H

Genetic Characterization of the Y-Chromosome Specific Minisatellite, MSY1, in a Multi-cultural Population of New York City

by

Pasquale Buffolino

A dissertation submitted to the Graduate Faculty in Biochemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

2002

UMI Number: 3074629

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

This manuscript has been read and accepted for the Graduate Faculty in Biochemistry in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

17 Dec 2002

Date

Chair of Examining Committee

Dec 17th 2002

Date

Executive Officer

Supervisory Committee

The City University of New York

Abstract

Genetic Characterization of the Y-Chromosome Specific Minisatellite, MSY1, in a Multi-cultural Population of New York City

by

Pasquale Buffolino

Advisor: Professor Lawrence Kobilinsky, Ph.D.

Minisatellite variant repeat mapping by PCR (MVR-PCR) has revealed a high degree of allelic variability at the single Y-chromosomal minisatellite locus, MSY1. MVR-PCR assays both the allelic length variability and internal structural variability of MSY1. Modular structures representing the interspersion pattern of variant repeat units within MSY1 have offered far greater levels of discrimination as compared to the analysis of allelic length. MVR-PCR has been used to determine the structure of MSY1 and degree of code diversity in four major ethnic groups of New York City. 180 complete codes and 192 modular structures were defined in Caucasians, African Americans, Asians, and Hispanics. Code and structural diversity was calculated using Nei's unbiased estimator. Caucasians and African Americans displayed the highest levels of genetic diversity when compared to global MSY1 data. Asians and Hispanics were surpassed by a single British population. Defined structural characteristics were observed within the four populations. These characteristics were compared to previous studies. American Caucasians and Hispanics were found to be genetically similar to western Europeans. African Americans showed a striking resemblance to the interspersion of repeats and structural diversity observed in African populations. The highest 3:1:3:4 structural frequency was found in New York City Asians. This structure was not specific to this population, with relatively equal frequencies

found in Caucasian (21.2%) and Hispanic (26.9%) populations. This correlates with previous data collected from all continents. Analysis of Molecular Variance (AMOVA) revealed a larger percent of within population (88.83%) than between-population (11.17%) genetic variance. Significant population differentiation was determined through F-statistics (F_{ST} = 0.11166) inferring population stratification with some gene flow between populations. The defined characteristics in New York City populations and degree of differentiation implies that the Y-chromosome is not as homogenized in urban populations as has been speculated.

A four-reaction semi-automated detection system was developed for the characterization of MSY1 in true forensic samples. In mixed samples, male specific DNA was detectable with as little as 2.5 ng of sperm cell DNA. A 72.9% amplification success rate was observed for postmortem bloodstains. Comparing this to a 97.2% amplification success rate observed in microsatellite analysis, the present study demonstrates the difficulties involved with the characterization of large minisatellite loci in forensic samples characterized by degradation and low concentrations of DNA. The power of discrimination of MSY1 is found to be greater than nine microsatellite loci combined. When combining both minisatellite and microsatellite data, all haplotypes in the present study were determined to be unique. The probative value of analyzing MSY1 for identification purposes is obvious. However, the method of detection must be validated and standardized before it can be accepted by the forensic community. The present study offers insights into the distribution of MSY1 in four major ethnic categories in United States forensic databases and its feasibility for application in forensic casework.

Acknowledgments

I would like to thank Dr. Horst Schultz for giving me the opportunity to participate in the biochemistry program while employed by the New York City Office of Chief Medical Examiner. I hope that I have lived up to all his expectations. Without him, none of this would have been possible. I would like to thank Dr. Howard Baum for his ideas and guidance with this project and Dr. Mecki Prinz for her knowledge and wisdom. Thank you to Mike Elmozino for his work with the microsatellite raw data and to Carole Meyers for her help with bloodstain sample preparations. Thank you to Dr. Robert Shaler for allowing me to conduct my research in the laboratories of the New York City Medical Examiner Office and for being the best boss anyone could ask for. I would like to thank all my friends and colleagues at the New York City Medical Examiner Office for their support and patience with me during some of the toughest and frustrating times of my life. I miss them all dearly since I have left the Department of Forensic Biology. I must extend an enormous thank you to Dr. Lawrence Kobilinsky who always had time to return my phone calls, especially when I knew he was overwhelmed. I thank him for his confidence in me and for assuring me of the importance of my work. My wife, whom I have loved dearly since we have met over seventeen years ago, deserves the most gratitude. Her support and understanding allowed me to continue at the pace which I have worked for many years. It is for her and my two beautiful children that I have worked so hard. Finally, I would like to thank my family, especially my mother and father, who have always had faith in my ability and always told me that I would succeed. I dedicate this thesis to all those who have helped me fulfill my academic and career goals.

Table of Contents

List of Tables

List of Figures

Section I: Introduction

The analysis of genetic variation has become a powerful tool for historians, archaeologists, paleontologists, and linguists in their search for the origin of man. The DNA we inherit from our ancestors accumulates mutations over time and these slowly evolving events provide information pertaining to our relatedness and genetic history (Jobling, *et al.*, 1995; Hammer *et al.*, 1998). Autosomal and haploid systems have supplied complimentary data tracing the origin of man to a common African ancestor dating back some 200,000 years (Vigilant *et al.*, 1991; Stoneking, *et al.*, 1993; Cann, *et al.*, 1987; Hammer, *et al.*, 1996; Jobling, *et al.*, 1998a). The mode of inheritance and lack of recombination in mitochondrial DNA (mtDNA) and Y chromosomal DNA have made the two systems prominent in the study of human evolution. The highest degree of genetic diversity has been observed in African populations with major differences between African and non-African populations (Jorde, *et al.*, 1997; Armour, *et al.*, 1996). These results correlate with the expansion of non-African populations from a diverse set of African populations (Jorde, *et al.*, 1997) thereby placing the first branch point between African and all non-African populations (Seielstad, *et al.*, 1999; Jorde, *et al.*, 2000).

Human mtDNA is a circular genome of approximately 16.5-Kbp in size. mtDNA is maternally inherited with little to no observed recombination. Therefore, all individuals along the same maternal lineage share a common mtDNA type. The complete nucleotide sequence of the extranuclear genome has been reported by Anderson et. al., 1981. Base-substitution mutation rates of mtDNA have been calculated at approximately ten times the rate of its autosomal counterpart resulting in a high degree of genetic diversity. The displacement loop (D-loop) which is situated within the mitochondrial control region is one of the most polymorphic regions in the entire genome. Two regions of increased polymorphism were discovered through sequence comparison of the 680 bp D-loop region between human, bovine, and rat genomes. These regions are referred to as the Hypervariable Region I and Hypervariable Region II. Initial evolutionary studies relied on the accumulation of sequence data from these two polymorphic regions (Vigilant, *et al.*, 1991; Stoneking, *et al.*, 1993; Hammer, *et al.*, 1997; Jobling, *et al.*, 1998a).

The Y chromosome is approximately 60-Mbp in size making it approximately four thousand times larger than mtDNA (Jobling, *et al.*, 1995). Inheritance is strictly paternal with no contribution of mtDNA, making the Y chromosome the male mitochondrial counterpart (de Kniff, 1997). The Y chromosome is not entirely exempt from recombination during meiosis (Jobling, 1994; Hammer, *et al.*, 1996). Recombination occurs at two small pseudoautosomal regions (PAR1 and PAR2) located at the tip of each arm (refer to Figure 1). PAR1 and PAR2 are homologous to sequences on the X chromosome and undergo normal recombination. The remainder of the Y chromosome displays normal paternal inheritance with no observed recombination (Hammer, *et al.*, 1996; Thangaraj, *et al.*, 1999).

The complexity and lack of mutation on the Y chromosome have been obstacles in the progress of Y chromosomal research (Hammer, *et al.*, 1996; Jeffreys, *et al.*, 1990; de Kniff, *et al.*, 1997). Since the development of the polymerase chain reaction (PCR) by Kary Mullis in 1985, there has been an explosion of novel Y chromosomal polymorphisms and methods used for their detection (Jobling, *et al.*, 1995; Hammer, *et al.*, 1996; Jobling, *et al.*, 1998a). As a result, a variety of binary and multi-allelic polymorphisms were detected. Binary markers contain low mutation rates and have been useful in defining groups of chromosomes in the population. These properties have made them the most advantageous type of polymorphism in the construction of Y chromosomal trees (Jobling, *et.al.*, 1997). Two types of multi-allelic markers, minisatellites and microsatellites (commonly referred to as short tandem repeats or STR), are categorized according to the number of tandem repeat units within them. Minisatellites contain tandem arrays of 10-50 bp repeat units, while microsatellites contain tandemly arranged core repeats consisting of 2-5 bp (Jobling, *et al.*, 1998a). There are approximately 173 microsatellites (Kayser, *et al.*, 2002a) and a single minisatellite identified on the Y chromosome (MSY1). These multi-allelic markers have a higher rate of mutation as compared to binary markers (Jobling, *et al.*, 1995; Hammer *et al.*, 1996; Roewer, *et al.*, 1992; Jeffreys, *et al.*, 1990; Jobling, *et al.*, 1997), with an estimated 10^{-3} mutations per locus per generation for microsatellites and an approximate 1% mutation rate per generation for MSY1, the single minisatellite identified on the Y chromosome (Jobling, *et al.*, 1997). The degree of variability in these multiallelic markers has enabled the study of intra-allelic differences within closely related populations defined by binary markers (Jobling, *et al.*, 1998a; Jobling, *et al.*, 1999).

Polymorphic markers on the Y chromosome have become a valuable tool in forensic studies (Jobling, *et al.*, 1999; Jeffreys, *et al.*, 1991; Hopkins, *et al.*, 1994; Jobling, *et al.*, 1997; Roewer, *et al.*, http://www.promega.com/geneticidproc/esusymp2proc/03.pdf). Based upon the National Center for Victims of Crime & Crime Victims Research and Treatment Center, 12.1 million American women have been victims of forcible rape. This averages 13% or one out of eight American women, translating to approximately 683,000 adult women who are forcibly raped each year (New York City Alliance Against Sexual Assault, http://www.nycagainstrape.org/index.html). Forensic evidence recovered in connection with sexual offenses is collected with the use of standardized rape kits. Sterile swabs are used to collect specimens from the victim. The swabs most often contain a mixture of DNA from the victim (epithelial cells from the vaginal wall) and assailant (sperm cells) (Gusmao, *et al.*, 1999). These cell mixtures can be separated using a differential lysis procedure commonly employed in forensic casework (Reynolds, *et al.*, 1991). With this procedure, DNA from the assailant can be separated from the victims DNA. The efficiency of the extraction procedure is dependant upon the ratio of sperm cells to female epithelial cells in the mixed sample. Assuming that each sperm cell contains 3 pg of DNA, a minimum of 500 sperm cells must be present in the extracted sample.[1] Below this level, the DNA profile of the male assailant can potentially be masked using autosomal STR systems. This phenomena has also been observed in mixed forensic samples containing higher concentrations of sperm cell DNA with predominant concentrations of female DNA and results from preferential amplification during PCR. Similar results have been observed in forensic samples contributed by azospermatic assailants (Gusmao, *et al.*, 1999). The development of highly informative Y chromosomal DNA testing systems have circumvented such limitations through the selection of male specific DNA (Gusmao, *et al.*, 1999; Prinz, *et al.*, 2001; Henke, *et al.*, 2001; Roewer, *et al.*, http://www.promega.com/geneticidproc/esusymp2proc/03.pdf).

The relatively small size of microsatellite markers make them well suited for the analysis of highly degraded DNA samples associated with forensic casework. This, along with the

---

[1]Calculation is based upon a final extraction volume of 200 μL with a final concentration of 0.15 ng of extracted sperm cell DNA per 20 μL.

capability of combining several Y chromosomal loci into a PCR multiplex system, have made microsatellites the best strategy to facilitate high throughput analyses required in modern forensic biology laboratories (Gusmao, *et al.*, 1999). Additionally, two online Y-chromosome STR haplotype reference databases (YHRD) exist, a European YHRD (European YHRD, http://ystr.charite.de/index_gr.html.) containing 12111 minimal haplotypes and a U.S. YHRD (Kayser, M. *et al.* 2002b) containing 1705 minimal haplotypes. Since males comprise the majority of sex offenders, Y chromosomal systems become a valuable investigatory tool, especially for the exoneration of the innocent. There is no doubt of innocence if the presumed suspect contains a dissimilar haplotype when compared to forensic evidence. When considering an inclusion, caution must be expressed when interpreting the forensic significance of Y chromosomal match probabilities. If a suspect contains the same DNA profile as compared to forensic evidence, one of two scenarios must be true; (1) the suspect was the contributor of the DNA from the forensic evidence, or (2) someone other then the suspect, who has the same DNA profile as the suspect, was the contributor of the DNA from the forensic evidence (NRC, 1996). For autosomal systems, the major consideration of an inclusion is the frequency of the DNA profile in the population. The less frequent the DNA profile is in a given population, the more likely the inclusion (Jobling, *et al.*, 1997). Due to the lack of recombination and its haploid fashion of inheritance, the product rule used to determine population frequencies from autosomal DNA databases cannot be applied to the Y chromosome. Frequencies of inclusion generated for Y chromosomal STR systems are significantly lower than that of autosomal STR systems since haploid frequencies are dependant upon database size (Jobling, *et al.*, 1997; Prinz, *et al.*, 2001). The structure of population databases commonly used by forensic biology laboratories must also be considered when interpreting match frequencies for Y haplotypes. Reference samples utilized for evolutionary or anthropological studies are carefully chosen to best reflect the ethnic category they are placed into. These samples originate from localized populations where common haplotypes are shared. Forensic biologists normally do not operate under these circumstances. The localization experienced by isolated populations does not exist in urban settings. Therefore, match probabilities derived from an "urban-like" population database would be underestimated when comparing an individual from an isolated sub-population. These concerns similarly arise in autosomal systems, however, they are exaggerated in Y chromosomal systems due to their haploid fashion of inheritance (Jobling, *et al.*, 1997). Nevertheless, the unique features inherent in Y specific systems have added a powerful degree of discrimination to forensic DNA testing.

Aim of Study

Since the discovery of microsatellite DNA, minisatellites have been frowned upon by the forensic community. The template length of minisatellite DNA makes the analysis of forensic samples associated with degradation and low concentrations of DNA a difficult chore (Armour, *et al.*, 1992). The single haploid minisatellite, MSY1, identified on the Y chromosome consists of an array of 48-114 AT-rich 25 bp repeats of at least five different variant types (refer to Figure 2a-b) (Jobling, *et al.*, 1997; Jobling, *et al.*, 1998a; Jobling, *et al.*, 1999). Therefore, this Y specific minisatellite varies in both the number of allele repeats and the interspersion of allele sequences. Diversity for the locus has been calculated at 99.9%. This degree of diversity in MSY1 makes the locus more discriminative then haplotypes derived from ten microsatellite loci (Jobling, *et al.*, 1998a).

The digital codes[2] and modular structures[3] represented in MSY1 can be characterized using minisatellite variant repeat PCR (MVR-PCR). MVR-PCR has been used to sequence or code several autosomal systems (Tamaki, *et al.*, 1993) and has made its way into forensic studies (Jeffreys, *et al.*, 1991; Hopkins, *et al.*, 1994). A four-state MVR-PCR for MSY1 has been designed by Jobling, *et. al.*, 1998a. The locus can be mapped bidirectionally creating complimentary digital codes (refer to Figure 2c). Four forward mapping and three reverse mapping discriminative primers were constructed with 3' homology to the specific repeat sequence. To increase specificity, the primer sequences were constructed slightly longer than the repeat unit and a 5' 'tag' sequence was added to each discriminator to maximize annealing temperatures (Jeffreys, *et al.*, 1991). 5' 'tag' addition served to increase the primer annealing temperatures by 2 °C for each A/T addition and by 4 °C for each G/C addition. Primer sequences are represented in Figure 2d. Four separate reactions, each containing a flanking primer and one specific ^{32}P labeled discriminator, are amplified in the presence of a purified MSY1 template. Digital codes are manually interpreted from autoradiographs in four separate lanes.

The current methods of detection and analysis do not make MSY1 a practical system in forensic casework. As previously stated, the degree of degradation and low concentration of DNA in forensic samples makes the amplification of the MSY1 problematic. Digital codes obtained from degraded samples yield uninformative results with only a few repeats typed from

---

[2]Digital or repeat codes are represented numerically by the type of repeat (type-1through type-4 respectively) followed by the frequency of the repeat in parenthesis. Thus, a 1(15) 3(39) 4 (19) refers to a sequence of MSY1 which reads as follows: 5'....15 type-1....39 type-3.... 19 type-4....3'.

[3]Modular structure is coded as the sequence of repeat types within the MSY1 array without quantitation of the number of repeats. Thus, the modular structure of a MSY1 code of 1(32) 3(2) 1(2) 3(32) 0(12) 4(2) would be represented as 1:3:1:3:0:4.

each end of the locus (Jobling, *et al.*, 1997). Increased sensitivity is required for samples of such nature. The pressures of high case loads experienced by major crime laboratories requires forensic systems to be semi-automated for high throughput. The analysis of MSY1 can require up to seven separate reactions per sample. Therefore, there is a limit to the number of forensic samples which can be batched. The use of MSY1 as a discriminative forensic system is also hindered by the lack of knowledge pertaining to the distribution of MSY1 in urban populations. Urban populations best reflect populations encountered by most forensic laboratories (Jobling, *et al.*, 1997). Without population data from these multi-cultural settings, the frequencies of particular MSY1 types and their respective modular structures can not be assessed.

The present study describes the methods employed in the development of a forensic based system for the analysis of MSY1. The main objectives of the study were: (1) to develop a fluorescent based semi-automated system for the detection of minisatellite variant repeat types, (2) to develop an MVR-PCR multiplex for the amplification of MSY1 to accommodate high throughput, (3) to optimize the initial MSY1 amplification to address system sensitivity related to forensic samples, and (4) to investigate the diversity of MSY1 in a multi-cultural population of New York City. These objectives are accomplished through the analysis of approximately 200 postmortem bloodstain samples from subjects grouped according to four major ethnic distinctions; Caucasian, African American, Asian, and Hispanic. The population study will be used to determine the possible presence of conserved MSY1 structures within each of the four major ethnic categories studied. Population databases assume that populations exhibit random mating. Although we do not choose our mates according to their genetic make-up, non-random mating based on physical attributes, religion, and racial preferences does exist (NRC, 1996).

Non-random mating results in population substructures which may not significantly effect autosomal systems but can have adverse effects on highly conserved haploid systems like the Y chromosome. The extent of population stratification and genetic variation will be investigated through the analysis of molecular variance (AMOVA) and F-statistic (Werner, P., http://online.sfsu.edu/~efc/classes/biol710/amova/amova.htm, Roewer *et al.*, 1996). Though the diversity of MSY1 codes is expected to be high, the diversity of modular structures are unknown. The possibility of a conserved modular structure identifying a specific ethnic group will be investigated. The findings concluded from this population study will be compared to global MSY1 studies and haplotype data derived from nine Y specific microsatellite loci.

Recent studies have addressed the analysis of MSY1 using fluorescent based semi-automated systems (Dupuy, *et al.*, 2001). However, these studies have not demonstrated the analysis of MSY1 in true forensic samples. The present study will address this issue through the analysis of non-probative evidence from five sexual assault cases. The results generated will be used to evaluate the use of the MSY1 locus on authentic forensic samples. The probative nature of MSY1 can not be disputed, however, the system employed in its detection must reflect the needs of a modern forensic biology laboratory to be accepted in forensic casework.

Section II: Project Design-Materials and Methods

Sample Collection and Ethnic Grouping

Blood samples were collected during autopsies of victims associated with violent crimes at the New York City Office of the Chief Medical Examiner. Postmortem blood samples collected during autopsies fell under the heading of "waste tissue" and therefore could be used for research purposes. Proper consent was obtained from the chief medical examiner and the attorney representing the New York City Medical Examiner office. The samples were collected between the years 1992 through 2000. Dried bloodstains were previously prepared by placing approximately 1 mL of postmortem blood onto UltraStain™ cards (GibcoBRL). The samples were allowed to dry under a biological fume hood overnight then stored at 4°C indefinitely. Blood samples were grouped into one of the four major ethnic categories; Caucasian, African American, Hispanic, and Asian. Ethnic data was obtained from descriptive questionnaires completed by relatives of the victims. Unidentified victims were placed into a specific ethnic group based upon physical characteristics, when possible. Samples which did not have a clear ethnic distinction were not used in this study.

DNA Extraction-bloodstains

Two procedures were investigated for the extraction of DNA from bloodstains. A Chelex-100 based extraction (Walsh, *et al.*, 1991) was compared to a phenol:chloroform:isoamyl alcohol (P:C:I) extraction. Both methods were modifications of standard protocols commonly used by forensic laboratories (FBI; OCMEa).

a. Chelex-100 DNA Extraction

1. Label sterile 1.5 mL centrifuge tubes with the unique sample identifiers. (Include a reagent blank for each extraction set).
2. Place a 4 x 6 mm clipping of a bloodstain sample into the appropriately labeled 1.5 mL centrifuge tube.
3. Add 1 mL of sterile deionized water (dH_2O), vortex for 5 - 10 sec. and incubate at room temperature for 30 min. Vortex samples every 5 min. during the lysis step.
4. Centrifuge samples for 5 min. at 14,000 rpm (13,000 x g).
5. Remove all but 25 μL of the supernatant without disturbing the pellet.
6. Add 75 μL of 20% Chelex-100 (BioRad) to each sample and vortex for 5 - 10 sec.
7. Incubate samples for 30 min. at 56°C.
8. Transfer samples into a 100°C heat block and incubate for an additional 8 min.
9. Centrifuge samples for 5 min. at 14,000 rpm (13,000 x g).
10. Prepare dilutions of the extracted samples at concentrations of 1:10 and 1:100 in sterile dH_2O for DNA quantitation.

b. Phenol:Choloroform:Isoamyl Alcohol [P:C:I 25:24:1] Extraction

1. Label sterile 1.5 mL centrifuge tubes with the unique sample identifiers. (Include a reagent blank for each extraction set).
2. Place a 4 x 6 mm clipping of a bloodstain sample into the appropriately labeled 1.5 mL centrifuge tube.
3. Create an extraction master-mix for n + 2 samples as follows:

300 μL Organic Extraction Buffer *(Tris 10 mM, pH 8.0, EDTA, Disodium Salt, Dihydrate 50 mM, pH 8.0, NaCl 100 mM)*

7.5 μL 20% SDS [1.5% ^{V}/v]

10.2 μL 20 mg/mL (20 U/mg) Proteinase K [4 U]

4. Add 300 μL of extraction mastermix to each sample and vortex for 10 - 15 sec.
5. Incubate samples in a rotating water bath for 4 - 24 hrs. at 56°C.
6. Centrifuge tubes for 1 min. at 14,000 rpm (13,000 x g) to remove condensate from the top of each tube.
7. Add 200 μL of Phenol:Choloroform:Isoamyl Alcohol [25:24:1] (P:C:I) (GibcoBRL). Vortex samples for 10 - 15 sec. Until a milky emulsion is formed.
8. Centrifuge samples for 5 min. at 14,000 (13,000 x g) rpm.
9. Insert a Microcon-100 (Millipore) concentrator column into a labeled tube for each sample.
10. Add 100 μL of hot (80 - 90°C) sterile dH_2O to the filter side of each concentrator.
11. Remove the aqueous phase (top layer) from each tube and transfer it to the appropriate concentrator. Do not draw up the proteinaceous interface.
12. Spin the Microcon-100 concentrators for 20 min. at 2,500 rpm (400 x g). (If additional spin times are necessary, increase speed to 3,000 rpm (600 x g) for an additional 20 min.
13. Discard wash from the lower chamber and return the filtrate cup to the concentrator.
14. Add 400 μL of hot (80 - 90°C) sterile dH_2O to the filter side of each concentrator.
15. Spin the Microcon-100 concentrators for 20 min. at 2,500 rpm (400 x g). (If additional spin times are necessary, increase speed to 3,000 rpm (600 x g) for an additional 20 min.
16. Label the appropriate number of fresh retention tubes.

17. Add 40 µL of sterile dH_2O to the filter side of each concentrator.

18. Invert sample reservoirs into the newly labeled retention tubes.

19. Spin the Microcon-100 concentrators for 3 min. at 3,500 rpm (800 x g) to collect the samples.

20. Discard sample concentrators and adjust sample volume to 100 µL.

21. Prepare dilutions of the extracted samples at concentrations of 1:10 and 1:100 in sterile dH_2O for DNA quantitation.

DNA Extraction-sperm cells

a. Sperm Cell Titration Assay

A sperm cell titration assay was designed to test the detection limits of the extraction procedure for mixed samples in the present study. Sperm cells obtained from a control subject were quantitated using a Hemacytometer (Clay Adams). Sperm cell dilutions were prepared in a final volume of 100 µL with phosphate buffered saline (PBS) [137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4-$7H_2O$, 1.4 mM KH_2PO_4] at the following concentrations: 200ng, 150 ng, 100 ng, 50 ng, 20 ng, 10 ng, and 5 ng of sperm cell DNA. Samples were prepared by adding the sperm dilutions to pre-prepared vaginal cavity swabs and dried overnight. A second set of dilutions were prepared on swabs lacking vaginal epithelial cells. The P:C:I based non-differential extraction procedure was performed as follows:

1. Label sterile 1.5 mL centrifuge tubes with unique sample identifiers. (Include a reagent blank and female negative control).

2. The entire swab was clipped and placed into the appropriately labeled 1.5 mL centrifuge

tube.

3. Add 1 mL of sterile dH_2O into each tube.

4. Place samples into a thermomixer set at room temperature and agitate samples for 30 min.

5. Label a new set of centrifuge tubes during the incubation step.

6. Flame the tip of a sterile 25 gauge needle for approximately 10 sec. and then puncture a hole in the base of the first sample tube. "Piggyback" the sample onto the appropriately labeled centrifuge tube. Repeat for the remainder of the samples using a fresh needle for each tube.

7. Spin the samples at 3,000 rpm (600 x g) for 5 minutes to isolate the supernatant from the swab. Discard the original tube containing the dried swab.

8. Create an extraction master-mix for n + 2 samples as follows:

200 µL	Organic Extraction Buffer *(Tris 10 mM, pH 8.0, EDTA, Disodium Salt, Dihydrate 50 mM, pH 8.0, NaCl 100 mM)*
2 µL	20 mg/mL (20 U/mg) Proteinase K [0.8 U]
10 µL	1M Dithiothreitol [50 mM]

9. Add 200 µL of extraction master-mix to each sample and vortex for 10 - 15 sec.

10. Incubate samples in a rotating water bath for 2 hrs. at 56°C. (Note: Overnight incubation will lead to sample degradation).

