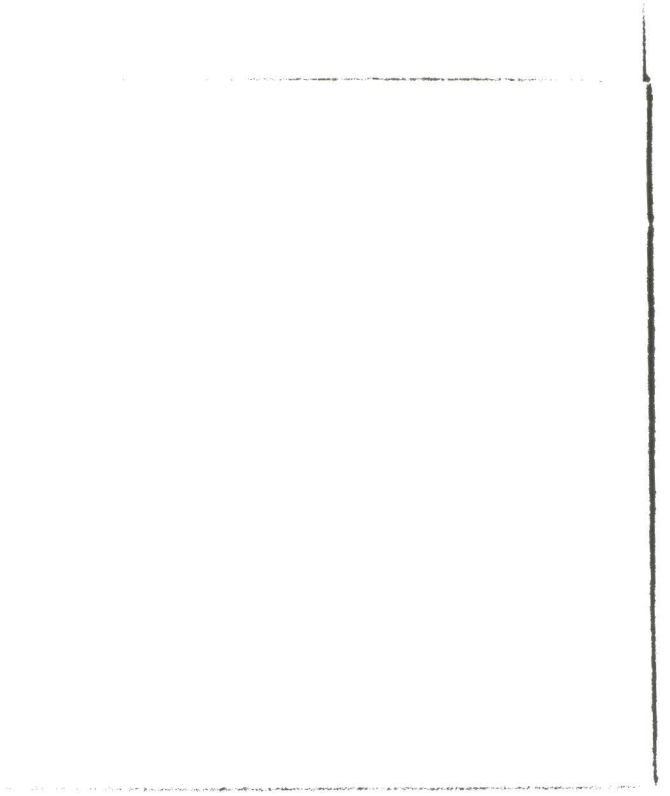
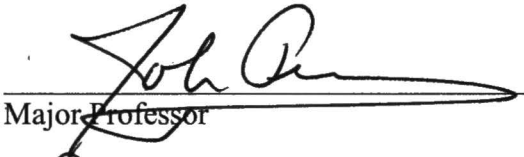

UNIVERSITY OF NORTH TEXAS
HEALTH SCIENCE CENTER AT FORT WORTH
1970



HAPLOTYPE FREQUENCIES FOR PROMEGA'S POWERPLEX®Y SYSTEM FOR SOUTHWESTERN HISPANIC AND ASIAN POPULATIONS

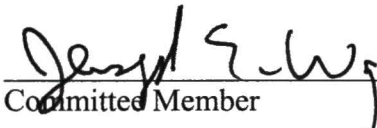
Renee Silvis, B.S.

APPROVED:



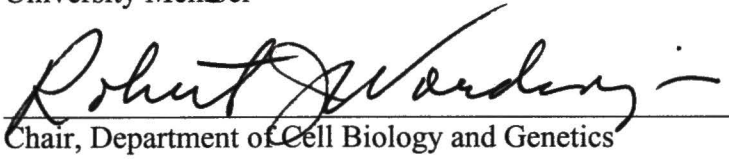
Major Professor

Committee Member

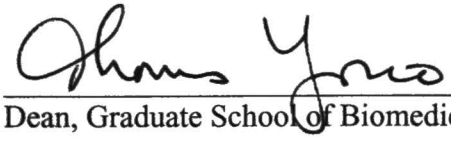


Committee Member

University Member



Chair, Department of Cell Biology and Genetics



Dean, Graduate School of Biomedical Sciences

HAPLOTYPE FREQUENCIES FOR PROMEGA"S POWERPLEX®Y SYSTEM FOR SOUTHWESTERN HISPANIC AND ASIAN POPULATIONS

THESIS

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

Renee Silvis, B.S.

Fort Worth, Texas

May 2004

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

Figure Page

INTRODUCTION:

The most common method of analysis for DNA in forensic and paternity labs is by using Short Tandem Repeat (STR) markers. These markers are more efficient at detecting degraded DNA templates than previously used typing methods, such as Restriction Fragment Length Polymorphisms (RFLP) due to the fact that the amplification products are less than 500bp long, (Saldanha,et.al., 2002).

For human identification purposes, the STRs used are from thirteen independent loci located on non-sex determining chromosomes, or autosomal DNA. The thirteen loci were recommended for forensic use by the Scientific Working Group for DNA Analysis Methods (SWGDAM), and were shown to have a combined average match probability rarer than one in a trillion among unrelated individuals (Butler, 2003). The autosomal core loci are: CSF1PO, FGA, THO1, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, and D21S11. The majority of the STR loci are composed of tetranulceotide repeat regions. These loci are made up of polymorphic markers that when combined, allows the analyst the ability to discriminate between samples (Butler, 2003).

The ability to discriminate between samples is possible, since, every individual has their own combination of alleles at these thirteen loci, because these autosomal loci interrogate DNA that is inherited from both parents in a Mendelian fashion. At each locus there will be an allele randomly donated from the father and an allele randomly donated from the mother. Even though a child should have some combination of the parents' DNA, mutation occasionally does occur. Normally, mutations consist of either a

base pair change, or variation in repeat length. The average rate of mutations for autosomal STRs occurs between 1-5 times in 1000 generational events (Butler, 2003).

In addition to the thirteen core loci, many commercially available kits also include an amelogenin marker for gender identification. This marker is used to tell what the sex of the contributor is, also it is helpful in determining if multiple DNA profiles were analyzed in a sample.

Y-chromosome loci have recently been of particular interest to forensic scientists. Unlike females that have two X-chromosomes, males have an X and a Y-chromosomes. The X-chromosome is inherited from the mother and the Y-chromosome is inherited from the father. The Y-chromosome is the only nonhomologous chromosome in the human genome, and it is one of the smallest human chromosomes (Gusmao and Carracedo, 2003). Half of the Y-chromosome is made up of tandemly repeated SATELLITE DNA the other half, most of which doesn't recombine, carries a few genes. Since Y- chromosomal loci do not change by recombination, mutation is the only way polymorphisms are developed. Mutations on the Y-chromosome were thought to be rare, since nucleotide base substitutions occurred at such a low rate that it was difficult to accurately quantify without very large data sets. With the availability of Y-chromosome sequence many Y-specific single nucleotide polymorphisms (SNP) have been identified along with repeat sequences (Jobling and Smith, 2003). These mutations may sometimes be lost by random drift genetic change, so demography and population structure will have large influences on the rate. It is known that patrilocality is practiced in about 70% of modern societies (Jobling and Smith, 2003). This means that most males live close to

their birthplace; so local differentiation of Y-chromosomes is enhanced. Figure 1 shows global distributions of Y haplogroups. The distribution of the haplogroups reflects possible migration patterns of their predecessors and that there are more similarities between haplotypes that are geographically closer then between haplotypes that are geographically more distant.

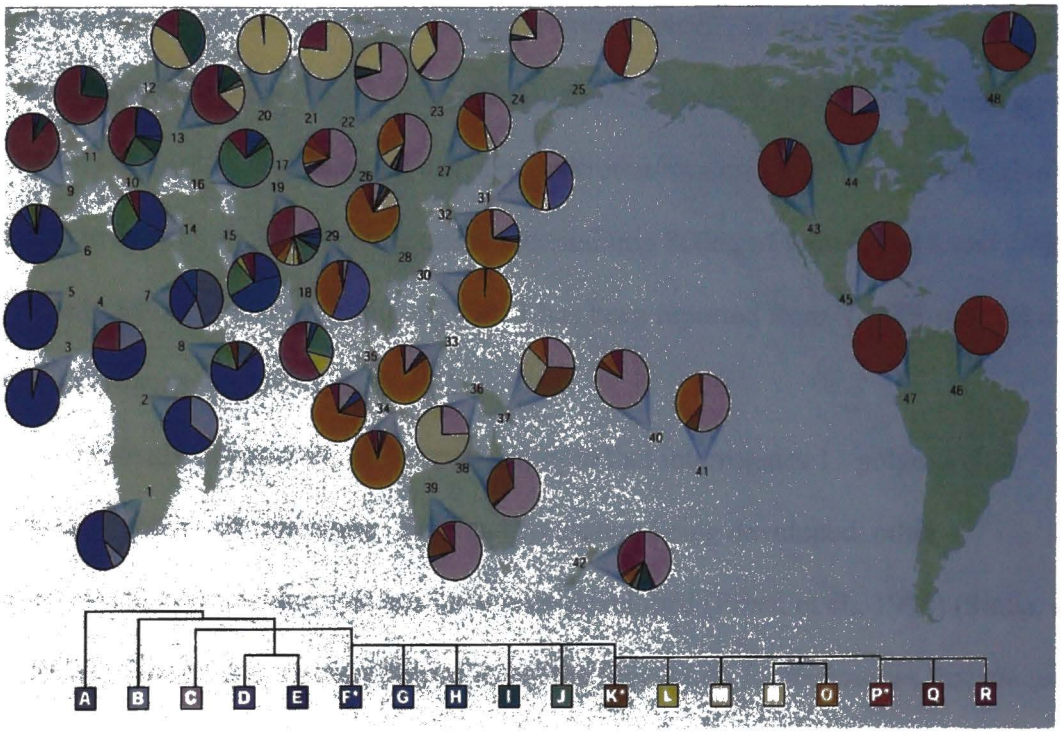


Figure 1 Global distribution of Y haplogroups. Each circle represents a population sample with the frequency of the 18 main Y haplogroups. Indicated by the colored sectors. Note the general similarities between neighboring populations but large differences between different parts of the world. (from Jobling and Smith. The Human Y Chromosome: an Evolutionary Marker Comes of Age. Genetics. 2003; 4:598-612)

Recently manufacturers have produced kits that are made up of STRs markers on the non-recombining region of the Y-chromosome (Y-STRs). These Y-STRs produce a haploid profile when male DNA is amplified. Like autosomal STRs, Y-STRs are

analyzed in a similar fashion and have a high level of diversity (Gusmao and Carracedo, 2003).

