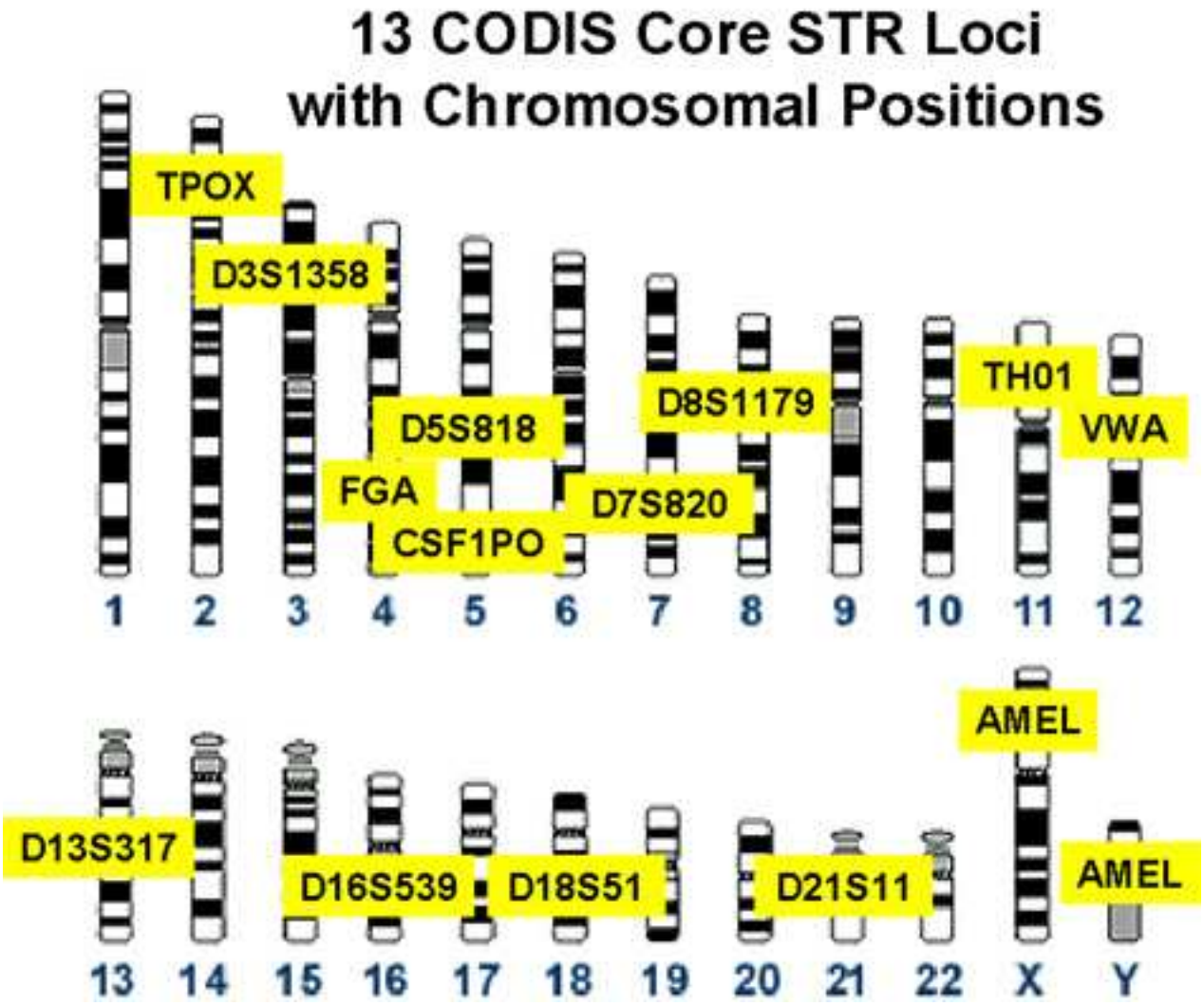

Figure 2.2: **A map of the chromosomes in the human genome. The labels correspond to chromosomal locations examined by the Profiler Plus® and COfiler® analysis kits (National Institute of Standards and Technology, 2007).**

be heterozygous for that locus. If a locus contains two copies of the same allele, the individual is said to be homozygous. Heterozygous loci are reported using both observed allele numbers as a pair, such as (11, 12) or (7, 9). Homozygous loci can be reported as a single allele, (11), or as a repeated pair, (11, 11). The Amelogenin locus is used to determine the sex of the contributor, returning (X, X) for a woman or (X, Y) for a man. The sum collection of all alleles detected in a sample is called the DNA profile or genotype of a sample.

2.2 Sources of DNA

A complete set of genetic instructions is found inside virtually every human cell. DNA can be found in (but is not limited to): blood, semen, skin cells, tissue, organs, muscle, brain cells, bone, teeth, saliva, mucus, perspiration, urine, and even feces (National Criminal Justice Reference Service, 2007). DNA has no clock and can exist for many years in an essentially constant state. The presence of a DNA profile says nothing about the time frame or the circumstances under which DNA was transferred to that item.

Epithelial skin cells are easily shed, facilitating DNA transfer. DNA profiles can be readily obtained from skin swabs or objects that have been handled (Oorschot and Jones, 1997). Studies have been performed showing that DNA can be transferred through passive means (Taylor, 2001). For example, if person A kisses person B on the cheek and person C touches person B on the cheek with a glove, DNA consistent with A and B will likely be present on the glove. DNA can also be passed on something

as simple as a damp towel. Other tests have shown that semen can be transferred in the laundry, adding a new dimension to rape, particularly incest, investigations (Kafarowski et al., 1996). The many possibilities for DNA profiles to arise illustrates the fact that a DNA test alone provides no insight as to what cell-types gave rise to a DNA profile or how long a particular profile has been associated with a sample.

2.3 DNA testing

DNA testing is a relatively recent advent in forensic science (Wambaugh, 1991). Original testing methods were first used in US courts in 1988 (Butler, 2001). Subsequent changes were made to improve the sensitivity and the resolving power of the analysis. With the introduction of Polymerase Chain Reaction (PCR), samples with very little starting material could be amplified enough to produce a full STR profile (Fregèau and Fourney, 1993; Kimpton et al., 1993). The FBI began examining STR regions for all of its forensic DNA analysis in 1998 (Butler, 2001).

The process of collecting a DNA sample and generating a genotype is a complex process (Butler, 2001). The DNA must first be extracted from a sample and quantified to determine how much starting material is present. The STR regions of the DNA are separated and amplified using Polymerase Chain Reaction (PCR). The PCR process acts as a sort-of Xerox machine for DNA, producing millions of copies of DNA (Mullis and Faloona, 1987; Saiki et al., 1988; Mullis, 1990).

Before PCR, genetic testing required a greater amount of starting material to produce a DNA profile. These early testing methods sometimes ran the risk of con-

suming an entire sample, which would prevent the sample from being re-tested. A failed DNA test could not always be reanalyzed, which presented a serious limitation in DNA testing technology.

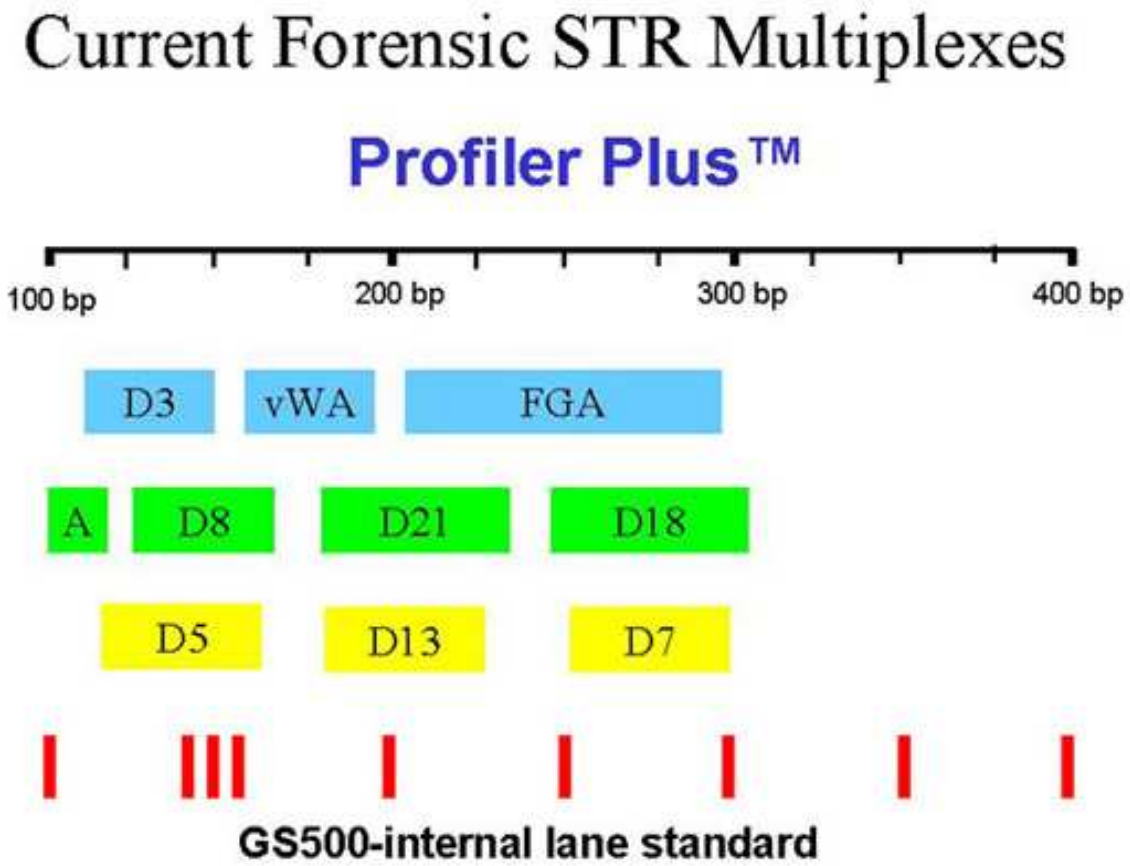

Figure 2.3: **The loci and associated florescent dyes (blue, green, yellow, and red) examined by the Profiler Plus® analysis kit (National Institute of Standards and Technology, 2007). The blue, green, and yellow dyes are used to identify the chromosomal origin of the DNA fragments being analyzed. The red dye is a size standard, used to identify the relative sizes of the DNA fragments.**

Fluorescent dye tags are integrated into the isolated STR fragments during the PCR amplification process. Since many of the STR DNA fragments from different loci

are observed in the same size ranges, different dyes allow the separation of fragments into distinct loci (see Figure 2.3). With most testing kits, the red dye, often referred to as the ROX channel, is a size standard used to calibrate the system to ensure that the fragments are sized properly. More recent testing kits, such as Identifiler®, utilize the red channel to examine a larger number of loci. These kits add an addition orange channel (called LIZ) that acts as the size standard (Applied Biosystems, 2005b).

Several DNA testing kits are available and each examine a different set of polymorphic loci, which are likely to differ from one individual to another. The FBI established the 13 CODIS loci as a standard of which loci should be examined, particularly when developing a database profile (Federal Bureau of Investigation, 2007). Many laboratories currently examine the 13 CODIS loci using Applied Biosystems' Profiler Plus® and COfiler® analysis kits (Applied Biosystems, 2000a; Applied Biosystems, 2005a). Profiler Plus® (see Figure 2.3) examines nine loci plus Amelogenin, while COfiler® examines six loci plus Amelogenin. The analysis kits analyze two of the same loci (D3S158 and D7S820), making the total unique loci examined 13 (plus Amelogenin). The latest analysis kit from Applied Biosystems is the Identifiler® kit, which examines 15 loci (D2S1338 and D19S433 were added) . Promega (Promega, 2007) provides the PowerPlex® 16 analysis kit, which analyzes 15 loci plus Amelogenin (PentaD and PentaE pentanucleotide loci were added).

Separation of the PCR amplification products generated with these various kits is typically accomplished by capillary electrophoresis on a genetic analyzer, such as the Applied Biosystems 310 or 3100 series (Applied Biosystems, 2007). The separation process begins with a capillary containing an electrode being inserted into a sample

vial. Since DNA is an intrinsically negatively charged molecule, the electric field induced by the electrode can effectively pull the DNA molecules through the capillary. The capillary contains a gel (or matrix) to control the flow of fragments. Small fragments will move more quickly than larger fragments due to interactions with the matrix. At the end of the capillary lies a laser and photo-detector. As STR fragments pass by the laser, the integrated dye tags fluoresce and the amount of light emitted is captured by the photo detector. The time at which the fragment is "seen" by the photo-detector is relative to the size of the fragment (smaller fragments are seen first). The amount of light fluoresced is relative to the quantity of DNA present in a fragment size. Finally, the color of the light fluoresced indicates from which locus the fragment originates. All of the data is recorded and stored in an electronic format on a computer attached to the genetic analyzer.

Software programs, such as Applied Biosystems' GeneScan® and GenoTyper®, analyze the raw electronic data generated by the genetic analyzer to produce interpretable results (Applied Biosystems, 2007). GeneScan® processes the raw electronic data to separate the electropherogram into its individual dye channels and determine the attributes of each detected peak, including its size, height, and area. After GeneScan® has processed the data, GenoTyper® adds the allele calls to the peaks and generates the final electropherograms used for interpretation. Applied Biosystems has more recently released a program called GeneMapper® ID, which is able to process the raw electronic data and create the final electropherogram using a single software package.

Determining whether samples have matching DNA profiles can be a fairly straight-

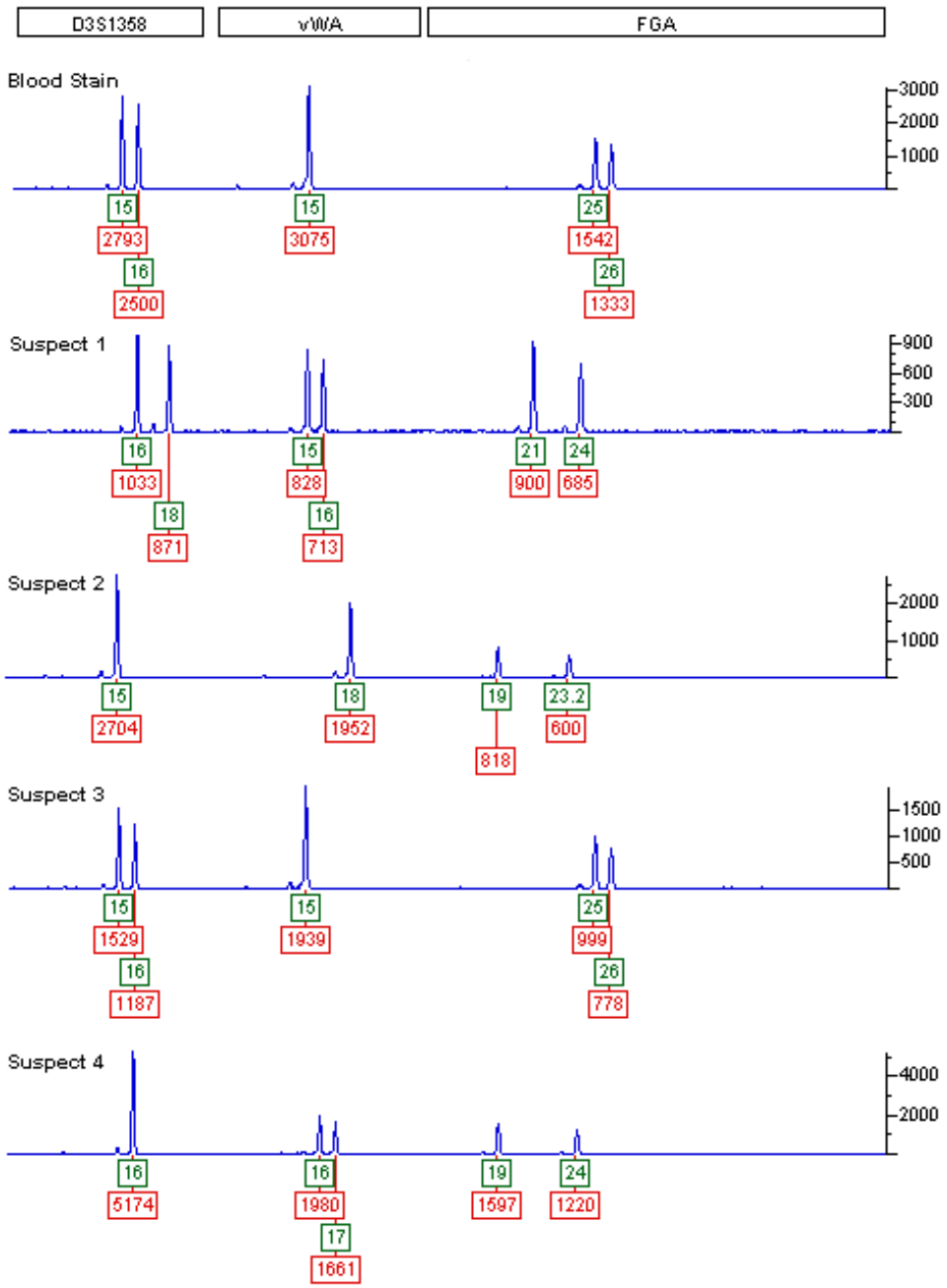

Figure 2.4: **A sample case involving a blood stain and four suspects. Suspect 3 is the only person who matches the blood stain in all locations shown. Therefore, he or she cannot be excluded as the donor of the sample. The profiles of suspects 1, 2, and 4 differ from the blood stain in at least one location, meaning they are excluded as possible donors.**

forward process. If every allele in a suspect's reference sample is present in an evidentiary sample, the suspect cannot be excluded as a possible contributor from the evidentiary sample. Simply put, there is no way to prove that the suspect did not leave his or her DNA on the item of evidence. Barring testing problems, if a single allele does not match the evidence sample, the suspect can be excluded from consideration as being a possible contributor to the sample. Figure 2.4 shows the electropherograms for an evidentiary sample (a blood stain) and four possible suspects. Only the third suspect matches the blood stain at all of the loci present, and thus is the only person that cannot be excluded as being a possible contributor to the evidence. While suspect three "matches" the evidence at the loci shown in Figure 2.4, a mismatch at any additional loci should be sufficient to qualify as an exclusion.

2.4 Ascertaining the weight of a DNA match

A complete DNA profile match means that an individual cannot be excluded from being a potential contributor to an item of evidence. In order to ascertain the weight associated with a DNA match, it is necessary to know the probability of observing a coincidental match. Several statistical methods have been developed to answer this question in a variety of situations. Due to differences in the methods presented and resulting controversy that developed, the following statistical sections are presented in chronological order to offer a historical perspective.

2.4.1 The first National Research Council report

In order to facilitate adoption of national standards for the review of DNA evidence, the National Research Council (NRC) wrote a report in 1992, titled "DNA technology in forensic science" (National Research Council, 1992). The report, often referred to as NRC I, lays the foundation of how DNA testing should be performed and how testing results should be evaluated. While the report appeared during the time when the earlier DNA typing kits using restriction fragment length polymorphism (RFLP) and variable nucleotide tandem repeat (VNTR) markers (utilizing 3-5 loci), it is written with the anticipation of more complex DNA testing technologies on the horizon. The text is written from the perspective that "estimates used in forensic science should avoid placing undue weight on incriminating evidence."

Statistical methods

The central question with any forensic evidence "match" is determining what weight to associate with a given piece of evidence. Specifically, determining the chance of a coincidental match. When using as few as three loci, it may be possible to determine the relative rarity of a DNA profile in the population simply by using the counting method: determining how often a DNA profile is observed in a given database. As more loci are utilized, the probability of observing a given profile quickly dwarfs even the largest currently-available DNA databases. Therefore, theoretical statistical models must be used to assess the weight of a DNA match.

The random match probability

Almost all DNA match statistics are based on the random match probability (RMP), which attempts to determine the probability of selecting a randomly-chosen unrelated individual from a given population that possesses the same DNA profile observed on a given piece of evidence. The RMP is based on the product rule, which assumes that all loci are independent (in linkage equilibrium) and that alleles are inherited randomly and are therefore independent as well. Allele frequencies are determined by examining the frequency of a given allele in a given DNA database population. States often create their own frequency databases and the FBI has published a national frequency database (Budowle and Moretti, 1999).

A single source sample will exhibit at most two alleles at each locus. Let the observation frequency in a given population of the first allele be P and the second be Q. The probability of selecting a randomly-chosen unrelated individual from the chosen population exhibiting the alleles P and Q at a given locus is:

$$Heterozygote : P_{PQ} = 2pq \tag{2.1}$$

If the evidence exhibits a single allele P (a homozygous locus), the formula is:

$$Homozygote : P_P = p^2 \tag{2.2}$$

The overall frequency is obtained by multiplying the frequencies observed at each locus.

Population substructure

The core assumption of the random match probability is that we are calculating the probability of selecting a random *unrelated* individual from the population. In reality, most individuals share some level of relatedness. Population substructure is more pronounced in closed populations, such as Native Americans living on a reservation and Amish communities. Population geneticists have debated the relative influence of population substructure. Some believe its effect is negligible (Chakraborty and Kidd, 1991). Others believe that if population substructure is to be ignored, then its absence must be proven empirically (Lewontin and Hartl, 1991).

The "practical and sound approach" provided by the National Research Council is to perform the ceiling principle (Lander, 1991). A conservative estimate of a given profile frequency can be obtained by utilizing the highest frequency observed in any population. For example, consider a DNA profile match where the African American, Caucasian, and Hispanic populations are examined. For each allele in the profile, the highest frequency observed in the three databases is used in the random match probability calculation. The resulting match statistic will not overestimate the rarity for any given population and the RMP value is independent of a source population. An alternative is to randomly select 100 individuals from 15 to 20 populations and determining the highest observed frequency for each allele. If no frequency is higher than 5%, then a 5% lower bound is chosen as the allele frequency to avoid overestimating the rarity of the allele due to sampling error.

Mixture statistics

The number of individuals that could be consistent with a mixture is much higher than a single-source sample, so the random match probability is not sufficient. Consider a locus with only three alleles, (12, 13, 14). This locus could contain material from any of the following genotypes: (12, 12), (13, 13), (14, 14), (12, 13), (12, 14), or (13, 14). In addition, there is no concrete indication as to the number of individuals present with any of the six possible genotypes (Paoletti et al., 2005).

The first National Research Council report suggests utilizing the combined probability of inclusion (CPI) to assess the weight of a DNA match with a mixture of two or more individuals (Devlin, 1992; Ladd et al., 2000). The CPI determines all possible genotypes for a locus and adds their frequencies together.

$$CPI = A_I A_J \ldots A_N : P_{IJ\ldots N} = (P_I + P_J + \ldots + P_N)^2 \tag{2.3}$$

As with the random match probability, the overall frequency is obtained by multiplying the frequencies observed for each locus. The CPI is typically several orders of magnitude greater than the random match probability of the suspect's profile. Some testing laboratories choose to report the combined probability of exclusion (CPE). The CPE is simply 1 - CPI.

Cold hit statistics

A criminal investigation may contain evidence with no specific suspect in question. If DNA is obtained from the available evidence, a DNA database search is typically

performed. A database match is called a cold hit. The question then becomes, what is the weight of the DNA evidence? A standard investigation utilizes probable cause to place a specific person at the scene of the crime. Here, a search of thousands (and perhaps millions) of DNA profiles are being searched for a possible match. The danger of a false inclusion is much higher than a traditional investigation. The first NRC report suggests that the initial match forms probable cause. The weight associated with a cold hit is found from the RMP calculation derived from testing additional loci. It is possible to either use a subset of the loci used for testing (e.g. the nine Profiler Plus® loci) or additional loci found in other testing kits (e.g. Identifiler® and PowerPlex® 16).

Familial searches

During a cold hit investigation, it is possible to identify a close, but not perfectly-matching DNA profile in a database. In these instances, one may be inclined to examine the relatives of the close match to determine if one of them is the true perpetrator. Due to the fact that the actions of a relative created probable cause (his or her inclusion in the DNA database) and not the suspect himself, privacy concerns caused the first NRC report to frown upon these types of investigations.

> To put it succinctly, DNA databanks have the ability to point not just to individuals but to entire families including relatives who have committed no crime. Clearly, this poses serious issues of privacy and fairness. . . . [I]t is inappropriate, for reasons of privacy, to search databanks of DNA from convicted criminals in such a fashion. Such uses should be prevented both

by limitations on the software for search and by statutory guarantees of privacy.

2.4.2 The second National Research Council report

In 1996, the National Research Council published a second report on DNA typing titled, "The evaluation of forensic DNA evidence" (National Research Council, 1996). The second report, often referred to as NRC II, attempts to clarify the statements made in the first report to attempt to eliminate any existing controversy. New methods were established for calculating the weight of DNA evidence in several different circumstances, including mixtures and cold hits. Since the report was created for clarification, its scope is more narrow than the first report.

Population substructure

The ceiling principle proposed in the first NRC report resulted in much debate. Law enforcement generally considered the ceiling principle to be overly conservative. In order to correct for population substructure, a correction factor, called theta, was introduced into the random match probability calculations.

$$Heterozygote : P_{PQ} = 2pq(1-\theta) \tag{2.4}$$

$$Homozygote : P_P = p^2 + p(1-p)\theta \tag{2.5}$$

For large populations, a theta value of 0.01 is recommended. For small, isolated populations, a theta value of 0.03 is recommended.

Mixtures

The second National Research Council abandoned the idea of the combined probability of inclusion (CPI) to assess the weight of a DNA match to a mixture in favor of a likelihood ratio. The suspect's profile is used to determine the relative likelihood of obtaining a mixed profile that matches the evidence if one of the contributors is the suspect, compared to the likelihood of obtaining such a profile from a randomly-selected pair of contributors from the salient population.

Consider the case of observing four alleles at a particular locus: A_1, A_2, A_3, and A_4, with the suspect's profile being (A_1, A_2). The two scenarios being compared are whether the DNA profile came from the suspect and one other individual compared to the profile being generated from two random individuals. The first likelihood is calculated with the equation $2p_3p_4$ because it is assumed that two of the alleles are A_1 and A_2 and come directly from the suspect. The denominator is calculated by first determining one set of potential contributors. Consider the two contributors (A_1, A_3) and (A_2, A_4). The resulting probability is $(2p_1p_3)(2p_2p_4) = 4p_1p_2p_3p_4$. Since there are six possible combinations of two individuals each contributing two alleles, the total probability is $24p_1p_2p_3p_4$. Therefore, the likelihood ratio is as follows:

$$LR = \frac{2p_3p_4}{24p_1p_2p_3p_4} = \frac{1}{12p_1p_2} \tag{2.6}$$

A likelihood ratio greater than one indicates that it is more likely that a given mixture profile would be observed if the source were the suspect and one other person, rather than two randomly-selected unrelated individuals. A larger likelihood ratio further

adds confidence to that assumption.

Cold hits

The NRC I proposal of testing additional loci to determine the weight of a cold hit was also determined to be overly conservative. Instead, the statistic associated with a cold hit is simply the expectation on the number of random matches. That is, the random match probability multiplied by the number of individuals in the database. Consider a cold hit profile with a random match probability of one in one million. If the DNA database contains a million profiles, then it is likely that a match will be found completely by chance. In this case, the value of $RMP \times n$ is one.

Familial searches

With regard to familial searches, the second National Research Council effectively reversed the position of the first report.

> If the possible contributors of the evidence sample include relatives of the suspect, DNA profiles of those relatives should be obtained. If these profiles cannot be obtained, the probability of finding the evidence profile in those relatives should be calculated.

The calculations mentioned are modifications of the random match probability to determine the chance of selecting a related individual of the suspect with the same DNA profile.

$$Heterozygote : P_{PQ} = 2pq + 2(p + q - 4pq)F \tag{2.7}$$

$$Homozygote : P_P = p^2 + 4p(1-p)F \tag{2.8}$$

The factor for F is determined by the level of relatedness. For parent and offspring, F = 1/4; for half-siblings, 1/8; for uncle and nephew, 1/8; for first cousins, 1/16.

Calculating the chance of two siblings possessing the same DNA profile is different because siblings are bilineal (both inheriting two alleles from their parents) rather than the unilineal (inheriting a single allele) situations described above.

$$Heterozygote : P_{PQ} = \frac{1 + p + q + 2pq}{4} \tag{2.9}$$

$$Homozygote : P_P = \frac{1 + 2p + p^2}{4} \tag{2.10}$$

2.4.3 Balding and Donnelly's approach to cold hits

The first and second NRC reports differed in their approach to assessing the weight of a cold hit match. In 1996, David Balding and Peter Donnelly wrote a response to both NRC reports arguing for a third method to assess the weight of a cold hit (Balding and Donnelly, 1996). The method is often simply referred to as the "Balding and Donnelly approach," and relies on likelihood ratios to determine the weight of a database match. Under the ideal conditions of observing a single DNA profile match in a database, the weight of that match is not reduced by the number of individuals present in the database, as the second NRC report suggests. In fact, a cold hit will result in a match statistic that is more impressive than the standard random match probability used in a probable cause scenario. The rationale is that for every cold

hit, you not only show that there is one individual that is the likely source of the evidence, but also that there are many individuals who have been excluded from being a contributor.

The Balding and Donnelly likelihood ratio can be written as:

$$LR = \frac{\text{P(DNA evidence|suspect is source)}}{\text{P(DNA evidence|suspect is not source)}} \tag{2.11}$$

Depending on the assumptions, the Balding and Donnelly formula can be reduced to the random match probability (DNA Advisory Board, 2000b). Let the probability of randomly choosing an unrelated person from the appropriate population be p_x. The likelihood ratio can then be expressed as the likelihood of an evidence match given that the source is the same individual (H_s) vs. the likelihood of an evidence match given that the source could come from two different individuals (H_d). The resulting equation is as follows:

$$LR = \frac{\text{P(DNA evidence}|H_s)}{\text{P(DNA evidence}|H_d)} = \frac{p_x}{p_x \times p_x} = \frac{1}{p_x} \tag{2.12}$$

As a result, the cold hit statistic reported is often similar to the random match probability (RMP).

2.4.4 The DNA Advisory Board

The DNA Advisory Board (DAB) is the oversight committee for the FBI. The DAB primarily established quality assurance standards for DNA testing labs (DNA Advi-

sory Board, 2000a). They have also determined an additional method of reporting the statistics associated with a cold hit (DNA Advisory Board, 2000b). The DAB suggests reporting the standard random match probability for the suspect along with the database match probability ($RMP \times n$) described in the second NRC report.

2.5 Issues with DNA testing

DNA testing, as in all scientific testing, is subject to issues that can adversely affect the analysis outcome (Thompson et al., 2003a; Thompson et al., 2003b). The process of sampling handling, analysis, and interpretation all constitute opportunities for errors to be introduced. What follows is a list of some of the most confounding issues.

2.5.1 Peak height imbalance

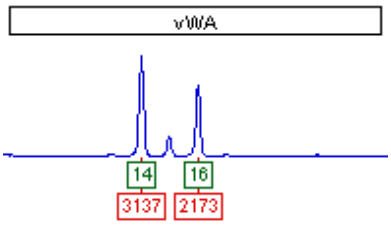

Figure 2.5: **Electropherogram exhibiting a peak height imbalance, identified by observing more than a 30 % difference in the peak heights of a heterozygote locus.**

The height of a peak is usually an accurate indication as to how much DNA is present in a sample, with the exception of degraded samples. As such, the amount of

DNA present from a single individual should be relatively constant, especially with fragments observed in the same locus. The resulting peaks should also be relatively equal with little variability. This assumption is supported in part by numerous validation studies (Frank et al., 2001; Holt et al., 2002; Leclair et al., 2004; Applied Biosystems, 2000a). General practice has found that "[t]he peak height ratio, as measured by dividing the height of the lower quantity peak in relative fluorescence units by the height of the higher quantity allele peak, should be greater than approximately 70% in a single source sample" (Butler, 2001). A large disparity between the two peaks at the same locus indicates that there are possibly two or more contributors to a sample (See Figure 2.5).

