
UNIVERSITY OF NORTH TEXAS
HEALTH SCIENCE CENTER AT FORT WORTH
1970

BETA TESTING AND THE POPULATION GENETICS OF PROMEGA'S PROTOTYPE POWERPLEX™ Y KIT

Ross A. Kirkendoll, B.S.

APPROVED:

Major Professor

Committee Member

Committee Member

University Member

Chair, Department of Cell Biology and Genetics

Dean, Graduate School of Biomedical Sciences

BETA TESTING AND THE POPULATION GENETICS OF PROMEGA'S PROTOTYPE POWERPLEX® Y KIT

INTERNSHIP PRACTICUM REPORT

Presented to the Graduate Council of the

Graduate School of Biomedical Sciences

University of North Texas

Health Science Center at Fort Worth

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

By

Ross A. Kirkendoll, B.S.

Fort Worth, Texas

August 2004

ACKNOWLEDGEMENTS

Promega Corporation

Ben Krenke, Ph.D.

Renee Silvis, M.S.

Xavier Aranda, M.S.

Joseph Warren, Ph.D.

John Planz, Ph.D.

Arthur Eisenberg, Ph.D.

Hriday Das, Ph.D.

TABLE OF CONTENTS

LIST OF TABLES.....v

LIST OF FIGURES.....vi

Chapter

1. INTRODUCTION.....1

2. MATERIALS AND METHODS.....10

3. RESULTS AND CONCLUSIONS.....15

REFERENCES.....25

LIST OF TABLES

Table

1. Y-STRs used in the PowerPlex™ Y Prototype.....6

2. Haplotype diversity comparisons.....15

3. Haplotypes observed more than once.....16

4. Shared haplotypes.....19

5. Mutation rates.....20

6. Mutations observed.....20

LIST OF FIGURES

Figure

1. Example of observed mutation in Caucasian population at DYS385a/b.....22

2. Example of observed mutation in Caucasian population at DYS437.....23

3. Example of mutation observed in African American population at DYS439...23

4. Example of mutation observed in African American population at DYS393...24

5. Off ladder allele at DYS438.....24

CHAPTER1

INTRODUCTION

The most commonly used genetic markers in use for forensic purposes today are microsatellites known as Short Tandem Repeats (STRs). STRs consist of short stretches of DNA which have a repeated sequence. The STRs used for human identification/forensic purposes generally contain 3-5 base pair repeat motifs with most being tetranucleotide repeats. These repeats can be categorized into three different categories. Simple repeats are repeat units of identical length and sequence, compound repeats have two or more simple repeats adjacent to each other, and complex repeats have several areas of intervening sequences along with blocks of repeats of various lengths(1).

STRs have a number of advantages over previously used forensic genetic systems including the ability to multiplex. Their small size also reduces allele dropout due to degradation of the DNA found in forensic samples. The four base pair repeats also show reduced stutter, making mixture interpretation easier (1).

Individual STRs are generally chosen according to the following criteria: high power of discrimination, linkage equilibrium, reproducibility and robustness, low amount of unwanted PCR products such as stutter (PCR product usually 1 repeat smaller and less intense then the true allele), low mutation rate, and small allele size to accommodate degraded samples found in evidentiary materials (1).

STRs specific to the Y-chromosome (Y-STRs) have several uses in forensic genetics. Since Y-STRs are passed down through paternal lineages, paternity testing can be done in cases in which the alleged father sample is missing. Instead a sample from an alleged grandfather or uncle could be used. This can also be extended into identification of human remains in a mass disaster when a reference sample for the victim can not be obtained. In some sexual assault cases, the assailants do not ejaculate or are azospermatic (do not ejaculate any sperm). When conventional autosomal testing is done, a mixture of female and male DNA may result. Due to the kinetics of the PCR reaction, the female DNA is preferentially amplified due to its greater quantity and as a result the male DNA would be amplified at a very low level making interpretation of data extremely difficult if not impossible. Y-STRs are a good solution to this problem since only the male profile is amplified during PCR (2-6).

The human Y-chromosome is approximately 60 Mb in size and is made up of two different regions, the pseudoautosomal region, and the nonrecombining region. The nonrecombining region can further be split up into euchromatin which contains the actively transcribed genes necessary for spermatogenesis and the heterochromatin which is transcriptionally inert (7-9).

The pseudoautosomal regions are a total of approximately 3Mb in size, with their composition similar to DNA found in autosomes. These regions are known to recombine with their X-chromosome counterparts during meiosis. If these regions are deleted or mutated, failure of meiosis and spermatogenesis will occur, which will lead to subsequent infertility (7,10).

The current hypothesis about the origin of the Y-chromosome, is that at one time the Y-chromosome was an autosome which at some point during evolution the Y-chromosome experienced as many as four inversions (11). Each inversion prevented recombination with the chromosome destined to become the X-chromosome. The first of these inversions is thought to have occurred approximately 240-320 million years ago and the last inversion occurred more recently during primate evolution. Of the original chromosome, only the pseudoautosomal regions are thought to remain (10-11).

Since the Y-chromosome is only inherited along a paternal lineage, and there is only one Y-chromosome, the effective population size is much smaller than an autosome. The effective population size is one quarter that of an autosome and only one third of the size of an X-chromosome. Thus the Y-chromosome is more prone to changes purely by chance (genetic drift) and is less diverse when compared to the other chromosomes (7).

The STRs used in this study are located on the Y-chromosome in the nonrecombining region (12). Because there is only one Y-chromosome, all Y-STRs are physically linked. As a result, the Y-STRs are inherited as a haplotype and have a much lower power of discrimination. Polymorphism in these microsatellites is through mutation and polymerase slippage. Mutation is thought to occur in STRs by a model known as the stepwise mutation model. In this model, a mutation occurs when the DNA polymerase slips during DNA synthesis resulting in a new allele. Most mutations are said to be one-step (one repeat unit), with addition of repeats being favored over deletions. One-step mutations are also favored over larger mutations, with larger mutations occurring only rarely (13-15). Mutations in Y-STRs have been reported to occur at a

greater rate in compound repeats. Also the greater the length of the repeat structure the more likely a mutation will occur. These findings are consistent with autosomal STRs from humans and also other species (14-15).