Prior to the introduction of Y-STRs, forensic scientists had a much more difficult time separating male DNA from female DNA in cases of a multi-contributor mixture. With autosomal STRs, mixed contributor samples are very difficult to interpret, since, it hard to differentiate between the minor contributor's profile from that of the stutter peaks of the major contributor (Sinha, et al., 2003). In such cases, Y-STR markers may prove to be very helpful by interrogating only the male contribution. Y-STRs are sensitive enough to identify male DNA in a mixture containing 800ng of female DNA and 400pg of male in DNA. A mixture ratio of 2000: 1 has been reported from Y-STR validation studies (Prinz, et al., 1996).

Promega Corporation developed a system that interrogates 12 polymorphic markers on the Y- chromosome. Prior to this system being developed, other Y-chromosomal multiplexing systems have been developed (Prinz, et al., 1996) (Sinha, et al., 2003). These other systems helped to lay the foundation for techniques that are used to observe Y-chromosomal DNA . The work on Y-chromosomal DNA done in Europe helped to develop the European Y Haplotype Reference Database (YHRD) (Willuweit and Roewer, 2004). This database was designed to "identify polymorphisms capable of discriminating between majority of unrelated lineages, establish a database representative of the geographical and ethnical structure, and to aim at a database size that would allow accurate frequency estimation for rare haplotypes" (Roewer, et al., 2001). The YHRD was started in 1994 to help analyze Y-STR haplotypes with 31 different forensic and

anthropological institutions contributing data. As of February of 2003 it contained 13,986 haplotypes and it has continued to grow becoming, the largest Y-STR database in the world. In Europe, nine loci are required for comparison (European Minimal Haplotype loci): DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, and DYS393. The markers incorporated into Promega's PowerPlex®Y System are DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, and DYS439. This system includes the nine European Minimal Haplotype loci plus two loci that were recommended by SWGDAM, DYS438 and DYS439. An extra locus, DYS437, was added to the PowerPlex®Y System to further increase the discriminating ability of this system.

The markers for PowerPlex®Y System are all located on the same chromosome and are very close together, so they are genetically linked. Figure 2 gives approximate locations for these loci.

Unlike autosomal STRs, where each allele is inherited independently, Y-STR alleles are inherited together. This pattern of inheritance is similar to how mitochondria are inherited along the mother's lineage (Alberts, et al., 2002). The linkage results in the ability of Y-STRs to discriminate between males in different paternal lineages but not between males that share a paternal lineage.

The amplified products for these markers are between 250 – 300 base pairs in length, and are mostly made up of tetrameric repeats except for one locus, DYS438, that has a pentanucleotide repeat motif (Saldanha, et al., 2002). Table 1 gives the locus specific information for the PowerPlex®Y System.

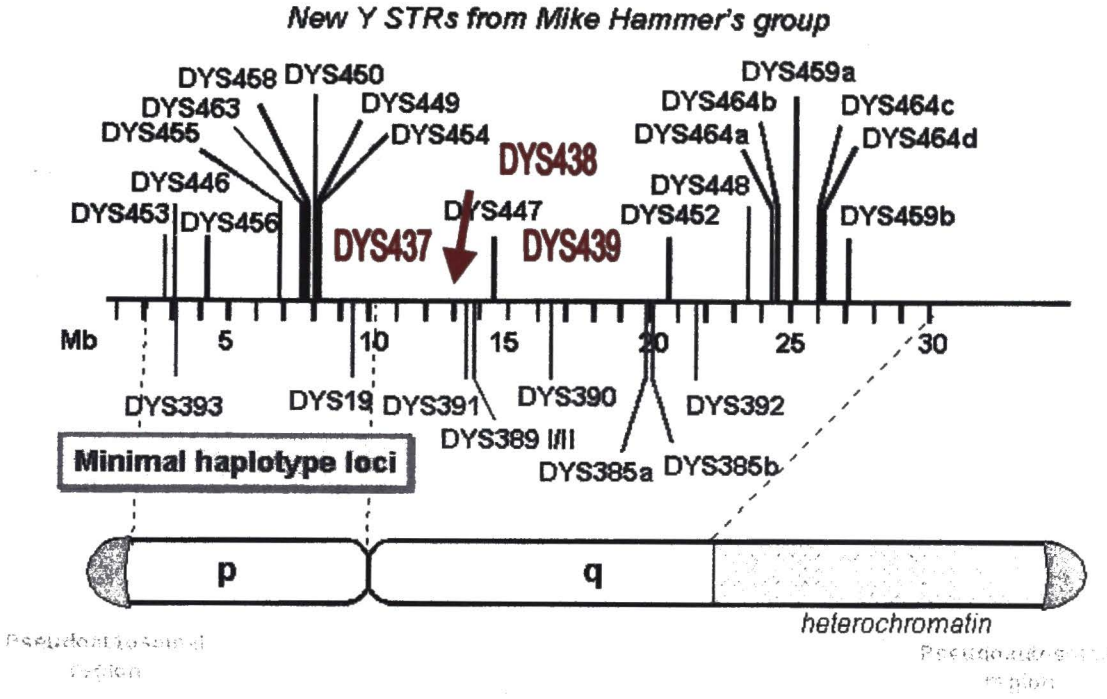


Figure 2: Y STR Positions along Y Chromosome: DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, and DYS393 (European Minimal Haplotype loci) are shown below the scale, the red arrow show the approximate locations of the three additional loci (DYS437, DYS438, and DYS439) included in the PowerPlexY®System. (adapted from http://www.cstl.nist.gov/div831/strbase/ystpos1.htm).

This project had tow objectives. The primary one was to establish the haplotype frequency data for the Asian and Southwestern Hispanic populations. This involved developing Y-STR profiles for the entire sample set, establishing haplotype frequencies from that data's allele frequencies for each of the loci interrogated. This information was then used to calculated the Power of Discrimination for the PowerPlex®Y System. The secondary objective was to calculate a mutation rate for this particular Y-STR panel by

analyzing established father – son pairs (each pair being one meiosis) and counting the number of differences between the pairs, noting which loci the mutations occurred in.

STR Locus	Label	Chromosomal Location	GenBank® Accession Number	Repeat Sequence[1] 5′→3′
DYS391	FL	Yq	G09613	TCTA (19)
DYS389I/II	FL	Yq	AF140635	[TCTG][TCTA] Complex (19)
DYS439	FL	Yq	AC002992	GATA (20)
DYS393	TMR	Yp	G09601	AGAT (19)
DYS390	TMR	Yq	AC011289	[TCTG][TCTA] Complex (19)
DYS385a/b	TMR	Yq	Z93950	GAAA (19)
DYS438	JOE	Yq	AC002531	TTTTC (20)
DYS437	JOE	Yq	AC002992	[TCTA][TCTG] Complex (20)
DYS19	JOE	Yp	X77751	TAGA Complex (19)
DYS392	JOE	Yq	G09867	TAT (19)

[1]The August 1997 report (21,22) of the DNA Commission of the International Society for Forensic Haemogenetics (ISFH) states, "1) for STR loci within coding genes, the coding strand shall be used and the repeat sequence motif defined using the first possible 5′ nucleotide of a repeat motif; and 2) for STR loci not associated with a coding gene, the first database entry or original literature description shall be used."

TMR = carboxy-tetramethylrhodamine

FL = fluorescein

JOE = 6-carboxy-4′,5′-dichloro-2′,7′-dimethoxyfluorescein

Table1: PowerPlex®Y System Locus-Specific Information (From Saldanha, et al., PowerPlex®Y System Validation Study. Promega. (2002); 1-37.)

METHODS:

Beta test kits were obtained from Promega Corporation for this study. These kits were also received by six other laboratories across North America as part of a beta test and database validation study for the PowerPlexY® System. The data collected from all of the laboratories included in this study, are being assembled to create a North American Y-STR haplotype database for the Hispanic, Caucasian, Asian, African American, and Native American populations (Budowle, et al., 2003).

DNA from 207 father – son pairs with Southwestern Hispanic ethnicity and 83 father – son pairs with Asian ethnicity, were obtained from University of North Texas

Health Science Center DNA Identity Lab. The father – son pairs had been previously examined with autosomal STR's to verify their relationship.

Extraction of DNA for Population Studies

Two methods of extraction were used for this study: FTA paper and Promega's DNA IQ™ System. Although organic extraction is very useful for evidentiary samples, FTA® paper is very useful for reference samples. The chemical treatment on the FTA® paper kills blood-borne pathogens on contact, immobilizes DNA within the matrix, protects DNA from degradation, and allows the DNA to be stored long term at room temperature. FTA paper allows DNA to be immobilized in the matrix while cellular debris is being washed away (Whatman BioScience, 2000). Samples for amplification consisted of 1.2mm punches from the sample area of each card, which contain approximately 60-70ng of DNA. The punch containing the entrapped DNA directly serves as a template for the Polymerase Chain Reaction (PCR) (Budowle, et al., 2000).

DNA extraction using the DNA IQ ™ System was also used. This system contains paramagnetic resin that captures a specific amount of DNA. The DNA is released from the resin particles with an elution buffer. 100ul of elution buffer was used to give a known concentration of DNA of approximately 1ng/ul; 1ul of this was used for PCR. The DNA elutant obtained from the DNA IQ ™ System directly serves as the template for the Polymerase Chain Reaction (Promega Corporation, 2002) to amply the Y-STR loci.