2.5.2 Mixtures

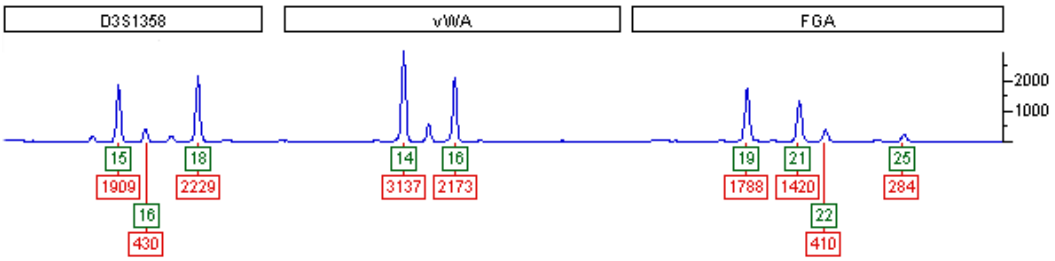

Figure 2.6: **An electropherogram of a mixture sample. Mixtures are identified by observing three or more alleles in a single locus. The additional peaks in the D3 and FGA loci indicate that at least two people contributed to this sample.**

Mixture samples involve two or more contributors and can be very difficult to interpret. Observing three or more alleles in a single locus is a clear indication that

a mixture is present (see Figure 2.6). However, allele counts alone are not a reliable indicator of the number of contributors to a sample for several reasons. First, a mixture can exhibit only one or two alleles in a single locus since people can have alleles in common, particularly if they are related. However, most mixtures exhibit three alleles in at least one of 13 loci (Paoletti et al., 2005). Second, stutter, pull-up, noise, and other artifacts can sometimes be interpreted as true alleles when they are in fact not. Artifacts can often be identified and removed from consideration, but there always remains the chance that an artifact could actually be hiding a true allele in its position.

Several methods have been proposed for the interpretation of forensic DNA mixtures (Curran et al., 1999; Clayton et al., 1998; Evett and Lambert, 1998; Evett et al., 1998; Gill et al., 1998; Perlin and Szabady, 2001; Wang et al., 2001). Many of these involve subtracting a known DNA profile from a mixture (such as the victim) before interpretation takes place. More detailed approaches enumerate the possible mixture combinations and evaluate them based on peak balance, assumptions of contributor profiles, and expected mixture ratios (Clayton et al., 1998). An extension of this approach attempts to eliminate further genotype combinations from the mixture by assessing the DNA contribution ratio for all contributors to a sample (the mixture proportion/ratio), and then minimizing the variance from this ratio among all contributors across all tested loci (Perlin and Szabady, 2001; Wang et al., 2001).

2.5.3 Contamination

One of DNA testing's greatest strengths is its remarkable sensitivity. The ability to obtain typable results from extremely small amounts of material also translates to the technology's greatest weakness. The mishandling of samples can inadvertently transfer DNA between pieces of evidence (Butler, 2001; Rudin and Inman, 2002). Every DNA test contains positive and negative control samples (DNA Advisory Board, 2000a). The positive control contains a sample with a known DNA profile and the negative control contains no DNA. Observing unexpected peaks in the positive control or any peaks in the negative control indicates possible contamination of the samples in question as well. If the original evidence samples have been mishandled and DNA transfer has occurred before the DNA has been amplified, the DNA test may not provide any clues as to possible contamination.

2.5.4 Degradation

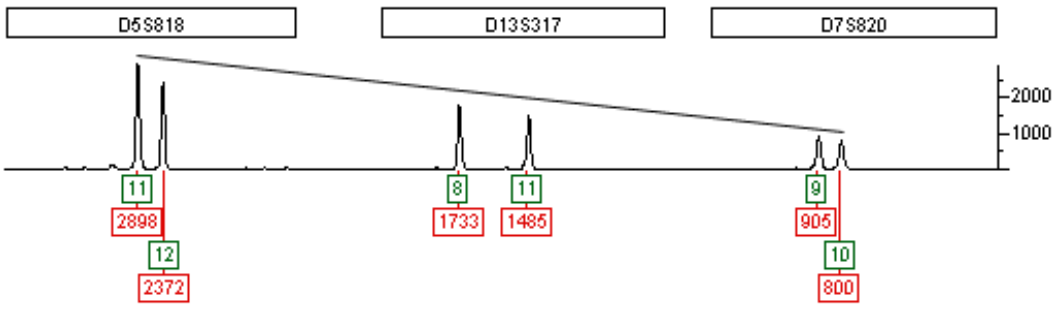

Figure 2.7: **An electropherogram of a degraded sample. Degradation is marked by observing progressively falling peak heights as the size of the DNA product increases.**

DNA can be degraded by being subjected to environmental elements, including UV sunlight and bacteria (Thompson et al., 2003a; Thompson et al., 2003b; Adams et al., 1991). Observing progressively smaller peak heights across loci in an electropherogram indicates that degradation may have occurred (see Figure 2.7). Larger DNA fragments are the first to be broken down because they represent the largest targets. In the figure, the peaks corresponding to the larger fragments contain less material and lower RFU peaks are observed toward the right side of the electropherogram. Smaller fragments are also affected, but usually to a lesser extent, thus higher peak heights are observed near the left side of the electropherogram.

Electropherograms of degraded samples skew the relationship between peak heights and amount of DNA present in a sample. One danger with interpreting degraded samples is the possibility of allelic dropout. If there is insufficient DNA to test, the fragment is not reported and consequently, the allele drops out. Difficulties in distinguishing between the effects of: physical damage to DNA molecules, the presence of chemicals that inhibit the PCR process (DeFranchis et al., 1988; Akane et al., 1994) and the fact that smaller DNA fragments are more efficiently amplified than larger ones during the PCR process (Walsh et al., 1992) have complicated efforts to develop objective standards of degradation.

2.5.5 Pull-up (bleed-through)

Pull-up, or bleed-through, occurs when the amount of fluorescence associated with a particular amplification product is so great that it saturates the photodetector

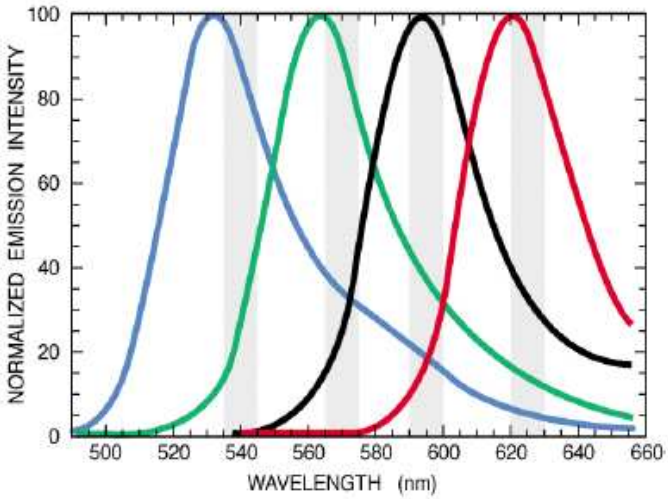

Figure 2.8: **The spectral calibration matrix for the Applied Biosystems genetic analyzers (Applied Biosystems, 2000b). Note the spectral overlap of blue, green, yellow, and red light. If the photo detector is saturated in one dye, light may be perceived in the overlapping spectra, resulting in pull-up peak artifacts.**

(Applied Biosystems, 2000b; Butler, 2001; Thompson et al., 2003a; Thompson et al., 2003b). The absorption spectrum of dyes used overlaps, meaning that when blue dye is seen, the green and yellow sensors also observe some of the light (see Figure 2.8). If enough of one dye is observed, the other dye sensors will record a substantial amount of light in their sensors and artifact peaks will be recorded.

Observing two peaks in two different dyes at the exact same time point can be indicative of pull-up peaks. Relatively large pull-up peaks are possible, which presents the danger of pull-up potentially being declared to be actual alleles. It is also possible for two valid peaks to be observed at approximately the same moment, so all instances of potential pull-up should be closely examined. Pull-up can sometimes be avoided by restricting the amount of DNA tested to ensure that the sensors are not saturated (Applied Biosystems, 2000b).

2.5.6 Stutter peaks

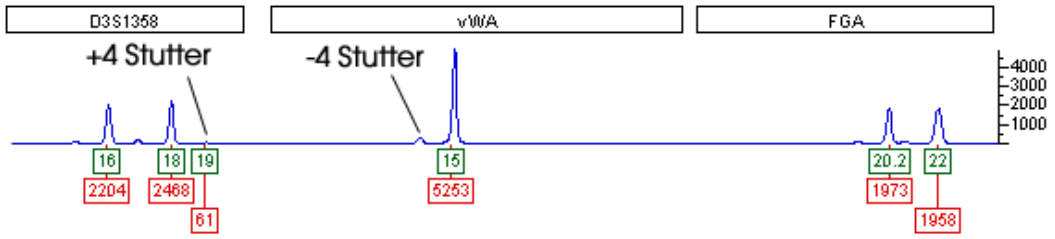

Figure 2.9: **An electropherogram exhibiting stutter alleles. -4 stutter peaks are the most common, occurring one repeat before the true allele. +4 stutter are more rare, occurring one repeat after the true allele.**

The process of DNA amplification via PCR can also result in artifact peaks (Ap-

plied Biosystems, 2000b; Butler, 2001). As the polymerase copies a strand of DNA during the PCR amplification, it is possible to slip forwards four base pairs, creating a smaller number of copies that contain one fewer repeat than the fragment being copied. The electropherogram will exhibit a small peak occurring one repeat before the actual fragment, called "-4 stutter" because the stutter peak is four base pairs shorter than the real DNA fragment. Conversely, the polymerase can slip backwards four base pairs during PCR amplification, creating a small number of fragments that are one repeat larger than the fragment being copied. The new artifact peaks occur directly after the real peak and are called "+4 stutter" (see Figure 2.9). It is also possible to observe +8 and -8 stutter, although they are relatively rare.

The main issue of stutter is that, in mixtures, it is possible for a second minor contributor to exhibit a profile that exists in a stutter position of the primary contributor. The result is that the primary contributor's stutter is labeled as such and is removed from the profile along with the minor contributor's allele. Mixtures, especially those with pronounced stutter, must be interpreted cautiously. Since it is often unapparent whether a peak is a stutter allele, a true allele, or both, frequency calculations, like the combined probability of inclusion (CPI), may include the stutter alleles in the set of possible combinations of DNA profiles present in a sample.

Stutter peaks can also result in a single-source sample being misclassified as a potential mixture due to the fact that three or more peaks are observed at a single locus. The most commonly used method to determine possible stutter is to use the stutter threshold cutoff of 15% (Butler, 2001). That is, any peak found in a possible stutter position and of a height that is less than 15% of the following peak is

categorized as possible stutter. GenoTyper® implements a -4 stutter filter, utilizing a specific set of stutter threshold values depending on the locus in which the stutter is observed.

2.5.7 Spikes, blobs, and other noise

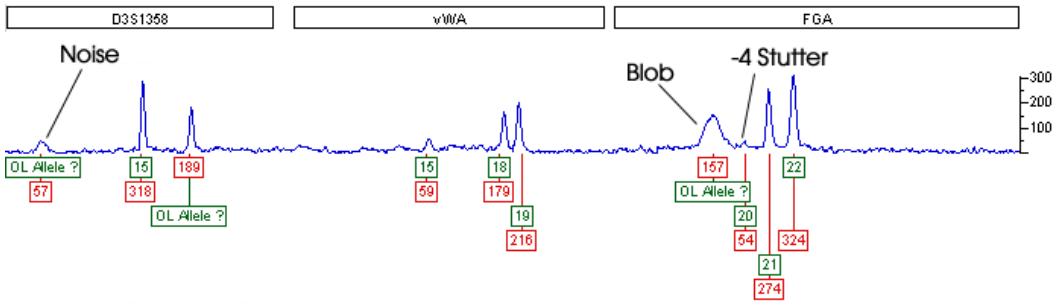

Figure 2.10: **An electropherogram exhibiting a blob, noise, and stutter alleles. Blobs are caused by large amounts of dye binding together. Stutter occurs due to slippage of the PCR polymerase. Noise can be due to dirt, bubbles, and other stochastic effects.**

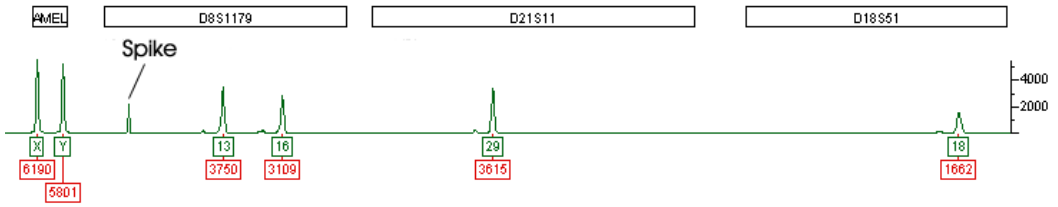

Figure 2.11: **Electropherogram exhibiting a spike, identified by observing a tall peak that is very thin.**

Dirt, air bubbles, urea crystals, and other contaminants can be present in a sample. The resulting electropherograms may contain anomalies (see Figure 2.10) (Applied

Biosystems, 2000b; Butler, 2001; Thompson et al., 2003a; Thompson et al., 2003b). Dye blobs typically occur when too many dye molecules have been added to a sample during the PCR amplification process. The unincorporated dye tags may bind together or to a contaminant and be detected by the genetic analyzer as a wide peak. Spikes are typically caused by the presence of particles in the polymer or voltage fluctuations occurring during the analysis process. Spikes often appear as very narrow peaks (see Figure 2.11). Dirt and other contaminants can also be detected by the genetic analyzer and may be presented as non-symmetrical peaks on the electropherogram. The danger with artifacts is that they may either be misinterpreted as actual alleles or they may, in the case of blobs and large noise, mask the presence of true alleles. Analysts generally examine peak shape in order to determine a peak's validity, but no definite objective standards are currently employed.

2.6 Data collection and analysis tools

After a DNA sample has been run through a genetic analyzer, its results are stored in electronic format. Sample analysis software, such as Applied Biosytems' GeneScan® and Genotyper® separates the raw data into dye channels separated by color, identifies potential peaks, and determines each peak's attributes and potential allele call. Peak detection algorithms may differ between software programs and even different versions of the same program (Gilder et al., 2004).

2.6.1 Genophiler®

Genophiler® is software tool that automates the operation of GeneScan® and Genotyper® and organizes and stores the testing results (Gilder, 2003; Ford et al., 2004). Genophiler® is made up of a suite of software programs written in Visual Basic (the graphical user interface), WinBatch (to automate the operation of GeneScan® and Genotyper®), and Perl (to collect, process, and organize the data). Genophiler® produces several output files in the form of HTML documents (for interactivity and cross-compatibility). These include the Quicklinks Navigator (providing access to all GeneScan® and Genotyper® analysis output), a summary table (providing a color-coded list of all samples and detected peaks), and a report (detailing potential issues found with each tested sample).

During the automation of GeneScan® and Genotyper®, Genophiler® creates a core data structure that stores all of the peak information for each tested sample. For each detected peak, the following information is stored:

> **Run name, Testing kit, RFU cutoff, Sample name, Dye color, Peak #, Time, Size, Peak height, Peak area, Data point, Locus, Allele call**

In addition, a second file contains user-entered data for sample classification and organization. For each tested sample, the following information is stored:

> **Run name, Sample name, Sample nickname, Show/Hide, Position, Defendant/Victim**

These data structures allow for automated large-scale analyses (Gilder, 2003; Gilder et al., 2004; Gilder et al., 2007b).

2.6.2 BatchExtract

The National Center for Biotechnology Information's (NCBI) BatchExtract software (National Center for Biotechnology Information, 2007) can be used to extract the the trace and peak data from Applied Biosystem's GeneScan® sample files. BatchExtract provides the height (in RFUs) of each data collection point (DCP) for each dye along a sample's electropherogram trace and is preented in the following form:

> Data collection point, Blue channel RFU, Green channel RFU, Yellow channel RFU, Red channel RFU, Orange channel RFU

BatchExtract also provides additional information associated with labeled peaks, including the data collection points where GeneScan® considered peaks to begin and end. The data contained in a sample's information window is also provided, including the date, time, and duration of analysis and the parameters used for electronic analysis.

Chapter 3

Using statistical distributions to classify forensic DNA profiling data

3.1 Preface

The work described here is also presented in the journal articles J. Gilder, T. Doom, K. Inman, D. Krane. "Run-specific limits of detection and quantitation for STR-based DNA testing." Journal of Forensic Sciences. 2007;52(1):97-101. and J. Gilder, T. Doom, M. Raymer, K. Inman, D. Krane. "Objective identification of degradation/inhibition in forensic STR-PCR DNA profiles." Journal of Forensic Sciences, to be submitted.

3.2 Introduction

The accuracy of statistical inference is highly dependent upon the number of samples observed. Many of the current methods used to identify potential technical artifacts and other anomalies in forensic DNA evidence rely upon internal laboratory validation studies performed using a small number of samples. With a large number of data points, distributions can be developed to classify what has been observed and determine a level of error associated with existing as well as new observations. Given a sufficient amount of data, the characteristics of clean, artifact-free electropherograms can be identified and parameterized. Distributions for such features as instrument noise can be estimated, and true signal can be differentiated from noise based on the observed distribution. A key advantage to this approach for identifying specific characteristics of DNA signal is that a statistical confidence can be assigned to the associated conclusions regarding the nature of the signal's source.

The heights of the peaks in an electropherogram are typically used to assess the quality of a DNA profile. Peak height is an approximation of the quantity of DNA present, so observing tall peaks indicates a "strong" profile. When dealing with reference samples consisting of abundant DNA in high quality, it is expected that a single strong profile will be observed. Difficulties arise when dealing with evidentiary samples due the variability in quantity of typable material, the quality of that material, and questions as to the origin of DNA found in a sample. Indications of contributors may be found at lower DNA quantities (and thus lower peak heights) and may provide important information in identifying the circumstances surrounding a DNA test

result.

Testing labs currently determine the level of background noise in the system through sensitivity studies performed during the initial validation process. An RFU threshold is then established that will be used for all future casework. A common value for this threshold is 150 RFUs. Changes in the operating environment over time (age of the genetic analyzer, chemicals, and level of staff experience) can lead to differences in the level of background noise in the system. A method is proposed to utilize the control samples tested in every run to mathematically determine the level of background noise in the system for a specific run using a distribution of points from the electropherogram that are not associated with amplified DNA product (i.e. signal arising from pure background noise). Run-specific RFU thresholds can then be derived that take into account the level of background noise in the system with a specified level of acceptable error.

Validation studies are often unable to deal with issues of sample integrity because they rely on the results of a small number of high quality samples. Samples tested in routine casework are often in less than optimal condition. The quantity of testable DNA product and quality of a sample's condition may pose additional issues with a sample's testing and interpretation. Environmental agents, such as moisture, bacteria, and UV sunlight can lead to the breakdown (degradation) of DNA molecules. Degraded samples have an increased risk of incomplete testing results (allelic dropout). Currently, no method is in place to objectively determine if a sample is potentially degraded. That judgment is left up to an analyst's "training, expertise, and experience." A method is proposed to determine if a given sample is statistically similar

to a distribution of established non-degraded samples. Samples that are statistically significantly dissimilar (with some level of specified error) from the population of non-degraded samples can be flagged for closer examination.

3.3 Run-specific limits of detection and quantitation for STR-based DNA testing

3.3.1 Introduction

STR-based DNA profiling methodology is effectively at the theoretical limit of detection in that typable results can be generated from as little starting material as a single cell (Findlay et al., 1997; Oorschot and Jones, 1997). However, one of the most challenging aspects of forensic DNA analysis is the interpretation of low-level testing results where it is difficult to reliably distinguish between noise and signal from template DNA that is associated with an evidence sample (Thompson et al., 2003a; Thompson et al., 2003b). This difficulty with minimal samples is often compounded by the consumptive nature of PCR-based DNA testing (Leclair et al., 2003; Fregèau and Fourney, 1993) when material is unavailable for replicate testing. Forensic DNA testing laboratories typically endeavor to minimize the effect of baseline noise and stochastic artifacts by relying upon very conservative minimum peak height thresholds (commonly fixed in the range of 50 to 200 relative fluorescent units; RFUs) that are established during the course of their validation processes (DNA Advisory Board, 2000a; Federal Bureau of Investigation, 2005; Scientific Working Group on DNA

Analysis Methods, 2000; Moretti et al., 2001a). However, the conservative nature of these commonly employed thresholds can also arbitrarily remove from consideration legitimate signal from trace and secondary contributors to an evidentiary sample - matters of critical importance in many criminal investigations.

Any measurement made with a light-detecting instrument, such as a genetic analyzer is subject to at least some level of background noise (Rubinson and Rubinson, 2000) - defined here as signal not associated with amplified DNA. Instrument-related factors that may contribute to background noise in DNA testing experiments are typically run-specific and include (but are not necessarily limited to): the age and condition of the polymer and capillary being used; dirty capillary windows; and dirty pump blocks (Applied Biosystems, 2000b). Background noise may also differ between instruments due to differences in CCD (charged couple device) detectors, laser effectiveness and alignment, and cleanliness and alignment of the optical components (Moretti et al., 2001a). Many amplification-related factors that contribute to background noise (such as analyst skill and stocks of chemicals) are also run-specific and might be reasonably expected to have varying impacts over time.

Many analytical disciplines aside from forensic DNA profiling have needed to rigorously account for background noise mixed with low levels of signal (Anderson, 1989; Thomsen et al., 2003). In the uncommon circumstances where background noise occurs at a constant level it can simply be subtracted from an analyzed signal to get true measurements of the tested material (Rubinson and Rubinson, 2000). It is much more common, however, for background noise, such as that associated with DNA testing results, to not be constant. In those instances, it is commonly

assumed that noise magnitude is independent of analyte signal and that noise levels are distributed in a Gaussian fashion that can be effectively characterized with a mean and a standard deviation (Rubinson and Rubinson, 2000; Anderson, 1989; Thomsen et al., 2003; Arinbuster et al., 1994). Two different signal-to-noise thresholds can be readily derived from the mean (μ) and standard deviation (σ) of the noise levels from a particular test and instrument: a limit of detection (LOD), and a limit of quantitation (LOQ) (Rubinson and Rubinson, 2000; Anderson, 1989; Thomsen et al., 2003; Arinbuster et al., 1994). The LOD is the smallest quantity of analyte that the analytical process can reliably detect. LOD is expressed as a statistical confidence limit of noise error, usually 99.7% (i.e. three standard deviations) or:

$$LOD = \mu_b + 3\sigma_b \tag{3.1}$$

where μ_b is the average amount of background noise and σ_b is the standard deviation associated with that value (Rubinson and Rubinson, 2000; Anderson, 1989; Thomsen et al., 2003; Arinbuster et al., 1994). The LOQ represents the threshold beneath which measurements of signal strength cannot be reliably used to determine the relative quantity of detected analyte (e.g. because such measurements may include an appreciable amount of signal arising from background noise). LOQ is commonly expressed as the average background signal plus ten standard deviations (Rubinson and Rubinson, 2000; Anderson, 1989; Thomsen et al., 2003; Arinbuster et al., 1994) or:

$$LOQ = \mu_b + 10\sigma_b \tag{3.2}$$

Forensic DNA testing laboratories routinely test a positive control, negative control, and reagent blank with every DNA analysis run (DNA Advisory Board, 2000a; Federal Bureau of Investigation, 2005; Scientific Working Group on DNA Analysis Methods, 2000). While these controls are utilized primarily as sentinels for gross failures of DNA testing processes, such as cross contamination of samples, as well as contamination or inappropriate activity of reagents, they also contain an abundance of subtle but important information about the running environment of the DNA testing system - particularly as it pertains to background noise. In this chapter section, a methodology is described that invokes generally accepted practices from other analytical disciplines and uses information associated with those ubiquitous controls to establish objective run-specific electropherogram peak height thresholds.

3.3.2 Materials and methods

Data set

Data for this study were obtained from 50 STR-based DNA testing runs generated by four analysts working at Forensic Analytical Specialties, Inc. (Hayward, CA) using the laboratory's validated standard protocols (e.g. no additional rounds of amplification were used as might be the case for low-copy-number analyses). All DNA profiles were generated with the Profiler Plus® commercial testing kit during the course of actual casework associated with approximately 150 cases conducted

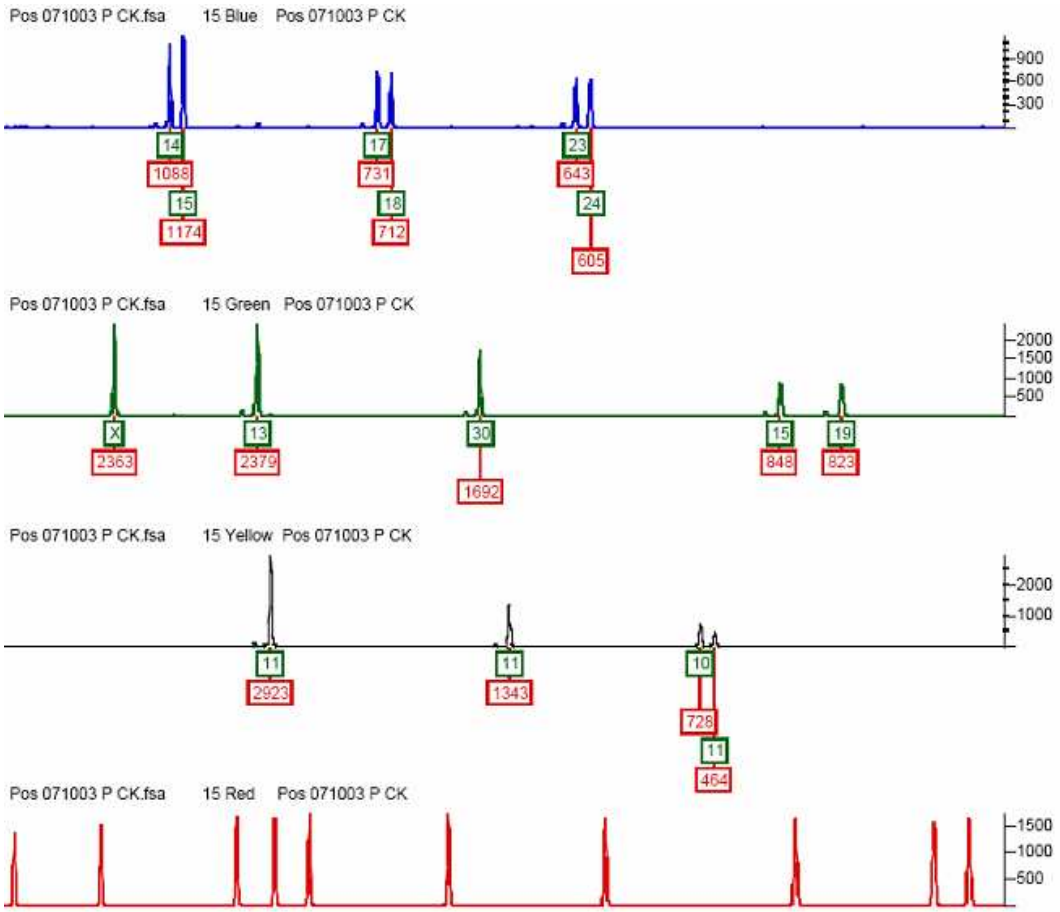

Figure 3.1: **The electropherogram for the 9947a positive control sample run using the Profiler Plus® test kit. To determine the average background signal, the peaks in the known profile are masked in all channels along with regions where -4 or +4 stutter peaks may occur.**

between 2004 and 2006. Each run was performed on the same Applied Biosystems 310 Genetic Analyzer and contained: a positive control; a negative control; and a reagent blank. A positive control consisted of template DNA from the 9947A immortal lymphoid cell line (Fregèau et al., 1995). This positive control DNA is provided by the manufacturer of the test kit and its STR genotype is well characterized 3.1. Negative controls begin at the amplification step and contain all of the reagents used for amplification (but no template DNA). A reagent blank is a sample that contains all of the reagents used from the beginning of the extraction of a sample through amplification and typing, but again containing no template DNA. When a single run contained more than one injection of a given control, the last injection was used. No other information associated with a run (e.g. that associated with reference or evidentiary samples) was used. Electronic data files associated with these control samples (with any case-specific information removed) are available on the Internet at: www.bioforensics.com/baseline/baseline.zip.

Baseline noise determination algorithm

The National Center for Biotechnology Information's (NCBI) BatchExtract software (National Center for Biotechnology Information, 2007) was used to obtain the trace and peak data from Applied Biosystem's GeneScan® sample files. BatchExtract provides the height (in RFUs) of each data collection point (DCP) for each dye along a sample's electropherogram trace. BatchExtract also provides additional information associated with labeled peaks, including the data collection points where GeneScan® considered peaks to begin and end. DCP regions containing a ROX size standard peak

were excluded (masked) from consideration in all dye colors to avoid any complications from spectral overlap artifacts (i.e. pull-up) (Butler, 2001; Applied Biosystems, 2000b). A total of 296,592 DCPs associated with the 50 negative controls ($\mu = 5,932$ DCP per run, $\sigma = 131$ DCP) and 297,315 DCPs associated with the 50 reagent blank controls ($\mu = 5946$ DCP per run, $\sigma = 87$ DCP) remained for inclusion in subsequent analyses after masking was completed. Similarly, DCP regions (plus and minus 55 DCPs to conservatively account for potential stutter artifacts) associated with the expected alleles for the 9947A immortal lymphoid cell line (Fregèau et al., 1995) were also masked in all dye colors for positive control samples. 120,762 DCPs associated with the 50 positive controls ($\mu = 2,415$ DCP per run, $\sigma = 198$ DCP) remained for inclusion in subsequent analyses after masking was completed.

Test mixture

A two-person mixture was created by combining the genomic DNA of two unrelated individuals with known genotypes in a ratio of approximately 10 to 1. The major contributor was known to be a female with the following STR-DNA profile: D3S1358 18, 18; vWA 16, 19; FGA 20, 21; D8S1179 13, 15; D21S11 32.2, 32.2; D18S51 15, 17; D5S818 11, 12; D13S317 11, 11; and D7S820 8, 10. The secondary contributor was known to be a male with an STR-DNA profile of: D3S1358 13, 17; vWA 17, 18; FGA 22, 24; D8S1179 11, 11; D21S11 28, 30; D18S51 12, 19; D5S818 11, 13; D13S317 10, 11; and D7S820 11, 12. The electropherograms for the mixed sample were generated with the same Applied Biosystems 310 Genetic Analyzer and protocols as those used to generate the control samples described above.