Knowledge of mutation rate is important for paternity testing. AABB guidelines require more than one non-matching system to exclude an individual in a parentage test making mutation rate important for analysis to ensure that the individuals tested are truly related (16). Kayser and Sajantila (17) have suggested that a minimum of three Y-STR exclusions are necessary to exclude an individual in a parentage test.

Y-STRs and other Y-chromosome systems have been used to study a number of questions in regards to human evolution and history. Scientists have considered dueling hypotheses in regards to human evolution. The multiregional hypothesis states that modern humans have evolved simultaneously from *Homo erectus* in various regions around the globe. The Out of Africa hypothesis states that modern humans evolved in Africa then migrated and replaced more archaic species around the globe (18). A number of studies have shown Y-chromosome markers to be in agreement with other genetic systems including mtDNA and autosomal data which support the Out of Africa hypothesis (7, 18).

Due to genetic linkages of the Y-STR markers and the haploid state of the Y-chromosome, the product rule (random match probabilities multiplied together) normally used in autosomal STR analysis can not be used. Instead haplotype analysis which relies on the counting method (the same method as mtDNA) with occurrence in a database used to establish point estimates. Anyone in the same paternal lineage (barring mutation) will

share the same haplotype as the person tested, therefore the ability to distinguish between two different individuals (power of discrimination) is lower than conventional autosomal STR testing (2, 5).

The PowerPlex® Y Kit used in this study is a Polymerase Chain Reaction (PCR) multiplex of Y-STRs (12). Multiplex PCR uses multiple pairs of primers which are fluorescently tagged to amplify a number of loci at one time (1). The PowerPlex® Y Kit uses 10 different primer pairs for 12 different loci (see Table 1 DYS389I/II and DYS385a/b use the same primer pairs) (12). Primer design for multiplex PCR is quite complex, in addition to amplifying a specific locus, the primers must not interact with any other primer set (1). Because the Y-chromosome has areas of homology to the X-chromosome, female artifacts may occur if the primers are poorly designed (9). The melting temperature which can be modulated with potassium and other ions must be similar for the multiplex PCR to work. Stutter (PCR product usually 1 repeat smaller and less intense then the true allele) needs to be minimized so that mixtures can be interpreted. Peak heights must also be well balanced for correct mixture analysis. Primer sequences must be constructed so that complete adenylation occurs, preventing unwanted peaks that show up as 1 base pair smaller than the true allele. A final extension step following cycling assists with the completion of adenylation (1, 12). Many of these factors have to be evaluated through trial and error since each primer pair has its own idiosyncrasies (1).

Nine of the loci (see Table 1) in the PowerPlex™ Y kit are based on the European Minimal Haplotype (12). The loci in the European Minimal Haplotype were chosen on

the basis of their ability to discriminate between individuals, well known molecular properties, known locus specific mutation rates based on studies of father/son pairs, a large body of haplotype data, and forensic developmental validation has been completed (19). Additionally the kit also contains two alleles (DYS438 and DYS439) recently required for use along with the European Minimal Haplotype in the United States by the Scientific Working Group on DNA Analysis Methods (SWGDAM) (20) and an additional locus, DYS437; added by Promega to further increase the ability of the kit to discriminating power of the panel between close lineages (12).

Table 1 Y-STRs used and their repeat structure (12).

Locus	Repeat Sequence
DYS391 *	TCTA
DYS389I/II *	TCTG TCTA
DYS439 **	GATA
DYS393 *	AGAT
DYS390 *	TCTG TCTA
DYS385a/b *	GAAA
DYS438 **	TTTTC
DYS437 ***	TCTA TCTG
DYS19 *	TAGA
DYS392 *	TAT

* = European Minimal Haplotype, ** = SWGDAM loci

*** = Locus added by Promega

Forensic validation is the process in which a technique or technology is evaluated for effectiveness and reliability. Two types of validation exist, developmental validation and internal validation.

Developmental validation is typically done by the manufacturer of the technique or technology. According to National Standards (21), the manufacturer must test for human specificity to ensure compliance with standards. In addition, the PowerPlex® Y kit must be shown to have male specificity because all of the loci are located on the Y-chromosome. Other necessary studies include mixture both male/female and male/male mixture studies, stability studies to show stability in the presence of environmental insults, and the focus of this study the construction of a population database.

Internal validation is the testing that each individual lab must do to ensure that the new technique or technology is working as expected within the laboratory (21).

These tests are done in part because forensic genetics must meet the scrutiny of the legal system. Two landmark decisions (22-23) may be used (depending on the jurisdiction) to determine the admissibility of scientific evidence. The first test is known as the Frye test (22), which is based on general acceptance. For evidence to be admitted based on this standard, the following questions must be addressed:

1. Is the theory generally accepted in the scientific community?
2. Is the method generally accepted in the scientific community?
3. Has the technique been applied correctly?

Based on these questions, the judge will decide whether or not to admit the scientific evidence.

The other legal standard which may apply is the Daubert standard (23). Instead of basing the decision to admit the evidence on general acceptance (Frye Standard) the judge answers the following questions:

1. Has the theory or technique been tested?
2. Has the theory or technique been peer reviewed?
3. What is the error rate of the method?
4. Do standards and controls exist and are they followed?
5. Is the technique or theory generally accepted in the scientific community?

Developmental and internal validation studies are completed to ensure new DNA technologies are admitted into court as well as comply with the National Standards. Standard 8.1.2.3 specifically requires population studies on all genetic systems used (21). Additionally, parameters such as mutation rates for each loci used must be determined for use in paternity testing. The American Association of Blood Banks (AABB) requires exclusions at more than one locus to determine parentage in any given parentage test (16). The mutation rate must be known in order to correctly analyze the paternity test data. Without these data, the PowerPlex® Y Kit cannot be used in a forensic DNA or paternity testing laboratory.

