This is to certify that the
thesis entitled

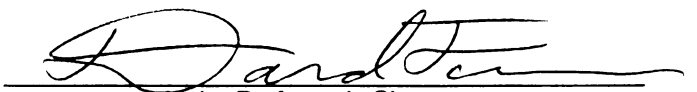
THE CHARACTERIZATION OF VARIANT ALLELES AT THE 13 CODIS STR LOCI FOR USE IN PATERNITY DISPUTE RESOLUTIONS

presented by

Catherine Therese Allor

has been accepted towards fulfillment
of the requirements for the

Master of Science	degree in	Criminal Justice with Specialization in Forensic Science



Major Professor's Signature

12/6/04

Date

THE CHARACTERIZATION OF VARIANT ALLELES AT THE 13 CODIS STR LOCI FOR USE IN PATERNITY DISPUTE RESOLUTIONS

By

Catherine Therese Allor

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

School of Criminal Justice

2004

ABSTRACT

THE CHARACTERIZATION OF VARIANT ALLELES AT THE 13 CODIS STR LOCI FOR USE IN PATERNITY DISPUTE RESOLUTIONS

By

Catherine Therese Allor

This study investigated rare, genetic anomalies known as variant alleles, and their role in resolving paternity disputes. Paternity analysis typically involves the comparison of DNA profiles from a mother, child and alleged father, and conclusions are drawn based on the genetic evidence. In the past, most variant alleles could not be utilized in these analyses, leading to weaker genetic support for the conclusions. A total of 32,671 DNA profiles were examined for variant alleles. A selection of affected samples were quantified, amplified and subjected to gel electrophoresis to confirm the presence of variant alleles. A total of 85 variant alleles at 12 of the 13 CODIS loci were confirmed in 757 samples. The affected samples were sorted by racial group in order to calculate allele frequencies. Twenty-eight of the variant alleles were observed in 5 or more samples, regardless of racial group, and were added to the Orchid GeneScreen allele frequency database for use in paternity calculations. Paternity analyses were performed on two cases to demonstrate that the use of variant allele data greatly increases the strength of the genetic evidence, providing further support for the paternity conclusions. It is expected that the regular use of these alleles may lead to a reduction in laboratory expenditures, minimize the need for additional testing and decrease the turn-around time for reporting results. In addition, 50 of the variant alleles were listed on the STRBase website as a reference for others in the paternity and forensic science communities.

To my family. Your love and support mean the world to me.

ACKNOWLEDGEMENTS

I would like to thank several people for helping to make this project a success. First, I would like to acknowledge my advisor, Dr. David Foran, for his guidance and assistance in editing this paper. I could not have done this without your help. I would also like to recognize Dr. Marco Scarpetta, Manager of Paternity and Associate Laboratory Director at Orchid GeneScreen as the creative genius and my mentor during this project. Without your direction, I might still be looking for thesis topic ideas. Dr. John Butler of the National Institute of Standards and Technology is owed a great deal of thanks for performing some additional tests on several of my samples. Your contributions were greatly appreciated. The laboratory staff at Orchid GeneScreen is owed a great deal of thanks for assisting with some of the preliminary data collection. Your efforts made a seemingly endless amount of work feel less overwhelming. Lastly, Dr. Jay Siegel, my professor and advisor is owed a tremendous amount of thanks for all of his encouragement and endless support throughout this process. It has been both a pleasure and an honor to have been one of your students, and I am extremely proud knowing that I have learned from the best!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES (cont'd)

LIST OF FIGURES

LIST OF FIGURES (cont'd)

LIST OF ABBREVIATIONS

AABB	- American Association of Blood Banks
AF	- Alleged Father
BP/bp	- Base Pair
C	- Child
CPI	- Cumulative Paternity Index
DNA	- Deoxyribonucleic Acid
M	- Mother
PCR	- Polymerase Chain Reaction
PI	- Paternity Index
RFU	- Relative Fluorescence Unit
STR	- Short Tandem Repeat

INTRODUCTION

DNA analysis is currently at the forefront of human identity testing, utilized by both the forensic and parentage testing communities. Current methods of DNA testing make it possible to distinguish one person from all other humans, living or dead, based on their genetic makeup. While forensic DNA analysis is typically performed for matters of a criminal nature, parentage testing involves the examination of DNA to determine familial relationships. The majority of parentage cases are paternity disputes, which involve the testing of a mother, child and alleged father to determine the likelihood that the tested man is the biological father of the child. Other parentage cases involve the establishment of maternity, where an alleged mother and child are tested, typically without the biological father. Maternity testing is commonly performed for adoption agencies that want to verify the relationship between the mother and the child to be adopted. A third form of parentage identification, sibship testing, is useful in determining whether two or more individuals have the same biological mother or father.

DNA profiles are generated for each individual involved in a parentage testing case. These profiles contain several pieces of genetic information useful in establishing identity. The more genetic information that can be gathered from the individuals involved in a parentage testing case, the stronger the conclusions become. Sometimes, one or more pieces of information are missing from a DNA profile, and the strength of the parentage conclusions diminishes. A lack of data in a DNA profile can be caused by several factors, including genetic variations not recognized by the software that generates the profiles. This project was an investigation into a specific type of rare genetic

anomaly found in certain individuals' DNA that can affect the results of parentage testing.

DNA Analysis - Orchid GeneScreen Paternity Laboratory

Orchid GeneScreen in East Lansing, Michigan is a high throughput parentage-testing laboratory that currently tests approximately 135,000 samples from 45,000 cases annually using DNA analysis. Although the laboratory performs paternity, maternity and sibship testing on a regular basis, approximately 96% of the samples are tested for paternity, and that nomenclature will be used throughout this paper. Nearly 65% of the paternity testing performed by Orchid GeneScreen is court-ordered for the purposes of providing financial support to children, with the remaining 35% performed for private accounts.

A standard paternity case involves 3 samples (one each for the mother, child and alleged father). Laboratory analysis involves the comparison of the child's DNA profile to those of the mother and alleged father. Every individual inherits two copies of their chromosomes, one from each biological parent, and specific chromosomal regions known as loci are examined in paternity analysis. These loci are comprised of short tandem repeats (STRs), 2 to 7 base pair (bp) repeats of DNA that vary in number among individuals. For example, the sequence 'AGAGAG' contains three copies of the 2bp repeat unit, AG. STR loci are often highly polymorphic, and a multiple number of alleles existing at each locus, facilitating the differentiation of individuals.

The STR loci commonly analyzed by Orchid GeneScreen contain repeat sequences comprised of 4bp. The repeat pattern can be simple, compound or complex.

A simple repeat pattern consists of a single STR unit repeated a variable number of times (n), such as $[AGAT]_n$. A compound repeat is comprised of two or more adjacent STR units such as $TCTA[TCTG]_3[TCTA]_n$. Complex repeat patterns contain variable DNA sequences amidst blocks of several different STR units, such as $[TCTA]_n[TCTG]_n\{[TCTA]_3TA[TCTA]_3TCA[TCTA]_2TCCATA\}[TCTA]_nTATCTA$ (Butler, 2001, Urquhart *et al.*, 1994).

The number of repeat units determines the size of each allele, which is measured using standards added to each sample prior to analysis. The nomenclature, or designation for each allele is then determined based on the number of repeats present. For example, a person may be typed as a '7/8', having inherited 7 repeat units from one parent and 8 from the other. Alleles are assigned repeat numbers based on the comparison of their size to the size of previously defined alleles found in an 'allelic ladder'. These ladders contain the most common alleles existing for each locus. The comparisons made between the DNA profiles of the child and alleged father result in a classification of 'inclusion' or 'exclusion' for each locus. A paternal inclusion occurs at a locus when the alleged father and the child share the obligate paternal allele, the allele that must be contributed by the biological father. A paternal exclusion occurs at that locus if the alleged father could not have contributed the obligate paternal allele to the child (Table 1).

Table 1: Examples of Paternity Inclusions and Exclusions

The numbers indicate the allele designation for the mother, child and alleged father for each locus. Heterozygous loci are indicated by two different allele designations, while homozygous loci have two copies of a single allele designation. A paternal inclusion occurs at a locus when the alleged father shares a common allele with the child that the mother did not contribute to the child, the obligate paternal allele. A paternal exclusion occurs at a locus when the alleged father could not have contributed the obligate paternal allele to the child.

Mother	Child	Alleged Father	Conclusion
9, 11	11, 12	12	Inclusion
9, 11	11, 12	13, 14	Exclusion
13	13	13, 15	Inclusion
13	13	14, 15	Exclusion
12	12, 16	15, 16	Inclusion
12	12, 16	14	Exclusion

An inclusion at a given locus can occur by chance, even if the tested man is not the biological father. Therefore, several different loci must be tested in order to reduce the possibility of a random match between unrelated individuals. Orchid GeneScreen regularly tests samples using between 9 – 13 different CODIS STR loci. Exclusions at a minimum of 2 loci are required by the American Association of Blood Banks (AABB) to conclude that the tested man is not the biological father. There are occasions when the biological father may differ from the child at one or two loci due to mutations that can occur in the sperm. At Orchid GeneScreen, single exclusions occur at a per-case ratio of 1:50, with double exclusions expected to occur at a 1:2500 ratio. For this reason, Orchid requires exclusions at three loci before rendering the conclusion that the tested man could not have fathered the child.

The ladder alleles for each locus have established frequencies that represent how often the allele is seen within a given racial group (Caucasian, African American, Hispanic or Other). The allele frequencies are utilized to calculate a paternity index (PI) value for each locus. A PI is a likelihood ratio defined as the probability that an event

will occur under certain conditions divided by the probability that the same event will occur under a different, mutually exclusive set of conditions (Traver, 1998). More specifically, the PI ratio is the likelihood that a non-excluded alleged father from one racial group is the biological father of a child divided by the likelihood that a man randomly selected from the same racial group is the biological father. The PIs for all loci are then multiplied together to calculate the cumulative paternity index (CPI). The CPI is a numerical value that indicates how strongly the genetic data support either the hypothesis that the alleged father is the biological father or the alternative hypothesis that another man is the biological father (Traver, 1998).

Theoretically speaking, CPI values can range from 0 to infinity. CPI values less than 1 indicate that the alleged father may not be the biological father, while a CPI value of 0 signifies that the alleged father cannot be the biological father and has been excluded. When the CPI value is 1 or greater, the genetic data support the hypothesis that the tested man could have fathered the child. As the CPI value increases, the genetic data more strongly support this hypothesis (Traver, 1998).

In addition to the CPI value, the probability of paternity is included with all paternity reports. This value is a measure of the strength of one's belief in the hypothesis that the tested man in the father (Traver, 1998). This probability is not only based on the genetic evidence (e.g. the CPI value), but also on the prior probability of paternity. This prior probability refers to the strength of one's belief that the tested man is the father based solely on non-genetic evidence, such as the mother's assertion that the correct man was tested (Traver, 1998). Orchid GeneScreen uses the commonly utilized prior

probability value of P = 0.5. As the CPI value increases, the probability of paternity increases proportionally.

Variant Alleles

The DNA profiles of the majority of individuals involved in a paternity dispute contain alleles found within the ladders. However, some people have rare variant alleles that do not correspond with those in the ladder. Variant alleles differ from the ladder alleles by one or more bp or one or more STR units. This variation is caused by mutations involving the insertion or deletion of nucleotides within the repeat sequence (Butler, 2001).

Some variants lie beyond the smallest or largest alleles found in the ladder. Crouse *et al.* (1999) proposed that the general nomenclature for these variants include a '>' or '<' sign in reference to the nearest allele. For example, if the largest ladder allele has 30 repeat units, a variant containing more than 30 repeats would be described as a '>30' allele.

Other variants are found amongst the alleles in the ladder, and differ from them by 1, 2 or 3bp. These are termed "microvariants" due to the small size deviation from the ladder alleles. There are several common microvariants, some of which are included in the ladder. The general nomenclature for microvariants contains the ladder allele designation that has the full number of repeats followed by a decimal value indicating the number of additional bp present (Crouse *et al.*, 1999). For example, an allele that is 2bp larger than allele 20 would be designated as 20.2. The highly polymorphic loci, including FGA, D21S11 and D18S51 contain the most microvariant alleles. These loci

contain larger sized compound and complex repeat patterns, with more locations for mutations to occur as compared to loci that have simple repeat patterns (Butler, 2001).

Samples containing variant alleles are revealed during the analysis of DNA profiles. Graphical representations of the alleles present in the DNA sample, known as electropherograms, are generated using software programs after the sample has been tested. These programs label the alleles with the repeat number based on the sizes and designations of the ladder alleles. When the software is unable to recognize an allele at a locus, the allele is labeled "off-ladder" (Figure 1).

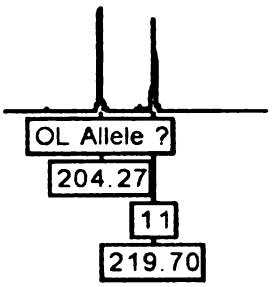


Figure 1: Electropherogram of a Locus Containing an Off-Ladder Allele
Each peak represents a single allele. The off-ladder allele is designated with the "OL Allele?" label. This off-ladder allele has a measured size of 204.27bp. Allele 11 is referred to as the "sister" allele, and has a measured size of 219.70bp. Although the actual bp size of the allele is a whole number, the analysis software estimates the size to 2 decimal points.

Alleles that are not recognized by the analysis software do not have established frequencies and cannot be used to calculate the CPI. As a result, all variant alleles not included in the ladder are of no value in establishing paternity.

Not all off-ladder alleles are caused by the presence of variant alleles. Several factors may lead to the presence of off-ladder alleles including an insufficient volume of sample loaded onto the gel, using excessive amounts of DNA for testing (which can cause extra, off-ladder peaks to be present in a locus), and weak signal intensity of the size standards used to assign allele designations. Therefore, certain steps must be taken

in order to confirm the presence of a variant allele. Butler (2001) proposed that the most common method of confirming a variant is the re-amplification of the sample in question, and the re-analysis of the DNA profile. If the off-ladder allele is present in the DNA profile after re-analysis and no other causative factors can be identified, the allele can be confirmed as variant.

Loss of Data for Paternity Calculations

If an alleged father cannot be excluded as the biological father of a child, the value of the CPI determines how well the genetic data support the conclusion of paternity. The more PI values used to calculate the CPI, the larger the CPI value becomes and the greater the likelihood of correctly identifying the biological father. However, the CPI value decreases when data from any loci are missing, providing weaker support for the hypothesis of paternity. Orchid GeneScreen requires data from a minimum of 6 loci with paternal inclusions and a minimum CPI value before a report of inclusion can be generated. Minimum CPI values range from 100 to 10,000 depending on the particular account to which a paternity case belongs. Cases that do not meet these criteria require the testing of additional loci before a paternity report can be issued. While supplementary testing is very costly in terms of resources, it also delays the reporting of results, ultimately leading to an increase in laboratory turn-around time.

The Investigation

In recent years, the number of samples submitted to Orchid GeneScreen for paternity analysis has steadily increased, and variant alleles have been observed more

frequently. This investigation was designed as a means to learn more about them and how often they occur. Samples containing suspected variants were re-amplified and re-analyzed to confirm their existence. Frequencies were calculated for confirmed variants found in each racial group. The ultimate goal of this investigation was to measure the extent to which the CPI values would increase in cases where the allele shared by the child and the alleged father was variant.

An additional aspect of this investigation involved the submission of confirmed variant alleles to the STRBase Internet website database (http://www.cstl.nist.gov/biotech/strbase/). An important function of this website is to provide information about variant alleles to members of the forensic science, paternity and general biology communities. By submitting these alleles, they are not only beneficial to Orchid GeneScreen, but also to any individual hoping to obtain information about them.

MATERIALS AND METHODS

Reviewing the Orchid GeneScreen DNA Profile Database

A total of 32,671 DNA profiles from 409 different AmpFℓSTR® Profiler Plus™ and COfiler™ (Applied Biosystems) gel runs were reviewed for variant alleles. A maximum of 90 DNA profiles can be found in each gel run. Using ABI Prism® Genotyper® software (Applied Biosystems), the elecropherograms (referred to herein as "original" electropherograms or DNA profiles) from each gel run were reviewed after applying the "Kazaam (20% filter)" macro. This macro re-assigns allele designation labels to all loci so that any off-ladder alleles are clearly marked as "OL Allele?" within the electropherograms (Figure 1 above). This step was necessary as the labels from loci containing off-ladder alleles are manually removed from the electropherograms prior to being saved in the database (Profiler Plus™ User Manual, 1997).

Data from samples containing off-ladder alleles were recorded based on the Orchid GeneScreen criteria for acceptable DNA profiles:

1) All 11 of the size standard peaks (GeneScan-500™ ROX, Applied Biosystems), ranging from 75 – 400bp must be present in each sample and have peak heights of at least 100 relative fluorescence units (RFUs) to yield accurate base pair sizing of the alleles.
2) The sample in question must have allele peak heights ≥150 RFUs. Alleles with peak heights <600 RFUs must be scrutinized for signal leakage from adjacent sample lanes.

3) Samples with high RFU values (>4000) must be checked for pull up, which appears as a peak of similar bp size in the dye color above and/or below the locus that contains the peak(s) with high RFU values (see example in Figure 2). These pull up peaks may be labeled as off-ladder alleles by the software and should not be mistaken for true alleles.

4) Two peaks within a locus are considered a pair, or sister alleles, only if the smaller peak's RFU value is at least 50% of the larger peak's RFU value.

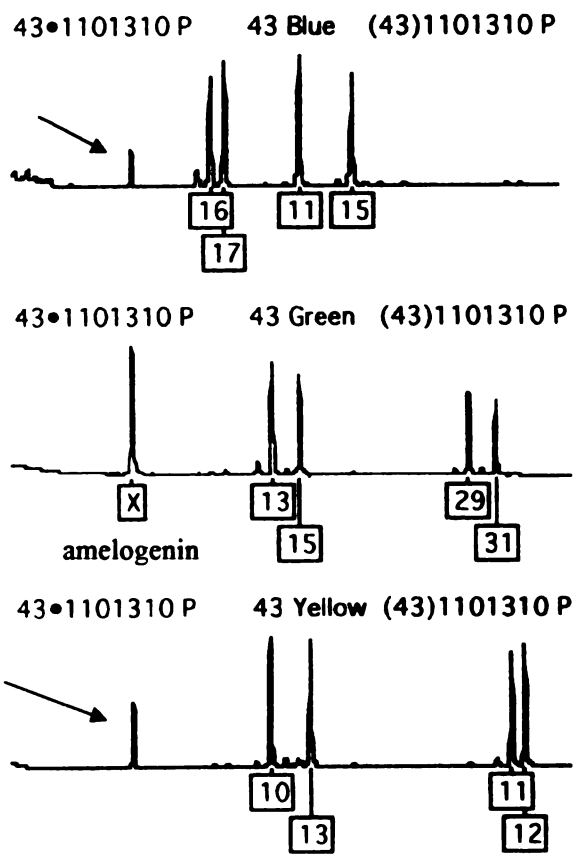


Figure 2: Electropherogram showing Pull-Up Peaks caused by Amelogenin Marker
The arrows indicate the pull-up peaks at the D3S1358 locus in the blue dye region (top panel) and D5S818 locus in the yellow dye region (bottom panel). These are caused by the amelogenin marker, the leftmost marker in the green dye region (center panel). Pull-up peaks are sometimes labeled as off-ladder alleles and may falsely resemble true allele. Peaks to the right of amelogenin are from other markers in the DNA profile.

Samples that met these criteria were classified as having "suspected" or "potential" variant alleles and several pieces of information were recorded for each (Table 2).

Table 2: Information Recorded for Samples with Suspected Variant Alleles
The amplification kit used, gel run number, sample barcode number and lane assignment were essential in order to locate the original electropherograms in the database. The sample barcode number also indicated where the sample could be located in the storage unit. The number of samples tested per gel was used to determine the total number of DNA profiles reviewed. The locus containing the suspected variant, the sizes of the suspected variant and the sister allele, as well as the sister allele designation were useful in approximating the suspected variant allele designation.

Identification Categories
Amplification Kit (Profiler Plus™ or COfiler™)
Gel Run #
Sample Barcode #
Lane Assignment #
Samples Tested on Gel
Locus Containing Suspected Variant Allele
bp Size of Suspected Variant Allele
Sister Allele Designation
Sister Allele's bp Size

Approximation of Allele Designations and Sample Selection

Samples containing potential variant alleles were sorted and grouped together per affected locus and size. Allele designations for those found within the ladder were approximated by comparing their sizes with the mean size of the ladder alleles listed in the Profiler Plus™ or COfiler™ User Manuals (1997). Designations for suspected variant alleles found outside of the ladder were extrapolated based on their sizes as compared to those of the smallest or largest ladder alleles.

In general, one representative sample from each variant allele grouping was re-amplified and re-analyzed. Two or more samples were re-analyzed for variant allele

groupings with a broader range of bp sizes to determine if the grouping contained only one allele, or two. The samples selected contained suspected variants with sizes near either end of the range. For example, the observed size range for the 44.2 microvariant in FGA was 326.40bp – 326.79bp. The two samples tested contained off-ladder alleles with sizes of 326.44bp and 326.79bp.

DNA Quantitation

DNA quantitation was performed on DNA samples that were previously organically extracted using in-house protocols (Orchid GeneScreen SOP Manual, 2001). A single probe Hamilton MICROLAB® 2200 Robot Liquid Handler and an Eclipse v. 4.1 software application (both Hamilton Co.) were used to transfer 80µL of TE (10mM Tris-HCL, 0.1mM Na_2EDTA, pH 8) to the wells of a 96-well microplate. A robot then added 20µL of each sample to the microplate. Twenty microliters of varying concentrations of K562 DNA standard was hand-added to the first 12 wells of the microplate (Table 3).