3.3.3 Results

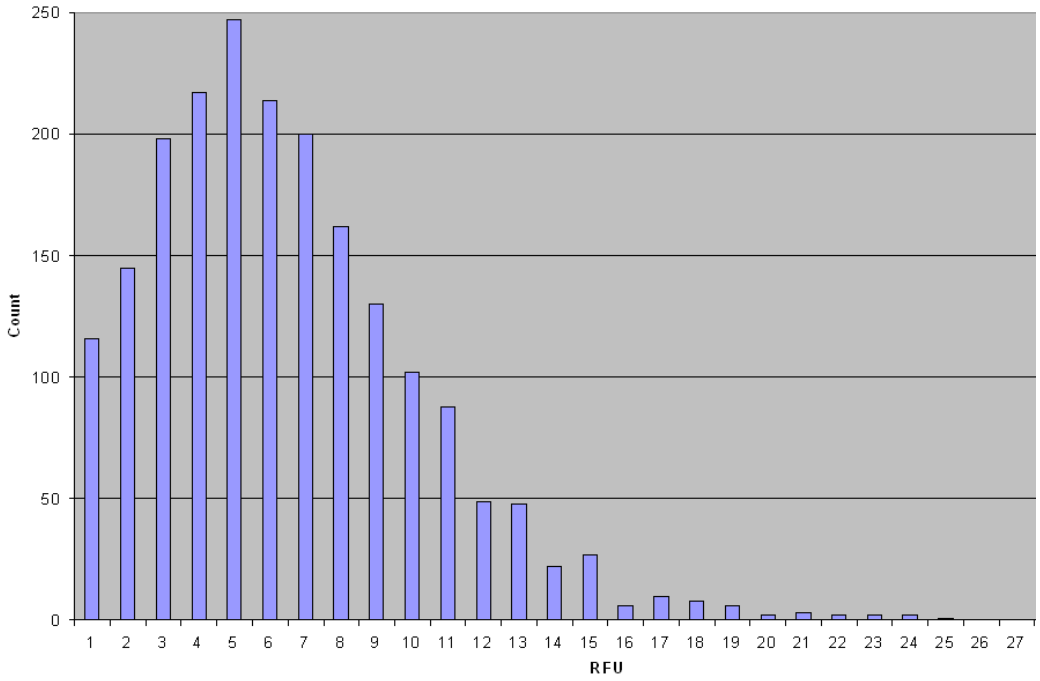

Figure 3.2: **A representative histogram taken from the distribution of measured RFU levels at all non-masked data collection points in the first of 50 negative control samples after masking. This distribution is from a blue channel and exhibits an average baseline approximately equal to that of the population's average baseline signal (5.5 RFUs).**

The distribution of baseline RFU level at each non-masked data collection point (DCP) was generally Gaussian for each of the 50 analyzed negative, reagent blank and positive controls (Figure 3.2). Histograms displaying the distribution of all three controls for all 50 runs included in this analysis can be found on-line at www.bioforensics.com/baseline/baseline.zip. Differences in the average baseline levels within each of the 50 analyzed runs were small between negative and positive control samples (with an average difference of the averages of only 0.60 RFUs). Differences

in the average baseline levels within each of the 50 analyzed runs were similarly small between negative and reagent blank controls (with an average difference of μ_b values of 0.41 RFUs) and between positive and reagent blank samples (with an average difference of μ_b values of 0.46 RFUs). While the inferred LOQ thresholds for all three controls were very similar within runs, average background noise values (μ_b) and standard deviations (σ_b) varied substantially between runs (Table 3.1) such that $\mu_b + 10\sigma_b$ (LOQ thresholds) derived from positive controls, negative controls and reagent blank controls ranged from: 27.7 to 75.7; 30.0 to 145.4; and 30.0 to 116.5 RFUs, respectively.

Positive Control	**Run Type**	μ	σ	$\mu + 3\sigma$	$\mu + 10\sigma$
	Maximum	6.7	6.9	27.4	75.7
	Average	5.0	3.7	16.1	42.0
	Minimum	3.7	2.4	10.9	27.7
Negative Control	**Run Type**	μ	σ	$\mu + 3\sigma$	$\mu + 10\sigma$
	Maximum	13.4	13.2	53.0	145.4
	Average	5.4	3.9	17.1	44.4
	Minimum	4.0	2.6	11.8	30.0
Reagent Blank	**Run Type**	μ	σ	$\mu + 3\sigma$	$\mu + 10\sigma$
	Maximum	6.5	11.0	39.5	116.5
	Average	5.3	4.0	17.3	45.3
	Minimum	4.0	2.6	11.8	30.0
Average of Baselines	**Run Type**	μ	σ	$\mu + 3\sigma$	$\mu + 10\sigma$
	Maximum	7.1	7.3	29.0	80.1
	Average	5.2	3.9	16.9	44.2
	Minimum	3.9	2.5	11.4	28.9

Table 3.1: **The maximum, minimum, and average baseline levels observed in the set of reagent blanks, negative controls, and positive controls (determined from controls in 50 different runs).**

All of the combined average limits of detection and quantitation fall below 100 RFUs. Baseline values were found to be generally homogeneous in that the minimum and average limits of detection and quantitation were within three standard deviations of each other for each of the 150 analyzed controls. The maximum values for μ_b were

generally similar in each of the three different control types, with a maximum observed difference within a run of only 8.8 RFUs (between a negative control and positive control). Single averages and standard deviations for each of the 50 analyzed runs were also generated by considering all DCP values for a run together (i.e. independent of which of the three different controls they came from). Standard deviations for these larger data sets were generally smaller than those observed when each of the three controls were considered separately though the calculated LOD and LOQ values were very similar to those obtained by considering the three controls for runs separately (Table 3.1).

A known mixed DNA profile from two unrelated individuals of an approximately 10:1 ratio was also examined using this methodology (Figure 3.3). The negative control tested in the same analysis run as the mixture yielded a limit of detection (LOD) of 29 RFUs and a limit of quantitation (LOQ) of 77 RFUs. Eleven alleles (including the Y allele at the amelogenin locus) associated with the known DNA profile of the minor contributor were not labeled for this mixed sample when a GeneScan® threshold of 150 RFUs was used. Eight alleles (including the Y allele at the amelogenin locus) associated with the male secondary contributor fall between the limit of quantitation and the commonly used 150 RFU threshold. Similarly, three additional alleles associated with the secondary contributor fall between the limit of detection and the limit of quantitation thresholds. The 17 allele (347 RFUs) at the D3 locus (which is in a stutter position relative to the major contributor's 3,509 RFU 18 allele at that locus) and the 10 allele (210 RFUs) at the D13 locus (which is in a stutter position relative to the major contributor's 2,670 RFU 11 allele at that locus) are the

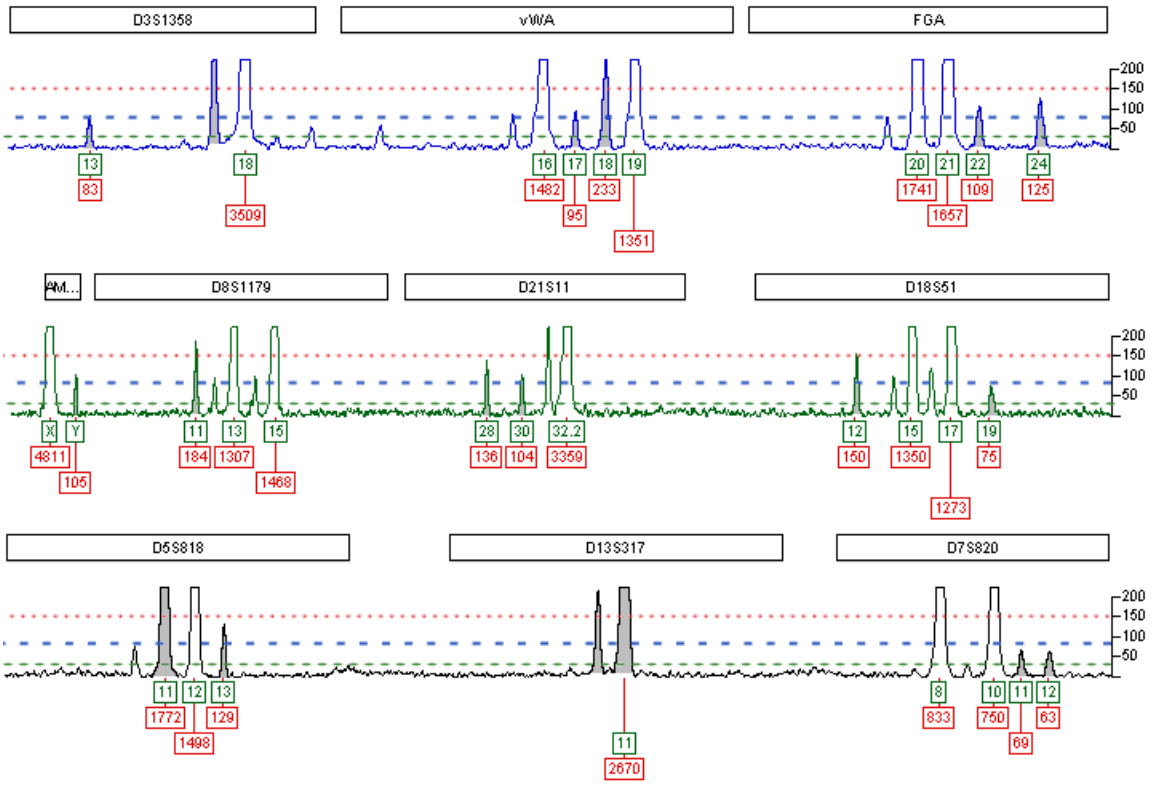

Figure 3.3: **Electropherograms from an approximately 10:1 mixture of two reference samples. Three different thresholds are shown: a minimum peak height threshold at 150 RFU (dotted line); a limit of quantitation (LOQ) threshold determined to be at 77 RFUs from the negative control for this electrophoresis run (dashed line); and a limit of detection (LOD) threshold determined to be at 29 RFUs for this electrophoresis run (small-dashed line). Genotyper assigned allele calls (with ABI stutter filters in place) are shown in boxes immediately below the electropherogram peaks while peak heights (in RFUs) are shown in boxes below those labels for all peaks with heights greater than the LOD. Peaks consistent with the known profile of the minor contributor are shaded.**

only allele of the secondary contributor that is not labeled by Genotyper® when the threshold is set to the limit of detection inferred from the negative control (29 RFUs) (Figure 3.3).

3.3.4 Discussion

The similarity of the baseline levels of samples that were expected to have a high signal amplitude arising from analyte (template DNA in the positive controls) and those expected to contain little or no analyte (the negative and reagent blank controls) indicates that noise magnitude in STR-based DNA testing is independent of the analyte signal. Baseline levels for each of the three different standard controls included in each DNA profiling electrophoresis run were also very similar within runs, but differed widely between runs. These observations suggest that the baseline noise associated with capillary electrophoresis of DNA profiles is comparable to that encountered in other analytical endeavors and that generally accepted means of determining limits of detection and quantitation can be applied.

The samples analyzed in this study were primarily positive, negative or reagent blank controls. It should be possible to evaluate evidentiary or reference samples included in the same capillary electrophoresis run with the LOD and LOQ values inferred from these controls. Any peaks in evidentiary or reference samples that exceed these thresholds (such as those associated with the secondary contributor in the mixture containing DNA of two unrelated individuals with known STR-DNA profiles; Figure 3.3) are unlikely to be due to baseline noise. All peaks above the

threshold would then require evaluation to ascertain whether they were signal from amplified genomic DNA, or if they may have originated from technical artifacts such as pull-up, voltage spikes or stutter.

It is worth noting that the maximum range of LOD thresholds (10.9 to 53.0 RFUs; Table 3.1) determined with this method in these 50 runs associated with casework performed by Forensic Analytical Specialties, Inc. is substantially below the minimum peak height threshold of 100 RFUs established by the laboratory during the course of their validation studies. Disregarding information associated with electropherogram peaks well above an analytical threshold of detection (and even above an analytical threshold of quantitation) might be considered abundantly conservative in some circumstances, given that DNA testing is a very sensitive process subject to a variety of technical artifacts such as pull-up, voltage spikes and stutter. However, in this abundance of caution, valid information about the presence of real DNA peaks is being discarded or ignored. In the instance of the mixture of two individuals with known STR-DNA profiles (Figure 3.3) the lower levels of the LOQ and LOD allowed reliable recognition of alleles arising from the genomic DNA of a secondary contributor while the commonly used 150 RFU minimum peak height threshold did not. In some investigations (e.g. a mixture of a victim and perpetrator that was small enough to require consumption of the entire sample) the observation of alleles associated with a secondary contributor using the LOD threshold methodology described here could constitute critically important information that would have not been available if only conservative minimum peak height thresholds were used.

The standard LOD/LOQ framework establishes a false positive rate for an indi-

vidual data point. It may be desirable to establish a false positive rate across an entire sample. The number of standard deviations necessary to maintain a fixed false positive rate can be calculated directly. First, the false positive rate per data point (p) for a given sample false positive rate (e.g. 5%) must be calculated:

$$(1 - \text{false positive rate}) \leq 1 - (1 - p)^n \tag{3.3}$$

The number of required standard deviations can be approximated by referencing a table of standard deviation confidence intervals or calculated directly from $z_{p/2}$ (using a standard normal distribution). For example, a typical analysis window consists of approximately 4000 data points per dye channel. If the overall sample false positive rate is chosen to be 5% and $n = 12000$, then the false positive rate for an individual data point is $p = 0.0002$ and approximately 3.7 standard deviations are required.

3.4 Objective identification of degradation/inhibition in forensic STR-PCR DNA profiles

3.4.1 Introduction

DNA is a relatively stable macromolecule and under certain circumstances has been known to persist for tens of thousands of years (Handt et al., 1994; von Wurmb-Schwark et al., 2003; Poinar, 1994). Samples of human DNA that are decades (Kevles,

2003; Gill et al., 1994; Ivanov et al., 1996) and even centuries (von Wurmb-Schwark et al., 2003) old have been amenable to genotyping for forensic purposes as well. However, the environmental conditions to which most evidentiary samples are exposed are usually much less conducive to the preservation of the information content of DNA molecules. Exposure to UV irradiation from sunlight, as well as to warm, moist environments have been found to result in degradation of DNA within a matter of hours (Adams et al., 1991). Evidence samples that begin with only trace amounts of DNA are particularly at risk of only being partially detected due to degradation and/or inhibition of PCR amplification.

STR-typing typically involves a polymerase chain reaction (PCR) amplification step followed by size fractionation of the resulting products and fluorescent signal detection and processing. While alternatives are available (Krenke et al., 2002; Moretti et al., 2001b), the separation of alleles from different STR loci is most commonly performed in the United States and Europe with Perkin Elmer-Applied Biosystems capillary electrophoresis equipment, such as the 310 and 3100 Genetic Analyzers (Moretti et al., 2001a). With both these (Fregèau et al., 1999; Wallin et al., 1999) and earlier typing systems (Adams et al., 1991; Holt et al., 2002) it has been widely observed that alleles corresponding to larger fragments of DNA typically exhibit weaker signals/intensity than smaller alleles after exposure to the environment, ostensibly because they provide a larger target for damage to be accumulated (Handt et al., 1994).

In the absence of degradation and stochastic effects due to small sample sizes, the amount of genomic template associated with any given locus in an evidence sample

should be equivalent (stoichiometric). Given that the amount of product generated during PCR amplification is generally proportional to the amount of starting template in multiplex reactions (Walsh et al., 1992), total peak height or area between alleles and loci should be roughly equivalent. As a result, progressively falling peak heights from small to large (left to right) DNA fragments on electropherograms are commonly considered to be an indication of degradation by forensic DNA testing laboratories. However, the absence of quantitative thresholds associated with these trends has made declarations of degradation subjective and commonly supported simply by an examiner's "past experience, training and expertise."

Difficulties in distinguishing between the effects of: physical damage to DNA molecules, the presence of chemicals that inhibit the PCR process (DeFranchis et al., 1988; Akane et al., 1994) and the fact that smaller DNA fragments are more efficiently amplified than larger ones during the PCR process (Walsh et al., 1992) have complicated efforts to develop objective standards of identifying potential degradation. We have directly addressed this issue by examining the trends in peak height relative to allele size using a best-fit linear regression for a set of 164 positive control samples. We use positive control samples as such samples are unlikely to have been affected by either degradation or inhibition. Evidence samples that display trends that are statistically significantly different from what is observed in this sampling of positive controls can be objectively described as being inconsistent and should be flagged for closer inspection. A case study where these objective thresholds are practically applied is also described.

3.4.2 Materials and methods

Underlying data

The data from electropherograms associated with 164 positive control samples that were completely genotyped at nine of the thirteen CODIS STR loci using the commercially available Profiler Plus® test kit were considered in this study. These 164 genotypings of the 9947A positive control were generated during the course of 44 different forensic investigations conducted between 1999 and 2003 by 23 different laboratories across the United States. Genomic template DNA from the 9947A immortal lymphoid cell line is included as a standard component of the Profiler Plus® test kits and contains both heterozygous and homozygous loci (Fregèau et al., 1995; Applied Biosystems, 2000a). Saturated samples (those with one or more peaks higher than 4,500 relative fluorescent units, RFUs) (Applied Biosystems, 2000b) are not considered. Similarly, only positive control samples where all expected peaks were observed to be greater than 200 RFUs are considered in order to minimize the contribution of stochastic effects. Peaks below 200 RFUs are subject to peak imbalance issues that could skew the trends observed in the data set.

Regression analysis

Linear regression, slope analyses, correlation of determination (r^2), and paired T-tests are performed using standard equations (Devore, 2000). Best-fit linear regressions are calculated using a total of six data points (two for each of three STR loci, Amelogenin was not included) for each of the three electropherograms (blue, green, and

yellow) associated with each sample. In other words, one data point is used for both the maternal and paternal contribution to genotype for both homo- and heterozygous loci. For heterozygous loci (D3S1358, vWA, FGA, D18S51, and D7S820 in the 9947A positive control) the y-coordinate is the height of each peak in RFUs while the x-coordinate is the peak's reported "data collection point." For homozygous loci (D8S1179, D21S11, D5S818, and D13S317 in the 9947A positive control) peak heights is assumed to be generally additive (a scatter plot of the average peak height at heterozygous loci vs. the average peak height at homozygous loci for these 164 genotypings had a best-fit linear regression of $y = 1.86x + 219.7$ with a correlation coefficient of $r^2 = 0.84$). Therefore, the height of each peak in RFUs at each homozygous locus is divided by two to determine y-coordinates comparable to those at heterozygous loci and paired with the x-coordinate (the peak's data collection point) two times to yield two data points for each homozygous locus. COfiler samples cannot be utilized as they lack sufficient data points to get slopes in all three dyes.

Normalized sums

Trends for each of the three sets of loci associated with each sample are considered additively after normalization of their values to assure equal weighting. Normalized values (mnorm) for the trends in each color were calculated separately with the following equation:

$$m_{norm} = \frac{m - m_{min}}{m_{max} - m_{min}} \tag{3.4}$$

where m is the observed value of the slope in a given color, m_{min} is the minimum slope observed in the sampling of 164 positive controls for that given color, and m_{max} is the maximum observed slope in the sampling of 164 positive controls for that color. The normalized sum is a single value which can be used to determine how consistent a given sample is with the sampling of positive controls.

3.4.3 Results

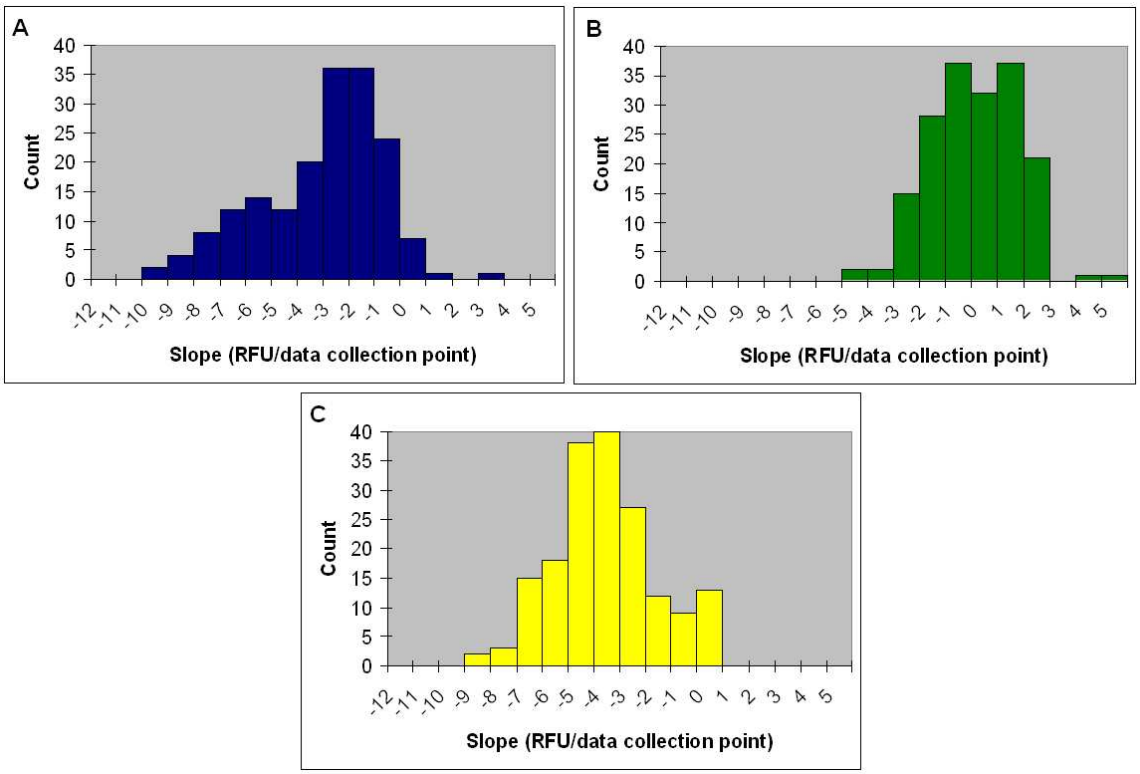

Figure 3.4: **Histograms displaying the distribution of observed slopes for each set of loci (blue, green and yellow) of the 164 positive control samples studied.**

The 164 slopes for the best-fit linear regressions generated for the positive control

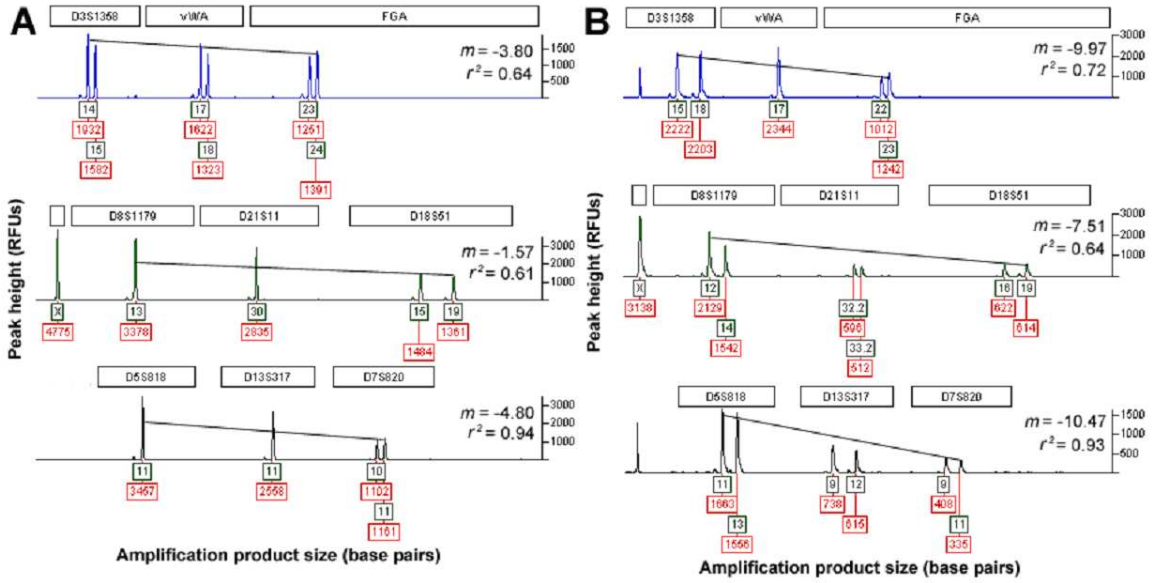

Figure 3.5: **Electropherograms associated with a high-quality genomic DNA template and with a genomic template that qualifies as being inconsistent with the sampling of positive controls. (A) Electropherograms from blue-, green- and yellow-labeled STR-amplification products associated with the genotyping of a 9947A positive control sample using the Profiler Plus® DNA typing kit. Boxes immediately above each of the three electropherograms indicate the loci being typed. Boxes immediately below each peak correspond to allele designations by Genotyper® for each locus while boxes below these designations display the observed height for each peak in RFUs. A best-fit linear regression line is shown for the heights associated with each allelic peak. The slope (m, in units of peak height RFUs/data collection point) determined for each regression line is shown on the right side of each electropherogram. (B) Electropherograms from a condom sample associated with a rape investigation. Electropherograms and regression lines are labeled as in (A).**

samples in this study appear to be approximately normally distributed for each dye color (blue, green and yellow) (Figure 3.4). The normal quantile plots for each dye appear to fall on a straight line, with correlation coefficient values of $r^2 = 0.96$, 0.99, and 0.98 for the blue, green, and yellow dyes, respectively (plots not shown). However, only the green distribution has evidence of normality using the Shapiro-Wilk test, with p-values of < 0.0001, 0.50, and 0.01 for the blue, green, and yellow dyes, respectively. Since the distributions appear to be nearly normal, the normal distribution has been used as the currently-available best estimate. Each of the six data points in each color for each sample generally contribute to internally consistent trends with average correlation coefficients of $r^2 = 0.66$ ($\sigma = 0.22$), 0.51 ($\sigma = 0.30$), and 0.93 ($\sigma = 0.08$) for the 164 sets of blue, green and yellow slopes, respectively as seen in a single typical example in Figure 3.5.

	Avg PH	**Avg PH Std Dev**	**Slope Avg**	**Slope Std Dev**	$\alpha = 0.05$	$\alpha = 0.01$
Blue	1324.72	512.44	-3.84	2.27	-7.58	-9.16
Green	1795.26	675.96	-1.02	1.67	-3.77	-4.93
Yellow	1575.30	588.99	-4.55	1.97	-7.81	-9.18
Normalized Sum			1.49	0.41	0.80	0.52

Table 3.2: **Summary statistics for best-fit linear regressions of peak height vs. data collection point for 164 positive control samples. Slope values are expressed in units of peak height RFUs per data collection point.**

The slopes in each of the three different colors only weakly correlate with each other ($r^2 = 1 \times 10^{-4}$, 0.29, and 0.11 for blue vs. green, blue vs. yellow, and green vs. yellow, respectively). PCR amplification product size is negatively correlated with signal strength in all three colors even in these positive control samples (Table 3.1). A paired T-test indicates that the distribution of slopes for the blue and yellow loci are dissimilar ($p < 1.4 \times 10^{-5}$) while the slopes for the green loci are generally less negative

than those for the blue loci ($p < 2.4 \times 10^{-26}$) and the yellow loci ($p < 4.5 \times 10^{-49}$).

The average and standard deviation values for each sampling of slopes is also used to determine thresholds for significance of departure at the $\alpha = 0.05$ and 0.01 levels (Table 3.2) for each of the three different data sets (blue, green, and yellow). Use of the threshold of significance at the $\alpha = 0.05$ or $\alpha = 0.01$ level indicates a classification error rate of 5% or 1%, respectively. Relatively few of the slopes observed in the 164 positive control samples fell beneath the $\alpha = 0.05$ significance levels (12 in blue, 8 in green, and 5 in yellow) and $\alpha = 0.01$ significance levels (3 in blue, 2 in green, and 1 in yellow).

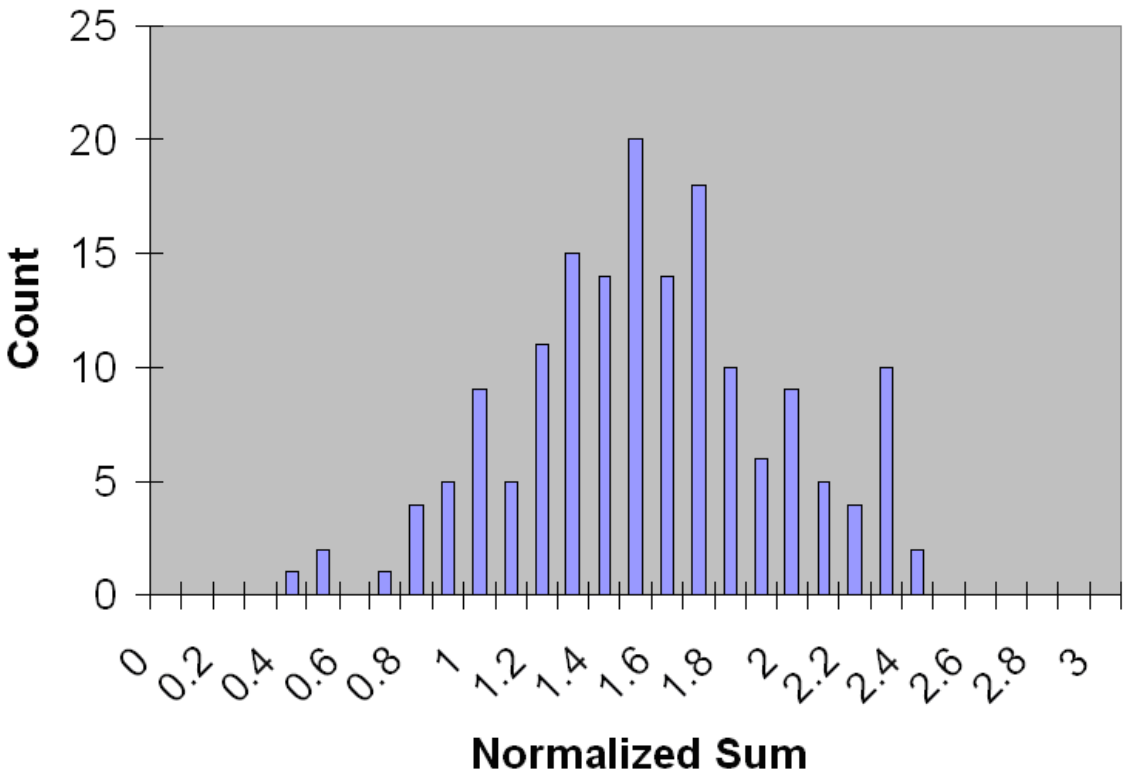

Figure 3.6: **A histogram displaying the distribution of the sum of normalized slopes for each of the 164 positive control samples studied.**

The 164 normalized sums of the slopes associated with the blue, green, and yellow STR loci for each positive control sample appear to be approximately normally distributed (Figure 3.6). The normal quantile plot appears to fall on a straight line, with a correlation coefficient of $r^2 = 0.99$ (plot not shown). However, the Shapiro-Wilk test only provides minimal support for normality, with a p-value of 0.06. Again, since the distribution appears to be nearly normal, the normal distribution has been used as the currently-available best estimate. The average normalized sum for the 164 positive controls is 1.49 with a standard deviation of 0.41 (Table 3.2). Samples with normalized sums of less than 0.67 would be significantly different from the general sampling of undegraded and uninhibited samples at $\alpha = 0.05$ while those with normalized sums of less than 0.41 would be significant at the $\alpha = 0.01$ level. Only eight of the 164 positive control samples were found to have normalized slopes that summed to less than 0.80 and three were observed to sum to less than 0.52 (sums = 0.49, 0.45, and 0.40).

3.4.4 Discussion

Genotypings of positive controls should exhibit little or no indications of degradation or inhibition. The genomic template (with known concentration, purity and source) for these controls is included as an integral component of the commercially available testing kits routinely used as part of standard forensic casework. Standard laboratory practices associated with the storage and use of these kits minimize the possibility that the positive control DNA or the reagents used to genotype would be compromised.

Further, the minimal quantities of reagents supplied with these test kits are intended to assure that they will be replaced frequently. Given that all of the positive control samples included in this study were run side by side with evidence samples associated with criminal investigations, it is unlikely that positive control DNA will have suffered appreciable degradation due to repeated freeze/thaw cycles (Ross et al., 1990) or exposure to the environment. However, it is still possible to observe various levels of degradation in positive control samples.

The slopes associated with the best fit linear regressions of each of the three sets of loci from these 164 positive controls appeared to be part of an approximately normal distribution (Figure 3.4) with relatively small standard deviations compared to their means (Table 3.2). The significant difference in the distribution observed for the loci labeled with green fluorescent dye relative to those labeled with blue ($p < 2.4 \times 10^{-26}$) or yellow ($p < 4.5 \times 10^{-49}$) dyes as well as the absence of significant correlations of slopes between colors suggests that the trends observed in each set of colors are independent. Interestingly, the slopes of all but 52 (2 blue and 50 green) of the 492 regression lines generated from these positive controls are negative (Figure 3.4). Given that the alleles from all tested STR loci are equally represented in the genomic template DNA associated with the 9947A positive control (Fregèau et al., 1995; Applied Biosystems, 2000a), these background negative correlations between signal strength (peak height) and amplification product size (data collection point) are consistent with the observation that the PCR process itself tends to preferentially amplify smaller fragments (Walsh et al., 1992).