In order to satisfy both the requirements of the National Standards and the scrutiny of the legal system, Promega Corporation assembled a collaboration of different laboratories to assist with the developmental validation of the PowerPlex™ Y Kit. This project was a small part of that collaboration. The DNA Identity Laboratory was chosen by Promega to assist with the construction of a population database because of the number of samples available and the need for confirmed father/son pairs.

The objectives of the study were to type ~200 father/son pairs from each of the Caucasian and African American races, and then determine the haplotype frequencies, haplotype diversities, and mutation rates for each race.

CHAPTER 2

MATRIALS AND METHODS

Sample Collection

203 father/son pairs from the African American population and 199 father son/pairs from the Caucasian population were provided by the UNT Health Science Center's DNA Identity Laboratory. All personal identifying information other than population affinity was removed. Population for these individuals was self declared. Each pair tested has had paternity confirmed previously with autosomal STR analysis and was selected at random without regard to genotypes. These pairs will be used for the mutation rate study.

For the haplotype frequency and diversity studies, only the data from the fathers was used to determine these statistics. All samples provided were previously extracted with either DNA IQ™ or FTA™ paper.

Promega's DNA IQ™ system utilizes magnetic beads to remove inhibitors and other contaminants. After lysing the DNA is bound to the beads, a magnet is used to draw the beads to the bottom of the tube and the inhibitors are easily removed. The beads are subsequently washed and eluted with various solutions and buffers. The DNA is then ready for PCR amplification (24).

FTA™ paper is a cellulose based paper which contains protein denaturants, UV protection agents, buffers and a free radical trap. This allows the samples to be stored for

prolonged periods of time without special storage conditions. During addition of the biological sample (blood or saliva) to the paper, the cells are lysed and any pathogens are killed. The DNA becomes intertwined with the matrix of the paper. To prepare a sample for PCR analysis the user only needs to wash the paper with FTA™ Wash Reagent then dry. A small punch (~1.2mm) of the FTA™ paper is added to the PCR tube along with the PCR mix and amplified with reduced PCR cycles to accommodate an increased input amount of DNA (1).

Amplification

Due to differences in DNA yields, different PCR cycling parameters and amplification set up were used for Promega's DNA IQ™ and FTA™ paper. For Promega's DNA IQ™ system a small modification from the manufacturer's protocol was used(1).18.45µL of nuclease-free water, 2.5µL of GoldSTAR 10X Buffer, 2.5µL 10X Primers, 0.55µL AmpliTaq Gold® DNA polymerase (2.75U), and 1 µL of 1ng/µL DNA extract were placed into each reaction tube. For the FTA™ reactions, the setup was done as follows, 9.38µL of nuclease-free water, 1.2µL GoldSTAR 10X buffer, 1.2µL of 10X primers, 0.22µL AmpliTaq Gold® DNA polymerase, and one FTA™ punch per reaction tube. In all cases master mixes were created and used.

An ABI GeneAmp® PCR System 9700 Thermal Cycler was used for all cycling. Parameters (manufacturer's protocol see ref 10) were as follows 95°C for 11 minutes, 96°C for 1 minute, then: ramp 100% to 94°C for 30s, ramp 29% to 60°C for 30s, ramp 23% to 70°C for 45s for 10 cycles, then: ramp 100% to 90°C for 30s, ramp 29% to 58°C

for 30s, ramp 23% to 70°C for 45s for 22 cycles, then: 60°C for 30 minutes followed by a 4°C soak. The temperatures and individual cycles did not change for FTA™ samples, but the second set of cycles was done 18 times rather than 22 to account for the increase in the amount of template DNA provided by the FTA™ paper.

Electrophoresis Conditions

An ABI Prism® 3100 capillary electrophoresis system was used to analyze the DNA fragments. After PCR the resulting fragments must be separated for analysis. The PCR products are mixed with formamide and a fluorescently tagged internal lane standard DNA denatured with heating and snap cooling then loaded into the instrument. Instead of using a gel apparatus the instrument uses a capillary filled with a polymer called POP-4. The DNA fragments undergo electrophoresis through the polymer with a buffer solution similar to that used in slab gels. As in gel electrophoresis, the smaller DNA fragments move faster than the larger fragments. The fragments are detected when a laser passing through a window in the capillary hits the fluorescent tag attached to the primer. The data are displayed as electropherograms with each peak being equivalent to a band on a gel. The ABI Prism® 3100 is capable of running 96 samples in about three hours (25).

Electrophoresis was carried out on an ABI 3100 by combining 7.8μL of formamide, 1.2μL of ILS-600 internal lane standard, and 1μL of PCR product or allelic ladder. To account for differences in DNA yields, DNA IQ™ samples were injected 7s at 3kV, and FTA™ injected 7s at 2kV. Total run time was 1200s. Filter set Z was used for

all samples. All other settings were left at their default values. The raw data was analyzed with GeneScan and PowerTyper Y macro in Genotyper.

Raw Data Analysis

The GeneScan® program uses the internal lane standard to generate a standard curve which is used to determine the size of all of the DNA fragments in the sample. It also assigns peak heights based on the intensity of fluorescence detected (26). After fragment size and peak heights are determined, the Genotyper® program with the PowerTyper Y macro (provided by Promega) is used to determine the alleles present in a sample. The software compares the sizes of peaks found in the sample with those found in an allelic ladder consisting of all common alleles found in the genetic system. After comparison the computer assigns an allele designation to a particular locus (27). The data are exported into a Microsoft Excel spreadsheet and the Arlequin software package (28) for further analysis.

Statistical Analysis

Haplotype frequencies were determined by the counting method (5). Mutation rate calculations were done as suggested by Kayser et al. (14) by dividing the number of mutations at a locus by the number of meioses analyzed. The haplotype diversity is a measure of diversity of haplotypes. The values for haplotype diversities range from 0 to 1 with the higher values being more diverse. This value is important because the higher the value is the easier it is to individualize a particular haplotype. Haplotype diversity was

determined by using the formula (N/N-1) (1-Σp^2) (29). Haplotype frequencies and diversities were determined with Arlequin software package (28).