STR Amplification

This kit included male specific primers that were fluorescently labeled. Each amplification reaction contained 2.5ul of Gold STR 10x Buffer; 2.5ul of PowerPlex®Y 10x Primer Pair Mix; 16.95 ul of Nuclease-Free Water; AmpliTaq Gold® DNA polymerase; and 2.5 ul of Template DNA. When using FTA paper an additional 2.5ul of Nuclease-Free Water was added to keep the total volume at 25ul. Amplification reactions were preformed in a Perkin-Elmer GeneAmp® PCR System 9700 Thermal Cycler. Cycling conditions were as follows: 95°C for 11 minutes then 96°C for 1 minute, then ramp100% to 94°C for 30 seconds then ramp 29% to 60 °C for 30 seconds, then ramp 23% to 70 °C for 45 seconds, this was done for 10 cycles. Then ramp100% to 90 °C for 30 seconds, then ramp 29% to 58 °C for 30 seconds, then ramp 23% to 70 °C for 45 seconds and this was done for 18 or 22 cycles. There are two cycles numbers since FTA paper contains far more DNA then normal amplification requires so the number of cycles was reduced to prevent too much product from being produced. A final temperature of 60 °C was then preformed for 30 minutes and it ended with a 4°C soak. A positive and negative control was amplified with every batch of amplification reactions (Saldanha, et al., 2002).

Analysis on ABI 3100 Genetic Analyzer

The samples for analysis were prepared by combining 1.0ul of Internal Lane Standard 600 (ILS600) with 24.0ul of deionized formamide. This was combined with 1.0ul of amplified product. The samples were denatured by heating them to 95 °C for 3 minutes then immediately chilling then on crushed ice for 3 minutes. The denatured

samples underwent electrophoresis on an ABI 3100 Genetic Analyzer using POP 4 polymer. The run time was approximately 45 min. Promega's PowerPlex® matrix standards, 3100 (Cat#D63380) were used (Saldanha, et al., 2002).

Data Analysis

The mutation rate was determined by evaluating father – son pairs and counting how many times the child's Y-STR profile differed from that of the father's. The number of observed meioses observed, was divided by the total number of pairs tested for each population tested.

The haplotype frequency was determined by counting how many times a Y-STR haplotype appeared in the population. Arlequin, (Schneider, et al., 2004), a genetic data analysis software package, was used to compile data from this study, determine the allele frequencies and the genetic diversity of the loci examined. Arlequin was also used to determine the allele frequencies and the genetic diversity of the loci examined. In order to compare the PowerPlexY® System to the European Minimal Haplotype loci for this study, the three extra loci that were added to the PowerPlexY® System's were subtracted from the complete PowerPlexY® System results and then reanalyzed with Arlequin. Additionally, the haplotype profiles that were observed multiple times with the PowerPlexY® System were entered into the YHRD (http://ystr.charite.de/index_gr.html) to compare the results of this study to that of the European population.

RESULTS:

Verification of Father- Son Pairs

Autosomal DNA is inherited half from the mother and half from the father. By examining the autosomal DNA of the father- son pairs it is possible to prove that the samples being tested were true father- son pairs. Figures 3 and 4 represent typical electropherograms of autosomal DNA using Promega's PowerPlex®16 System which interrogates 16 STR loci.

Figure 5 and represent typical electropherograms of Y haplotypes from father – son pairs, using the PowerPlex®Y System for Southwestern Hispanic and Asian populations, respectively.

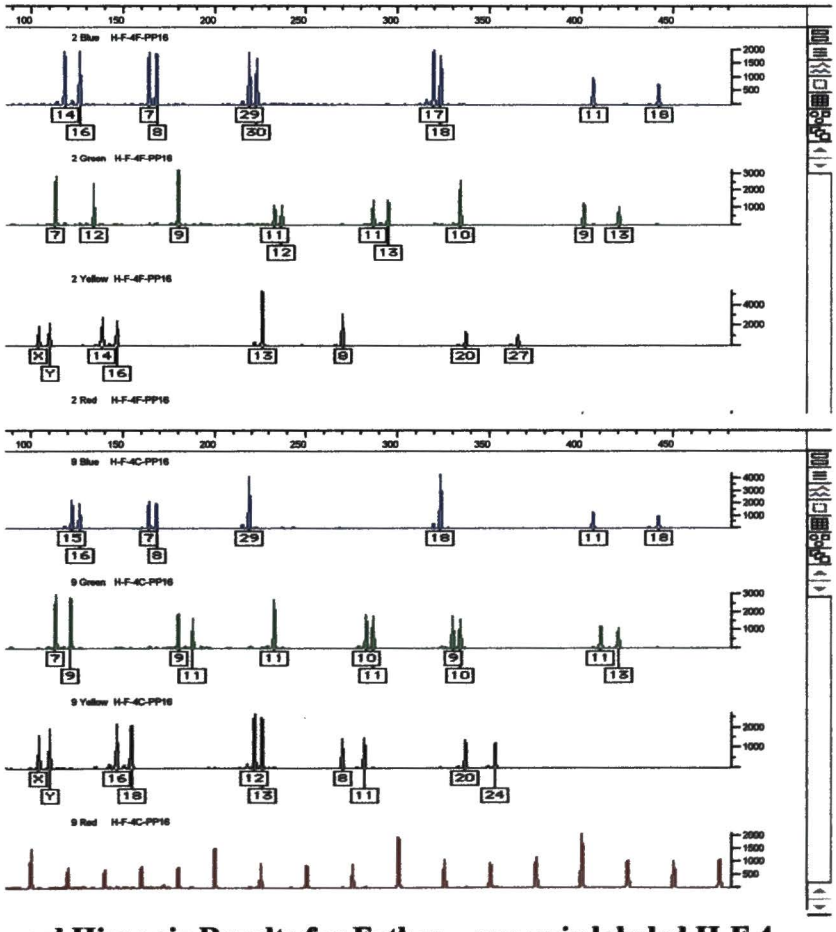


Figure 3: Autosomal Hispanic Results for Father – son pair labeled H-F 4. Data is arranged so that the father is seen first then the child

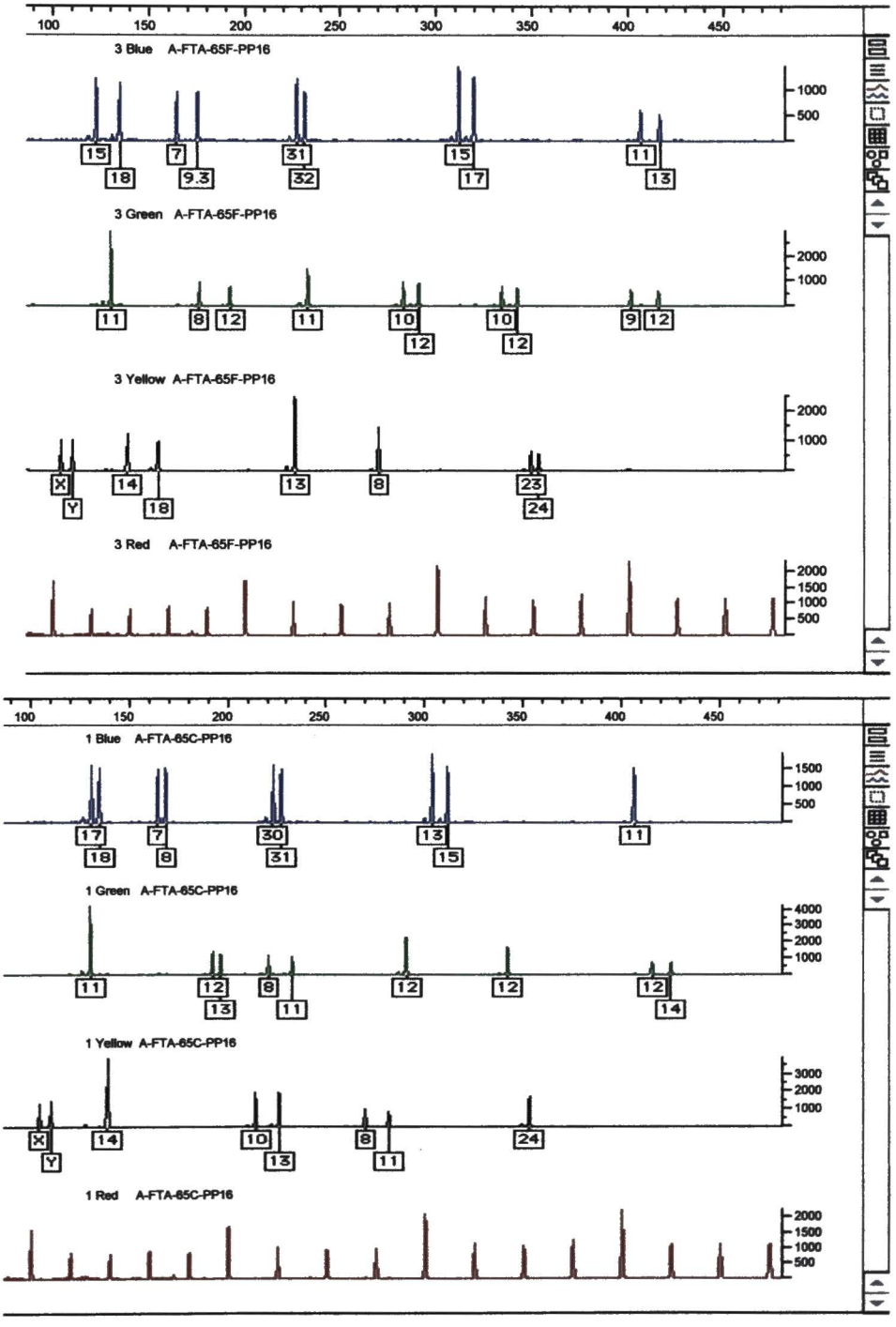


Figure 4: Autosomal Asian Results for Father – son pair labeled A-FTA-65. Data is arranged so that the father is seen first then the child