Table 3: Concentrations and Positions of Quantitation K562 DNA Standards
Each of the K562 DNA standards was added to two wells, and the measured average of each standard concentration was used to determine the DNA concentration of each sample.

Well Assignments	DNA Concentration (ng/µL)
1 and 2	0.0 (TE only)
3 and 4	0.5
5 and 6	2.5
7 and 8	5.0
9 and 10	7.5
11 and 12	10.0

A solution containing 100µL TE and 0.5µL PicoGreen® dsDNA quantitation reagent (Molecular Probes) was made for each DNA sample and standard, and 100µL of this solution was added to each.

The microplate was placed inside an Fmax Microplate Quantifier (Molecular Devices) and the samples were allowed to incubate for 4 minutes. An Fmax Quantifier is a fluorometer with an excitation maximum set at 485nm and an emission maximum set at 538nm. After excitation, a SoftMAX Pro v. 1.3.1 software program (Molecular Devices) records the amount of light emitted by each sample. A SoftMAX Pro program created a standard curve based on the measured concentration of each K562 DNA standard and determined the DNA concentration of each unknown sample accordingly. The concentration data were then exported from a SoftMax Pro program to a Laboratory Information Management System (LIMS) software program v. 1.5 (Blue Sabre Systems). LIMS was then used to calculate the amount of TE required for a DNA concentration of approximately 0.25 – 0.30ng/µL, and the samples were hand diluted accordingly.

DNA Amplification and Polymerase Chain Reaction (PCR)

A Hamilton robot and an Eclipse software amplification program were used to transfer 6µL of each sample into the corresponding wells of a 96-well microplate. In addition, 6µL each of NANOpure® dH_2O or AmpFℓSTR® Control DNA 9947A were also transferred to the microplate. The dH_2O served as a negative control, while the 9947A served as a positive control.

An amplification master mix of DNA primers, deoxynucleotide triphosphates (dNTPs) and DNA polymerase was made for each sample or control (Table 4). Nine microliters of the master mix were added to each for a total PCR volume of 15μL.

Table 4: Volumes of Amplification Kit Components Added Per Sample
DNA primers, reaction mix containing dNTPs and DNA polymerase were combined to form the amplification master mix. The DNA primers varied depending on whether the Profiler Plus™ or COfiler™ amplification kit was used.

Master Mix Component	Volume per sample (μL)
AmpFℓSTR® Profiler Plus™/COfiler™ Primer Set	3.3
AmpFℓSTR® PCR Reaction Mix	6.3
AmpliTaq Gold® DNA Polymerase	0.3

The samples were then placed into a PTC-100™ Programmable Thermal Controller (MJ Research, Inc.) for hot start PCR using the cycling parameters listed in Table 5.

Table 5: Thermal Cycler Temperatures and Times
During PCR, the samples underwent an initial incubation of 11 minutes at 95° prior to 28 cycles of denaturation, annealing and extension. The final extension step lasted 45 minutes at 60° before the final soak at 25°.

PCR Step	Temperature (°C)	Time (min.)
Initial Incubation	95	11
28 cycles: 1) Denature	94	1
2) Anneal	59	1
3) Extend	72	1
Final Extension	60	45
Final Soak	25	Hold

Gel Electrophoresis

At the completion of amplification, 4μL of each sample were transferred from the amplification microplate to the corresponding wells of a 96-well loading microplate.

Four microliters of AmpFℓSTR® Profiler Plus™ or COfiler™ Allelic Ladder were added to an empty well in the loading plate. A loading buffer solution was then made for the samples, negative and positive controls and ladder. This solution contained the ROX sizing standard, formamide loading solution and Hi-Di™ Formamide (Table 6). Each sample or control was combined with 3.5μL of the loading buffer solution, and 4μL of the solution was added to the ladder.

Table 6: Sample Loading Buffer Solution Components and Volumes

A loading buffer solution comprised of ROX, formamide loading solution and Hi-Di™ Formamide was added to each sample, control or ladder prior to electrophoresis.

Loading Buffer Component	Volume Per Sample (μL)
GeneScan-500™ ROX	0.55
Formamide Loading Solution	1.75
Hi-Di™ Formamide	1.75

The loading microplate was placed on a 95°C heat block for 3 minutes to denature the samples, and then immediately placed on ice for 4 minutes.

A Long Ranger® Singel® (BioWhittaker Molecular Applications) pack was mixed and poured in between two 36cm well-to-read plates according to manufacturer's instructions. The gel was allowed to polymerize for a minimum of 2 hours at room temperature prior to loading. Electrophoresis was performed using an ABI Prism® 377 DNA Sequencer and an ABI Prism® 377-96 Collection software v. 2.5 application (Applied Biosystems). The running buffer used was 0.5X TBE (47mM Tris, 47mM H_3BO_3, 1mM Na_2EDTA, pH 8). Half of the DNA samples were loaded into their respective lanes using a 96-lane Sharks Tooth Comb before the control reagents and

ladders were loaded. The loading volumes for the samples, controls and ladder are listed in Table 7.

Table 7: Loading Volumes for Samples, Controls and Ladder

A total sample volume of 0.8μL was loaded in each lane. 1.5μL of allelic ladder was loaded in each of the 4 designated ladder lanes, while 1μL of the negative and positive controls was loaded in the designated control lanes.

DNA Source	Volume Per Lane (μL)
Sample	0.8
Negative/Positive Controls	1.0
Allelic Ladder	1.5

Electrophoresis was performed for 1 minute prior to loading the remaining DNA samples into their respective lanes. This was done to slightly stagger the DNA banding patterns of the samples, which makes visualization of the lane assignments easier during post-electrophoresis analysis. The gel run continued for 2.5 hours using the settings listed in Table 8. The remaining samples were tested with two additional Profiler Plus™ gel runs and one COfiler™ run.

Table 8: Electrophoresis Parameters for the ABI Prism® 377 DNA Sequencer

These are the optimal electrophoresis settings described in the Profiler Plus™ and COfiler™ User Manuals (1997).

Parameter	Setting
Electrophoresis Voltage (kV)	3.0
Electrophoresis Current (mA)	60.0
Electrophoresis Power (W)	200.0
Laser Power (mW)	40.0
Gel Temperature (°C)	51.0

Data Analysis

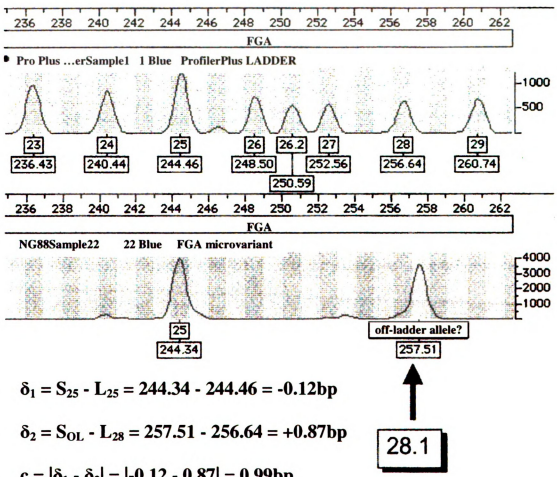
At the completion of electrophoresis, the resultant data were first analyzed using a GeneScan® software program. Following the guidelines in the Profiler Plus™ User Manual (1997), the "Auto-Track" function was used to place a tracking line through each sample's DNA banding pattern. The "Extract" function was utilized to determine the allele sizes before the data were imported into a Genotyper® software application. The "Check GS500" macro was used to analyze the ROX peaks in the samples amplified using the Profiler Plus™ kit, while the "Check GS350" macro was used for samples amplified using the COfiler™ kit. Allele designations were assigned to the peaks in each locus using the "Kazaam (20% Filter)" macro.

The sample DNA profiles were reviewed for the presence of off-ladder alleles at the affected loci; any found were classified as "confirmed" variant alleles. The barcode numbers of samples not containing off-ladder alleles were recorded in order to review their original electropherograms. This was done to determine reasons why these alleles were off-ladder originally, but were recognized as one of the ladder alleles after re-analysis. Both peaks in an affected locus were manually labeled with their bp sizes.

Allele designations were calculated for each variant found within the size range of the ladder to confirm whether or not the approximated designations were correct. Gill *et al.* (2001) described three calculations used to compare the relative size difference (δ) between the sample alleles (S) and the ladder alleles (L) run under the same electrophoretic conditions:

$$\delta_1 = S_Y - L_Y \qquad \delta_2 = S_{OL} - L_X \qquad c = |\delta_1 - \delta_2|$$

δ_1 represents the size difference between non-variant sister allele Y, (S_Y) and ladder allele Y (L_Y). δ_2 signifies the size difference between the off-ladder variant allele (S_{OL}) and ladder allele X (L_X), the smaller allele adjacent to the variant. The c value is the relative size shift between alleles in the sample and the ladder alleles, and indicates how many additional bp are present in each variant (Figure 3).



From Butler (2001)

Figure 3: Electropherogram showing Microvariant Allele Calculations at the FGA Locus
The sample in the bottom panel is compared to the ladder shown in the top panel. Peaks are labeled with the allele designations and the bp size. The δ_1 calculation shows the size difference between sample allele 25 and ladder allele 25. The δ_2 calculation shows the size difference between the off-ladder sample allele and ladder allele 28. The c calculation shows the absolute value of the difference between δ_1 and δ_2, and this indicates how many additional bp are present in the off-ladder allele.

The c value in Figure 3 is 0.99bp, indicating that the relative peak shift between the off-ladder allele and allele 25 in the sample is approximately 1bp. The off-ladder allele in this sample was therefore designated as a 28.1 microvariant.

Butler (2001) stated that allele designations must be extrapolated for variant alleles found outside the expected ladder range, as the above calculations are intended only for variants within the ladder. Therefore, allele designations for the beyond ladder variants were extrapolated using their sizes as compared to the smallest or largest ladder alleles.

Some suspected variant allele designations were adjusted after the calculations and extrapolations were performed, as several errors were made when the designations were initially assigned. These errors occurred because the designations were not calculated using the formulas from Figure 3, but were instead approximated by the author by comparing the sizes from the original run with the mean size of the ladder alleles listed in the Profiler Plus™ or COfiler™ User Manuals (1997). As a result, some samples were grouped with allele designations that were inaccurate. The adjustments that were made involved combining samples from two groups of variant alleles (e.g. combining 21.3 samples with the 21.2 samples), dividing one group into two groups (e.g. dividing the 16.2 samples into 16.1 and 16.2 samples), and assigning a different variant allele designation to a group (e.g. changing the 24.3 samples into 24.2 samples). Additional details on these adjustments can be found in the discussion.

Off-site SGM Plus™ Analysis

Sixteen samples containing above and below ladder variant alleles at the FGA locus were sent to Dr. John Butler at the National Institute of Standards and Technology (NIST) laboratory for SGM Plus™ analysis. This SGM kit contains additional alleles in the FGA ladder not included in the Profiler Plus™ kit. The goal of this testing was to see if they were in the kit and to demonstrate whether or not the variant allele designations were accurately extrapolated. Only samples containing variant alleles that were within 4bp of the SGM ladder alleles were submitted for testing to avoid further extrapolations of the designations. The samples were analyzed via capillary electrophoresis using an ABI Prism® 3100 Genetic Analyzer (Applied Biosystems) according to NIST's standard operating procedures.

Classification of Confirmed Variant Alleles

The confirmed variant alleles were sorted into 5 different categories based on their sizes relative to the smaller, adjacent allele (X) with a full number of repeats. For microvariant alleles below and above the ladder, X was not an allele included in the ladder, but represented the smaller, adjacent, extrapolated allele. The first category, X.1 microvariants, included alleles within the ladder that were 1bp larger than the smaller, adjacent allele. X.2 microvariants included alleles that were 2bp larger than the smaller allele. The X.3 microvariants were 3bp larger than the smaller allele. Categories 4 and 5 were comprised of all variant alleles (including relevant X.1, X.2 and X.3 microvariants) that were smaller or larger than the ladder alleles.

Submission to the STRBase Website

The confirmed variant alleles were compared to those already listed on the STRBase website to determine which could be submitted as previously unreported alleles. It was relevant to note whether or not the alleles already listed on the STRBase website had been tested using an ABI 377 because the way allele sizes are measured can vary slightly depending on the instrument platform used for electrophoresis. These variations are caused by differences in the type and concentration of the gel used and the electrophoresis conditions (Profiler Plus™ User Manual, 1997). For example, the STRBase listing for the 33.1 microvariant in D21S11 indicates a size of 221.35bp when electrophoresed using an ABI 310. In contrast, one of the 33.1 microvariants re-analyzed for this project had a size of 224.73bp, appearing to be more than 3bp larger than the 33.1 microvariant tested on the ABI 310. The alleles were sorted into 3 categories:

1) Variant alleles not already listed on the STRBase website AND not included in any of the allelic ladders used by Orchid GeneScreen.
2) Variant alleles already listed on the STRBase website that were electrophoresed using an instrument other than an ABI 377, AND were not included in any of the allelic ladders used by Orchid GeneScreen.
3) Variant alleles already listed on the STRBase website that were electrophoresed using an ABI 377, OR were included in any of the allelic ladders used by Orchid GeneScreen.

All variant alleles classified under categories 1 and 2 were submitted to the STRBase website.

Racial Data and Variant Allele Frequencies

Orchid GeneScreen collects self-described racial data from both the mother (M) and alleged father (AF) prior to testing, while the child's (C) race can only be assigned after paternity is established. This data was necessary as allele frequencies vary among racial groups, and the categorization of the alleles by group influences the calculated frequency values. Using a LIMS software application, each sample containing a confirmed variant allele was assigned to one of 4 racial groups: African American, Caucasian, Hispanic and Other. The "Other" category included anyone who could not be grouped with the other three or did not specify their race on the chain of custody form that accompanied the samples to the lab. Racial data for individuals with confirmed variant alleles were recorded using a LIMS software application. LIMS was then used to sort the 32,671 DNA profiles reviewed for this project to yield the following numbers of individuals (N):

- Caucasian N = 16,257
- African-American N = 14,160
- Hispanic N = 1,895
- Other N = 359

Racial data for all children were designated using the 9 different classifications as follows:

1. If M and AF were the same race, then AF race was used for C, regardless of AF inclusion or exclusion in paternity.
2. If M and AF were different races and AF was included in paternity, then AF race was used for C.
3. If M and AF were different races and AF was excluded in paternity, then M race was used for C.
4. If M was not tested and AF race was known, then AF race was used for C regardless of AF inclusion or exclusion in paternity.
5. If M race was known and AF was not tested, then M race was used for C.
6. If M race was known, AF race was Other and AF was excluded in paternity, then M race was used for C.
7. If M race was known, AF race was Other and AF was included in paternity, then C race was Other.
8. If M race was Other, AF race was known and AF was excluded in paternity, then C race was Other.
9. If both M and AF races were Other, then C race was Other.

For each racial group, observed variant allele frequencies were calculated by counting the number of occurrences and dividing by 2N, the approximated total number of alleles (n) found in the above samples.

Variant Alleles and the Orchid GeneScreen Allele Frequency Database

The variant allele occurrences and frequencies were added to the Orchid GeneScreen allele frequency database prior to their use in CPI calculations. The laboratory directors decided that confirmed variants found in 5 or more samples, regardless of race, would be included in the database. In addition, only those alleles within 12bp of the smallest and largest ladder alleles were chosen in order to minimize the use of those with extrapolated designations. These extrapolated designations may or may not be completely accurate as they were simply estimated based on the sizes of the variant alleles and the directors wanted to avoid adding potentially erroneous alleles to the database. The directors also decided that the number of variant allele observations, and not the complete DNA profiles from affected individuals would be added to the database. Thus, the addition of the alleles did not increase the number of individuals (N') in the Orchid database, but rather increased the total allele occurrences (n') per locus.

The number of individuals reviewed in this investigation was roughly twice the number of individuals in the Orchid database. The numbers of allele observations were therefore approximately halved prior to adding to the Orchid database so that the observed allele frequencies would not change. These adjusted values represented the number of times each would have occurred had they been found in the Orchid database samples. Because the observed allele frequencies were determined by dividing the number of observations by 2N, the modified number of observations was calculated by multiplying each frequency by 2N'. For example, if a variant allele was found 7 times in African Americans with a frequency of 0.00024718, the extrapolated number of observations amongst 7419 African Americans in the database was calculated as 3.67,

rounded off to 4 occurrences. The calculated observations were then added to the n' values for each locus in the database. Frequencies for all alleles, including variants, were re-calculated by dividing the number of occurrences by the new n' values.

One of the goals in calculating CPI values is to use allele frequencies that are scientifically conservative. That is, they must not overestimate the strength of an inference that the tested man is the biological father. According to the National Research Council (1996), allele frequencies can be very inaccurate if the allele is so rare that it is only observed one or a few times in a database. Therefore, conservative frequency values must be used so that the power of the tests is not compromised. These values are used in paternity calculations for alleles found 5 times or less in a racial group. This value was calculated using 5 allele observations and dividing by the total number of alleles included in the database, and was determined for the loci tested in each group using 5/2N'. The original database is included in Appendix 1, and the modified database with the 5/2N' values is found in Appendix 2.

RESULTS

Approximation of Suspected Variant Allele Designations

A total of 32,671 DNA profiles were reviewed for the presence of variant alleles. Of these, 892 samples were found to contain 117 prospective variants at 12 of the 13 CODIS loci. The allele designations for these variants were approximated (see Materials and Methods), and are included in Table 9.

Table 9: 117 Suspected Variant Alleles

These potential variants were found at 12 of the 13 CODIS loci. Allele designations were approximated by comparing their bp sizes to the sizes of the alleles found in the ladders.

Locus	Suspected Variant Allele
D3S1358	9, 9.1, 12.3, 13.1, 13.3, 14.1, 14.3, 15.1, 15.3, 16.2, 17.1, 17.2, 18.2, 20, 20.1
vWA	13.3, 16.1, 18.3
FGA	14.3, 15.3, 16.1, 16.2, 19.3, 20.1, 20.3, 21.2, 21.3, 22.1, 22.2, 22.3, 23.3, 24.1, 24.2, 24.3, 25.2, 25.3, 31.2, 32.2, 33.2, 34.1, 34.2, 41.2, 42.2, 43.1, 43.2, 44.2, 45.2, 45.3, 46.2, 46.3, 47.2, 47.3, 48.3, 49.3
D8S1179	6.2, 7.3, 9.2, 12.3, 14.3
D21S11	24.3, 25.3, 27.1, 29.1, 29.3, 30.3, 31.1, 32.1, 33.1, 34.1, 35.1, 36.2
D18S51	7, 11.2, 12.2, 12.3, 13.3, 14.1, 15.2, 16.2, 17.2, 17.3, 18.1, 18.2, 20.2, 20.3, 21.2, 21.3, 27.1, 28.1
D5S818	11.1, 12.3, 18
D13S317	5, 6, 7.1, 10.3, 13.1
D7S820	5.2, 6.3, 7.3, 8.1, 8.3, 9.1, 9.3, 10.1, 10.3, 11.1, 11.3, 12.1, 13.1
TH01	4, 8.3, 13.3
TPOX	7.3, 14
CSF1PO	10.2, 16

Gel Electrophoresis Data

A total of 134 samples, which encompassed all observed variant allele sizes, were re-tested on 3 Profiler Plus™ gels and 1 COfiler™ gel. Ninety-eight of these samples contained 85 different confirmed variant alleles (explanations for the other 36 samples below). Of these 98 samples:

- 66 contained 1 unique variant each for a total of 66 variants;
- 4 contained 2 unique variants each, for a total of 8 variants;
- 22 contained 11 unique variants, as two samples were tested for each of these variants; and
- 6 contained 5 variants, none of which were unique, as they were included with the other sample groupings mentioned above.

All variants were observed at heterozygous loci. Fifty-two of the alleles fell amongst the ladder alleles, while the remaining 33 were smaller or larger than the ladder alleles. Once a variant was confirmed, all samples with an allele of the same size were grouped with that variant, although they were not re-tested. Therefore, a total of 757 samples contained the 85 confirmed alleles (Table 10).

Table 10: Confirmed Variant Alleles

A total of 85 different variant alleles were confirmed at 12 CODIS loci. The sizes are from the DNA profiles of the re-analyzed samples. Some of the variant alleles are listed more than once, as more than one sample containing the variant allele was tested. Alleles in bold fell outside the allelic ladder and have extrapolated allele designations.