While thresholds of significant departures from the trends observed in these 164

positive control samples at the $\alpha = 0.05$ and $\alpha = 0.01$ levels were determined for each of the three colored fluorescent dyes independently (Table 3.2), they were also determined cumulatively for sums of the normalized slopes for the three sets of loci (Table 3.2). Randomly selected electropherograms from this sampling of 164 ostensibly undegraded samples and others like it are unlikely to exhibit electropherograms with best-fit linear regression slopes that fall beneath these thresholds in any or all of the three different sets of labeled loci.

Evidentiary samples may differ from the sampling of positive controls considered in this study in three important ways that might prevent direct application of the thresholds of significance that were determined. First, only positive controls where all peaks were between 200 and 4,500 RFUs were considered. Stochastic effects associated with low level peaks as well as unreliable determination of peak height in saturated samples could both have substantial impact on the slopes observed in evidence sample electropherograms. Second, by their nature and design, positive controls contain genomic template that is derived from a single individual. In contrast, evidence samples often constitute mixtures of the DNA of two or more individuals that can make it difficult to determine the exact contribution to observed peak heights for any single contributor - especially if one or more of the contributors' DNA profiles are unknown. Third, the positive control samples considered here do not exhibit consistent indications of degradation/inhibition. Evidence samples that have been exposed to the environment may in fact be degraded or inhibited to the point that allelic drop out may occur - especially for the alleles associated with the largest amplification products. Observation of alleles associated with these largest amplification

products with peak heights in excess of at least 200 RFUs in each of the three sets of fluorescently labeled STR loci should minimize concern that allelic drop out may have occurred. However, when evidentiary samples: 1) have peak heights that all fall between 200 and 4,500 RFUs, and 2) do not appear to be mixtures, it should be possible to objectively compare the observed individual and normalized sum slopes to this sampling of positive control samples and apply the same thresholds of significance to them.

It may be desirable to develop a threshold for similarity to a population of non-degraded samples using the positives controls in a given run or case. Doing so may better capture the state of the system, including the condition of the capillary, reagents, polymer, and genetic analyzer. However, there will not be enough data points for the positive controls in a single run to develop a meaningful threshold. Utilizing the positive controls for a large run or several runs carried out during the same time frame may be sufficient if one wishes to factor in the current conditions of the working environment.

3.4.5 Practical application

A Coroner's inquest into the death of Jaidyn Leskie provided an interesting opportunity to apply the thresholds indicative of degradation/inhibition determined from this study of 164 positive control samples. The Victoria Police Forensic Services Centre in Australia used the Profiler Plus® test kit to generate STR DNA profiles from two evidentiary samples associated with the deceased. The DNA profiles that

were detected were subsequently found to be consistent with the DNA profile of a rape victim associated with a distinctly separate investigation for which DNA testing was performed by the same laboratory within hours of one another. Evidence at the inquest suggested that the rape victim could not have been involved in the death of Jaidyn Leskie. The testing laboratory suggested that the correspondence between DNA profiles of at least seven (and as many as 12, after additional testing and review) STR CODIS loci associated with the evidence samples in the two cases may be a result of an "adventitious" (coincidental) match rather than due to contamination between the two analyses.

	Condom Max PH	**Condom Min PH**	**Condom Slope**	**Condom p-value**	**Pos Control Max PH**	**Pos Control Min PH**	**Pos Control Slope**	**Pos Control p-value**
Blue	2222	1012	-9.97	< 0.01	805	293	-4.05	> 0.10
Green	2129	512	-7.51	< 0.01	1091	487	-0.23	> 0.10
Yellow	1663	335	-10.5	< 0.01	883	334	-0.44	> 0.10
Normalized Sum			0.29	< 0.01			2.01	> 0.10

Table 3.3: **Slope values are expressed in units of peak height RFUs per data collection point. The positive control used is 9947A and is analyzed on the same instrument and at the same time as the evidentiary sample.**

The small quantities of template available for PCR amplification from the investigation samples associated with the deceased, coupled with apparent degradation/inhibition resulted in several peaks associated with the largest amplification products falling below 200 RFUs. One issue raised during the course of the Coroner's inquest was whether the most likely source of contamination from the rape investigation (an unmixed sample of the complainant on a condom) also qualified as being degraded/inhibited. Comparison of the trends in peak height vs. data collection point for the condom sample (Figure 3.5) were found to be significantly different

than those of the sampling of 164 ostensibly non-degraded/inhibited positive control samples associated with this study or the positive control associated with the rape investigation (Table 3.3). The condom falls below the threshold for variance in the sampling of positive controls at the $\alpha = 0.01$ level. Thus, there is less than a 1% chance that a sample consistent with the sampling of positive controls (and thus presumably undegraded) would, by chance, exhibit the significant difference noted in the condom. The Coroner ruled that the DNA in the Leskie investigation was caused by contamination.

> The match to the bib occurred as a result of contamination in the laboratory and was not an adventitious match. The samples from the two cases were examined by the same scientist within a close time frame (Johnstone, 2006).

Chapter 4

Resolution of forensic DNA mixtures

4.1 Preface

The work described here comes from the journal article submission: J. Gilder, T. Doom, M. Raymer, K. Inman, D. Krane. "Resolution of forensic DNA mixtures." Journal of Forensic Sciences, to be submitted.

4.2 Introduction

In the case of a high-quality, single source sample and barring the possibility of error, STR analysis can provide compelling statistical evidence that an observed correspondence between an evidentiary sample and a particular individual is very unlikely to be the result of coincidence (National Research Council, 1996). However, many eviden-

tiary samples are comprised of mixtures of two or more individuals' DNA and their interpretation can be significantly more challenging (Paoletti et al., 2005). Consider a locus where three alleles (such as the D3S1358 locus in Figure 1 with a 13, 17 and 18 allele) are observed. Even if it is known that exactly two persons contributed genetic material to this sample, six different pair-wise combinations of genotypes are qualitatively consistent with the observation of these three alleles: (1) 13, 13 and 17, 18; (2) 13, 17 and 18, 18; (3) 13, 17, and 17, 18; (4) 13, 17 and 13, 18; (5) 13, 18 and 17, 17; and (6) 13, 18 and 17, 18. Interpretation becomes even more difficult when no assumption regarding the number of contributors to a mixed DNA sample is made (e.g. the three alleles observed at the D3S1358 locus in Figure 4.1 could represent a mixture of three individuals with genotypes: 13, 13; 13, 17; and 17, 18). Unfortunately, the potential for alleles to be shared between individuals limits the ability of simple counting techniques to correctly infer the number of contributors to mixed samples (Paoletti et al., 2005). All interpretation methods involve the analyst forming a working hypothesis regarding the apparent number of contributors. Methods which explicitly state the probability of error in the working hypothesis are far superior to those in which such error is implicit.

The rarity of single source samples among unrelated individuals is commonly estimated as a random match probability (RMP) using the following equations:

$$homozygotes : A_iA_i : P_{ii} = p_i^2 + p_i(1 - p_i)\theta_{ii} \tag{4.1}$$

$$heterozygotes : A_iA_j : P_{ij} = 2p_ip_j(1 - p_i)\theta_{ij} \tag{4.2}$$

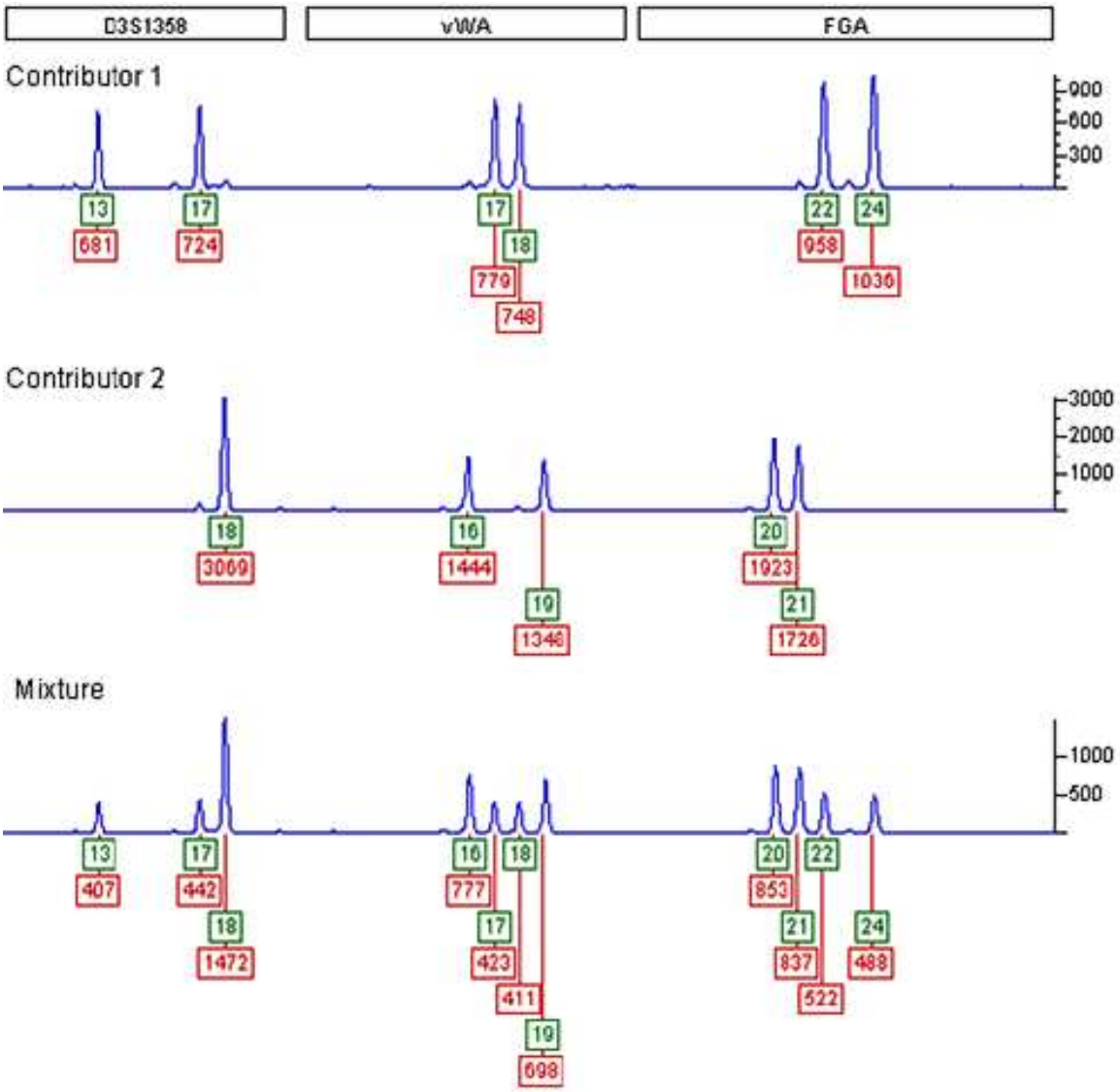

Figure 4.1: **Partial electropherograms of two single source samples and their corresponding 1:2 mixture. Allelic designations for each peak appear immediately below it with corresponding peak height information (in relative fluorescence units, RFUs) immediately below that. Locus names are in the boxes above.**

where p_i and p_j are the frequencies of alleles in a relevant population of alternative suspects and θ is an allowance for population substructure (National Research Council, 1996). The numerous alternative hypotheses associated with mixed samples should have a marked impact on the probative value of mixed evidentiary samples. The chance of a randomly selected, unrelated individual not being excluded as a possible contributor to a mixed evidentiary sample (the combined probability of inclusion or CPI) (National Research Council, 1992) is arrived at with the equation:

$$CPI = A_i A_j \ldots A_n : P_{ij\ldots n} = (p_i + p_j + \ldots + p_n)^2 \tag{4.3}$$

where p_i through p_n are the frequencies of alleles in a relevant population of alternative suspects (National Research Council, 1996). Typical RMP values for single source samples that have been genotyped at 13 standard CODIS STR loci are in the range of 1 in 1012 to 1050 (Butler, 2001) while CPI values for 13 locus mixed STR genotypes where no more than four alleles are observed across all tested loci (and thus consistent with a two person mixture) are typically in the range of 1 in 106 to 1010. For example, in the case of the D3S1358 locus shown in Figure 4.1, the RMP value for "contributor 1" using a Caucasian database (Budowle et al., 1999) is 1 in 192 and is 1 in 38 for "contributor 2" as single source samples, while the CPI value for their mixture is 1 in 7.

The striking difference in the weight of the DNA evidence associated with single source and mixed evidentiary samples has motivated the development of approaches that attempt to elucidate the genotypes of the individual contributors from mixed

evidentiary samples (Curran et al., 1999; Clayton et al., 1998; Evett and Lambert, 1998; Evett et al., 1998; Gill et al., 1998; Perlin and Szabady, 2001; Wang et al., 2001; Weir et al., 1997). These existing approaches have generally attempted to formalize and objectify a series of *ad hoc* rules employed by DNA analysts tasked with assessing which peaks at each tested locus are associated with each other and, thereby, with individual contributors. The most obvious approach to resolution involves grouping pairs of alleles according to their respective peak heights. For example, in the mixture at the D3S1358 locus in Figure 4.1, the heights of the 13 and 17 allele peaks are similar to each other, whereas the 18 allele is approximately three times the size. This approach is ultimately based on the assumption that a pair of peaks from a heterozygote should contribute relatively equal amounts of DNA. This assumption is supported in part by numerous validation studies (Frank et al., 2001; Holt et al., 2002; Leclair et al., 2004; Applied Biosystems, 2000a) that suggest that when an individual is heterozygous at a locus the peak heights of the alleles tend to be within a certain percentage of each other. An extension of this approach attempts to eliminate further genotype combinations from the mixture by assessing the DNA contribution ratio for all contributors to a sample (the mixture proportion/ratio), and then minimizing the variance from this ratio among all contributors across all tested loci (Perlin and Szabady, 2001; Perlin, 1999). Erroneous assignment of peaks to contributors can occur as the result of potentially incorrect assumptions (e.g. peak heights are strictly additive, similar amounts of genomic template will yield similar peak heights, artifacts can be reliably identified, mixture ratios are constant across all loci) and/or complications arising from similar amounts of DNA being contributed by two or more

individuals (in which case the resulting assignment may not be any more likely than other potential assignments).

The reality of mixed STR DNA profiles is that some loci cannot be resolved into two single genotypes because the observed electropherogram data provides equivalent or very similar support for two or more of the competing alternative hypotheses of genotype combinations that could account for all the detected alleles. This manuscript describes and tests a novel methodology that provably determines which alternative hypotheses of genotype combinations are mathematically feasible (in light of peak height balance and additivity expectations) and which should be eliminated from consideration due to its failure to satisfy one or more objective rules. The approach rests primarily on the same two principal assumptions of existing resolution methods: (1) that the number of contributors is known (or explicitly hypothesized) and (2) that alleles from the same individual will be present at approximately the same intensity ("in balance" within a specified margin of error). Each locus is considered separately. When only a single combination of genotypes is supported by the underlying data, RMP calculations can be used to describe the rarity of those individual genotypes. In instances where some but not all alternative hypotheses of contributor genotypes can be eliminated from consideration, a CPI-with-constrained-hypotheses value for a mixed DNA profile can be calculated.

4.3 Materials and methods

4.3.1 Satisfiability approach

Hypothesis	Contributor 1	Contributor 2	Mixture Condition 1	Mixture Condition 2
#1	P_4P_3	P_2P_1	$P_4 \leq c \times P_3$	$P_2 \leq c \times P_1$
#2	P_4P_2	P_3P_1	$P_4 \leq c \times P_2$	$P_3 \leq c \times P_1$
#3	P_4P_1	P_3P_2	$P_4 \leq c \times P_3$	$P_2 \leq c \times P_1$

Table 4.1: **The three genotype combination hypotheses that can explain the observation of four peaks at a single locus in a mixture of exactly two individuals. Since each contributor must contribute two different alleles, only peak height balance conditions need to be considered (there is no opportunity for additivity).**

For each locus, our approach to mixture resolution postulates all possible genotype combinations and tests each for compliance with the predicted conditions that must be satisfied in order for that genotype combination to be acceptable. Consider the case of there being exactly two contributors to a mixed sample. The n peaks present at a given locus are ranked by height and labeled: P_1, P_2, ..., P_n, where P_1 is a peak of minimal height and P_n a peak of maximal height. Name assignment is arbitrary for peaks of equal height. All potential contributor genotype combinations are then listed. For example, at a locus with four peaks (labeled P_1-P_4) the possible set of genotypes for two individuals that could explain the observation of all four peaks are: [(P_4, P_3), (P_2, P_1)], [(P_4, P_2), (P_3, P_1)] and [(P_4, P_1), (P_3, P_2)] (Table 4.1).

Individuals normally contribute two alleles per locus though it is possible for the two alleles to be indistinguishable from each other (i.e. homozygous). A second contributor to a mixture might posses a genotype at a given locus that includes one or both of the peaks in the first contributor's genotype. For any two individuals,

qualitatively, there may be 1 (two homozygous for the same allele), 2, 3, or 4 (two different heterozygotes) alleles.

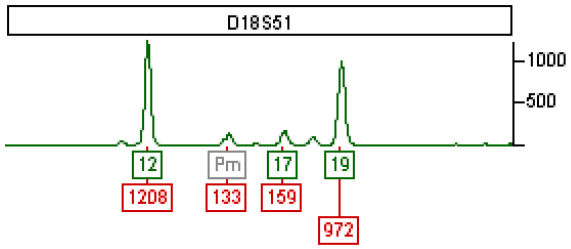

Figure 4.2: **A mixture containing a peak that exists below a minimum peak height threshold of 150 RFUs. The 15 allele at 133 RFUs falls below 150 RFUs so it is not reported by the DNA analysis software. The 17 allele at 159 RFUs can be paired with P_m at 150 RFUs, so the mixture resolution results in (12, 17) and (17, P_m).**

When three or fewer alleles are observed at a particular locus, it is sometimes also possible that alleles possessed by one or both contributor are present at levels below the detection capability of the equipment used for genotyping (allelic drop out). The label P_m is used to represent potential peaks below the minimum peak height threshold that may need to be considered in order to evaluate all possible contributor profiles (Figure 4.2). Minimum peak height thresholds are typically discrete values (often 150 relative fluorescent units; RFUs) that are determined in the course of a testing laboratory's validation studies (DNA Advisory Board, 2000a; Federal Bureau of Investigation, 2005; Scientific Working Group on DNA Analysis Methods, 2000; Moretti et al., 2001a). Genetic analyzers used for genotyping also have specific maximum measurement thresholds beyond which relative fluorescent unit levels are not reliably measured (saturation). For example, ABI genetic analyzer user manuals

specify that peaks greater than 4,000 RFUs in height are an indication of "Too much sample injected into capillary" (Applied Biosystems, 2000b).

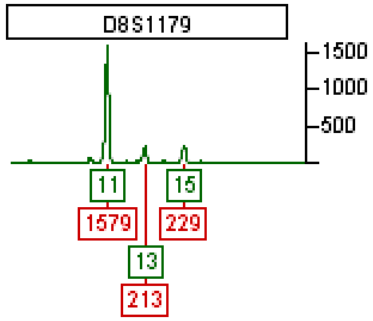

Figure 4.3: **Interpretation of three peaks at a single locus. The 11 allele is labeled P_1, the 15 allele is labeled P_2, and the 13 allele is labeled P_3. The peaks P_2 and P_3 may have arisen from the same source as their heights are within 30% of each other ($P_3 \leq 1.43 \times P_2$). However, the peaks P_1 and P_2 cannot represent alleles from the same contributor, as the peaks are not balanced (P_1 is more than 30% higher than P_2).**

These three observations [1) each individual contributes two alleles per locus, 2) some alleles may not be detected when present at low levels, and 3) saturation occurs above machine specific thresholds] allow a determination of all pairs of possible contributor profiles that can explain the observed data. Each possible pairing of contributor genotypes represents a hypothesis that is tested for satisfiability against determined conditions of peak height balance and additivity. Peak height balance, for example, demands that two peaks from the same contributor must have peak heights within a specific constant multiplier of each other (Butler, 2001; Frank et al., 2001; Holt et al., 2002; Leclair et al., 2004; Applied Biosystems, 2000a; Rudin and Inman, 2002). Thus, in order for a profile containing (P_2, P_1) to satisfy peak height balance,

it must be true that

$$P_2 \leq c \times P_1 \tag{4.4}$$

for the specific value of c appropriate for the measurement technology used in analyzing the sample (Figure 4.3. General practice has found that "[t]he peak height ratio, as measured by dividing the height of the lower quantity peak in relative fluorescence units by the height of the higher quantity allele peak, should be greater than approximately 70% in a single source sample" (Butler, 2001). Therefore, we use 1.43 as a representative value of c (representing a peak height ratio of 70%) in this study. Peak additivity assumes that the observed product from multiple contributors is approximately the same as the summation of each contributor's allele height (in RFUs) if each contributor was tested separately (Perlin and Szabady, 2001; Wang et al., 2001).

If either of the satisfiability conditions (peak height balance and additivity) fail for a given hypothesis, that hypothesis is removed from further consideration. If all but one alternative hypothesis for a given locus has been eliminated from consideration, then the remaining hypothesis represents an unambiguous genotype.

4.3.2 Derivation for loci with four observable alleles

If exactly two contributors are present and four alleles are observed at a locus, then all contributor alleles are accounted for and there is no possibility for additional peaks to exist below the minimum peak height threshold. Thus:

Hypothesis	Contributor 1	Contributor 2	Mixture Condition 1	Mixture Condition 2
#1	P_4P_3	P_2P_1	$P_4 \leq c \times P_3$	$P_2 \leq c \times P_1$
#2	P_4P_2	P_3P_1	$P_4 \leq c \times P_2$	$P_3 \leq c \times P_1$
#3	P_4P_1	P_3P_2	$P_4 \leq c \times P_1$	$P_3 \leq c \times P_2$

Table 4.2: **The three genotype combination hypotheses that can explain the observation of four peaks at a single locus in a mixture of exactly two individuals. Since each contributor must contribute two different alleles, only peak height balance conditions need to be considered (there is no opportunity for additivity).**

$$Saturation \geq P_4 \geq P_3 \geq P_2 \geq P_1 \geq P_m \tag{4.5}$$

Only three genotype combination hypotheses can account for the observation of four alleles at a single locus where there are exactly two contributors (Table 4.2).

Of the three possible interpretations in Table 4.2, only the first hypothesis leads to an unambiguously separable pair of genotypes (e.g. the mixture arises from the combination of an individual whose genotype is P_3, P_4 and another individual whose genotype is P_1, P_2). Since the peaks are ordered by their height, if solution row two is satisfied, then the first row is satisfied as well. Similarly, if the third solution row is satisfied, then all peaks are within the balance range of each other and all mixture combinations are possible. Thus, a four-peak locus from a two-contributor mixed sample can only be unambiguously resolved when the mixture conditions in row one of the table are satisfied, and those in rows two and three are not.

By producing a logical disjunction of each of these three satisfiability results and simplifying using the rules of Boolean algebra, a locus with four observed peaks can only be unambiguously resolved when:

$$(P_4 \leq c \times P_3) AND (P_2 \leq c \times P_1) AND (P_3 > c \times P_1) \tag{4.6}$$

This is intuitively obvious as, in order for only one hypothesis to be satisfied the highest and lowest peaks must be in balance with the next highest and lowest peaks (respectively) but out of balance with all other peaks.

4.3.3 Derivation for loci with three observable alleles

Hypothesis	Contributor 1	Contributor 2	Mixture Condition 1	Mixture Condition 2
#1	P_3P_3	P_2P_1	None	$P_2 \leq c \times P_1$
#2	P_3P_2	P_3P_1	$P_3 \leq c \times (P_2 + P_1)$	$P_3 \geq (1/c) \times (P_2 + P_1)$
#3	P_3P_2	P_2P_1	$P_2 \leq c \times (P_3 + P_1)$	$P_2 \geq (1/c) \times (P_3 + P_1)$
#4	P_3P_2	P_1P_m	$P_3 \leq c \times P_2$	$P_1 c \times P_m$
#5	P_3P_2	P_1P_1	$P_3 \leq c \times P_2$	None
#6	P_3P_1	P_2P_m	$P_3 \leq c \times P_1$	$P_2 \leq c \times P_m$
#7	P_3P_1	P_2P_2	$P_3 \leq c \times P_1$	None
#8	P_3P_1	P_2P_1	$P_1 \leq c \times (P_3 + P_2)$	$P_1 \geq (1/c) \times (P_3 + P_2)$
#9	P_3P_m	P_2P_1	$P_3 \leq c \times P_m$	$P_2 \leq c \times P_1$

Table 4.3: **The nine genotype combination hypotheses that can explain the observation of three alleles at a single locus in a mixture of exactly two individuals. Homozygotes are automatically in balance, so no peak balance calculations need to be performed. With only three observable alleles, one of the contributor's alleles is either shared with the other contributor or falls below the minimum peak height threshold (P_m). The value for P_m is the minimum peak height threshold for a given sample. When an allele is hypothesized as shared, it is necessary to sum their unshared contributions to determine balance with the shared allele.**

When only three alleles are observed at a locus in a two-contributor mixture, only two potential values for the "missing" fourth allele must be considered. As with equation 4, the three observable peaks satisfy the condition:

$$Saturation \geq P_3 \geq P_2 \geq P_1 \geq P_m \tag{4.7}$$

The fourth allele may be below an established minimum peak height threshold (P_m) and thus not reliably observed but this is only possible if such an allele would be "balanced" with an observed allele that has a peak height sufficiently close to the minimum peak height detection threshold. The value for P_m is the minimum peak height threshold for a given sample. Alternatively, the fourth allele may be indistinguishable from another allele and thus be represented as a single (and proportionally higher) observed peak. Consequently, nine different genotype combination hypotheses can account for the observation of three alleles at a single locus where there are exactly two contributors (Table 4.3).

When three alleles are observed at a single locus in a two-contributor mixture the implications of peak additivity must be considered. For example, in the hypothetical mixture of genotype P_3, P_2 with genotype P_3, P_1, the peak P_3 is shared by the two contributors. In this hypothesis, the first contributor accounts for a portion of the genetic material detected in allele P_3 that must be in peak height balance with allele P_2. Likewise, the second contributor accounts for the remaining material detected for allele P_3, and that amount must be in peak height balance with the P_1 allele. Thus, the height of allele P_3 must be in balance with the sum of the heights of P_2 and P_1. As it is not known whether the sum of the heights of alleles P_2 and P_1 is greater or less than the height of peak for allele P_3, both

$$P_3 \leq c \times (P_2 + P_1) \tag{4.8}$$

and

$$P_3 \geq (1/c) \times (P_2 + P_1) \tag{4.9}$$

must be true.

4.3.4 Derivation for loci with two observable alleles

Hypothesis	Contributor 1	Contributor 2	Mixture Condition 1	Mixture Condition 2
#1	P_2P_2	P_1P_1	None	None
#2	P_2P_2	P_2P_1	None	None
#3	P_2P_1	P_2P_m	$P_2 \leq c \times (P_1 + P_m)$	$P_2 \geq (1/c) \times (P_1 + P_m)$
#4	P_2P_m	P_1P_m	$P_2 \leq c \times P_m$	$P_1 \leq c \times P_m$
#5	P_2P_1	P_1P_1	$P_2 \leq c \times P_1$	None
#6	P_2P_1	P_2P_1	$P_2 \leq c \times P_1$	$P_2 \leq c \times P_1$
#7	P_2P_m	P_1P_1	$P_2 \leq c \times P_m$	None
#8	P_2P_2	P_1P_m	None	$P_1 \leq c \times P_m$
#9	P_2P_1	P_1P_m	$P_1 \leq c \times (P_2 + P_m)$	$P_1 \geq (1/c) \times (P_2 + P_m)$
#10	P_2P_1	P_mP_m	$P_2 \leq c \times P_1$	None

Table 4.4: **The ten genotype combination hypotheses that can explain the observation of two alleles at a single locus in a mixture of exactly two individuals. The first two mixture combinations are always possible and cannot be eliminated from consideration because of peak height imbalance or additivity constraints. If the two observed alleles are in balance, then five of the mixture combinations are possible. The remaining five mixture combinations rely on one or both of the observed alleles being in balance with the minimum peak height threshold (meaning that a third or fourth allele may not be reliably observable).**

When only two alleles are observed at a given locus in a two-contributor sample, potential values for two "unobserved" alleles must be considered. As a result, ten different genotype combination hypotheses can account for the observation of two alleles at a single locus where there are exactly two contributors (Table 4.4). Much like Equations 4.5 and 4.7, the two observable alleles satisfy the condition:

$$Saturation \geq P_2 \geq P_1 \geq P_m \tag{4.10}$$

Loci with only two alleles can never be completely resolved with the satisfiability approach described here in that it is impossible to determine if the larger of the two peaks has a heterozygous contribution to the second peak or if the observed alleles represent two homozygotes. The observed peaks must be sufficiently higher than the minimum peak height threshold in order to discount the possibility that the contributor of a particular allele also contributed alleles that are not observed because they fall below the minimum peak height threshold. The number of viable alternative hypotheses can also be narrowed if the two observed alleles are outside of the balance range of each other (they do not result from either heterozygous or roughly equivalent homozygous contributors).

4.3.5 Derivation for loci with one observed allele

Hypothesis	Contributor 1	Contributor 2	Mixture Condition 1	Mixture Condition 2
#1	P_1P_1	P_1P_1	None	None
#2	P_1P_m	P_1P_1	None	None
#3	P_1P_m	P_1P_m	$P_1 \leq c \times P_m$	$P_1 \leq c \times P_m$
#4	P_1P_m	P_mP_m	$P_1 \leq c \times P_m$	None

Table 4.5: **The four genotype combination hypotheses that can explain the observation of just one allele at a single locus in a mixture of exactly two individuals. The first two combination hypotheses can never be eliminated simply because of peak height balance and additivity constraints. If the observed allele is in balance with the minimum peak height threshold, then all four hypotheses are viable.**

Loci at which only one allele is observed in a two-contributor mixture cannot be resolved unambiguously though some alternative genotype combination hypotheses can be eliminated from consideration when the height of the single allele is out of balance with potential peaks below the minimum peak height threshold. Four different

genotype combination hypotheses can account for the observation of just one allele at a single locus where there are exactly two contributors (Table 4.5).

4.3.6 Testing the approach with known DNA mixtures

Mixed DNA profiles from two unrelated individuals with previously determined genotypes were generated using the Profiler Plus® testing kit with an ABI 310 Genetic Analyzer. Approximate mixture ratios of 5 to 1; 3 to 1; 1 to 1; 1 to 2; 1 to 3; 1 to 6; 1 to 10; and 1 to 28 were confirmed by an evaluation of average peak heights in the resulting electropherograms.

4.3.7 Additional considerations for stutter artifacts

The basic mixture resolution framework assumes that all peaks originate from amplified template DNA and are free of technical artifacts. The introduction of stutter product can increase the reported height of a peak in stutter position and alter the interpretation of a profile, particularly when dealing with low-level contributors. It is also possible for a contributor to be present in stutter position at a quantity small enough for the stutter filter to be invoked. Therefore, it is important to recognize when a potential contributor may be present in a stutter position and consider the appropriate alternative hypotheses.

While it is possible for a minor contributor's allele to be masked by stutter, one must be mindful of the basic assumptions of the number of contributors and the use of the minimum peak height threshold. The maximum number of observed peaks is equal

to twice the number of contributors. Therefore, the consideration of peaks in stutter position is mainly applicable when the number of peaks in non-stutter positions falls below the maximum number of expected alleles. It is also important to remember that peaks in a stutter position falling below the minimum peak height threshold are already considered with the P_m designation, which considers the presence of potential allelic dropout.