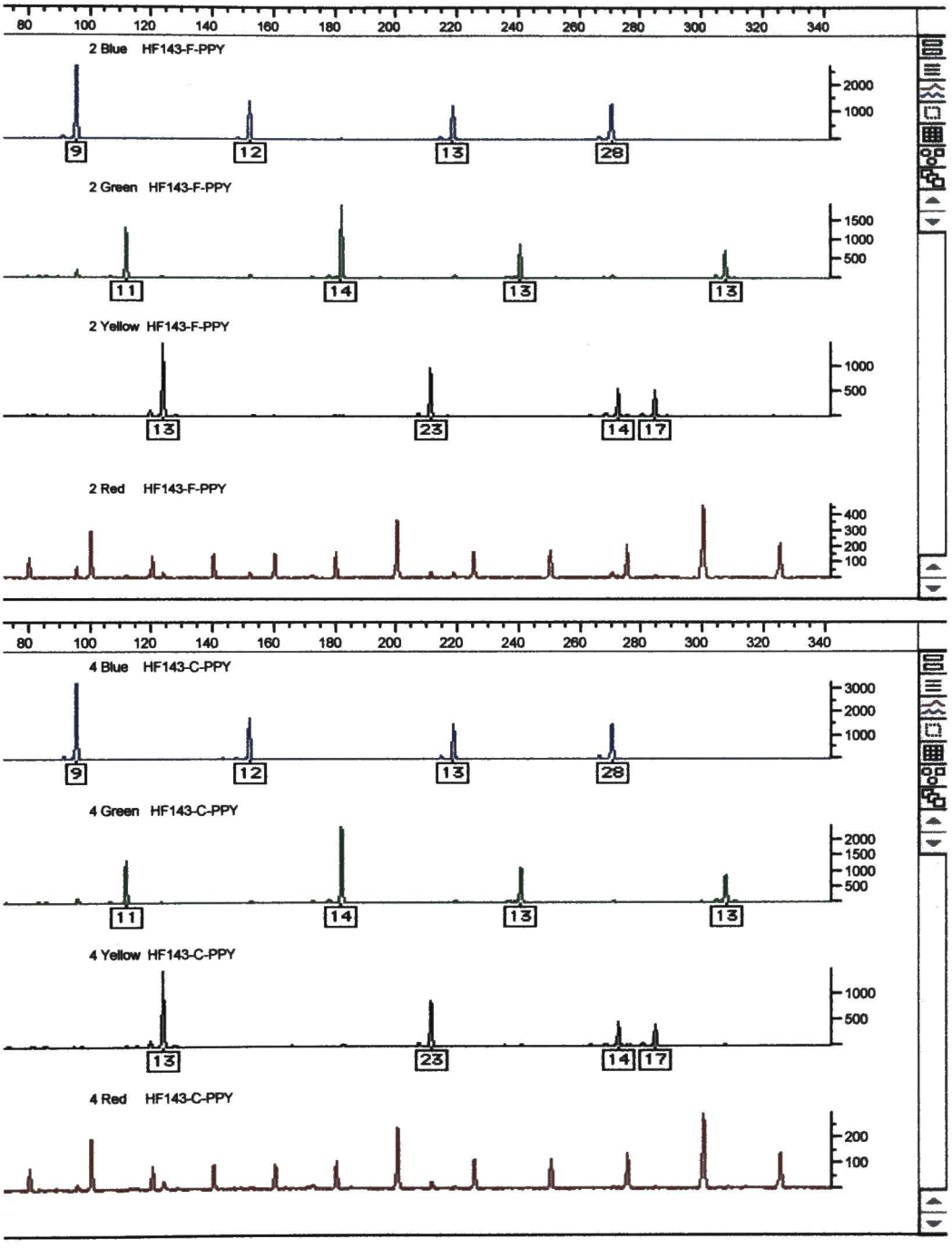


Figure 5: Typical father – son pair results in the Hispanic population, sample set used was HF 143. Data is arranged so that the father is seen first then the child

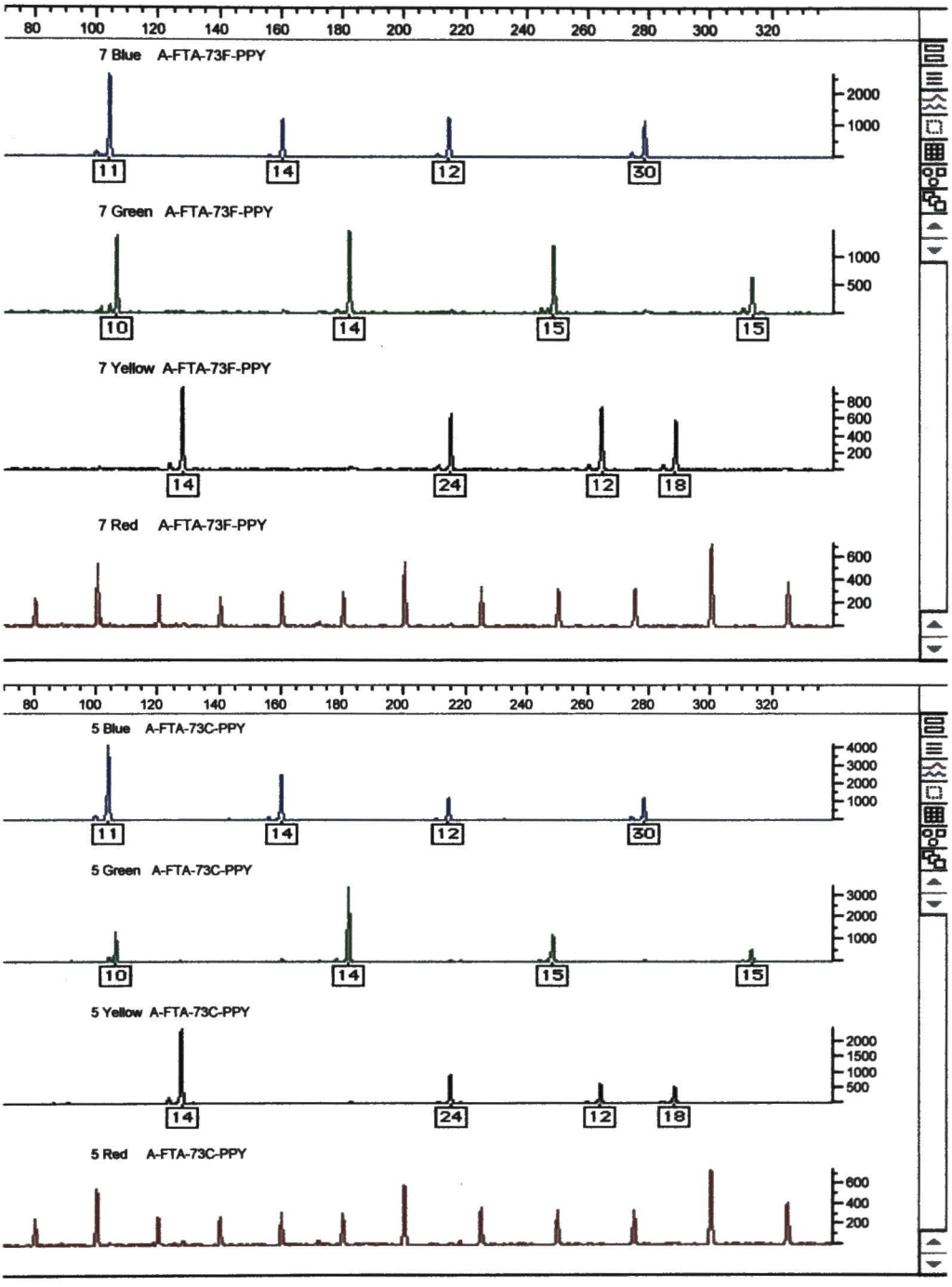


Figure 6: Typical father – son pair results in the Asian population, sample used was A-FTA-73. Father's profile is on top and son's profile is on bottom.

Mutation Rate

Within the 83 Asian pairs, three mutations where observed giving a frequency of mutation of 0.0361. Among the 207 Southwestern Hispanic pairs two mutations were observed giving a frequency of mutation of 0.0097. Mutations were seen at DYS389I, DYS389II, DYS439, DYS385a/b, and DYS391. Only DYS389II exhibited a mutation in both populations and no other locus had multiple mutations. Figure 7 demonstrates a typical father – son mutation where the child's allele at DYS391c showed a change of one repeat motif from an 11 allele to a 10 allele. Figure 8 is an example of a mutation that occurred at the DYS389II locus. Figure 9 is an example of how a mutation at DYS389I appears.

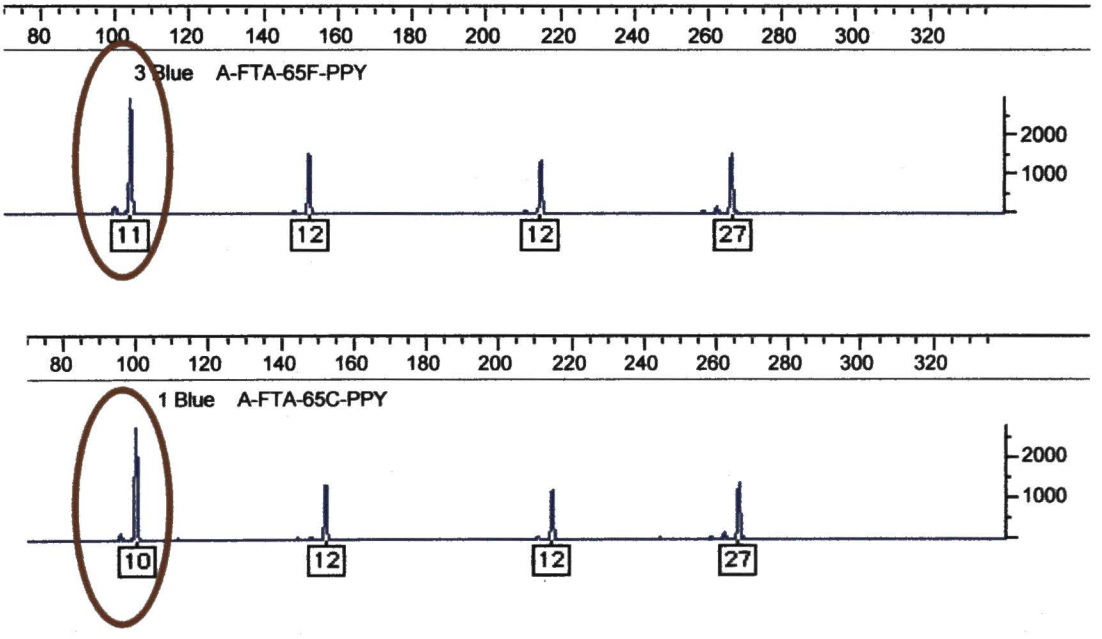


Figure 7: Examples of mutation at the DYS391 locus. Sample used was A-FTA-65. Father's profile is on top and son's profile is on bottom. Note that mutations are surrounded by red circle.