Locus	Variant Allele	Size (bp)
D3S1358	**9**	102.45
	15.1	126.31
	16.2	132.58
	17.1	135.74
	20.1	147.95
vWA	18.3	188.02

Table 10 (cont'd)

Locus	Variant Allele	Size (bp)
FGA	**14.3**	207.79
	15.3	211.88
	16.1	212.91
	16.2	214.02
	19.3	227.49
	20.1	229.62
	20.3	231.61
	20.3	231.71
	21.1	233.70
	22.3	239.93
	22.3	239.89
	23.3	244.05
	24.1	246.21
	24.3	248.31
	25.1	250.34
	25.3	252.19
	31.2	274.64
	31.2	274.60
	32.2	278.49
	33.1	281.74
	34.1	285.70
	34.2	286.32
	41.2	314.37
	42.2	318.48
	43.2	322.56
	43.2	322.59
	44.2	326.65
	44.2	326.63
	45.2	330.90
	45.2	330.73
	45.2	330.82
	46.2	334.89
	46.2	334.88
	47.2	339.05
	47.2	339.05
	47.2	339.30
	48.2	343.45
	49.2	347.53
D8S1179	**7**	123.15
D21S11	24.3	189.99
	25.3	194.09
	27.1	200.19
	29.1	208.37
	29.3	210.55
	29.3	210.54
	30.3	214.60
	31.1	216.59
	31.1	216.60
	32.1	220.70
	33.1	224.73
	33.1	224.83
	33.1	224.79
	34.1	228.89
	35.1	233.01

Table 10 (cont'd)

Locus	Variant Allele	Size (bp)
D21S11 (cont'd)	35.1	232.95
	36.1	237.06
D18S51	**7**	266.53
	11.2	294.09
	12.2	288.06
	13.3	293.18
	15.2	299.92
	16.1	303.01
	16.2	303.85
	17.2	308.06
	17.3	309.05
	18.1	311.10
	20.2	320.17
	21.2	324.37
	21.2	324.37
	27	347.35
	28.1	351.52
D5S818	12.3	158.72
	18	179.32
D13S317	**5**	195.09
	6	199.16
	7.1	204.27
D7S820	**5.2**	257.78
	6.3	262.56
	8.1	268.33
	8.3	270.34
	9.1	272.26
	9.1	272.25
	9.3	274.15
	9.3	274.17
	10.1	276.10
	10.3	278.01
	11.1	280.02
	11.3	281.89
	12.1	283.75
	12.1	283.85
	13.1	287.66
TH01	**4**	165.61
	8.3	184.63
	13.3	204.89
TPOX	7.3	225.67
	14	251.00
CSF1PO	10.2	298.86
	16	320.64

The calculation or extrapolation of the confirmed allele designations (see Materials and Methods) led to minor adjustments of 29 inaccurate designations (Table 9 above). Eighteen of these were combined together as 9 alleles. The suspected 16.2 microvariants at the D18S51 locus became confirmed 16.1 and 16.2 microvariants. The remaining 10 changed by a matter of 1 – 2bp after re-analysis (Table 11).

Table 11: Adjusted Variant Allele Designations

Re-testing of samples led to several adjustments of the suspected allele designations. Two values in the suspected variant allele column indicate designations that were combined as one confirmed variant. Two values in the confirmed variant allele column indicate designations that resulted from the division of a single suspected variant grouping. Single values in both columns indicate the designation changed after being calculated or extrapolated.

Locus	Suspected Variant Allele(s)	Confirmed Variant Allele(s)
D3S1358	9, 9.1	9
	20, 20.1	20.1
FGA	21.2	21.1
	24.2	24.1
	25.2	25.1
	33.2	33.1
	43.1, 43.2	43.2
	45.2, 45.3	45.2
	46.2, 46.3	46.2
	47.2, 47.3	47.2
	48.3	48.2
	49.3	49.2
D8S1179	6.2	7
D21S11	36.2	36.1
D18S51	12.2, 12.3	12.2
	16.2	16.1, 16.2
	18.2	18.1
	20.2, 20.3	20.2
	21.2, 21.3	21.2
	27.1	27

A total of 24 suspected variant alleles were not off-ladder after re-amplification and re-analysis because they were alleles that were then recognized by the Genotyper® software (Table 12). These 24 were found in the 36 samples that were re-tested, and 135 of the samples originally recorded. Although only these 36 samples were re-tested, the

original DNA profiles for all 135 were reviewed to verify that none contained true variant alleles. It was determined that even if all of these samples had been re-tested, none would have contained true variant alleles. Detailed explanations for why these alleles were originally off-ladder are found in the Discussion.

Table 12: Suspected Variant Alleles That Were Not Confirmed

A total of 24 suspected variant alleles were not confirmed after re-amplification and re-analysis. The observed sizes (or range of sizes if >1 tested) are from their original DNA profiles. The common allele column refers to the allele that was called when the samples were re-analyzed. Common allele bp sizes are mean values from the Profiler Plus™ User Manual. The number of samples containing these suspected alleles ranged from 1 to 35, and testing was performed between 2 and 5 times for each suspected allele.
[a] Suspected allele was pull up peak from D5S818 locus (see Discussion).

Locus	Suspected Variant Allele	Observed bp Range/Size	Common Allele	Common Allele Size (bp)	Total Found	Times Re-tested
D3S1358	12.3	117.68 – 117.78	13	118.24	5	3
	13.1	118.37	13	118.24	1	1
	13.3	121.63 – 121.96	14	122.27	35	3
	14.1	122.36 – 122.53	14	122.27	5	1
	14.3	125.45 – 125.95	15	126.34	31	4
	15.3	129.97 – 130.22	16	130.45	13	2
	17.2	136.70	17	134.59	1	1
	18.2	139.97 – 140.46	18	138.75	2	2
vWA	13.3	168.36	14	168.88	1	1
	16.1	177.37	16	176.76	1	1
FGA	21.3	236.32	22	236.48	1	1
	22.1	237.49	22	236.48	1	1
	22.2	238.64	23	240.60	1	2
	24.1	245.79	24	244.75	1	1
D8S1179	7.3	126.65	8	127.06	2	2
	9.2	132.43	9	131.17	1	1
	12.3	147.86	13	148.58	1	1
	14.3[a]	155.78	-	-	1	1
D18S51	14.1	294.48	14	293.90	1	1
	18.1	310.48	18	309.80	1	1
D5S818	11.1	152.43	11	151.80	1	1
D13S317	10.3	219.21	11	219.54	1	1
	13.1	227.87 – 227.99	13	227.63	20	1
D7S820	7.3	266.73 – 266.76	8	267.18	3	3

Among the 135 samples that did not contain variants were 4 that were originally grouped with microvariants that were confirmed using other samples. Initially, a total of 14 samples were grouped with the 23.3 variant at FGA while 8 were grouped with the 9.3 microvariant at D7S820. After re-testing 2 samples containing 23.3 and 3 containing 9.3, 1 of each did not contain off-ladder alleles. All of the original DNA profiles for the 22 samples grouped with these alleles were reviewed to determine if any problems had occurred when the samples were initially tested. The two samples, as well as an additional 2 grouped with the 9.3 microvariant had experienced anomalies during electrophoresis, and were erroneously included in this study. As a result, a total of 13 samples remained grouped with the 23.3 microvariant, while 5 were grouped with the 9.3 microvariant. Further explanation on these anomalies is found in the Discussion.

Off-site SGM Plus™ Analysis

Electropherograms for 11 of the 16 samples submitted for SGM Plus™ analysis were obtained from NIST. Eight of these contained alleles found in the SGM Plus™ FGA ladder. These 8 are considered to be variants for the purposes of this study, as they are not found in the FGA ladder of the Profiler Plus™ kit. The other 3 samples contained microvariants not found in the SGM Plus™ FGA ladder, and their allele designations were determined at NIST (Table 13). All of these allele designations matched those approximated using the Profiler Plus™ data, indicating that they were extrapolated correctly.

Table 13: Confirmed FGA Variants – SGM Plus™ Analysis

A total of 11 samples tested at NIST using the SGM Plus™ kit yielded variant alleles at the FGA locus. These alleles ranged from 31.2 to 48.2 repeats. The bp sizes ranged from 272.88bp and 343.31bp and were measured using an ABI 3100 Genetic Analyzer. [a] Allele is found in the SGM Plus™ FGA ladder.

Variant Allele	Variant Allele Size (bp)
31.2[a]	272.88
32.2[a]	277.02
33.1	280.28
34.1	284.35
34.2	285.12
42.2[a]	319.50
43.2[a]	323.79
44.2[a]	327.92
45.2[a]	331.93
47.2[a]	339.78
48.2[a]	343.31

The remaining 5 samples submitted to NIST did not yield useable data under the PCR conditions used, and results were not received. These 5 included 2 variants that were below allele 18, the smallest in the Profiler Plus™ FGA ladder, and 3 variants that were above allele 30, the largest in the Profiler Plus™ FGA ladder (Table 14). Of these, only the 46.2 microvariant is included in the SGM Plus™ ladder. Without SGM Plus™ verification, it was assumed that the allele designations were extrapolated correctly.

Table 14: Extrapolated FGA Variant Allele Designations

These 5 variant allele designations were extrapolated using the sizes from the Profiler Plus™ gel runs. They ranged from 16.1 repeats to 49.2 repeats, with bp sizes ranging from 212.91 – 347.53. The 46.2 microvariant is included twice, and size data from both runs are included. [a] Allele is found in the SGM Plus™ FGA ladder.

Variant Allele	Size (bp)
16.1	212.91
16.2	214.02
41.2	314.37
46.2[a]	334.88
46.2[a]	334.89
49.2	347.53

Classification of Confirmed Variant Alleles

The 85 variant alleles were sorted into 5 different categories based on their sizes: X.1 microvariants, X.2 microvariants, X.3 microvariants, variant alleles below the ladder, and variant alleles above the ladder (Tables15 – 19). The latter two categories included their own X.1, X.2 and X.3 microvariants. The totals for each category were:

- Twenty-two X.1 microvariants in 182 samples;
- Nine X.2 microvariants in 97 samples;
- Twenty-one X.3 microvariants in 134 samples;
- Twelve below ladder variants in 99 samples; and
- Twenty-one above ladder variants in 245 samples.

Table 15: X.1 Microvariant Alleles

A total of 22 variant alleles in 182 samples were categorized as X.1 microvariants. The number in parentheses in the Locus column indicates the total number of X.1 microvariants observed at that locus. The size range and average size of each microvariant were determined using the allele sizes from both the original DNA profiles and the DNA profiles generated after being re-tested. The number of samples containing each microvariant range from 1 to 76. [a] Allele has a published repeat structure.

Locus (Total Found)	Microvariant	Observed Size Range (bp)	Avg. Size (bp)	# Found	Original Reference
D3S1358	15.1	127.25 – 127.30	127.27	1	-
(2)	17.1	135.69 – 135.85	135.74	3	SGM Plus™
FGA	20.1	229.56 – 229.62	229.59	1	SGM Plus™
(4)	21.1	233.70 – 233.82	233.76	1	-
	24.1	246.02 – 246.21	246.11	5	SGM Plus™
	25.1	250.24 – 250.34	250.29	1	SGM Plus™
D21S11	27.1	200.09 – 200.19	200.15	5	STRBase
(8)	29.1[a]	208.37 – 208.44	208.40	1	SGM Plus™
	31.1	216.59 – 216.61	216.60	2	SGM Plus™
	32.1	220.68 – 220.70	220.69	4	SGM Plus™
	33.1	224.69 – 224.88	224.80	76	Profiler Plus™
	34.1	228.84 – 229.01	228.90	20	Profiler Plus™
	35.1	232.89 – 233.05	232.96	11	STRBase
	36.1	237.01 – 237.06	237.03	1	STRBase
D18S51	16.1	-	303.01	1	-
(2)	18.1	311.01 – 311.10	311.04	3	Profiler Plus™
D7S820	8.1	268.32 – 268.36	268.34	2	STRBase
(6)	9.1	272.14 – 272.46	272.28	11	STRBase
	10.1	275.93 – 276.20	276.07	17	STRBase
	11.1	279.85 – 280.08	279.97	9	STRBase
	12.1	283.74 – 283.85	283.77	3	STRBase
	13.1	287.66 – 287.78	287.73	4	STRBase

Table 16: X.2 Microvariant Alleles

A total of 9 variant alleles in 97 samples were categorized as X.2 microvariants. The number in parentheses in the Locus column indicates the total number of X.2 microvariants observed at that locus. The size range and average size of each microvariant were determined using the allele sizes from both the original DNA profiles and the DNA profiles generated after re-analysis. The number of samples containing each microvariant range from 1 to 41. [a] Allele has a published repeat structure.

Locus (Total Found)	Microvariant	Observed Size Range (bp)	Avg. Size (bp)	# Found	Original Reference
D3S1358 (1)	16.2	132.58 – 132.74	132.67	4	Budowle *et al.* (1997)
D18S51	11.2	284.09 – 284.24	284.17	5	STRBase
(7)	12.2	287.94 – 288.28	288.11	8	SGM Plus™
	15.2[a]	299.75 – 299.92	299.85	41	Barber and Parkin (1996)
	16.2	303.85 – 303.94	303.89	1	SGM Plus™
	17.2[a]	307.86 – 308.06	307.96	2	Gill *et al.* (1996)
	20.2	320.05 – 320.30	320.16	11	SGM Plus™
	21.2	324.03 – 324.44	324.23	24	SGM Plus™
CSF1PO (1)	10.2	-	298.86	1	STRBase

Table 17: X.3 Microvariant Alleles

A total of 21 variant alleles in 134 samples were categorized as X.3 microvariants. The number in parentheses in the Locus column indicates the total number of X.3 microvariants observed at that locus. The size range and average size of each microvariant were determined using the allele sizes from both the original DNA profiles and the DNA profiles generated after re-analysis. The number of samples containing each microvariant range from 1 to 29. [a] Allele has a published repeat structure.

Locus (Total Found)	Microvariant	Observed Size Range (bp)	Avg. Size (bp)	# Found	Original Reference
vWA (1)	18.3	-	188.02	1	SGM Plus™
FGA (6)	19.3	227.47 – 227.53	227.50	2	SGM Plus™
	20.3	231.59 – 231.71	231.65	3	SGM Plus™
	22.3	239.81 – 240.12	239.93	8	Gill *et al.* (1996)
	23.3	243.91 – 244.27	244.09	13	SGM Plus™
	24.3	248.17 – 248.35	248.24	7	SGM Plus™
	25.3	252.11 – 252.57	252.27	7	SGM Plus™
D21S11 (4)	24.3	189.93 – 190.07	190.02	29	Profiler Plus™
	25.3	194.05 – 194.11	194.08	2	Profiler Plus™
	29.3	210.43 – 210.62	210.54	15	Profiler Plus™
	30.3	214.60 – 214.68	214.65	4	Profiler Plus™
D18S51 (2)	13.3	293.03 – 293.18	293.11	4	-
	17.3	308.97 – 309.05	309.00	2	SGM Plus™
D5S818 (1)	12.3	158.72 – 158.76	158.74	2	-
D7S820 (5)	6.3	262.44 – 262.56	262.51	7	Profiler Plus™
	8.3	270.28 – 270.34	270.31	1	STRBase
	9.3	274.10 – 274.20	274.14	5	STRBase
	10.3	277.94 – 278.15	278.03	16	STRBase
	11.3	281.89 – 281.98	281.92	2	STRBase
TH01 (1)	8.3[a]	184.59 – 184.72	184.64	3	Brinkmann *et al.* (1996)
TPOX (1)	7.3	225.65 – 225.67	225.66	1	-

Table 18: Variant Alleles Below the Ladder

A total of 12 variant alleles in 99 samples were smaller than alleles in the ladder. The number in parentheses in the Locus column indicates the total number of these variants observed at that locus. The size range and average size of each variant were determined using the allele sizes from both the original DNA profiles and the DNA profiles generated after re-analysis. The number of samples containing each microvariant range from 1 to 50. [a] Allele has a published repeat structure. [b] Allele is found in SGM Plus™ ladder.

Locus (Total Found)	Variant Allele	Observed Size Range (bp)	Avg. Size (bp)	# Found	Original Reference
D3S1358 (1)	9	102.36 – 102.60	102.46	50	Profiler Plus™
FGA (4)	14.3	207.79 – 207.87	207.83	2	-
	15.3	211.87 – 211.92	211.89	3	-
	16.1[a]	212.91 – 213.10	213.02	21	Griffiths *et al.* (1998)
	16.2[a]	213.96 – 214.20	214.09	8	Profiler Plus™
D8S1179 (1)	7[a]	122.98 – 123.15	123.05	2	Griffiths *et al.* (1998)
D18S51 (1)	7[b]	266.45 – 266.53	266.49	2	SGM Plus™ (Ladder)
D13S317 (3)	5	194.99 – 195.09	195.05	2	Profiler Plus™
	6	199.10 – 199.16	199.12	2	STRBase
	7.1	204.27 – 204.34	204.32	3	STRBase
D7S820 (1)	5.2	257.77 – 257.78	257.77	1	-
TH01 (1)	4[a, b]	165.61 – 165.66	165.63	3	Griffiths *et al.* (1998)

Table 19: Variant Alleles Above the Ladder

A total of 21 variant alleles in 245 samples were larger than the alleles in the ladder. The number in parentheses in the Locus column indicates the total number of above ladder variants observed at that locus. The size range and average size of each variant were determined using the allele sizes from both the original DNA profiles and the DNA profiles generated after re-analysis. The number of samples containing each microvariant range from 1 to 41. [a] Allele has a published repeat structure. [b] Allele is found in SGM Plus™ ladder.

Locus (Total Found)	Variant Allele	Observed Size Range (bp)	Avg. Size (bp)	# Found	Original Reference
D3S1358 (1)	20.1	147.86 – 148.09	147.92	40	STRBase
FGA (14)	31.2[a, b]	274.35 – 274.71	274.55	41	Griffiths *et al.* (1998)
	32.2[a, b]	278.35 – 278.59	278.46	11	Griffiths *et al.* (1998)
	33.1	281.70 – 281.79	281.74	2	-
	34.1	285.64 – 285.70	285.67	2	-
	34.2[a]	286.32 – 286.42	286.36	2	Barber *et al.* (1996)
	41.2	314.31 – 314.41	314.35	5	-
	42.2[a, b]	318.28 – 318.60	318.37	21	Griffiths *et al.* (1998)
	43.2[a, b]	322.32 – 322.65	322.47	31	Griffiths *et al.* (1998)
	44.2[a, b]	326.40 – 326.79	326.55	28	Griffiths *et al.* (1998)
	45.2[a, b]	330.53 – 330.94	330.77	22	Griffiths *et al.* (1998)
	46.2[a, b]	334.64 – 335.09	334.86	18	Barber *et al.* (1996)
	47.2[a, b]	338.85 – 339.49	339.16	8	Griffiths *et al.* (1998)
	48.2[a, b]	343.11 – 343.45	343.29	3	Griffiths *et al.* (1998)
	49.2	347.53 – 347.54	347.53	1	SGM Plus™
D18S51 (2)	27[a, b]	347.33 – 347.35	347.34	3	Barber and Parkin (1996)
	28.1	351.48 – 351.52	351.50	1	-
D5S818 (1)	18	179.32 – 179.39	179.35	2	-
TH01 (1)	13.3[a, b]	204.89 – 204.93	204.92	2	Gene *et al.* (1996)
TPOX (1)	14[a]	250.84 – 251.00	250.92	1	Huang *et al.* (1995)
CSF1PO (1)	16[a]	320.56 – 320.64	320.60	1	Margolis-Nunno *et al.* (2001)

Submission to STRBase Website

The 85 confirmed variant alleles were sorted into three categories before being added to the STRBase database. Twenty-four variant alleles at 8 loci were grouped under category 1, which included those not already listed on the STRBase website and not included in any of the allelic ladders used by Orchid GeneScreen. Twenty-six variant alleles at 8 loci were grouped under category 2, which included those already listed on the STRBase website that were electrophoresed using an instrument other than an ABI 377, and that were not included in any of the allelic ladders used by Orchid GeneScreen. The remaining 35 variants at 9 loci

were grouped under category 3, which included those already listed on the STRBase website that were electrophoresed using an ABI 377, or that were included in any of the allelic ladders used by Orchid GeneScreen (Table 20). All 50 of the category 1 and category 2 variant alleles were submitted and are now listed on the STRBase website.

Table 20: STRBase Categorization of Variant Alleles

Twenty-four of the variant alleles were included in category 1, 26 alleles were included in category 2 and 35 alleles were included in category 3.

Locus	Category 1	Category 2	Category 3
D3S1358	15.1	17.1 20.1	9 16.2
vWA	-	-	18.3
FGA	14.3 15.3 16.2 20.1 20.3 21.1 33.1 34.1 34.2 41.2 49.2	19.3 23.3 24.1 25.1	16.1 22.3 24.3 25.3 31.2 32.2 42.2 43.2 44.2 45.2 46.2 47.2 48.2
D8S1179	7	-	-
D21S11	25.3 30.3	24.3 27.1 29.3 32.1 33.1 36.1	29.1 31.1 34.1 35.1
D18S51	13.3 16.1 17.2 17.3 28.1	11.2 18.1 20.2	7 12.2 15.2 16.2 21.2 27
D5S818	12.3 18	-	-
D13S317	-	6 7.1	5
D7S820	5.2	8.3 9.3 10.1 10.3 11.3 12.1	6.3 8.1 9.1 11.1 13.1
TH01	-	8.3	4 13.3
TPOX	7.3	-	14
CSF1PO	-	10.2 16	-

Racial Group Data and Variant Allele Frequencies

The samples containing confirmed variants were sorted by racial group prior to calculating the allele frequencies. African Americans had the highest number of samples containing variants, with a total of 511. This comprised 67.5% of the 757 samples containing variant alleles and 1.6% of the 32,671 profiles reviewed. One hundred forty-eight Caucasian samples contained variant alleles, representing 19.6% of the 757 containing variants and 0.45% of the total profiles reviewed. The 63 Hispanic samples comprised 8.3% of the 757 containing variants and 0.19% of the total profiles reviewed. Thirty-four samples from the Other racial category contained variants, representing 4.5% of the 757 containing variants and 0.10% of the total profiles reviewed. For each group, observed variant allele frequencies were calculated and are listed with the number of occurrences in Table 21.