11. Centrifuge tubes for 1 min. at 14,000 rpm (13,000 x g) to remove condensate from the top of each tube.

12. Add 200 µL of Phenol:Choloroform:Isoamyl Alcohol [25:24:1] (P:C:I). Vortex samples for 10 - 15 sec. until a milky emulsion is formed.

13. Centrifuge samples for 5 min. at 14,000 rpm (13,000 x g).

14. Insert a Microcon-100 concentrator column into a labeled tube for each sample.
15. Add 100 µL of hot (80 - 90°C) sterile dH_2O to the filter side of each concentrator.
16. Remove the aqueous phase (top layer) from each tube and transfer it to the appropriate concentrator. Do not draw up the proteinaceous interface.
17. Spin the Microcon-100 concentrators for 20 min. at 2,500 rpm (400 x g). (If additional spin times are necessary, increase speed to 3,000 rpm (600 x g) for an additional 20 min.
18. Discard wash from the lower chamber and return the filtrate cup to the concentrator.
19. Add 400 µL of hot (80 - 90°C) sterile dH_2O to the filter side of each concentrator.
20. Spin the Microcon-100 concentrators for 20 min. at 2,500 rpm (400 x g). (If additional spin times are necessary, increase speed to 3,000 rpm (600 x g). for an additional 20 min.
21. Label the appropriate number of fresh retention tubes.
22. Add 15 µL of sterile dH_2O to the filter side of each concentrator.
23. Invert sample reservoirs into the newly labeled retention tubes.
24. Spin the Microcon-100 concentrators for 3 min. at 3,500 rpm (800 x g) to collect sample.
25. Discard sample concentrators and adjust the sample volume to 20 µL. (Note: The final concentrations of the titration samples are assumed to be 100 ng, 75 ng, 50 ng, 25 ng, 10 ng, 5 ng, 2.5 ng per 10 µL respectively).

b. Non-probative Case Study

Five cases were chosen for the present study. Each case was classified as "non-probative casework" since the defendants have been tried and convicted in the criminal courts of New York City. Therefore, this post-conviction biological evidence no longer carries probative value.

Representative samples from the most frequently encountered casework samples were chosen. These included two internal cavity swabs, an external swab from the victim's body, and two non-mixed samples recovered from crime scene evidence. For each of the cases described above bloodstain exemplars (known blood standards) from the suspects in question were processed. The unique identifiers and the sample descriptions for each non-probative item are depicted in Table 1. DNA from bloodstain exemplars (N6-N8) was obtained using the P:C:I extraction procedure outlined in the *DNA Extraction-bloodstain* section. Semen-positive case evidence (N1-N5) was extracted using the non-differential extraction procedure outlined in the *DNA Extraction-sperm cells* section.

Estimation of DNA Quantity from QuantiBlot Analysis (Perkin Elmer Corp., 1996; OCMEb)

a. Sample Blotting

1. Vortex all samples including DNA standards and calibrators #1 and #2. Centrifuge briefly to bring the contents to the bottom of the tube.
2. Label enough microfuge tubes for all samples and standards.

   Pipet samples and standards into the microfuge tubes, using the following amounts of each:

   a. DNA Standards and Calibrators - 5 μL
   b. Extracted samples- 1/10 and 1:100 dilutions in sterile distilled water (with a final volume of 20 μL).
3. Heat a shaking water bath to 50°C. Heat a stationary water bath to between 37°C and 50°C. Warm the QuantiBlot Hybridization Solution [5X SSPE (5 mM EDTA, 20 mL 10 N NaOH, 50 mM Sodium Phosphate, Monobasic, 0.9 M NaCl), 0.5% SDS] and the QuantiBlot Wash

Solution [2.5X SSPE, 0.1% SDS] in the water bath.

4. Add 150 µL of Spotting Solution [0.00008% Bromothymol Blue in 75 mL Pre-Wetting Solution (0.4 N NaOH, 25 mM EDTA)] to each tube. Vortex and centrifuge briefly to bring the contents to the bottom of the tube.

5. While wearing gloves, cut a piece of Biodyne B membrane to 11.5 x 7.9 cm. Cut a small notch in the upper left corner to mark its orientation. Place the membrane in a container containing 50 mL of Pre-Wetting Solution and incubate at room temperature for 1-30 minutes.

6. Using forceps, remove the membrane from the Pre-Wetting solution. Place the membrane on the gasket of the slot blotter, then place the top plate of the slot blotter on top of the membrane. Turn on vacuum pump to a vacuum pressure of approximately 200 to 250 mm Hg. Turn off the sample vacuum and turn on the clamp vacuum on the slot blot apparatus. Push down to ensure a tight seal.

7. Using a new pipete tip for each sample, apply all of each sample into the wells of the slot blotter in the following order:

Slot	Sample	Slot	Sample
1A	10 ng standard	1G	0.15625 ng standard
1B	5 ng standard	1H	plate negative control (sterile dH_2O)
1C	2.5 ng standard	2A	3.5 ng Calibration 1 Std.
1D	1.25 ng standard	2B	0.5 ng Calibration 2 Std.
1E	0.625 ng standard	2C	0.15625 ng standard
1F	0.3125 ng standard	2D-6H	samples and controls

8. After all the samples have been applied, slowly turn on the sample vacuum. Leave the sample vacuum on until all samples have been drawn through the membrane. Inspect each slot that contains a sample for a uniform blue band. Turn off the sample vacuum, the clamp vacuum, then the vacuum source.
9. Disassemble the slot blotter and remove the membrane. Proceed immediately to pre-hybridization. Do not allow the membrane to dry out.
10. Transfer the membrane to 100 mL of pre-warmed QuantiBlot Hybridization Solution in the hybridization tray (Applied Biosystems). Add 5 mL of 30% H_2O_2. Place the lid on the tray. Put the tray into the 50°C shaking water bath. Place a weight on the covered tray to prevent the tray from sliding or floating. Shake at 50°C for 15 min. at 50-60 rpm. Pour off the solution.

b. Hybridization

11. Add 30 mL of pre-warmed QuantiBlot Hybridization Solution to the tray. Tilt the tray to one side and add 20 µL of QuantiBlot D17Z1 Probe to the QuantiBlot Hybridization Solution. Cover tray with lid and weight. Shake at 50°C for 20 min. at 50-60 rpm. Pour off the solution.
12. Add 100 mL of pre-warmed QuantiBlot Wash Solution to the tray. Rinse by rocking for several seconds, then pour off the solution.
13. Add 30 mL of pre-warmed QuantiBlot Wash Solution to the tray. Tilt the tray to one side and add 180 µL of Enzyme Conjugate (Applied Biosystems). Cover tray with lid and weight. Shake at 50°C for 10 min. at 50-60 rpm. Pour off the solution.

14. Add 100 mL of QuantiBlot Wash Solution to the tray. Rinse by rocking for 1 minute, then pour off the solution. Repeat for a total two washes.

15. Add 100 mL of QuantiBlot Wash Solution to the tray. Cover tray with lid and weight. Shake at room temperature for 15 min. at 100-125 rpm. Pour off the solution. During this time, prepare the Color Development Solution.

c. Color Development

16. In a glass flask, prepare the Color Development Solution. Add the reagents in order:

60 mL of Citrate Buffer	[60 mM Trisodium Citrate, 25 mM Citric Acid, pH 5.0]
3 mL Chromogen	[3,3',5,5'-Tetramethylbenzidine in 30 mL 100% ETOH] (Applied Biosystems)
60 μL 3% H_2O_2	

Mix thoroughly by swirling (do not vortex). Note: Do not prepare the Color Development Solution more than 10 minutes before use.

17. Add 100 mL of Citrate Buffer to the tray. Rinse by rocking for several seconds, then pour off the solution.

18. Add the Color Development Solution to the tray. Cover tray with lid. Develop the membrane by shaking at room temperature for 20-60 min. at 50-60 rpm. Pour off the solution.

19. Stop the color development by washing in approximately 100 mL deionized H_2O. Repeat several times. After the last wash, store membrane in deionized H_2O. Cover tray with lid and proceed with photography.

d. Photography

20. Photograph the membrane while wet. Place the membrane on a dark, flat, non-absorbent surface.
21. Use a Polaroid MP4 camera system with type 667 or 664 film and a Wratten 23A or 22 (orange) filter.
22. Turn on the flood lights. Adjust the height of the camera and focus so that the membrane fills the entire viewing frame.
23. Photograph at 1/125 seconds and f8 for type 667 film. Photograph at 1/2 second and f5.6 for type 664 film. Develop the film for 30-60 seconds. Use photo for DNA concentration interpretations.

Amplification of MSY1

Amplification conditions of MSY1 defined by Jobling, *et al.*, 1998a were compared to amplification conditions in the present study. Several factors were varied to achieve optimal amplification of forensic evidence and postmortem bloodstains. Theses factors included: (1) DNA polymerase comparisons based upon their fidelity; Ampli*Taq* Gold DNA polymerase (Applied Biosystems) vs. *Taq* 2000™ DNA Polymerase (Stratagene) vs. *Taq*Plus® Long DNA Polymerase (Stratagene), (2) a modified PCR buffer system, (3) varied concentrations of template DNA , (4) amplification cycle number, and (5) reaction final volumes.

a. Amplification Conditions as per Jobling, et al, 1998a

100 ng of P:C:I extracted control DNA (bloodstain cards prepared from live male

donors) was amplified in a PCR buffer system containing: 1X PCR Buffer System-2 (Applied Biosystems), 45 mM Tris-HCl (pH 8.8), 11 mM $(NH_4)_2SO_4$, 6.7 mM β-Mercaptoethanol, 4.5 μM EDTA, 110 μg/mL BSA (Sigma), 4.5 mM $MgCl_2$, and 1 mM each dNTP (Applied Biosystems). MSY1 flanking primers Y1A and Y1B (refer to figure 2d for primer sequences) were used at a concentration of 1 μM each. To this reaction 5U of Ampli*Taq* Gold DNA Polymerase was added. The final reaction volume of 10 μL was amplified in an Applied Biosystems GeneAmp® PCR System 9700 thermocycler set for 9600 thermocycling speed. Amplification was conducted under the following conditions:

Hot start:	96°C, 11 min.
Cycle:	95°C, 2 min. 66°C, 3.5 min. 25 cycles
Link: 4°C ∞	

Amplified fragments were separated on a 1.0 % DNA typing grade agarose gel (Sigma) using a 1X TBE buffer (GibcoBRL); electrophoresis was conducted at 0.4 V/cm^2 for approximately 45 min.; DNA fragments were visualized under ultra violet light (365 nm).

b. Optimization of MSY1 Amplification

The MSY1 amplification buffer was optimized for the use of Ampli*Taq* Gold DNA polymerase. The initial PCR buffer was designed as follows: 1 μM of each flanking primer Y1A and Y1B, 1X PCR Buffer System-2, 5 U Ampli*Taq* Gold DNA polymerase, 200 μM each dNTP, 1.5 mM $MgCl_2$, 160 μg/mL BSA. Varying amounts of template DNA (25 - 50 ng) were amplified in a final reaction volume of 10 μL. Amplification was carried out as described in the

Amplification Conditions as per Jobling, et. al., 1998 section above. Amplified fragment yields were improved by increasing the number of cycles to thirty. Amplification sensitivity was tested using four concentrations of DNA; 25 ng, 10 ng, 5 ng, and 1 ng. Specificity was tested using four types of DNA polymearases; Ampli*Taq* Gold DNA polymerase, *Taq* 2000™ DNA Polymerase, and *Taq*Plus® Long DNA Polymerase. *Taq* 2000™ DNA Polymerase displayed the highest specificity and sensitivity. Amplification was present down to 1 ng of DNA in the presence of 160 μg/mL BSA. These results could not be duplicated with new lots of *Taq* 2000™, therefore, Ampli*Taq* Gold DNA polymerase, with inherent hot-start activity, was chosen for the final optimization of MSY1 amplification. The PCR buffer consisted of the following: 1 μM of each flanking primer Y1A and Y1B, 1X PCR Buffer System-2, 5 U Ampli*Taq* Gold DNA polymerase, 200 μM each dNTP, 1.5 mM $MgCl_2$, 200 μg/mL BSA. The final reaction volume was increased to 20 μL to accommodate increased template addition without the need for sample concentration. With these differences, amplification was carried out as described in the *Amplification Conditions as per Jobling, et. al., 1998* section above. Amplified fragments were visible down to 2.5 ng of extracted DNA making the system practical for forensic studies since the quantity of DNA isolated from biological evidence is often in this range.

c. Isolation of MSY1 from 1.0 % Agarose gels

Amplified fragments were separated on a 1.0 % agarose (DNA typing grade) gel using a modified 1X TBE buffer containing: 40 mM Tris-acetate, pH 8.0, 0.1 mM Na_2 EDTA (Sigma); electrophoresis was conducted at 0.4 V/cm^2 for approximately 45 min.; DNA fragments were visualized under ultra violet light. Gel slices containing MSY1 were cut out using a sterile single edge razor and placed into an Ultrafree®-DA purification column (Millipore). The DNA fragments were purified by centrifugation at 5,000 x g for 10 min. The column was discarded and the final sample volume was adjusted to 50-250 µL with sterile dH_2O. The final sample volume was dependant upon amplification efficiency.

d. Optimization of MSY1 MVR-PCR (separate reactions)

Initial MVR-PCR was performed using up to four separate reactions, each containing a specific fluorescent labeled discriminating primer (refer to Figure 2d). Forward mapping reactions were performed as follows: 100 nM unlabeled Y1A, one of four fluorescent labeled type discriminating primers (tag-1, tag-2, tag-3, tag-4) at 100 nM, 5 U *Ampli*Taq Gold DNA polymerase, 200 µM each dNTP, 1.5 mM $MgCl_2$ and 200 µg/mL BSA. Reverse mapping reactions were performed using the same concentration of reaction components but were substituted with 100 nM unlabeled Y1B primer and one of three labeled type discriminating primers (tag-1R, tag-3R, tag-4R). 2 µL of the purified MSY1 fragment was amplified in a final reaction volume of 10 µL. Amplification was performed in an Applied Biosystems GeneAmp® PCR System 9700 thermocycler set for 9600 thermocycling speed. Amplification was conducted according to Joblin et. al., 1998a as follows:

Hot start: 96°C, 11 min.

Primary Phase:
Cycle: 94°C, 8 sec.
64°C, 1 min.
68°C, 3 min.
3 cycles
Secondary Phase:
Cycle: 94°C, 8 sec.
68°C, 4 min. +4 sec./cycle
28 cycles

Link: 4°C ∞

e. Sample preparation, gel electrophoresis, fragment analysis

Reactions were dried to completeness in a Revco SpeedVac for 20 min. under low heat. A second set of reactions (forward and reverse, separately) were pooled and then dried to completeness. All samples were resuspended in 6 μL of blue dextran/formamide (5:1) buffer (Applied Biosystems), denatured at 95°C for three min., and then stored on ice until gel loading. Modular structures were characterized using an Applied Biosystems Prism™ DNA Sequencer. Electrophoresis was performed with 2.5 μL of product through a 2.5% acrylamide gel, 34-well square tooth comb, 36 cm well-to-read distance, under the following parameters: GeneScan Run36A-2400, 3000 V, 60 mA, 200 W, 51°C, laser power 40 mW, 2.5 hr. run. Fragment analysis was conducted using GeneScan® software version 2.1. To cancel fluorescent overlap emission, matrix files were created using fragment results from control experiments.

Modification to electrophoresis conditions were made to improve resolution of MSY1 detection into the 3' end of the repeat array. The well-to-read distance was increased to 48 cm using 52 cm gel plates (Applied Biosystems). Gels were prepared as follows:

18 g	Urea
5.0 mL	10X TBE
2.5 mL	50% Acrylamide (Fisher Brand)
20 mL	dH_2O

→Stir with mild heat until dissolved
→Adjust volume to 50 mL with sterile dH_2O
→Filter through a 0.2 μ filter
→Add 300 μL 10% Ammonium Persulfate (APS) and 40 μL of N,N,N',N'-Tetramethylethylenediamine (TEMED)

Electrophoresis was conducted at lower voltages and longer run times as follows: 3% acrylamide, SEQ Run 48A-1200, 2400 V, 50 mA, 200 W, 51°C, laser power 40 mW, 10.0 hr. run. Fragment analysis was conducted using GeneScan® software version 2.1.

f. Optimization of MSY1 MVR-PCR (multiplex)

Optimization of multiplex MVR-PCR began by adding 100 nM of each type discriminating primer and the appropriate unlabeled flanking primer to the reaction buffer described in the *Optimization of MSY1 MVR-PCR (separate reactions)* section. Thermocycling conditions were not altered. Preferential amplification of the type-1 repeat was present throughout the entire array with non-specific type-1 repeats far into the 3' end with no presence of type-2, 3, or 4 repeats (refer to Figure 3a). Difficulties with specificity were also observed in reverse mapping experiments. Type-1R and 3R repeats were specific to the array, however, low stringent type-3R fragments were present within the type-4R repeat block. Variations of annealing temperatures, primer concentrations, and PCR buffer concentrations did not eliminate non-specific amplification associated with the initial multiplex. Optimization proceeded with the combination of tag-1 and tag-4 in a duplex reaction. Six separate reactions were prepared with tag-4 at 100 nM and tag-1 at the following concentrations: 7.5 nM, 10 nM, 25 nM, 50 nM, 75 nM, and 100 nM. Optimal fluorescent intensity was present with a tag-1 concentration at 25 nM. Tag-3 was then added to the multiplex at concentrations of 7.5 nM, 10 nM, 25 nM, 50 nM, 75 nM, and 100 nM in six separate reactions containing 25 nM tag-1 and 100 nM tag-4. Tag-3 had an inhibitory effect on the presence of type-1 and type-4 repeats in the triplex. The inhibitory effect of tag-3 on the presence of type-4 repeats was eliminated at tag-3 concentrations below 75 nM. At tag-3 concentrations between 25 nM and 100 nM, non-specific type-3 repeats were present within the 5' type-1 repeat block. Below 25 nM tag-3, specificity was returned to the 5' type-1 repeat block, however, type-3 repeats were no longer present in the central repeat block. In an attempt to achieve triplex specificity, variations in primary phase annealing temperatures were performed. Three annealing temperatures and three tag-3 concentrations per temperature were investigated using the optimal concentrations of tag-1 and tag-4 discussed above. 10 nM, 25 nM and 50 nM concentrations of tag-3 were amplified at primary annealing temperatures of 66°C, 64°C, and 62 °C, with no improvements in triplex specificity. In separate reactions, 5% glycerol and varied concentrations of dimethyl sulfoxide (DMSO) were added in an attempt to disrupt the non-specific binding of tag-3. Addition of these denaturants led to complete destablization of the reaction. To attribute multiplex disruption to the presence of tag-3, varied concentrations of tag-2 (100 nM, 75 nM, 50 nM, 25 nM, 10 nM, and 2.5 nM) were amplified in the presence of 25 nM tag-1 and 100 nM tag-4 in six separate optimization reactions. The addition of tag-2 did not inhibit the presence of type-1 or type-4 specific repeat blocks, nor did the presence of tag-2 result in the presence of non-specific amplification throughout the entire repeat array. Optimal fluorescent intensities were observed with a 100 nM tag-2 concentration (refer to Figure 3b).

A FailSafe™ PCR optimization kit (EPICENTRE) was employed to address the difficulties encountered with the introduction of tag-3 into the multiplex. The FailSafe™ PCR kit consists of two components: (1) a high fidelity PCR enzyme mix and (2) twelve FailSafe PCR premixes which contain buffer, dNTPs, and various amounts of $MgCl_2$ and FailSafe™ PCR Enhancer (with bataine). The enhancer specifically functions to increase PCR specificity, sensitivity, and consistency (Grunenwald, 2001) A multiplex containing 25 nM tag-1, 100 nM tag-2, 25 nM tag-3, and 100 nM tag-4, and a separate control experiment lacking the tag-3 primer, were tested using the FailSafe™ PCR optimization kit. Cycling conditions were modified to meet manufactures specifications. This entailed the alteration of the initial hot start to a temperature of 98°C for 2 min. The remaining reaction followed the normal MVR-PCR cycling conditions described in the *Optimization of MSY1 MVR-PCR (separate reactions)* section. Both sets of reactions yielded incomplete forward mapping with the best set of results containing eight repeats from the 5' end of the array. Reverse mapping displayed non-specific artifacts throughout the array.

Inability to optimize the MVR-PCR multiplex led to the investigation of alternate forms of the tag-3 primer sequence. Five different variations of the NED-labeled tag-3 primer were designed. The goal of the experiment was to slightly decrease the stability of the tag-3 primer, thereby reducing the degree of non-specific tag-3 hybridization in the type-1 repeat block. Table 2 illustrates the primer designs and the melting temperature (Tm) associated with them. The base prior to the specific base position discriminating the type-1 repeat from the type-3 repeat was altered in two separate primer designs. Tag-3A was constructed by introducing a T→A transversion and tag-3B was constructed by introducing a T→G transversion. No significant Tm differences were expected, however, the substitutions were expected to cause a decrease in type-3 specificity for the type-1 repeat due to two consecutive base miss-matches with the type-1 repeat sequence. 5' and 3' deletions were introduced into primers tag-3C-E, also in an attempt to destabilize the primer structure. The 20 bp tag sequence incorporated into the 5' end of each discriminative primmer served to increase the primer Tm. This increase in Tm is required for the two step MVR-PCR amplification which combines the annealing and extension steps in a single 68°C step. At this elevated temperature and increased primer length, the flanking primer and the tag lengthened discriminator cannot bind internally within the products of previous cycles; a problem associated with the inherent AT rich MSY1 repeat units (Jobling, *et al.*, 1998a). PCR collapse is avoided at this elevated annealing temperature and increased primer length. By decreasing the primer length at the 3' end of the tag-3 primer (tag-3C-D), primer sensitivity would be affected during the primary phase of MVR-PCR. Reduction of non-specific tag-3 binding must be accomplished during this primary phase, thereby, minimizing the numbers of nonspecific templates available in the second phase of MVR-PCR.

The sensitivity and specificity of the newly constructed tag-3 discriminator primers were assessed through six separate MVR-PCR reactions using 100 nM of primer and 2 µL of positive control MSY1 template. The reaction conditions are specified in the *Optimization of MSY1 MVR-PCR (separate reactions)* section. The tag-3A, B and E primers yielded results specific to the positive control sample, however, at decreased sensitivity. No repeat units were present in the tag-3C and D primer reactions (refer to Figure 4a). Of the three functional primers, tag-3B displayed the highest sensitivity. All 37 type-3 repeats specific to the positive control were present well above fluorescent background levels, allowing for complete mapping of the central type-3 block. The complete central block was also present in the primer tag-3A and E reactions, however at lower resolution. Complete mapping would not be expected from forensic samples associated with varied degrees of degradation. Experimentation with tag3A, B and E was expanded into four separate reactions at concentrations of 25 nM, 50 nM, 100 nM, and 200 nM for each primer. Low levels of amplified repeats were present in all tag-3A and E reactions. The fluorescent signal for tag-3B was approximately three-times that of the tag-3A and tag-3E primer reaction, with optimal results present at 100 nM of tag-3B.

The interactions of primers tag-3A, B and E with tag-1 and tag-4 were investigated through multiplex MVR-PCR analysis. Tag-3A, B and E concentrations of 25 nM, 50 nM, 75 nM, and

100 nM were added to 25 nM tag-1 and 100 nM tag-4 in four separate reactions. Primary annealing temperatures of 58°C, 60°C, and 62°C respectively were investigated. [Previous multiplex MVR-PCR analysis of tag-3B at 64°C displayed high levels of non-specific type-3 repeats in the 5' type-1 repeat block]. Although no single reaction displayed results specific to the positive control repeat code, increased type-1 specificity was present at an annealing temperature of 62°C. The type-1 repeats displayed higher fluorescent emission as compared to the non-specific type-3 repeats, but this level of non-specificity resulted in ambiguous mapping of MSY1 (refer to Figure 4b).

The final MVR-PCR procedure was performed using as many as four separate reactions. Forward products were mapped using a triplex containing 25 nM tag-1, 100 nM tag-2, and 100 nM tag-4 and a single reaction containing 25 nM tag-3. Dependant upon the degree of sample degradation, samples displaying 3' signal drop-out were supplemented with a maximum of 200 nM tag-4. Reverse mapping was performed using two separate reactions containing 25 mM tag-3R and 100 mM tag-4R. Complete reverse mapping was not necessary beyond the type-3 type-4 junction.

Microsatellite Analysis

Y-STR data was generated for all population samples at nine microsatellite loci. Microsatellite data was generated according to Elmoznino and Prinz, (http://ystr.charite.de/text/locus_pcr2_3.html).

Data Analysis

a. Haploype Frequency and Genetic Diversity

The distribution of MSY1 in New York City populations was examined within each ethnic group and across all ethnic borders. Repeat codes, repeat numbers and modular structures were placed into a spreadsheet created in Corel Quattro Pro version 8.0, formatted to create all graphic representations. Haplotype frequencies were calculated according to Nei, *et al.*, 1973 and Nei, 1978, using the formula $p = x_i / n$, where x_i represents the number of occurrences of the i-th haplotype and n represents the total number of haplotypes. Diversity was calculated using Nei's unbiased estimator (Nei, *et al.*, 1973; Nei, 1978), using the following formula:

$$h = \frac{n}{n - 1} (1 - \Sigma p_i^2)$$

Σp_i^2 represents the summation of the squares of all haplotype frequencies and is equivalent to population homozygosity. Nei's unbiased estimator is used in place of the common method of estimating average heterozygosity, $\zeta = (1 - g)$, where g represents population homozygosity. Nei's unbiased estimator is used to subtract systematic bias introduced by small sample numbers (NEI, 1978). All haplotype frequencies and diversity indices were calculated using Arlequin 2.0 Population Genetics Data Analysis Software (Schneider, *et. al.*, 2000).

Microsatellite Y-STR data was organized and examined in the same manner as MSY1 data. The allele frequencies of the nine specific loci have previously been determined in Caucasian, African American, Asian, and Hispanic populations in New York City (Zawacki, 2001). The genctic diversity of both systems were compared to each other. Population samples with matching haplotypes were combined with MSY1 data to identify true or fortuitous matches. The combined system heterozygosity was calculated from this combined data. These results were used to determine the levels of Y chromosome diversity in New York City.

b. Analysis of Molecular Variance (AMOVA)

An AMOVA was performed to determine the genetic variance within and between all ethnic groups studied. AMOVA calculations included F-statistics (F_{ST}) which is a measure of the difference between the mean heterozygosity among subdivisions in a population. F_{ST} calculations ≥ 0.1 indicate differentiation between the overall population and its subdivisions. All calculations were performed using the Arlequin 2.0 genetic analysis package. Population data was entered according to Zawacki 2001, except for "population grouping" for AMOVA and F_{ST}. Caucasian, African American, Asian, and Hispanic population samples were grouped into a single population according to Schneider, *et. al.*, 2000. The calculated genetic inferences from both minisatellite and microsatellite data will be used to assess the structural requirements of a forensic MSY1 haplotype database in an "urban- like" population.