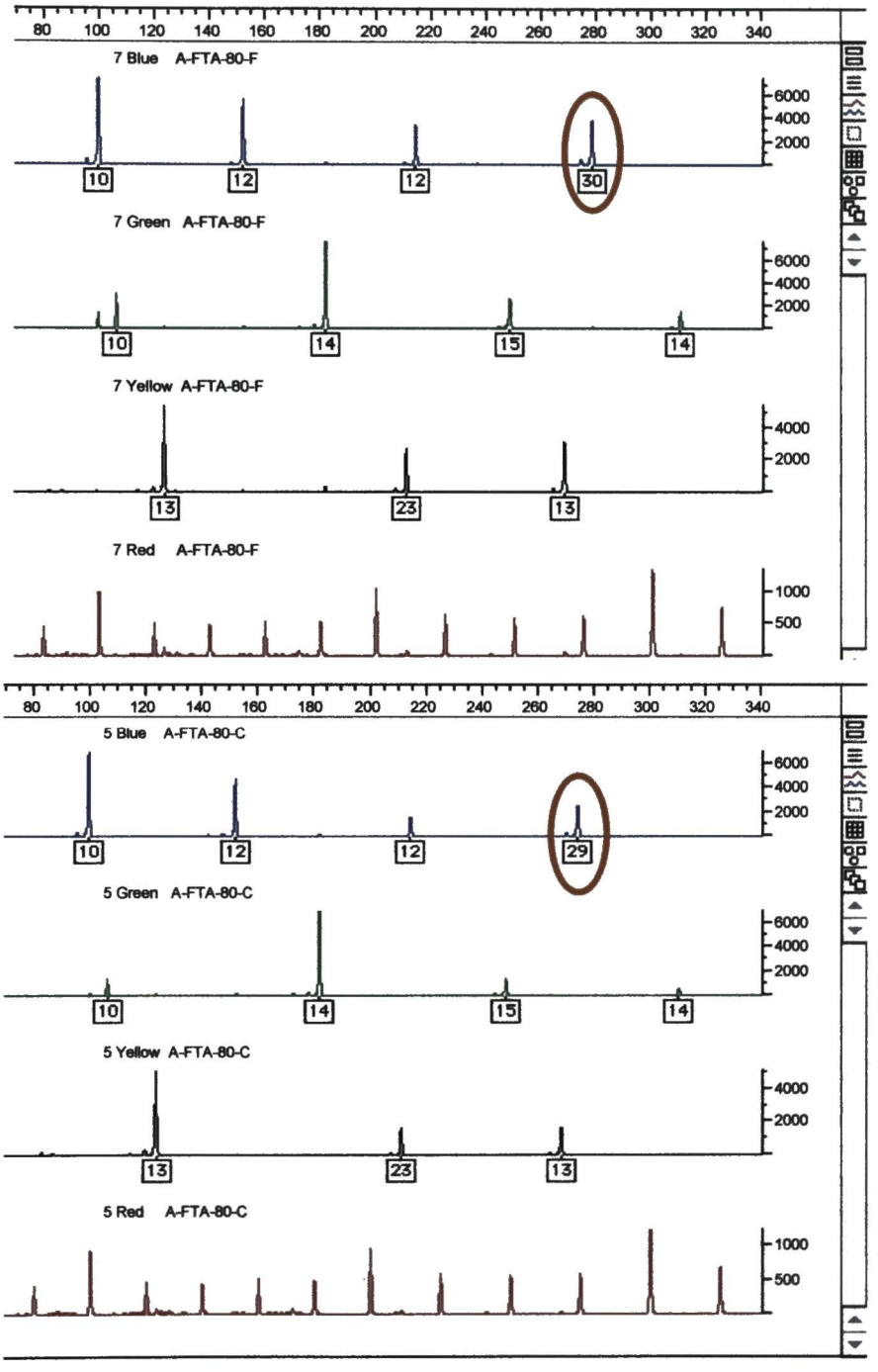


Figure 8: Example of a mutation at the DYS389II locus. Sample used was A-FTA-80. Father's profile is on top and son's profile is on bottom. Note that mutations are surrounded by red circle.

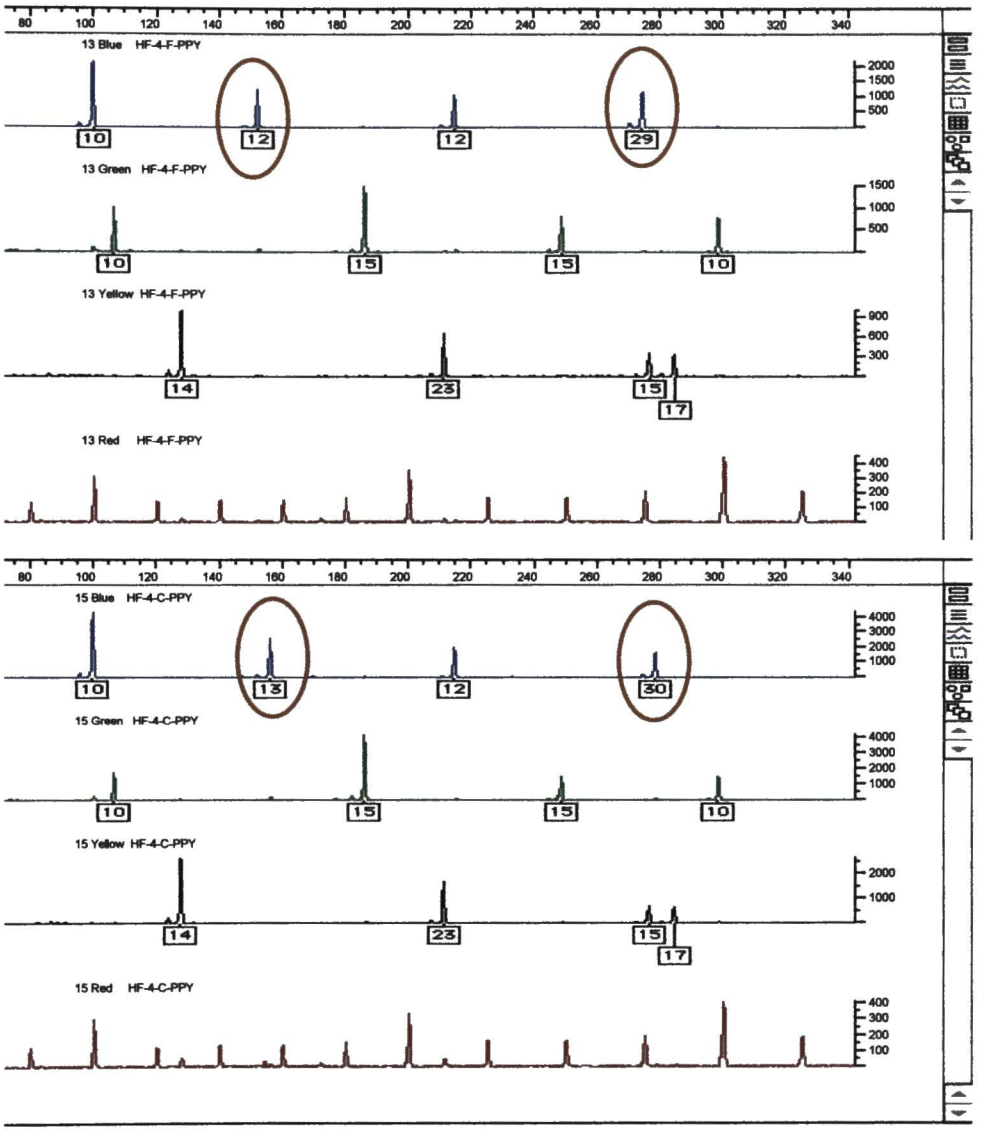


Figure 9: Examples of mutation at the DYS389I locus, which also results in a change at DYS389II.. Sample used was HF-4. Father's profile is on top and son's profile is on bottom. Note that mutations are surrounded by red circle.

Population Studies

Southwestern Hispanic (n =192) and Asian (n = 73) population groups were profiled using PowerPlex®Y System. The haplotypes are built on data from fathers

collected for the mutation study. Summaries of alleles and haplotypes data performed using Arlequin software. The allelic frequencies for Y-specific STR loci are shown in Table 2.

Table 2: Allele frequencies in Asian and Southwestern Hispanic populations using PowerPlex®Y System.