Case 1: a peak in stutter position that rises above the stutter threshold

Hypothesis	Contributor 1	Contributor 2	Mixture Condition 1	Mixture Condition 2
#1a	P_2P_1	P_3P_3	$P_2 \leq c \times P_1$	None
#1b	P_2P_1	P_3P_3	$(P_2 - s \times P_3) \leq c \times P_1$	$(P_2 - s \times P_3) \geq (1/c) \times P_1$

Table 4.6: **Expanding a mixture hypothesis to consider potential stutter. In this example, P_2 is in stutter position of P_3. The first hypothesis is unchanged. A second hypothesis is created that subtracts the expected level of stutter (s) of P_3 from P_2 to determine if P_2 is in balance with P_1.**

The range of expected stutter peaks has been well-researched (Holt et al., 2002; Leclair et al., 2004; Kinsey and Hormann, 2000). A peak that is large enough to avoid being filtered can either be unusually large stutter or the presence of another contributor. The height of a contributor's allele in stutter position may of sufficient height to make the contribution of stutter negligible. However, low-level peaks can be elevated to the point at which they are no longer balanced with the remaining contributor's peak. The latter scenario can be accommodated with the introduction of an additional mixture hypothesis that subtracts the contribution of stutter. Consider a two-person mixture with three peaks: P_1, P_2, and P_3, with P_2 in the stutter position of P_3. The mixture hypothesis of (P_1, P_2) and (P_3, P_3) can now be tested with the two

equations shown in Table 4.6. Hypothesis 1a follows the standard mixture resolution framework. Hypothesis 1b considers the contribution of stutter by subtracting the expected stutter percentage (s) of P_3 from P_2. If $(P_2 - s \times P_3)$ is less than the minimum peak height threshold, then hypothesis 1b is no longer necessary because the mixture hypothesis of (P_1, P_m) and (P_3, P_3) is already present in the original framework.

Case 2: a peak in stutter position that falls below the stutter threshold

Since stutter is relatively common, it is likely that most mixtures will exhibit peaks in stutter position that fall below the stutter threshold. If all of these peaks were considered to originate from potential contributors, then the mixture resolution framework would not be functional because the number of peaks would likely exceed twice the number of assumed contributors. In the circumstance where the number of observed peaks is fewer than twice the number of contributors, filtered peaks in stutter position can be considered one at a time.

For example, a two-person mixture with three labeled peaks will result in each filtered peak in stutter position being considered separately so that no more than four peaks are considered at a single time. A two-person mixture with two labeled peaks can have up to two filtered peaks in stutter position being considered at one time. Again, any peak in stutter position with a height less than the minimum peak height threshold does not need to be explicitly considered due to the presence of P_m in the original mixture resolution framework.

The interpretation of potential contributors in stutter position should be carried out with caution. Determining the precise contribution of stutter can be quite chal-

lenging. In certain circumstances, it may be more desirable to eliminate any potential error by simply considering all possible mixture contributors at a given locus (at the expense of potentially gaining information from a mixture resolution).

4.3.8 Statistical estimates

Statistical estimates of the rarity of observed DNA profiles in the known mixtures of two unrelated individuals were determined after the application of this approach to mixture resolution was applied. When the approach eliminated all but one hypothesized genotype combination at a given locus, the locus was designated "resolved" and random match probability (RMP) statistics for each genotype were generated using the allele frequencies reported by the FBI for US Caucasians (Budowle et al., 1999). In those instances where the approach eliminated some but not all hypothesized genotype combinations for a given locus, the locus was designated "constrained" and a combined probability of inclusion (CPI) was calculated. When the approach failed to eliminate any of the alternative hypotheses of genotype combinations for a locus the locus was designated "unconstrained."

The CPI for constrained loci only considers the alternative hypotheses of genotype combinations that have not been eliminated from consideration. For unconstrained loci, the CPI calculation is performed with all possible genotypes. For example, the D3S1358 locus from the 1:3 mixture ratio could not be fully resolved using $c = 1.43$ because the 13 and 17 alleles were found to be out of balance with each other and the remaining 18 allele. In this case, where P_i is the probability of observing allele i in

an unrelated random population, the CPI value considers all potential contributors to the mixture:

$$CPI_{unresolved} = (P_{13} + P_{17} + P_{18})^2 \tag{4.11}$$

For loci where the approach restricted the list of potential genotypes, only those genotypes are included in the CPI calculation. For example, at the D5S818 locus the 1:3 mixture ratio, all hypotheses for loci with three alleles are eliminated except hypotheses #2 and #5 (Table 4.3) corresponding to the two possible mixture combinations: [(11, 12), (11, 13)] and [(11, 12), (13, 13)]. Therefore, the partially resolved CPI only considers the genotypes (11, 12), (11, 13), and (13, 13) or:

$$CPI_{partiallyresolved} = 2P_{11}P_{12} + 2P_{11}P_{13} + P_{13}^2 \tag{4.12}$$

For fully resolved loci, only two genotypes can contribute to the mixture, so the CPI for the locus is the sum of the random match probabilities for those two genotypes. For example, at the vWA locus the 1:3 mixture ratio is found to have four alleles and all hypotheses for loci with four alleles are eliminated except hypothesis #1 (Table 4.2). The only genotypes consistent with the observed data corresponded to (16, 19) and (17, 18), so the mixture calculation becomes:

$$CPI_{resolved} = 2P_{16}P_{19} + 2P_{17}P_{18} \tag{4.13}$$

The cumulative CPI is achieved by taking the product of the CPI values calculated

for all loci. The result is a compromise between the random match probability for a contributor and the standard CPI mixture calculation which takes into account all possible genotype combinations without considering the quantitative aspects of the data.

4.4 Results

The two unrelated individuals whose DNA was mixed for the purposes of this study share only two alleles (an 11 allele at the D5S818 locus and an 11 allele at the D13S317 locus) across the nine polymorphic STR loci (D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317 and D7S820) genotyped for this study (Table 4.7). Four of the tested loci (vWA, FGA, D18S51, and D7S820) exhibited four alleles in the mixed samples. Four other loci (D3S1358, D8S1179, D21S11 and D5S818) had three alleles. Only one locus (D13S317) displayed two alleles and there is no locus that displays only one allele. Of the 80 instances in which one or both of the contributors to a mixed sample was a heterozygote and neither allele was present in the other contributor's genotype for that locus, eight instances of peak heights falling outside of the 70% peak height imbalance expectation were observed (average observed peak height ratio was 0.87, $\sigma = 0.12$). Homozygous peaks for a contributor were approximately twice the height of that contributor's heterozygous peaks for that same mixture (e.g. contributor 1's 11 allele at the D8S1179 locus and contributor 2's 32.2 allele at the D21S11 locus averaged 1.92 times the height of contributor 1's 28 and 30 alleles at the D21S11 locus and contributor 2's 13 and 15 alleles at the D8S1179

	D3	vWA	FGA	D8	D21	D18	D5	D13	D7
Contributor 1	13, 17	17, 18	22, 24	11, 11	28, 30	12, 19	11, 13	10, 11	11, 12
Contributor 2	18, 18	16, 19	20, 21	13, 15	32.2, 32.2	25, 17	11, 12	11, 11	8, 10
5:1	13 (1197)	16 (234)	20 (190)	11 (2422)	28 (1450)	12 (1208)	11 (1225)	10 (1125)	8 (90) 10
	17 (1245)	17 (1146)	21 (283)	13 (194)	30 (1398)	15 (133)	12 (250)	11 (1391)	(131) 11
	18 (607)	18 (1053)	22 (1241)	15 (154)	32.2	17 (159)	13 (1038)		(698) 12
		19 (152)	24 (1082)		(333)	19 (972)			(687)
3:1	13 (644)	16 (270)	20 (255)	11 (1579)	28 (931)	12 (813)	11 (1105)	10 (779)	8 (110)
	17 (632)	17 (865)	21 (291)	13 (213)	30 (899)	15 (239)	12 (287)	11 (1061)	10 (156)
	18 (489)	18 (589)	22 (966)	15 (229)	32.2	17 (207)	13 (805)		11 (528)
		19 (210)	24 (955)		(409)	19 (741)			12 (554)
1:1	13 (504)	16 (515)	20 (554)	11 (1070)	28 (702)	12 (630)	11 (1204)	10 (706)	8 (231)
	17 (501)	17 (536)	21 (500)	13 (387)	30 (667)	15 (495)	12 (510)	11 (1314)	10 (266)
	18 (914)	18 (546)	22 (701)	15 (393)	32.2	17 (370)	13 (697)		11 (425)
		19 (421)	24 (672)		(984)	19 (563)			12 (355)
1:2	13 (407)	16 (777)	20 (853)	11 (763)	28 (537)	12 (515)	11 (1324)	10 (491)	8 (418)
	17 (442)	17 (423)	21 (837)	13 (603)	30 (507)	15 (723)	12 (792)	11 (1634)	10 (381)
	18 (1472)	18 (411)	22 (522)	15 (597)	32.2	17 (638)	13 (476)		11 (313)
		19 (698)	24 (488)		(1470)	19 (390)			12 (276)
1:3	13 (306)	16 (905)	20 (1195)	11 (468)	28 (339)	12 (344)	11 (1474)	10 (427)	8 (560)
	17 (448)	17 (264)	21 (1116)	13 (818)	30 (368)	15 (1038)	12 (1115)	11 (2138)	10 (567)
	18 (2272)	18 (371)	22 (408)	15 (848)	32.2	17 (936)	13 (318)		11 (225)
		19 (974)	24 (317)		(2172)	19 (268)			12 (214)
1:6	17 (115)	16 (590)	20 (648)	11 (187)	28 (74)	12 (106)	11 (923)	10 (146)	8 (338)
	18 (885)	17 (74)	21 (659)	12 (70)	30 (124)	15 (640)	12 (723)	11 (1021)	10 (326)
		18 (105)	22 (87)	13 (714)	32.2	17 (583)	13 (116)		12 (56)
		19 (524)	24 (114)	15 (670)	(1170)	19 (110)			
1:10	13 (83)	16 (1482)	20 (1741)	11 (184)	28 (136)	12 (150)	11 (1772)	11 (2760)	8 (833)
	18 (3509)	17 (95)	21 (1657)	13 (1307)	30 (104)	15 (1351)	12 (1498)		10 (750)
		18 (233)	22 (109)	15 (1468)	32.2	17 (1273)	13 (129)		11 (69)
		19 (1350)	24 (125)		(3359)	19 (75)			12 (63)
1:28	13 (82)	16 (1566)	20 (1791)	11 (90)	28 (77)	15 (1523)	11 (1879)	12 (2641)	8 (821)
	18 (3476)	18 (155)	21 (1653)	13 (1363)	32.2	17 (1443)	12 (1741)		10 (837)
		19 (1318)	22 (62)	15 (1324)	(3235)		13 (71)		11 (55)
			24 (64)						

Table 4.7: **Genotypes of two unrelated mixture contributors and observed peak heights (in RFU) for each at eight tested mixture ratios. The contributor profiles are followed by the results of eight mixture ratios tested. The mixture profiles list called alleles for each locus (followed by the height of the corresponding peak in RFUs). Mixture ratios are expressed in terms of genomic template from contributor 1 relative to that from contributor 2.**

locus across all mixture ratios) (Table 4.7).

The RMP among randomly selected, unrelated individuals for contributor 1 and contributor 2's nine locus STR profile using the FBI's Caucasian allele frequencies (Budowle et al., 1999) unmixed DNA profiles are approximately 1 in 59 trillion and 1 in 213 billion, respectively. The CPI value for an unresolved mixture of contributor 1 and 2's nine locus STR profile (Table 6) is approximately 1 in 28,200 randomly chosen, unrelated US Caucasians as determined from the FBI's allele frequency database (Budowle et al., 1999).

Application of our approach to mixture resolution to the mixed DNA profile information presented in Table 4.7 did not allow the complete resolution of more than six of the nine STR loci that were tested. When the DNA mixture ratio was 1:28, only one locus that could be completely resolved (Table 4.8). The partially resolved CPI values for each of the eight tested mixture ratios ranged from approximately 1 in 1,850,000 (1:1 mixture) to 1 in 4,100,000 (1:6 mixture) and were consistently more discriminating than the 1 in 28,200 unresolved CPI value (Table 4.8).

The D18S51 locus was the easiest to separate. It was completely resolved in six of the eight different mixture ratios; in all instances at least one of its four alleles was consistently out of peak height balance with those of the secondary contributor. Only one of the nine tested loci, D13S317, was never resolved in any of the eight mixture ratios. This is due to that fact that the mixture at that locus contains only two alleles, which always contains multiple legitimate hypotheses as discussed above (Table 4.4). Only one of the eight tested mixture ratios, 1:2, resulted in at least a partial resolution for all nine of the tested loci. Three of the eight observed peak

Mix Ratio	Solved	Number of loci where Narrowed	Not narrowed or solved	Incorrect	RMP 1	RMP 2	Partially Resolved CPI
5:1	3	3	3	0	1 / 6,110,000	1 / 12,800	1 / 3,390,000
3:1	4	4	1	0	1 / 10,800,000	1 / 70,900	1 / 12,200,000
1:1	2	5	2	0	1 / 1,930	1 / 270	1 / 1,850,000
1:2	5	4	0	0	1 / 28,900,000	1 / 1,920,000	1 / 44,400,000
1:3	6	2	1	0	1 / 1,200,000,000	1 / 166,000,000	1 / 145,000,000
1:6	3	3	1	1	1 / 18,000	1 / 14,700	1 / 4,100,000
1:10	3	3	2	1	1 / 4,400	1 / 16,700	1 / 28,000,000
1:28	1	5	0	3	1 / 20	1 / 20	1 / 51,900,000

Table 4.8: **Mixture resolution outcomes. This approach to mixture resolution was applied to the DNA profile information for the eight mixture ratios presented in 4.7. "Only one hypothesis valid" reports the number of loci for which only one of the alternative hypothesis combination of genotypes satisfies the imposed peak height imbalance and additivity constraints. "Hypotheses constrained" reports the number of partially resolved loci - loci at which at least one hypothesis has been eliminated from consideration but more than one remain. "All hypotheses valid" reports the number of loci at which none of the hypotheses can be eliminated. RMP 1 and RMP 2 represent the random match probabilities for the solved loci for contributor 1 and 2, respectively. The CPI-with-constrained-hypotheses is derived from the remaining alternative hypotheses of genotype combinations for all loci, including those that have been fully resolved. "Incorrect loci" were caused by a minor contributor's allele being removed by the ABI stutter filter, a stutter peak exceeding the ABI stutter threshold, or a minor contributor present in stutter position causing the peak to be imbalanced with its paired allele. All "incorrect loci" become "not narrowed or solved" when the additional considerations for stutter peaks are taken into account.**

height imbalances that exceeded the 70% threshold in this data set were observed in the 5:1 mixture. All eight instances of unexpectedly large peak height imbalance contributed significantly to the method's inability to eliminate alternative hypotheses of genotype combinations from consideration.

There are six instances in which the resolution method did not include the correct contributor pair in its list of hypotheses for a given locus. The resolution method failed to include the correct hypothesis for the D3S1358 locus in the 1:6 and 1:10 mixture ratios due to the 17 allele from the minor contributor being removed by the ABI stutter filter (Table 4.7). The 1:6 mixture ratio also failed to resolve the D21S11 locus due to the presence of stutter exceeding the stutter threshold, which made a three allele locus appear to contain four alleles (Table 4.7). The mixture ratio of approximately 1:28 failed to include the correct hypothesis in three loci (D3S1358, vWA, and D21S11). The 17 allele in DS1358 and the 30 allele in D21S11 were removed by the ABI stutter filters. In the vWA locus, the minor 17 allele fell below the 50 RFU threshold, but the paired 18 allele was raised by the stutter contribution from the major contributor's 19 allele.

In the instances where a contributor fell below the RFU threshold and was not affected by stutter filters or contributions, the resolution method was able to identify the presence of allelic dropout. In the 1:6 and 1:10 mixture ratios, one locus exhibited allelic dropout and was identified as a possible hypothesis. The mixture ratio of approximately 1:28 exhibited three additional loci with allelic dropout that were identified. The minor contributor was completely absent from the reported results of the D18S51 locus, yet that hypothesis was included.

4.5 Discussion

Resolution of individual STR DNA profiles from samples containing mixtures of material from multiple contributors can be challenging. It has been our experience that when forensic testing laboratories are confronted with mixed evidence samples that they routinely report combined probability of inclusion statistics and/or rely upon their experience, training and expertise to visually separate alleles into a major (and sometimes also a minor) contributor. As long as a single genotype can be so resolved, the lab reports a random match probability. This effectively treats the mixture as two 'single-stains'. The mixture resolution method described here uses an objective approach to resolve a locus under an explicitly hypothesized number of contributors. The net effect is to generate a list of 'restrained' genotypes for the contributors. We advocate the use of a CPI statistic that is generated from only the restrained genotypes. In our opinion, this approach more accurately reflects the discriminating power of the DNA profiling technique when the evidence sample is mixed. It differs significantly from other approaches in that it utilizes the quantitative aspects of the data and a series of mathematical calculations to provide an objective list of possible genotypes after eliminating those which fail to satisfy objective criteria.

The methods proposed by Perlin et al. and Wang et al. determine a likely mixture separation based on the minimization of an error metric (Perlin and Szabady, 2001; Wang et al., 2001). The profile combination with the least amount of measured error may or may not be the correct contributor profile set and there is little one can do to ascertain the correctness of the choice absent additional information. The method

proposed by Clayton et al. is similar to that which is proposed here: all possible contributor profiles are enumerated and evaluated (Clayton et al., 1998). However, Clayton's method relies heavily on mixture ratio assumptions and fails to consider the presence of allelic dropout.

This approach to mixture resolution is intended to form the basis of a mathematical approach for objectively interpret mixed forensic DNA samples. Herein we explicitly hypothesize that there are two (and only two) individuals contributing alleles to each tested locus though a similar framework could in principle be extended to evaluate mixtures presumed to have more than two contributors. The approach rests upon the fact that in a heterozygote the two alleles should give rise to two signal strengths that are within a fixed parameter (c) of each other and that peak heights are generally additive (at least within a specified range of signal strength, such as between 150 and 4,000 RFUs). The formulas presented in the Materials and Methods section of this manuscript incorporate the threshold for peak height balance as a variable (c) that can be easily substituted for whatever value a testing laboratory's validation studies have determined to be appropriate (such as the commonly used 70% threshold that was utilized here for illustration purposes).

The ability for mixtures to be resolved diminishes when the ratio is either very similar (close to 1:1) or far apart (greater than 1:5 in the results presented here). One reason for the latter is the preferential amplification of the major contributor. Another reason is that for low DNA concentrations, a minor contributor can be masked by or mistaken for stutter. For example, the 17 allele at D3S1358 in the 1:10 ratio is in stutter position of the major 18 allele and falls below the ABI stutter threshold

(Table 4.7). It may be necessary to consider certain mixtures with the stutter filters turned off in order to identify all possible mixture contributors. Contributor peaks in stutter position may also be observed to be higher than if they were observed in non-stutter positions. Care should be taken when examining such contributor profiles as the contribution from stutter may put the minor contributor peaks out of balance. For example, the 13 allele at D3S1358 in the 1:6 ratio falls below the 50 RFU threshold, yet its paired 17 allele is raised to 115 RFUs due to the stutter product from the major contributor's 18 allele (Table 4.7). All of the "incorrect loci" became "not narrowed or solved" when the additional considerations for stutter peaks were taken into account (either labeling filtered stutter peaks or subtracting the expected amount of stutter from labeled stutter peaks). As with all sample evaluation, care should be taken to determine the appropriate peak detection threshold for a specific analysis run (Gilder et al., 2007b).

The mixture resolution method outlined here is implemented in a freely-downloadable software package called GenoStat™ (Gilder et al., 2007a). Once a user has entered the RFU values for a given DNA sample, GenoStat™ performs the mixture resolution method as well as calculates the random match probability (RMP) for each contributor's fully resolved loci, the unconstrained combined probability of inclusion (CPI), and the combined probability of inclusion (CPI) for constrained loci using only the contributor hypotheses that the resolution method has deemed plausible.

Chapter 5

Computational analyses of simulated DNA databases

5.1 Preface

The work described here comes from the journal article submissions: J. Gilder, T. Doom, M. Raymer, D. Krane. "Assessing the implications of the presence of related individuals in DNA databases." Journal of Forensic Sciences, to be submitted and J. Gilder, T. Doom, M. Raymer, D. Krane. "A practical approach for conducting familial searches of DNA databases." Journal of Forensic Sciences, to be submitted.

5.2 Introduction

If a criminal investigation results in a DNA profile with no suspect, law enforcement may search that DNA profile against a database of DNA profiles from previous offend-

ers. A DNA match acquired in such a manner is called a "cold hit." A DNA database search is simply a string comparison operation. Each state has its own DNA database and there is a centralized national database as well (Federal Bureau of Investigation, 2007). There is currently debate as to what the evidential weight of a cold hit DNA match should be (National Research Council, 1992; National Research Council, 1996; DNA Advisory Board, 2000b; Balding and Donnelly, 1996). All currently employed methods fail to consider the possibility of related individuals in a database. Related individuals naturally have more alleles in common as opposed to unrelated individuals and are therefore more likely to result in a coincidental match, especially when partial profiles are being considered.

In some instances, a DNA database search will result in a close, but ultimately non-matching DNA profile. In those cases, law enforcement may decide to investigate close relatives of the best matching DNA profile in their database. Few methods are available for determining when it is highly likely that a relative will match the evidence in this situation. Law enforcement currently relies on the number of alleles or loci that match the evidence. A likelihood ratio approach has been developed (Paoletti et al., 2006), but it has not been extensively validated or compared to existing methods of familial searches.

The CODIS and state DNA databases are currently not available to the public. Therefore database research must be performed on simulated databases. The FBI has made their databases used to generate allele frequencies publically available (Budowle and Moretti, 1999). Databases can be randomly generated to contain any number of individuals, yet still retain the allele frequencies found in the general population.

Virtual families can be generated by creating random individuals and creating offspring through standard Mendelian inheritance (Paoletti et al., 2005). The resulting databases can be examined to determine the effect of introducing related individuals. Collections of related individuals can be studied to determine how well existing familial search methods work and how they can be improved.

5.3 Assessing the implications of the presence of related individuals in DNA data banks

5.3.1 Introduction

One of the principal advantages of using STR genotypes for the purposes of human identification is their amenability to archiving in searchable databases such as the Combined DNA Index System (CODIS) (Federal Bureau of Investigation, 2007). More than two million complete, 13 locus STR-DNA profiles of convicted offenders have already been entered into the CODIS database in the United States and similar databases are maintained by European countries and Australia. These databases were created as investigative tools for law enforcement agencies tasked with identifying suspects in cases where a perpetrator has left biological material at the scene of a crime but few or no additional leads are available. Perfect matches between evidentiary material and an individual in such a database are known as "cold hits" and the appropriate way to describe the significance of such a DNA profile match in such cases has been a topic of considerable debate (National Research Council,

1992; National Research Council, 1996; DNA Advisory Board, 2000a; Balding and Donnelly, 1996).

It would not be surprising to learn that a given DNA database contained related individuals. However, the effect of related individuals in a DNA database has not been extensively examined. In this study, hypothetical DNA databases of randomly-generated individuals were created to observe the levels of allele and locus sharing amongst unrelated individuals. Virtual families were then created to generate siblings, half-siblings, cousins, and parent-child pairs to observe what effect their introduction to the database would have on the overall levels of genotype sharing. The introduction of siblings increased the observed levels of allele and locus sharing, while the other degrees of related individuals displayed similar trends to those observed in unrelated individuals. Current methods for evaluating a DNA match found in a database rely on forms of the random match probability, which assumes that the population in question is unrelated. It may be necessary to apply a correction when a DNA database exhibits indications of containing related individuals.

5.3.2 Materials and methods

The databases in this study were generated from the 196 published FBI Caucasian 13 locus genotypes (Budowle and Moretti, 1999). The dataset was used to generate the allele frequencies for the Caucasian population and was analyzed for Hardy-Weinberg equilibrium (Budowle et al., 1999).

All initial databases were created to be unambiguously unrelated by randomly

selecting alleles from the initial FBI dataset. Allele frequencies in this randomized dataset are the same as in the original dataset but individuals are unequivocally unrelated by descent (alleles are not the same because they have been faithfully passed from a common ancestor). Instead, any allele sharing can arise only through identity by state (alleles are the same because there is a finite number of different alleles that can be detected). Each locus was considered independently during the production of randomized genotypes.

All individuals in the original dataset are assumed to have two and only two alleles per locus (rare conditions resulting in unusual allele counts such as null alleles, triploidy or chimerism are beyond the scope of this study). Similarly, all simulated genotypes are considered to be free of any typing errors that might further complicate the interpretation.

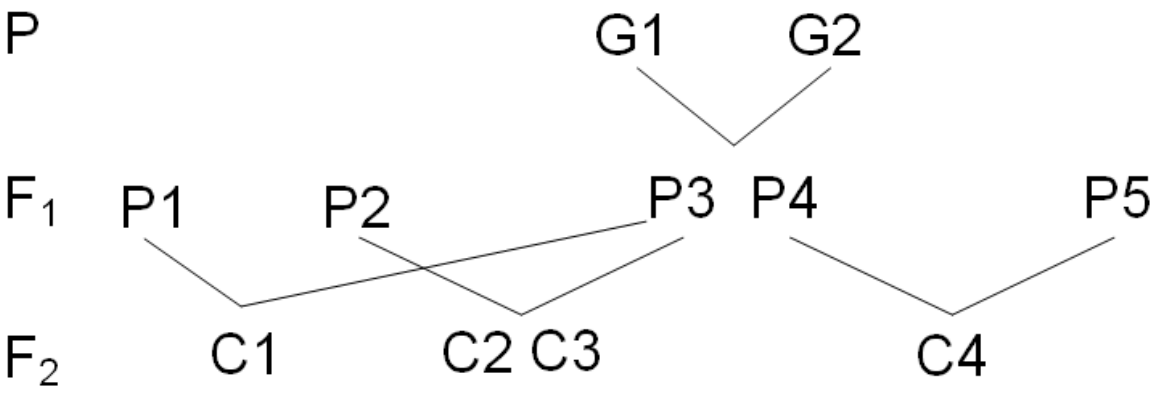

Figure 5.1: **The "virtual family" consisting of two grandparents, five parents, and four children. Individuals G1, G2, P1, P2, and P5 are randomly-generated (unrelated). Individuals C2 and C3 are selected as siblings, C1 and C2 are half-siblings, C3 and C4 are cousins, and G1 and P3 form the parent-child pair.**

Related individuals were created by generating "virtual families" (Paoletti et al., 2005). Each family consists of two grandparents, five parents, and four children (Figure 5.1). Each child was generated by randomly selecting one allele from each parent at each locus. With three generations in a virtual family, it was possible to create siblings, half-siblings, cousins, and parent-child pairs.

Allele/Locus sharing

For each database, every individual was compared with every other individual to determine the number of alleles and loci in common between all individuals. Homozygotes were deemed to share two alleles with other homozygotes of the same genotype. The number of pairwise comparisons can be calculated by $\frac{n(n-1)}{2}$, where n is the number of individuals in the database.

Databases of unrelated individuals

Twenty-one randomly-generated databases of individuals were generated containing between 1000 and 100000 unrelated individuals in increments of 5000 individuals (starting at the increment with 5000 individuals). Each database was generated independently.

Databases of constant size with sibling pairs

Twelve randomly-generated databases of 10000 unrelated individuals were created. Each database had a portion of its individuals randomly removed and replaced with sibling pairs. The number of sibling pairs varied from zero (completely unrelated) to

5000 (all sibling pairs) in 500 sibling pair increments. In addition, a single database was generated containing 100 sibling pairs. Sibling pairs were generated independently, so each sibling has only one other sibling in the database.

Databases of different sibling ratios

Thirty-three randomly-generated databases were created containing between 1000 and 100000 unrelated individuals. There are 25 databases between 1000 and 25000 individuals (in 1000 individual increments) and eight databases between 30000 and 100,000 individuals (in 10000 individual increments). For each database of unrelated individuals, two additional databases were created replacing 1% and 10% of randomly-chosen individuals with sibling pairs. Siblings were generated independently and each sibling has only one other sibling in the database.

Databases with different degrees of related individuals

Five of the previously-generated databases were chosen containing 5000, 10000, 15000, 20000, and 25000 unrelated individuals. For each database of unrelated individuals, four additional databases were created replacing 10% of the randomly-chosen individuals with pairs of siblings, half-siblings, cousins, and parent-child pairs. Four additional replicates of each database were generated. The average degree of allele and locus sharing was calculated across the five databases for each population size

In addition, 53 randomly-generated databases of 10000 unrelated individuals were generated. Each database had a portion of its individuals randomly removed and replaced with pairs of siblings, parents and children, uncles and nephews, and cousins.

The number of related pairs varied from zero (completely unrelated) to 5000 (all related pairs) in 500 related pair increments. In addition, databases were generated containing 10, 50, and 100 related pairs. Related pairs were generated independently, so each relative has only one other relative in the database. The original database of unrelated individuals was also examined for profile similarity. Profile similarity thresholds were set at nine or more loci and twenty-one or more alleles.

5.3.3 Results

Databases of unrelated individuals

The databases of unrelated individuals exhibit higher degrees of allele and locus sharing as the size of the database increases (Figures 5.2 and 5.3). Matches at nine or more loci are first observed in the database of 10000 individuals (four pairs of individuals). With a population of 50000 individuals, there are 60 pairs of individuals matching at nine or more loci. With the final population of 100000 individuals, there are 249 pairs of individuals matching at nine or more loci.