Section III: Results

Extraction and Amplification of MSY1

a. Bloodstain extraction

Postmortem bloodstains extracted using the Chelex-100 based procedure were compared to postmortem samples extracted using the P:C:I extraction procedure (refer to Figure 5). 25 ng of this DNA was amplified using the methods described in the *Optimization of MSY1 Amplification* section (refer to Figure 5a). A titration of P:C:I extracted control DNA was included to determine the detection limits of the method. In the P:C:I extraction experiment, amplified MSY1 fragments were present down to 2.5 ng of control DNA (refer to Figure 5b). Similar success rates of postmortem amplification were observed in both procedures, however, far higher concentrations of amplified MSY1 fragments were present in the P:C:I extraction procedure. Based upon these results, the P:C:I extraction procedure was used to extract all postmortem bloodstains.

b. Sperm cell extraction

Increased background was present in the control mixed sample extraction experiment when compared to the unmixed sperm cell extraction experiment (refer to Figure 6). This increase in background was attributed to the abundant presence of unamplified female DNA which coextracts with the male DNA. This is supported by the presence of increased background in the unmixed vaginal cell control and absence in the reagent blank. MSY1 amplification was present between the concentrations of 10 ng and 2.5 ng of sperm cell DNA, with low levels of amplification in the 2.5 ng control sample. In the unmixed sperm cell control extraction experiment, amplified fragments specific to MSY1 were present between the concentrations of 25 ng and 2.5 ng of sperm cell DNA, with low levels of amplification in the 5 ng and 2.5 ng control samples. Non-specific amplification products were present in all concentrations above 25 ng of sperm cell DNA. This degree of non-specificity at these concentrations of DNA was not noted in the bloodstain extractions. The precise cause of this non-specificity is unknown, but is unlikely to be the result of the extraction procedure. As a result, the mixed samples examined in the non-probative case study were minimized to ≤25 ng of DNA[4].

MSY1 Population Data Analysis

A total of 247 samples were amplified for MSY1: 60 Caucasian, 72 African American, 51 Asian, and 63 Hispanic. Of the 247 amplifications a total of 180 complete repeat codes were obtained: 47 Caucasian, 45 African American, 38 Asian, and 50 Hispanic. This equates to a 72.9% amplification success rate for postmortem bloodstains. The 180 allele codes are represented symbolically in Figures 7a-d. Symbol nomenclature was adopted from Jobling, *et al.*, 1998a, including null alleles which failed to code because of additional base substitutions not coded for in the four-state system (Jobling, *et al.*, 1998b). Allele length diversity is represented electrophoretically in Figures 8a-d. Figures 9a-d graphically represent the repeat number diversity within the four populations. Caucasian samples displayed repeat numbers between 38 and 91 repeats with the highest repeat frequency (n = 4) shared between 71, 73-75

[4]Sample N4 of the non-probative case study (refer to table 1 and the Forensic Applications section of the Discussions) contained an estimated sample concentration above 25 ng of male specific DNA. However, this concentration was presumed to be an overestimation due to bacterial degradation.

repeats. African American samples displayed the highest range of repeat numbers, ranging between 21 and 93 repeats with the highest repeat frequency (n = 54) shared between 55 and 58 repeats. Asian samples displayed the tightest range of repeat numbers with repeats between 58 and 89 repeats and 67 repeats being the most frequent at n = 5. Hispanic samples displayed repeat numbers between 48 and 84 repeats with the highest repeat frequency (n = 8) shared between 73 and 74 repeats. Figure 10a-b represents the New York City combined MSY1 repeat numbers for all four populations. The combined data displayed a repeat number range between 21 and 93 repeats for the 180 samples studied, with 60% of these samples residing within 66 and 78 repeats. African American population samples display a significant decrease in allele length as compared to Caucasian, Asian, and Hispanic populations, with a total of 23 samples (51.1%) concentrated between 55 and 62 repeats.

Of the 247 samples amplified for MSY1, 192 modular structures were obtained: 47 Caucasian, 53 African American, 39 Asian, and 53 Hispanic. This equates to a 77.7% rate for the determination of MSY1 modular structure in postmortem bloodstains. The examination of modular structure adds a higher level of discrimination to MSY1 as compared to length diversity (Jobling, *et al.,* 1998a). The modular structure of each sample along with the numerical representation of each repeat code is represented in Table 3a-d. Graphical representation of modular structure diversity is displayed in Figures 11a-d. A defined modular structure was present within the Caucasian, Asian, and Hispanic populations. Caucasian and Hispanic populations displayed the highest levels of structure similarity. The 1:3:4 modular structure was the most common with 23 (48.9%) in Caucasian and 24 (45.3%) in Hispanic population samples. The 3:1:3:4 modular structure was second most common structure between the two populations, with 11 (23.4%) in Caucasian and 14 (26.4%) in Hispanic population samples; observations approximately two-fold less than the most common 1:3:4 structure. This 3:1:3:4 modular structure was the most prevalent structure in Asian population samples with a total of 23 (60.0%). The second most common repeat structure was a 3:1:3:1:3:4 modular structure with a total of 6 (15.4%); a probable result of a mutational extension of the major 3:1:3:4 structure (Jobling, *et al.*, 1998a). The 1:3:4 and 3:1:3:4 modular repeats remain the most common structures observed in New York City when data from all four populations are combined. The 1:3:4 repeat structure was observed in 29.2% and the 3:1:3:4 was observed in 27.1% of combined New York City modular repeat types. The combined distribution of MSY1 repeats are represented graphically in Figure 12. A common repeat structure could not be determined in the African American population, however, a distinct 3' repeat characteristic was significantly present. The 3' repeat block was broken into three main structures: 3'-0, 3'-0:4, and 3'-3:4. The 3'-0 represent repeat codes which end in null alleles and the 3'-0:4 represents repeat codes which contain a string of null alleles prior to the final type-4 repeat block. Figures 13a-d represents the graphical examination of the 3' repeat block in all four populations. 36 (69.2%) African American samples contained the 3'-0:4 repeat block structure. Figures 14a-e is representative of two African American samples which displayed this distinct 3' fragment pattern. Four electropherograms and the corresponding gel slice view have been included. This specific structure was only present in 5 Caucasians, 1 Asian, and 8 Hispanics, a total of 14 within the three populations combined. Of the 50 combined samples grouped within the 3'-0:4 repeat block, African American samples accounted for 72% of the total. These results are consistent with data published by Jobling, *et al.*, 1998b, which determined the repeat code and modular structure of 16 African continent individuals: 8 Kenyan and 8 Nigerian samples.

New York City MSY1 repeat code diversity within each ethnic group was calculated using Table 3a-d. Samples with identical repeat codes were highlighted in boldface type. The results are represented in Table 4. Caucasian and African American population samples displayed the highest code diversity (h = 1) with no matching codes. Asian population samples displayed the second highest level of repeat code diversity with 37 different codes in 38 samples (h = 0.9986). Hispanic population samples displayed the lowest diversity with 48 different repeat codes in 50 samples (h = 0.9984). Combined MSY1 repeat code diversity was calculated using Table 5. Samples with identical repeat codes were highlighted in boldface type. The results are represented in Table 4. 170 different repeat codes were observed in a total of 180 samples. The diversity was calculated to be h = 0.9992, a high level of diversity expected from the urban population of New York City.

The genetic diversity of MSY1 modular structures was calculated according to repeat code data. Table 6 represents the total number of modular structures within each population and across all ethnic borders. The calculated diversities are listed in Table 4. African American population samples displayed the highest levels of modular structure diversity (h = 0.9376). Hispanic population samples displayed the second highest level of modular structure diversity (h = 0.7300), followed by Caucasians with a diversity of h = 0.7123. The lowest levels of modular structure diversity were observed in Asian population samples with a diversity of h = 0.6343. Overall modular structure diversity was calculated at h = 0.8358. This level of diversity is far greater than worldwide modular structure diversities determined by Jobling, *et al.* 1998a, results expected from urban populations of New York City.

Y-Chromosome Microsatellite Population Data Analysis

Amplification success rate was calculated using samples which produced complete haplotypes. Of the 247 samples amplified for the minisatellite nanoplex, a total of 240 complete haplotypes were obtained: 57 Caucasian, 68 African American, 52 Asian, and 63 Hispanic. This equates to an amplification success rate of 97.2% for postmortem bloodstains. The 247 haplotypes are arranged within each ethnic group according to allele number in Table 7a-d. Samples with identical haplotypes were highlighted in boldface type. Microsatellite diversity calculations are represented in Table 8. The highest levels of microsatellite diversity were observed in Caucasian population samples with 55 different haplotypes in 57 samples (h = 0.9987). Asian population samples followed Caucasians with a slightly lower diversity, displaying 50 different haplotypes in 52 samples (h = 0.9985). African American population samples displayed the third highest level of diversity with 64 haplotypes in 68 samples (h = 0.9978). The lowest levels of haplotype diversity were observed in Hispanic population samples, with 57 haplotypes in 63 samples (h = 0.9949). This low level of diversity may be due to the fact that a large number of New York City Hispanics originate from Puerto Rico. Combined haplotype diversity was calculated using Table 9. Samples with identical haplotypes were highlighted in boldface type. The combined haplotype diversity calculation is represented in Table 8. A total of 221 haplotypes were observed in 240 samples with a calculated diversity of h = 0.9991.

Haplotype Sharing

The degree of MSY1 code and microsatellite haplotype sharing was investigated within and among population groups (Table 10a-b). A total of three MSY1 types were shared within New York City populations; one pair within Asian samples and two pair within Hispanic samples. The sum of MSY1 code sharing increased to nine codes among all populations; the previous pair in Asian samples remained specific to Asians, one of the previous two Hispanic pairs remained specific to Hispanics, a total of five pairs among Caucasians and Hispanics, and two shared code structures (a pair and a trio) among African American and Hispanic samples. For microsatellite data, a total of ten haplotypes were shared within New York City populations; two pair within Caucasians, two pair and a trio within African Americans, two pair within Asians, and a pair, trio, and four haplotype share within Hispanic samples. The sum of microsatellite haplotype sharing increased by a single haplotype among all populations; both Caucasian haplotypes remained specific to Caucasians, one of the previous pairs and the trio in African Americans remained specific to this group, one of the previous pairs and the trio in Hispanics remained specific to this group, a pair among Caucasians and Hispanic populations, a pair and a six haplotype share among African American populations. The most interesting observation was the lack of haplotype sharing between Asians and the remaining populations. All shared haplotypes were grouped with their complimentary system (Table 11a-b) to access the power of discrimination of a combined MSY1/microsatellite nanoplex system. When data from both systems was available, all shared haplotypes were determined to be fortuitous.

Analysis of Molecular Variance (AMOVA)

The comparison of allele frequencies and genetic diversities do not take into account the structural relationships between haplotypes, and thus cannot infer genetic relatedness between groups of populations (Roewer, 1996). The number of mutational events separating haplotypes and the populations they reside in were calculated by AMOVA. For MSY1 modular structure data, AMOVA revealed a larger mean portion of within population (88.83%) than between-population (11.17%) genetic variance (Table 12a). Significant population differentiation was determined through F-statistics (F_{ST} = 0.11166) implying population stratification with some gene flow between populations. AMOVA of microsatellite haplotypes revealed significantly similar data with a larger mean portion of within population (89.89%) than between-population (10.11%) genetic variance (Table 12b). A slightly lower level of population differentiation (F_{ST} = 0.10108) was determined for microsatellite data, but the implications are the same as MSY1 population data. Lack of haplotype sharing between Asians and the remaining populations lead to the examination of population differentiation between the two groups. The highest levels of population differentiation (F_{ST} = 0.15753) was determined between Asian and the remaining New York City populations for MSY1 repeat codes. Relatively similar data was observed in microsatellite data with Asian populations (F_{ST} = 0.11893) insignificantly surpassed only by African Americans (F_{ST} = 0.12064).

Non-probative Case Studies

MSY1 amplification was successful in two of the five non-probative cases examined. A single internal orifice swab (N2-anal swab) and the external swab (N3-dried secretions from "left breast") yielded sufficient concentrations of amplified MSY1 fragments for isolation and MVR-PCR. Figure 15a represents the product gel containing the mixed samples and bloodstain exemplars. Table 13 displays the results of MVR-PCR. N2 and N3 shared a common repeat code which was consistent with the repeat code of the exemplar bloodstain, N6. These results are consistent with previous autosomal STR systems characterizing fifteen loci and one sex determining locus, amelogenin. Complete profiles were obtained from N2 and N3 using this autosomal system. In the MSY1 system a distinct characteristic was observed in the forward reaction of both mixed samples. N2 and N3 coding results displayed a decrease in fluorescence intensity of all type-1 and type-3 repeats prior to the 3'-0:4 terminal block code (Figures 15b and 15d). This decrease in fluorescence intensity is consistent with additional base substitutions within the type-1 and type-3 repeat sequence which, as previously stated, is not coded for in the present four-state MVR-PCR system (Jobling, *et al.*, 1998a; Jobling, *et al.*, 1998b). Identical results are present in N6, the included exemplar bloodstain. This repeat code characteristic adds another level of discrimination to the inclusion. The repeats which failed to code as the usual type-3 repeats in the forward reaction were not effected in the reverse mapping reaction (Figures 15c, e, and g).

Section IV: Discussion

The present study offers a novel approach to the analysis of the Y-specific minisatellite locus MSY1. The incorporation of a fluorescent based MVR-PCR triplex and modifications to standard gel electrophoresis conditions has resulted in a system delivering higher efficiency for the purposes of MSY1 population databasing and forensic studies. The current methods have reduced the number of reactions required to characterize MSY1 and have increased system sensitivity. The four reaction system described in this study was capable of amplifying MSY1 with as little as 2.5 ng of DNA; DNA concentrations frequently encountered in forensic casework.

MSY1 Code Diversity

The genetic distribution of MSY1 in urban populations of the United States were unknown prior to this study. The present study is the first to address this in four major ethnic divisions used in United States forensic DNA laboratory databases; Caucasian, African American, Hispanic, and Asian. Prior to this, high MSY1 code and structural diversities were assumed in urban populations, like New York City (Jobling, *et al.*, 1997). These multi-cultural populations are, for the most part, assumed to contain fewer mating constraints than are normally associated with isolated rural populations. As a result, Y-haplotype association within a specific ethnic group will break down. The population data from the present study does not support this theory since significant differentiation was observed between populations. A tight correlation was observed within MSY1 ($F_{ST} = 0.11166$) and microsatellite ($F_{ST} = 0.10108$) data, both displaying F-statistic values implying structural differentiation between populations placed into a single group.

In comparison to eight selected populations in Jobling, *et al.*, 1998a, New York City Caucasian and African American populations displayed the highest MSY1 code diversity while Asian ($h = 0.9986$) and Hispanic ($h = 0.9984$) populations displayed the least diversity, surpassed only by the British population displaying 39 unique codes in 40 samples studied ($h = 0.999$). These diversity calculations are most probably artificial and are expected to decrease if the size of the database was increased. New York City populations have a diversity of >0.999 as compared to >0.95 reported by Jobling, *et al.*, 1998a, with the exception of a South American Surui population ($h = 0.492$). Allele length diversity was also found to be less restricted with 60% of samples having between 66 and 78 repeats as compared to 80% having between 58 and 77 repeats reported by Jobling, *et al.*, 1998a. New York City population data contained a higher frequency of individuals who contained a total repeat number observed only once in the sampled population, spanning between 21 and 93 repeat numbers. Again the frequency of these "singlets" are presumed artificial due to the size of the database in the present study.

Haplotype Sharing

Significant haplotype sharing was observed within Caucasian and Hispanic MSY1 population data (Table 10b). These results correlated with New York City microsatellite data reported by Zawacki, 2001, further supporting the occurrence of Caucasian genetic admixture in Hispanic populations. The lack of microsatellite haplotype sharing between Caucasian and Hispanic populations in the present study is directly related to database size, with 196 fewer samples in the present study. Equal numbers of shared haplotypes are observed between African

American and Hispanic populations in both minisatellite and microsatellite data. This may be explained by the large contingency of New York Hispanics originating from Puerto Rico and the Dominican Republic (Zawacki, 2001). In comparison to data reported by Jobling, *et al.*, 1998a, five of the eight groups studied contained shared codes; one trio and four pairs. The code sharing in these particular groups was believed to be a result of population sub-structuring of the Y-chromosome. The sharing observed in New York City Caucasians and Hispanics cannot be explained by this sub-structuring but may be the result of "genetic urbanization" or homoginization of the Y-chromosome within these two ethnic groups. New York City Asians displayed the highest levels of genetic uniqueness with no shared haplotypes between populations in both minisatellite and microsatellite studies. F-statistic calculations supported this data with the highest levels of population differentiation determined among Asian (F_{ST} = 0.15753) and the remaining New York City populations for MSY1 repeat codes.

Modular Structure Diversity

The examination of modular structures offers a more informative method of estimating the diversity of MSY1 (Jobling, *et al.*, 1998a). The 1:3:4 and 3:1:3:4 structures were found to be the most common in New York City populations, with a 29.2% 1:3:4 observance and a 27.1% 3:1:3:4 observance. Similar results were observed in a survey of populations living on 5 continents as reported in Jobling, *et al.*, 1998a, and in Polynesians examined by Hurles, *et al.*, 1998. The 1:3:4 is found at high frequencies only in western Europeans and peoples of the Americas (Hureles, *et al.*, 1998). This data is strikingly similar to Caucasian and Hispanic populations of New York City (Figures 12a and 12d). The 1:3:4 structure is the most predominant structure observed within the two populations. As in population data from Europeans previously referenced above, the 3:1:3:4 structure is the only other significant structure observed within Caucasians and Hispanics. The 3:1:3:4 was found to be the most prevalent structure in New York City Asians (44.2%) as observed by Jobling *et al.*, 1998a. Although the highest 3:1:3:4 structural frequency was found in New York City Asians, the structure was not exclusive to this group. Relatively equal frequencies were found in Caucasian (21.2%) and Hispanic (26.9%) populations. This correlates with population data from all continents studied by Jobling *et al.*, 1998a. The present study revealed the lowest frequency of the 3:1:3:4 modular structure in African Americans (7.7%) as did African population data from Jobling *et al.*, 1998a.

In Jobling, *et al.*, 1998a, distinct modular structures were found in Africans as compared to all populations living on other continents. The prevalent 1:3:4 and 3:1:3:4 structures were rarer by proportion in this population. African derived Y-chromosomes contained repeat codes which lacked the normal type-1 and type-3 repeats and instead contained the type-2 repeat and other unknown nulls. Of the four New York City population samples containing the rare type-2 allele, three are specific to African Americans and the fourth is found in a Caucasian individual (C029 Figure 7a). The repeat code in this single Caucasian individual is more consistent with those observed in San and Zimbabwean populations reported by Jobling, *et al.*, 1998a, a finding which most probably reflects the problems associated with haplotype databases established in urban populations. In Jobling, *et al.*, 1998a and Jobling *et al.*, 1998b, samples containing a distinct string of null alleles prior to the final type-4 repeat block were prevalent in the Y-chromosomes of Africans. 69.2% of New York City African Americans displayed this distinct

3' repeat block. When combining all four populations, African Americans accounted for 72% of the total.

Minisatellite diversity has been used by anthropologists to study the ancestral relationships between human chromosomes. Where genetic diversity calculated by allele length provides little information on the relationships between specific alleles, the interspersion pattern of variant repeats along an allele can provide these relationships (Armour, *et al.*, 1996). This diversity in the minisatellite locus MS205 (D16S309) reported by Armour, *et al.* 1996, supports a recent African origin for modern humans. The assumption was based upon a clear difference in allelic diversity between African and non-African populations. Similar results were observed in African populations by Jobling, *et al.*, 1998a, who concluded that the highly diverged structures within this group suggests that they may represent deep branches within the Y-chromosomal phylogenetic tree. The larger portions of null alleles in African populations contributes to this degree of genetic diversity. African American populations in New York City display the highest levels of diversity (Table 4) and occurrences of null alleles when compared to Caucasian, Asian, and Hispanic populations. In comparison to global MSY1 structural diversity reported by Jobling, *et al.*, 1998a, the African American modular structure diversity was far greater than that reported in eight worldwide populations. The high degree of structural diversity and distinct repeat structures illustrate a clear African Y-chromosomal ancestry in New York City African Americans.

Forensic Applications

The examination of population samples from four New York City ethnic groups in the present study demonstrates the discriminative value of MSY1 in urban populations. The degree of diversity was calculated to be greater than nine microsatellite loci combined (Table 4). The amount of shared codes observed within such a small population size precludes MSY1 as a marker for genetic uniqueness but the degree of code and structural diversity makes the locus a powerful exclusionary tool. The numbers of shared haplotypes between populations was found to be comparable with microsatellite data (Table 10b). The current system has enabled the amplification of MSY1 in mixed samples with as little as 2.5 ng of sperm DNA. Optimal results were observed between the concentrations of 5 and 10 ng of sperm DNA. These values are comparable to the template concentrations used for Y-STR nanoplex databasing by Elmozino & Prinz (http://ystr.charite.de/text/locus_pcr2_3.html), but could not match the 250 pg minimal nanoplex amplification concentration in forensic samples. A complete four-state multiplex was unattainable, however, a four reaction system has been developed for the semi-automated detection of MSY1. This, opposed to a total of seven separate reactions specified by Jobling, *et al.*, 1998a, increases the throughput of MSY1 analysis; a critical requirement for forensic laboratories.

The non-probative case studies involving mixed samples demonstrated the degree of difficulty associated with the analysis of minisatellite DNA in forensic samples. The template length inherently makes the analysis of MSY1 in forensic samples problematic, especially when low concentrations of degraded DNA are involved. This difficulty was exemplified in the postmortem bloodstain study which resulted in an amplification success rate of 72.9%. Similar results were observed by Hureles, *et. al.*, 1998, with an 84.6% success rate in live individuals. Only two of the five non-probative mixed samples resulted in a complete MSY1 code type. The amplification failure observed in two of the four mixed samples may be attributed to sample degradation. Table 1 depicts the condition in which sample N4 and N5 were found. Regardless of the estimated concentrations of DNA calculated from prior analysis (refer to table 14), the presence of mold on the two samples reduced the likelihood of successfully characterizing MSY1. The discriminative power of MSY1 was demonstrated in the characterization of mixed samples N2 and N3. The results obtained from the unknown samples were linked to the blood exemplar N6 by three degrees of discrimination. The blood exemplar contained an identical allele length, an identical repeat code, and most importantly a distinct repeat characteristic, as compared to the unknown samples. The occurrence of reduced fluorescence emission was consistent with additional base substitutions within the allele which resulted in null alleles in the type-1 and type-3 repeat blocks prior to the final 3'-0:4 block. This characteristic added a higher level of uniqueness to the inclusion.

Although general structural characteristics were observed in all four populations studied, the characteristics were not entirely specific to each population. As a result ethnic inferences cannot be predicted from the MSY1 type of unknown DNA donors in forensic cases. Caucasian and Hispanics cannot be distinguished from each other due to their high degree of genetic similarity. The predominant 3:1:3:4 structure in Asians is precluded as an Asian specific marker since the 3:1:3:4 is also found significantly in Caucasians and Hispanics. The highly distinct repeat codes observed in African global MSY1 studies and New York City African Americans is also displayed in 28% of Caucasians, Asians, and Hispanics combined. These findings were expected in urban populations. However, this study presents sufficient data to show that the degree of homogenization previously speculated for multi-cultural populations does not exist in New York City populations.

Section V: Conclusions

The discovery of microsatellite DNA has restricted the developmental evolution of highly informative minisatellite DNA markers in forensic casework. Rightfully so, with the capability of examining several microsatellites in a single multiplex and their ease of interpretation, microsatellites deserve the respect that they have earned for their role in forensic identification. In mixed samples containing minimal concentrations of male specific DNA, Y-chromosomal STR systems have added to the arsenal of tools available to the forensic biologist. These systems are limited by their mode of inheritance and therefore, cannot be used to infer genetic uniqueness as autosomal STR systems do. However, when combining the minisatellite locus MSY1 with the microsatellite nanoplex, all shared haplotypes were determined to be fortuitous when data from both systems was available. In reality, determining the true degree of MSY1 haplotype frequency would require a population database with far more samples than described here. The creation of an MSY1 specific population database and a forensic system incorporating the characterization of MSY1 would provide a power of discrimination equivalent to that of eighteen microsatellite loci. This power of discrimination would strengthen the likelihood of inclusion using haplotype systems in forensic cases. The present study provides a more efficient approach to the analysis of MSY1 for both population and forensic studies. Reduction of the number of reactions required to characterize MSY1 and the development of a fluorescent based detection system with increased sensitivity can potentially lead the way for MSY1 research where the distribution of this highly informative locus remains unknown.

Section VI: Table of Results

Table 1: Non-probative case sample summary.

Sample Identifier	Sample description	Comment
N1	anal swab	-----
N2	anal swab	-----
N3	dried secretions from "left chest"	-----
N4	mattress top	mold present
N5	bed spread	mold present
N6	suspect bloodstain exemplar for samples N2-N4	-----
N7	suspect bloodstain exemplar for sample N1	-----
N8	suspect bloodstain exemplar for sample N5	-----

Table 1: Sample identifiers assigned to non-probative samples. N1-2 correspond to the mixed sample internal cavity swabs, N3 corresponds to the external body swab, N4-5 correspond to non-mixed samples recovered from crime scene evidence, and N6-8 correspond to bloodstain exemplars from suspects in question. The specific sample(s) the suspect's are in question for are marked in the sample description section.

Table 2: Modified Tag-3 sequences and associated Tm.