Locus	Allele	Asian	Hispanic
DYS19	10	0.0137	0.0000
	12	0.0137	0.0052
	13	0.0137	0.2566
	14	0.2877	0.4740
	14.2	0.0000	0.0052
	15	0.5068	0.1458
	16	0.1370	0.0729
	17	0.0274	0.0312
DYS389I	11	0.0137	0.0104
	12	0.5890	0.1667
	13	0.3151	0.6510
	14	0.0822	0.1719
	26	0.0137	0.0052
	27	0.1918	0.0260
DYS389II	28	0.2603	0.1146
	29	0.2192	0.3385
	30	0.2603	0.3542
	31	0.0411	0.1302
	32	0.0137	0.0312
DYS390	21	0.0000	0.0469
	22	0.0548	0.0417
	23	0.4110	0.2812
	24	0.3836	0.5312
	25	0.1507	0.0938
	26	0.0000	0.0052
DYS391	9	0.0274	0.0677
	10	0.7397	0.6146
	11	0.2329	0.3021
	12	0.0000	0.0156

Locus	Allele	Asian	Hispanic
DYS385-1	9	0.0137	0.0000
	10	0.0000	0.0052
	11	0.1370	0.2917
	12	0.3562	0.0781
	13	0.3425	0.1562
	13.2	0.0000	0.0052
	14	0.0548	0.1823
	15	0.0959	0.1302
	16	0.0000	0.1146
	17	0.0000	0.0260
	18	0.0000	0.0052
DYS385-2	12	0.0274	0.0052
	13	0.0411	0.0521
	14	0.0274	0.2917
	15	0.0137	0.0573
	16	0.1507	0.1250
	17	0.1370	0.1458
	18	0.1644	0.1406
	19	0.1781	0.0521
	20	0.0822	0.0156
	21	0.0137	0.0104
DYS437	13	0.0137	0.0052
	14	0.5342	0.6198
	15	0.4247	0.3385
	16	0.0274	0.0365
DYS438	8	0.0000	0.0052
	8.2	0.0000	0.0052
	9	0.0000	0.0938
	10	0.6575	0.2344
	11	0.3151	0.3177
	12	0.0274	0.3281
	13	0.0000	0.0156

Locus	Allele	Asian	Hispanic
DYS392	10	0.0000	0.0156
	11	0.0822	0.3125
	12	0.1096	0.0312
	13	0.3425	0.3698
	14	0.4110	0.1510
	15	0.0411	0.0625
	16	0.0137	0.0469
	17	0.0000	0.0052
	18	0.0000	0.0052
DYS393	12	0.5068	0.1562
	13	0.2603	0.7135
	14	0.2055	0.1146
	15	0.0274	0.0104

Locus	Allele	Asian	Hispanic
DYS439	9	0.0000	0.0104
	10	0.0274	0.1302
	11	0.2055	0.0365
	12	0.6027	0.3073
	13	0.1370	0.1302
	14	0.0274	0.0156

The haplotype frequency distributions were also determined for the PowerPlex®Y System in both the Asian and Southwestern Hispanic populations. Of the 192 samples that were typed in the Southwestern Hispanic population, 165 different haplotypes were determined. The Southwestern Hispanic population had four haplotypes that were observed four or more times and 150 haplotypes that were observed one time. In the 73 samples that were typed in the Asian population, 68 different haplotypes were found. Of the 68 different haplotypes only one haplotype was observed four times, two haplotypes were observed two times, and 65 haplotypes were observed one time. Table 3 A and B presents the haplotypes distributions that are shared between individuals in each population and a point estimate of the haplotype frequencies. Some of the haplotypes have a question mark at the locus DYS385b, this means that no allele was observed.

The nine European minimal haplotype loci (DYS19, DYS385A/B, DYS389I/II, DYS390, DYS391, DYS392, and DYS393) are included in the PowerPlex®Y System. To calculate haplotype frequencies for the European minimal haplotype frequencies for

complete PowerPlex®Y System's loci were analyzed on Arlequin, than the three additional loci were subtracted from further analysis. Of the 192 samples that were typed in the Southwestern Hispanic population, 161 different haplotypes were observed. 16 haplotypes were observed two time, two were observed three times, one was observed four times, one was observed nine times, one was observed 11 times, and 140 haplotypes were observed one time. In the 73 samples that were typed in the Asian population, 71 different haplotypes were observed. One haplotype was observed four times, five haplotypes were observed two time, and 65 haplotypes were observed one time.

Table 3A: Southwestern Hispanic Population (n = 192)
Y Haplotype (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385-1, DYS385-2, DYS437, DYS438, DYS439, N = Number of times observed, Freq = Frequency of Occurrence)

Sample Id	Haplotype												N	Freq
H7F	13	14	30	24	9	11	13	13	?	14	10	10	4	0.0208
H79F	13	12	28	23	10	13	14	14	18	14	11	12	2	0.0104
H76F	14	13	29	23	11	13	13	11	?	15	12	13	6	0.0313
H75F	14	13	29	24	10	13	13	11	14	15	12	12	6	0.0313
H70F	14	13	28	24	11	13	13	11	13	15	12	12	2	0.0104
H66F	17	13	28	23	10	11	13	12	13	15	10	11	2	0.0104
H60F	15	13	29	23	9	11	13	13	16	14	9	12	2	0.0104
H46F	14	13	29	24	11	13	13	11	14	15	12	12	6	0.0313
H38F	13	13	30	24	10	11	13	16	19	14	10	13	2	0.0104
H34F	16	14	30	24	11	13	13	11	13	14	12	12	2	0.0104
H28F	14	13	29	24	10	13	13	11	14	14	12	12	2	0.0104
H24F	16	14	29	23	10	11	13	12	?	15	10	12	2	0.0104
H172F	13	13	30	24	11	14	14	15	16	14	11	12	2	0.0104
H138F	14	13	29	24	10	13	13	11	14	15	12	11	2	0.0104
H118F	14	13	29	24	10	13	13	11	14	15	12	12	2	0.0104

*? = No allele observed at locus

Table 3B: Asian Population (n = 73)
Y Haplotype (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385-1, DYS385-2, DYS437, DYS438, DYS439, N = Number of times observed, Freq = Frequency of Occurrence)

Sample Id	Haplotype												N	Freq
A-33-F	15	12	27	25	10	13	12	12	20	14	10	13	2	0.0274
A-19-F	14	12	28	23	10	14	12	15	18	15	11	12	4	0.0548
A-62-F	14	13	30	23	10	14	13	13	?	14	10	12	2	0.0274

*? = No allele observed at locus

The genetic diversity of the PowerPlex®Y System and the European Minimal Haplotypes were calculated for the Asian population, of 0.9970, and for the Southwestern Hispanic population had a of 0.9988. The genetic diversity is calculated by finding the expected haplotype frequency, and multiplying it by the correction for sample size bias. Looking at the same dataset and considering only the European Minimal Haplotype loci, the genetic diversity was 0.9958 and 0.9935 for the Asian and Southwestern Hispanic populations, respectively.

DISCUSSION:

The PowerPlex®Y System allows amplification of twelve Y chromosome-specific STR loci. Nine of the loci are also included in the European minimal haplotype, DYS438, DYS439 have been recommended by SWGDAM (Scientific Working Group on DNA Analysis Methods) along with the European minimal haplotype loci to comprise the base set of loci to be used in the United States (Saldanha, et al., 2002), and DYS437, was included by Promega Corporation to increase the discriminating power.

The mutational study found that mutations did occur among the Y-STR loci in both populations. A mutation is an error that occurs which changes the nucleotide sequence (Alberts, et al., 2002). Changes within the nucleotide sequences usually occur when normal mechanisms to copy or repair fail. A typical error that is seen is a base pair change such as a Single Nucleotide Polymorphisms (SNP), insertion, or deletion. Another error that is commonly seen when dealing with Short Tandem Repeats is a change in repeat length. The differences in repeat length is inherited by the offspring of the individual. These differences in repeat length are what is being observed when analyzing STR's (Nadir, et al., 1995). Both types of mutations typically are a result of the polymerase slipping or gaps formed between the template strand and the strand that is being extended. Figure 10 demonstrates one way that a difference in repeat length can occur.

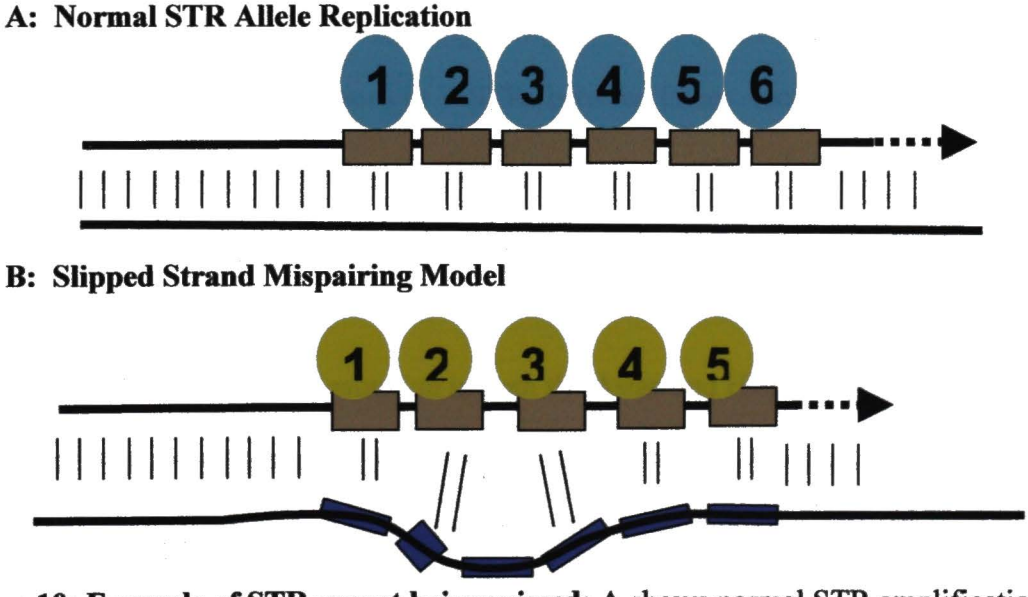


Figure 10: Example of STR repeat being missed: A shows normal STR amplification. B shows the template DNA strand breathing, allowing a repeat to be missed during amplification. (Walsh *et al* (1996) *Nucleic Acids Res.* 24: 2807-2812)