Databases of constant size with sibling pairs

The degree of allele and locus sharing increases with the number of sibling pairs (Figures 5.4 and 5.5). The databases of unrelated individuals and ten sibling pairs exhibit only four pairs of individuals matching at nine or more loci. With 500 sibling pairs, there are 17 pairs of individuals matching at nine or more loci. With half the database composed of siblings (2500 pairs), there are 64 pairs of individuals matching

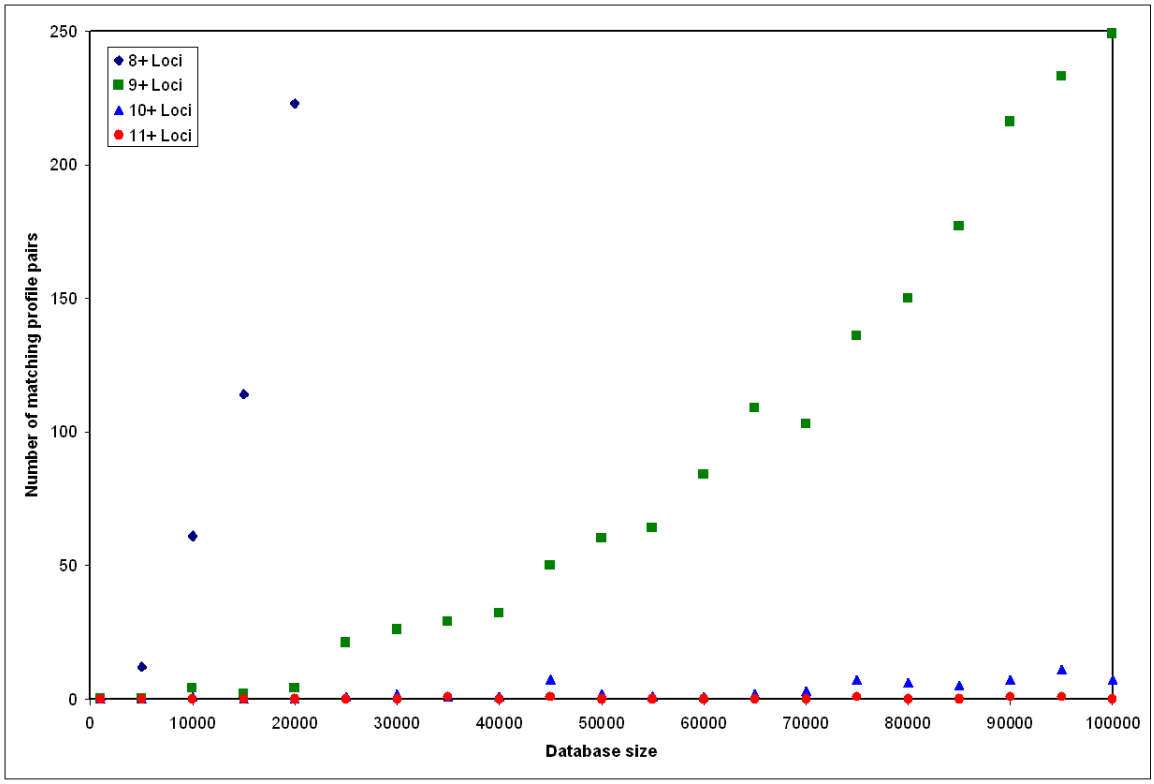

Figure 5.2: **The number of pairs of profiles that are consistent across 8, 9, 10, and 11 or more loci from a set of hypothetical databases containing between 1000 and 100,000 unrelated individuals.**

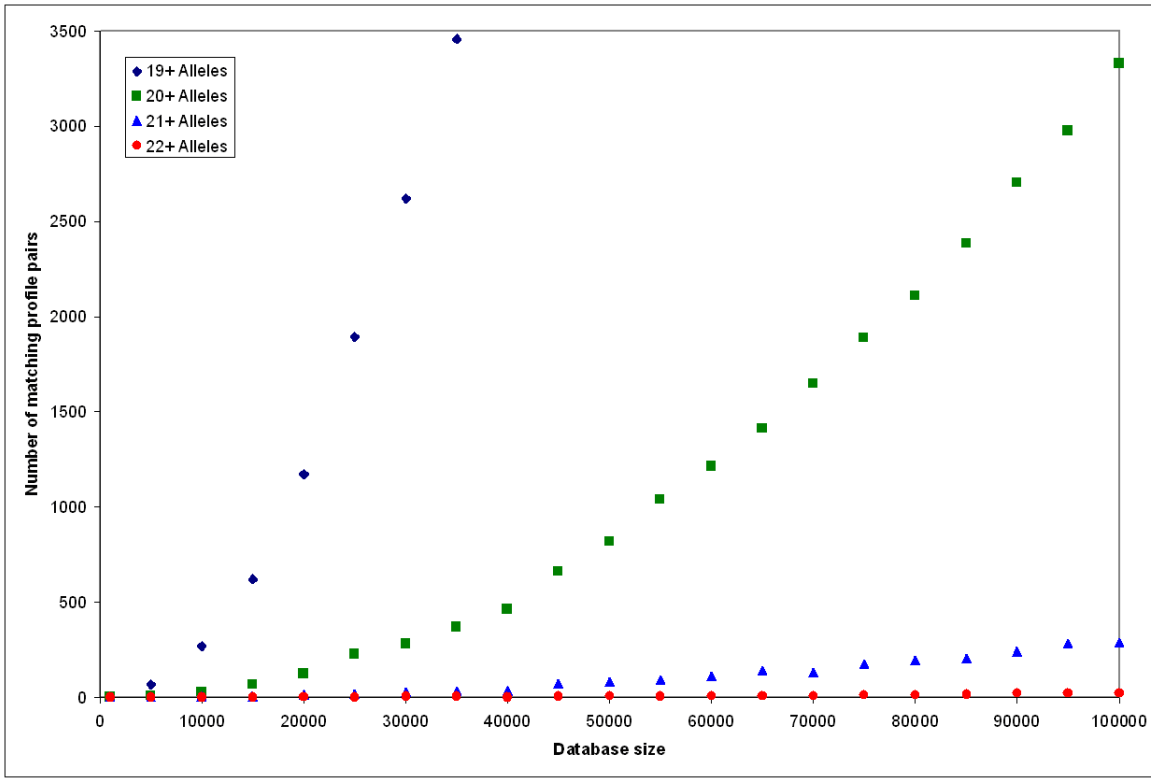

Figure 5.3: **The number of pairs of profiles that are consistent across 19, 20, 21, and 22 or more alleles from a set of hypothetical databases containing between 1000 and 100,000 unrelated individuals.**

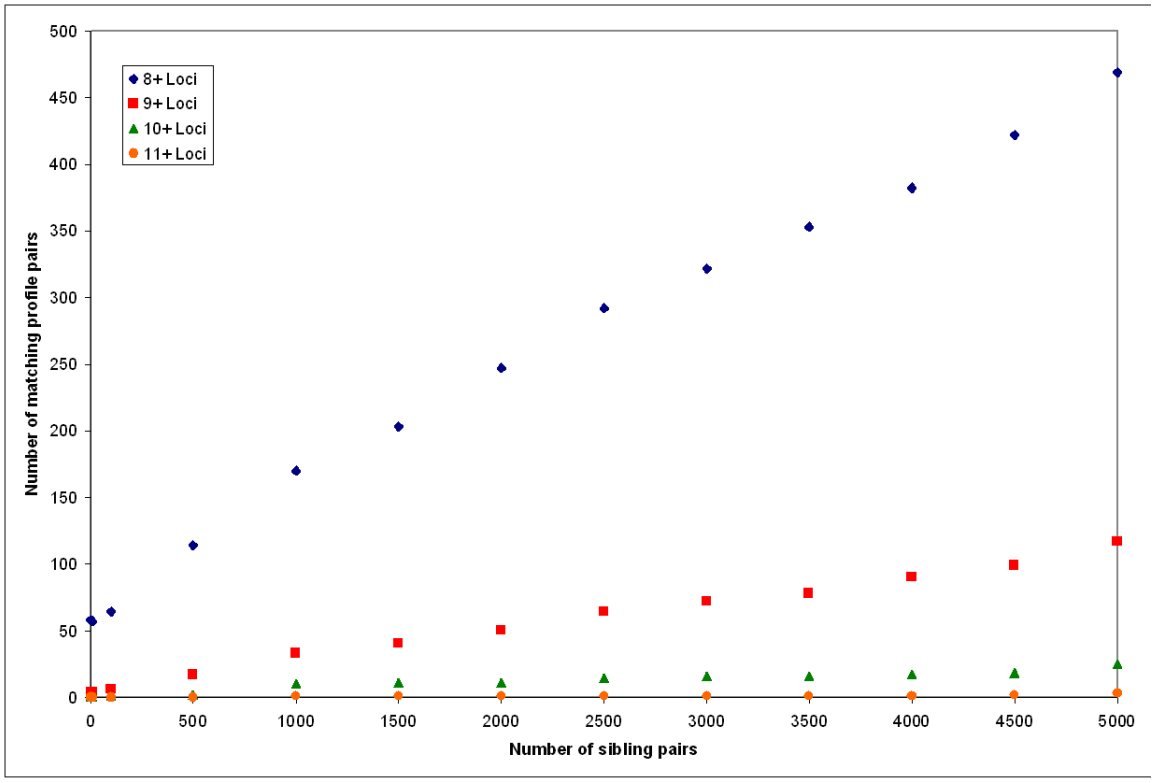

Figure 5.4: **The number of pairs of profiles that are consistent across 8, 9, 10, and 11 or more loci from a set of hypothetical databases containing 10000 individuals with between 0 and 5000 pairs of siblings.**

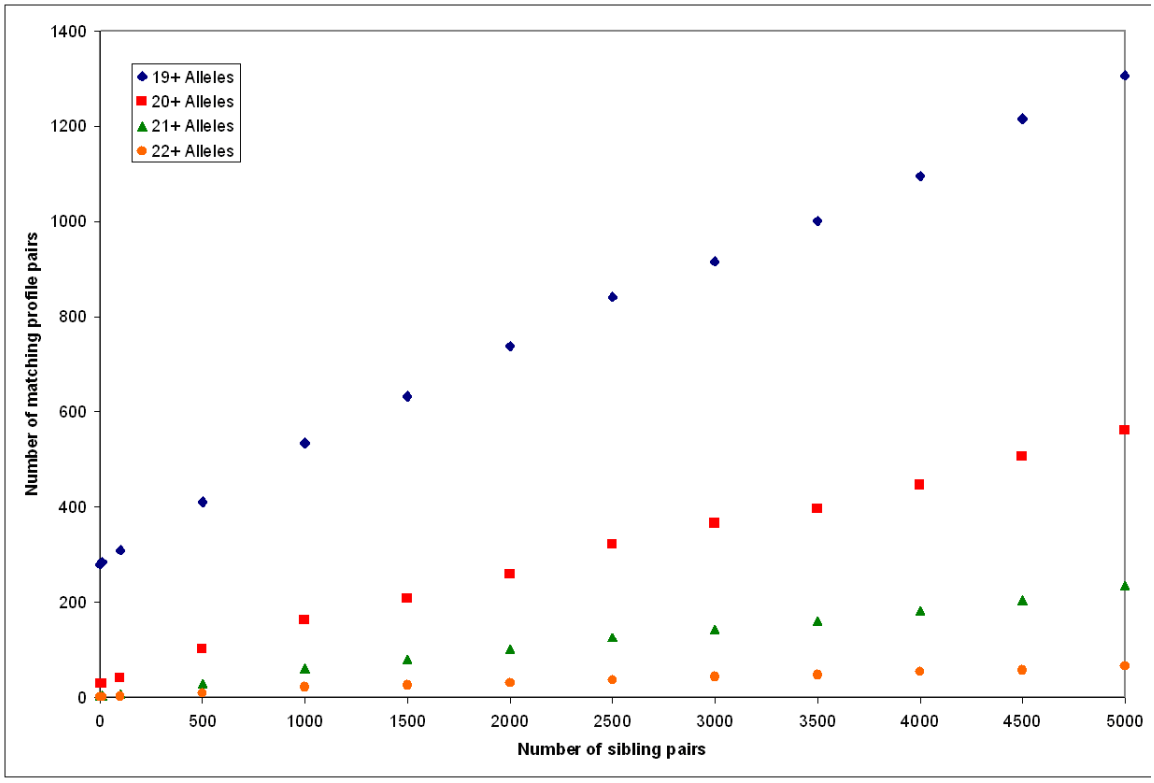

Figure 5.5: **The number of pairs of profiles that are consistent across 19, 20, 21, and 22 or more alleles from a set of hypothetical databases containing 10000 individuals with between 0 and 5000 pairs of siblings.**

at nine or more loci. With the entire database composed of sibling pairs (5000 pairs), there are 117 pairs of individuals matching at nine or more loci.

Number of matching loci	Slope (matching profile pairs / number of sibling pairs)	y intercept (matching profile pairs with no siblings)	r^2	Number of matching alleles	Slope (matching profile pairs / number of sibling pairs)	y intercept (matching profile pairs with no siblings)	r^2
8	0.08	71	0.991	**19**	0.2	304	0.997
9	0.02	6	0.993	**20**	0.1	42	0.996
10	0.004	2	0.926	**21**	0.04	7	0.996
11	0.0004	0.02	0.764	**22**	0.01	4	0.986

Table 5.1: **The linear regression models for the data presented in figures 5.2 and 5.2.**

The relationship between the number of sibling pairs in a database and the number of pairs of individuals matching at a given number of alleles or loci is illustrated by a strong correlation (Table 5.1). The linear regressions generate r-squared values above 0.9 for all locus and allele sharing levels except for the number of pairs of individuals matching at eleven or more loci ($r^2 = 0.76$). The slope shows the rate of growth based on the number of siblings in the database. The y-intercept shows the approximate number of pairs of matching individuals for a given locus or allele sharing level when the database is completely unrelated.

Databases of different sibling ratios

The databases of unrelated individuals and those containing one percent sibling pairs display similar levels of allele and locus sharing across all databases (Figures 5.6 and 5.7). A significantly higher degree of allele and locus sharing is observed when ten percent of the database is composed of sibling pairs. Profile matches at nine or more loci are first observed with the database of 3000 individuals containing ten percent

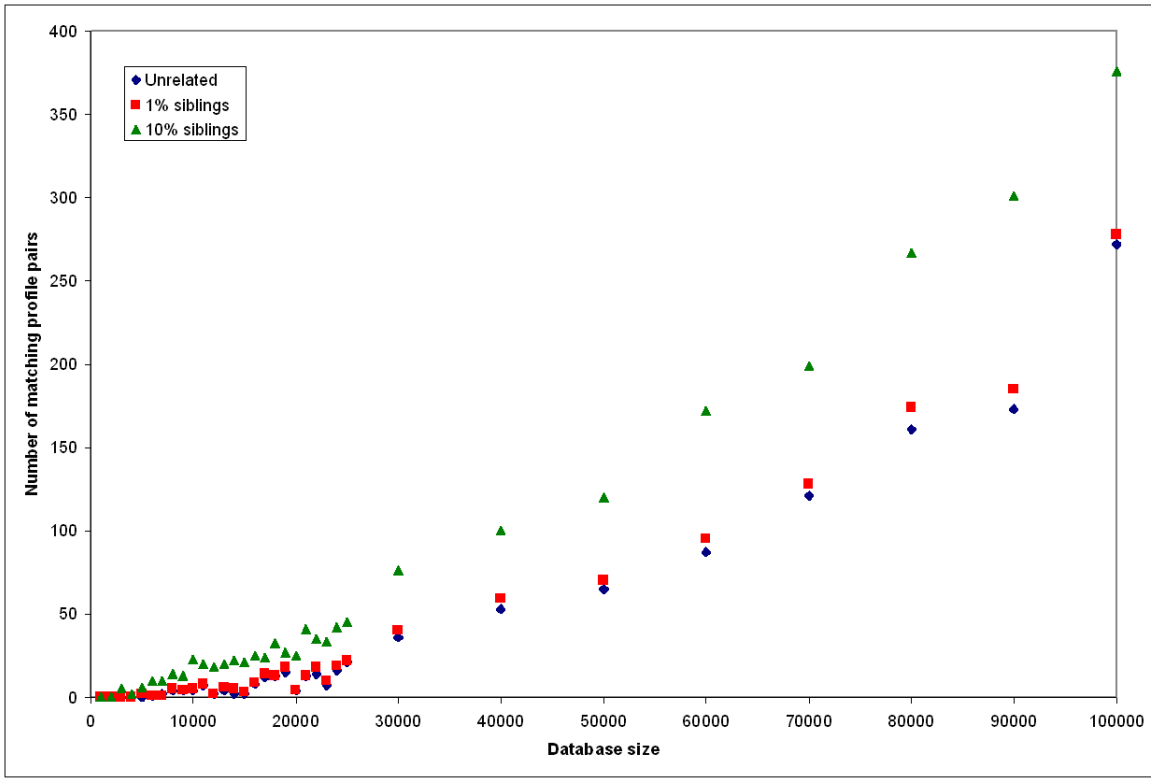

Figure 5.6: **The number of pairs of profiles that are consistent across nine or more loci in databases containing between 1000 and 100,000 individuals and consisting of no, 1%, and 10% siblings.**

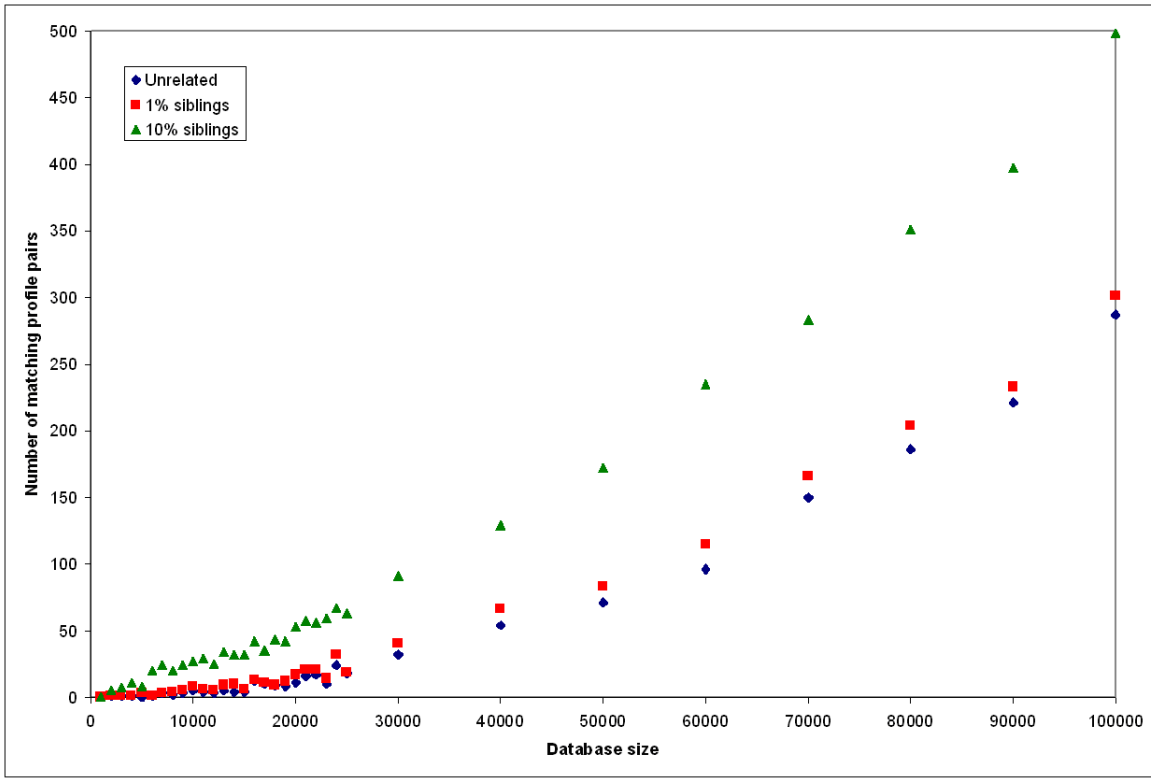

Figure 5.7: **The number of pairs of profiles that are consistent across 21 or more alleles in databases containing between 1000 and 100,000 individuals and consisting of no, 1%, and 10% siblings.**

sibling pairs (five pairs of individuals). With a population of 50000 individuals, there are 65 pairs of individuals in the unrelated database, 70 pairs of individuals in the database with one percent sibling pairs, and 120 pairs of individuals in the database containing ten percent sibling pairs that are consistent across nine or more loci. With the final database of 100000 individuals, there are 272 pairs of individuals in the unrelated database, 278 pairs of individuals in the database with one percent sibling pairs, and 376 pairs of individuals in the database containing ten percent sibling pairs that are consistent across nine or more loci.

Databases with different degrees of related individuals

	Database size				
	5000	**10000**	**15000**	**20000**	**25000**
Unrelated	0, 0	2.8, 1.2	4.8, 1.9	7.4, 3.1	16.9, 2.2
Siblings	5.6, 1.0	15.0, 5.5	22.0, 4.4	28.6, 3.7	41.8, 2.3
Parent-child	0, 0	3.0, 1.9	4.8, 2.2	8.0, 3.0	16.0, 2.4
Half-siblings	0, 0	3.8, 2.0	5.0, 1.4	9.8, 2.8	15.2, 1.7
Cousins	0.2, 0.4	2.8, 1.2	4.0, 1.3	7.8, 2.9	17.6, 2.5

Table 5.2: **The numbers of pairs of profiles that are consistent across nine or more loci within databases containing between 5000 and 25000 individuals with 10% of the database consisting of siblings, half-siblings, cousins, or parent-child pairs. The results are presented in the form of (average, standard deviation), taken from five replicates of the databases.**

	Database size				
	5000	**10000**	**15000**	**20000**	**25000**
Unrelated	0.4, 0.5	3.0, 1.3	6.6, 2.2	10.2, 2.6	18.2, 2.9
Siblings	10.4, 2.6	25.0, 2.8	36.2, 6.4	54.0, 7.4	64.6, 4.1
Parent-child	1.4, 1.0	4.6, 1.6	6.0, 1.7	14.2, 3.1	17.4, 1.2
Half-siblings	0.8, 0.7	3.6, 1.6	6.2, 1.8	13.4, 3.1	15.4, 2.4
Cousins	1.0, 1.5	3.8, 1.6	5.6, 1.4	10.6, 3.0	17.8, 0.7

Table 5.3: **The numbers of pairs of profiles that are consistent across 21 or more alleles within databases containing between 5000 and 25000 individuals with 10% of the database consisting of siblings, half-siblings, cousins, or parent-child pairs. The results are presented in the form of (average, standard deviation), taken from five replicates of the databases.**

The databases containing pairs of half-siblings, cousins, and parent-child pairs exhibit allele and locus sharing at approximately the same rate as populations of unrelated individuals (Tables 5.2 and 5.3). With 5000 individuals, all but the sibling databases had approximately no individuals matching at nine or more loci (the sibling databases had an average of 5.6 individuals ($\sigma = 1.0$)). With a 15000 individuals, there are an average of 4.8, 4.8, 5.0, and 4.0 pairs of individuals matching at nine or more loci for the unrelated, parent-child, half-sibling, and cousin databases, respectively. The sibling database of 15000 individuals has an average of 22.0 pairs of individuals matching at nine or more loci. With the final population size of 25000 individuals, there are an average of 16.9, 16.0, 15.2, and 17.6 pairs of individuals matching at nine or more loci for the unrelated, parent-child, half-sibling, and cousin databases, respectively. The sibling database of 25000 individuals has an average of 41.8 pairs of individuals matching at nine or more loci.

The number of profile matches at nine or more loci or 21 or more alleles increases greatly with the presence of sibling pairs (Figures 5.8 and 5.9. With a locus threshold of nine or more loci, all levels of relatedness (aside from siblings) display similar levels of profile similarity (including unrelated individuals). The observed trends are similar with an allele threshold of 21 or more alleles. However, an allele similarity allows for a finer distinction between profiles, so relatively small increases in profile similarity are observed with the related populations (aside from siblings) compared to the unrelated population.

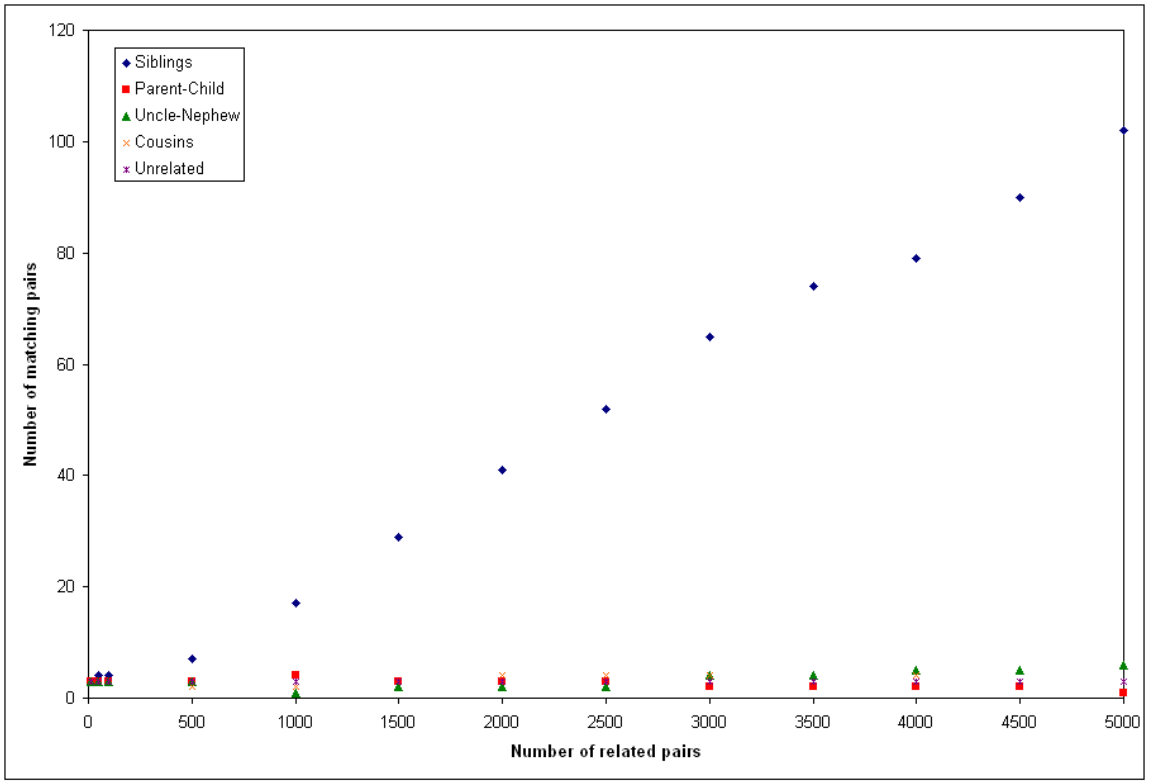

Figure 5.8: **The number of pairs of profiles matching at 9+ loci from a set of hypothetical databases containing 10000 individuals with between 0 and 5000 pairs of related individuals.**

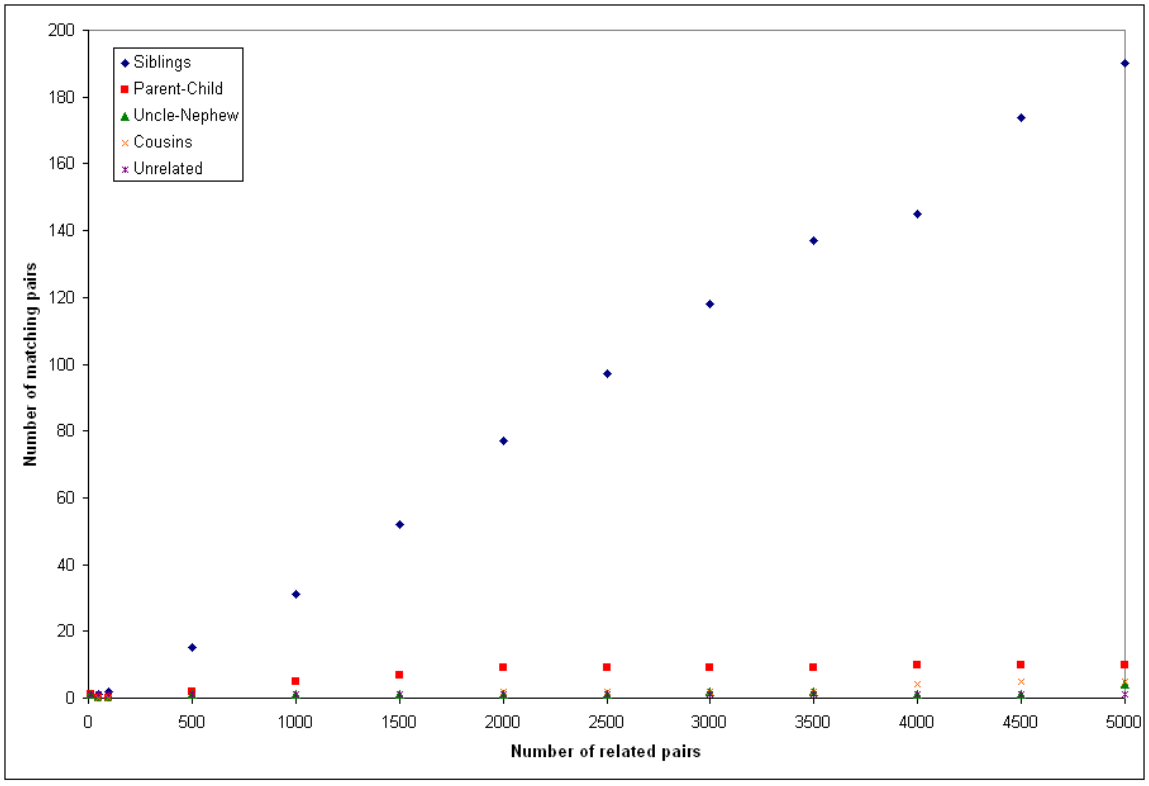

Figure 5.9: **The number of pairs of profiles matching at 21+ alleles from a set of hypothetical databases containing 10000 individuals with between 0 and 5000 pairs of related individuals.**

5.3.4 Discussion

Presence of related individuals in a database

The analysis of allele and locus sharing among profiles in a database can provide insights into the presence of related individuals. However, observing increased allele and locus sharing at higher locus or allele thresholds is almost solely attributed to the presence of siblings. Other degrees of related individuals exhibit similar allele and locus sharing patterns to that of unrelated individuals. Therefore, the use of genotype information alone is likely to only allow for identifying the possible presence of siblings in a database.

Familial searches

Law enforcement agencies sometimes perform database searches of "low" or "moderate stringency" to identify close, but not perfect DNA matches to an article of evidence (Section 5.4). A family member of a close database match may be investigated to see if he or she provides a perfect match to the evidence (Paoletti et al., 2006; Bieber et al., 2006). This process is known as a familial search. Different jurisdictions have their own thresholds to determine when a search of a family member is warranted. In Florida, an individual matching at least 21 out of 26 alleles is close enough to investigate a relative (Paoletti et al., 2006). It is possible that many individuals could match 21 or more alleles simply by chance. There are 80 pairs of individuals that match at 21 or more alleles in the database of 50000 unrelated individuals. There are 288 pairs of individuals who match at 21 or more alleles in the

database of 100000 unrelated individuals. Clearly, there is a potential for unrelated individuals to match at a large number of alleles completely by chance.

Cold hit statistics

There is significant controversy regarding the appropriate means of describing the significance of a match between the DNA profile observed in an evidence sample and an individual whose DNA profile is maintained in the databank (see *US v Jenkins*). Most of the statistical approaches that have been suggested are rooted in formulae that generate a random match probability (RMP) (National Research Council, 1992; National Research Council, 1996; DNA Advisory Board, 2000a; Balding and Donnelly, 1996). The random match probability explicitly describes the chance of picking a random *unrelated* individual from the population with a given DNA profile. The logical foundation for using RMP-based statistics is undermined to the extent to which related individuals are found to exist in a databank.

Performing pairwise database searches of real-world offender databases can provide insight into the presence of related individuals. For example, as part of the laboratory's quality assurance policy, the Arizona Department of Public Safety routinely performs a search of their DNA database against itself in a similar manner to that employed by this study. In 2001, the Arizona DPS discovered a nine locus match between two unrelated individuals in their database of approximately 20000 individuals (Troyer et al., 2001; Johnson, 2005). In 2005, a pairwise search of the their database of 65393 individuals yielded 144 pairs of individuals matching at nine or more loci (Johnson, 2005). The question then becomes: is the Arizona DNA

database representative of an unrelated population? In the study presented here, the simulated database of 65000 unrelated individuals has 109 pairs of individuals matching at nine or more loci. Based on the results shown in Figures 5.2 and 5.4, it is likely that there are approximately 1000 pairs of siblings present in the Arizona database.

It may be necessary to implement a correction to the random match probability formula when a database contains related individuals. One conservative method would be to implement a relatively large value of theta for the random match probability to account for increased population substructure (Section NRC2theta). It is currently unknown what value of theta would be appropriate to use in a database containing a relatively large percentage of related individuals. A much more conservative approach would be to replace the random match probability formula with the sibling match probability formula (National Research Council, 1996). The sibling match probability determines the likelihood of choosing a random sibling that has a given DNA profile. The sibling match probability still typically generates impressive results and guarantees that if siblings are present in the database, they are being evaluated appropriately.

5.4 A practical approach for conducting familial searches of DNA databases

A perfect match between the STR DNA profile of an evidence sample and an individual whose genotype has been entered into a database has clear utility as an investigative tool. Lack of concordance between the alleles of an evidence sample and an individual's DNA profile is also commonly used as an investigative tool in that the individual can be excluded as a source of the biological material in the evidence sample. Very similar but nonetheless non-matching DNA profiles between an evidence sample and an individual in a DNA database also have the potential to provide useful information by suggesting that a close relative of the individual may be the actual source of the evidence sample (Paoletti et al., 2006; Bieber et al., 2006). However, limitations of the software currently available for searching the CODIS database make it much more likely to miss the DNA profile of a close relative in a database than to generate a useful investigative lead. Similarly, an approach that relies upon a minimum number of matching alleles across all loci only identified a small fraction of true familial hits unless the threshold was set so low that false positives became problematic. A likelihood ratio approach (Paoletti et al., 2006) is much better suited for familial searches, particularly for parent/child and sibling relationships but also for the other degrees of relatedness that were considered, both in terms of its true and false positive rates.

5.4.1 Introduction

The National Research Council (NRC) was among the first to formally consider the possibility of using DNA databases for the purpose of familial searches. It's summary report (National Research Council, 1992) clearly articulated concerns about a suspect being investigated not because of their own actions but rather because of the actions of a relative that led to that relative being included in a DNA database. These privacy and fairness concerns caused the NRC to go so far as to suggest that the software used for database searches be specifically designed to not allow familial searches.