Primer	Sequence	Tm
Tag-1	[6-FAM] 5'-tca tgc gtc cat ggt ccg gaT GTG TAT AAT ATA CA**T** (C)AT GTA TAT TG-3'	77.8°C
Tag-3	[NED] 5'-tca tgc gtc cat ggt ccg gaT GTG TAT AAT ATA CA**T** (G)AT GTA TAT TG-3'	77.8°C
Tag-3A	[NED] 5'-tca tgc gtc cat ggt ccg gaT GTG TAT AAT ATA CA**A** (G)AT GTA TAT TG-3'	77.2°C
Tag-3B	[NED] 5'-tca tgc gtc cat ggt ccg gaT GTG TAT AAT ATA CA**G** (G)AT GTA TAT TG-3'	78.3°C
Tag-3C	[NED] 5'-tca tgc gtc cat ggt ccg gaT GTG TAT AAT ATA CAT (G)AT GTA TAT T -3'	76.6°C
Tag-3D	[NED] 5'-tca tgc gtc cat ggt ccg gaT GTG TAT AAT ATA CAT (G)AT GTA TAT -3'	76.4°C
Tag-3E	[NED] 5'—a tgc gtc cat ggt ccg gaT GTG TAT AAT ATA CAT (G)AT GTA TA TTG-3'	76.0°C

Table 2: Modified Tag-3 MVR-PCR primer sequences and associated Tm. Tag-1 and Tag-3 sequences were included as references. The primer base position discriminating the type-1 from the type-3 repeat is marked by parenthesis. The base prior to the discriminating position is represented in **boldface type**. Tag-3A was constructed by introducing a T→A transition prior to the discriminatory base. Tag-3B was constructed by introducing a T→G transversion prior to the discriminatory base. Tag -3C was constructed by the 3' deletion of the final G residue. Tag-3D was constructed by the 3' deletion of the final two TG residues. Tag-3E was constructed by the 5' deletion of the TC residues. All primers were purchased pre-labeled from Applied Biosystems.

Table 3a: Caucasian MSY1 repeat codes and modular structures. Data sorted by modular structures.

Sample ID	MSY1 Code	Modular Structure
C053	0(1) 1(9) 3(31) 0(15) 4(1)	0:1:3:0:4
C029	0(5) 2(7) 0(1) 2(6) 0(4) 4(49)	0:2:0:2:0:4
C040	0(1) 3(1) 1(21) 0(9) 3(26) 4(16)	0:3:1:0:3:4
C041	0(1) 3(1) 1(9) 3(25) 0(11) 4(1)	0:3:1:0:3:4
C017	1(19) 3(2) 1(4) 3(10) 1(2) 3(31) 4(19)	1:3:1:3:1:3:4
C009	1(15) 3(11) 1(2) 3(31) 4(18)	1:3:1:3:4
C008	1(23) 3(26) 1(1) 3(25) 4(16)	1:3:1:3:4
C024	1(14) 3(37) 4(20)	1:3:4
C054	1(15) 3(36) 4(19)	1:3:4
C002	1(15) 3(39) 4(19)	1:3:4
C031	1(16) 3(36) 4(21)	1:3:4
C050	1(16) 3(37) 4(20)	1:3:4
C034	1(16) 3(39) 4(18)	1:3:4
C012	1(16) 3(39) 4(19)	1:3:4
C013	1(16) 3(39) 4(20)	1:3:4
C016	1(16) 3(40) 4(19)	1:3:4
C011	1(16) 3(41) 4(18)	1:3:4
C022	1(16) 3(43) 4(17)	1:3:4
C037	1(17) 3(35) 4(22)	1:3:4
C025	1(17) 3(37) 4(23)	1:3:4
C056	1(17) 3(38) 4(20)	1:3:4
C032	1(20) 3(52) 4(15)	1:3:4
C027	1(21) 3(49) 4(18)	1:3:4
C046	1(21) 3(50) 4(17)	1:3:4
C006	1(22) 3(34) 4(15)	1:3:4
C045	1(22) 3(38) 4(20)	1:3:4
C007	1(22) 3(48) 4(14)	1:3:4
C028	1(22) 3(49) 4(18)	1:3:4
C047	1(22) 3(51) 4(16)	1:3:4
C018	1(22) 3(53) 4(14)	1:3:4
C039	1(18) 3(35) 4(1) 3(1) 4(16)	1:3:4:3:4
C020	1(14) 3(38) 4(1) 3(2) 4(1) 3(2) 4(18)	1:3:4:3:4:3:4
C052	3(1) 1(19) 0(3) 3(19) 4(12)	3:1:0:3:4
C023	3(1) 1(19) 0(17) 4(1)	3:1:0:4
C059	3(2) 1(3) 3(1) 1(5) 3(36) 4(19)	3:1:3:1:3:4
C038	3(4) 1(5) 3(1) 1(6) 3(38) 4(20)	3:1:3:1:3:4
C019	3(1) 1(17) 3(31) 4(20)	3:1:3:4
C042	3(1) 1(8) 3(41) 4(19)	3:1:3:4
C030	3(2) 1(17) 3(28) 4(20)	3:1:3:4
C048	3(2) 1(12) 3(40) 4(17)	3:1:3:4

Table 3a: Caucasian MSY1 repeat codes and modular structures (continued). Data sorted by modular structures.

Sample ID	MSY1 Code	Modular Structure
C026	3(2) 1(12) 3(59) 4(12)	3:1:3:4
C001	3(2) 1(17) 3(27) 4 (21)	3:1:3:4
C043	3(3) 1(13) 3(41) 4(21)	3:1:3:4
C060	3(3) 1(14) 3(41) 4(23)	3:1:3:4
C036	3(4) 1(13) 3(31) 4(19)	3:1:3:4
C014	3(4) 1(14) 3(31) 4(19)	3:1:3:4
C004	3(6) 1(11) 3(37) 4 (18)	3:1:3:4

Note: No results were obtained for samples C003, 005, 010, 015, 021, 033, 035, 044, 049, 051, 055 and 057-058.

Table 3a-d:

MSY1 Code = Type of repeat is followed by frequency in parenthesis. Thus, 1(15) 3 (39) 4 (19) refers to a sequence of MSY1which reads as follows: 5'...15 type 1...39 type 3... 19 type 4...3'.

Modular Structure = Sequence of repeat type without quantitation. Thus, 1(15) 3 (39) 4 (19) is expressed as 1:3:4.

(?) = total number of repeats could not be determined.

Identical repeat codes are represented in ***italicized bold face type***.

Table 3b: African American MSY1 repeat codes and modular structures. Data sorted by modular structures.

Sample ID	MSY1 Code	Modular Structure
AA052	0(1) 1(8) 3(32) 0(13) 4(2)	0:1:3:0:4
AA054	0(1) 1(19) 3(1) 1(3) 3(2) 1(2) 3(20) 4(24)	0:1:3:1:3:1:3:4
AA065	0(2) 1(13) 3(2) 1(3) 3(42) 4(22)	0:1:3:1:3:4
AA060	0(2) 2(7) 0(1) 2(6) 0(1) 4(?)	0:2:0:2:0:4
AA067	0(1) 2(22) 0(1) 4(55)	0:2:0:4
AA050	0(4) 2(7) 0(?) 4(2)	0:2:0:4
AA044	0(1) 3(1) 0(?)	0:3:0:4
AA046	0(1) 3(1) 0(?) 4(2)	0:3:0:4
AA042	0(1) 3(1) 1(6) 3(29) 0(10) 4(2)	0:3:1:3:0:4
AA028	0(1) 3(1) 1(7) 3(31) 0(16) 4(2)	0:3:1:3:0:4
AA016	0(1) 3(1) 1(8) 3(27) 0(16) 4(2)	0:3:1:3:0:4
AA003	1(9) 0(1) 1(1) 0(2) 1(1) 0(1) 1(1) 0(1) 1(1) 0(3)	1:0:1:0:1:0:1:0:1:0
AA020	1(11) 3(27) 0(15) 4(2)	1:3:0:4
AA007	1(16) 3(38) 0(1) 4(20)	1:3:0:4
AA034	1(17) 3(23) 0(17) 4(2)	1:3:0:4
AA025	1(8) 3(34) 0(11) 4(2)	1:3:0:4
AA023	1(1) 3(1) 1(8) 3(33) 0(15) 4(2)	1:3:1:3:0:4
AA033	1(1) 3(2) 1(5) 3(26) 0(15) 4(1)	1:3:1:3:0:4
AA012	1(1) 3(2) 1(5) 3(33) 0(15) 4(2)	1:3:1:3:0:4
AA030	1(1) 3(2) 1(6) 3(33) 0(13) 4(2)	1:3:1:3:0:4
AA035	1(1) 3(2) 1(7) 3(29) 0(15) 4(1)	1:3:1:3:0:4
AA024	1(1) 3(2) 1(7) 3(35) 0(15) 4(2)	1:3:1:3:0:4
AA031	1(1) 3(2) 1(1) 3(2) 1(8) 3(32) 0(16)	1:3:1:3:1:3:0
AA017	1(1) 3(2) 1(6) 3(2) 1(1) 3(29)0(16) 4(2)	1:3:1:3:1:3:0:4
AA038	1(21) 3(3) 1(4) 3(9) 1(2) 3(30) 4(24)	1:3:1:3:1:3:4
AA068	1(15) 3(37) 4(18)	1:3:4
AA039	1(15) 3(4) 0(?)	1:3:4
AA069	1(16) 3(38) 4(20)	1:3:4
AA013	1(16) 3(40) 4(18)	1:3:4
AA029	1(17) 3(40) 4(21)	1:3:4
AA010	1(17) 3(41) 4(14)	1:3:4
AA022	1(20) 3(31) 4(18)	1:3:4
AA047	3(2) 0(?) 4(2)	3:0:4
AA049	3(3) 0(?) 4(2)	3:0:4
AA051	3(3) 0(?) 4(2)	3:0:4
AA064	3(2) 1(7) 3(27) 0(14) 4(2)	3:1:3:0:4
AA056	3(2) 1(7) 3(32) 0(15) 4(1)	3:1:3:0:4
AA055	3(2) 1(7) 3(33) 0(14) 4(2)	3:1:3:0:4
AA002	3(3) 1(13) 3(36) 0(1) 4(16)	3:1:3:0:4
AA053	3(3) 1(5) 3(33) 0(14) 4(2)	3:1:3:0:4

Table 3b: African American MSY1 repeat codes and modular structures (continued). Data sorted by modular structures.

Sample ID	MSY1 Code	Modular Structure
AA037	3(3) 1(7) 3(33) 0(15) 4(2)	3:1:3:0:4
AA058	3(3) 1(8) 3(29) 0(16) 4(2)	3:1:3:0:4
AA001	3(4) 1(12) 3(1) 1(1) 3(27) 0(17) 4(13)	3:1:3:1:0:4
AA057	3(3) 1(5) 3(2) 1(1) 3(33) 0(16) 4(1)	3:1:3:1:3:0:4
AA070	3(3) 1(6) 3(1) 1(1) 3(25) 0(18) 4(1)	3:1:3:1:3:0:4
AA066	3(3) 1(7) 3(2) 1(1) 3(21) 0(16) 4(2)	3:1:3:1:3:0:4
AA040	3(3) 1(7) 3(2) 1(1) 3(26) 0(10) 4(2)	3:1:3:1:3:0:4
AA032	3(3) 1(7) 3(2) 1(1) 3(31) 0(15) 4(2)	3:1:3:1:3:0:4
AA026	3(2) 1(7) 3(8) 1(5) 3(3) 1(1) 3 (16) 0(12) 4(2)	3:1:3:1:3:1:3:0:4
AA021	3(1) 1(15) 3(27) 4(26)	3:1:3:4
AA072	3(2) 1(17) 3(35) 4(24)	3:1:3:4
AA061	3(3) 1(12) 3(35) 4(18)	3:1:3:4
AA071	3(3) 1(13) 3(40) 4(21)	3:1:3:4

Note: No results were obtained for AA04-006, 008-009, 011, 014-015, 018-019, 27, 36, 41, 43, 45, 48, 59 and 62-63.

Table 3c: Asian MSY1 repeat codes and modular structures. Data sorted by modular structures.

Sample ID	Repeat Code	Modular Structure
A036	0(2) 1(12) 0(2) 1(4) 0(38) 4(12)	0:1:0:1:0:4
A043	0(1) 1(16) 3(34) 0(7)	0:1:3:0
A024	0(1) 1(21) 3(33) 4(12)	0:1:3:4
A016	0(1) 3(2) 1(11) 3(26) 4(18)	0:3:1:3:4
A014	1(16) 3(36) 4(20)	1:3:4
A044	1(18) 3(40) 4(8)	1:3:4
A019	1(19) 3(38) 4(9)	1:3:4
A038	3(2) 1(15) 3(?) 0(?)	3:1:3:0
A022	3(5) 1(1) 3(1) 1(11) 3(1) 1(2) 3(32) 4(19)	3:1:3:1:3:1:3:4
A023	3(2) 1(19) 3(2) 1(1) 3(38) 4(13)	3:1:3:1:3:4
A042	3(3) 1(1) 3(1) 1(11) 3(41) 4(10)	3:1:3:1:3:4
A009	3(3) 1(2) 3(2) 1(10) 3(40) 4(9)	3:1:3:1:3:4
A051	3(3) 1(2) 3(2) 1(11) 3(42) 4(10)	3:1:3:1:3:4
A011	3(4) 1(14) 3(2) 1(1) 3(52) 4(11)	3:1:3:1:3:4
A017	3(5) 1(16) 3(16) 1(8) 3(38) 4(6)	3:1:3:1:3:4
A004	3(1) 1(11) 3(50) 4(11)	3:1:3:4
A008	3(1) 1(12) 3(51) 4(15)	3:1:3:4
A018	3(1) 1(13) 3(51) 4(18)	3:1:3:4
A039	3(1) 1(17) 3(39) 4(9)	3:1:3:4
A030	3(1) 1(18) 3(41) 4(7)	3:1:3:4
A025	3(1) 1(18) 3(41) 4(8)	3:1:3:4
A020	3(2) 1(15) 3(45) 4(13)	3:1:3:4
A040	3(2) 1(16) 3(37) 4(13)	3:1:3:4
A002	3(2) 1(16) 3(52) 4(8)	3:1:3:4
A005	3(3) 1(12) 3(49) 4(9)	3:1:3:4
A050	3(3) 1(15) 3(38) 4(12)	3:1:3:4
A003	3(3) 1(15) 3(45) 4(14)	3:1:3:4
A012	3(3) 1(16) 3(36) 4(14)	3:1:3:4
A013	3(3) 1(16) 3(38) 4(14)	3:1:3:4
A031	***3(4) 1(12) 3(52) 4(10)***	***3:1:3:4***
A034	***3(4) 1(12) 3(52) 4(10)***	***3:1:3:4***
A015	3(4) 1(12) 3(54) 4(10)	3:1:3:4
A033	3(4) 1(13) 3(25) 4(20)	3:1:3:4
A010	3(4) 1(13) 3(40) 4(10)	3:1:3:4
A041	3(4) 1(13) 3(41) 4(12)	3:1:3:4
A037	3(4) 1(14) 3(44) 4(13)	3:1:3:4
A032	3(4) 1(8) 3(54) 4(14)	3:1:3:4
A001	3(1) 1(18) 3(43) 4(5)	3:1:3:4
A035	3(3) 1(17) 3(39) 4(1) 3(2) 4(11)	3:1:3:4:3:4

Note: No results were obtained for A006-007, 021, 026-029 and 045-049.

Table 3d: Hispanic MSY1 repeat codes and modular structures. Data sorted by modular structures.

Sample ID	MSY1 Code	Modular Structure
H049	0(1) 1(16) 3(31) 0(1) 3(1) 0(2) 4(1) 0(9) 4(1)	0:1:3:0:3:0:4:0:4
H010	0(1) 3(1) 1(6) 3(33) 0(16) 4(1)	0:3:1:3:0:4
H033	0(1) 3(1) 1(7) 3(32) 0(16) 4(2)	0:3:1:3:0:4
H037	0(1) 3(2) 1(6) 3(1) 1(1) 3(18) 0(17) 4(2)	0:3:1:3:1:3:0:4
H063	1(19) 3(3) 0(4) 3(28) 4(23)	1:3:0:3:4
H052	1(19) 3(5) 0(1) 3(38) 4(17)	1:3:0:3:4
H057	1(11) 3(1) 1(1) 3(29) 0(18) 4(2)	1:3:1:3:0:4
H013	1(15) 3(2) 1(2) 3(39) 4(18)	1:3:1:3:4
H004	***1(20) 3(1) 1(3) 3(35) 4(25)***	***1:3:1:3:4***
H054	***1(20) 3(1) 1(3) 3(35) 4(25)***	***1:3:1:3:4***
H019	1(12) 3(43) 4(16)	1:3:4
H005	1(14) 3(37) 4(20)	1:3:4
H042	1(15) 3(?) 4(20)	1:3:4
H061	1(15) 3(28) 4(19)	1:3:4
H048	1(15) 3(38) 4(20)	1:3:4
H051	1(15) 3(39) 4(19)	1:3:4
H016	1(15) 3(39) 4(20)	1:3:4
H043	1(16) 3(?) 4(20)	1:3:4
H060	1(16) 3(34) 4(20)	1:3:4
H009	1(16) 3(36) 4(21)	1:3:4
H056	1(16) 3(37) 4(18)	1:3:4
H002	1(16) 3(37) 4(20)	1:3:4
H053	1(16) 3(38) 4(20)	1:3:4
H029	1(16) 3(39) 4(21)	1:3:4
H014	1(16) 3(40) 4(17)	1:3:4
H001	***1(16) 3(40) 4(18)***	***1:3:4***
H008	***1(16) 3(40) 4(18)***	***1:3:4***
H015	1(16) 3(40) 4(19)	1:3:4
H012	1(16) 3(40) 4(21)	1:3:4
H062	1(17) 3(35) 4(19)	1:3:4
H026	1(17) 3(38) 4(14)	1:3:4
H024	1(17) 3(39) 4(17)	1:3:4
H023	1(17) 3(42) 4(15)	1:3:4
H031	1(18) 3(36) 4(20)	1:3:4
H050	3(2) 1(4) 0(8) 3(27) 0(13) 4(1)	3:1:0:3:0:4
H022	3(1) 1(13) 0(7) 3(30) 4(22)	3:1:0:3:4
H058	3(3) 1(6) 3(33) 0(15) 4(2)	3:1:3:0:4
H006	3(3) 1(7) 3(1) 1(1) 3(29) 0(15) 4(2)	3:1:3:1:3:0:4
H036	3(2) 1(1) 3(1) 1(5) 3(2) 1(2) 3(1) 1(2) 3(36) 4(26)	3:1:3:1:3:1:3:1:3:4
H045	3(1) 1(15) 3(23) 4(14)	3:1:3:4

Table 3d: Hispanic MSY1 repeat codes and modular structures (continued). Data sorted by modular structures.

Sample ID	MSY1 Code	Modular Structure
H032	3(1) 1(15) 3(44) 4(14)	3:1:3:4
H034	3(1) 1(17) 3(31) 4(20)	3:1:3:4
H030	3(1) 1(17) 3(31) 4(25)	3:1:3:4
H055	3(1) 1(19) 3(36) 4(17)	3:1:3:4
H041	3(2) 1(17) 3(?) 4(20)	3:1:3:4
H011	3(2) 1(18) 3(36) 4(19)	3:1:3:4
H018	3(2) 1(18) 3(42) 4(16)	3:1:3:4
H025	3(3) 1(11) 3(29) 4(17)	3:1:3:4
H003	3(3) 1(13) 3(42) 4(13)	3:1:3:4
H021	3(4) 1(11) 3(36) 4(15)	3:1:3:4
H028	3(4) 1(14) 3(27) 4(19)	3:1:3:4
H007	3(5) 1(11) 3(30) 4(16)	3:1:3:4
H059	3(5) 1(19) 3(32) 4(15)	3:1:3:4

Note: No results were obtained from H017, 020, 027, 035, 038-040, 044 and 046-047.

Table 3a-d:

MSY1 Code = Type of repeat is followed by frequency in parenthesis. Thus, 1(15) 3 (39) 4 (19) refers to a sequence of MSY1which reads as follows: 5'...15 type 1...39 type 3... 19 type 4...3'.

Modular Structure = Sequence of repeat type without quantitation. Thus, 1(15) 3 (39) 4 (19) is expressed as 1:3:4.

(?) = total number of repeats could not be determined.

Samples with identical repeat codes are represented in ***italicized bold face type***.

Table 4: MSY1 standard diversity (*h*) indices calculated within Caucasian, African American, Asian, and Hispanic ethnic groups

Ethnic Group	n[a]	No. of Codes	*h* (codes)	n[b]	No. of Modular structures	*h* (modular structure)
Caucasian	47	47	1.0	47	12	0.7123
African American	45	45	1.0	53	20	0.9376
Asian	38	37	0.9986	39	10	0.6343
Hispanic	50	48	0.9984	53	13	0.7300
Combined Ethnic Groups	180	170[c]	0.9992	192	38[d]	0.8358

Table 4: Diversity calculations determined by Nei's unbiased estimator. *h* calculations of 1 represent the most diverse populations which did not display any shared MSY1 codes.

[a]Represents the total number of samples which yielded complete repeat codes.
[b]Represents the total number of samples which yielded complete modular structures.
[c]Represents the total number of different repeat codes observed within all ethnic groups combined.
[d]Represents the total number of different modular structures observed within all ethnic groups combined.

Table 5: Combined MSY1 Codes sorted by repeat number.

Sample ID	MSY1 CODE
AA052	0(1) 1(8) 3(32) 0(13) 4(2)
C053	0(1) 1(9) 3(31) 0(15) 4(1)
H049	0(1) 1(16) 3(31) 0(1) 3(1) 0(2) 4(1) 0(9) 4(1)
A043	0(1) 1(16) 3(34) 0(7)
AA054	0(1) 1(19) 3(1) 1(3) 3(2) 1(2) 3(20) 4(24)
A024	0(1) 1(21) 3(33) 4(12)
AA067	0(1) 2(22) 0(1) 4(55)
AA044	0(1) 3(1) 0(?)
AA046	0(1) 3(1) 0(?) 4(2)
C040	0(1) 3(1) 1(21) 0(9) 3(26) 4(16)
AA042	0(1) 3(1) 1(6) 3(29) 0(10) 4(2)
H010	0(1) 3(1) 1(6) 3(33) 0(16) 4(1)
AA028	0(1) 3(1) 1(7) 3(31) 0(16) 4(2)
H033	0(1) 3(1) 1(7) 3(32) 0(16) 4(2)
AA016	0(1) 3(1) 1(8) 3(27) 0(16) 4(2)
C041	0(1) 3(1) 1(9) 3(25) 0(11) 4(1)
A016	0(1) 3(2) 1(11) 3(26) 4(18)
H037	0(1) 3(2) 1(6) 3(1) 1(1) 3(18) 0(17) 4(2)
A036	0(2) 1(12) 0(2) 1(4) 0(38) 4(12)
AA065	0(2) 1(13) 3(2) 1(3) 3(42) 4(22)
AA060	0(2) 2(7) 0(1) 2(6) 0(1) 4(?)
AA050	0(4) 2(7) 0(?) 4(2)
C029	0(5) 2(7) 0(1) 2(6) 0(4) 4(49)
AA023	1(1) 3(1) 1(8) 3(33) 0(15) 4(2)
AA031	1(1) 3(2) 1(1) 3(2) 1(8) 3(32) 0(16)
AA033	1(1) 3(2) 1(5) 3(26) 0(15) 4(1)
AA012	1(1) 3(2) 1(5) 3(33) 0(15) 4(2)
AA017	1(1) 3(2) 1(6) 3(2) 1(1) 3(29)0(16) 4(2)
AA030	1(1) 3(2) 1(6) 3(33) 0(13) 4(2)
AA035	1(1) 3(2) 1(7) 3(29) 0(15) 4(1)
AA024	1(1) 3(2) 1(7) 3(35) 0(15) 4(2)
H057	1(11) 3(1) 1(1) 3(29) 0(18) 4(2)
AA020	1(11) 3(27) 0(15) 4(2)
H019	1(12) 3(43) 4(16)
C024	***1(14) 3(37) 4(20)***
H005	***1(14) 3(37) 4(20)***
C020	1(14) 3(38) 4(1) 3(2) 4(1) 3(2) 4(18)
H042	1(15) 3(?) 4(20)
C009	1(15) 3(11) 1(2) 3(31) 4(18)
H013	1(15) 3(2) 1(2) 3(39) 4(18)
H061	1(15) 3(28) 4(19)
C054	1(15) 3(36) 4(19)
AA068	1(15) 3(37) 4(18)
H048	1(15) 3(38) 4(20)
C002	***1(15) 3(39) 4(19)***
H051	***1(15) 3(39) 4(19)***
H016	1(15) 3(39) 4(20)
AA039	1(15) 3(4) 0(?)
H043	1(16) 3(?) 4(20)

Table 5: Combined MSY1 Codes sorted by repeat number (continued).

Sample ID	MSY1 CODE
H060	1(16) 3(34) 4(20)
A014	1(16) 3(36) 4(20)
C031	1(16) 3(36) 4(21)
H009	1(16) 3(36) 4(21)
H056	1(16) 3(37) 4(18)
C050	***1(16) 3(37) 4(20)***
H002	***1(16) 3(37) 4(20)***
AA007	1(16) 3(38) 0(1) 4(20)
AA069	***1(16) 3(38) 4(20)***
H053	***1(16) 3(38) 4(20)***
C034	1(16) 3(39) 4(18)
C012	1(16) 3(39) 4(19)
C013	1(16) 3(39) 4(20)
H029	1(16) 3(39) 4(21)
H014	1(16) 3(40) 4(17)
AA013	***1(16) 3(40) 4(18)***
H001	***1(16) 3(40) 4(18)***
H008	***1(16) 3(40) 4(18)***
C016	***1(16) 3(40) 4(19)***
H015	***1(16) 3(40) 4(19)***
H012	1(16) 3(40) 4(21)
C011	1(16) 3(41) 4(18)
C022	1(16) 3(43) 4(17)
AA034	1(17) 3(23) 0(17) 4(2)
H062	1(17) 3(35) 4(19)
C037	1(17) 3(35) 4(22)
C025	1(17) 3(37) 4(23)
H026	1(17) 3(38) 4(14)
C056	1(17) 3(38) 4(20)
H024	1(17) 3(39) 4(17)
AA029	1(17) 3(40) 4(21)
AA010	1(17) 3(41) 4(14)
H023	1(17) 3(42) 4(15)
C039	1(18) 3(35) 4(1) 3(1) 4(16)
H031	1(18) 3(36) 4(20)
A044	1(18) 3(40) 4(8)
C017	1(19) 3(2) 1(4) 3(10) 1(2) 3(31) 4(19)
H063	1(19) 3(3) 0(4) 3(28) 4(23)
A019	1(19) 3(38) 4(9)
H052	1(19) 3(5) 0(1) 3(38) 4(17)
H004	***1(20) 3(1) 1(3) 3(35) 4(25)***
H054	***1(20) 3(1) 1(3) 3(35) 4(25)***
AA022	1(20) 3(31) 4(18)
C032	1(20) 3(52) 4(15)
AA038	1(21) 3(3) 1(4) 3(9) 1(2) 3(30) 4(24)
C027	1(21) 3(49) 4(18)
C046	1(21) 3(50) 4(17)
C006	1(22) 3(34) 4(15)
C045	1(22) 3(38) 4(20)

Table 5: Combined MSY1 Codes sorted by repeat number (continued).