Table 21: Variant Allele Observations and Frequencies

The total number of variant allele observations per locus is in bold face. The number of observations per racial group is included with the corresponding observed allele frequency in parentheses (given as a percentage). Frequencies were calculated based on the number of observations in each group and the total number of allele occurrences at each locus (approximated as 2N). The totals for all variant allele observations are included at the bottom of each racial group column.

Locus	Variant Allele	Total Found	Caucasian N = 16,257	African American N = 14,160	Hispanic N = 1895	Other N = 359
D3S1358	9	**50**	2 (0.006151)	46 (0.162429)	-	2 (0.278552)
	15.1	**1**	-	1 (0.003531)	-	-
	16.2	**4**	-	4 (0.014124)	-	-
	17.1	**3**	3 (0.009227)	-	-	-
	20.1	**40**	28 (0.086117)	4 (0.014124)	5 (0.131926)	3 (0.417827)
vWA	18.3	**1**	1 (0.003076)	-	-	-
FGA	14.3	**2**	2 (0.006151)	-	-	-
	15.3	**3**	2 (0.006151)	-	-	1 (0.139276)
	16.1	**21**	-	21 (0.074153)	-	-
	16.2	**8**	-	8 (0.028249)	-	-
	19.3	**2**	2 (0.006151)	-	-	-
	20.1	**1**	1 (0.003076)	-	-	-
	20.3	**3**	-	2 (0.007062)	1 (0.026385)	-
	21.1	**1**	1 (0.003076)	-	-	-
	22.3	**8**	-	6 (0.021186)	2 (0.052770)	-
	23.3	**13**	-	12 (0.042373)	1 (0.026385)	-
	24.1	**5**	-	5 (0.017655)	-	-
	24.3	**7**	1 (0.003076)	6 (0.021186)	-	-
	25.1	**1**	-	-	1 (0.026385)	-
	25.3	**7**	-	7 (0.024718)	-	-
	31.2	**41**	37 (0.113797)	-	-	4 (0.557103)
	32.2	**11**	1 (0.003076)	10 (0.035311)	-	-
	33.1	**2**	-	2 (0.007062)	-	-
	34.1	**2**	-	2 (0.007062)	-	-
	34.2	**2**	-	2 (0.007062)	-	-
	41.2	**5**	-	3 (0.010593)	1 (0.026385)	1 (0.139276)
	42.2	**21**	-	20 (0.070621)	1 (0.026385)	-
	43.2	**31**	1 (0.003076)	30 (0.105932)	-	-
	44.2	**28**	2 (0.006151)	26 (0.091808)	-	-
	45.2	**22**	-	18 (0.063559)	4 (0.105541)	-
	46.2	**18**	-	15 (0.052966)	2 (0.052770)	1 (0.139276)
	47.2	**8**	1 (0.003076)	7 (0.024718)	-	-
	48.2	**3**	-	3 (0.010593)	-	-
	49.2	**1**	-	1 (0.003531)	-	-
D8S1179	7	**2**	-	2 (0.007062)	-	-
D21S11	24.3	**29**	-	23 (0.081215)	2 (0.052770)	4 (0.557103)
	25.3	**2**	-	2 (0.007062)	-	-
	27.1	**5**	-	5 (0.017655)	-	-
	29.1	**1**	1 (0.003076)	-	-	-
	29.3	**15**	8 (0.024605)	7 (0.024718)	-	-
	30.3	**4**	-	3 (0.010593)	1 (0.026385)	-

Table 21 (cont'd)

Locus	Variant Allele	Total Found	Caucasian N = 16,257	African American N = 14,160	Hispanic N = 1895	Other N = 359
D21S11 (cont'd)	31.1	**2**	-	-	2 (0.052770)	-
	32.1	**4**	2 (0.006151)	2 (0.007062)	-	-
	33.1	**76**	6 (0.018454)	60 (0.211864)	4 (0.105541)	6 (0.835655)
	34.1	**20**	3 (0.009227)	16 (0.056497)	1 (0.026385)	-
	35.1	**11**	-	11 (0.038842)	-	-
	36.1	**1**	-	1 (0.003531)	-	-
D18S51	7	**2**	-	-	1 (0.026385)	1 (0.139276)
	11.2	**5**	-	5 (0.017655)	-	-
	12.2	**8**	-	8 (0.028249)	-	-
	13.3	**4**	-	4 (0.014124)	-	-
	15.2	**41**	2 (0.006151)	36 (0.127119)	3 (0.079156)	-
	16.1	**1**	-	-	-	1 (0.139276)
	16.2	**1**	-	1 (0.003531)	-	-
	17.2	**2**	-	-	-	2 (0.278552)
	17.3	**2**	-	-	2 (0.052770)	-
	18.1	**3**	3 (0.009227)	-	-	-
	20.2	**11**	-	10 (0.035311)	-	1 (0.139276)
	21.2	**24**	1 (0.003076)	22 (0.077684)	-	1 (0.139276)
	27	**3**	1 (0.003076)	-	2 (0.052770)	-
	28.1	**1**	-	-	1 (0.026385)	-
D5S818	12.3	**2**	-	2 (0.007062)	-	-
	18	**2**	-	2 (0.007062)	-	-
D13S317	5	**2**	1 (0.003076)	1 (0.003531)	-	-
	6	**2**	2 (0.006151)	-	-	-
	7.1	**3**	-	-	3 (0.079156)	-
D7S820	5.2	**1**	-	1 (0.003531)	-	-
	6.3	**7**	7 (0.021529)	-	-	-
	8.1	**2**	-	-	2 (0.052770)	-
	8.3	**1**	-	1 (0.003531)	-	-
	9.1	**11**	6 (0.018454)	2 (0.007062)	-	3 (0.417827)
	9.3	**5**	-	3 (0.010593)	1 (0.026385)	1 (0.139276)
	10.1	**17**	3 (0.009227)	12 (0.042373)	2 (0.052770)	-
	10.3	**16**	1 (0.003076)	-	15 (0.395778)	-
	11.1	**9**	8 (0.024605)	1 (0.003531)	-	-
	11.3	**2**	-	2 (0.007062)	-	-
	12.1	**3**	3 (0.009227)	-	-	-
	13.1	**4**	1 (0.003076)	2 (0.007062)	1 (0.026385)	-
TH01	4	**3**	-	2 (0.007062)	1 (0.026385)	-
	8.3	**3**	3 (0.009227)	-	-	-
	13.3	**2**	2 (0.006151)	-	-	-
TPOX	7.3	**1**	-	-	1 (0.026385)	-
	14	**1**	-	1 (0.003531)	-	-
CSF1PO	10.2	**1**	-	-	-	1 (0.139276)
	16	**1**	-	-	-	1 (0.139276)
TOTALS	-	**757**	**148**	**511**	**63**	**34**

Variant Alleles Added to Orchid GeneScreen Allele Frequency Database

A total of 28 variant alleles were found in five or more samples, regardless of racial group, and were selected for inclusion in the Orchid GeneScreen allele frequency database (Appendix A). These were observed in a total of 521 samples, comprising 68.8% of the 757 samples containing variant alleles. Not surprisingly, each of these alleles was found at the Profiler Plus™ loci, as the laboratory uses the COfiler™ kit less frequently (Table 22). Variant alleles found in fewer than 5 samples were considered too rare for inclusion in the database, as it was unlikely these variant alleles would be encountered in DNA profiles on a regular basis.

Table 22: 28 Variant Alleles Incorporated into Orchid GeneScreen Database

These 28 variant alleles were selected for inclusion in the Orchid GeneScreen allele frequency database. The size range of each variant was determined using the allele sizes from both the original DNA profiles and the DNA profiles generated after re-analysis. The number of samples containing each variant allele ranged from 5 to 76.

Locus	Variant Allele	Size Range (bp)	Total Found
D3S1358	9	102.36 – 102.60	50
	20.1	147.86 – 148.09	40
FGA	16.1	212.91 – 213.10	21
	16.2	213.96 – 214.20	8
	22.3	239.81 – 240.12	8
	23.3	243.91 – 244.27	13
	24.1	246.02 – 246.21	5
	24.3	248.17 – 248.35	7
	25.3	252.11 – 252.57	7
	31.2	274.35 – 274.71	41
	32.2	278.35 – 278.59	11
D21S11	24.3	189.93 – 190.07	29
	27.1	200.09 – 200.19	5
	29.3	210.43 – 210.62	15
	33.1	224.69 – 224.88	76
	34.1	228.84 – 229.01	20
	35.1	232.89 – 233.05	11
D18S51	11.2	284.09 – 284.24	5
	12.2	287.94 – 288.28	8
	15.2	299.75 – 299.92	41
	20.2	320.05 – 320.30	11
	21.2	324.03 – 324.44	24
D7S820	6.3	262.44 – 262.56	7
	9.1	272.14 – 272.46	11
	9.3	274.10 – 274.20	5
	10.1	275.93 – 276.20	17
	10.3	277.94 – 278.15	16
	11.1	279.85 – 280.08	9

Seven FGA variants were not added to the database, even though they were observed in 5 or more samples. These seven included 41.2, 42.2, 43.2, 44.2, 45.2, 46.2 and 47.2. All but the 41.2 and 46.2 microvariants were confirmed using the SGM Plus™ kit and the 41.2 microvariant is the only one not found in its ladder. Each of these alleles was more than 44bp larger than allele 30, the largest in the Profiler Plus™ FGA ladder, and only those within 12bp of the highest and lowest ladder alleles were added to the database. This was done because as the variants became larger than allele 30, their observed size ranges expanded, and the chance for extrapolation errors increased. If necessary, future samples containing these 7 FGA variants can be tested using the SGM Plus™ kit, as Orchid GeneScreen now uses it on a regular basis for additional loci.

Due to the ambiguity of the Other racial group, only data from the African American, Caucasian and Hispanic racial groups were added to the database. The variant allele observations were roughly halved, as the number of individuals in the Orchid database was nearly half the number of individuals reviewed for this investigation. These values were rounded to the nearest whole number, while values that were <0.5 were rounded to 1 (Table 23).

The allele frequencies were re-calculated using the adjusted observations and are included in Table 23. The minor disparities that were evident between the original variant allele frequencies (Table 21, above) and the re-calculated frequencies were attributed to rounding error. For instance, microvariant allele 29.3 at the D21S11 locus was observed 7 times in the African Americans reviewed for this project, with an observed allele frequency of 0.00024718. The calculated number of observations for

African Americans in the database was 4, with a re-calculated allele frequency of 0.00026023 and a rounding error of 0.00001305.

Thirty-seven of these variant alleles were observed 5 times or fewer in Caucasians, African Americans and Hispanics, and their frequencies were smaller than the conservative 5/2N' values (Table 23). Therefore, these 5/2N' values, and not the observed frequencies would be used for CPI calculations. The modified Orchid database, which includes the 5/2N' values is found in Appendix B.

Table 23: 28 Adjusted Variant Allele Frequencies and Observations

The total number of recalculated variant allele observations per locus is in bold face. The number of occurrences per racial group is included, with the corresponding allele frequency in parentheses (given as a percentage). N' indicates the number of individuals in the database from each race that had genotypes for a given locus. The n' value indicates the total number of alleles observed at the locus. The 5/2N' value (given as a percentage) indicates the minimum allele frequency that can be used for CPI calculations.

[a] Frequency is smaller than the 5/2N' value.

D3S1358 Variants	**Total Found**	**Caucasian** N' = 7636 n' = 15274 5/2N' = 0.0327	**African American** N' = 7602 n' = 15244 5/2N' = 0.0329	**Hispanic** N' = 690 n' = 1383 5/2N' = 0.362
9	**26**	1 (0.00654707) [a]	25 (0.163999)	0.0
20.1	**17**	13 (0.085112)	2 (0.0131199) [a]	2 (0.144613) [a]
FGA Variants	**Total Found**	**Caucasian** N' = 7674 n' = 15368 5/2N' = 0.0326	**African American** N' = 7419 n' = 14947 5/2N' = 0.0337	**Hispanic** N' = 939 n' = 1879 5/2N' = 0.266
16.1	**11**	0.0	11 (0.0735934) [a]	0.0
16.2	**4**	0.0	4 (0.0267612) [a]	0.0
22.3	**4**	0.0	3 (0.0200709) [a]	1 (0.0532198) [a]
23.3	**7**	0.0	6 (0.0401418) [a]	1 (0.0532198) [a]
24.1	**3**	0.0	3 (0.0200709) [a]	0.0
24.3	**4**	1 (0.00650703) [a]	3 (0.0200709) [a]	0.0
25.3	**4**	0.0	4 (0.0267612) [a]	0.0
31.2	**17**	17 (0.1106195)	0.0	0.0
32.2	**6**	1 (0.00650703) [a]	5 (0.0334515) [a]	0.0
D21S11 Variants	**Total Found**	**Caucasian** N' = 7730 n' = 15467 5/2N'= 0.0323	**African American** N' = 7652 n' = 15371 5/2N'= 0.0327	**Hispanic** N' = 683 n' = 1369 5/2N' = 0.366
24.3	**13**	0.0	12 (0.078069)	1 (0.073046) [a]
27.1	**3**	0.0	3 (0.019517) [a]	0.0
29.3	**8**	4 (0.025862) [a]	4 (0.026023) [a]	0.0
33.1	**36**	3 (0.019396) [a]	32 (0.208184)	1 (0.073046) [a]
34.1	**11**	1 (0.0064654) [a]	9 (0.058552)	1 (0.073046) [a]
35.1	**6**	0.0	6 (0.039035)	0.0
D18S51 Variants	**Total Found**	**Caucasian** N' = 7628 n' = 15258 5/2N' = 0.0328	**African American** N' = 7463 n' = 14974 5/2N' = 0.0335	**Hispanic** N' = 647 n' = 1296 5/2N' = 0.386
11.2	**3**	0.0	3 (0.0200347) [a]	0.0
12.2	**4**	0.0	4 (0.026713) [a]	0.0
15.2	**21**	1 (0.0065539) [a]	19 (0.1268866)	1 (0.0771605) [a]
20.2	**5**	0.0	5 (0.0333912) [a]	0.0
21.2	**13**	1 (0.0065539) [a]	12 (0.0801389)	0.0
D7S820 Variants	**Total Found**	**Caucasian** N' = 7685 n' = 15376 5/2N' = 0.0325	**African American** N' = 7612 n' = 15235 5/2N' = 0.0328	**Hispanic** N' = 939 n' = 1887 5/2N' = 0.266
6.3	**3**	3 (0.019511) [a]	0.0	0.0
9.1	**4**	3 (0.019511) [a]	1 (0.0065638) [a]	0.0
9.3	**3**	0.0	2 (0.013128) [a]	1 (0.0529942) [a]
10.1	**8**	1 (0.0065036) [a]	6 (0.039383)	1 (0.0529942) [a]
10.3	**8**	1 (0.0065036) [a]	0.0	7 (0.3709592)
11.1	**5**	4 (0.026015) [a]	1 (0.0065638) [a]	0.0

DISCUSSION

Overview

The goal of this investigation was to gather information on rare, variant alleles that differ from those found in the allelic ladder by 1bp or more. The use of these alleles in paternity calculations may decrease the need for additional testing, and provide further support that the correct man has been identified as the biological father. Eighty-five different variant alleles at 12 of the 13 CODIS loci were confirmed during the course of this investigation. Twenty-eight of these were observed a minimum of 5 times, and were added to the Orchid GeneScreen allele frequency database. They represent 68.8% of all samples containing a variant, and approximately 1.6% of the 32,671 DNA profiles reviewed for this investigation. This indicates that an estimated 2,160 samples out of the 135,000 tested annually will contain variant alleles that can now be used in CPI calculations.

X.1 Microvariant Alleles

A total of 22 microvariant alleles from 182 samples were categorized as X.1 microvariants, which were 1bp larger than the smaller, adjacent allele. All of these were found within the Profiler Plus™ ladder. These represented 24% of samples containing variant alleles, and 0.6% of all DNA profiles reviewed for this investigation. The X.1 microvariants were found at the D3S1358, FGA, D21S11, D18S51 and D7S820 loci (Table 15, above). Currently, none of the allelic ladders used by Orchid GeneScreen contain X.1 microvariants.

Sequence analysis is necessary to determine the type of mutation present and whether it occurred within the repeat region or in the sequences that flank the repeat. Previous research has identified a mutational mechanism resulting in some X.1 microvariants. According to Egyed *et al.* (2000), X.1 microvariants at the D7S820 locus can be caused by variability in a poly-T region located in the 3' flanking region, 13 bases downstream of the core $[GATA]_n$ repeat. The D7S820 ladder alleles contain 9Ts in this region, while X.1 microvariants have a T insertion (Figure 4).

Allele Designation	5' Flanking Region	Repeat Sequence	3' Flanking Region
9	[redacted] -	$[GATA]_9$ -	[redacted] - $(T)_9ATCT$ - [redacted]
9.1	[redacted] -	$[GATA]_9$ -	[redacted] - $(T)_{10}ATCT$ - [redacted]

From Egyed *et al.* (2000)

Figure 4: Example of an X.1 Microvariant at the D7S820 Locus
Sequence structure of allele 9 and microvariant allele 9.1 at the D7S820 locus. Both the 9 and 9.1 alleles contain the same repeat sequence. The additional bp arises from the insertion of a single T in the 3' flanking region of the 9.1 microvariant. Allele 9 contains 9Ts, while microvariant allele 9.1 contains 10Ts. The dark blocks of different size represent non-repeated DNA sequences.

X.2 Microvariant Alleles

A total of 9 alleles from 97 samples were categorized as X.2 microvariants, which were 2bp larger than the smaller, adjacent allele. All of these were sized amongst the alleles in Profiler Plus™ or COfiler™ ladders. They represented 13% of samples containing variant alleles, and 0.3% of all DNA profiles reviewed for this investigation. Three different loci contained X.2 microvariants within their ladder ranges: D3S1358, D18S51 and CSF1PO (Table 16, above). The ladders for the highly polymorphic FGA, D21S11 and D18S51 loci contain several X.2 microvariants. All 9 of these microvariants have been previously reported. Seven were found at the D18S51 locus, which has a

simple tetranucleotide repeat motif of $[AGAA]_n$. Barber and Parkin (1996) observed X.2 microvariants at this locus with a deletion of 2bp, AG, from the 3'-flanking region.

X.3 Microvariant Alleles

Twenty-one different X.3 microvariants from 133 samples were sized amongst the Profiler Plus™ or COfiler™ ladder alleles. These were 3bp larger than the smaller, adjacent allele. They represented 18% of the samples containing confirmed variants, and 0.4% amongst all DNA profiles reviewed. Eight different loci contained X.3 microvariants: vWA, FGA, D21S11, D18S51, D5S818, D7S820, TH01 and TPOX (Table 17, above). One of the most common X.3 microvariants, 9.3 at the TH01 locus, is found in the TH01 ladder in both the COfiler™ and SGM Plus™ amplification kits. No other ladders used by Orchid GeneScreen contain X.3 microvariants.

Previous research has identified the mechanisms behind the additional nucleotides found in some X.3 microvariants. Similar to the X.1 microvariants at D7S820, the X.3 microvariants at this locus can be caused by variability in a poly-T region 13 bases downstream from the core GATA repeat. According to Egyed *et al.* (2000), the D7S820 ladder alleles contain 9Ts in this region, while X.3 microvariants contain only 8Ts, indicating a deletion of a single T nucleotide. Similarly, several of the X.3 microvariants at TH01 have been observed with a deletion of a single "A" nucleotide. Brinkmann *et al.* (1996) determined that one repeat sequence for the 9.3 microvariant is $[AATG]_5ATG[AATG]_3$, as compared with allele 9, which has a repeat sequence of $[AATG]_9$.

Variant Alleles Below the Ladder

Twelve different variants that sized below the smallest ladder alleles were found in 99 samples, representing 13% of those samples containing confirmed variants, and 0.3% amongst all DNA profiles reviewed. These variants included X.1, X.2 and X.3 microvariants as well as alleles with a full number of repeats, and ranged from 2bp – 13bp below the smallest alleles in the ladder. Seven different loci contained these below ladder alleles: D3S1358, FGA, D8S1179, D18S51, D13S317, D7S820 and TH01 (Table 18, above).

Because the X.1, X.2 and X.3 microvariants are included in this category, multiple mechanisms exist for the sequence variation. For example, all of the below ladder alleles at the FGA locus were microvariants. The repeat structure for alleles at FGA is $[TTTC]_3$ TTTT TTCT $[CTTT]_n$ CTCC $[TTCC]_2$. According to Butler (2001), the X.2 microvariants at FGA, both below and above the ladder can be caused by the deletion of 2bp, CT, from the TTCT sequence adjacent to the core $[CTTT]_n$ repeat. This is evident in the repeat structure of the 16.2 microvariant:

$$[TTTC]_3 \text{ TTTT TT } [CTTT]_9 \text{ CTCC } [TTCC]_2.$$

Allele 4 at the TH01 locus and allele 7 at D8S1179 comprised 2 of the 6 variants that contained complete repeat units with sizes smaller than those of the ladder alleles. The repeat sequence of variant allele 4 is $[AATG]_4$, and is one STR unit smaller than allele 5 in the COfiler™ ladder for TH01 (Griffiths *et al.*, 1998), although it is found in the SGM Plus™ ladder. Similarly, variant allele 7 has a repeat sequence of $[TCTA]_7$, and is one STR unit smaller than allele 8 in the D8S1179 ladder (Griffiths *et al.*, 1998).