CHAPTER 3

RESULTS AND CONCLUSIONS

Haplotype studies for the Caucasians consisted of a total of 193 haplotypes typed. These data revealed a total of 169 different haplotypes. 15 haplotypes were observed more than once. These haplotypes are listed in Table 2 along with the corresponding haplotype frequency. The African American database also consisted of 93 haplotypes with 181 different haplotypes were observed in this database. Eleven haplotypes were observed more than once and are listed in Table 3.

Haplotype diversity is a measure of how diverse a particular population is. The closer the value is to 1, the more differences that will be found in a database making it more likely to narrow down an individual to a particular lineage. The haplotype diversity calculated for the Caucasian database is 0.9981. The haplotype diversity calculated for the African American database was 0.9993. These values are larger than those previously reported (0.9974 and 0.9981 respectively for the European Minimal Haplotype (19). Given the extra loci found in the PowerPlex™ Y system, these values would not be unexpected.

Table 2 Haplotype diversity comparisons.

Race	Calculated All Loci	Calculated EMH Loci	Calculated SWGDAM Loci	EMH Database	Y-Plex(SWGDAM Loci)
Caucasian	0.9981	0.9924	0.9975	0.9974	0.9946
African American	0.9993	0.9978	0.9990	0.9981	0.9991

Table 3 Haplotypes observed more than once

Race	Haplotype											n=193
	DYS391	DYS389I	DYS439	DYS389II	DYS438	DYS437	DYS19	DYS392	DYS393	DYS390	DYS385a/b	Freq.
Caucasian	11	13	12	29	12	15	14	14	13	25	11,13	0.016
	10	14	12	30	12	15	14	13	13	23	11,14	0.010
	11	13	12	29	12	15	14	13	13	24	11,14	0.021
	11	13	12	29	12	15	14	13	13	23	11,14	0.021
	10	12	11	28	10	16	14	11	13	22	13,14	0.016
	11	13	13	29	12	15	14	13	13	24	11,14	0.016
	10	13	13	29	12	15	14	13	13	24	11,14	0.010
	10	13	12	29	12	15	14	13	13	24	11,14	0.010
	11	13	11	29	12	15	14	13	13	23	11,14	0.010
	10	13	12	30	12	15	14	13	13	25	11,14	0.010
	11	13	12	29	12	15	14	13	13	24	11,15	0.010
	10	13	12	29	12	15	14	13	13	23	11,14	0.021
	10	13	11	29	12	15	14	13	13	24	11,14	0.010
	10	12	11	28	10	16	14	11	13	23	15,15	0.010
	10	12	11	29	10	16	14	11	13	22	13,14	0.010
African American	11	13	12	31	11	14	15	11	13	21	16,17	0.010
	11	13	13	29	12	15	14	13	13	24	11,14	0.010
	10	13	11	31	11	14	15	11	14	21	16,16	0.010
	11	13	11	29	12	15	14	13	13	24	11,14	0.016
	10	13	11	31	11	14	15	11	13	21	16,17	0.010
	10	13	12	30	11	14	17	11	14	21	17,18	0.010
	10	11	12	28	11	14	16	11	13	21	16,16	0.010
	10	14	12	32	11	14	15	11	13	21	16,17	0.010
	10	13	12	29	12	15	14	13	13	24	11,15	0.010
	10	13	11	30	11	13	17	11	13	21	18,18	0.010
	10	14	11	29	12	14	16	13	13	24	13,15	0.010

To make a more direct comparison to the European Minimal Haplotype, the DYS437, DYS438, and DYS439 loci data was removed from the dataset and haplotype diversities were recalculated. The respective haplotype diversities (Table 2) were recalculated. The haplotype diversity for the Caucasian database was 0.9924 and the African American haplotype diversity was 0.9978. Both of these values are less than the European Minimal Haplotype. The difference of 0.0005 for Caucasian and 0.0003 for African American databases suggest that the database constructed for the PowerPlex™ kit is very comparable to that of the European Minimal Haplotype despite being much smaller in size.

Reliagene manufactures two Y-STR kits (Y-PLEX 5 and Y-PLEX 6) which collectively contain all of the SWGDAM loci (30). To compare the new PowerPlex™ Y to the Reliagene SWGDAM database, the DYS437 locus was removed from the dataset for a direct comparison. With DYS437 removed, the haplotype diversities (Table 3) for the Caucasian and African American databases were 0.9975 and 0.9990 respectively. These values are also quite comparable to the Y-PLEX database values of 0.9946 and 0.9991 (30).

Both then European Minimal Haplotype and the Y-PLEX databases are larger (~20,000 and ~500). The addition of the DYS437 to the PowerPlex™ Y system appears to increase haplotype diversity with a smaller database. It would be interesting to see the haplotype diversities calculated from the entire collaboration and compare the to the European Minimal Haplotype and Y-PLEX databases. Unfortunately the report on the entire Promega PowerPlex™ Y database is still in press and is not currently available to the forensic genetics community.

Eight haplotypes are shared between the Caucasian and African American populations (see Table 4). A search of the Y-STR Haplotype Reference Database (19) shows that these haplotypes based on the European Minimal Haplotype are seen predominately in Europe, although a few have also been observed in Africa.

These haplotypes raise the possibility of admixture. Destro-Bisol et al. (31) used autosomal STRs and Alu to generate a distance estimate between different racial groups. Multidimensional analysis of these distances showed African Americans to be closer to Europeans than Africans. Subsequent likelihood curves provided an admixture estimate

of 0.26. Parrra et al. (32) looked at admixture among African Americans in various locales in South Carolina by using a number of different systems including autosomal, mtDNA, and Y-chromosome markers to also look for Native American influence. Estimates range from 3.5% in the coastal areas of South Carolina to 17.7% in Columbia. Chakraborty et al (33) used blood plasma protein markers taken from African Americans and Nigerians to estimate admixture. Their analysis showed and admixture of approximately 25.2% (33). Kayser et al (22) calculated a range of 27.5-33.6% using the Y-STRs found in the European Minimal Haplotype.