Sample ID	MSY1 CODE
C007	1(22) 3(48) 4(14)
C028	1(22) 3(49) 4(18)
C047	1(22) 3(51) 4(16)
C018	1(22) 3(53) 4(14)
C008	1(23) 3(26) 1(1) 3(25) 4(16)
AA025	1(8) 3(34) 0(11) 4(2)
AA003	1(9) 0(1) 1(1) 0(2) 1(1) 0(1) 1(1) 0(1) 1(1) 0(3)
A004	3(1) 1(11) 3(50) 4(11)
A008	3(1) 1(12) 3(51) 4(15)
H022	3(1) 1(13) 0(7) 3(30) 4(22)
A018	3(1) 1(13) 3(51) 4(18)
H045	3(1) 1(15) 3(23) 4(14)
AA021	3(1) 1(15) 3(27) 4(26)
H032	3(1) 1(15) 3(44) 4(14)
C019	***3(1) 1(17) 3(31) 4(20)***
H034	***3(1) 1(17) 3(31) 4(20)***
H030	3(1) 1(17) 3(31) 4(25)
A039	3(1) 1(17) 3(39) 4(9)
A030	3(1) 1(18) 3(41) 4(7)
A025	3(1) 1(18) 3(41) 4(8)
A001	3(1) 1(18) 3(43) 4(5)
C023	3(1) 1(19) 0(17) 4(1)
C052	3(1) 1(19) 0(3) 3(19) 4(12)
H055	3(1) 1(19) 3(36) 4(17)
C042	3(1) 1(8) 3(41) 4(19)
C030	3(2) 1(17) 3(28) 4(20)
A023	3(2) 1(19) 3(2) 1(1) 3(38) 4(13)
AA047	3(2) 0(?) 4(2)
H036	3(2) 1(1) 3(1) 1(5) 3(2) 1(2) 3(1) 1(2) 3(36) 4(26)
C048	3(2) 1(12) 3(40) 4(17)
C026	3(2) 1(12) 3(59) 4(12)
A038	3(2) 1(15) 3(?) 0(?)
A020	3(2) 1(15) 3(45) 4(13)
A040	3(2) 1(16) 3(37) 4(13)
A002	3(2) 1(16) 3(52) 4(8)
H041	3(2) 1(17) 3(?) 4(20)
C001	3(2) 1(17) 3(27) 4 (21)
AA072	3(2) 1(17) 3(35) 4(24)
H011	3(2) 1(18) 3(36) 4(19)
H018	3(2) 1(18) 3(42) 4(16)
C059	3(2) 1(3) 3(1) 1(5) 3(36) 4(19)
H050	3(2) 1(4) 0(8) 3(27) 0(13) 4(1)
AA064	3(2) 1(7) 3(27) 0(14) 4(2)
AA056	3(2) 1(7) 3(32) 0(15) 4(1)
AA055	3(2) 1(7) 3(33) 0(14) 4(2)
AA026	3(2) 1(7) 3(8) 1(5) 3(3) 1(1) 3 (16) 0(12) 4(2)
AA049	3(3) 0(?) 4(2)
AA051	3(3) 0(?) 4(2)
A042	3(3) 1(1) 3(1) 1(11) 3(41) 4(10)

Table 5: Combined MSY1 Codes sorted by repeat number (continued).

Sample ID	MSY1 CODE
H025	3(3) 1(11) 3(29) 4(17)
AA061	3(3) 1(12) 3(35) 4(18)
A005	3(3) 1(12) 3(49) 4(9)
AA002	3(3) 1(13) 3(36) 0(1) 4(16)
AA071	3(3) 1(13) 3(40) 4(21)
C043	3(3) 1(13) 3(41) 4(21)
H003	3(3) 1(13) 3(42) 4(13)
C060	3(3) 1(14) 3(41) 4(23)
A050	3(3) 1(15) 3(38) 4(12)
A003	3(3) 1(15) 3(45) 4(14)
A012	3(3) 1(16) 3(36) 4(14)
A013	3(3) 1(16) 3(38) 4(14)
A035	3(3) 1(17) 3(39) 4(1) 3(2) 4(11)
A009	3(3) 1(2) 3(2) 1(10) 3(40) 4(9)
A051	3(3) 1(2) 3(2) 1(11) 3(42) 4(10)
AA057	3(3) 1(5) 3(2) 1(1) 3(33) 0(16) 4(1)
AA053	3(3) 1(5) 3(33) 0(14) 4(2)
AA070	3(3) 1(6) 3(1) 1(1) 3(25) 0(18) 4(1)
H058	3(3) 1(6) 3(33) 0(15) 4(2)
H006	3(3) 1(7) 3(1) 1(1) 3(29) 0(15) 4(2)
AA066	3(3) 1(7) 3(2) 1(1) 3(21) 0(16) 4(2)
AA040	3(3) 1(7) 3(2) 1(1) 3(26) 0(10) 4(2)
AA032	3(3) 1(7) 3(2) 1(1) 3(31) 0(15) 4(2)
AA037	3(3) 1(7) 3(33) 0(15) 4(2)
AA058	3(3) 1(8) 3(29) 0(16) 4(2)
H021	3(4) 1(11) 3(36) 4(15)
AA001	3(4) 1(12) 3(1) 1(1) 3(27) 0(17) 4(13)
A031	***3(4) 1(12) 3(52) 4(10)***
A034	***3(4) 1(12) 3(52) 4(10)***
A015	3(4) 1(12) 3(54) 4(10)
A033	3(4) 1(13) 3(25) 4(20)
C036	3(4) 1(13) 3(31) 4(19)
A010	3(4) 1(13) 3(40) 4(10)
A041	3(4) 1(13) 3(41) 4(12)
A011	3(4) 1(14) 3(2) 1(1) 3(52) 4(11)
H028	3(4) 1(14) 3(27) 4(19)
C014	3(4) 1(14) 3(31) 4(19)
A037	3(4) 1(14) 3(44) 4(13)
C038	3(4) 1(5) 3(1) 1(6) 3(38) 4(20)
A032	3(4) 1(8) 3(54) 4(14)
A022	3(5) 1(1) 3(1) 1(11) 3(1) 1(2) 3(32) 4(19)
H007	3(5) 1(11) 3(30) 4(16)
A017	3(5) 1(16) 3(16) 1(8) 3(38) 4(6)
H059	3(5) 1(19) 3(32) 4(15)
C004	3(6) 1(11) 3(37) 4 (18)

Table 5:

Sample ID is coded by ethnic group followed by the sample identification number.

C = Caucasian; AA = African American; A = Asian; H = Hispanic

Samples with identical repeat codes are represented in ***italicized bold face type***.

Table 6: Population specific and combined MSY1 modular structures

Repeat Structure	Caucasian	African American	Asian	Hispanic	n =
0:1:0:1:0:4			1		1
0:1:3:0			1		1
0:1:3:0:3:0:4:0:4				1	1
0:1:3:0:4	1	1			2
0:1:3:1:3:1:3:4		1			1
0:1:3:1:3:4		1			
0:1:3:4			1		1
0:2:0:2:0:4	1	1			2
0:2:0:4		2			2
0:3:0:4		2			2
0:3:1:3:0:4	2	3		2	7
0:3:1:3:1:3:0:4				1	1
0:3:1:3:4			1		1
1:0:1:0:1:0:1:0:1:0		1			1
1:3:0		1			1
1:3:0:3:4				2	2
1:3:0:4		4			4
1:3:1:3:0:4		6		1	7
1:3:1:3:1:3:0		1			1
1:3:1:3:1:3:0:4		1			1
1:3:1:3:1:3:4	1	1			2
1:3:1:3:4	2			3	5
1:3:4	23	6	3	24	56
1:3:4:3:4	1				1
1:3:4:3:4:3:4	1				1
3:0:4		3			3
3:1:0:3:0:4				1	1
3:1:0:3:4	1			1	2
3:1:0:4	1				1
3:1:3:0			1		1
3:1:3:0:4		7		1	8
3:1:3:1:3:0:4		6		1	7
3:1:3:1:3:1:3:0:4		1			1
3:1:3:1:3:1:3:1:3:4				1	1
3:1:3:1:3:1:3:4			1		1
3:1:3:1:3:4	2		6		8
3:1:3:4	11	4	23	14	52
3:1:3:4:3:4			1		1
$n_{(total)}$ =	47	53	39	53	192
No. of structures:	12	20	10	13	38*

*Total number of different modular structures observed within all ethnic groups combined.

Table 7a: Caucasian Y chromosome microsatellite haplotype data in sorted format.

SAMPLE ID	DYS 19	DYS 389 I	DYS 389 II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385 I	DYS 385 II
C052	13	10	17	24	10	11	13	16	18
C023	13	10	17	25	10	11	14	14	17
C017	13	11	16	24	10	11	13	13	14
C060	14	9	16	22	10	11	13	14	14
C044	14	9	16	23	10	11	13	14	15
C011	14	9	16	23	11	13	13	12	14
C040	14	9	19	24	10	11	13	15	16
C009	14	10	15	24	10	13	13	11	14
C010	14	10	16	22	12	13	13	11	14
C014	14	10	16	23	10	11	12	14	17
C022	14	10	16	23	10	13	13	11	14
C034	14	10	16	23	11	13	12	11	14
C054	14	10	16	23	11	13	13	12	14
C050	14	10	16	24	11	13	12	12	14
C056	14	10	16	24	11	13	13	11	15
C013	***14***	***10***	***16***	***24***	***11***	***13***	***13***	***12***	***14***
C020	***14***	***10***	***16***	***24***	***11***	***13***	***13***	***12***	***14***
C033	14	10	16	24	11	13	14	12	14
C031	14	10	16	25	11	14	13	11	13
C057	14	10	17	23	10	11	12	13	15
C016	14	10	17	23	10	11	12	13	18
C021	14	10	17	23	11	13	13	10	12
C024	14	10	17	24	11	13	13	11	14
C035	14	11	16	24	11	13	12	11	14
C002	14	11	17	24	10	13	13	11	15
C043	15	9	16	22	10	11	13	13	14

Table 7a: Caucasian Y chromosome microsatellite haplotype data in sorted format (continued).

SAMPLE ID	DYS 19	DYS 389 I	DYS 389 II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385 I	DYS 385 II
C026	15	9	16	23	10	13	12	11	21
C048	15	9	17	21	10	11	14	17	15
C058	15	9	17	22	10	11	14	12	14
C038	***15***	***10***	***16***	***23***	***9***	***11***	***12***	***13***	***17***
C039	***15***	***10***	***16***	***23***	***9***	***11***	***12***	***13***	***17***
C059	15	10	16	24	10	11	12	13	15
C008	15	10	17	25	10	11	13	11	15
C005	15	10	18	26	9	11	14	11	14
C053	15	10	19	21	11	11	13	16	17
C029	15	10	19	24	10	11	13	11	12
C012	15	11	17	24	10	13	13	11	14
C004	15	11	18	23	10	11	13	13	16
C036	15	11	18	23	10	12	14	15	15
C042	16	9	15	24	10	11	13	14	14
C003	16	9	16	24	10	11	12	14	17
C037	16	9	16	24	10	13	12	12	14
C030	16	9	17	24	11	11	13	14	15
C025	16	9	17	25	10	11	13	11	14
C006	16	10	15	25	10	11	12	11	11
C015	16	10	16	25	10	11	13	11	15
C027	16	10	17	24	10	11	13	11	14
C032	16	10	17	24	11	11	13	11	14
C018	16	10	17	25	11	11	13	11	13
C028	16	10	17	25	11	11	14	10	15
C019	16	10	18	24	10	11	13	13	15
C001	16	10	18	24	11	11	13	12	15

Table 7a: Caucasian microsatellite haplotype data in sorted format (continued).

SAMPLE ID	DYS 19	DYS 389 I	DYS 389 II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385 I	DYS 385 II
C041	16	11	17	21	10	11	13	16	16
C046	16	11	17	25	11	11	13	11	15
C047	17	10	16	24	10	11	13	11	14
C055	17	10	17	25	10	11	13	11	14
C007	17	10	18	25	10	11	13	10	13
C045	NR	9	17	25	10	11	13	11	14
C049	NR	NR	NR	NR	NR	NR	NR	NR	NR
C051	NR	NR	NR	NR	NR	NR	NR	NR	NR

NR = no result

Samples with identical repeat codes are represented in ***italicized bold face type***.

Table 7b: African American Y chromosome microsatellite haplotype data in sorted format.

SAMPLE ID	DYS 19	DYS 389 I	DYS 389 II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385 I	DYS 385 II
AA072	13	09	15	22	10	11	13	13	14
AA010	13	10	17	22	09	11	12	12	16
AA038	13	11	16	23	09	11	13	12	15
AA071	14	09	16	23	10	11	13	13	14
AA021	14	10	15	24	10	10	14	13	18
AA039	14	10	16	22	10	13	13	11	14
AA029	14	10	16	24	10	13	13	11	14
AA011	***14***	***10***	***16***	***24***	***11***	***13***	***13***	***11***	***14***
AA069	***14***	***10***	***16***	***24***	***11***	***13***	***13***	***11***	***14***
AA007	14	10	17	23	11	13	13	12	14
AA004	14	10	17	24	10	13	13	11	14
AA018	14	10	18	24	11	13	13	12	15
AA019	14	11	16	23	10	10	14	12	17
AA022	14	11	16	24	10	13	13	11	14
AA068	14	11	17	24	11	13	13	11	15
AA067	14	11	18	20	11	11	14	17	18
AA066	15	09	17	21	10	11	13	16	16
AA065	15	09	17	22	10	11	13	15	16
AA048	15	09	18	21	09	11	13	14	17
AA005	15	09	18	21	10	11	13	16	18
AA040	15	09	19	21	10	11	12	16	18
AA063	15	10	17	21	10	11	13	16	17
AA015	15	10	17	21	10	11	14	15	15
AA020	15	10	17	21	10	11	14	16	16
AA037	15	10	17	21	10	11	14	17	19

Table 7b: African American Y chromosome microsatellite haplotype data in sorted format (continued).

SAMPLE ID	DYS 19	DYS 389 I	DYS 389 II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385 I	DYS 385 II
AA008	***15***	***10***	***17***	***21***	***10***	***11***	***15***	***16***	***17***
AA025	***15***	***10***	***17***	***21***	***10***	***11***	***15***	***16***	***17***
AA045	15	10	17	21	11	11	13	16	16
AA003	15	10	17	21	11	11	14	15	18
AA064	15	10	17	22	10	11	14	17	17
AA049	15	10	18	21	10	10	13	16	18
AA062	15	10	18	21	10	11	13	15	16
AA047	15	10	18	21	10	11	13	16	17
AA026	15	10	18	21	10	11	14	17	18
AA027	15	10	18	22	10	11	13	17	18
AA061	15	10	18	22	10	11	14	14	14
AA002	15	10	18	23	10	12	15	14	15
AA070	15	10	NR	21	11	NR	13	NR	NR
AA060	15	10	NR	24	10	11	13	11	12
AA013	15	11	16	23	11	13	13	11	13
AA009	15	11	17	21	10	11	13	14	17
AA042	15	11	17	21	10	11	13	16	17
AA033	15	11	17	21	10	11	14	15	16
AA041	15	11	18	21	10	11	13	16	16
AA043	15	11	18	21	10	11	15	16	16
AA058	15	11	18	21	11	11	13	16	16
AA054	16	09	17	21	10	11	12	13	14
AA001	16	09	17	23	10	11	13	12	15
AA028	16	10	17	21	10	11	14	16	18
AA012	***16***	***10***	***17***	***21***	***10***	***11***	***15***	***17***	***18***
AA051	***16***	***10***	***17***	***21***	***10***	***11***	***15***	***17***	***18***
AA053	***16***	***10***	***17***	***21***	***10***	***11***	***15***	***17***	***18***

Table 7b: African American Y chromosome microsatellite haplotype data in sorted format (continued).

SAMPLE ID	DYS 19	DYS 389 I	DYS 389 II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385 I	DYS 385 II
AA014	16	10	17	21	10	11	15	17	20
AA006	16	10	17	21	10	11	15	18	18
AA044	16	10	18	22	10	11	15	16	18
AA050	16	10	18	24	10	11	13	11	12
AA059	16	10	19	21	10	11	15	16	17
AA056	16	10	20	21	10	12	13	16	18
AA036	16	11	17	22	10	11	14	17	17
AA046	16	11	18	21	10	11	14	14	20
AA035	17	08	18	21	10	11	12	15	16
AA024	17	09	18	21	10	11	14	17	18
AA057	17	10	16	21	10	11	13	16	17
AA055	17	10	17	21	10	11	13	17	18
AA016	17	10	17	21	10	11	14	17	17
AA031	17	10	17	21	10	11	15	17	17
AA032	17	10	18	21	10	11	13	15	19
AA017	17	10	18	21	10	11	13	17	17
AA052	17	10	18	21	10	11	15	17	18
AA030	17	10	18	21	10	11	15	17	19

NR = no result

Samples with identical repeat codes are represented in ***italicized bold face type***.

Table 7c: Asian Y chromosome microsatellite haplotype data in sorted format.

SAMPLE ID	DYS 19	DYS 389 I	DYS 389 II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385 I	DYS 385 II
A007	13	10	15	25	09	15	14	16	21
A024	13	11	16	23	06	14	14	16	19
A042	14	09	16	23	10	12	12	11	16
A005	14	09	16	25	10	14	12	13	19
A011	14	09	17	23	10	14	12	15	20
A029	14	09	18	22	10	14	13	12	13
A038	14	10	16	23	10	10	14	12	17
A028	14	10	16	23	10	13	13	11	17
A014	14	10	16	24	10	13	13	11	15
A020	14	10	16	24	10	14	12	13	17
A050	14	10	16	24	10	14	12	13	21
A047	14	10	17	24	11	14	13	11	12
A036	14	12	16	23	10	14	12	12	12
A009	15	09	16	23	10	12	13	13	17
A001	***15***	***09***	***16***	***23***	***10***	***14***	***13***	***13***	***13***
A025	***15***	***09***	***16***	***23***	***10***	***14***	***13***	***13***	***13***
A051	15	09	16	24	10	12	12	12	16
A033	15	09	16	24	10	13	15	15	16
A037	15	09	16	24	10	14	12	14	14
A052	15	09	16	24	10	14	12	14	17
A003	***15***	***09***	***16***	***24***	***10***	***14***	***12***	***14***	***18***
A041	***15***	***09***	***16***	***24***	***10***	***14***	***12***	***14***	***18***
A030	15	09	17	23	10	14	13	13	13
A026	15	09	17	23	11	13	13	13	13
A039	15	09	17	23	11	14	13	13	13
A046	15	09	17	24	10	11	12	11	12
A002	15	09	17	25	09	13	12	10	22
A017	15	09	18	25	09	13	12	12	19

Table 7c: Asian Y chromosome microsatellite haplotype data in sorted format (continued).

SAMPLE ID	DYS 19	DYS 389 I	DYS 389 II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385 I	DYS 385 II
A048	15	10	16	21	09	11	14	18	19
A013	15	10	16	23	10	13	14	12	21
A040	15	10	17	24	11	13	14	13	18
A023	15	10	18	24	11	13	14	13	18
A012	15	10	18	25	11	13	14	13	17
A034	15	11	15	23	10	13	13	10	18
A010	15	11	16	22	10	13	13	10	20
A022	15	11	16	23	10	11	12	17	20
A016	15	11	17	25	10	11	13	12	13
A044	16	08	17	24	10	14	13	13	14
A004	16	09	16	23	09	13	12	12	17
A032	16	09	16	23	10	13	12	12	19
A006	16	09	16	23	11	14	13	12	13
A021	16	09	16	25	10	10	12	12	21
A031	16	10	15	22	10	14	12	11	11
A018	16	10	15	25	10	10	13	12	20
A015	16	10	16	22	10	14	12	11	11
A035	16	10	16	23	11	13	14	14	18
A049	16	12	15	23	10	13	13	10	18
A019	17	09	17	23	10	14	13	13	13
A008	17	09	17	25	10	13	12	14	18
A054	17	09	18	24	10	14	12	12	17
A043	17	11	17	26	10	11	14	13	19
A027	17	11	18	24	10	11	13	13	18
A045	NR	NR	NR	NR	NR	NR	NR	NR	NR
A053	NR	NR	NR	NR	NR	NR	NR	NR	NR

NR = no result

Samples with identical repeat codes are represented in ***italicized bold face type***.

Table 7d: Hispanic Y chromosome microsatellite haplotype data in sorted format.

SAMPLE ID	DYS 19	DYS 389 1	DYS 389 11	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385 1	DYS 385 11
H052	13	09	16	22	08	11	13	12	16
H011	13	10	16	24	10	14	14	16	16
H045	13	10	16	24	11	14	13	16	17
H021	13	10	17	23	10	11	13	17	18
H040	13	10	17	24	10	11	13	15	18
H027	13	10	17	24	10	16	13	13	14
H049	13	10	19	23	10	11	13	15	19
H063	13	11	16	24	09	11	13	13	14
H018	13	11	17	22	10	13	13	13	16
H030	14	09	16	24	11	11	14	17	19
H031	14	09	18	23	11	13	12	11	14
H062	14	10	15	23	10	13	13	11	14
H026	14	10	15	24	11	13	14	11	14
H015	14	10	16	23	10	13	13	11	14
H012	14	10	16	23	11	13	13	15	16
H016	14	10	16	24	11	13	13	09	14
H002	14	10	16	24	11	13	13	10	11
H039	***14***	***10***	***16***	***24***	***11***	***13***	***13***	***11***	***13***
H048	***14***	***10***	***16***	***24***	***11***	***13***	***13***	***11***	***13***
H005	***14***	***10***	***16***	***24***	***11***	***13***	***13***	***11***	***14***
H013	***14***	***10***	***16***	***24***	***11***	***13***	***13***	***11***	***14***
H020	***14***	***10***	***16***	***24***	***11***	***13***	***13***	***11***	***14***
H060	***14***	***10***	***16***	***24***	***11***	***13***	***13***	***11***	***14***
H001	14	10	16	24	11	14	13	11	14
H051	14	10	16	25	11	13	13	11	14
H036	14	10	17	23	10	11	13	14	15
H007	14	10	17	23	10	12	12	13	17
H035	14	10	17	24	10	13	13	11	11

Table 7d: Hispanic Y chromosome microsatellite haplotype data in sorted format (continued).

SAMPLE ID	DYS 19	DYS 389 I	DYS 389 II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385 I	DYS 385 II
H046	14	10	17	24	11	13	13	12	15
H028	14	10	18	23	10	11	12	13	18
H009	14	10	18	24	11	13	13	11	14
H047	14	10	18	25	10	14	13	15	17
H034	14	10	19	24	10	14	13	14	17
H044	14	11	15	24	11	14	13	11	14
H014	***14***	***11***	***16***	***24***	***11***	***13***	***13***	***11***	***14***
H042	***14***	***11***	***16***	***24***	***11***	***13***	***13***	***11***	***14***
H043	***14***	***11***	***16***	***24***	***11***	***13***	***13***	***11***	***14***
H008	14	11	16	24	11	13	13	12	14
H061	14	11	16	25	10	14	13	11	14
H032	14	11	16	25	10	14	13	14	18
H024	14	11	16	25	11	13	13	11	14
H053	14	11	17	24	11	13	13	11	14
H038	15	10	16	23	11	13	13	11	14
H017	15	10	16	24	10	13	13	11	14
H029	15	10	16	25	11	14	13	11	14
H058	15	10	17	21	10	11	13	15	16
H059	15	10	17	23	10	13	13	15	16
H056	15	10	17	23	11	13	13	11	14
H022	15	10	17	24	10	11	13	17	18
H057	15	10	18	21	10	11	13	17	17
H006	15	10	18	21	10	11	14	15	19
H037	15	10	18	21	11	11	13	16	17
H019	15	10	19	21	10	11	13	16	17
H023	15	11	16	23	11	13	13	11	13
H050	15	11	17	21	10	11	14	17	19
H055	15	11	17	24	10	12	13	15	18

Table 7d: Hispanic Y chromosome microsatellite haplotype data in sorted format (continued).

SAMPLE ID	DYS 19	DYS 389 I	DYS 389 II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385 I	DYS 385 II
H003	15,17	11	16	21	10	11	14	13	14
H010	16	10	17	21	10	11	15	12	13
H041	16	10	17	24	11	11	14	14	15
H004	16	11	16	23	11	13	13	12	14
H054	16	11	16	23	12	13	13	12	14
H025	17	10	15	23	10	11	13	12	12
H033	17	11	17	21	11	11	15	18	18

NR = no result

Samples with identical repeat codes are represented in ***italicized bold face type***.

Table 8: Y-chromosome Microsatellite haplotype diversity (*h*) values calculated within Caucasian, African American, Asian, and Hispanic ethnic groups.

Ethnic Group	n	No. of haplotypes	*h** (haplotype)
Caucasian	57	55	0.9987
African American	68	64	0.9978
Asian	52	50	0.9985
Hispanic	63	57	0.9949
Combined Ethnic Groups	240	221	0.9991

*Diversities calculated for samples containing complete haplotypes. Samples with partial profiles were removed from analysis.

Table 9: Combined Y chromosome microsatellite haplotype data in sorted format.

SAMPLE ID	DYS 19	DYS 389 1	DYS 389 11	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385 1	DYS 385 11
AA072	13	09	15	22	10	11	13	13	14
H052	13	09	16	22	08	11	13	12	16
A007	13	10	15	25	09	15	14	16	21
H011	13	10	16	24	10	14	14	16	16
H045	13	10	16	24	11	14	13	16	17
AA010	13	10	17	22	09	11	12	12	16
H021	13	10	17	23	10	11	13	17	18
H040	13	10	17	24	10	11	13	15	18
C052	13	10	17	24	10	11	13	16	18
H027	13	10	17	24	10	16	13	13	14
C023	13	10	17	25	10	11	14	14	17
H049	13	10	19	23	10	11	13	15	19
A024	13	11	16	23	06	14	14	16	19
AA038	13	11	16	23	09	11	13	12	15
H063	13	11	16	24	09	11	13	13	14
C017	13	11	16	24	10	11	13	13	14
H018	13	11	17	22	10	13	13	13	16
C060	14	09	16	22	10	11	13	14	14
AA071	14	09	16	23	10	11	13	13	14
C044	14	09	16	23	10	11	13	14	15
A042	14	09	16	23	10	12	12	11	16
C011	14	09	16	23	11	13	13	12	14
H030	14	09	16	24	11	11	14	17	19
A005	14	09	16	25	10	14	12	13	19
A011	14	09	17	23	10	14	12	15	20
A029	14	09	18	22	10	14	13	12	13
H031	14	09	18	23	11	13	12	11	14
C040	14	09	19	24	10	11	13	15	16

Table 9: Combined Y chromosome microsatellite haplotype data in sorted format (continued).