In this study, the Asian population had a higher frequency of mutation than the Southwestern Hispanics. Within the Asian population, three mutations were observed; in the Southwestern Hispanic population two mutations were observed. It is interesting to note that Asian population consisted of a smaller number of sample meioses, so it would be expected to have displayed fewer mutations than the Southwestern Hispanic dataset. The frequency of mutation in both populations for the PowerPlex®Y System were higher than that of autosomal DNA, which is between 0.001 – 0.005 (Butler, 2003). These results agree with findings from other studies that report that there is a higher rate of accumulation of variants on the Y-chromosome, more than likely due to having a higher number of meiosis during gametogenesis (Scozzari, et al., 1997). The datasets examined are not sufficient to determine precise locus-specific mutation rates, hundreds more father – son pairs would be needed to determine accurate locus-specific mutational rates. The fact that mutations were observed with such a small data size was a significant finding in itself. The observation of mutations in Y-STRs between father – son pairs will affect how scientist determine true paternal lineages, since, they can't exclude an individual from a particular lineage if one locus shows to be different from the paternal reference sample.

DYS389II presented a mutation in both the Asian and Hispanic populations, single mutations were observed at the four other loci. One father- son pair presented two mutations at DYS389I and DYS389II loci. This was somewhat expected since both loci were amplified with the same primer set, other studies (DNA Heritage, 2004) have found

that DYS389I and DYS389II occasionally show concurrent mutations. Figure 11 shows the composition of the DYS389 locus.

Allele distributions were different between the two populations with some alleles observed for the Asian population that did not occur in the Southwestern Hispanic population, and some alleles that were observed in the Southwestern Hispanic population that did not occur in the Asian population (See table 2). At one locus, DYS385b a ? appears in the haplotype designation in some samples. This indicates that a discrete second allele was not detected at this locus and may reflect tow copies of the DYS385a allele are present. Another explanation of this phenomenon may be a primer binding mutation on that site or loss of allelic signal due to degradation. These causes, however did not seem evident in this dataset or were not tested further.

Haplotype distribution is the standard way to analyze markers that are uni-parental inherited such as for Y-STR or mtDNA. The method used to determined haplotype distribution is called the "Counting Method," since counts of how many times a particular haplotype appears within the dataset are recorded. Population size greatly affects the results obtained by using the Counting Method. The larger the population size, the higher the probability of seeing rarer haplotypes is, along with establishing other haplotypes are more common.

There were no haplotypes shared between the Asian and Southwestern Hispanic populations. These results went against the initial hypothesis that there would be haplotype profiles that were shared between the two populations, due to a major influence of Asian haplotypes in Native American populations (Zegura, et al., 2004). An

explanation for this finding is that the Asian population that was used for this study was specifically Hong Kong Chinese. The Asian populations that crossed the land bridge are thought to have had origins from the Altai Mountain region (Zegura, et al., 2004).

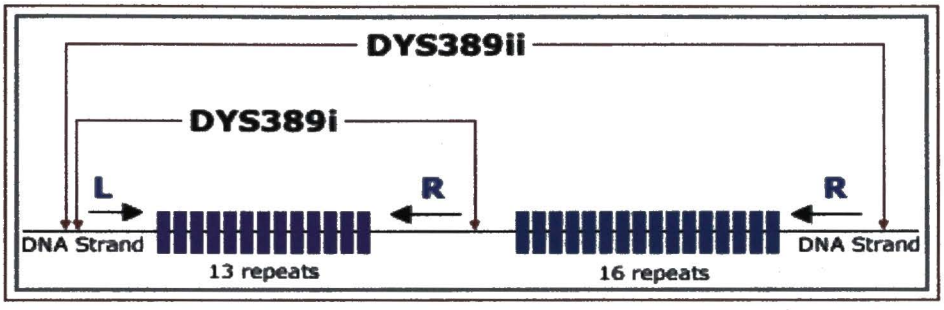


Figure 11 DYS389 Loci: This figure shows how the DYS389I/II loci are amplified with the same primer sets. The "L" primer bind only in one location, but the "R" primers bind at two allowing for two loci to be amplified. If a mutation occurs within the first primer region (DYS389I), the amplified product from the second region (DYS389II) will be affected (see Figure 9), so only one mutation will be counted. If a mutation occurs at DYS389II, DYS389I will not be affected (see Figure 8). (Figure from DNA Heritage, 4/12/2004, available online at: http//www.dnaheritage.com/markers.asp)

The Southwestern Hispanic population had more haplotypes that were shared between individuals then the Asian population. This may be the result of the small population size of the Asian population and that the Asian population studied was from a distinct biogeographical entity. The Southwestern Hispanic population had four haplotypes that were observed four or more times, where the Asian population only had one haplotype that was observed four times. Some of the haplotypes that occurred more than one time, did not have an observed allele at the locus DYS385b. This might be due to degradation more likely it is due to how the DYS385b locus is amplified, as mentioned

previously DYS385a and DYS385b are two tandemly duplicated loci, similar to DYS464 shown in figure 12.

The haplotype profiles observed multiple times were evaluated on the online European Database. This database has the largest collection of Y-Short Tandem Repeat profiles. As of September of 2000 it contained 4688 haplotypes (Roewer, et al., 2001) and is currently made up of 13,986 haplotypes.

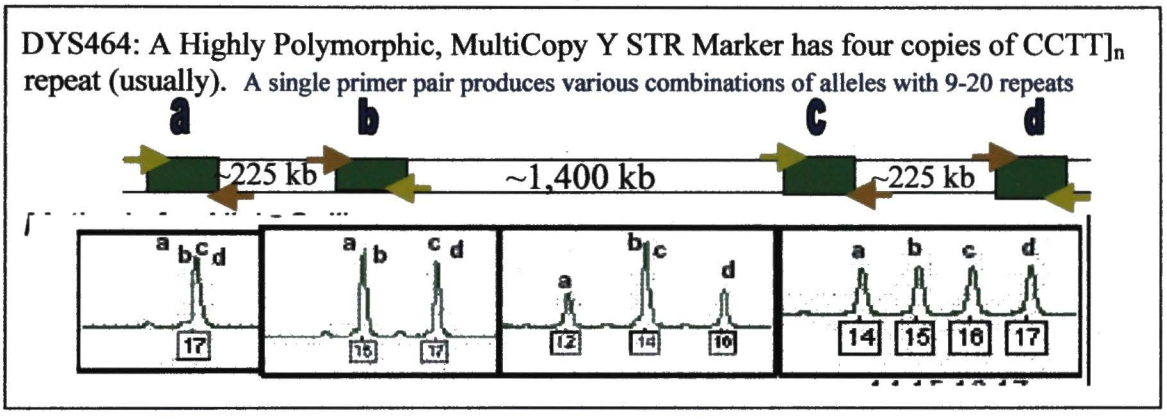


Figure 12:Amplification of locus DYS464. Figure from Butler. AAFS Y-STR workshop (2004). Internet. 2004. Available: http://www.cstl.nist.gov/div831/strbase/pub_pres/ButlerAAFS2004Y, PDF

Five of the 15 haplotypes entered were observed in the European Database. This suggests that the haplotypes that these lineages may have a European ancestry. Those that did not appear in the European Database may have a Native American influence however this could not be evaluated since a North American database is currently not available. None of the multiple haplotypes found in the Asian population appeared in the European database. This is understandable since the Asian population that was tested had their origins from China not from Europe and reinforces the biographical distinction of these regions. In the Southwestern Hispanic population, only one haplotype had a similar

frequency of occurrence in both this study and in the European Database. This haplotype profile was DYS19 -14, DYS389 I – 13, DYS389 II – 29, DYS390 – 24, DYS391 – 11, DYS392 –13, DYS393 – 13, DYS385a – 11, and DYS385b – 14. It had a frequency of occurrence in our study of 0.0313 and in the European Database it had a frequency of occurrence of 0.0317. The similar frequency of occurrence for this haplotype not only shows that this is a common haplotype, but even though this study had a small sample size it is comparable to a much larger database up to the indicating that the sample size used for this population is sufficient for determining point estimates, at least, the more common haplotypes. Table 4 A and B show the observed frequency of distribution of the most frequent minimal haplotypes determined for these data sets analyzed against the European Y-STR Haplotype Reference Database.