An admixture study using the PowerPlex™ Y data is really not possible. Such a study would require family groups (this study used convenience samples), a larger population size, autosomal STR data, and preferably mtDNA data (not available on entire data set) as well (J. Planz, personal communication). It is important to note that the population affinities for the individuals used in this study are self reported. Consequently the individuals may not know the entire extent of their heritage, thus introducing error into an admixture study and may account for observed shared haplotypes.

Mutation rates were calculated for both Caucasian and African American populations. Two mutations (see Tables 5 and 6) were observed in the Caucasian population (DYS385a/b and DYS437). Mutation rates for both of these loci were 5×10^{-3}. Six mutations (see Tables 5 and 6) were observed in the African American population with mutation rates ranging from 4.9×10^{-3}-9.85×10^{-3}.

Table 4 Shared haplotypes observed in Caucasian and African American populations.

	Haplotype											*n*=193	
Sample Numbers	DYS391	DYS389I	DYS439	DYS389II	DYS438	DYS437	DYS19	DYS392	DYS393	DYS390	DYS385a/b	Cau.Freq	A.A. Freq
CF120, B179, B995, BF19	11	13	11	29	12	15	14	13	13	24	11,14	0.0052	0.016
CF71, CF141, CF153, CF172, BF7	11	13	12	29	12	15	14	13	13	24	11,14	0.021	0.0052
C87, CF156, C193, B107, BF111	11	13	13	29	12	15	14	12	12	24	11,14	0.021	0.0052
CF144, CF159, CF194, BF62	10	12	11	28	10	16	14	11	13	22	13,14	0.016	0.0052
CF167, CF182, BF179	10	13	12	29	12	15	14	13	13	24	11,13	0.01	0.0052
CF49, BF24	10	13	13	29	12	15	14	13	13	24	11,14	0.0052	0.0052
CF61, BF189	10	14	11	32	10	14	15	12	14	23	15,15	0.0052	0.0052
CF80, BF119	11	13	12	29	12	14	14	13	13	23	11,14	0.0052	0.0052

All mutations were confirmed by reamplifying with the PowerPlex® Y system to confirm the result as well as retesting with an autosomal system (usually Profiler Plus ID) to confirm paternity and ensure that no sample switch occurred. The mutation rates for both the Caucasian and African American are higher than those calculated by Kayser et al (14) and the Y-STR Haplotype Reference Database ($2.01 \times 10^{-3} - 3.96 \times 10^{-3}$) (19). The size of this study is much smaller when compared to those referenced previously, but given the calculated haplotype diversities calculated for this study being comparable to larger studies published, sampling error is probably not an issue.

All of the mutations observed are consistent with the stepwise mutation model. Both of the mutations observed in the Caucasian population are one repeat unit larger(see Figures1-2 and Table 5). Four of the mutations observed in the African American population are one repeat unit larger, while the other two are one repeat unit smaller (see Figure 3-4 and Table 5).

Table 5 Mutation rates Found in the Caucasian and African American populations.

Race	Locus	Number of Mutations	Mutation Rate
Caucasian	DYS385a/b	1	0.00503
	DYS437	1	0.00503
African American	DYS19	2	0.00985
	DYS385a/b	1	0.00493
	DYS393	1	0.00493
	DYS437	1	0.00493
	DYS439	1	0.00493

Table 6 Mutations observed in the Caucasian and African American populations. Mutations are in bold.

Race	Sample ID	Haploptype										
		DYS391	DYS389I	DYS439	DYS389II	DYS438	DYS437	DYS19	DYS392	DYS393	DYS390	DYS385a/b
Caucasian	CF55-C	11	13	13	29	12	15	14	13	13	24	**12,15**
	CF55-F	11	13	13	29	12	15	14	13	13	24	**12,14**
	CF71-C	11	13	12	29	12	**16**	14	13	13	24	11,14
	CF71-F	11	13	12	29	12	**15**	14	13	13	24	11,14
African American	B71C	10	14	11	29	12	14	**17**	13	13	24	13,15
	B71F	10	14	11	29	12	14	**16**	13	13	24	13,15
	BF82-C	10	13	11	30	11	13	**16**	11	13	21	18,18
	BF82-F	10	13	11	30	11	13	**17**	11	13	21	18,18
	BF93-C	10	13	12	30	11	14	16	11	**15**	21	16,18
	BF93-F	10	13	12	30	11	14	16	11	**14**	21	16,18
	BF41-C	11	13	11	28	12	**14**	14	12	12	24	11,15
	BF41-F	11	13	11	28	12	**15**	14	12	12	24	11,15
	BF25-C	10	13	**12**	30	11	14	15	11	14	21	15,15
	BF25-F	10	13	**13**	30	11	14	15	11	14	21	15,15
	BF106-C	10	13	13	30	11	14	15	11	13	21	**15,17**
	BF106-F	10	13	13	30	11	14	15	11	13	21	**14,17**

One off ladder allele was observed at the DYS438 locus in a father/son pair (CF199). Genotyper and GeneScan size this peak at ~127bp (see Figure 5). The largest DYS438 allele on the PowerPlex® Y ladder is a 12 with a corresponding size of ~117bp. DYS438 is a five base pair repeat STR, thus the peak found at ~127bp can be characterized as a 14. Promega will need to look at all of the population data from the other collaborators on the project to determine the significance of this. If this allele is observed a statistically significant number of times, Promega may want to consider adding additional alleles to the allelic ladder. Alternatively, Promega could make

adjustments in their PowerTyper Y macro by setting new bins to account for this off ladder allele.

Is the PowerPlex® Y Prototype ready for forensic use? A number of other studies in addition to population data must be completed. Sensitivity, specificity, mixture studies, and stability studies must be done to complete the forensic validation of this kit. To the best of my knowledge, these studies have been completed by the collaborative group assembled by Promega, however the results of these studies are still in press and not readily available to the forensic genetics community at large. Promega will want to also perform a concordance study against Reliagene's Y-Plex™ kits to determine concordance (alleles should be the same at the same locus). I believe these experiments have been performed by Promega's group, but I am unaware of the status of these studies or any resulting publications.