SAMPLE ID	DYS 19	DYS 389 I	DYS 389 II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385 I	DYS 385 II
H062	14	10	15	23	10	13	13	11	14
AA021	14	10	15	24	10	10	14	13	18
C009	14	10	15	24	10	13	13	11	14
H026	14	10	15	24	11	13	14	11	14
AA039	14	10	16	22	10	13	13	11	14
C010	14	10	16	22	12	13	13	11	14
A038	14	10	16	23	10	10	14	12	17
C014	14	10	16	23	10	11	12	14	17
C022	***14***	***10***	***16***	***23***	***10***	***13***	***13***	***11***	***14***
H015	***14***	***10***	***16***	***23***	***10***	***13***	***13***	***11***	***14***
A028	14	10	16	23	10	13	13	11	17
C034	14	10	16	23	11	13	12	11	14
C054	14	10	16	23	11	13	13	12	14
H012	14	10	16	23	11	13	13	15	16
AA029	14	10	16	24	10	13	13	11	14
A014	14	10	16	24	10	13	13	11	15
A020	14	10	16	24	10	14	12	13	17
A050	14	10	16	24	10	14	12	13	21
C050	14	10	16	24	11	13	12	12	14
H016	14	10	16	24	11	13	13	09	14
H002	14	10	16	24	11	13	13	10	11
H039	***14***	***10***	***16***	***24***	***11***	***13***	***13***	***11***	***13***
H048	***14***	***10***	***16***	***24***	***11***	***13***	***13***	***11***	***13***
AA011	***14***	***10***	***16***	***24***	***11***	***13***	***13***	***11***	***14***
AA069	***14***	***10***	***16***	***24***	***11***	***13***	***13***	***11***	***14***
H005	***14***	***10***	***16***	***24***	***11***	***13***	***13***	***11***	***14***
H013	***14***	***10***	***16***	***24***	***11***	***13***	***13***	***11***	***14***
H020	***14***	***10***	***16***	***24***	***11***	***13***	***13***	***11***	***14***

Table 9: Combined Y chromosome microsatellite haplotype data in sorted format (continued).

SAMPLE ID	DYS 19	DYS 389 I	DYS 389 II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385 I	DYS 385 II
H060	***14***	***10***	***16***	***24***	***11***	***13***	***13***	***11***	***14***
C056	14	10	16	24	11	13	13	11	15
C013	***14***	***10***	***16***	***24***	***11***	***13***	***13***	***12***	***14***
C020	***14***	***10***	***16***	***24***	***11***	***13***	***13***	***12***	***14***
C033	14	10	16	24	11	13	14	12	14
H001	14	10	16	24	11	14	13	11	14
H051	14	10	16	25	11	13	13	11	14
C031	14	10	16	25	11	14	13	11	13
C057	14	10	17	23	10	11	12	13	15
C016	14	10	17	23	10	11	12	13	18
H036	14	10	17	23	10	11	13	14	15
H007	14	10	17	23	10	12	12	13	17
C021	14	10	17	23	11	13	13	10	12
AA007	14	10	17	23	11	13	13	12	14
H035	14	10	17	24	10	13	13	11	11
AA004	14	10	17	24	10	13	13	11	14
C024	14	10	17	24	11	13	13	11	14
H046	14	10	17	24	11	13	13	12	15
A047	14	10	17	24	11	14	13	11	12
H028	14	10	18	23	10	11	12	13	18
H009	14	10	18	24	11	13	13	11	14
AA018	14	10	18	24	11	13	13	12	15
H047	14	10	18	25	10	14	13	15	17
H034	14	10	19	24	10	14	13	14	17
H044	14	11	15	24	11	14	13	11	14
AA019	14	11	16	23	10	10	14	12	17
AA022	14	11	16	24	10	13	13	11	14
C035	14	11	16	24	11	13	12	11	14

Table 9: Combined Y chromosome microsatellite haplotype data in sorted format (continued).

SAMPLE ID	DYS 19	DYS 389 I	DYS 389 II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385 I	DYS 385 II
H014	***14***	***11***	***16***	***24***	***11***	***13***	***13***	***11***	***14***
H042	***14***	***11***	***16***	***24***	***11***	***13***	***13***	***11***	***14***
H043	***14***	***11***	***16***	***24***	***11***	***13***	***13***	***11***	***14***
H008	14	11	16	24	11	13	13	12	14
H061	14	11	16	25	10	14	13	11	14
H032	14	11	16	25	10	14	13	14	18
H024	14	11	16	25	11	13	13	11	14
C002	14	11	17	24	10	13	13	11	15
H053	14	11	17	24	11	13	13	11	14
AA068	14	11	17	24	11	13	13	11	15
AA067	14	11	18	20	11	11	14	17	18
A036	14	12	16	23	10	14	12	12	12
C043	15	09	16	22	10	11	13	13	14
A009	15	09	16	23	10	12	13	13	17
C026	15	09	16	23	10	13	12	11	21
A001	***15***	***09***	***16***	***23***	***10***	***14***	***13***	***13***	***13***
A025	***15***	***09***	***16***	***23***	***10***	***14***	***13***	***13***	***13***
A051	15	09	16	24	10	12	12	12	16
A033	15	09	16	24	10	13	15	15	16
A037	15	09	16	24	10	14	12	14	14
A052	15	09	16	24	10	14	12	14	17
A003	***15***	***09***	***16***	***24***	***10***	***14***	***12***	***14***	***18***
A041	***15***	***09***	***16***	***24***	***10***	***14***	***12***	***14***	***18***
AA066	15	09	17	21	10	11	13	16	16
C048	15	09	17	21	10	11	14	17	15
AA065	15	09	17	22	10	11	13	15	16
C058	15	09	17	22	10	11	14	12	14
A030	15	09	17	23	10	14	13	13	13

Table 9: Combined Y chromosome microsatellite haplotype data in sorted format (continued).

SAMPLE ID	DYS 19	DYS 389 I	DYS 389 II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385 I	DYS 385 II
A026	15	09	17	23	11	13	13	13	13
A039	15	09	17	23	11	14	13	13	13
A046	15	09	17	24	10	11	12	11	12
A002	15	09	17	25	09	13	12	10	22
AA048	15	09	18	21	09	11	13	14	17
AA005	15	09	18	21	10	11	13	16	18
A017	15	09	18	25	09	13	12	12	19
AA040	15	09	19	21	10	11	12	16	18
A048	15	10	16	21	09	11	14	18	19
C038	***15***	***10***	***16***	***23***	***09***	***11***	***12***	***13***	***17***
C039	***15***	***10***	***16***	***23***	***09***	***11***	***12***	***13***	***17***
A013	15	10	16	23	10	13	14	12	21
H038	15	10	16	23	11	13	13	11	14
C059	15	10	16	24	10	11	12	13	15
H017	15	10	16	24	10	13	13	11	14
H029	15	10	16	25	11	14	13	11	14
H058	15	10	17	21	10	11	13	15	16
AA063	15	10	17	21	10	11	13	16	17
AA015	15	10	17	21	10	11	14	15	15
AA020	15	10	17	21	10	11	14	16	16
AA037	15	10	17	21	10	11	14	17	19
AA008	***15***	***10***	***17***	***21***	***10***	***11***	***15***	***16***	***17***
AA025	***15***	***10***	***17***	***21***	***10***	***11***	***15***	***16***	***17***
AA045	15	10	17	21	11	11	13	16	16
AA003	15	10	17	21	11	11	14	15	18
AA064	15	10	17	22	10	11	14	17	17
H059	15	10	17	23	10	13	13	15	16
H056	15	10	17	23	11	13	13	11	14

Table 9: Combined Y chromosome microsatellite haplotype data in sorted format (continued).

SAMPLE ID	DYS 19	DYS 389 I	DYS 389 II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385 I	DYS 385 II
H022	15	10	17	24	10	11	13	17	18
A040	15	10	17	24	11	13	14	13	18
C008	15	10	17	25	10	11	13	11	15
AA049	15	10	18	21	10	10	13	16	18
AA062	15	10	18	21	10	11	13	15	16
AA047	15	10	18	21	10	11	13	16	17
H057	15	10	18	21	10	11	13	17	17
H006	15	10	18	21	10	11	14	15	19
AA026	15	10	18	21	10	11	14	17	18
H037	15	10	18	21	11	11	13	16	17
AA027	15	10	18	22	10	11	13	17	18
AA061	15	10	18	22	10	11	14	14	14
AA002	15	10	18	23	10	12	15	14	15
A023	15	10	18	24	11	13	14	13	18
A012	15	10	18	25	11	13	14	13	17
C005	15	10	18	26	09	11	14	11	14
H019	15	10	19	21	10	11	13	16	17
C053	15	10	19	21	11	11	13	16	17
C029	15	10	19	24	10	11	13	11	12
AA070	15	10	NR	21	11	NR	13	NR	NR
AA060	15	10	NR	24	10	11	13	11	12
A034	15	11	15	23	10	13	13	10	18
A010	15	11	16	22	10	13	13	10	20
A022	15	11	16	23	10	11	12	17	20
AA013	***15***	***11***	***16***	***23***	***11***	***13***	***13***	***11***	***13***
H023	***15***	***11***	***16***	***23***	***11***	***13***	***13***	***11***	***13***
AA009	15	11	17	21	10	11	13	14	17
AA042	15	11	17	21	10	11	13	16	17

Table 9: Combined Y chromosome microsatellite haplotype data in sorted format (continued).

SAMPLE ID	DYS 19	DYS 389 I	DYS 389 II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385 I	DYS 385 II
AA033	15	11	17	21	10	11	14	15	16
H050	15	11	17	21	10	11	14	17	19
H055	15	11	17	24	10	12	13	15	18
C012	15	11	17	24	10	13	13	11	14
A016	15	11	17	25	10	11	13	12	13
AA041	15	11	18	21	10	11	13	16	16
AA043	15	11	18	21	10	11	15	16	16
AA058	15	11	18	21	11	11	13	16	16
C004	15	11	18	23	10	11	13	13	16
C036	15	11	18	23	10	12	14	15	15
H003	15,17	11	16	21	10	11	14	13	14
A044	16	08	17	24	10	14	13	13	14
C042	16	09	15	24	10	11	13	14	14
A004	16	09	16	23	09	13	12	12	17
A032	16	09	16	23	10	13	12	12	19
A006	16	09	16	23	11	14	13	12	13
C003	16	09	16	24	10	11	12	14	17
C037	16	09	16	24	10	13	12	12	14
A021	16	09	16	25	10	10	12	12	21
AA054	16	09	17	21	10	11	12	13	14
AA001	16	09	17	23	10	11	13	12	15
C030	16	09	17	24	11	11	13	14	15
C025	16	09	17	25	10	11	13	11	14
A031	16	10	15	22	10	14	12	11	11
A018	16	10	15	25	10	10	13	12	20
C006	16	10	15	25	10	11	12	11	11
A015	16	10	16	22	10	14	12	11	11
A035	16	10	16	23	11	13	14	14	18

Table 9: Combined Y chromosome microsatellite haplotype data in sorted format (continued).

SAMPLE ID	DYS 19	DYS 389 I	DYS 389 II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385 I	DYS 385 II
C015	16	10	16	25	10	11	13	11	15
AA028	16	10	17	21	10	11	14	16	18
H010	16	10	17	21	10	11	15	12	13
AA012	***16***	***10***	***17***	***21***	***10***	***11***	***15***	***17***	***18***
AA051	***16***	***10***	***17***	***21***	***10***	***11***	***15***	***17***	***18***
AA053	***16***	***10***	***17***	***21***	***10***	***11***	***15***	***17***	***18***
AA014	16	10	17	21	10	11	15	17	20
AA006	16	10	17	21	10	11	15	18	18
C027	16	10	17	24	10	11	13	11	14
C032	16	10	17	24	11	11	13	11	14
H041	16	10	17	24	11	11	14	14	15
C018	16	10	17	25	11	11	13	11	13
C028	16	10	17	25	11	11	14	10	15
AA044	16	10	18	22	10	11	15	16	18
AA050	16	10	18	24	10	11	13	11	12
C019	16	10	18	24	10	11	13	13	15
C001	16	10	18	24	11	11	13	12	15
AA059	16	10	19	21	10	11	15	16	17
AA056	16	10	20	21	10	12	13	16	18
H004	16	11	16	23	11	13	13	12	14
H054	16	11	16	23	12	13	13	12	14
C041	16	11	17	21	10	11	13	16	16
AA036	16	11	17	22	10	11	14	17	17
C046	16	11	17	25	11	11	13	11	15
AA046	16	11	18	21	10	11	14	14	20
A049	16	12	15	23	10	13	13	10	18
AA035	17	08	18	21	10	11	12	15	16
A019	17	09	17	23	10	14	13	13	13

Table 9: Combined Y chromosome microsatellite haplotype data in sorted format (continued).

SAMPLE ID	DYS 19	DYS 389 I	DYS 389 II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385 I	DYS 385 II
A008	17	09	17	25	10	13	12	14	18
AA024	17	09	18	21	10	11	14	17	18
A054	17	09	18	24	10	14	12	12	17
H025	17	10	15	23	10	11	13	12	12
AA057	17	10	16	21	10	11	13	16	17
C047	17	10	16	24	10	11	13	11	14
AA055	17	10	17	21	10	11	13	17	18
AA016	17	10	17	21	10	11	14	17	17
AA031	17	10	17	21	10	11	15	17	17
C055	17	10	17	25	10	11	13	11	14
AA032	17	10	18	21	10	11	13	15	19
AA017	17	10	18	21	10	11	13	17	17
AA052	17	10	18	21	10	11	15	17	18
AA030	17	10	18	21	10	11	15	17	19
C007	17	10	18	25	10	11	13	10	13
H033	17	11	17	21	11	11	15	18	18
A043	17	11	17	26	10	11	14	13	19
A027	17	11	18	24	10	11	13	13	18
C045	NR	09	17	25	10	11	13	11	14
A045	NR	NR	NR	NR	NR	NR	NR	NR	NR
A053	NR	NR	NR	NR	NR	NR	NR	NR	NR
C049	NR	NR	NR	NR	NR	NR	NR	NR	NR
C051	NR	NR	NR	NR	NR	NR	NR	NR	NR

NR = no result

Samples with identical repeat codes are represented in ***italicized bold face type***.

Table 10a: Number of shared repeat codes and haplotypes within each population.

Population(s)	Number of Shared Codes/Haplotypes	
	MSY1	**Microsatellites**
Caucasian	0	2
African American	0	3
Asian	1	2
Hispanic	2	3
Total Shared Haplotypes	3	10

Minisatellite Data (MSY1) n = 180
Microsatellite Data n = 240

Table 10b: Number of shared repeat codes and haplotypes across all ethnic borders.

Population(s)	Number of Shared Codes/Haplotypes	
	MSY1	**Microsatellites**
Caucasian	0	2
African American	0	2
Asian	1	2
Hispanic	1	2
Caucasian\Hispanic	5	1
African American\Hispanic	2	2
Total Shared Haplotypes	9	11

Minisatellite Data (MSY1) n = 180
Microsatellite Data n = 240

Table 11a: Minisatellite Fortuitous Matches

SAMPLE ID	MSY1 Code	DYS19	DYS389 1	DYS389 11	DYS390	DYS391	DYS392	DYS393	DYS 385 1	DYS 385 11
C024	1(14) 3(37) 4(20)	14	10	17	24	11	13	13	11	14
H005	1(14) 3(37) 4(20)	14	10	16	24	11	13	13	11	14
C002	1(15) 3(39) 4(19)	14	11	17	24	10	13	13	11	15
H051	1(15) 3(39) 4(19)	14	10	16	25	11	13	13	11	14
C050	1(16) 3(37) 4(20)	14	10	16	24	11	13	12	12	14
H002	1(16) 3(37) 4(20)	14	10	16	24	11	13	13	10	11
AA069	1(16) 3(38) 4(20)	14	10	16	24	11	13	13	11	14
H053	1(16) 3(38) 4(20)	14	11	17	24	11	13	13	11	14
AA013	1(16) 3(40) 4(18)	15	11	16	23	11	13	13	11	13
H001	1(16) 3(40) 4(18)	14	10	16	24	11	14	13	11	14
H008	1(16) 3(40) 4(18)	14	11	16	24	11	13	13	12	14
C016	1(16) 3(40) 4(19)	14	10	17	23	10	11	12	13	18
H015	1(16) 3(40) 4(19)	14	10	16	23	10	13	13	11	14
H004	1(20) 3(1) 1(3) 3(35) 4(25)	16	11	16	23	11	13	13	12	14
H054	1(20) 3(1) 1(3) 3(35) 4(25)	16	11	16	23	12	13	13	12	14
C019	3(1) 1(17) 3(31) 4(20)	16	10	18	24	10	11	13	13	15
H034	3(1) 1(17) 3(31) 4(20)	14	10	19	24	10	14	13	14	17
A031	3(4) 1(12) 3(52) 4(10)	16	10	15	22	10	14	12	11	11
A034	3(4) 1(12) 3(52) 4(10)	15	11	15	23	10	13	13	10	18

Table 11b: Microsatellite Fortuitous Matches

SAMPLE ID	MSY1 Code	DYS19	DYS389 I	DYS389 II	DYS390	DYS391	DYS392	DYS393	DYS 385 I	DYS 385 II
C038	3(4) 1(5) 3(1) 1(6) 3(38) 4(20)	15	10	16	23	9	11	12	13	17
C039	1(18) 3(35) 4(1) 3(1) 4(16)	15	10	16	23	9	11	12	13	17
C013	1(16) 3(39) 4(20)	14	10	16	24	11	13	13	12	14
C020	1(14) 3(38) 4(1) 3(2) 4(1) 3(2) 4(18)	14	10	16	24	11	13	13	12	14
AA008	NR	15	10	17	21	10	11	15	16	17
AA025	1(8) 3(34) 0(11) 4(2)	15	10	17	21	10	11	15	16	17
AA012	1(1) 3(2) 1(5) 3(33) 0(15) 4(2)	16	10	17	21	10	11	15	17	18
AA051	3(3) 0(?) 4(2)	16	10	17	21	10	11	15	17	18
AA053	3(3) 1(5) 3(33) 0(14) 4(2)	16	10	17	21	10	11	15	17	18
A001	3(1) 1(18) 3(43) 4(5)	15	9	16	23	10	14	13	13	13
A025	3(1) 1(18) 3(41) 4(8)	15	9	16	23	10	14	13	13	13
A003	3(3) 1(15) 3(45) 4(14)	15	9	16	24	10	14	12	14	18
A041	3(4) 1(13) 3(41) 4(12)	15	9	16	24	10	14	12	14	18
H039	NR	14	10	16	24	11	13	13	11	13
H048	1(15) 3(38) 4(20)	14	10	16	24	11	13	13	11	13
C022	1(16) 3(43) 4(17)	14	10	16	23	10	13	13	11	14
H015	1(16) 3(40) 4(19)	14	10	16	23	10	13	13	11	14
H014	1(16) 3(40) 4(17)	14	11	16	24	11	13	13	11	14
H042	1(15) 3(?) 4(20)	14	11	16	24	11	13	13	11	14
H043	1(16) 3(?) 4(20)	14	11	16	24	11	13	13	11	14
AA013	1(16) 3(40) 4(18)	15	11	16	23	11	13	13	11	13
H023	1(17) 3(42) 4(15)	15	11	16	23	11	13	13	11	13

Table 11b: Microsatellite Fortuitous Matches (continued)

SAMPLE ID	MSY1 Code	DYS19	DYS389 1	DYS389 11	DYS390	DYS391	DYS392	DYS393	DYS 385 1	DYS 385 11
AA011	NR	14	10	16	24	11	13	13	11	14
AA069	1(16) 3(38) 4(20)	14	10	16	24	11	13	13	11	14
H005	1(14) 3(37) 4(20)	14	10	16	24	11	13	13	11	14
H013	1(15) 3(2) 1(2) 3(39) 4(18)	14	10	16	24	11	13	13	11	14
H020	NR	14	10	16	24	11	13	13	11	14
H060	1(16) 3(34) 4(20)	14	10	16	24	11	13	13	11	14

NR = No result

Table 12a: MSY1 AMOVA calculations.

Source of Variation	Percent Variation
Among Populations	11.17
Within Populations	88.83
F-statistic	F_{ST} : 0.11166

Table 12b: Microsatelite AMOVA calculations

Source of Variation	Percent Variation
Among Populations	10.11
Within Populations	89.89
F-statistic	F_{ST} : 0.10108

Table 13: Non-probative case study results.

Sample ID	Description	MSY1 Result
N1	Q	NR
N2	Q	0(1) 3(1) 1(7) 3(30) 0(13) 4(2)
N3	Q	0(1) 3(1) 1(7) 3(30) 0(13) 4(2)
N4	Q	NR
N5	Q	NR
N6	K	0(1) 3(1) 1(7) 3(30) 0(13) 4(2)
N7	K	1(16) 3(39) 4(21)
N8	K	0(1) 3(1) 1(6) 3(29) 0(13) 4(1)

Q = forensic unknown sample
K = known exemplar sample
NR = no result

Table 14: Previous differential extraction procedure concentrations of non-probative case samples.

Sample ID	Fraction	ng/µL	Yield (ng)/ 200 µL	MSY1 amplification estimate (ng/10µL)
N1	sperm fraction	0.075	15	5
N2	sperm fraction	0.031	6.2	2.1
N3	swab remain fraction*	0.25	50	16.6
N4	sperm fraction	0.625	125	41.7#
N5	sperm fraction	0.0075	1.5	0.5#

Table 14: Conversion of extraction yields from previous differential procedures performed during casework analysis on mixed samples N1-N5. Yields have been converted to estimated MSY1 one-step amplification concentrations in a 10 µL template volume. Estimates are based upon extractions performed on a 1/3 clipping of a cotton swab or a 3 mm^2 clipping of a crime scene substrate (OCMEb). Based upon control experiments, sufficient yields are expected in samples N1-N4. Sample degradation is expected in samples N4 and N5 due to the presence of mold on substrates (refer to table 1).

* Swab remain fraction refers to the extraction performed on residual cells which inherently adhere to the sample substrate.

#Sample concentration presumed to be far less than estimated value do to bacterial degradation.

Section VII: Figures

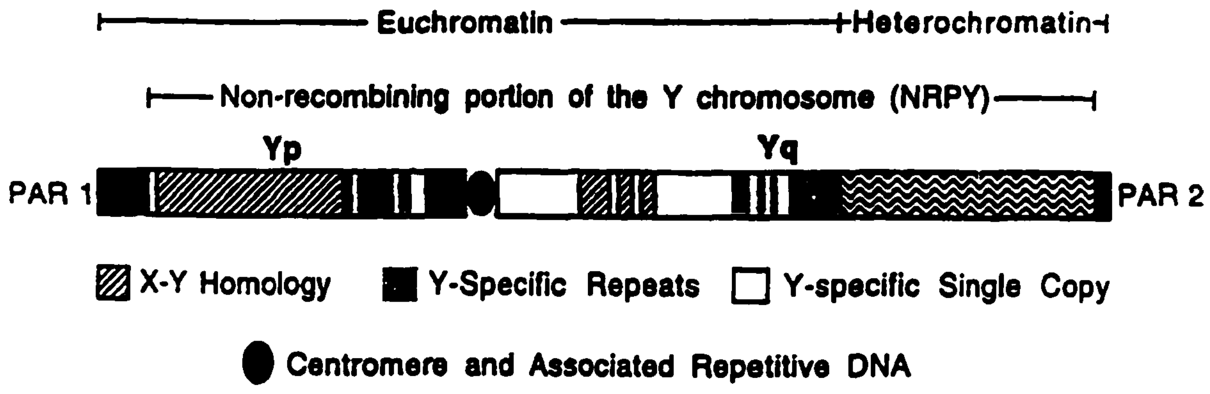


Figure 1: Schematic representation of the human Y chromosome. Recombination occurs at two small pseudoautosomal regions which are represented by darkened boxes at the two ends of the chromosome (PAR1 and PAR2). PAR1 and PAR2 are homologous to sequences on the X chromosome and undergo normal recombination. The remainder of the chromosome displays normal paternal inheritance with no observed recombination. Figure 1 was adapted from Hammer, *et al.*, 1997.

Figure 2a:

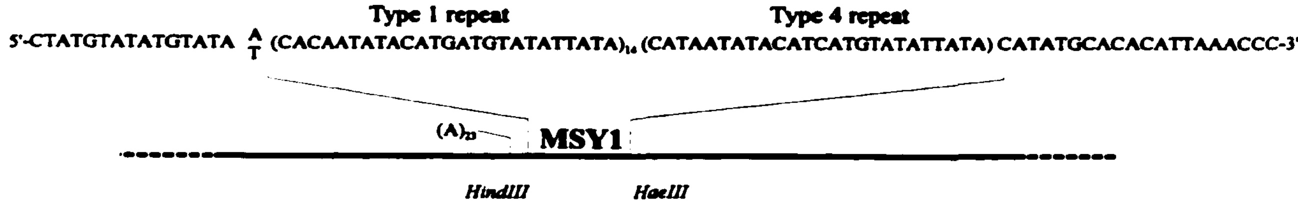


Figure 2b:

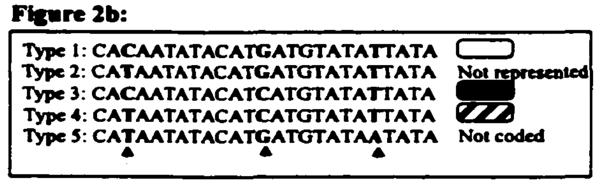


Figure 2c:

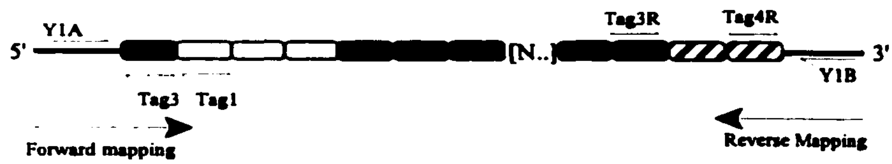


Figure 2d:

Primer	Sequence	Emission
Y1A	5'-ACA GAG GTA GAT GCT GAA GCG GTA TAG C-3'	None
Tag1	[6 FAM[1]] 5'-tca tgc gtc cat ggt ccg gaT GTG TAT AAT ATA CAT CAT GTA TAT TG-3'	Blue
Tag2	[VIC[2]] 5'-tca tgc gtc cat ggt ccg gaC ATC ATG TAT ATT ATG TAT AAT ATA CAT C-3'	Green
Tag3	[NED[3]] 5'-tca tgc gtc cat ggt ccg gaT GTG TAT AAT ATA CAT GAT GTA TAT TG-3'	Yellow
Tag4	[ROX[4]] 5'-tca tgc gtc cat ggt ccg gaC ATG ATG TAT ATT ATG TAT AAT ATA CAT G-3'	Red
Y1B	5'-GCA ACT CAA GCT AGG ACA AAG GGA AAG G- 3'	None
Tag1R	[6 FAM] 5'-tca tgc gtc cat ggt ccg gaC ATG ATG TAT ATT ATA CAC AAT ATA CAT G-3'	Blue
Tag3R	[NED] 5'-tca tgc gtc cat ggt ccg gaC ATC ATG TAT ATT ATA CAC AAT ATA CAT C-3'	Yellow
Tag4R	[ROX] 5'-tca tgc gtc cat ggt ccg gaC ATC ATG TAT ATT ATA CAT AAT ATA CAT C-3'	Red

Figure 2: Organization of the MSY1 locus. (2a) Schematic representation of the restriction map depicting an 890 bp HindIII-HaeIII subclone containing MSY1. Two types of repeat units are represented in the fragment. (2b) Sequences representing the 5 variant repeat types identified. (2c) MVR-PCR displaying both forward and reverse typing. A single tagged discriminator is used in each reaction with the appropriate flanking primer. Type 2 and type 5 coding are not represented. (2d) Primer sequences of the MVR-PCR four-state system. Flanking primers (Y1A and Y1B) and type discriminating primers (Tag1-4 and Tag 1R, 3R-4R) were purchased from Applied Biosystems (ABI). The 5'-end label associated with each type discriminator is represented in brackets. A 5' tag sequence incorporated into each discriminating primer is represented in lowercase font. Figures 2a-c were adapted from Jobling, *et. al.*, 1998a.