Variant Alleles Above the Ladder

Twenty-one different variant alleles with sizes larger than the ladder alleles were found in 245 samples. These comprised the largest category of observed variants, representing 32% of samples containing confirmed variants, and 0.7% of all DNA profiles reviewed. Similar to the below ladder variants, these included X.1, X.2 and X.3 microvariants, as well as alleles with complete STR units, and they ranged in size from 4bp – 78bp larger than the ladder alleles. The D3S1358, FGA, D18S51, D5S818, TH01, TPOX and CSF1PO loci contained these above ladder variants (Table 19, above).

Similar to those found below the FGA ladder, the above ladder X.2 microvariants can be caused by a 2bp deletion from the region adjacent to the core $[CTTT]_n$ repeat sequence of FGA (Butler, 2001). The published repeat sequences for FGA ladder alleles, ranging from 42.2 – 48.2 repeats, also contain additional repeated sequences not included in smaller FGA ladder alleles. For example, the 42.2 allele has a repeat structure of $[TTTC]_4$ TTTT TT $[CTTT]_8$ **$[CTGT]_4$ $[CTTT]_{13}$ $[CTTC]_3$ $[CTTT]_3$** CTCC $[TTCC]_4$ (Griffiths *et al.*, 1998). The sequences in bold face indicate the additional nucleotide repeat units. The cause of these sequences has not been established, although it is possible that several single-nucleotide mutations within the core CTTT sequence may have occurred (e.g. T to G and T to C). Consequently, the larger microvariant alleles at FGA may be the result of multiple mutations, rather than a single insertion or deletion of one or more nucleotides.

The 13.3 microvariant at the TH01 locus also has a unique repeat structure, which also may have arose from multiple mutations. While smaller alleles with complete repeat units have a core structure of $[AATG]_n$, the 13.3 microvariant contains an additional

sequence of AACG amid the core repeat unit [AATG] [AACG] $[AATG]_8$ ATG $[AATG]_3$ (Griffiths *et al.*, 1998). Similar to the larger alleles at FGA, it is possible that a T to C mutation in the AATG sequence is the cause for this additional sequence, indicating that the 13.3 microvariant may contain more than one mutation. It is unclear why this additional sequence is present in this allele, and not in the smaller TH01 alleles, including the 9.3 microvariant. Just like 9.3, the partial repeat in 13.3 arises from a deletion of a single 'A' nucleotide in one of the AATG repeats (Gene *et al.*, 1996, Griffiths *et al.*, 1998). At both FGA and TH01, it is evident that as the alleles become larger in size, more possibilities for sequence variation exist.

An irregularity was observed in the one sample containing the 28.1 microvariant at the D18S51 locus that was not observed in any other variant allele from this investigation. When reviewing the electropherograms generated after re-analysis, the Genotyper® software did not label this allele off-ladder. In fact, this allele did not have a label, and only the sister allele peak was labeled with its designation. The reason for this was that the size of microvariant allele 28.1 was larger than the programmed size range for the D18S51 locus in the Genotyper® software. The sample containing this microvariant was re-tested twice with the same result (Appendix C). It is likely that the size range for alleles at D18S51 is larger than originally believed, and may need to be broadened in order to accurately assign designations to these larger alleles.

Novel Variant Alleles

Fourteen of the variant alleles at 6 different loci were not previously reported by other researchers, and are presumed to be novel alleles. These alleles encompass each of the 5 different variant allele categories. All are considered to be extremely rare, as they were found in only 1 – 5 samples each. Currently, these alleles are being sequenced to confirm their designations and to determine where within the repeat structure the mutation occurred. These 14 alleles are summarized in Table 24, and their electropherograms are included in Appendix C.

Table 24: 14 Novel Variant Alleles

These 14 variants have not been previously reported and are considered to be novel alleles. The number in parentheses indicates the total number of novel variants observed at each locus. The number of samples containing each variant is included to indicate the relative rarity of each.

Locus (Total)	Variant Allele	# Found
D3S1358 (1)	15.1	1
FGA (6)	14.3	2
	15.3	3
	21.1	1
	33.1	2
	34.1	2
	41.2	5
D18S51 (3)	13.3	4
	16.1	1
	28.1	1
D5S818 (2)	12.3	2
	18	2
D7S820 (1)	5.2	1
TPOX (1)	7.3	1

Miscalled Suspected Variant Alleles

While 85% of the samples contained variant alleles, the remaining 15% did not. The 135 samples comprising these 15% were recorded in error (Table 12, above). After re-reviewing the original DNA profiles for these samples, it was evident that several

amplification or electrophoresis anomalies were overlooked during the initial data collection. While the author should have detected the majority of these problems before re-testing, others would not have been apparent until after re-analysis. Each of the 135 samples was placed into four different anomaly categories:

- Missing or Low RFU Size Standards;
- DNA Fragment Migration Anomalies;
- Pull-up; and
- Incomplete Adenylation.

Missing or Low RFU Size Standards

Missing or low RFU size standards represented the largest problem observed in these DNA profiles, which affected 89 samples. The GeneScan® software is programmed to recognize size standards at 150RFUs or higher (GeneScan® Reference Guide, 1997). Standards with peak heights below this value (but above 50RFUs) were classified as low RFU standards. In addition, the Genotyper® software is programmed to remove labels from standard peaks when their heights are below 50RFUs (Genotyper® User Manual, 1998). These were classified as missing standards. The intensity of the light emitted by the fluorescently labeled standards may be too low if an insufficient sample volume is loaded onto the gel (Profiler Plus™ User Manual, 1997, Butler, 2001).

When the size standard RFU values are below these settings, some alleles may be correctly labeled with their designations, while others may be called off-ladder. Genotyper® uses the two standard peaks to the left of the allele and the two to the right to assign allele designations. For example, allele 13 at D3S1358 is sized using both the

75bp and 100bp peaks to the left of the allele and the 139bp and 150bp peaks to the right. If one of these standards has low RFUs or is missing, allele 13 may be labeled off-ladder, falsely resembling a variant allele (Figure 5).

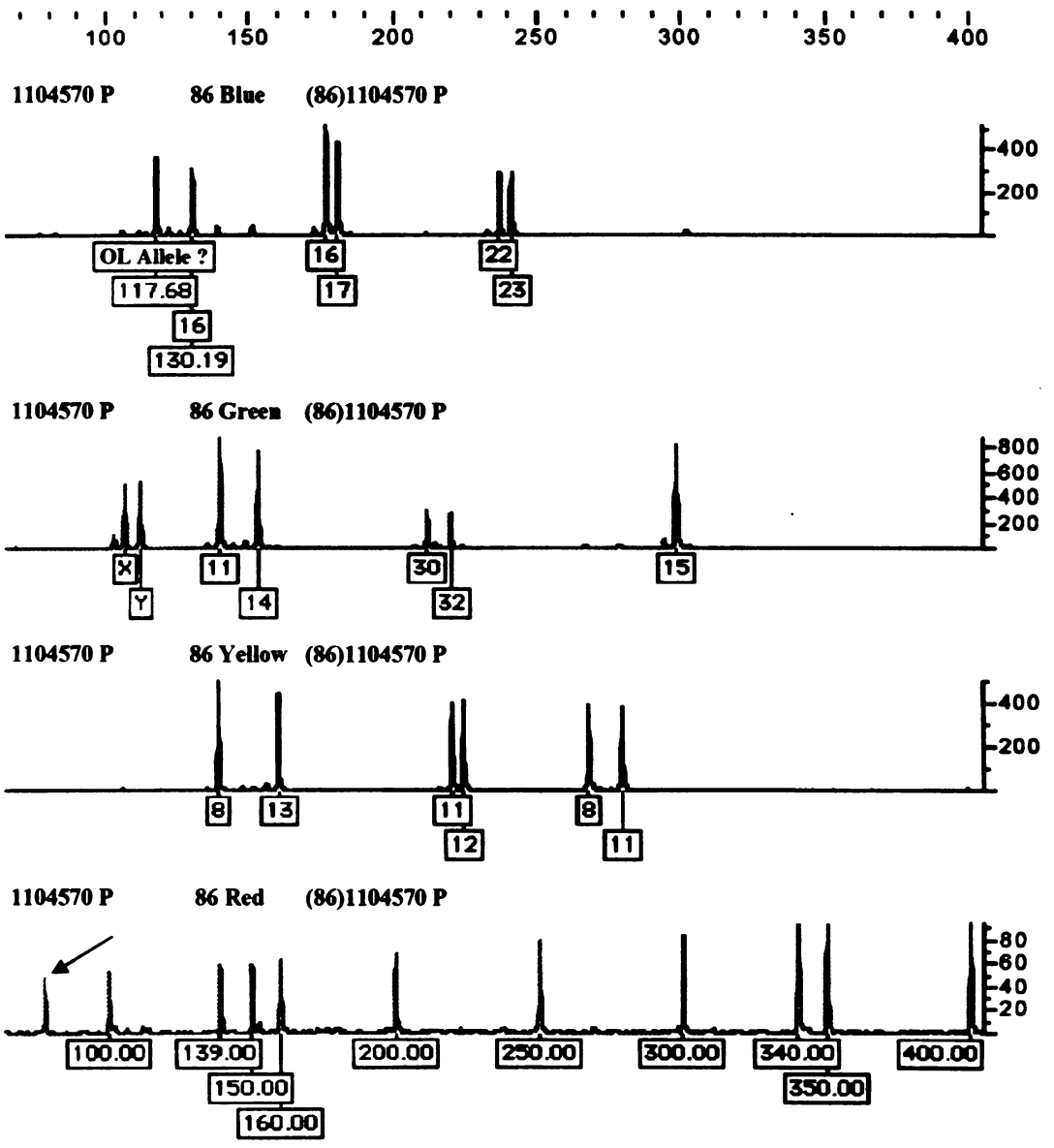


Figure 5: Missing 75bp Size Standard Peak
The suspected 12.3 microvariant at the D3S1358 locus (top panel) was caused by a missing 75bp size standard (bottom panel), indicated by the arrow. Although a peak is present where the 75bp peak should be, the height of the peak was below 50 RFUs, and was not labeled by the Genotyper® software. Therefore, accurate sizing of this allele was not possible. The size of this off-ladder peak was 117.68bp, while allele 16 was correctly sized at 130.19bp. Additional peaks are from markers in the DNA profile that were correctly assigned allele designations.

Samples with missing size standards included those where the peak was not visibly evident in the electropherogram and those with a visible peak whose height was below 50 RFUs. Eighty-three of these samples were missing only the 75bp peak, two of these were missing the 75bp, 100bp and 139bp peaks and the remaining two were missing only the 139bp peak. This anomaly only affected alleles at the D3S1358 and D8S1179 loci. At D3S1358, missing standards caused all of the suspected 12.3, 15.3, 17.2 and 18.2 microvariants. In addition, missing standards caused all but two of the suspected 13.3 microvariants and all but one of the suspected 14.3 microvariants at D3S1358 (causes for miscalls in the other samples containing these suspected microvariants discussed below). This anomaly also caused all of the suspected 7.3 and 9.2 microvariants at the D18S51 locus.

Two samples contained low RFU size standards and included suspected microvariants 13.3 at vWA and 18.1 at D18S51. Because all of the standards for these 2 samples were labeled with their bp sizes, they were overlooked as the cause of the off-ladder alleles in the DNA profiles (Figure 6).

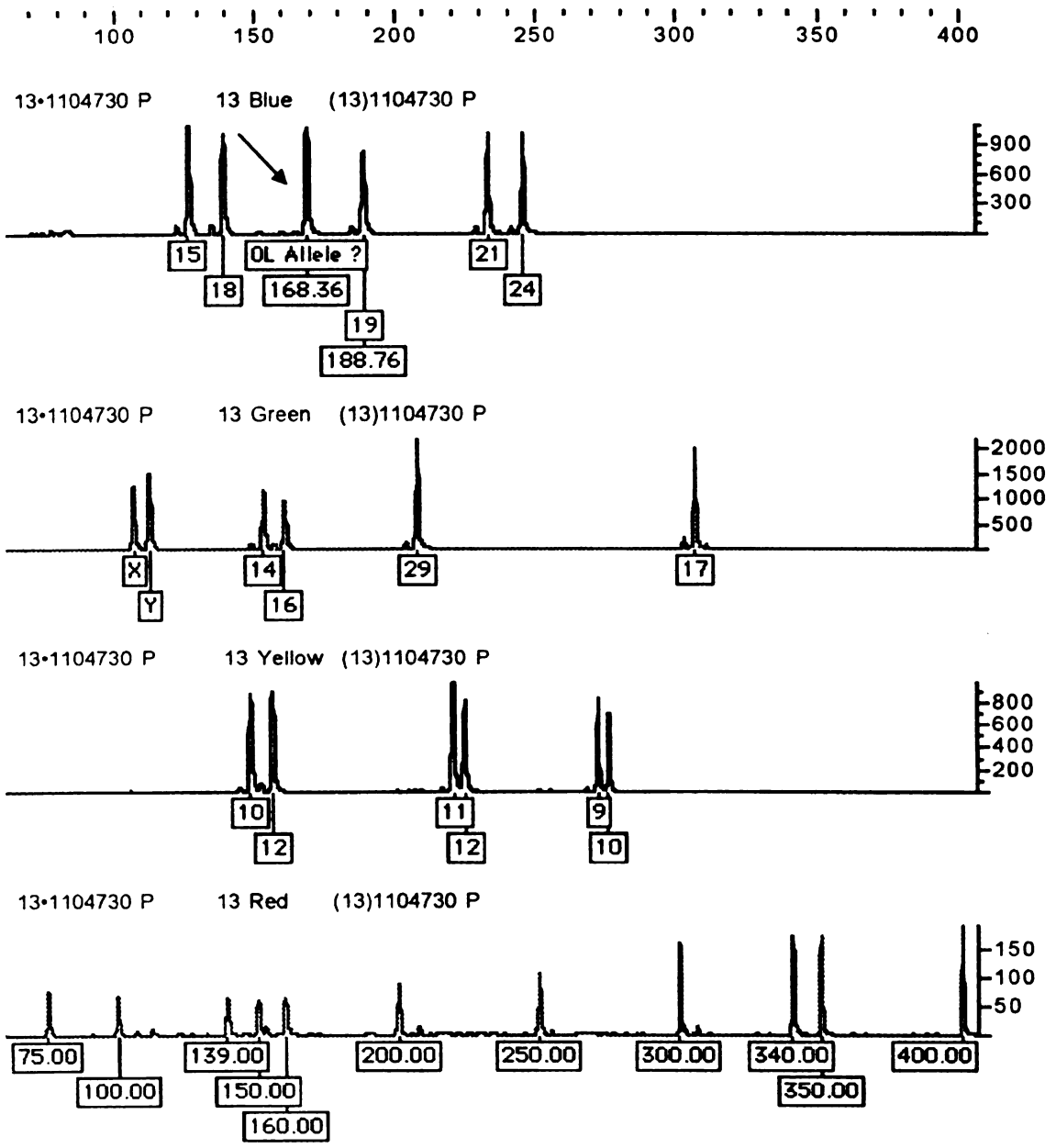


Figure 6: Low Size Standard RFUs

The 75bp – 250bp size standard peak heights in this sample (bottom panel) were all below 150RFUs. Below this value, the Genotyper® software does not always accurately assign designations. As the arrow indicates, the off-ladder allele at the vWA locus (top panel) was caused by the insufficient RFUs values of the size standards. As a result, the off-ladder allele was presumed to be a 13.3 microvariant allele, when it actually was an allele 14. Peaks in the middle two panels are from other markers.

DNA Fragment Migration Anomalies

A total of 44 different samples experienced migration anomalies when they were originally electrophoresed, and the measured sizes of the alleles were incorrect. In 8 of the samples, these anomalies were detected in the size standards (the others are discussed below). During electrophoresis, a laser scans one specific region of the gel for DNA fragments. Based on settings such as temperature and voltage, the fragments are expected to migrate through this scan region within a specific time range. This time range varies depending on the size of the fragments and is programmed in the GeneScan® software. If the size standards do not migrate within their expected time range, the software cannot assign their sizes accurately, leading to allele miscalls (GeneScan® Reference Guide, 1997).

Inaccurate sizing of the 75bp size standard resulted in one 14.3 and two 13.3 miscalls at D3S1358. In addition, two samples containing the suspected 7.3 microvariants at D7S820 were miscalled due to inaccurate sizing of both the 250bp and 300bp size standards (Figure 7). Alleles in 3 samples that were grouped with the 9.3 microvariant at D7S820 (which was confirmed using other samples) were labeled off-ladder because of inaccurate sizing of the 300bp size standard peak.

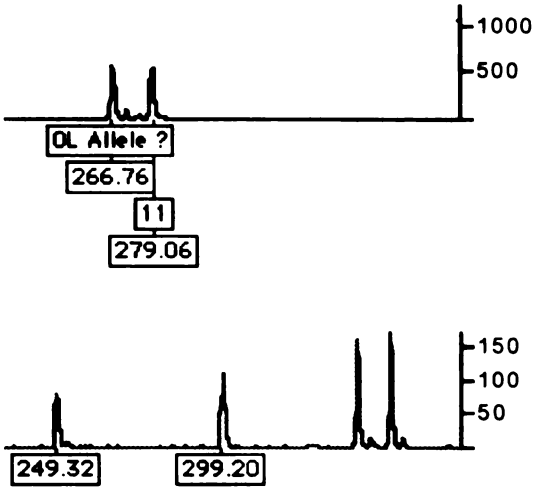


Figure 7: Migration Anomalies in the Size Standards
The incorrect sizes of the 250bp and 300bp size standards, labeled as 249.32bp and 299.20bp (bottom panel), led to inaccurate sizing of an allele at D7S820 (top panel). This allele was labeled off-ladder because of an increased migration of the size standards, which caused them to appear smaller than they actually were. The off-ladder allele was thought to be the 7.3 microvariant, but was actually an allele 8.

Migration of DNA fragments during electrophoresis is regulated by several factors, including electrophoresis temperature and current, pH and conductivity of the TE buffer and polyacrylamide gel consistency (Shewale *et al.*, 2000). Any of these factors may cause the DNA to migrate faster or slower through the gel than the expected migration times. While the exact cause for these anomalies in the size standards is uncertain, one of the most likely causes is a lack of uniformity in the gel. Minute contaminants or bubbles within the gel may retard the migration rate of the DNA, while initiating electrophoresis before the gel has completely polymerized may increase the migration rate of the DNA.

Migration anomalies were assumed to be responsible for the off-ladder alleles in 36 samples. As recommended by the Genotyper® User's Manual (1998), each allele category in the Genotyper® software is ±0.50bp wide. The size of each allele must

therefore be within ±0.50bp of the reference allele size programmed in the software before an allele designation can be assigned. Peaks that do not meet this criterion are labeled off-ladder. When alleles migrate faster or slower through the gel than expected, their measured sizes may erroneously fall outside of the actual allele categories. The purported causes of these anomalies are the same as those that affected the size standard migration rates (see above). Allele migration anomalies cannot typically be detected without re-testing, as other visible problems with the DNA profiles may not be present.

An increase or decrease in allele migration affected 6 different loci. All of the X.1 microvariants in this category had delayed migration rates, appearing to be larger than they actually were. (Figure 8, below). These were:

- 13.1 and 14.1 at D3S1358;
- 6.1 at vWA;
- 22.1 and 24.1 at FGA;
- 14.1 at D18S51;
- 11.1 at D5S818; and
- 13.1 at D13S317.

Similarly, all of the suspected X.3 microvariants in this category experienced increased migration rates, and appeared to be smaller than they actually were. These were 21.3 and 23.3 at FGA, and 10.3 at D13S317.