> To put it succinctly, DNA databanks have the ability to point not just to individuals but to entire families including relatives who have committed no crime. Clearly, this poses serious issues of privacy and fairness. . . . [I]t is inappropriate, for reasons of privacy, to search databanks of DNA from convicted criminals in such a fashion. Such uses should be prevented both by limitations on the software for search and by statutory guarantees of privacy (National Research Council, 1992).

The second National Research Council report was published in 1996 and effectively reversed the position of the first report.

> If the possible contributors of the evidence sample include relatives of the suspect, DNA profiles of those relatives should be obtained. If these profiles cannot be obtained, the probability of finding the evidence profile in those relatives should be calculated (National Research Council, 1996).

Currently available CODIS database software allows for three different database search stringencies: high, moderate, and low (Rudin and Inman, 2002). These stringencies are rooted in a desire to identify suspects whose profiles are in DNA databases even when the information from an evidence sample is compromised either by degradation/inhibition and/or by arising from small amounts of starting material High stringency searches require a perfect match, so both the database and search profiles must contain exactly the same alleles

High stringency match: locus profile **A,B** matches **A,B**

A moderate stringency search also requires that the database profile contains all of the alleles found in the search profile (and vice versa), but homozygous loci can be considered a match to a heterozygote if one allele is in common.

Moderate stringency match: locus profile **A,A** matches **A,B**; **A,B** matches **B,B**

Finally, a low stringency search only requires that one allele is found to be in common across all loci.

Low stringency match: locus profile **A,B** matches **B,C**

The United States have steadily shifted toward inclusion of all felons in government controlled DNA databases and federal and six U.S. state laws now include some provision for the inclusion of those who are just arrested or indicted. Statutes governing the use of DNA databases vary significantly from state to state (www.aslme.org) with some specifically disallowing familial searches, while others specifically encourage

them. It has been reported (Willing, 2005) that the threshold of similarity that must be exceeded to warrant an actual investigation of relatives is also variable. Those thresholds however, are generally rooted in a certain minimum number of alleles being in common between an evidence sample and a similar but non-matching DNA profile in a database (Willing, 2005).

The DNA of biological relatives is commonly used to perform indirect genetic kinship analyses to assist with missing person identifications and for identifications in mass disasters (Brenner and Weir, 2006; Brenner, 2006; Budowle et al., 2005). Monte Carol analyses have suggested that such methods could be used to detect suspects who are the parents, children, or siblings of those whose profiles are in forensic databases (Bieber et al., 2006). For instance, a child of an individual in a database containing 50,000 unrelated individuals was found to have the greatest likelihood of a parent-child relationship of all individuals in a database about half the time and has an 80% chance of being in the top 10 leads (Bieber et al., 2006). However, it is worth noting that the larger the database, the less well siblings will rise to the very top. Regardless, investigating all of the parents/children and/or siblings of the very best candidates in a database without defined statistical thresholds can easily result in a prohibitively large number of false investigative leads (Williams and Johnson, 2005) at the same time that compelling questions are raised about the balance between collective security and individual privacy (Lazer and Meyer, 2004).

Paoletti et al. (2006) consider the effect of the alternative suspect pool and provide explicit formulae for calculating likelihood ratios of the actual perpetrator being a sibling or a parent/child (versus a randomly chosen, unrelated member of the pop-

ulation) for any arbitrary pair of suspect and evidence profiles. Paoletti et al. (2006) also describe the results of simulations that provide statistical boundaries on both the number and rarity of the alleles shared between an evidence sample and an excluded suspect necessary to determine if a significant shadow of suspicion is cast upon the excluded suspect's relatives. Paoletti et al. (2006) generate threshold values that allow correct prediction with a stated degree of confidence to provide a useful framework for using these formulae and provide empirical guidelines for such thresholds through extensive simulation. Alleles possessed by the initial suspect that are not found in the evidence sample (as well as the size of the database searched) play no role whatsoever in this evaluation beyond excluding them as a possible contributor of the evidence sample. Their analysis provides an objective framework for law enforcement agencies as well as a trier of fact for determining what level of similarity between an evidence sample and a single non-matching initial suspect constitutes sufficient grounds to warrant a specific investigation of a close relative of an initial suspect (e.g. the taking and genotyping of their DNA). Two important parameters, the size of the reasonable alternative suspect pool and the tolerance for false positives/negatives, are considered to be beyond the scope of forensic scientists and are left to be determined on a jurisdictional (and even case-by-case) basis.

5.4.2 Materials and methods

Simulated CODIS database

A set of complete STR-DNA profiles of individuals whose allele frequencies and allelic distributions are considered to be representative of those found in a larger United States population was needed for an evaluation of familial searching approaches. Such a dataset from the FBI, used for the determination of allele frequencies in Caucasians at the 13 CODIS loci, has already been analyzed for Hardy-Weinberg equilibrium (Budowle et al., 1999), and is publicly available (http://www.fbi.gov/hq/lab/fsc/backissu/july1999/dnaloci.txt.). All of the simulated database profiles in this analysis were generated from these 194 Caucasian 13 locus genotypes. The original FBI Caucasian dataset contains typing information for a larger number of individuals, but any with incomplete information (i.e. allele '0' at one or more loci) were discarded. The additional two loci needed to produce the fifteen locus PowerPlex® 16 profiles were generated using the Virginia Department of Forensic Science Caucasian population data (Commonwealth of Virginia Department of Forensic Science, 2006). The frequencies were derived from 101 Caucasians for the PentaD locus and 120 Caucasians for the PentaE locus.

Ten thousand pairs of unrelated individuals were created by randomly selecting alleles from the initial FBI dataset. Allele frequencies in this randomized dataset are the same as in the original dataset but individuals are unequivocally unrelated by descent (alleles are not the same because they have been faithfully passed from a common ancestor). Instead, any allele sharing can arise only through identity by

state (they are the same because there is a limited number of different alleles that can be detected). All individuals in the original dataset are assumed to have two and only two alleles per locus (rare conditions resulting in unusual allele counts such as null alleles, triploidy or chimerism are beyond the scope of this study).

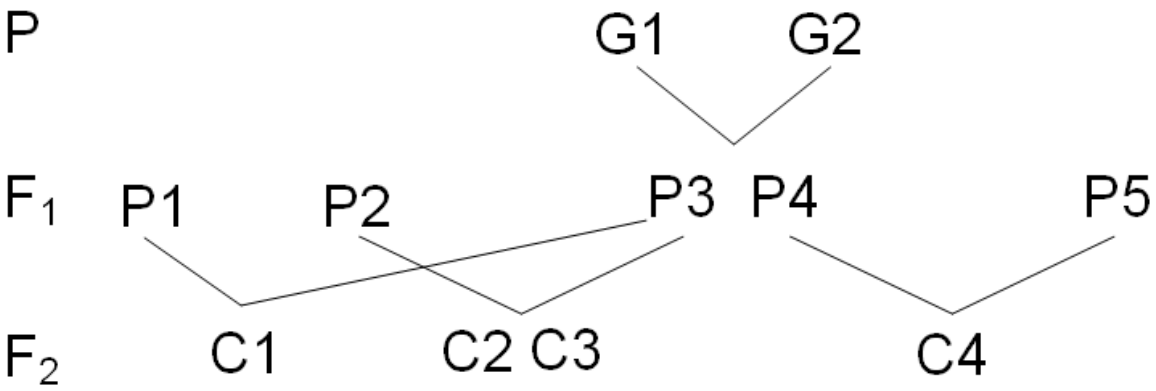

Figure 5.10: **The "virtual family" consisting of two grandparents, five parents, and four children. Individuals G1, G2, P1, P2, and P5 are randomly-generated (unrelated). Individuals C2 and C3 are selected as siblings, C1 and C2 are half-siblings, C3 and C4 are cousins, C3 and P4 form the uncle-nephew pair, P1 and C1 form the parent-child pair, and G1 and C2 form the grandparent-grandchild pair.**

Related individuals were created by generating "virtual families" as described by Paoletti et al. (Paoletti et al., 2005). Each family consists of two individuals in a P generation, five in an F_1 generation, and four in an F_2 generation (Figure 5.10). Virtual family members that did not arise from simulated matings within these families (i.e. grand parents) were drawn randomly without replacement from the set of 10,000 unrelated individuals. Each child was generated by randomly selecting one allele from each parent at each locus. Each family contained the following numbers (shown in parentheses) of pairs of specific familial relationships: grandparent-grandchild (8);

parent-child (8); half-sibling (2); sibling (1); aunt/uncle-niece/nephew (3) and cousin (3). A sufficient number of virtual families were created to generate 10,000 pairs of each of the different familial relationships. One individual from each pair was arbitrarily designated as the "source" of an evidentiary sample while the other individual was included in a searchable database. A simulated search of the database was considered to have generated a true positive "hit" when the evidentiary sample was matched with the sample with which it was originally paired.

CODIS search simulation

Each of the 10,000 pairs of individuals with the six specific familial relationships was evaluated using the existing CODIS database search criteria. Low stringency matches were counted by determining if the two profiles contained at least one allele in common across each of the thirteen CODIS loci. Medium stringency matches were counted by determining if both alleles were in common across all loci, with the caveat that a homozygous individual is considered to be a perfect match to a second individual if that individual contains one or both of the homozygous individual's alleles. Since it is very unlikely that high stringency matches (26/26 alleles) would be observed by chance alone among 10,000 pairs of individuals (regardless of relatedness), average random match probabilities were calculated with appropriate corrections for the degree of relatedness (National Research Council, 1996). In addition, the average degree of allele sharing, the average number of loci containing at least one allele in common, and the number of profiles pairs containing twenty or more alleles in common were counted.

Likelihood ratio approach

The likelihood ratio approach utilizes a likelihood ratio to determine if the source of the evidence is more likely a related individual of the partially-matching profile or an unrelated individual (Paoletti et al., 2006). The form of the likelihood ratio is:

$$LR = \frac{P(E|relative)}{P(E|random)} \tag{5.1}$$

where $P(E|relative)$ is the probability of choosing an individual of a given relation and $P(E|random)$ is the probability of choosing a random, unrelated individual. The denominator is simply the random match probability (National Research Council, 1996).

$$P(E|random) = P_a \cdot P_b \cdot HF \tag{5.2}$$

where P_a and P_b are the expected frequencies for the two alleles observed at a locus and the heterozygosity factor (HF) designates if the locus is homozygous ($HF = 1$) or heterozygous ($HF = 2$).

The appropriate equation for the numerator, $P(E|relative)$, depends on the inferred relationship and the number of alleles in common between the two partially-matching profiles at a given locus. To identify a sibling:

$$P(E|sib) = \begin{cases} \frac{P_a \cdot P_b \cdot HF}{4}, & if \quad shared = 0 \\ \frac{P_b + P_a \cdot P_b \cdot HF}{4}, & if \quad shared = 1 \\ \frac{1 + P_a + P_b + P_a \cdot P_b \cdot HF}{4}, & if \quad shared = 2 \end{cases} \tag{5.3}$$

To identify a parent or child:

$$P(E|parent/child) = \begin{cases} 0, & if \quad shared = 0 \\ \frac{P_b}{2}, & if \quad shared = 1 \\ \frac{P_a+P_b}{2}, & if \quad shared = 2 \end{cases} \tag{5.4}$$

The formulas for grandparents and grandchildren, aunts/uncles and nieces/nephews, and half-siblings are all the same:

$$P(E|GG/AUNN/HS) = \begin{cases} \frac{2 \cdot P_a \cdot P_b \cdot HF}{4}, & if \quad shared = 0 \\ \frac{P_b+2 \cdot P_a \cdot P_b \cdot HF}{4}, & if \quad shared = 1 \\ \frac{P_a+P_b+2 \cdot P_a \cdot P_b \cdot HF}{4}, & if \quad shared = 2 \end{cases} \tag{5.5}$$

To identify first cousins:

$$P(E|cousins) = \begin{cases} \frac{6 \cdot P_a \cdot P_b \cdot HF}{8}, & if \quad shared = 0 \\ \frac{P_b+6 \cdot P_a \cdot P_b \cdot HF}{8}, & if \quad shared = 1 \\ \frac{P_a+P_b+6 \cdot P_a \cdot P_b \cdot HF}{8}, & if \quad shared = 2 \end{cases} \tag{5.6}$$

The likelihood ratio with simulated databases

The likelihood ratio approach was also used for the same sets of 10,000 pairs of individuals. The number of profile pairs with likelihood ratios greater than 1 and 10,000 were counted (both indicating the greater likelihood of a related individual versus a randomly chosen, unrelated individual). In order to examine the effect of using fewer loci, all likelihood ratios were calculated using the nine Profiler Plus®

loci, the 13 CODIS loci, and the 15 PowerPlex® 16 loci.

Mixture studies

A separate database of 10,000 unrelated individuals was created for the purpose of mixture analyses. Simulated two-person mixtures were generated for each of the 10,000 pairs of individuals of each of the different familial relationships by adding the alleles of a single randomly chosen individual from the unrelated database to the alleles in the evidentiary sample. Three-person mixtures were generated by adding the alleles of two randomly chosen unrelated individuals. Likelihood ratios for simulated mixtures were calculated as described for unmixed samples but with a combined probability of inclusion (CPI; (Devlin, 1992; Ladd et al., 2000)) being used in place of the random match probability in the denominator. The form of the combined probability of inclusion for n observed alleles at a given locus is:

$$CPI : P_a P_b ... P_n = (P_a + P_b + ... + P_n)^2 \tag{5.7}$$

Likelihood ratios were calculated for each mixture to determine if the relative contained within the mixture would be identified using a likelihood ratio threshold of one or 10,000. In some instances, the likelihood ratio threshold was exceeded in the mixed sample, but not in the case of the single-source sample without the additional profile (in two person mixtures) or profiles (in three person mixtures). Additional counts were made for these instances.

False positives

An additional set of 10,000 pairs of unrelated individuals was created to determine false positive rates with the likelihood ratio approach using a range in numbers of loci. False positives were considered to have occurred when an unrelated pair exceeded the likelihood ratio threshold (either 1 or 10,000) for a given kind of familial relationship. When more than 15 loci were needed to achieve a false positive rate below 5% or 0.5%, additional loci were added to the existing genotypes of the 10,000 pairs of unrelated individuals by randomly choosing one of the 15 loci used and using the allele frequencies for that locus to generate an additional two alleles for each individual in the pairs.

Use of non-cognate databases

The racial or ethnic group of a contributor to an evidentiary sample is often unknown and/or a matter in dispute. In order to assess the implications of using an incorrect allele frequency database to generate likelihood ratios, 10,000 pairs of siblings, parent-child, and unrelated individuals that were generated with the FBI's and Virginia DFS' Caucasian dataset were also evaluated using the African American and Hispanic allele frequencies (Budowle and Moretti, 1999; Commonwealth of Virginia Department of Forensic Science, 2006). A consensus approach to using the likelihood ratio approach to recognizing true familial hits was also evaluated wherein the likelihood ratio had to exceed 1 (or 10,000) using all three frequency databases in order to be considered a "hit." A ceiling approach was also employed in which a single likelihood ratio was

calculated where the allele frequencies which were used to generate the likelihood ratio were the highest observed in any of the three different allele frequency databases.

5.4.3 Results

CODIS search simulation

A low stringency CODIS search drew attention to all 10,000 virtual parent-child pairs (as expected since since Mendelian genetics assures that they all at least one allele in common) (Table 5.4). However, a low stringency search only drew attention to 1813 of the known sibling pairs and less than 3% of the grandparent-grandchild, uncle-nephew, and half-sibling pairs were identified. Low stringency CODIS searches also resulted in "matches" between fifteen pairs of cousins and one of the 10,000 randomly generated pairs of unrelated individuals. No familial "hits" were observed when either moderate or high stringency searches were performed upon the sets of 10,000 pairs of individuals (Table 5.4).

A matching allele count approach using a threshold of 20 out of a possible 30 alleles identified only 4223 and 1882 of the 10,000 virtual sibling and parent-child pairs, respectively. The half-sibling, uncle-nephew, and grandparent-grandchild exhibited similar numbers of threshold matches with 59, 54, and 64 pairs, respectively. Only eight cousins exceeded the allele threshold. No pairs of unrelated individuals were identified with the allele threshold.

Relationship	Average Alleles	Std Dev Alleles	Average Loci	Std Dev Loci	CODIS High	CODIS Medium	CODIS Low	20+ allele matches
Siblings	19.0	2.47	13.4	1.20	3.58E-07	0	1813	4223
Parent-Child	18.1	1.57	15.0	0	3.03E-11	0	10000	1882
Half-sibling	13.9	2.26	11.7	1.57	7.96E-14	0	236	59
Cousins	11.8	2.36	10.1	1.80	1.17E-15	0	15	8
Uncle/Nephew	13.9	2.24	11.8	1.58	7.92E-14	0	236	54
Grandparent-Grandchild	14.0	2.26	11.8	1.58	7.96E-14	0	246	64
Unrelated	9.7	2.33	8.5	1.87	2.48E-18	0	1	0

Table 5.4: **CODIS search simulation using sets of 10,000 pairs of individuals of varying degrees of relatedness. The average loci value is the average number of loci having one or more alleles in common between the pair of individuals. The CODIS high stringency search value is the average random match probability for each set of individuals.**

	9 Loci		15 Loci		Only 9	Only 10
Relationship	**LR** > 1	**LR** > 10000	**LR** > 1	**LR** > 10000	**LR** > 1	**LR** > 10000
Actual Siblings : Unrelated	9932	2267	9991	6364	6	19
Actual Parent/Child : Unrelated	10000	668	10000	5611	0	3
Actual Half-Siblings : Unrelated	7342	0	7803	7	583	0
Actual Cousins : Unrelated	5726	0	5834	0	932	0
Actual Uncle/Nephew : Unrelated	7422	0	7842	9	617	0
Actual Grandparent/Grandchild : Unrelated	7367	0	7875	6	572	0
Incorrectly assumed siblings : Actual unrelated	354	0	141	0	269	0
Incorrectly assumed parent/child : Actual unrelated	42	0	1	0	41	0
Incorrectly assumed uncle/half-sib/grandparent : Actual unrelated	1449	0	872	0	836	0
Incorrectly assumed cousin : Actual unrelated	2484	0	1891	0	1097	0
Incorrectly assumed sibling : Actual parent/child	812	0	443	0	521	0
Incorrectly assumed parent/child : Actual sibling	1472	0	737	0	889	0

Table 5.5: **Likelihood ratio approach using sets of 10,000 pairs of individuals of varying degrees of relatedness using the nine Profiler Plus® and fifteen PowerPlex® 16 loci. The "only 9" columns show the number profiles that exceeded the likelihood ratio threshold with the nine Profiler Plus® loci, but not all fifteen PowerPlex® 16 loci. False positive rates are shown with the incorrectly assumed relations.**

Likelihood ratio approach

The likelihood ratio approach successfully identified more than 99% of both the sibling and parent-child pairs using a threshold of 1 (Table 5.5). When a much more stringent likelihood ratio threshold of 10,000 was utilized, 64% of the sibling and 56% of the parent-child pairs were still identified. More than 78% of the half-sibling, uncle-nephew, and grandparent-grandchild pairs were identified using a threshold of 1, however virtually none were identified with a threshold of 10,000. More than 58% of cousins were successfully identified with a threshold of 1, but none were identified with a threshold of 10,000.

False positive rates using unrelated individuals were lowest with incorrectly assumed parent-child and sibling pairs (0.01% and 1%, respectively). False positive rates grow as the level of relatedness decreases, with incorrectly assumed uncles/grandparents/cousins at 9% and incorrectly assumed cousins at 19%. No false positives were observed when using a likelihood ratio threshold of 10,000.

Non-cognate database study

The use of non-cognate frequency databases resulted in increases in related individuals being identified as well as unrelated individuals mistakenly identified as being related (Table 5.6). A ceiling approach resulted in the fewest number of false positives, but it also produced fewer true positives than the cognate and consensus frequency approach. The consensus frequency approach resulted in fewer false positives than if the cognate database was used alone, but it also resulted in slightly fewer true

Relationship	LR > 1	LR > 10000
Actual Caucasian Siblings : Unrelated Caucasian (all Caucasian Freq)	9991	6364
Actual Caucasian Siblings : Unrelated Caucasian (all Afr. Amer. Freq)	9999	8232
Actual Caucasian Siblings : Unrelated Caucasian (all Hispanic Freq)	9996	7472
Actual Caucasian Siblings : Unrelated Caucasian (Ceiling)	9926	3834
Sibling:Unrelated threshold exceeded in all three databases	9990	6249
Actual Caucasian Parent/Child : Unrelated Caucasian (all Caucasian Freq)	10000	5611
Actual Caucasian Parent/Child : Unrelated Caucasian (all Afr. Amer. Freq)	10000	8635
Actual Caucasian Parent/Child : Unrelated Caucasian (all Hispanic Freq)	10000	7399
Actual Caucasian Parent/Child : Unrelated Caucasian (Ceiling)	9995	1823
Parent:Unrelated threshold exceeded in all three databases	10000	5364
Incorrectly assumed Caucasian Siblings : Actual Caucasian Unrelated (all Caucasian Freq)	141	0
Incorrectly assumed Caucasian Siblings : Actual Caucasian Unrelated (all Afr. Amer. Freq)	1229	20
Incorrectly assumed Caucasian Siblings : Actual Caucasian Unrelated (all Hispanic Freq)	990	14
Incorrectly assumed Caucasian Siblings : Actual Caucasian Unrelated (Ceiling)	18	0
Incorrect Sib:Unrelated threshold exceeded in all three databases	127	0
Incorrectly assumed Caucasian Parent/Child : Actual Caucasian Unrelated (all Caucasian Freq)	1	0
Incorrectly assumed Caucasian Parent/Child : Actual Caucasian Unrelated (all Afr. Amer. Freq)	1	1
Incorrectly assumed Caucasian Parent/Child : Actual Caucasian Unrelated (all Hispanic Freq)	1	1
Incorrectly assumed Caucasian Parent/Child : Actual Caucasian Unrelated (Ceiling)	1	0
Incorrect Parent:Unrelated threshold exceeded in all three databases	1	0
Incorrectly assumed Caucasian Uncle : Actual Caucasian Unrelated (all Caucasian Freq)	872	0
Incorrectly assumed Caucasian Uncle : Actual Caucasian Unrelated (all Afr. Amer. Freq)	4150	9
Incorrectly assumed Caucasian Uncle : Actual Caucasian Unrelated (all Hispanic Freq)	3658	5
Incorrectly assumed Caucasian Uncle : Actual Caucasian Unrelated (Ceiling)	173	0
Incorrect Uncle:Unrelated threshold exceeded in all three databases	805	0
Incorrectly assumed Caucasian Cousin : Actual Caucasian Unrelated (all Caucasian Freq)	1891	0
Incorrectly assumed Caucasian Cousin : Actual Caucasian Unrelated (all Afr. Amer. Freq)	6401	0
Incorrectly assumed Caucasian Cousin : Actual Caucasian Unrelated (all Hispanic Freq)	5904	0
Incorrectly assumed Caucasian Cousin : Actual Caucasian Unrelated (Ceiling)	467	0
Incorrect Cousin:Unrelated threshold exceeded in all three databases	1748	0

Table 5.6: **The likelihood ratio approach using non-cognate allele frequencies. The profiles were generated using the FBI Caucasian allele frequencies for the CODIS loci and Virginia DFS Caucasian frequencies for the additional two PowerPlex® 16 loci. All profiles were evaluated using the FBI African American, Caucasian, and Hispanic allele frequencies. In addition, the ceiling principle was used to calculate the likelihood ratio using the highest observed allele frequency for each allele in a given profile. Finally, a consensus count was calculated by determining the number of profiles that exceeded the likelihood ratio using all three frequency databases.**

positives than the cognate database.

Mixture studies

The number of correctly identified sibling and parent-child pairs remained above 99% for single-source samples and two and three-person mixtures (Table 5.7). Half-sibling, uncle-nephew, and grandparent-grandchild pairs rates went from more than 75% for single-source samples to more than 86% for two and three-person mixtures. Cousins ranged from more than 58% for single-source samples to more than 74% for two and three-person mixtures.

Parent-child pairs experience the fewest false positives with 0.01% for single-source samples and two-person mixtures, and 0.07% for three-person mixtures. The second fewest false positives were observed in siblings, but the observed range was far greater. False identification of siblings ranged from 1% for single-source samples to 10% for two-person mixtures to 15% for three-person mixtures. Incorrectly assumed half-sibling/uncle-nephew/grandparent-grandchild pairs ranged from 9% for single-source samples to 22% for two-person mixtures to 30% for three-person mixtures. The largest false positive rate was observed with incorrectly assumed cousins, with rates of 19% for single-source samples, 41% for two-person mixtures, and 49% for three-person mixtures.

False positive locus study

Use of 13 loci allows for less than a 5% false positive rate for the identification of sibling, parent-child, half-sibling, uncle-nephew, and grandparent-grandchild pairs

	Single-source		2-person mixture		3-person mixture	
Relationship	**LR** > 1	**LR** > 10000	**LR** > 1	**LR** > 10000	**LR** > 1	**LR** > 10000
Actual Siblings : Unrelated	9991	6364	9999	8216	9999	8301
Actual Parent/Child : Unrelated	10000	5611	10000	7727	10000	7782
Actual Half-Siblings : Unrelated	7803	7	8653	48	8699	53
Actual Cousins : Unrelated	5834	0	7422	0	7721	0
Actual Uncle/Nephew : Unrelated	7842	9	8730	78	8828	82
Actual Grandparent/Grandchild : Unrelated	7875	6	8803	68	8915	83
Incorrectly assumed siblings : Actual unrelated	141	0	1054	6	1546	14
Incorrectly assumed parent/child : Actual unrelated	1	0	1	0	7	3
Incorrectly assumed uncle/half-sib/grandparent : Actual unrelated	872	0	2241	0	2952	0
Incorrectly assumed cousin : Actual unrelated	1891	0	4081	0	4918	0

Table 5.7: **The likelihood ratio approach with single-source, two-person, and three-person simulated mixtures.**

Incorrectly assumed relationship (actually unrelated)	Number of loci needed to observe $\leq 5\%$ error	Number of loci needed to observe $\leq 0.5\%$ error
Sibling	7	24
Parent/Child	5	11
Uncle/Half-Sib/Grandparent	13	46
Cousin	26	103

Table 5.8: **The number of loci needed to observe less than a 5% false positive rate. Loci were randomly added until the false positive rate for a given relation was less than 5% and 0.5%.**

(Table 5.8) using the likelihood ratio approach and a threshold of 1. Twenty-six loci are required to successfully identify cousins with this approach at the same false positive rate. If the desired false positive rate is at or below 0.5%, only parent-child pairs can be identified using the currently available 13 CODIS loci. False positive rates for siblings only fall below 0.5% when the equivalent of 24 loci (e.g. the 13 CODIS loci plus an additional 11 loci with similar discriminating power) are used while half-sibling, uncle-nephew, and grandparent-grandchild pairs require 46 loci, and cousins require 106 loci.

5.4.4 Discussion

Simulations of the CODIS search software suggest that it does not lend itself to the task of performing familial searches. Of all relatives tested, only the majority of parent-child pairs were successfully counted and only when performing low-stringency searches. Siblings are often of greater interest in familial searches and less than a quarter were counted using the same search method. Less than 5% of all other relatives were identified using the approach supported by the current CODIS software.

The use of pre-established threshold likelihood ratios in the kinship analyses per-

formed here allow a wide range of false positive and negative investigative leads to be generated. The appropriate balance between the two is something that is beyond the scope of a forensic scientist to determine and should be left to a policy-making body - perhaps at the level of a state legislature. Utilizing a likelihood ratios of 1 (i.e. suggesting that a single, randomly chosen sibling of an individual in a database is at least as likely as a single, randomly chosen individual from a given reference population to match the DNA profile associated with an evidence sample) may result in an appreciable number of false positives, while a likelihood ratio of 10,000 has a much lower risk of generating false investigative leads. The establishment of a pre-determined threshold should also take in to consideration the implications of performing multiple searches on databases with large numbers of individuals (the larger the database, the larger the threshold needs to be to achieve the same balance of false positives and negatives). It might also be reasonable to perform additional testing (e.g. generate Y-STR profiles) in some instances where a pre-established threshold has not been surpassed but another minimal threshold has been exceeded. Given that the allele frequency database that should be used in the course of kinship analysis is often unknown or a matter of dispute, it will probably be necessary to generate likelihood ratios using either the ceiling or consensus approach described in this study.

The success of familial searches in a relatively small number of cases will undoubtedly lead to a more widespread adoption among the law enforcement community. However, current methods such as allele counting result are arbitrary and inefficient and can easily lead to the generation of a prohibitively large number of false investigative leads. A false lead not only wastes the time and resources of law enforcement, but

it may be seen as an intrusion of privacy and civil rights. A likelihood ratio kinship analysis approach with pre-established thresholds is much more likely to reveal the presence of a familial hit, even in the seemingly difficult case of an evidential mixture or in circumstances where an appropriate reference population is not available for the determination of allele frequencies.

Chapter 6

Conclusions and future work

The work presented here is a collection of methods that improve upon the state-of-the-art by introducing statistically sound models for the identification of several commonly-encountered issues in the analysis and interpretation of forensic DNA evidence and thereby limit the introduction of subjective judgment. The frameworks provide solutions that can be integrated into the operating protocols of any laboratory or expert. To facilitate adoption, software packages have been made available for the determination of a run-specific RFU threshold and the resolution of forensic DNA mixtures (www.bioforensics.com). Software packages for the remaining studies may be released to the public in the future.

The RFU threshold is the height (in RFUs) at which the DNA analysis software will identify peaks for interpretation. According to validation standards, each testing laboratory is supposed to develop their own RFU threshold based on the sensitivity of their particular testing environment. However, most testing laboratories have adopted a single RFU threshold (typically 150 RFUs) that is utilized for all of their

subsequent DNA analyses. Changes in the age of the reagents, machinery, and staff can lead to different levels of background noise. There is the potential for useful information to be hidden by an overly high threshold or for noise to be observed with an overly low threshold. A run-specific limit of detection and quantitation allows an analyst to create a statistically-based RFU threshold that takes into consideration the level of noise in the system during a particular analysis run. A commonly-employed limit of detection (LOD) is the average baseline value plus three standard deviations. The LOD is the point at which signal can be reliably (with 99.7% confidence) be distinguished from noise. A commonly-employed limit of quantitation (LOQ) is the average baseline value plus ten standard deviations. If signal exceeds the LOQ, the contribution of background signal is relatively small enough to allow not only the detection of signal, but also its measurement. An LOD-LOQ validation study was performed utilizing 150 control samples from 50 analysis runs. In addition, software was developed to allow any laboratory or expert to employ this procedure.

DNA is a relatively stable molecule, but environmental factors, such as UV sunlight and bacteria, can lead to its breakdown. STR loci are a variety of different sizes, so the rate of degradation differs, with larger loci being the first to break down. The danger with any degraded sample is that important information may be in such small quantities that they are not able to be detected by the genetic analyzer (a phenomenon known as allelic dropout). Therefore, samples exhibiting potential degradation must be interpreted with caution. There is currently no established method for identifying samples that may exhibit degradation. Degraded samples typically exhibit a downward slope when examining the heights of the peaks from left to right. A slope can be

calculated using a best-fit linear regression. A population of 164 positive control sample slopes was developed to create a distribution of undeniably non-degraded samples for comparison. If a given sample's slope falls outside the range of a given confidence interval in the distribution, then it can be said to be sufficiently different from the population of non-degraded samples to warrant further examination.

The interpretation of DNA mixtures is notoriously problematic due to the large number of potential contributors and contributor profiles. A mathematical framework has been developed to enumerate all possible mixture combinations and determine which contributor combinations are possible using a small set of user-defined assumptions. Special considerations like allelic dropout are included in the calculations. Additional information, such as the victim's profile, can be used to eliminate potential mixture combinations. A validation study was performed on a set of known mixtures. The method was able to correctly resolve at least one locus in every mixture examined. Errors were only introduced when the mixtures entered a relatively extreme range of 1:6 or greater. The issues were caused by the presence of a minor contributor at low enough levels to be confused with a stutter artifact (and be filtered by the analysis software) or the contribution of stutter from a major contributor to cause the minor contributor's peaks to become out of balance. In addition, a software package called GenoStat™ was developed to provide anyone the ability to perform a mixture resolution using this method and calculate DNA statistics.