[1]6-Carboxyfluorescine
[2,3]Trademarks of ABI. Available only through ABI.
[4](6)-Carboxy-x-rhodamine

Figure 3: MVR-PCR optimization results.

Figure 3a: Electropherogram representation of the forward (A) and reverse (B) multiplex reactions. (A) Preferential amplification of the type-1 repeat was present throughout the entire array with non-specific type-1 repeats far into the 3' end. Type-2, 3, or 4 repeats were not observed in the reaction. (B) Non-specific mapping results were also present in the reverse reaction. Type-1R and 3R repeats were specific to the array, however, low stringent type-3R fragments were present within the type-4R repeat block (captured within the white bracket).

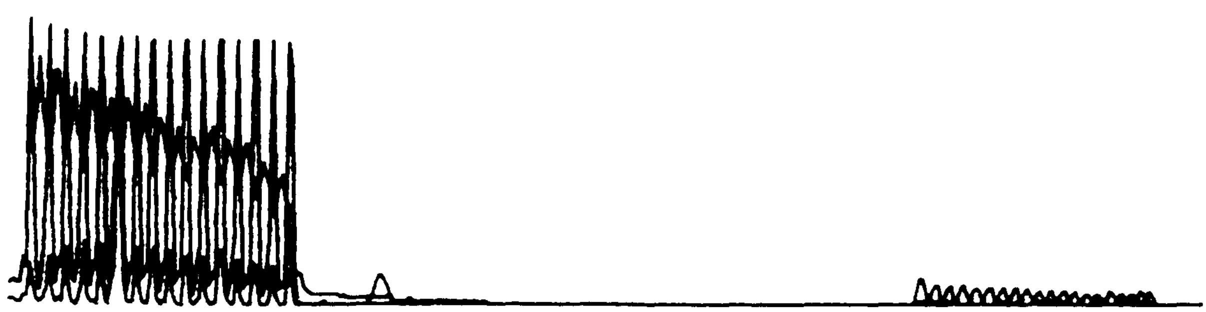


Figure 3b: Electropherogram representation of the tag-1,2,4 triplex. The addition of tag-2 did not inhibit the presence of type-1 or type-4 specific repeat blocks, nor did the addition of the tag-2 primer result in the presence of non-specific amplification throughout the entire repeat array as discussed in Figure 3a. Optimal fluorescent intensities were observed with a 100 nM tag-2 concentration. The electropherogram also depicts the system resolution using the modified gel conditions in the present study. The final type-4 fragment represents the 82nd repeat unit of this 2050 bp MSY1 allele.

Figure 4a: Modified Tag-3 experimentation.

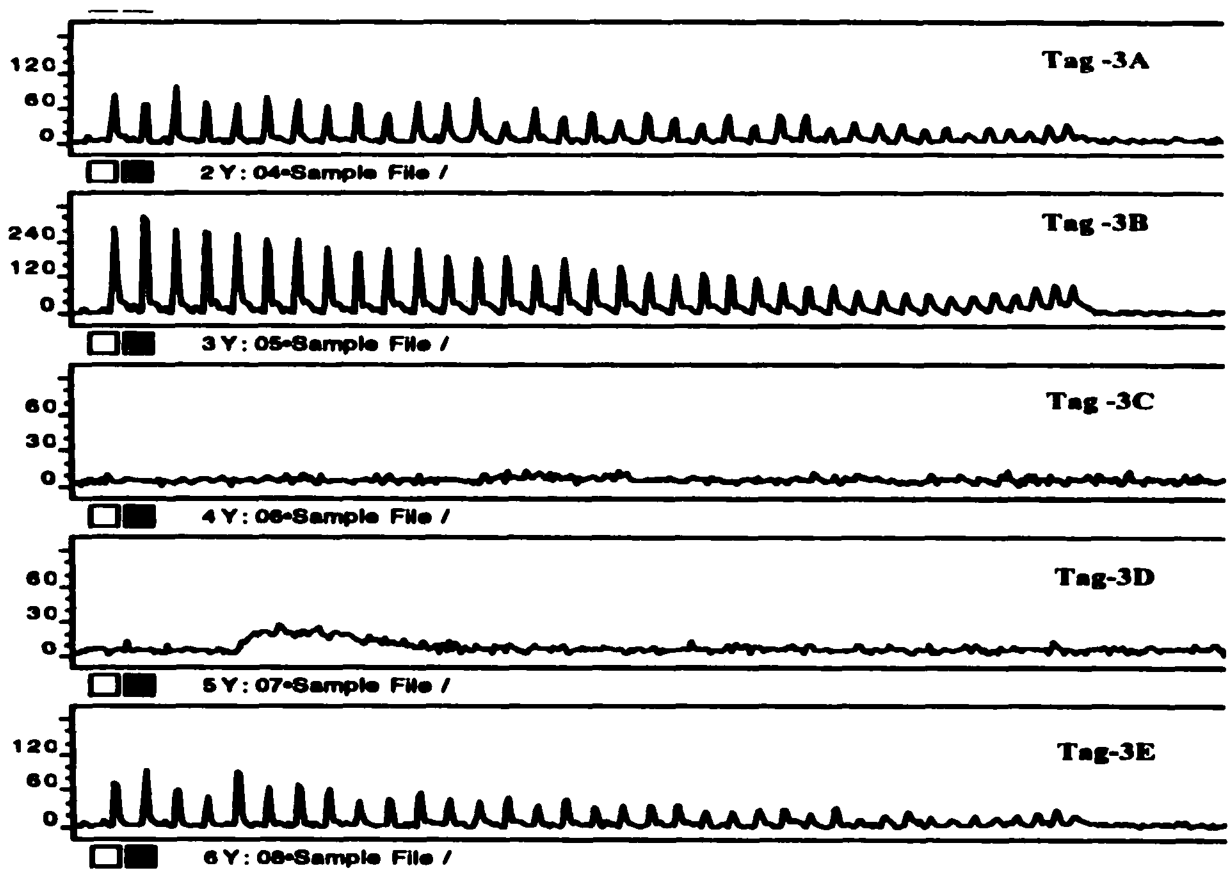


Figure 4a: Electropherogram representation of modified tag-3 MVR-PCR results. MVR-PCR was performed with 100 nM of each modified tag-3 primer (A-E). The tag-3A, B and E primers yielded results specific to the positive control sample, however, at decreased sensitivity. No repeat units were present in the tag-3C or D primer reactions. Of the three functional primers, tag-3B displayed the highest sensitivity. All 37 type-3 repeats specific to the positive control were present well above fluorescent background levels, allowing for complete mapping of the central type-3 block. The complete central block was also present in the primer tag-3A and E reactions, however at lower resolution.

Figure 4b: Modified Tag-3 experimentation (continued).

A
B
C
D

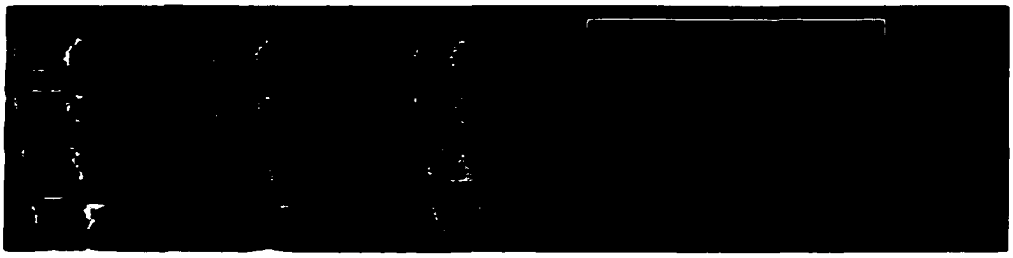


Figure 4b: MVR-PCR at an annealing temperature of 62°C with 25 nM tag-1, 100 nM tag-2, 75 nM tag-3B and increasing concentrations of tag-4 (lane A-25 nM, lane B-50 nM, lane C-75 nM, and lane D-100 nM). Increased type-1 specificity was present in the reactions at an annealing temperature of 62°C. The type-1 repeats displayed higher fluorescent emission as compared to the non-specific type-3 repeats, but this level of non-specificity resulted in ambiguous mapping of MSY1. A substantial decrease in type-3 fluorescent emission was noted. This decrease in emission resulted in the failure to code for the entire type-3 block. The first two type-3 repeats are captured within the white bracket.

Figure 5: Comparison of Chelex-100 and P:C:I extraction procedures.

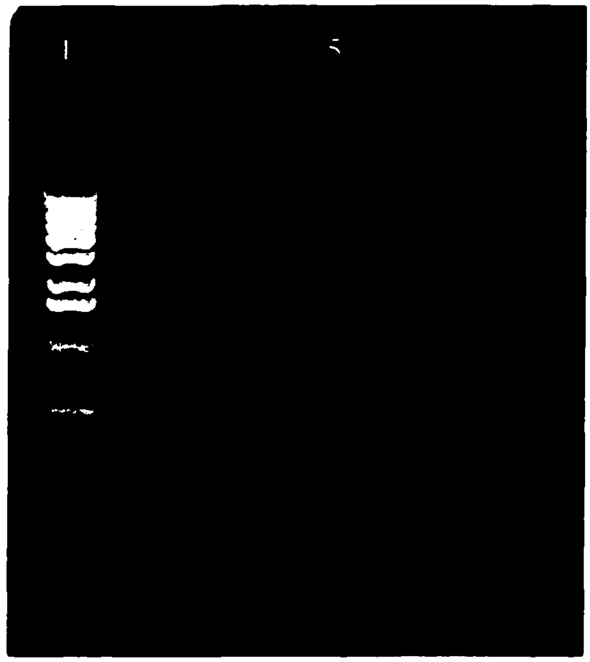


Figure 5a: Product gel containing MSY1 amplified fragments from PCI extracted DNA (lane 2) and Chelex-100 extracted DNA (lanes 4-8). MSY1 specific fragments are noted along with the invariant 203 bp homologue of MSY1 which coamplifies in the reaction (Jobling, *et al.*, 1998a).

Figure 5b: Product gel containing P:C:I extracted titration results (62.5 ng, 50 ng, 25 ng, 10 ng, 5 ng, 2.5 ng and 1 ng, lanes 3-9 respectively) and postmortem amplifications (lanes 11-14). Amplified MSY1 fragments were present down to 2.5 ng of control DNA (lane 8). Postmortem bloodstains displayed higher concentrations of amplified MSY1 fragments in the P:C:I extraction procedure (lanes 11-14) as compared to the Chelex-100 based extraction depicted above.

Figure 6: Product gel representation of sperm cell extractions. Comparison of mixed sample extraction and female negative sperm cell extraction.

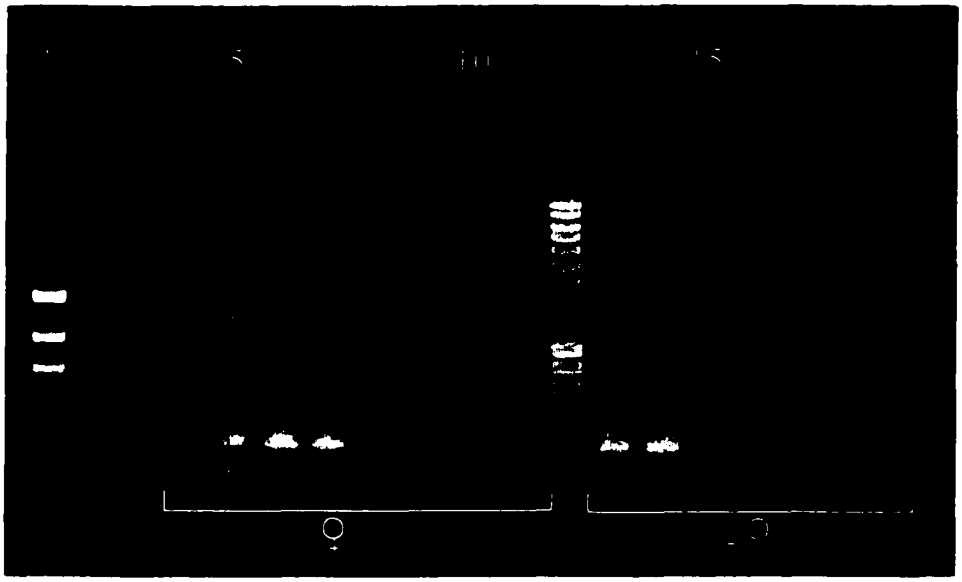


Figure 6: Product gel containing the mixed sample extractions and unmixed sperm cell control extractions. The lanes contain the following: (lane 2) amplification negative, (lane 3) reagent blank, (lane 4) unmixed vaginal cells (female-negative control), (lanes 5-11) the mixed samples containing100 ng, 75 ng, 50 ng, 25 ng, 10 ng, 5 ng, and 2.5 ng sperm cell DNA respectively, (lanes 13-19) the unmixed sperm cell extraction experiments containing 100 ng, 75 ng, 50 ng, 25 ng, 10 ng, 5 ng, and 2.5 ng sperm cell DNA respectively. Increased background was present in the mixed sample extraction experiment when compared to the unmixed sperm cell extraction experiment. This increase in background is attributed to unamplified degraded female template DNA. MSY1 amplification was present between the concentrations of 10 ng and 2.5 ng of sperm cell DNA, with low levels of amplification in the 2.5 ng control sample (the fragment was faintly visible to eye under UV but could not be represented photographically). In the unmixed sperm cell extraction experiment, amplified fragments specific to MSY1 were present between the concentrations of 25 ng and 2.5 ng of sperm cell DNA, with low levels of amplification in the 5 ng and 2.5 ng control samples (the fragment was faintly visible to eye under UV but could not be represented photographically). Non-specific amplification products were present in all concentrations above 25 ng of sperm cell DNA in the absence of female DNA (lanes 13-15).

Figure 7a: Caucasian MSY1 repeat code structure sorted by repeat number.

Sample ID	Repeat Number
C023	38
C041	48
C052	54
C053	57
C059	66
C001	67
C036	67
C030	67
C014	68
C042	69
C019	69
C054	70
C024	71
C048	71
C039	71
C006	71
C004	72
C029	72
C050	73
C034	73
C002	73
C031	73
C037	74
C038	74
C012	74
C040	74
C016	75
C056	75

MSY1 Code

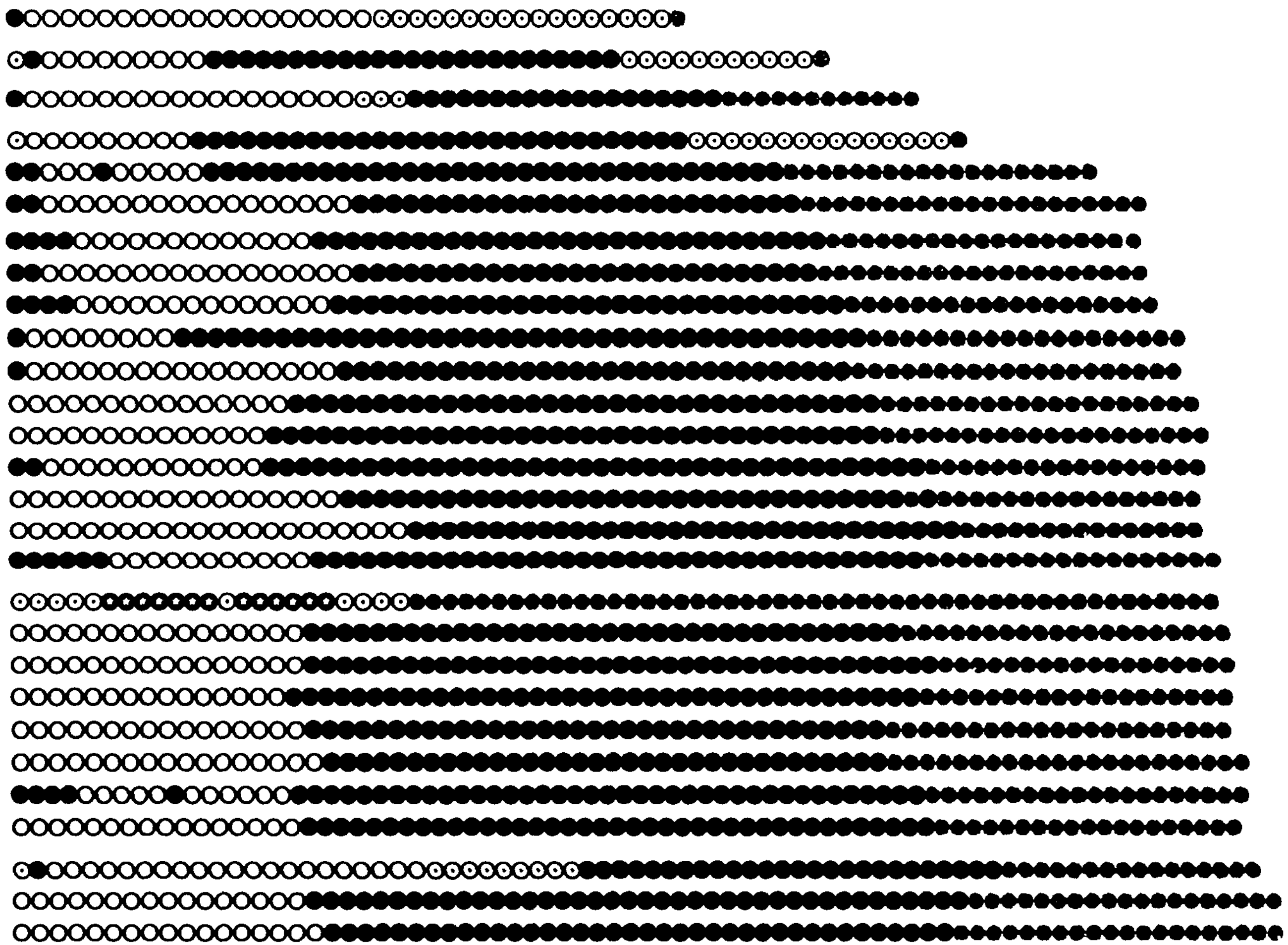
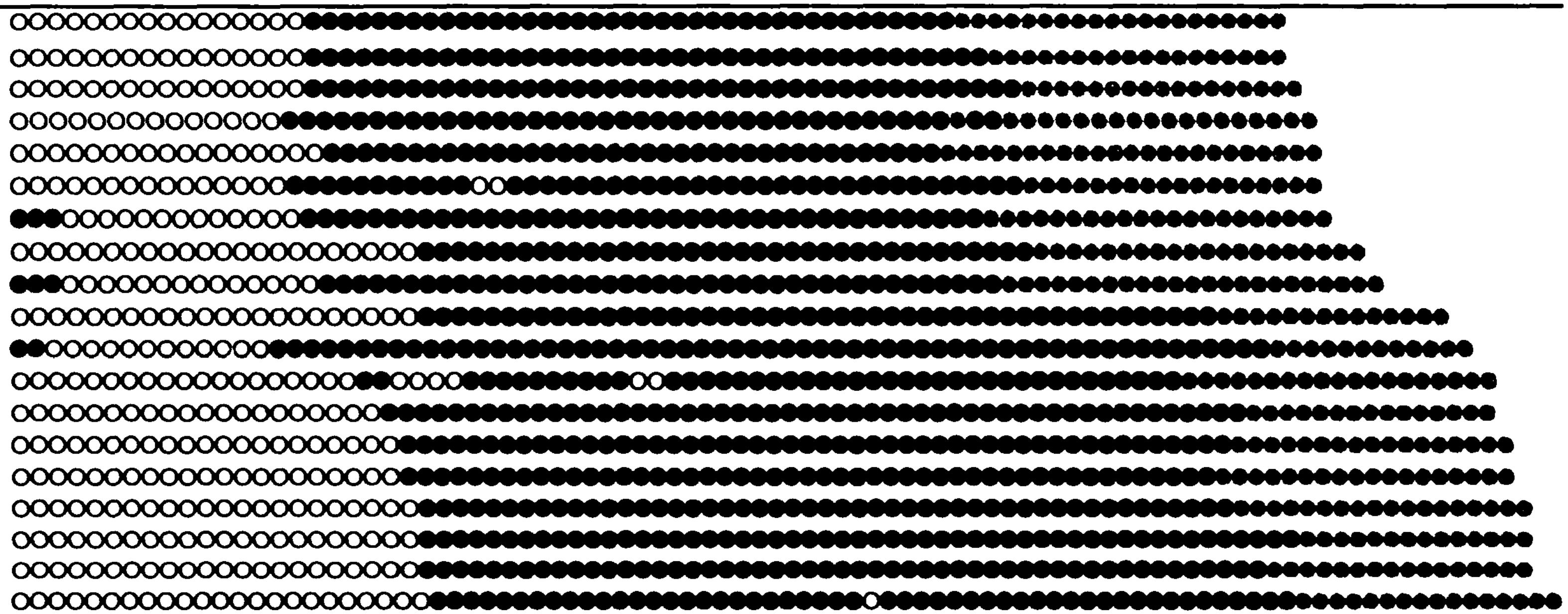


Figure 7a: Caucasian MSY1 repeat code structure sorted by repeat number (continued).

Sample ID	Repeat Number	MSY1 Code
C013	75	
C011	75	
C022	76	
C020	76	
C025	77	
C009	77	
C043	78	
C045	80	
C060	81	
C007	84	
C026	85	
C017	87	
C032	87	
C046	88	
C027	88	
C028	89	
C018	89	
C047	89	
C008	91	



Key: Type 1 Repeat ○; Type 2 Repeat ◘; Type 3 Repeat ●; Type 4 Repeat •; Null Repeat ⊙

No results were obtained for samples C003, 005, 010, 015, 021, 033, 035, 044, 049, 051, 055 and 057-58.

Figure 7b: African American MSY1 repeat code structure sorted by repeat number.

Sample ID	Repeat Number
AA003	21
AA042	49
AA033	50
AA040	51
AA066	52
AA064	52
AA025	55
AA035	55
AA070	55
AA020	55
AA016	55
AA026	56
AA052	56
AA053	57
AA056	57
AA030	57
AA028	57
AA012	58
AA055	58
AA058	58
AA034	59
AA017	59
AA023	60
AA037	60

MSY1 Code

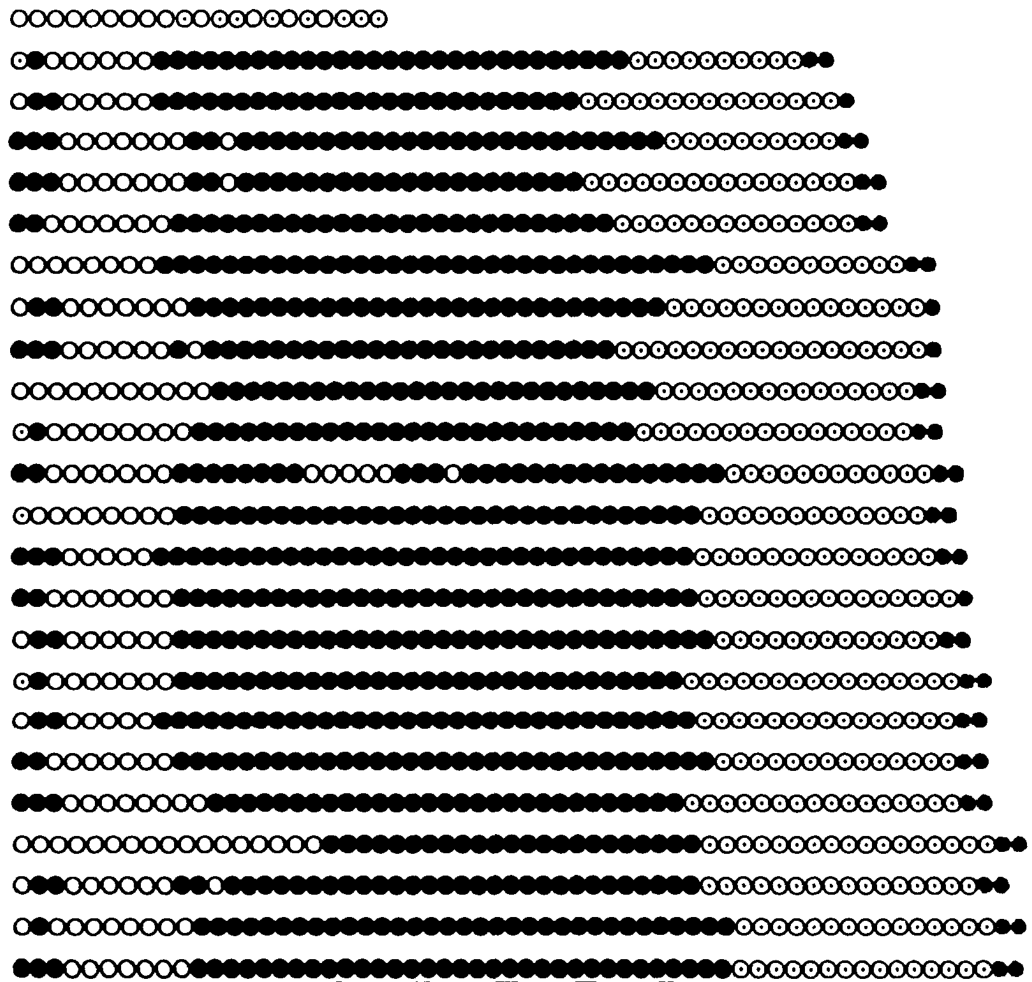


Figure 7b: African American MSY1 repeat code structure sorted by repeat number (continued).