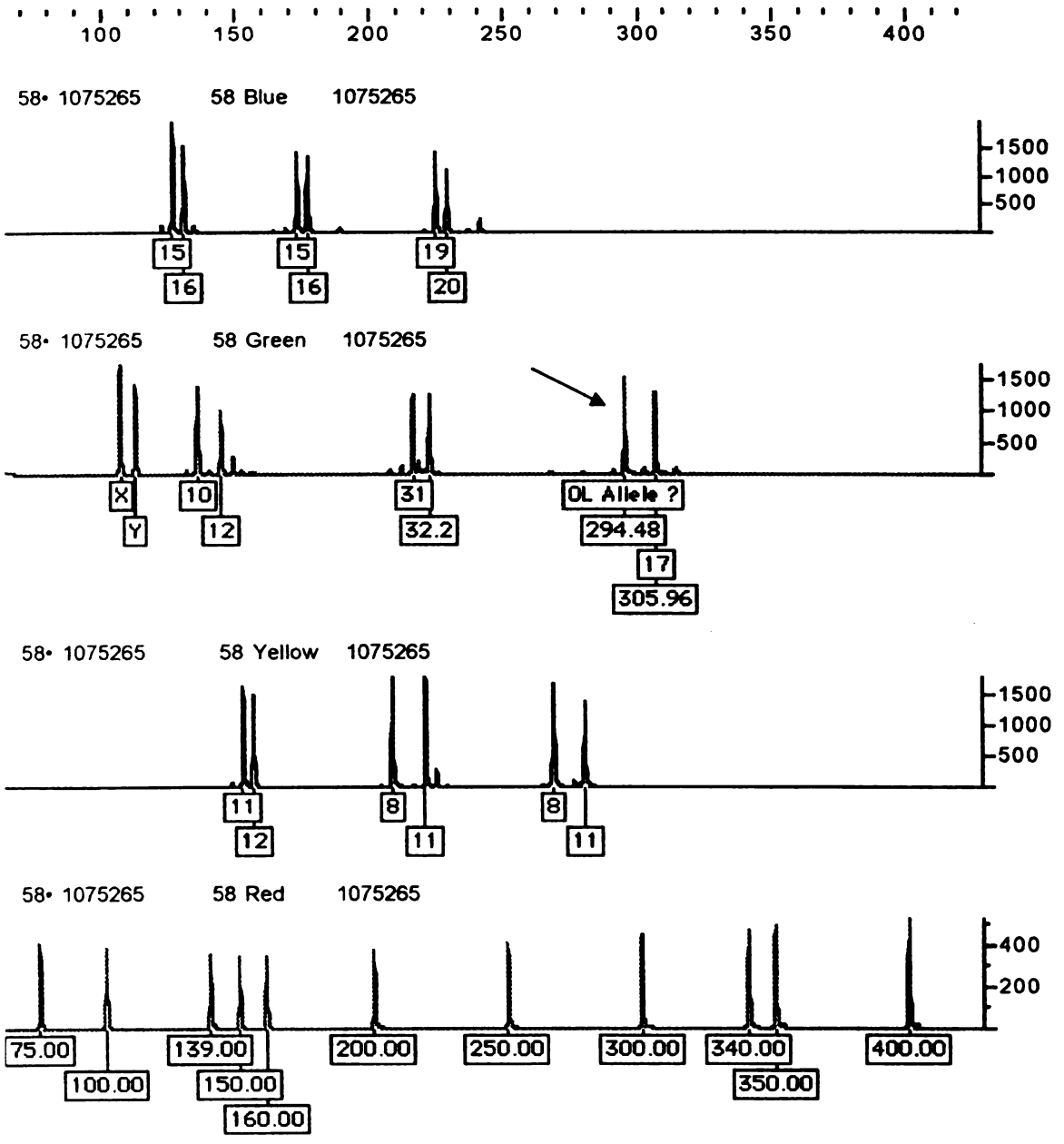


Figure 8: Allele Migration Anomaly at the D18S51 Locus
The arrow indicates the suspected 14.1 microvariant at D18S51 (second panel from top) that was a result of delayed migration of allele 14. The migration of the allele was likely slowed by minute contaminants in the gel, leading to the presence of the off-ladder allele. Peaks to the left of D18S51 in the top three panels are from other markers, while the size standards are shows in the bottom panel.

Pull-Up Peaks

Pull-up was a factor in one of the samples recorded as having a suspected variant allele. It results from a failure of the GeneScan® software to accurately resolve the fluorescent dyes used to label the DNA fragments prior to electrophoresis. While each of the dye colors emits its maximum fluorescence at a different wavelength, a substantial overlap occurs between several of the dyes, and the software is used to subtract out this overlap. When the sample is overloaded and allele peak heights exceed the detection limits for the instrument, the software may not be able to accurately subtract out the overlapping dyes, and several loci may be affected. As a result, an allele peak from one dye color can be "pulled-up" into the dye color(s) directly above or below the affected locus (Profiler Plus™ User Manual, 1997, Butler, 2001).

The off-ladder allele in the sample's original DNA profile was the suspected 14.3 microvariant at D8S1179, which was paired with an allele 11. All of the loci had peak heights averaging over 2500 RFUs each, indicating an excess of DNA. Allele 12 at D5S818 had a peak height of over 3000 RFUs, and the signal overload from this allele resulted in a pull-up peak at D8S1179 (Figure 9). After re-analysis, the only peak at D8S1179 was allele 11. While pull-up peaks in a DNA profile are typically easy to identify, the pull-up peak in this particular DNA profile was overlooked.

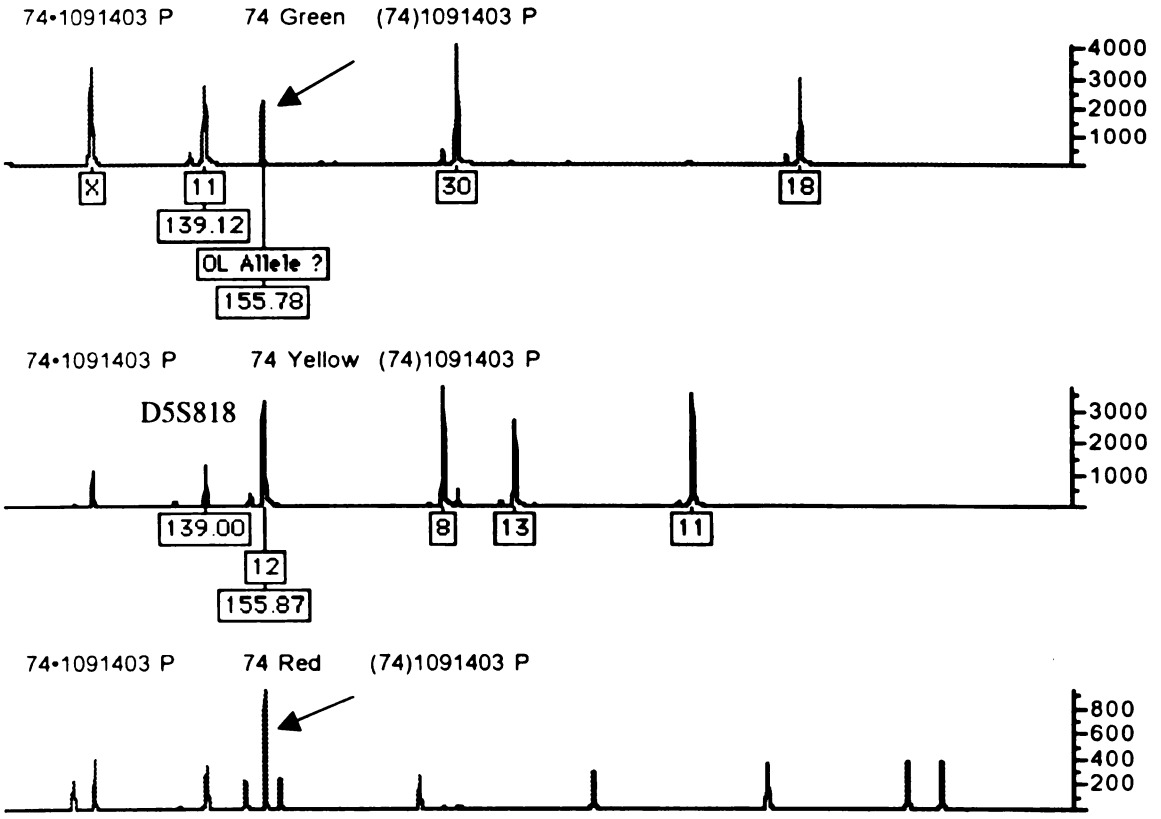


Figure 9: Off-Ladder Pull-Up Peak at the D8S1179 Locus
The top arrow indicates the pull-up peak at D8S1179 (top panel) caused by allele 12 at the D5S818 locus (middle panel), the leftmost locus in the yellow dye color. The sample was highly concentrated as evidenced by the high RFU values in the scale to the right of the allele peaks. The size of the pull-up peak (155.78bp) was very close to the size of allele 12 (155.87bp). The pull-up peak from allele 12 is also evident in the size standards (bottom panel), as indicated by the arrow. A smaller pull-up peak (labeled 139.00bp) at the D5S818 locus was caused by allele 11 in D8S1179, which had a size of 139.12bp. After re-running, the only peak present at D8S1179 was allele 11. Peaks to the right of the alleles in the top and middle panels are from other markers.

Incomplete Adenylation

The final cause for miscalled alleles in the original DNA profiles was incomplete adenylation. During PCR, a single nucleotide is added to the 3' end of the DNA target sequence. This nucleotide is typically adenosine (A) and PCR products with this additional nucleotide are in the "+A" form. The computer software assigns designations to peaks in the "+A" form. The final step of PCR gives the DNA polymerase extra time to complete this process. When an excess of DNA is present in the sample, the 45

minutes allotted for this extension may not be sufficient. As a result, some of the DNA will be in the "+A" form, and the remaining portion of DNA will be without the additional adenosine, or "-A" form. Alleles that containing both "+A" and "-A" sequences are evident as split peaks, where one peak is 1bp shorter than the "+A" target sequence. Due to the closeness in size of these two peaks, they may not be completely resolved, and the smaller peak may be labeled off-ladder (Profiler Plus™ User Manual, 1997).

A split peak at the D7S820 locus appeared to be a suspected 7.3 microvariant paired with an allele 8. This sample contained a high DNA concentration, and the peaks at this locus had heights of approximately 2000RFUs. The microvariant was 1bp smaller than its sister allele. After re-analysis, only allele 8 was present at the locus. It was evident that the suspected 7.3 microvariant was not a true allele, and instead was a peak in the "-A" form. The target sequence, allele 8, was in the "+A" form (Figure 10).

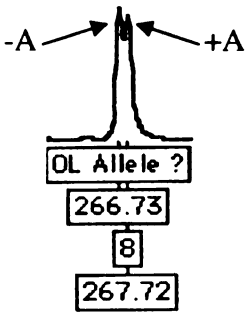


Figure 10: Incomplete Adenylation at the D7S820 Locus
The suspected 7.3 microvariant (left arrow) was caused by incomplete adenylation of allele 8 (right arrow), the target sequence. This resulted in a single allele with split peaks, where the smaller peak was in the "-A" form, and the larger peak was in the "+A" form. The suspected 7.3 allele had a size of 266.73bp, nearly 1bp smaller than allele 8, which had a size of 267.72.

Racial Group Variant Allele Data

During calculation of the variant allele frequencies, it became evident that African Americans contained the highest number of different variant alleles. Of the 85 confirmed variants, 26 were observed only in African Americans, while an additional 21 were observed predominantly in the group (i.e. at least 2 or more additional observations over other racial groups). In contrast, only 13 variants were observed exclusively in Caucasians, with 4 observed most often in this group. Seven of the variants were observed only in Hispanics, and 1 occurred predominantly in this group (Table 21, above). More Caucasian samples (16,257) were reviewed for this investigation than African American samples (14,160), and intuitively one would expect a greater number of variants in the former group. However, people of African descent have existed for a longer period of time when compared with other groups, and are known to show more variability in their DNA (Campbell, 1996).

Orchid GeneScreen Allele Frequency Database

The frequencies of the alleles already in the Orchid database decreased slightly with the addition of the 28 variant alleles. In addition, two of the 28 variants, 31.2 and 32.2 at FGA, are ladder alleles in the SGM Plus™ kit. These alleles are not found in the SGM database and the manual suggests a conservative frequency value of 1.3% be used in forensic and paternity calculations. In contrast, the 31.2 and 32.2 alleles were found in 41 and 11 samples, respectively, in this investigation (not surprising, as the number of Caucasians reviewed for this study was nearly 82 times greater than in the SGM database, and nearly 73 times greater for African Americans). The adjusted frequency

value for the 31.2 microvariant in Caucasians was 0.1106195%, which was larger than 0.0326%, the FGA 5/2N' value for Caucasians in the Orchid database, but smaller than 1.3%. In contrast, both observed frequencies for the 32.2 microvariant (0.00650703% for Caucasians and 0.0334515% for African Americans) were less than the 5/2N' values in the Orchid database (0.0326% for Caucasians and 0.0337% for African Americans), and smaller than 1.3%. Therefore, the Orchid frequencies will be used for these alleles, and will lead to higher PI values than if 1.3% were used.

Utilization of Variant Alleles for Paternity Index Calculations

Recently, protocols were put into place by the Orchid laboratory directors for the detection and utilization of the 28 variant alleles added to the database. The laboratory's LIMS software was reprogrammed to recognize these variants for use in paternity calculations. When reviewing DNA profiles containing off-ladder alleles, laboratory technicians refer to a list of the variants with their observed size ranges to determine whether or not the allele is variant. If the size of the off-ladder allele falls within the corresponding size range, it is considered variant and the peak is manually labeled with its designation and size. Data from this locus can then be used in CPI calculations. However, if the size of the off-ladder allele is very close to the observed size range but does not fall within the range, the labels from the peaks are manually removed and data from that locus are not used. The laboratory supervisors also review the DNA profiles and have the final say whether or not to use the variant allele data.

The genetic results of two paternity cases tested by Orchid GeneScreen are included below to demonstrate the increase in the CPI value when using loci with variant

alleles (Tables 25 and 26). Calculations used to determine the PI value for each locus are included in Appendix D. The first example in Table 25 uses frequency data from the 33.1 microvariant at D21S11, which was found in more samples (76) than any other variant during this investigation. This demonstrates the expected increase in CPI when one of the more "common" microvariants is the obligate paternal allele.

Table 25: CPI Calculations – Most Common Variant Allele

Allele designations are listed for the mother, child and alleged father for each of the loci with data for a case where the alleged father was included. PI values were determined using calculations from Traver, 1998 (Appendix D). The values in both PI columns were calculated using the African American allele frequencies listed in the Orchid GeneScreen allele frequency database. The second PI column also includes variant data from the D21S11 locus in bold face. The CPI values for both columns were calculated by multiplying the PI values together. The calculated probability of paternity is included at the bottom of each of these columns.

Locus	M	C	AF	PI Without Variant Allele	PI With Variant Allele
D13S317	12, 13	12	11, 12	1.17	1.17
D18S51	13, 17	12, 13	12, 19	8.03	8.03
D7S820	8, 12	8, 12	8	3.21	3.21
D8S1179	10, 15	14, 15	14	2.99	2.99
FGA	22, 25	20, 25	20, 28	8.12	8.12
vWA	16, 15	16, 17	15, 17	2.46	2.46
D21S11	30, 31	30, **33.1**	31, **33.1**	-	**240.17**
				CPI = 1,801	**CPI = 432,601**
				Probability of Paternity = 99.94%	**Probability of Paternity = 99.99%**

The addition of the D21S11 33.1 PI value of 240.17 increased the CPI value from 1,801 to 432,601. This indicates a probability of 1 in 432,601 of selecting a random, unrelated man from the same racial group as the biological father. While the CPI value of 1,801 was sufficient for reporting an inclusion for this case, another case may have required a higher CPI value, and additional testing would have been necessary prior to reporting the CPI. Because variant alleles are relatively rare when compared to the ladder alleles, their

frequencies are small. The smaller these frequencies are, the larger the CPI value becomes. The dramatic increase in the CPI value in the case above clearly demonstrates that the use of loci containing even the most common variant allele will only provide added support for the conclusion of paternity when an inclusion is indicated.

In Table 25, the probability of paternity was calculated to be 99.94% before the inclusion of the variant allele, and 99.99% after the inclusion. Orchid GeneScreen guarantees a minimum probability of paternity of 99.0% with indication of parentage, so probability values higher than this only increase the confidence in the paternity conclusions. The 0.05% increase in the probability of paternity further strengthens the hypothesis that the correct man has been identified as the biological father.

While the CPI using the most common variant allele increased by a factor of over 200, the CPI value for a case containing one of the "rarest" variant alleles is even more compelling, as shown in Table 26. The 9.3 microvariant at the D7S820 locus was found in only 5 samples, the minimum number required for inclusion in the database.

Table 26: CPI Calculations – Least Common Variant Allele

Allele designations are listed for the mother, child and alleged father for each of the loci with data for a case where the alleged father was included. PI values were determined using calculations from Traver, 1998 (Appendix D). The values in both PI columns were calculated using the African American allele frequencies listed in the Orchid GeneScreen allele frequency database. The second PI column also includes variant data from the D7S820 locus in bold face. The CPI values for both columns were calculated by multiplying the PI values together. The calculated probability of paternity is included at the bottom of each of these columns. [a] PI calculated using 5/2N' allele frequency (0.0328%) instead of observed allele frequency (0.013128%)

Locus	M	C	AF	PI Without Variant Alleles	PI With Variant Alleles
D13S317	12	8, 12	8, 13	19.23	19.23
D18S51	15, 18	17, 18	17, 18	3.06	3.06
D21S11	28, 29	28, 30	28, 30	2.69	2.69
D3S1358	15	15, 16	16, 17	1.52	1.52
D5S818	9, 12	9, 12	12, 13	1.34	1.34
D8S1179	10, 16	15, 16	15	5.15	5.15
FGA	22, 23	19, 22	19, 25	7.57	7.57
vWA	15	15	15	4.69	4.69
D7S820	9, 12	**9.3**, 12	**9.3**, 12	-	**1524**[a]
				CPI = 58,949	**CPI = 89,838,459**
				Probability of Paternity = 99.99%	**Probability of Paternity = 99.99%**

The addition of the D7S820 9.3 PI value of 1524 increased the CPI value from 58,949 to 89,838,459. This indicates a probability of 1 in 89,838,459 of selecting a random, unrelated man from the same racial group as the biological father. The 5/2N' frequency for the D7S820 locus was used to calculate the PI, as the observed frequency for 9.3 was sufficiently small. If the observed frequency had been used, the PI value would have more than doubled to 3809, and the CPI value would have increased to 224,516,759. This indicates the results of using a more conservative value when the allele is considerably rare in the racial group. The probability of paternity was calculated to be 99.99% both before and after the addition of the D7S820 locus. While there was no change at two decimal places, the CPI alone indicates that the strength of the genetic evidence increased because of the addition of the D7S820 PI value.

In both cases, the obligate paternal allele was a variant. This is not a requirement in order to use data from a locus containing a variant allele in CPI calculations. Data from any locus containing one of the 28 variant alleles, whether obligate or not, can now be used in CPI calculations. The use of these variant alleles in CPI calculations ensures that data are not discarded solely because the analysis software does not recognize them.

Reductions in Orchid GeneScreen Laboratory Expenditures

The decline in additional testing required for paternity cases containing variant alleles is expected to cause a decrease in laboratory expenditures. Currently, the exact role these variants will play in the reduction of laboratory costs is unknown. Nearly 12% of the samples tested annually require additional loci before paternity conclusions can be made. The 28 variant alleles added to the Orchid database are found in approximately 1.6% of samples, or 2,160 samples tested yearly. It is probable that the retest rate could decrease to 10.4% with the use of variant allele data. A decrease in the number of samples requiring additional testing should lead to the purchase of fewer amplification kits annually.

Other factors that affect laboratory costs are more difficult to pinpoint. These include the number of combined man-hours needed to complete the testing on a case and whether or not the case belongs to a public or private account. It is likely that the cost saving benefits of utilizing variant allele data will not be known until after they have been used for at least one year.

Suggestions for Future Research

Although this project uncovered valuable information about variant alleles and their usefulness in paternity dispute resolutions, several avenues are still in need of pursuit. For instance, plans are being discussed to establish a more accurate size range for each of the 28 variants added to the database. Currently, the observed size range is used to determine whether or not a variant allele is present. However, there have been circumstances where the size of an off-ladder allele is very close to the observed variant size range, but does not fall within the range. Recently, a sample contained an off-ladder allele with a size of 200.07bp at the D21S11 locus. This allele was 0.02bp smaller than the size range of 200.09 – 200.19bp for the 27.1 microvariant, and was therefore not considered a variant. It is possible that this off-ladder allele was a 27.1 microvariant, but the laboratory protocols dictate that variant alleles must size within the observed range.

In order to more accurately define the size ranges for these alleles, the laboratory directors have proposed that each sample containing one of these variants be re-amplified between 5 – 10 times and electrophoresed on the same number of gels. According to the Profiler Plus™ User Manual (1997), allele sizes may differ between gel runs on the same instrument due to variation in gel concentration and thickness, run temperature, and the distance between the wells and the laser-scanning region. In addition, the slight procedural and reagent variations that may affect several different gels can result in greater size variation than would occur with samples analyzed on the same gel. While the size of the variants found in these samples may measure within the observed ranges, it is anticipated that some will have sizes beyond the range. In this way, variant alleles in

future samples that would have sized outside of the current observed range will be recognized as variant, and their data will not be discarded.

Conclusions

This investigation uncovered many rare alleles that are found in individuals of different racial descent, which can play a large role in resolving paternity disputes. The author speculates that the use of these additional alleles may someday benefit those involved in forensic DNA testing. Perhaps many of these will be included in allelic ladders used for DNA analysis in the future, or technological advances will allow these alleles to be readily recognized. It is important that scientists continue to research these and other rare alleles, as an increase in the number of known alleles will only improve the discriminatory power of human identity testing.

APPENDICES

APPENDIX A

Appendix A contains the alleles with their corresponding frequencies found in the Orchid GeneScreen Allele Frequency Database (prior to the addition of the 28 variant alleles).

Table 27: Original Orchid GeneScreen Allele Frequency Database

Allele frequencies are given as percentages and were calculated prior to the addition of the variant allele data. The N values indicate the number of individuals of a given racial group that had genotypes at that locus. For each racial group, frequencies were calculated based on the total number of alleles present in each locus.