Most investigations involve a known suspect or set or suspects and one or more pieces of evidence. When no known suspect is available, a database search is typically performed to attempt to identify the perpetrator. These investigations are referred

to as cold hit cases. The weight of DNA evidence is measured in the probability of a coincidental match. In a traditional case, the random match probability is used to determine the chance of randomly selecting an unrelated person from the population with a given profile. With cold hit cases, there is currently a debate as to the weight to associate with the evidence identified with a DNA database search (see Chapter Two). Almost all currently-available methods rely on the assumption that the population of alternative suspects is completely unrelated. DNA databases may contain related individuals that may adversely affect the chance of a coincidental match. The size of the database will also affect the chance of a coincidental match. A pairwise search of a DNA database against itself can yield insight into the presence of related individuals and the chance of a coincidental match. Several simulated databases were created of different sizes. Simulated relatives were then added to determine the effect of related individuals on the relative chance of a coincidental match. The presence of siblings made the only appreciable difference in the rate of profile sharing. The number of related individuals in a real-world database can be inferred by comparing the pairwise analysis results of a real-world database to the simulation results.

The success of familial searches in a relatively small number of cases will undoubtedly lead to a more widespread adoption among the law enforcement community. However, current methods such as allele counting result in several candidates for investigation. A false lead not only wastes the time and resources of law enforcement, but it may be seen as some as an intrusion of privacy and civil rights. A likelihood ratio approach is more scientifically sound and is more likely to reveal the presence of a familial match, even in the seemingly difficult case of an evidential mixture. The

number of false positives can be limited (and in many cases eliminated) by increasing the likelihood ratio threshold beyond one. Additionally, if Y-STR profiles are available, then in many circumstances a familial relationship can be confirmed without the need for locating and testing relatives of individuals in a database. The race of the true perpetrator may not be known, especially if the perpetrator is unrelated to the database profile. A consensus among several population databases is able to produce similar results to that of the correct database being used alone. The framework described here can be adopted by any law enforcement agency and could be incorporated into the CODIS DNA database software for automated analysis.

6.1 Future work

The work presented here address several open questions, but the underlying issues are not completely resolved. Each project can be expanded to improve the reliability of dealing with testing issues. The newly-developed techniques may provide the research framework for creating methodologies for identifying additional issues that currently lack rigorous statistical standards of identification.

The run-specific RFU threshold method presented in Chapter Three utilizes an existing methodology that has been in place for many years in the field of analytical chemistry. Many disciplines face the issue of detecting signal in the presence of noise. More modern techniques should be assessed to determine if they can capture more information than a relatively simple limit of detection or quantitation. In addition, the issue of baseline noise is only one of many technical artifacts that can arise during

testing (see Chapter Two). Prior work (Gilder, 2003) could be expanded to better characterize other technical artifacts, such as spikes and blobs, using metrics such as the peak-area to peak-height ratio.

The run-specific RFU threshold methodology could be experimentally validated to determine if the LOD and LOQ adequately classify the levels of noise observed in actual casework samples. The hypothesis is that the LOD/LOQ is a conservative estimate of the level of noise present in a sample due to the fact that in order for noise to be confused with true signal, it must be present in the typable range of the allelic ladder and be of a sufficient shape to be labeled as a peak by the DNA analysis software's peak detection algorithm. The validation would entail injecting a known single-source sample 50 times over five runs. There would be ten injections per run and each would include a negative control, positive control, and reagent blank. A population of 50 samples is large enough to satisfy the SWGDAM guidelines for validation (Scientific Working Group on DNA Analysis Methods, 2000) and the requirements of most statistical tools. The controls would be used to determine the LOD and LOQ for each run, which would then be used as the RFU threshold for the GeneScan®/Genotyper® analysis. The number of on-ladder peaks not consistent with the template DNA profile or technical artifacts not due to baseline noise (e.g. stutter, pull-up) would be determined. The number of data collection points falling above the LOD and LOQ would also be counted to provide a more direct comparison to the LOD/LOQ framework. Dividing the samples into multiple runs with multiple injections allows for a more complete examination of the variability of baseline noise within and between analysis runs.

Additional studies could be performed to examine the utility of the run-specific RFU threshold framework in a variety of casework situations. An additional study using a known 1:10 two-person mixture could be used to determine how well the LOD/LOQ framework is able to detect the profile of the minor contributor in a situation that is typically difficult with a static threshold. The level of baseline noise may increase with additional rounds of PCR amplification, such as in low copy number (LCN) testing. The issue could be addressed by performing 34 rounds of PCR amplification on a known single-source sample (and respective controls) and repeating the above experiment. In addition, the data from any of these studies could be used to study the variability of peak heights over multiple injections.

A framework for identifying potentially degraded and inhibited samples has been established in Chapter Three, but work still remains in developing the ideal statistical thresholds for identification. The population of samples could be explored to determine why the current population deviates from normality and develop a new population that better approximates a normal distribution. In addition, data from the systematic degradation of known DNA samples could be used to evaluate the current methodology and determine the threshold where a sample no longer falls within the distribution of non-degraded positive controls. Additional issues such as mixtures with one or more degraded components could be explored to better identify when allelic dropout is likely to occur.

The mixture resolution framework presented in Chapter Four can be expanded to incorporate additional knowledge beyond peak balance and additivity. Additional work could be done to examine the variability of mixture ratios across loci. Examin-

ing intra-locus peak balance may also assist in determining the most likely contributor profiles. A better understanding of the expected variances can place additional boundaries on the possible mixture contributor profiles. Understanding expected inter-locus variability can also assist in the detection of degraded components of mixtures. Degraded contributors often exhibit larger peak variability, which could then be accounted for. Finally, a stutter artifact from one contributor can inflate the height of another individual's peak, which can adversely affect the interpretation of low-level contributors. Additional work could be done to consider the possible contributions of stutter and create an additional mixture resolution assessment.

The presence of related individuals in a DNA database has been shown to adversely affect the chance of a coincidental match in a database, but the work presented in Chapter Five partially elucidates the extent of this problem. Additional work could be done to evaluate DNA databases currently in use and develop a statistical correction to better represent the conditions of relatedness commonly observed. Future work also includes developing a mathematical model to approximate the number of related individuals in a database without the need for a comparison with simulation results.

A proposed framework for conducting a familial search is presented in Chapter Five. Additional work could include the evaluation of the DNA profiles of relatives with DNA databases currently in use. In addition, the evaluation of databases of different sizes could assist in determining the ideal likelihood ratio threshold for identifying a related individual that results in the fewest false leads.

Allelic dropout is a major issue when evaluating a DNA profile. The probability of experiencing allelic or locus dropout has not been extensively studied. Combining

information gained from the degradation identification and run-specific RFU threshold models from Chapter Three may provide valuable information in determining whether allelic dropout has occurred. Allelic dropout is more likely to occur when dealing with degraded samples with a large amount of baseline noise. Allelic dropout is particularly of interest in mixed samples, where it may not as noticeable. A mixture validation study may provide the information to determine the mixture ratios (and more importantly mixture DNA component quantities) where allelic dropout is most likely to occur. With enough information, a probabilistic model could be developed that could be incorporated into the existing statistics formulae to better determine the probability of a coincidental profile match.

The area of forensic DNA interpretation is an expanding field with new challenges continually presenting themselves. New methodologies will give rise to the development of new techniques to address additional issues. As more issues are identified, the underlying body of data will grow and previous issues can be re-examined to develop even more reliable thresholds. These tools can be used to provide the purveyors of justice the information necessary to evaluate the evidence with the most scientific means available.

The confidence in testing results will increase with the number of issues identified and the introduction of subjective judgment will be greatly limited. This increase in objectivity and statistical metric has the very real potential to have a significant long term impact on justice in our society.

Appendices

Glossary of forensic DNA terminology

A majority of these definitions were taken from DNA Technology in Forensic Science, (1992) National Research Council, Washington, D.C.: National Academy Press and STRBase (National Institute of Standards and Technology, 2007).

Adenine : A purine base; one of the four molecules containing nitrogen present in the nucleic acids DNA and RNA; designated by letter A.

Allele : An alternative form of a gene.

Allele Frequency : The proportion of a particular allele among the chromosomes carried by individuals in a population.

Amino acid : Any of a class of 20 molecules that are combined to form proteins in living things. The sequence of amino acids in a protein and hence protein function are determined by the genetic code.

Amplification : An increase in the number of copies of a specific DNA fragment; can be in vivo or in vitro.

Autosome : A chromosome not involved in sex determination. The diploid human genome consists of 46 chromosomes, 22 pairs of autosomes, and one pair of sex chromosomes (the X and Y chromosomes).

Basepair : Two complementary nucleotides joined by hydrogen bonds; basepairing occurs between A and T and between G and C.

Base sequence : The order of nucleotide bases in a DNA molecule.

Base sequence analysis : A method, sometimes automated, for determining the base sequence.

Biotechnology : A set of biological techniques developed through basic research and now applied to research and product development.

Chromosome :The structure by which hereditary information is physically transmitted from one generation to the next.

Complementary sequences : Nucleic acid base sequences that form a double-stranded structure by matching base pairs; the complementary sequence to G-T-A-C is C-A-T-G.

Cytosine : A pyrimidine base; one of the four molecules containing nitrogen present in the nucleic acids DNA and RNA; designated by letter C.

Denaturation : the process of splitting the complementary double strands of DNA to form single strands

Deoxyribonucleic acid (DNA) : The genetic material of organisms, usually double-stranded; a class of nucleic acids identified by the presence of deoxyribose, a sugar, and the four nucleobases.

DNA sequence : The relative order of base pairs, whether in a fragment of DNA, a gene, a chromosome, or an entire genome.

Double Helix : The shape that two linear strands of DNA assume when bonded together.

Electrophoresis : a technique in which molecules are separated by their velocity in an electric field

Enzyme : A protein that can speed up a specific chemical reaction without being changed or consumed in the process.

Gel : semisolid matrix (usually agarose or acrylamide) used in electrophoresis to separate molecules

Gene : the basic unit of heredity; a sequence of DNA nucleotides on a chromosome

Gene frequency : the relative occurrence of a particular allele in a population

Gene mapping : Determination of the relative positions of genes on a DNA molecule (chromosome or plasmid) and of the distance, in linkage units or physical units, between them.

Genetics : The study of the patterns of inheritance of specific traits.

Genome : All the genetic material in the chromosomes of a particular organism; its size is generally given as the total number of base pairs.

Genome projects : Research and technology development efforts aimed at mapping and sequencing some or all of the genome of an organism.

Genotype : the genetic makeup of an organism, as characterized by its physical appearance or phenotype

Guanine : a purine base; one of the four molecules containing nitrogen present in the nucleic acids DNA and RNA; designated by letter G

Heredity : the transmission of characteristics from one generation to the next

Heterozygosity : The presence of different alleles at one or more loci on homologous chromosomes.

Homologies : Similarities in DNA or protein sequences between individuals of the same linear sequences, each derived from one parent.

Homologous chromosomes : A pair of chromosomes containing the same linear gene sequences, each derived from one parent.

In vitro : Outside a living organism

Kilobase (kb) : Unit of length for DNA fragments equal to 1000 nucleotides.

Linkage : The proximity of two or more markers (genes, etc.) on a chromosome; the closer together the markers are, the lower the probability that they will be separated during DNA repair or replication process, and hence the greater the probability that they will be inherited together.

Localize : Determination of the original position (locus) of a gene or other marker on a chromosome.

Locus (pl. loci) : The specific physical location of a gene on a chromosome.

Marker : A gene of known location on a chromosome and phenotype that is used as a point of reference in the mapping of other loci.

Megabase (Mb) : Unit of length for DNA fragments equal to one million nucleotides.

Mitosis : The process of nuclear division in cells that produces daughter cells that are genetically identical to each other and to the parent.

Multiplexing : A sequencing approach that uses several pooled samples simultaneously, greatly increasing sequencing speed.

Mutation : Any inheritable change in DNA sequence.

Nucleic acid : A nucleotide polymer that DNA and RNA are major types.

Nucleotide : A unit of nucleic acid composed of phosphate, ribose or deoxyribose, and a purine or pyrimidine base.

Nucleus : The cellular organelle in eukaryotes that contains the genetic material.

Oncogene : A gene, one or more forms of which is associated with cancer. Many oncogenes are involved, directly or indirectly, in controlling the rate of cell growth.

Physical map : A map of the locations of identifiable landmarks on DNA. Distance is measured in base pairs.

Polymerase chain reaction (PCR) : An in vitro process that yields millions of copies of desired DNA through repeated cycling of a reaction involving the DNA polymerase enzyme.

Polymerase, DNA or RNA : Enzymes that catalyze the synthesis of nucleic acids on preexisting nucleic acid templates, assembling RNA from ribonucleotides or DNA from deoxyribonucleotides.

Polymorphism : Difference in DNA sequence among individuals. Genetic variations occuring in more than 1% of a population would be considered useful polymorphisms for linkage analysis.

Population : A group of individuals residing in a given area at a given time.

Primer : Short preexisting polynucleotide chain to which new deoxyribonucleotides can be added by DNA polymerase.

Probe : Single-stranded DNA or RNA of a specific base sequence, labeled either radioactively or immunology, that are used to detect the complementary base sequence by hybridization.

Protein : A large molecule composed of one or more chains of amino acids in a specific order; the order is determined by the base sequence of nuceotides in the gene coding for the protein. Proteins are required for the structure, function, and regulation of the body cells, tissues, organs, and each protein has unique functions.

Recombinant DNA technologies : Procedurese used to join together DNA sequences in a cell-free system. Under appropriate conditions, a recombinant DNA molecule can enter a cell and replicate there, either autonomously or after it has become integrated into a cellular chromosome.

Resolution : Degree of molecular detail on a physical map of DNA.

Restriction enzyme : A protein that recognizes specific, short nucleotide sequences

and cuts DNA at the those sites. Bacteria contain over 400 such enzymes that recognize and cut over 100 DNA sequences.

Restriction fragment length polymorphism (RFLP) : Variation between individuals in DNA fragment sizes cut by specific restriction enzymes; polymorphic sequences that result in RFLPs that are used as markers on both physical maps and genetic linkage maps. RFLPs are usually caused by mutation at a cutting site.

Sequencing : Determination of the order of nucleotides (base sequences) in a DNA or RNA molecule or the order of amino acids in a protein.

Sex chromosomes (X and Y chromosomes) : Chromosomes that are different in the two sexes and involved in sex determination.

Short tandem repeats (STR) : Multiple copies of an identical DNA sequence arranged in direct succession in a particular region of a chromosome.

Southern blotting : Transfer by absorbtion of DNA fragments separated in electrophoretic gels to membrane filters for detection of specific base sequences by radiolabeled complementary probes.

Tandem repeat sequences : Multiple copies of the same base sequence on a chromosome; used as a marker in physical mapping.

Thymine : a pyrimidine base; one of the four molecules containing nitrogen present in the nucleic acids DNA and RNA; designated by letter T

Variable number tandem repeats (VNTR) : repeating units of a DNA sequence which number varies between individuals

Bibliography

Bibliography

Adams, D., Presley, L., Baumstark, A., Hensley, K., Hill, A., Anoe, K., Campbell, P., McLaughlin, C., Budowle, B., Giusti, A., Smerick, J., and Baechtel, F. (1991). Deoxyribonucleic acid (DNA) analysis by restriction fragment length polymorphisms of blood and other body fluid stains subjected to contamination and environmental insults. *Journal of Forensic Sciences*, 36(5):1284–1298.

Akane, A., Matsubara, K., Nakamura, H., Takahashi, S., and Kimura, K. (1994). Identification of the heme compound copurified with deoxyribonucleic acid (DNA) from bloodstains, a major inhibitor of polymerase chain reaction (PCR) amplification. *Journal of Forensic Sciences*, 39(2):362–372.

Anderson, N. (1989). Determination of the lower limit of detection [letter]. *Clinical Chemistry*, 35:2152–2153.

Applied Biosystems (2000a). *AmpFlSTR Profiler Plus PCR amplification kit user's manual*. Foster City, CA.

Applied Biosystems (2000b). *Chemistry reference for the ABI Prism 310 genetic analyzer*. Foster City, CA.

Applied Biosystems (2005a). *AmpFlSTR COfiler PCR amplification kit user's bulletin*. Foster City, CA.

Applied Biosystems (2005b). *AmpFlSTR Identifiler PCR amplification kit user's manual*. Foster City, CA.

Applied Biosystems (2007). Applied Biosystems home page. *http://www.-appliedbiosystems.com*.

Arinbuster, D., Tillman, M., and Hubbs, L. (1994). Limit of Detection (LOD)/Limit of Quantitation (LOQ): Comparison of the empirical and the statistical methods exemplified with GC-MS assays of abused drugs. *Clinical Chemistry*, 40:1233–1238.

Balding, D. and Donnelly, P. (1996). Evaluating DNA profile evidence when the suspect is identified through a database search. *Journal of Forensic Sciences*, 41(4):603–607.

Bieber, F., Brenner, C., and Lazer, D. (2006). Finding criminals through DNA of their relatives. *Science*, 312:1315–1316.

Brenner, C. (2006). Some mathematical problems in the DNA identification of victims in the 2004 tsunami and similar mass fatalities. *Forensic Science International*, 157:172–180.

Brenner, C. and Weir, B. (2006). Issues and strategies in the identification of World Trade Center victims. *Theoretical Population Biology*, 63(3):173–178.

Budowle, B., Bieber, F., and Eisenberg, A. (2005). Forensic aspects of mass disasters: Strategic considerations for DNA-based human identification. *Legal Medicine*, 7(4):230–243.

Budowle, B. and Moretti, T. (1999). Genotype profiles for six population groups at the 13 CODIS short tandem repeat core loci and other PCR based loci. *Forensic Science Communications*, 1(2).

Budowle, B., Moretti, T., Baumstark, A., Defenbaugh, D., and Keys, K. (1999). Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. *Journal of Forensic Sciences*, 44(6):1277–1286.

Butler, J. M. (2001). *Forensic DNA Typing*. Academic Press, San Diego, CA.

Campbell, N. A. (1996). *Biology*. Benjamin/Cummings Publishing Company, New York, NY, 4th edition.

Chakraborty, R. and Kidd, K. (1991). The utility of DNA typing in forensic work. *Science*, 254:1735–1739.

Clayton, T., Whitacker, J., Sparkes, R., and Gill, P. (1998). Analysis and interpretation of mixed forensic stains using DNA STR profiling. *Forensic Science International*, 91:55–70.

Commonwealth of Virginia Department of Forensic Science (2006). Fluorescent detection PCR-based STR DNA protocol: Powerplex® 16 BIO system. *http://-www.dfs.virginia.gov/manuals/manuals.cfm?id=5*, (3).

Curran, J., Triggs, C., Buckleton, J., and Weir, B. (1999). Interpreting DNA mixtures in structured populations. *Journal of Forensic Sciences*, 44(5):987–995.

DeFranchis, R., Cross, N., Foulkes, N., and Cox, T. (1988). A potent inhibitor of Taq polymerase copurifies with human genomic DNA. *Nucleic Acids Research*, 16(21):10355.

Dennis, C., Gallagher, R., and Watson, J., editors (2002). *The Human Genome*. Palgrave Macmillan, New York, NY.

Devlin, B. (1992). Forensic inference from genetic markers. *Statistical Methods in Medical Research*, 2:241–262.

Devore, J. (2000). *Probability and Statistics for Engineering and the Sciences*. Pacific Grove, Duxbury, 5th revised edition.

DNA Advisory Board (2000a). Quality assurance standards for forensic DNA testing laboratories. *Forensic Science Communications*, 2(3).

DNA Advisory Board (2000b). Statistical and population genetics issues affecting the evaluation of the frequency of occurrence of DNA profiles calculated from pertinent population database(s). *Forensic Science Communications*, 2(3).

Edwards, A., Hammond, H., Lin, J., Caskey, C., and Chakraborty, R. (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics*, 12:241–253.

Evett, I., Foreman, L., Lambert, J., and Emes, A. (1998). Using a tree diagram to interpret a mixed DNA profile. *Journal of Forensic Sciences*, 43(3):472–476.

Evett, I. and Lambert, P. G. J. (1998). Taking account of peak areas when interpreting mixed DNA profiles. *Journal of Forensic Sciences*, 43(1):62–69.

Federal Bureau of Investigation (2005). National DNA Index System (NDIS) data acceptance standards.

Federal Bureau of Investigation (2007). FBI Combined DNA Index System (CODIS). *http://www.fbi.gov/hq/lab/codis/index1.htm*.

Findlay, I., Taylor, A., Quirke, P., Frazier, R., and Urquhart, A. (1997). DNA fingerprinting from single cells. *Nature*, 389:555–556.

Ford, S., Gilder, J., Rowland, C., Kelly, B., and Raymer, M. (2004). Validation study of the Genophiler® automated data review system. In *The 16th International Symposium on Human Identification*.

Frank, W., Llewellyn, B., Fish, P., Riech, A., Marcacci, T., Gandor, D., Parker, D., RR, R. C., and Thibault, S. (2001). Validation of the AmpFLSTR™ Profiler Plus PCR Amplification Kit for use in forensic casework. *Journal of Forensic Sciences*, 46(3):642–646.

Fregèau, C., Aubin, R., Elliot, J., Gill, S., and Fourney, R. (1995). Characterization of human lymphoid cell lines gm9947 and gm9948 as intra- and interlaboratory reference standards for dna typing. *Genomics*, 28:184–197.

Fregèau, C., Bowen, K., and Fourney, R. (1999). Validations of highly polymorphic fluorescent multiplex short tandem repeat systems using two generations of DNA sequencers. *Journal of Forensic Sciences*, 44(1):133–166.

Fregèau, C. and Fourney, R. (1993). DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. *BioTechniques*, 15:100–119.

Gilder, J. (2003). Developing an expert system and discovering new standards for forensic DNA analysis. Master's thesis, Wright State University, Dayton, Ohio.

Gilder, J., Anderson, P., Doom, T., Raymer, M., and Krane, D. (2007a). GenoStat software homepage. *http://www.bioforensics.com/genostat*.

Gilder, J., Doom, T., Inman, K., and Krane, D. (2007b). Run-specific limits of detection and quantitation for STR-based DNA testing. *Journal of Forensic Sciences*, 52(1):97–101.

Gilder, J., Ford, S., Doom, T., Raymer, M., and Krane, D. (2004). Systematic differences in electropherogram peak heights reported by different version of the GeneScan® software. *Journal of Forensic Sciences*, 49(1):85–92.

Gill, P., Ivanov, P., Kimpton, C., Piercy, R., Benson, N., Tully, G., Evett, I., Hagelberg, E., and Sullivan, K. (1994). Identification of the remains of the Romanov family by DNA analysis. *Nature Genetics*, 6(2):130–135.

Gill, P., Sparkes, R., Pinchin, R., Clayton, T., Whitaker, J., and Buckleton, J. (1998). Interpreting simple STR mixtures using allele peak area. *Forensic Science International*, 91:41–53.

Handt, O., Hoss, M., Krings, M., and Paabo, S. (1994). Ancient DNA: methodological challenges. *Experientia*, 50(6):524–529.

Holt, C., Buoncristiani, M., Wallin, J., Nguyen, T., Lazaruk, K., and Walsh, P. (2002). TWGDAM validation of the AmpFlSTR™ PCR Amplification Kits for forensic casework analysis. *Journal of Forensic Sciences*, 47(1):66–96.

Ivanov, P., Wadhams, M., Roby, R., Holland, M., Weedn, V., and Parsons, T. (1996). Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nature Genetics*, 12(4):417–420.

Johnson, R. (2005). In the matter of the application of the state of California for an order requiring: Custodian of Records Arizona Department of Public Safety product documents/DNA database unit. Transcript of proceedings.

Johnstone, G. (2006). Inquest into the death of Jaidyn Raymond Leskie. *Coroners Case Number: 007/98*.

Kafarowski, E., Lyon, A., and Sloan, M. (1996). The retention and transfer of spermatozoa in clothing by machine washing. *Canadian Society of Forensic Science Journal*, 20(1):7–11.

Kevles, D. (2003). Ownership and identity. *The Scientist*, 17(1).

Kimpton, C., Gill, P., Walton, A., Millican, E., and Adams, M. (1993). Automated DNA profiling employing multiplex amplification of short tandem repeat loci. *PCR Methods and Applications*, 3:13–22.

Kinsey, P. and Hormann, S. (2000). Modification of the stutter position label-filtering macro in the PE Biosystems Genotyper® version 2.5 software package: resolution of stutter-filter back talk. *Forensic Science Communications*, 2(3).

Krenke, B., Tereba, A., Anderson, S., Buel, E., Culhane, S., Finis, C., Tomsey, C., Zachetti, J., and Sprecher, C. (2002). Validation of a 16-locus fluorescent multiplex system. *Journal of Forensic Sciences*, 47(4):773–785.

Ladd, C., Lee, H., and Beiber, F. (2000). Probability of exclusion estimates in forensic analysis of complex DNA mixtures. In *Proceedings of the American Academy of Forensic Sciences 52nd annual meeting*.

Lander, E. (1991). Lander reply. (letter.). *American Journal of Human Genetics*, 49:899–903.

Lazer, D. and Meyer, M. (2004). *DNA and the Criminal Justice System*. Cambridge, MA.

Leclair, B., Fregèau, C., Bowen, K., and RM, R. F. (2004). Systematic analysis of stutter percentages and allele peak height and peak area ratios at heterozygous STR loci for forensic casework and database samples. *Journal of Forensic Sciences*, 49(5):1–13.

Leclair, B., Sgueglia, J., Wojtowicz, P., Juston, A., Fregèau, C., and Fourney, R. (2003). STR DNA typing: increased sensitivity and efficient sample consumption using reduced PCR reaction volumes. *Journal of Forensic Sciences*, 48(5):1001–1013.

Lewontin, R. and Hartl, D. (1991). Population genetics in forensic DNA typing. *Science*, 254:1745–1750.

Moretti, T., Baumstark, A., Defenbaugh, D., Keys, K., Brown, A., and Budowle, B. (2001a). Validation of STR typing by capillary electrophoresis. *Journal of Forensic Sciences*, 46(3):661–676.

Moretti, T., Baumstark, A., Defenbaugh, D., Keys, K., Smerick, J., and Budowle, B. (2001b). Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. *Journal of Forensic Sciences*, 46(3):647–660.

Mullis, K. (1990). The unusual origin of the polymerase chain reaction. *Scientific American*, pages 56–65.

Mullis, K. and Faloona, F. (1987). Specific synthesis of DNA in vitro via a polymerasecatalyzed chain reaction. *Methods in Enzymology*, 155:335–350.

National Center for Biotechnology Information (2007). BatchExtract download page. *ftp://ftp.ncbi.nlm.nih.gov/pub/forensics/*.

National Criminal Justice Reference Service (2007). Forensic science resources – summary. *http://www.ncjrs.org/forensic/summary.html*.

National Institute of Standards and Technology (2007). STRBase: Short tandem repeat internet database. *http://www.cstl.nist.gov/biotech/strbase/*.

National Research Council (1992). *DNA Technology in Forensic Science*. National Academy Press, Washington DC.

National Research Council (1996). *The Evaluation of Forensic DNA Evidence*. National Academy Press, Washington DC.

Oorschot, R. A. V. and Jones, M. K. (1997). DNA fingerprints from fingerprints. *Nature*, 387.

Paoletti, D., Doom, T., Krane, C., Raymer, M. L., and Krane, D. E. (2005). An empirical analysis of the STR profiles resulting from conceptual three-contributor mixtures. *Journal of Forensic Sciences*, 50(6):1361–1366.

Paoletti, D., Doom, T., Raymer, M., and Krane, D. (2006). Assesing the implications for close relatives in the event of similar but non-matching DNA profiles. *Jurimetrics*, 46(2):161–175.

Perlin, M. (1999). US Patent No. 5876933. *Washington, DC: U.S. Patent and Trademark Office*.

Perlin, M. and Szabady, B. (2001). Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. *Journal of Forensic Sciences*, 46(6):1372–1378.

Poinar, G. (1994). The range of life in amber: significance and implications in DNA studies. *Experientia*, 50(6):536–542.

Promega (2007). Promega home page. *http://www.promega.com*.

Ross, K., Haites, N., and Kelly, K. (1990). Repeated freezing and thawing of peripheral blood and DNA in suspension: effects on DNA yield and integrity. *Journal of Medical Genetics*, 27(9):569–570.

Rubinson, K. and Rubinson, J. (2000). *Contemporary Instrumental Analysis*. Prentice Hall, Upper Saddle River.

Rudin, N. and Inman, K. (2002). *An Introduction to Forensic DNA Analysis*. New York, NY, 2nd edition.

Saiki, R., Gelfand, D., Stoffel, S., Scharf, S., Higuchi, R., Horn, G., Mullis, K., and Erlich, H. (1988). Primer-directed enzymatic amplification of DNA with thermostable DNA polymerase. *Science*, 239:487–491.

Scientific Working Group on DNA Analysis Methods (2000). Short Tandem Repeat (STR) interpretation guidelines. *Forensic Science Communications*, 2(3).

Stacey, R. (2005). Report on the erroneous fingerprint individualization in the Madrid train bombing case. *Forensic Science Communications*, 7(1).

Taylor, M. (2001). TA case #1458, Commonwealth v. Dirk K. Geineder. DNA Analysis Report.

The National Health Museum (2007). Access Excellence: Structure of DNA. *http://-www.accessexcellence.org/AB/GG/structure.html*.

Thompson, W., Ford, S., Doom, T., Raymer, M., and Krane, D. (2003a). Evaluating forensic DNA evidence: Essential elements of a competent defense review part 1. *The Champion*, 27(3).

Thompson, W., Ford, S., Doom, T., Raymer, M., and Krane, D. (2003b). Evaluating forensic DNA evidence: Essential elements of a competent defense review part 2. *The Champion*, 27(4).

Thomsen, V., Schatzlein, D., and Mercuro, D. (2003). Limits of detection in spectroscopy. *Spectroscopy*, 18(12):112–114.

Troyer, K., Gilboy, T., and Koeneman, B. (2001). A nine locus match between two apparently unrelated individuals using AmpFlSTR® Profiler Plus™ and COfiler™. In *The 12th International Symposium on Human Identification*.

Voet, D., Voet, J. G., and Pratt, C. W. (1999). *Fundamentals of Biochemistry*. Wiley, New York, NY.

von Wurmb-Schwark, N., Harbeck, M., Wiesbrock, U., Schroeder, I., Ritz-Timme, S., and Oehmichen, M. (2003). Extraction and amplification of nuclear and mitochondrial DNA from ancient and artificially aged bones. *Legal Medicine (Tokyo)*, 5(1):S169–S172.

Wallin, J., Buoncristiani, M., Lazaruk, K., Fildes, N., Holt, C., and Walsh, P. (1999). TWGDAM validation of the AmpFlSTR™ Blue PCR Amplification Kit for forensic casework analysis. *Journal of Forensic Sciences*, 43(4):854–870.

Walsh, P., Erlich, H., and Higuchi, R. (1992). Preferential PCR amplification of alleles: mechanisms and solutions. *PCR Methods and Applications*, 1:241–250.

Wambaugh, J. (1991). *The Blooding*. William Morrow.

Wang, T., Xue, N., Radar, M., and Birdwell, J. (2001). Mixture STR/DNA deconvolution using allele peak area data and the least square technique. In *The 12th International Symposium on Human Identification*.

Weir, B., Triggs, Starling, L., Stowell, L., Walsh, K., and Buckleton, J. (1997). Interpreting DNA mixtures. *Journal of Forensic Sciences*, 42(2):213–222.

Williams, R. and Johnson, P. (2005). Inclusiveness, effectiveness and intrusiveness: issues in the developing uses of DNA profiling in support of criminal investigations. *Journal of Law, Medicine and Ethics*, 33(3):545–558.

Willing, R. (2005). Suspects get snared by a relative's DNA. *USA Today*.