In order for the PowerPlex™ Y kit to be used and admitted into court, it must pass either the Frye or Daubert standards previously discussed. Given the current state of the PowerPlex™ Y Prototype, admissibility of the kit is an open question. Without these studies published, a judge might not be inclined to admit the system into court since it would fail the conditions of both Frye and Daubert standards. Ultimately, Promega should be able to get their PowerPlex™ Y kit admitted into court as a result of their collaboration.

The population database, it is quite comparable to that of previously published works, particularly the haplotype diversities. Mutation rates calculated for this study were higher than those previously reported, which is probably not due to sampling error given

the haplotype diversities. Eight haplotypes were shared between Caucasians and African Americans. Admixture may be present, but because convenience samples were used and also the study size, admixture studies were not possible.

Figure 1 Example of Observed Mutation in Caucasian population at the DYS385a/b locus. The son's electropherograms is above the father's electropherogram.

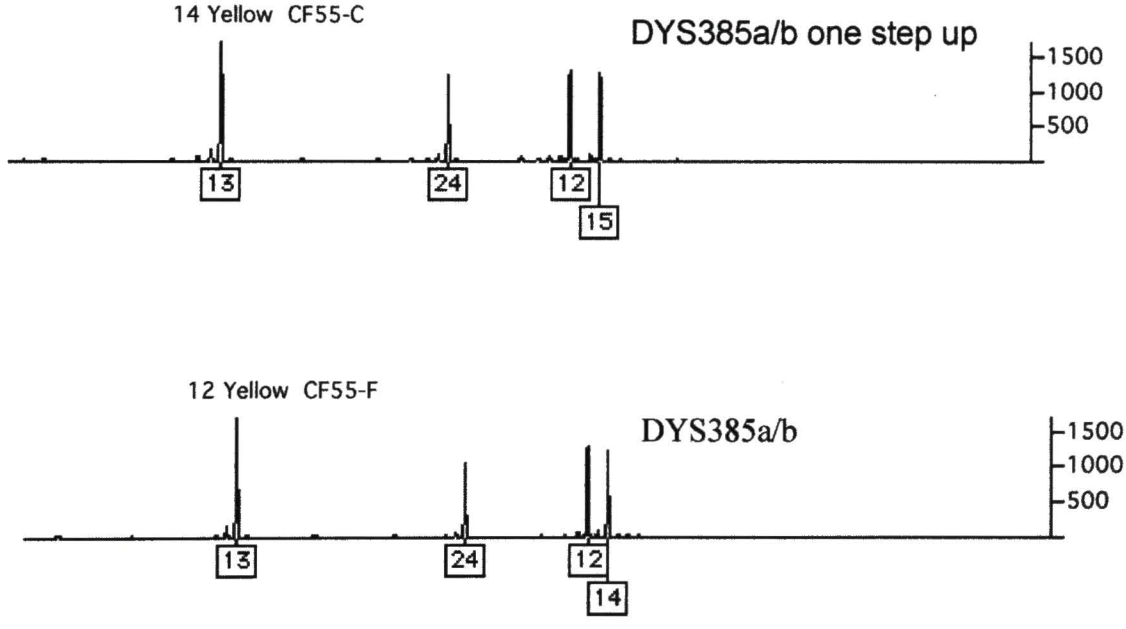


Figure 2 Example of mutation observed in Caucasian population at the DYS437 locus. The son's electropherogram is shown above the father's electropherogram.

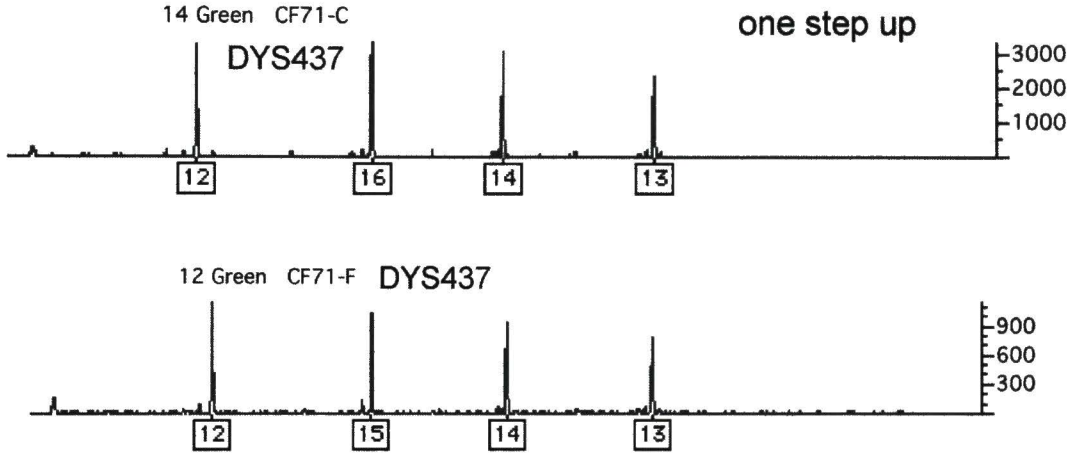


Figure 3 Example of mutation observed in the African American population at the locus DYS439. The son's electropherograms is displayed above the father's electropherogram.