D3S1358	**Caucasian** N' = 7636	**African American** N' = 7602	**Hispanic** N' = 690
11	0.0851	0.0329	0.2174
12	0.0655	0.454	0.145
13	0.308	0.77	0.507
14	12.40	9.05	7.50
15	26.90	29.20	36.60
15.2	0.0	0.0987	0.0
16	24.30	33.0	28.60
17	20.00	20.70	15.70
18	14.60	6.30	9.93
19	1.25	0.48	0.797
20	0.0131	0.0	0.0725
vWA	**Caucasian** N' =7837	**African American** N' = 7854	**Hispanic** N' = 703
11	0.0128	0.739	0.213
12	0.0447	0.115	0.142
13	0.0957	1.69	0.498
14	9.65	6.32	6.26
15	9.90	21.30	10.20
16	21.80	26.0	31.50
17	27.0	20.30	26.40
18	21.40	13.90	16.60
19	8.47	6.97	6.69
20	1.55	2.23	1.42
21	0.109	0.427	0.0
22	0.0128	0.07	0.0

Table 27 (cont'd)

FGA	Caucasian N' = 7674	African American N' = 7419	Hispanic N' = 939
15	0.0	0.0	0.0532
17	0.0847	0.148	0.161
17.2	0.0	0.0809	0.0
18	1.62	0.748	0.959
18.2	0.0	1.24	0.0
19	6.36	6.65	8.15
19.2	0.0065	0.283	0.107
20	14.70	6.20	10.5
20.2	0.15	0.135	0.0
21	17.70	11.20	12.10
21.2	0.319	0.148	0.0
22	17.70	18.0	14.30
22.2	1.0	0.202	0.213
23	14.10	17.8	14.70
23.2	0.437	0.101	0.373
24	13.40	16.8	15.40
24.2	0.111	0.0135	0.0532
25	8.50	9.89	13.2
25.2	0.013	0.0212	0.0
26	3.0	5.06	6.50
26.2	0.013	0.0135	0.0
27	0.593	3.19	2.29
28	0.156	1.25	0.639
29	0.0261	0.499	0.107
29.2	0.0	0.0135	0.0
30	0.013	0.175	0.107
30.2	0.0	0.135	0.0
31	0.0065	0.472	0.0532
D8S1179	**Caucasian** N' = 7731	**African American** N' = 7796	**Hispanic** N' = 679
8	1.69	0.237	0.663
9	1.27	0.487	0.957
10	9.15	2.32	8.62
11	7.24	4.45	6.11
12	14.70	11.90	12.20
13	32.60	20.30	29.50
14	19.80	33.50	25.30
15	10.60	19.40	13.20
16	2.67	6.06	3.09
17	0.362	1.19	0.368
18	0.0194	0.122	0.0736
19	0.0	0.0128	0.0

Table 27 (cont'd)

D21S11	Caucasian N' = 7730	African American N' = 7652	Hispanic N' = 683
24.2	0.0776	0.0	0.0
25	0.0129	0.0327	0.0
25.2	0.0776	0.0131	0.0
26	0.207	0.124	0.146
26.2	0.0194	0.0	0.0
27	3.32	5.60	1.98
28	15.80	24.30	10.70
28.2	0.0194	0.0065	0.0
29	21.20	18.90	21.40
29.2	0.104	0.0588	0.366
30	25.80	18.70	25.60
30.2	3.53	1.92	3.0
31	7.37	8.09	8.27
31.2	8.96	5.20	11.70
32	1.42	1.720	0.952
32.2	8.54	6.94	10.60
33	0.207	0.457	0.146
33.2	2.90	3.08	4.25
34	0.0065	0.588	0.293
34.2	0.349	0.314	0.0732
35	0.0194	3.14	0.366
35.2	0.0388	0.0261	0.0
36	0.0065	0.608	0.0732
36.2	0.0065	0.0	0.0
37	0.0	0.137	0.0732
38	0.0	0.0523	0.0
D18S51	**Caucasian** N' = 7628	**African American** N' = 7463	**Hispanic** N' = 647
8	0.0	0.0067	0.0
9	0.078700	0.127	0.0
10	0.904600	0.194	0.773
10.2	0.0	0.161	0.0773
11	1.120900	0.496	1.16
12	14.486100	6.25	12.20
13	12.624500	4.76	11.10
13.2	0.013100	0.496	0.0773
14	16.439400	6.58	17.0
14.2	0.0	0.369	0.0
15	14.407400	16.60	14.10
16	12.814600	18.10	11.80
17	11.602000	16.40	17.10
18	7.734700	11.60	6.65
19	4.404800	9.13	3.09
19.2	0.0	0.0402	0.0
20	1.730500	5.37	1.93
21	0.976700	2.22	1.16
22	0.458800	0.864	1.31
23	0.144200	0.214	0.309
24	0.052400	0.0335	0.155
25	0.0	0.0201	0.0773
26	0.006600	0.0	0.0

Table 27 (cont'd)

D5S818	Caucasian N' = 7829	African American N' = 7845	Hispanic N' = 705
3	0.0	0.0064	0.0
6	0.0064	0.0	0.0
7	0.204	0.331	4.82
8	0.326	5.55	1.42
9	3.55	1.90	4.75
10	5.84	5.98	4.89
11	36.90	23.70	40.60
12	36.20	35.30	30.50
13	15.90	25.10	12.30
14	0.926	1.75	0.638
15	0.102	0.293	0.0
16	0.0	0.102	0.0
17	0.0	0.0191	0.0
D13S317	**Caucasian** N'= 7814	**African American** N' = 7833	**Hispanic** N' = 703
7	0.0256	0.0128	0.0
8	12.0	2.60	7.89
9	7.54	2.18	16.80
10	6.18	2.73	7.89
11	31.10	29.40	22.0
12	28.30	42.90	26.60
13	10.40	15.20	12.70
14	4.43	4.86	5.97
15	0.141	0.102	0.142
16	0.0	0.0191	0.0
D7S820	**Caucasian** N' = 7685	**African American** N'= 7612	**Hispanic** N' = 939
6	0.013	0.105	0.0
7	2.08	0.769	1.38
8	15.60	20.40	12.80
9	16.50	11.40	8.63
10	26.80	33.90	26.60
11	20.40	20.40	28.0
12	14.50	10.80	18.40
13	3.33	2.02	3.73
14	0.709	0.21	0.479
15	0.0195	0.0	0.0
16	0.0065	0.0	0.0

APPENDIX B

Appendix B contains the alleles with their corresponding frequencies found in the Orchid GeneScreen Allele Frequency Database (after addition of the 28 variant alleles).

Table 28: Modified Orchid GeneScreen Allele Frequency Database

Only loci containing variant alleles are shown. Allele frequencies are given as percentages, with variant allele data in bold face. The N' values indicate the number of people of a given racial group that had genotypes for a given locus. The n' values indicate the total number of alleles (including variants) present in the database. Allele frequencies were re-calculated by dividing the number of observations by n'. The conservative 5/2N' values vary as N' differed for each locus.

D3S1358	Caucasian N' = 7636 n' = 15274 5/2N' = 0.0327	African American N' = 7602 n' = 15244 5/2N' = 0.0329	Hispanic N' = 690 n' = 1383 5/2N' = 0.362
9	**0.00654707**	**0.163999**	**0.0**
11	0.0850889	0.0328137	0.216928
12	0.0654914	0.4528087	0.144685
13	0.3079597	0.7679795	0.5059
14	12.3983763	9.026253	7.483731
15	26.8964777	29.1233797	36.520607
15.2	0.0	0.098441	0.0
16	24.2968181	32.9134086	28.537961
17	19.9973812	20.6456835	15.665944
18	14.5980883	6.2834689	9.90846
19	1.2498363	0.4787405	0.795271
20	0.0130983	0.0	0.072343
20.1	**0.085112**	**0.0131199**	**0.144613**
FGA	**Caucasian** N' = 7674 n' = 15368 5/2N' = 0.0326	**African American** N' = 7419 n' = 14947 5/2N' = 0.0337	**Hispanic** N' = 939 n' = 1879 5/2N' = 0.266
15	0.0	0.0	0.0531717
16.1	**0.0**	**0.0735934**	**0.0**
16.2	**0.0**	**0.0267612**	**0.0**
17	0.0845898	0.1469207	0.1609143
17.2	0.0	0.08031	0.0
18	1.6178917	0.7425453	0.9584896
18.2	0.0	1.2309574	0.0
19	6.3517231	6.6015053	8.1456626
19.2	0.00649154	0.2809362	0.1069431
20	14.6808693	6.1547869	10.4944119
20.2	0.1498048	0.1340155	0.0
21	17.6769651	11.1183247	12.0935604
21.2	0.3185849	0.1469207	0.0
22	17.6769651	17.8687362	14.2923896
22.2	0.9986986	0.2005269	0.2128866
22.3	**0.0**	**0.0200709**	**0.0532198**
23	14.0816502	17.6701947	14.6921767
23.2	0.4364313	0.1002635	0.3728015
23.3	**0.0**	**0.0401418**	**0.0532198**
24	13.3825612	16.6774871	15.3918042
24.1	**0.0**	**0.0200709**	**0.0**
24.2	0.1108555	0.0134016	0.0531717
24.3	**0.00650703**	**0.0200709**	**0.0**
25	8.4889381	9.8178778	13.192975
25.2	0.0129831	0.0210454	0.0
25.3	**0.0**	**0.0267612**	**0.0**
26	2.9960958	5.0231003	6.4965407
26.2	0.0129831	0.0134016	0.0
27	0.5922283	3.1667371	2.2887813
28	0.155797	1.2408845	0.6386599

Table 28 (cont'd)

FGA (cont'd)	Caucasian N' = 7674 n' = 15368 5/2N' = 0.0326	African American N' = 7419 n' = 14947 5/2N' = 0.0337	Hispanic N' = 939 n' = 1879 5/2N' = 0.266
29	0.026066	0.4953611	0.1069431
29.2	0.0	0.0134016	0.0
30	0.0129831	0.1737238	0.1069431
30.2	0.0	0.1340155	0.0
31	0.00649154	0.468558	0.0531717
31.2	**0.1106195**	**0.0**	**0.0**
32.2	**0.00650703**	**0.0334515**	**0.0**
D21S11	**Caucasian** N' = 7730 n' = 15467 5/2N' = 0.0323	**African American** N' = 7652 n' = 15371 5/2N' = 0.0327	**Hispanic** N' = 683 n' = 1369 5/2N' = 0.366
24.2	0.077565	0.0	0.0
24.3	**0.0**	**0.078069**	**0.073046**
25	0.012894	0.032557	0.0
25.2	0.077565	0.013043	0.0
26	0.206906	0.12346	0.14568
26.2	0.019391	0.0	0.0
27	3.318497	5.57559	1.975661
27.1	**0.0**	**0.019517**	**0.0**
28	15.792849	24.19408	10.676552
28.2	0.019391	0.0064717	0.0
29	21.190405	18.817618	21.353104
29.2	0.103953	0.058544	0.365198
29.3	**0.025862**	**0.026023**	**0.0**
30	25.788324	18.618489	25.543901
30.2	3.528402	1.911631	2.993426
31	7.366665	8.054737	8.251877
31.2	8.955945	5.177334	11.674361
32	1.419357	1.712503	0.949914
32.2	8.536135	6.90975	10.576771
33	0.206906	0.455008	0.14568
33.1	**0.019396**	**0.208184**	**0.073046**
33.2	2.898688	3.066575	4.240687
34	0.0064971	0.585437	0.292358
34.1	**0.0064654**	**0.058552**	**0.073046**
34.2	0.348842	0.312631	0.07304
35	0.019391	3.126313	0.365198
35.1	**0.0**	**0.039035**	**0.0**
35.2	0.038782	0.025986	0.0
36	0.0064971	0.60535	0.07304
36.2	0.0064971	0.0	0.0
37	0.0	0.136403	0.07304
38	0.0	0.052072	0.0

Table 28 (cont'd)

D18S51	Caucasian N' = 7628 n' = 15258 5/2N' = 0.0328	African American N' = 7463 n' = 14974 5/2N' = 0.0335	Hispanic N' = 647 n' = 1296 5/2N' = 0.386
8	0.0	0.00667852	0.0
9	0.07869	0.1265929	0.0
10	0.904481	0.1933781	0.7718071
10.2	0.0	0.1604839	0.0771807
11	1.120753	0.49441	1.1582099
11.2	**0.0**	**0.0200347**	**0.0**
12	14.484201	6.2299653	12.1811728
12.2	**0.0**	**0.026713**	**0.0**
13	12.622845	4.7447416	11.0828704
13.2	0.013098	0.49441	0.0771807
14	16.437245	6.5589074	16.9737654
14.2	0.0	0.3678171	0.0
15	14.405511	16.5467878	14.0782407
15.2	**0.0065539**	**0.1268866**	**0.0771605**
16	12.81292	18.0419794	11.7817901
17	11.600479	16.3474289	17.0736111
18	7.733686	11.5628155	6.6397377
19	4.404223	9.1007333	3.0852315
19.2	0.0	0.0400711	0.0
20	1.730273	5.3527862	1.9270216
20.2	**0.0**	**0.0333912**	**0.0**
21	0.976572	2.2128837	1.1582099
21.2	**0.0065539**	**0.0801389**	**0.0**
22	0.45874	0.8612304	1.3079784
23	0.144181	0.213314	0.3085231
24	0.052393	0.0333926	0.1547608
25	0.0	0.0200356	0.0771807
26	0.0065991	0.0	0.0
D7S820	**Caucasian** N' = 7685 n' = 15376 5/2N' = 0.0325	**African American** N' = 7612 n' = 15235 5/2N' = 0.0328	**Hispanic** N' = 939 n' = 1887 5/2N' = 0.266
6	0.012995	0.104924	0.0
6.3	**0.019511**	**0.0**	**0.0**
7	2.079188	0.768445	1.3734181
8	15.593913	20.385271	12.7389507
9	16.493561	11.391769	8.5888394
9.1	**0.019511**	**0.0065638**	**0.0**
9.3	**0.0**	**0.013128**	**0.0529942**
10	26.789542	33.875523	26.473132
10.1	**0.0065036**	**0.039383**	**0.0529942**
10.3	**0.0065036**	**0.0**	**0.3709592**
11	20.39204	20.385271	27.8664547
11.1	**0.026015**	**0.0065638**	**0.0**
12	14.494342	10.792202	18.3122417
13	3.328701	2.018542	3.7122099
14	0.708723	0.209848	0.4767154
15	0.019492	0.0	0.0
16	0.0064975	0.0	0.0

APPENDIX C

Appendix C contains electropherograms for the 14 novel variant alleles observed during this investigation.

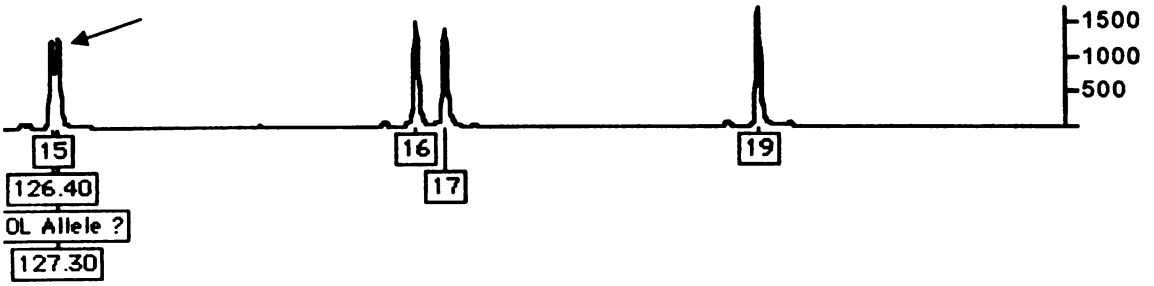


Figure 11: 15.1 Microvariant at the D3S1358 Locus
As indicated by the arrow, this 15.1 microvariant at the D3S1358 locus was the larger sized peak in a split peak that also contained allele 15. It had a size of 127.30bp with a height of approximately 1500RFUs. While not completely resolved due to the closeness in size to its sister allele, the 15.1 microvariant allele was not a result of incomplete adenylation. This novel allele was 1bp larger than allele 15, which has an average size of 126.34bp when analyzed on an ABI 377, and allele 15 in this sample had a size of 126.40bp. Peaks to the right of D3S1358 are from other markers.

Figure 12: Below Ladder 14.3 Microvariant at the FGA Locus
As indicated by the arrow, this 14.3 microvariant at the FGA locus had a size of 207.79bp with a height of approximately 2500RFUs. This novel allele was 13bp smaller than allele 18, which has an average size of 220.00bp when analyzed on an ABI 377. The microvariant was paired with a 21.2 allele, which had a size of 234.85bp and a height of approximately 2500RFUs. Peaks to the left of FGA are from other markers.

Figure 13: Below Ladder 15.3 Microvariant at the FGA Locus
The arrow indicates the 15.3 microvariant at the FGA locus, which had a size of 211.88bp and a height of approximately 1100RFUs. This novel allele was 9bp smaller than allele 18, which has an average size of 220.00bp when analyzed on an ABI 377. The microvariant was paired with an allele 24, which had a size of 245.11bp and a height of approximately 1200RFUs. Peaks to the left of FGA are from other markers.

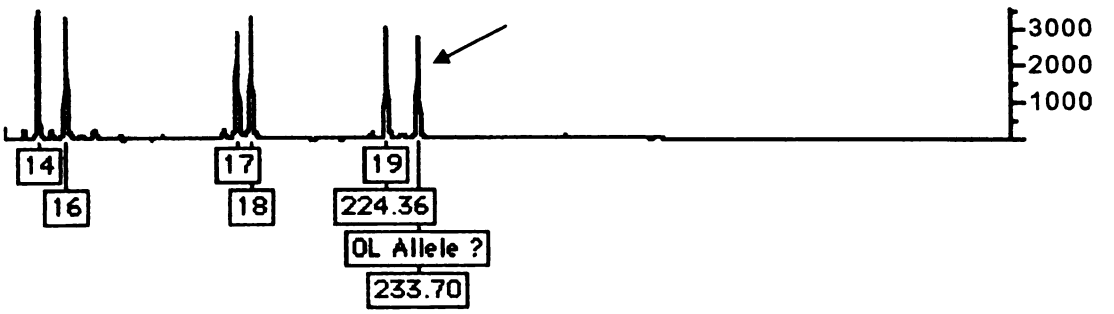


Figure 14: 21.1 Microvariant at the FGA Locus
The arrow indicates the 21.1 microvariant at the FGA locus, which had a size of 233.70bp and a height of approximately 2500RFUs. This novel allele was 1bp larger than allele 21, which has an average size of 232.38bp when analyzed on an ABI 377. The microvariant was paired with an allele 19, which had a size of 224.36bp and a height of approximately 3000RFUs. Peaks to the left of FGA are from other markers.

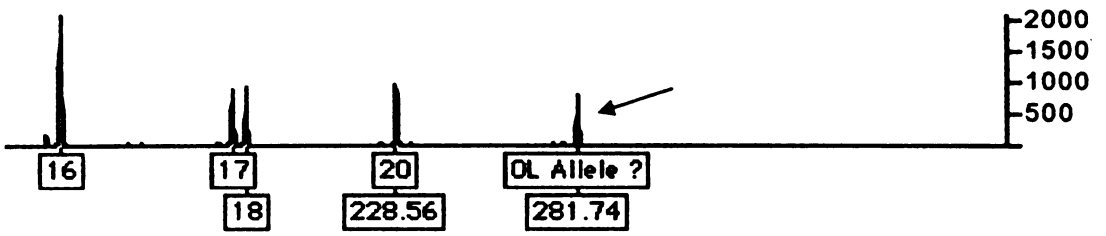


Figure 15: Above Ladder 33.1 Microvariant at the FGA Locus
As indicated by the arrow, this 33.1 microvariant at the FGA locus had a size of 281.74bp with a height of approximately 1000RFUs. This novel allele was 13bp larger than allele 30, which has an average size of 268.62bp when analyzed on an ABI 377. The microvariant was paired with an allele 20, which had a size of 228.56bp and a height of approximately 1200RFUs. Peaks to the left of FGA are from other markers.

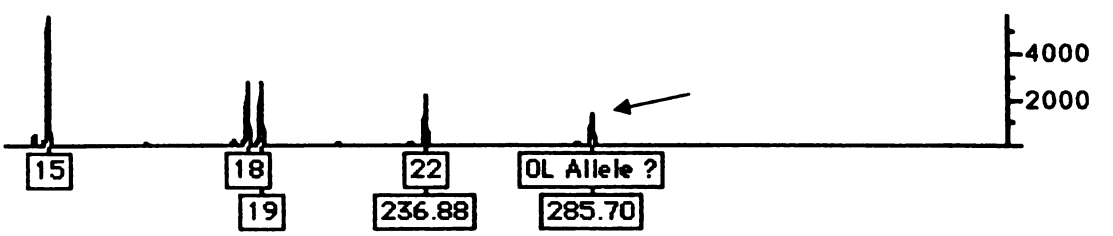


Figure 16: Above Ladder 34.1 Microvariant at the FGA Locus
The arrow points to the 34.1 microvariant allele at the FGA locus, which had a size of 285.70bp and a height of approximately 2000RFUs. This novel allele was 17bp larger than allele 30, which has an average size of 268.62bp when analyzed on an ABI 377. The microvariant was paired with an allele 22, which had a size of 236.88bp and a height of approximately 2500RFUs. Peaks to the left of FGA are from other markers.

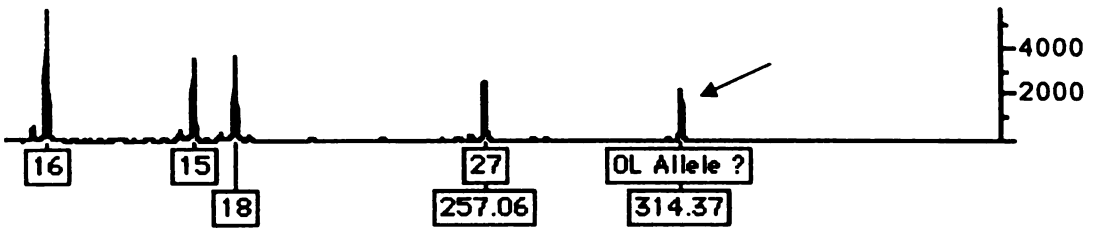


Figure 17: Above Ladder 41.2 Microvariant at the FGA Locus
The arrow indicates the 41.2 microvariant at the FGA locus, which had a size of 314.37bp and a height of approximately 2000RFUs. This novel allele was 46bp larger than allele 30, which has an average size of 268.62bp when analyzed on an ABI 377. The microvariant was paired with an allele 27, which had a size of 257.06bp and a height of approximately 3000RFUs. Peaks to the left of FGA are from other markers.