Table 4A: Minimal Haplotype for Southwestern Hispanic Population found in European database (n = 13986 European population) (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385-1, DYS385-2, N = Number of Times Observed in the European Database, Freq = Frequency of Occurrence in European database)

Sample Id	Haplotype									N	Freq
H7F	13	14	30	24	9	11	13	13	?	0	0.0000
H79F	13	12	28	23	10	13	14	14	18	0	0.0000
H76F	14	13	29	23	11	13	13	11	?	0	0.0000
H75F	14	13	29	24	10	13	13	11	14	190	0.0136
H70F	14	13	28	24	11	13	13	11	13	4	0.0003
H66F	17	13	28	23	10	11	13	12	13	0	0.0000
H60F	15	13	29	23	9	11	13	13	16	2	0.0001
H46F	14	13	29	24	11	13	13	11	14	443	0.0317
H38F	13	13	30	24	10	11	13	16	19	15	0.0011
H34F	16	14	30	24	11	13	13	11	13	0	0.0000
H24F	16	14	29	23	10	11	13	12	?	0	0.0000
H172F	13	13	30	24	11	14	14	15	16	0	0.0000

*? = No allele observed at locus

Table 4B: Minimal Haplotype for Asian Population found in European database (n = 13986 European population) (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385-1, DYS385-2, N = Number of Times Observed in the European Database, Freq = Frequency of Occurrence in European database)

Sample Id	Haplotype									N	Freq
A-33-F	15	12	27	25	10	13	12	12	20	0	0.0000
A-19-F	14	12	28	23	10	14	12	15	18	0	0.0000
A-62-F	14	13	30	23	10	14	13	13	?	0	0.0000

*? = No allele observed at locus

The European Minimal Haplotype was calculated for all of the samples in this study and was compared to the results that were obtained from the PowerPlex®Y System. Fewer individuals sharing haplotypes were observed with the PowerPlex®Y System than the European Minimal Haplotype Loci, indicated that the addition of extra loci included in this system increased the discriminating ability of the kit (Figure 13). This finding agrees with other studies, such as Hall and Ballantyne, (2003), which used an 18-locus Y-STR system in order to increase individualization of unrelated males. Kayser, et.al., (2002) showed that in the Hispanic population dataset of 478 individuals, 354 haplotypes were observed when analyzed with the European Minimal Haplotype, which produced a haplotype diversity of 0.9948.

The basis of the genetic diversity calculations is to give an estimate of discriminating power of the panel of markers. The higher the genetic diversity the more variation there is within this panel. Both the European Minimal Haplotype loci and PowerPlex®Y System had high levels of genetic diversity. The extra loci in the PowerPlex®Y System helped to increase the genetic diversity 0.12% and 0.53% from that of the European Minimal Haplotype Loci in the Asian and Southwestern Hispanic

populations, respectively. The European Database overall gave a genetic diversity of 0.9972 (Willuweit and Roewer, 2004), which is similar to genetic diversity determined using PowerPlex®Y System for both populations. This shows that even with the smaller population size most of the alleles at the loci tested were observed.

From the haplotype databases, point estimates for observed haplotypes in the Southwestern Hispanic and Asian populations can be made. These values are highly dependent on the database size and a confidence intervals (CI) should be placed around each point estimate. The equation for 95% CI is $p \pm 1.96[(p\ (1\text{-}p))/N]^{1/2}$ where p equals the point estimate and N equals population database size or database size. An example of point estimate and CI for a haplotype detected in the Hispanic population only one time is 0.005 ± 0.010. An Asian population haplotype observed one time has a point estimate and CI of 0.014 ± 0.0270. It is interesting to note how population size affects these calculations. For casework, if a haplotype was not observed in the database, an upper bound confidence level would be determined. The formula for this is $1 - \alpha^{(1/N)}$, α is equal to the confidence level and N is equal to the population database size. This calculation is also affected by population size. These calculations are also used for mtDNA haplotype frequency estimates.

Upon completion of this study, the Promega PowerPlex®Y System it was found to be a good tool for Y-STR analysis. Although some mutations were observed, the panel still contained a high level of genetic diversity, and paternal lineages can be determined using this system. Upon completion of a North American database, the

confidence intervals around point estimates can be narrowed, allowing it to be used this system to be used for actual casework

Figure13A: Comparison of Frequency Distribution Between PowerPlex®Y Systems Loci and the Minimal Haplotype Loci for the Southwestern Hispanic population

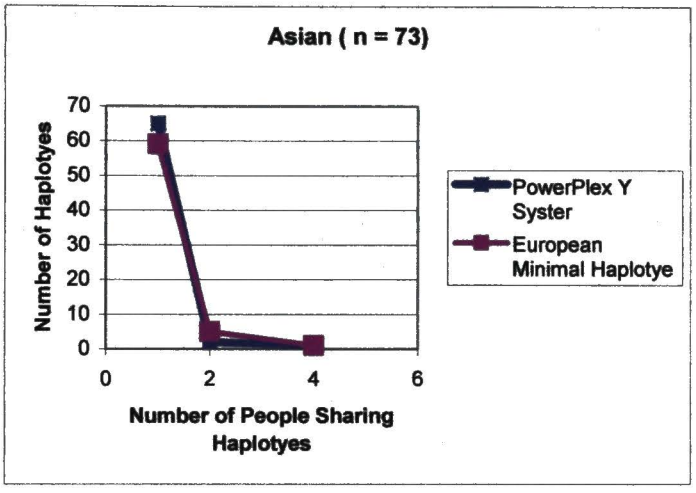
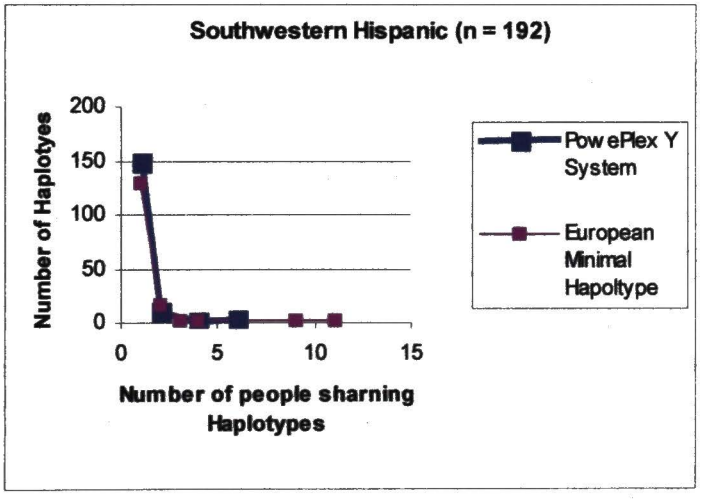


Figure 13, B: Comparison of Frequency Distribution Between PowerPlex®Y Systems Loci and the Minimal Haplotype Loci for the Asian population



REFERENCES:

Alberts B, Johnson A, Lewis J, Raff M, Walter P. Molecular Biology of the Cell. Garland Science, 2002; 18, 198, 816-818, 1135, 1140-1142.

Budowle B, Smith J, Moretti T, DiZinno J. DNA Typing Protocols: Molecular Biology and Forensic Analysis. Natick: Eaton Publishing, 2000

Budowle B, Y-STR Studies, Presented at the Fourteenth International Symposium on Human Identification 2003; Internet.2004. Available: www.promega.com
Butler, Forensic DNA Typing. London: Academic Press, 2003; 55-97, 117-134.

Butler J, AAFS 2004 Y-STR Workshop. Internet. 2004. Available: http://www.cstl.nist.gov/div831/strbase/pub_pres/ButlerAAFS2004Y, PDF

DNA Heritage, Internet. 2004. Available: http//www.dnaheritage.com/markers.asp

Gusmao L, Carracedo A, Y-STR. Profiles in DNA. Internet. 2004. Available: www.promega.com

Hall A, Ballantyne J. The development of an 18-locus Y-STR system for forensic casework. Anal. Bioanal. Chem., 2003; 376:1234-1246.

Hammer's group. YSTR Positions along Y Chromosome. Internet. 2004. Available:http://www.cstl.nist.gov/div831/strbase/ystpos1.htm

Jobling M, Smith C. The Human Y Chromosome: An evolutionary marker comes of age. Genetics. 2003; 4:598-612.

Kayser M, Brauer S, Willuweit S, Schadlich H, Batzer M, Zawacki J, et al. Online Y – chromosomal Short Tandem Repeat Haplotype Reference Database (YHRD) for U.S. Populations. Journal of Forensic Sciences. 2002; 47(3):513-519.

Nadir E, Margali H, Galliey T, Ben-Sasson S. Microsatellite spreading in the human genome: Evolutionary mechanisms and structural implications. Evolution.1996; 93:6470-6475.

National Research Council. DNA Technology in Forensic Science. Washington D.C: National Academy Press. 1992; 104-105.

Prinz M, Boll K, Baum H, Shaler B. Multiplexing of Y chromosome specific STRs and performance for mixed samples. Forensic Science International. 1997; 85:209-218.

Promega Corporation. DNA IQ™ System – Database Protocol. 2002; Internet. 2004. Available: www.promega.com

Quality Assurance Standards for Forensic DNA Testing Laboratories. 1998; Internet. 2004. Available:www.csh.nist.gov

Roewer L, Krawczak M, Willuweit S, Nagy M, Alves C, Amorin A, et al. Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. Forensic Science International. 2001; 118:106-113.

Saldanha G, Budowle B, Arcot S, Richey S, Chakraborty R, Jones M. PowerPlex®Y System Validation Study. Promega. 2002; 1-37.

Sinha S, Budowle B, Arcot S, Richey S, Chakraborty R, Jones M, et al. Development and Validation of a Multiplexed Y-Chromosome STR Genotyping System, Y-PLEX™6, for Forensic Casework. Journal of Forensic Science. 2003; 48(1):93-103.

Schnieder A, Roessli D, Excoffer L. Arlequin vs.2.000.[computer porgram] University of Geneva (Switzerland) 1997; Internet. 2004. Available: http://anthropologie.unige.ch/arlequin/

Scozzari R, Cruciani F, Malaspina P, Santolamazza P, Ciminelli B, Torroni A, et al. Differential structuring of human populations for homologous X and Y microsatellite loci. American Journal of Human Genetics. 1997; 61: 719-733.

Walsh S, Fildes N, Reynolds R. Nucleic Acids Res. 1996; 24:2807-2812.

Whatman BioScience. Whatman BioScience DNA Pure & Simple. Catalogue. 2000

Willuweit S, Roewer L. Y-STR Haplotype Reference Database. Internet. 2004. Available: http://ystr.charite.de/index_gr.html

Zegura S, Karafet T, Zhivotovsky L, Hammer M. High-Resolution SNPs and Microsatellite Haplotypes Point to a Single, Recent Entry of Native American Y Chromosomes into the Americas. Molecular Biology and Evolution. 2004; 21(1):164-175.