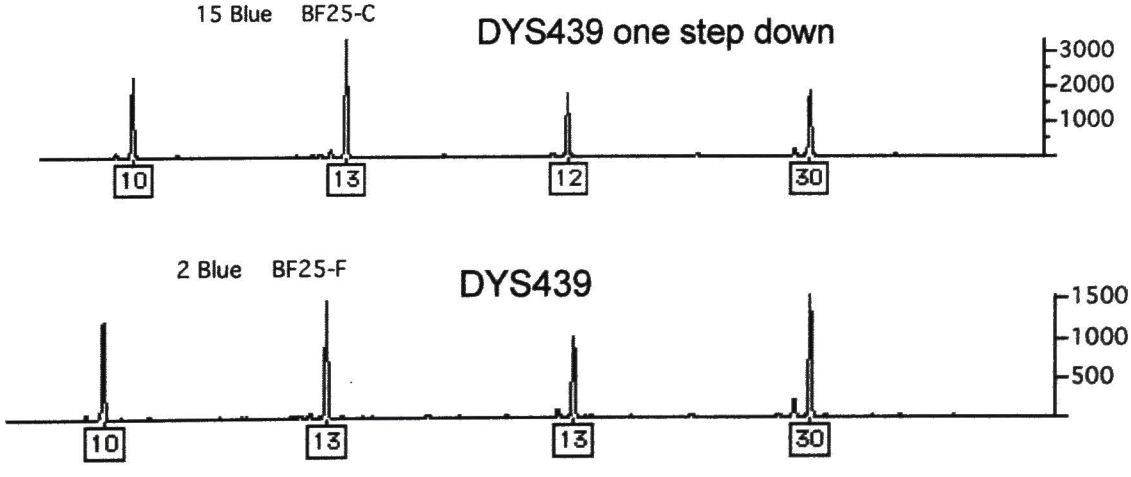


Figure 4 Example of mutation observed in the African American population at the locus DYS393. The son's electropherogram is above the father's electropherogram.

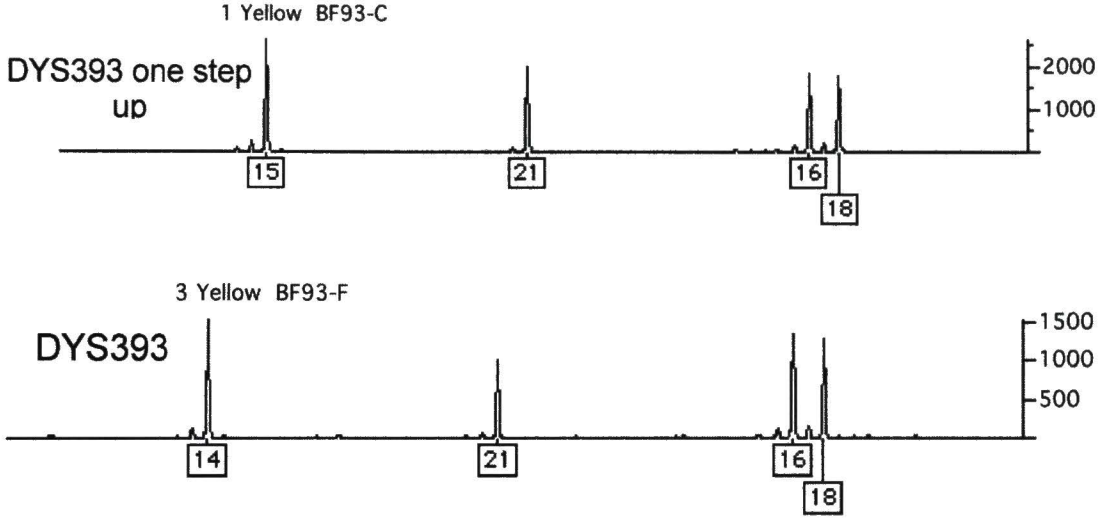
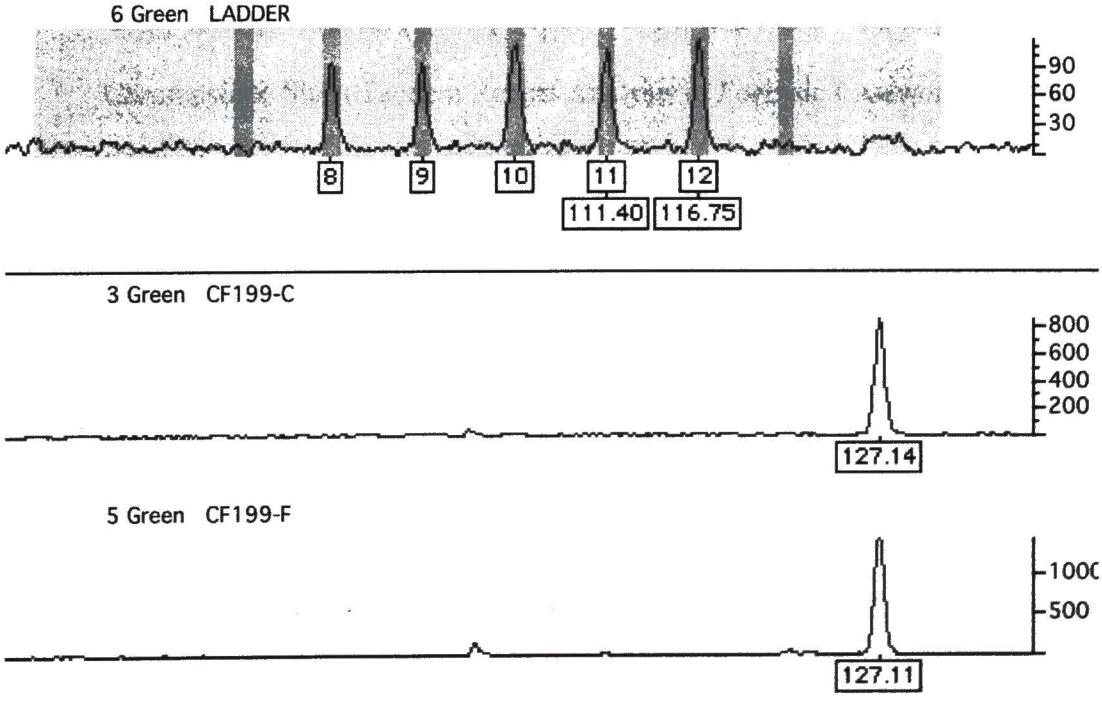


Figure 5 Off ladder allele found at DYS438.