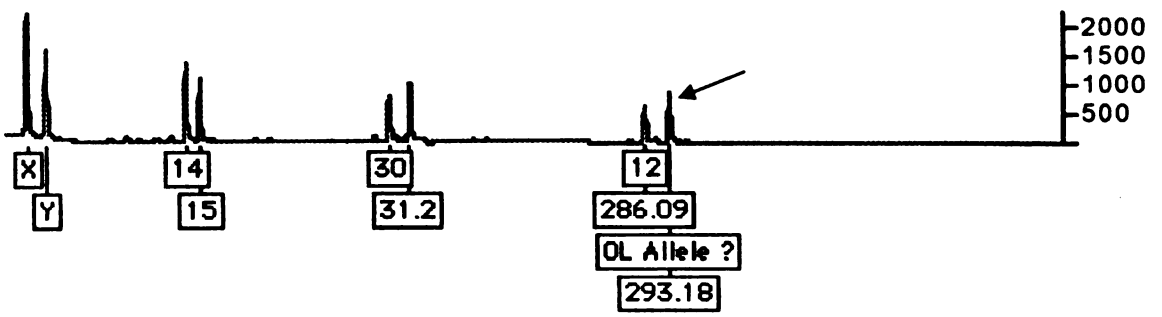


Figure 18: 13.3 Microvariant at the D18S51 Locus
The arrow points to microvariant allele 13.3 at the D18S51 locus, which had a size of 293.18bp and a height of approximately 1000RFUs. This novel allele was 1bp smaller than allele 16, which has an average size of 301.81bp when analyzed on an ABI 377. It was paired with an allele 12, which had a size of 286.09bp and a height of approximately 800 RFUs. Peaks to the left of D18S51 are from other markers.

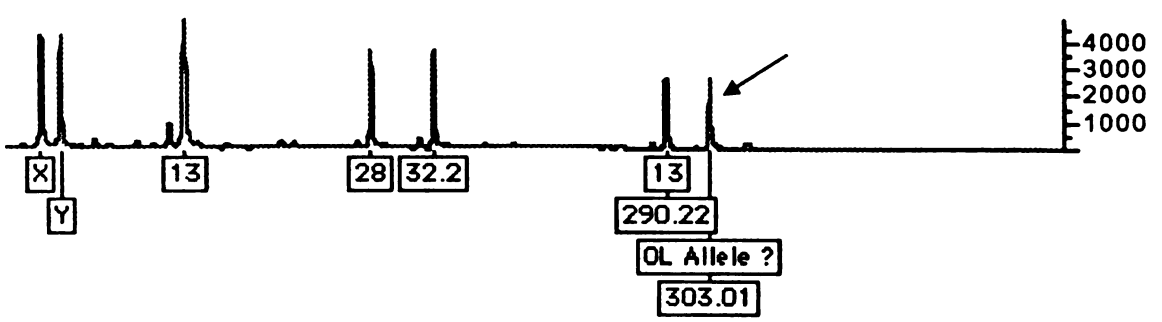


Figure 19: 16.1 Microvariant at the D18S51 Locus
The arrow points to microvariant allele 16.1 at the D18S51 locus, which had a size of 303.01bp and a height of approximately 3000RFUs. This novel allele was 1bp larger than allele 16, which has an average size of 301.81bp when analyzed on an ABI 377. It was paired with an allele 13, which had a size of 290.22bp and a height of approximately 3000 RFUs. Peaks to the left of D18S51 are from other markers.

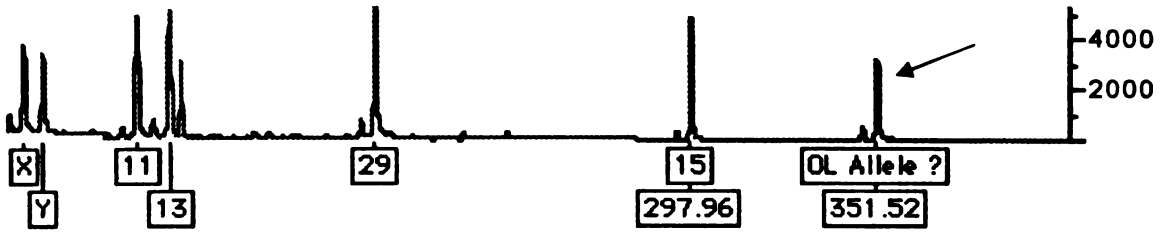


Figure 20: Above Ladder 28.1 Microvariant at the D18S51 Locus
The arrow indicates the 28.1 microvariant at the D18S51 locus, which had a size of 351.52bp and a height of approximately 2500RFUs. This novel allele was 9bp larger than allele 26, which has an average size of 342.71bp when analyzed on an ABI 377. It was paired with an allele 15, which had a size of 297.96bp and a height of approximately 5000 RFUs. Peaks to the left of D18S51 are from other markers.

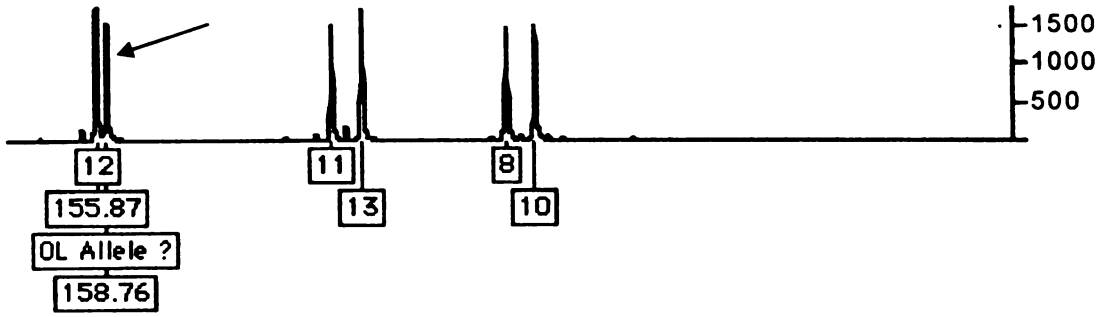


Figure 21: 12.3 Microvariant at the D5S818 Locus
As indicated by the arrow, this 12.3 microvariant at the D5S818 locus had a size of 158.76bp and a height of approximately 1500RFUs. This novel allele was 3bp larger than allele 12, which has an average size of 155.72bp when analyzed on an ABI 377. It was paired with an allele 12, which had a size of 155.87bp and an approximate height of 1700RFUs in this sample. Peaks to the right of D5S818 are from other markers.

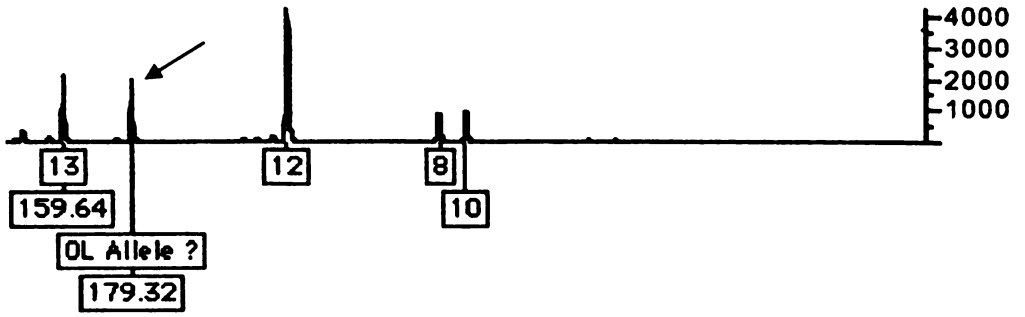


Figure 22: Above Ladder Variant Allele 18 at the D5S818 Locus
The arrow points to variant allele 18 at the D5S818 locus, which had a size of 179.32bp and a height of approximately 2500RFUs. This novel allele was 8bp larger than allele 16, which has an average size of 171.32bp when analyzed on an ABI 377. It was paired with an allele 13, which had a size of 159.64bp and an approximate height of 2500RFUs. Peaks to the right of D5S818 are from other markers.

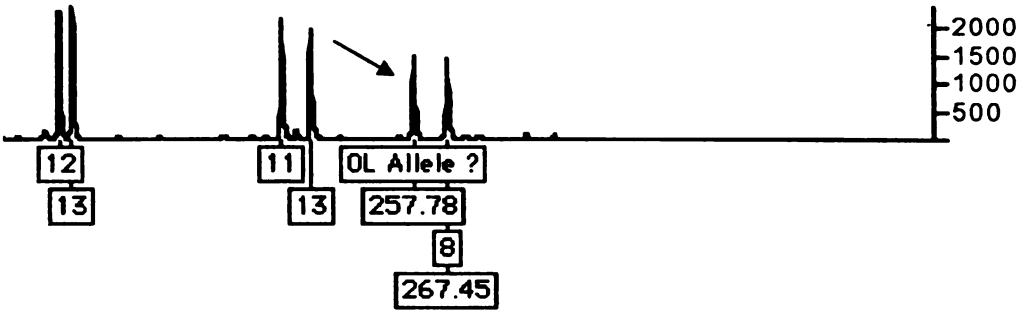


Figure 23: Below Ladder 5.2 Microvariant at the D7S820 Locus
The arrow points to the 5.2 microvariant at the D7S820 locus, which had a size of 257.78bp and a height of approximately 1500RFUs. This novel allele was 2bp smaller than allele 6, which has an average size of 259.42bp when analyzed on an ABI 377. It was paired with an allele 8, which had a size of 267.45bp and an approximate height of 1500RFUs. Peaks to the left of D7S820 are from other markers.

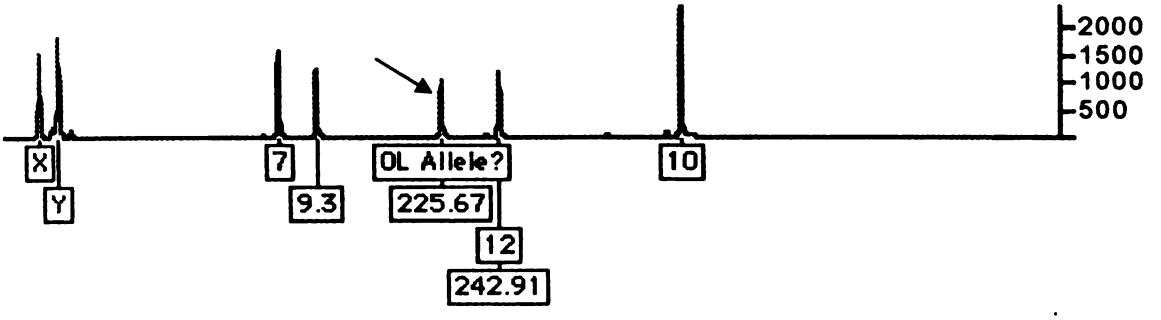


Figure 24: 7.3 Microvariant at the TPOX Locus
As indicated by the arrow, the 7.3 microvariant at the TPOX locus had a size of 225.67bp and a height of approximately 1000RFUs. This novel allele was 3bp larger than allele 7, which has an average size of 222.40bp when analyzed on an ABI 377. It was paired with an allele 12, which had a size of 242.91bp and an approximate height of 1500RFUs. Peaks to the left and right of TPOX are from other markers.

APPENDIX D

Appendix D contains the formulas used to calculate the Paternity Index (PI) values for each locus.

Table 29: Formulas for Paternity Indices

The genotypes for the Mother (M), Child (C) and Alleged Father (AF) are given using letters A-D, which represent different allele designations in a single locus. Allele frequencies for alleles A and B are represented by a and b, respectively. Single letters in the first three columns indicate homozygous loci, while double letters represent heterozygous loci. Dashes in the M Genotype column indicate the PI value was determined using only the AFs allele frequencies. Data from Traver (1998).

M Genotype	C Genotype	AF Genotype	PI Formula
BD	AB	AC	1/2a
BC	AB	AC	1/2a
BC	AB	AB	1/2a
BC	AB	A	1/a
B	AB	AC	1/2a
B	AB	AB	1/2a
B	AB	A	1/a
AB	AB	AC	1/[2(a+b)]
AB	AB	AB	1/(a+b)
AB	AB	A	1/(a+b)
AB	A	AC	1/2a
AB	A	AB	1/2a
AB	A	A	1/a
A	A	AB	1/2a
A	A	A	1/a
-	AB	AC	1/4a
-	AB	AB	(a+b)/4ab
-	AB	A	1/2a
-	A	AC	1/2a
-	A	A	1/a

BIBLIOGRAPHY

BIBLIOGRAPHY

Applied Biosystems Corporation. (1997). "ABI PRISM® GeneScan® Reference Guide: Chemistry Reference for the ABI PRISM® 377 and ABI™ 373 DNA Sequencers." Foster City, CA.

Applied Biosystems Corporation. (1998). "ABI PRISM® Genotyper® 2.5 User's Manual." Foster City, CA.

Applied Biosystems Corporation. (1997). "AmpFℓSTR® COfiler™ PCR Amplification Kit: User's Manual." Foster City, CA.

Applied Biosystems Corporation. (1997). "AmpFℓSTR® Profiler Plus™ PCR Amplification Kit: User's Manual." Foster City, CA.

Applied Biosystems Corporation. (2001). "AmpFℓSTR® SGM Plus™ PCR Amplification Kit: User's Manual." Foster City, CA.

Barber, M.D., McKeown, B.J. and Parkin, B.H. (1996). "Structural Variation in the Alleles of a Short Tandem Repeat System at the Human Alpha Fibrinogen Locus." International Journal of Legal Medicine. 108: 180 – 185.

Barber, M.D. and Parkin, B.H. (1996). "Sequence Analysis and Allelic Designation of the Two Short Tandem Repeat Loci D18S51 and D8S1179." International Journal of Legal Medicine. 109: 62 – 65.

Brinkmann, B., Sajantila, A., Goedde, H.W., Matsumoto, H., Nishi, K. and Wiegand, P. (1996). "Population Genetic Comparisons Among Eight Populations Using Allele Frequency and Sequence Data from Three Microsatellite Loci." European Journal of Human Genetics. 4: 175 – 182.

Budowle, B., Nhari, L.T., Moretti, T.R., Kanoyangwa, S.B., Masuka, E., Defenbaugh, D.A. and Smerick, J.B. (1997). "Zimbabwe Black Population Data on the Six Short Tandem Repeat Loci – CSF1PO, TPOX, TH01, D3S1358, VWA, and FGA." Forensic Science International. 90(3): 215 – 221.

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

Campbell, N.A. (1996). Biology. Menlo Park: The Benjamin/Cummings Publishing Company, Inc.

Crouse, C.A., Rogers, S., Amiott, E., Gibson, S. and Masibay, A. (1999). "Analysis and Interpretation of Short Tandem Repeat Microvariants and Three-Banded Allele Patterns Using Multiple Allele Detection Systems." Journal of Forensic Sciences. 44: 87 – 94.

Egyed, B., Furedi, S., Angyal, M., Boutrand, L., Vandenberghe, A., Woller, J. and Padar, Z. (2000). "Analysis of Eight STR Loci in Two Hungarian Populations." Forensic Science International. 113(1 – 3): 25 – 27.

Gene, M., Huguet, E., Moreno, P., Sanchez, C., Carracedo, A. and Corbella, J. (1996). "Population Study of the STRs HUMTH01 (Including a New Variant) and HUMVWA31A in Catalonia (Northeast Spain)." International Journal of Legal Medicine. 108: 318 – 320.

Gill, P., Urquhart, A., Millican, E.S., Oldroyd, N.J., Watson, S., Sparkes, R. and Kimpton, C.P. (1996). "A New Method of STR Interpretation Using Inferential Logic – Development of a Criminal Intelligence Database." International Journal of Legal Medicine. 109: 14 – 22.

Griffiths, R.A.L., Barber, M.D., Johnson, P.E., Gillbard, S.M., Haywood, M.D., Smith, C.D., Arnold, J., Burke, T., Urquhart, A. and Gill, P. (1998). "New Reference Allelic Ladders to Improve Allelic Designation in a Multiplex STR System." International Journal of Legal Medicine. 111(5): 267 – 272.

Huang, N.E., Schumm, J.W. and Budowle, B. (1995). "Chinese Population Data on Three Tetrameric Short Tandem Repeat Loci – HUMTH01, TPOX, and CSF1PO – Derived Using Multiplex PCR and Manual Typing." Forensic Science International. 71: 131 – 136.

Margolis-Nunno, H., Brenner, L., Cascardi, J. and Kobilinsky, L. (2001). "A New Allele of the Short Tandem Repeat (STR) Locus, CSF1PO." Journal of Forensic Sciences. 46: 1480 – 1483.

National Research Council. (1996). The Evaluation of Forensic DNA Evidence. The National Academy Press.

Orchid Gene Screen. (2001). "Standard Operating Procedures for Paternity Laboratory Testing." East Lansing, MI.

Shewale, J.G., Richey, S.L. and Sinha, S.K. (2000). "Detection and Correction of a Migration Anomaly on a 310 Genetic Analyzer." Journal of Forensic Sciences. 45(6): 1339 – 1342.

STRBase Internet Website. (1997). http://www.cstl.nist.gov/biotech/strbase/. NIST Biotechnology Division.

Traver, M. Appendices 5 ("Numeric Statements of the Strength of the Genetic Evidence") and 6 ("Formulas for Paternity Index and RMNE for Simple Codominant Systems"). (1998). Parentage Testing Accreditation Requirements Manual, Third Edition. American Association of Blood Banks.

General References

Bär, W., Brinkman, B., Budowle, B., Carracedo, A., Gill, P., Lincoln, P., Mayr, W. and Olaisen, B. (1997). "DNA Recommendations – Further Report of the DNA Commission of the ISFH Regarding the Use of Short Tandem Repeat Systems." International Journal of Legal Medicine. 110: 175 – 176.

Brinkmann, B., Meyer, E. and Junge, A. (1996). "Complex Mutational Events at the HumD21S11 Locus." Human Genetics. 98: 60 – 64.

Brito, R.M., Ribeiro, T., Viriato, L., Vieira-Silva, C., Espinheira, R., Pinto-Ribeiro, I. and Geada, H. (2000). "Sequence Variation of New Alleles at the Short Tandem Repeat D19S253 Locus." Journal of Forensic Sciences. 45: 932 – 934.

Budowle, B., Moretti, T.R., Baumstark, A.L., Defenbaugh, D.A. and Keys, K.M. (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: 1277 – 1286.

Espinheira, R., Geada, H., Ribeiro, T. and Reys, L. (1996). "STR Analysis – HUMTH01 and HUMFES/FPS for Forensic Application. Advances in Forensic Haemogenetics. 6: 528.

Henke, J. and Henke, L. (1999). "Mutation Rate in Human Microsatellites." American Journal of Human Genetics. 64(5): 1473 – 1474.

Junge, A. and Madea, B. (1998). "Validation Studies and Characterization of Variant Alleles at the Short Tandem Repeat Locus D12S391." International Journal of Legal Medicine. 112: 67 – 69.

Melvin, J.R., Jr., Kateley, J.R., Jr., Oaks, M.K., Simson, L.R., Jr. and Maldonado, W.E. (1998). "Paternity Testing." Forensic Science Handbook, Volume II. Ed. R. Saferstein. Englewood Cliffs: Prentice Hall.

Mizuno, N., Sekiguchi, K., Sato, H. and Kasai, K. (2003). "Variant Alleles on the Penta E Locus in the PowerPlex® 16 Kit." Journal of Forensic Sciences. 48: 1 – 4.

Mornhinweg, E., Luckenbach, C., Fimmers, R. and Ritter, H. (1998). "D3S1358: Sequence Analysis and Gene Frequency in a German Population." Forensic Science International. 95(2): 173 – 178.

Rudin, N. and Inman, K. (2002). An Introduction to Forensic DNA Analysis. Boca Raton: CRC Press.

Ruitberg, C.M., Reeder, D.J. and Butler, J.M. (2001). "STRBase: A Short Tandem Repeat DNA Database for the Human Identity Testing Community." Nucleic Acids Research. 29: 320 – 322.

Schanfield, M.S. (2000). "Parentage Testing." Encyclopedia of Forensic Sciences. 504 – 515.

Schanfield, M.S. (2000). "Polymerase Chain Reaction – Short Tandem Repeats." Encyclopedia of Forensic Sciences. 526 – 535.

Sprecher, C.J., Puers, C., Lins, A.M. and Schumm, J.W. (1996). "General Approach to Analysis of Polymorphic Short Tandem Repeat Loci." BioTechniques. 20: 266 – 276.

Urquhart, A., Oldroyd, N.J., Kimpton, C.P. and Gill, P. (1995). Highly Discriminating Heptaplex Short Tandem Repeat PCR System for Forensic Identification. BioTechniques. 18: 116 – 121.

Zhou, H.G., Sato, K., Nishimaki, Y., Fang, L. and Hasekura, H. (1997). The HumD21S11 System of Short Tandem Repeat DNA Polymorphisms in Japanese and Chinese. Forensic Science International. 86(1 – 2): 109 – 188.