REFERENCES

1. Butler JM. Forensic DNA Typing. Academic Press, New York 2001.
2. Jobling MA, Pandya A, Tyler-Simth C. The Y Chromosome in Forensic Analysis and Paternity Testing. International Journal of Legal Medicine 1997; 110:118-124.
3. Betz A, Bäbler G, Dietl G, Steil X, Weyermann G, Pflug W. DYS STR Analysis with the Epithelial Cells in a Rape Case. Forensic Science International 2001; 118:126-130.
4. Parson W, Niederstatter H, Kochl S, Steinlechner M, Berger B. When Autosomal Short Tandem Repeats Fail: Optimized Primer and Reaction Design for the Y-Chromosome Short Tandem Repeat Analysis in Forensic Casework. Croation Medical Journal 2001; 42(3)285-287.
5. Gill P, Brenner C, Brinkmann B, Budowle B, Carracedo A, Jobling MA, et al. DNA Commission of the International Society of Forensic Genetics: Recommendations on Forensic Analysis Using Y-Chromosome STRs. Forensic Science International 2001; 124:5-10.
6. Cerri N, Ricci U, Verzeletti A, De Ferrari F. Mixed Stains from Sexual Assault Cases: Autosomal or Y-Chromosome Short Tandem Repeats? Croation Medical Journal 2003; 44(3):289-292.

7. Jobling MA, Tyler-Smith C. The Human Y Chromosome: An Evolutionary Marker Comes of Age. Nature Reviews Genetics 2003; 4:598-612.
8. Lahn BT, Pearson NM, Jegalian K. The Human Y Chromosome, in the Light of Human Evolution. Nature Reviews Genetics 2001; 2:207-216.
9. Hall A, Ballantyne J. The Development of an 18-locus Y-STR System for Forensic Casework. Annals of Bioanalytical Chemisty 2003; 376:1234-1246.
10. Graves JAM. Mammals that Break the Rules: Genetics of Marsupials and Monotremes. Annual Reviews in Genetics 1996; 30:233-260.
11. Lahn BT, Page C. Four Evolutionary Strata on the Human X Chromosome. Science 1999; 286:964-967.
12. Promega Corporation. PowerPlex™ Y System Prototype Part #GE066. Promega Corporation, Madison, 2002.
13. Schötterer C, Tautz D. Slippage Synthesis of Simple Sequence DNA. Nucleic Acids Research 1992; 20(2):211-215.
14. Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, et al. Characteristics and Frequency of Germline Mutations at Microsatellite Loci from the Human Y Chromosome, as Revealed by Direct Observation in Father/Son Pairs. American Journal of Human Genetics 2000; 66:1580-1588.
15. Brinkmann B, Klintschar M, Neuhuber F, Hühne J, Rolf B. Mutation Rate in Human Microsatellites: Influence of the Structure and Length of the Tandem Repeat. American Journal of Human Genetics 1998; 62:1408-1415.

16. American Association of Blood Banks, Accreditation Information Manual 4th edition; 2001.

17. Kayser M, Sajantila A. Mutations at Y-STR Loci: Implications for Paternity Testing and Forensic Analysis. Forensic Science International 2001; 118:116-121

18. Seielstad M, Bekele E, Ibrahim M, Toure A, Traore M. A View of Modern Human Origins from the Y Chromosome Microsatellite Variation. Genome Research 1999; 9:558-567.

19. Roewer L, Krawczak M, Willuwiet S, Nagy M, Alves C, Anslinger K, et al. Online Refernece Database of European Y-Chromosomal Short Tandem Repeat (STR) Haplotypes. Forensic Science International 2001; 118(2-3):106-113. http://www.ystr.org

20. Scientific Working Group on DNA Analysis Methods. Report on the Current Activities of the Scientific Working Group on DNA Analysis Methods Y-STR Subcommittee. Forensic Science Communications. 2004; 6:3.

21. Federal Bureau of Investigation. FBI DNA Quality Assurance Document. Federal Bureau of Investigation, Washington, 2000.

22. *Frye v. United States* (D.C. Circuit 1923) 293 F.10

23. *Daubert v. Merell-Dow Pharmaceuticals, Inc.* (1993) 509 U.S. 579.

24. Promega Corporation. DNA IQ™ System Small Sample Casework Protocol Technical Bulletin No. 296. Promega, Madison, 2002.

25. Applied Biosystems. ABI Prism® 3100 Genetic Analyzer User Guide. Applied Biosystems, Foster City, 2001.

26. Applied Biosystems. ABI Prism® GeneScan® Analysis Software User Manual. Applied Biosystems, Foster City, 2001.

27. Applied Biosystems. ABI Prism® Genotyper® Software User Manual. Applied Biosystems, Foster City, 2001.

28. Schneider S, Roessli D, Excoffer L. Arlequin: A Software for Population Genetics Data Analysis. Version 2.00. Genetics and Biometry Lab, Department of Anthropology, University of Geneva, 2000.

29. Hedrick PW. Genetics of Populations. Jones and Bartlett, Sudbury, 2000.

30. Sinha S, Budowle B, Chakraborty R, Paunovic A, DeVille Guidry R, Larsen C, et al. Utility of the Y-STR Typing Systems Y-PLEX™ 6 and Y-PLEX™ 5 in Forensic Casework and 11 Y-STR Haplotype Database for the Three Major Population Groups in the United States. Journal of Forensic Science 2004; 49(4):691-700.

31. Destro-Bisol G, Maviglia R, Caglià A, Boschi I, Spedini G, Pascali V, et al. Estimating European Admixture in African Americans by Using Microsatellites and a Microsatellite Haplotype (CD4/Alu). Human Genetics. Human Genetics 1999; 104:149-157.

32. Parra EJ, Kittles RA, Argyropolous G, Pfaff CL, Hiester K, Bonilla C, et al. Ancestral Proportions and Admixture Dynamics in Geographically Defined African Americans Living in South Carolina. American Journal of Physical Anthropology 2001; 114:18-29.

33. Chakraborty R, Kamboh MI, Nwanko M, Ferrell, R. Caucasian Genes in American Blacks: New Data. American Journal of Human Genetics 1992; 50:145-155.

34. Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, et al. Estimating African American Admixture Proportions by Use of Population-Specific Alleles. American Journal of Human Genetics 1998; 63:1839-1851.