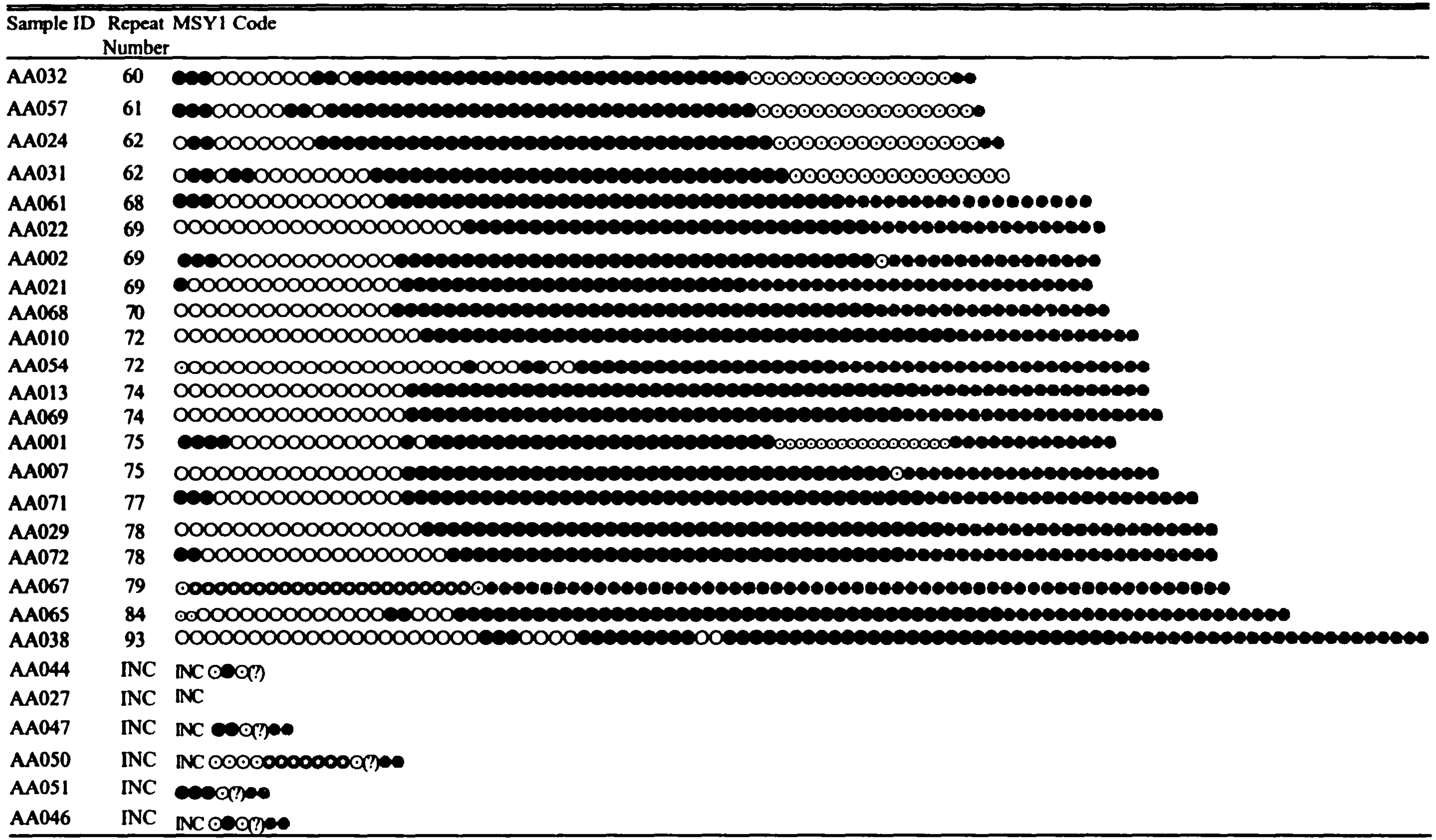


Figure 7b: African American MSY1 repeat code structure sorted by repeat number (continued).

Sample ID	Repeat Number	MSY1 Code
AA039	INC	INC ○○○○○○○○○○○○○○○●●●●⊙(?)
AA049	INC	INC ●●●⊙(?)•●
AA060	INC	INC ⊙⊙◉◉◉◉◉◉◉⊙◉◉◉◉◉◉⊙•(?)

Key: Type 1 Repeat ○; Type 2 Repeat ◉; Type 3 Repeat ●; Type 4 Repeat •; Null Repeat ⊙; Inconclusive INC

No results were obtained for AA004-006, 008-009, 011, 014-015, 018-019, 036, 041, 043, 045, 048, 059 and 062-063.

Figure 7c: Asian MSY1 repeat code structure sorted by repeat number.

Sample ID	Repeat Number
A043	58
A016	58
A033	62
A044	66
A039	66
A019	66
A009	66
A024	67
A042	67
A030	67
A001	67
A010	67
A025	68
A050	68
A040	68
A012	69
A051	70
A036	70
A041	70
A013	71
A022	72
A014	72
A035	73
A005	73
A004	73
A020	75
A023	75
A037	75
A003	77
A034	78

MSY1 Code

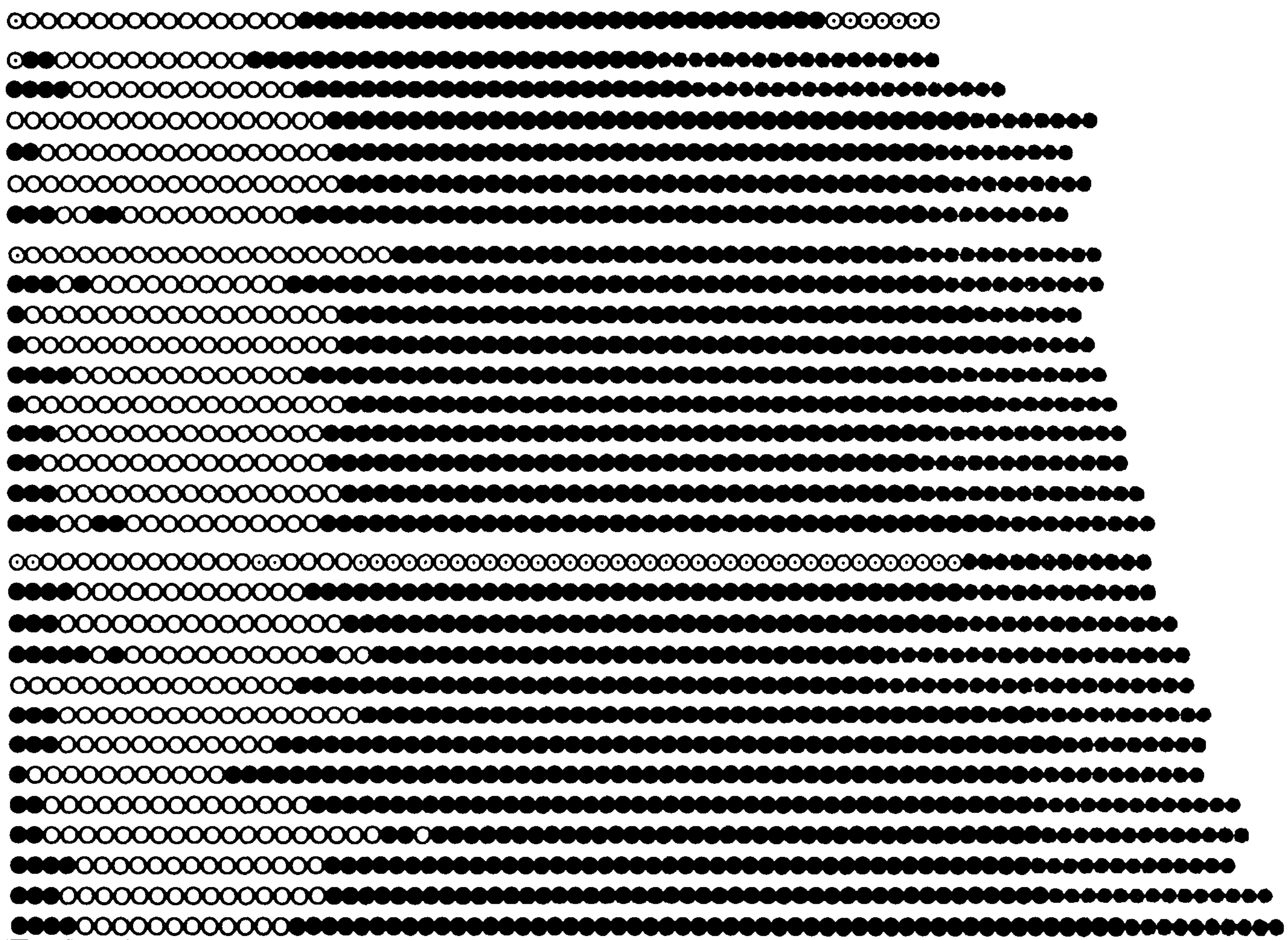
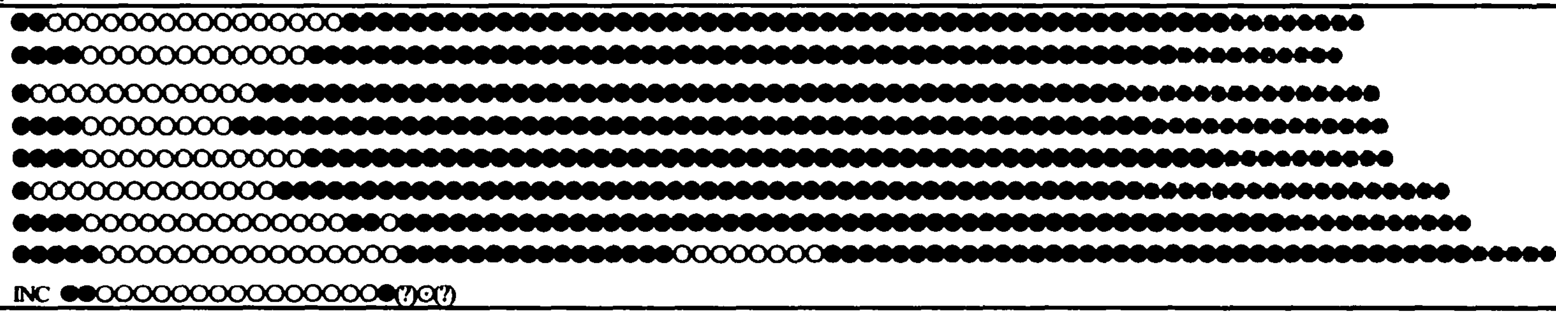


Figure 7c: Asian MSY1 repeat code structure sorted by repeat number (continued).

Sample ID	Repeat Number	MSY1 Code
A002	78	
A031	78	
A008	79	
A032	80	
A015	80	
A018	83	
A011	84	
A017	89	
A038	INC	INC



Key: Type 1 Repeat ○; Type 3 Repeat ●; Type 4 Repeat •; Null Repeat ⊙; Inconclusive INC

No results were obtained for A006-007, 021, 026-029 and 045-49.

Figure 7d: Hispanic MSY1 repeat code structure sorted by repeat number.

Sample ID	Repeat Number	MSY1 Code
H037	48	
H045	53	
H050	55	
H010	58	
H006	58	
H058	59	
H033	59	
H025	60	
H007	62	
H057	62	
H061	62	
H049	63	
H028	64	
H021	66	
H026	69	
H034	69	
H060	70	
H056	71	
H019	71	
H059	71	
H003	71	
H005	71	
H062	71	
H024	73	
H022	73	
H051	73	
H014	73	
H009	73	
H048	73	

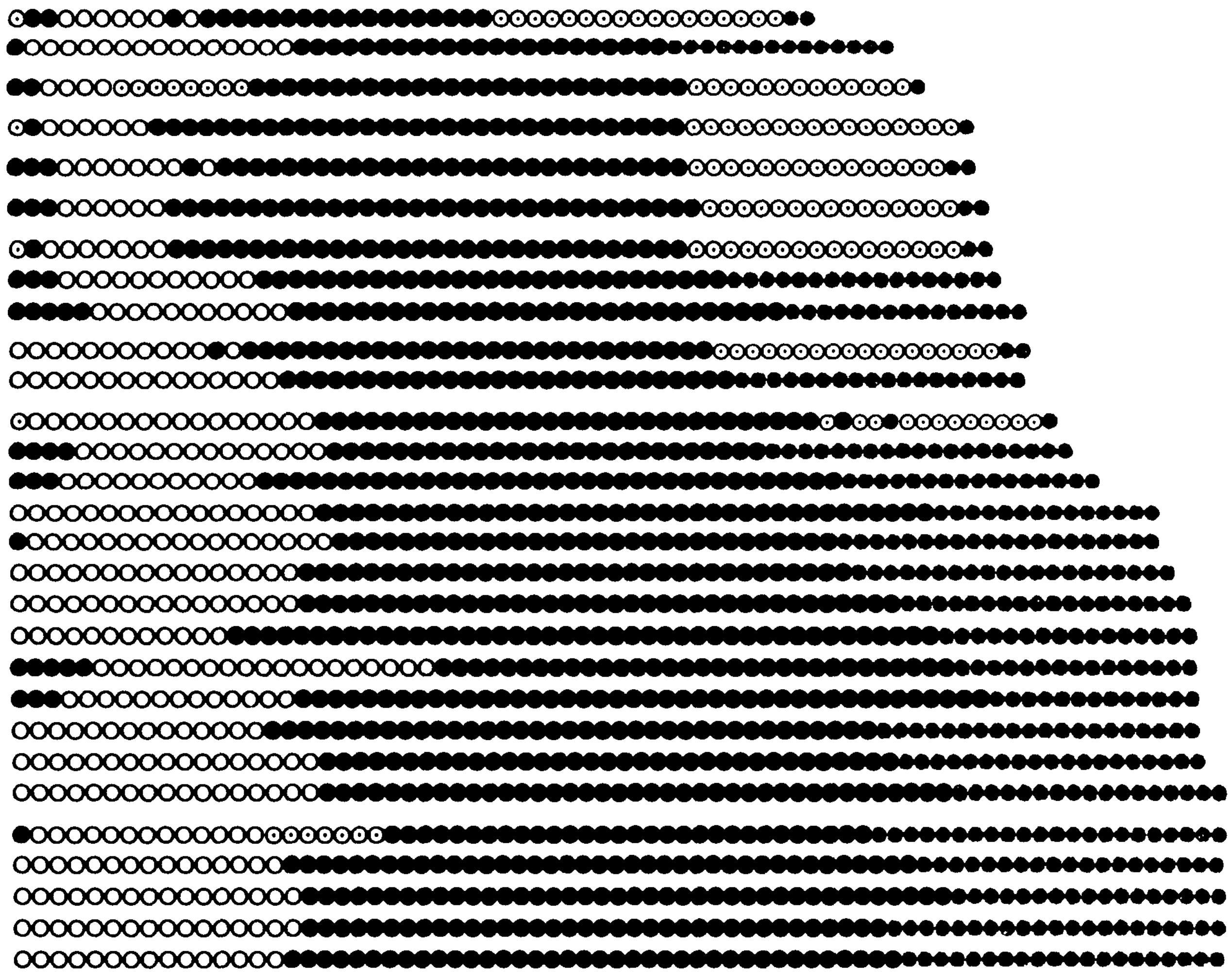


Figure 7d: Hispanic MSY1 repeat code structure sorted by repeat number (continued).

Sample ID	Repeat Number	MSY1 Code
H055	73	
H002	73	
H030	74	
H001	74	
H008	74	
H031	74	
H016	74	
H032	74	
H053	74	
H023	74	
H015	75	
H011	75	
H013	76	
H029	76	
H063	77	
H012	77	
H036	78	
H018	78	
H052	80	
H054	84	
H004	84	
H041	INC	INC
H042	INC	INC
H043	INC	INC

Key: Type 1 Repeat ○; Type 3 Repeat ●; Type 4 Repeat •; Null Repeat ⊙; Inconclusive INC

No results were obtained for H017, 020, 035, 038-040, 044 and 046-047.

Figure 8: MSY1 variability in New York City Caucasian, African American , Asian, and Hispanic populations.

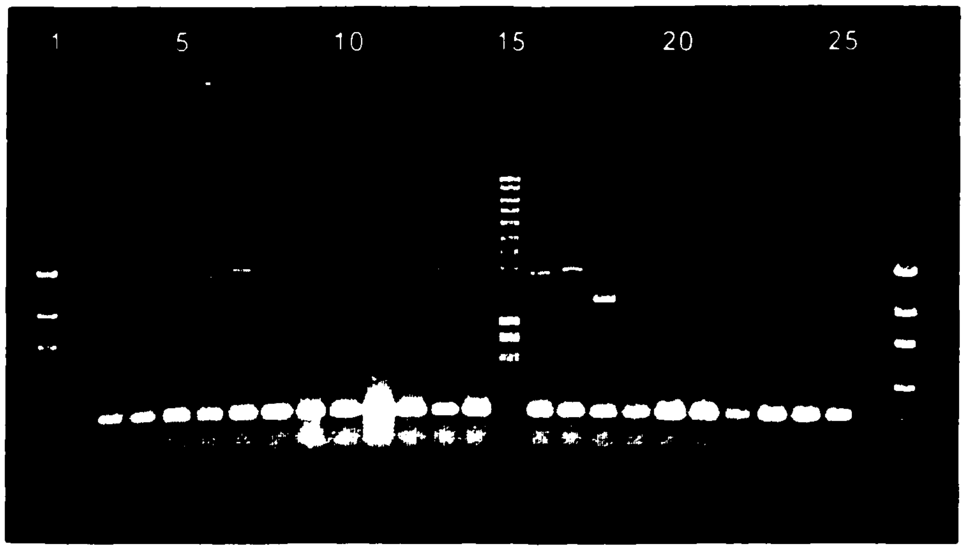


Figure 8a: MSY1 variability in Caucasian samples.

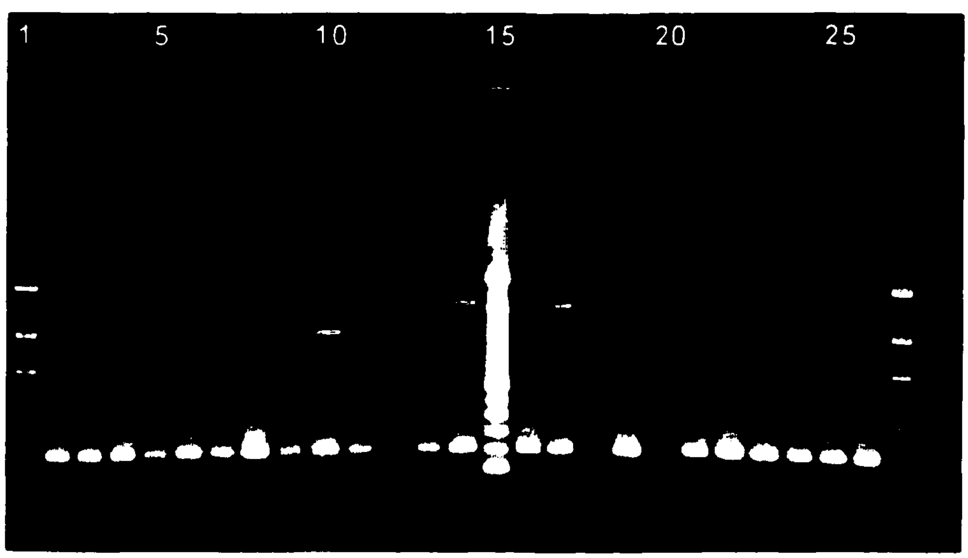


Figure 8b: MSY1 variability in African American samples

Figure 8: MSY1 variability in New York City Caucasian, African American , Asian, and Hispanic populations (continued).

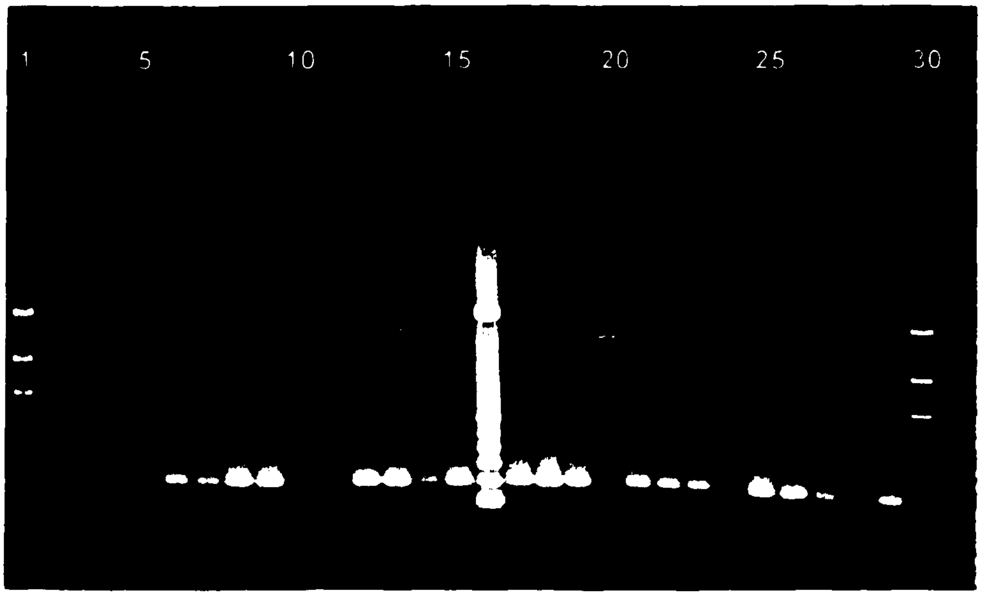


Figure 8c: MSY1 variability in Asian samples.

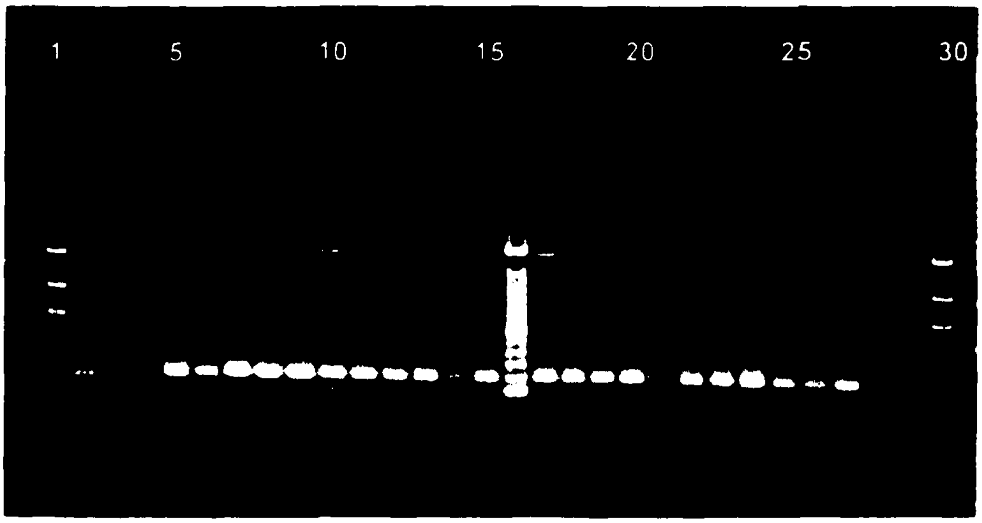


Figure 8d: MSY1 variability in Hispanic samples.

Figure 9a: Graphic representation of repeat number diversity in New York City Caucasian populations.

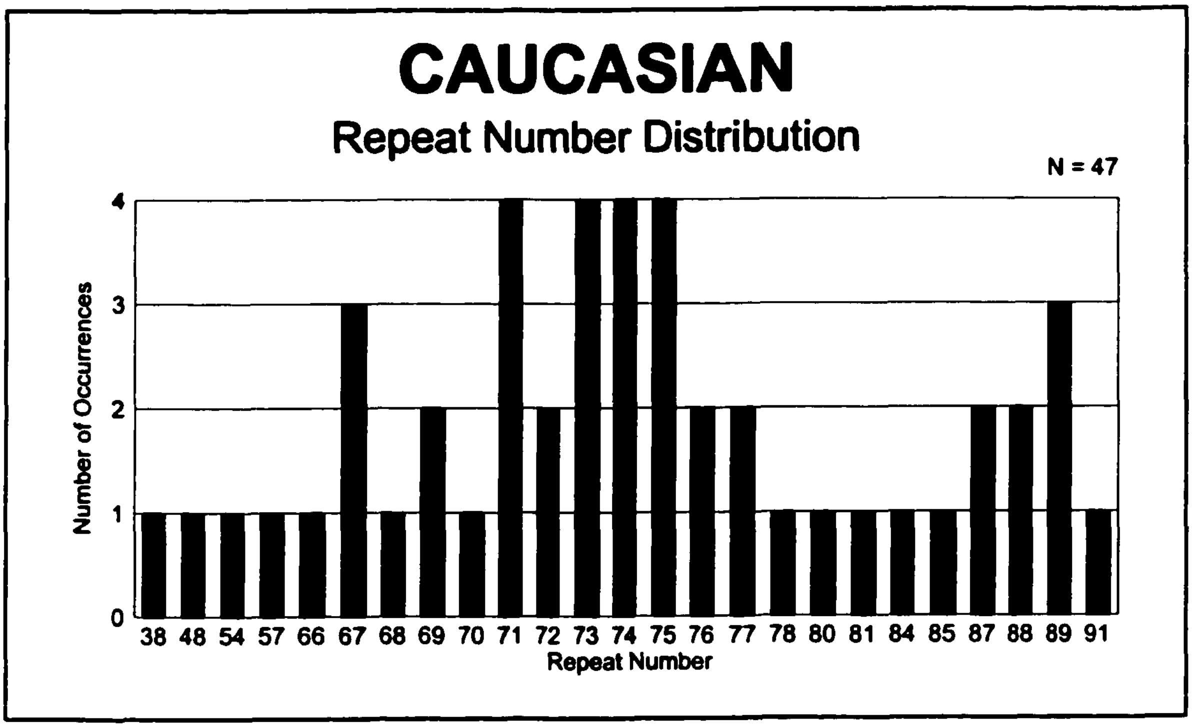


Figure 9a: The Caucasian population displayed repeat numbers between 38 and 91 repeats with the highest repeat frequency (n = 4) shared between 71, 73, 74 and 75 repeats.

Figure 9b: Graphic representation of repeat number diversity in New York City African American populations.

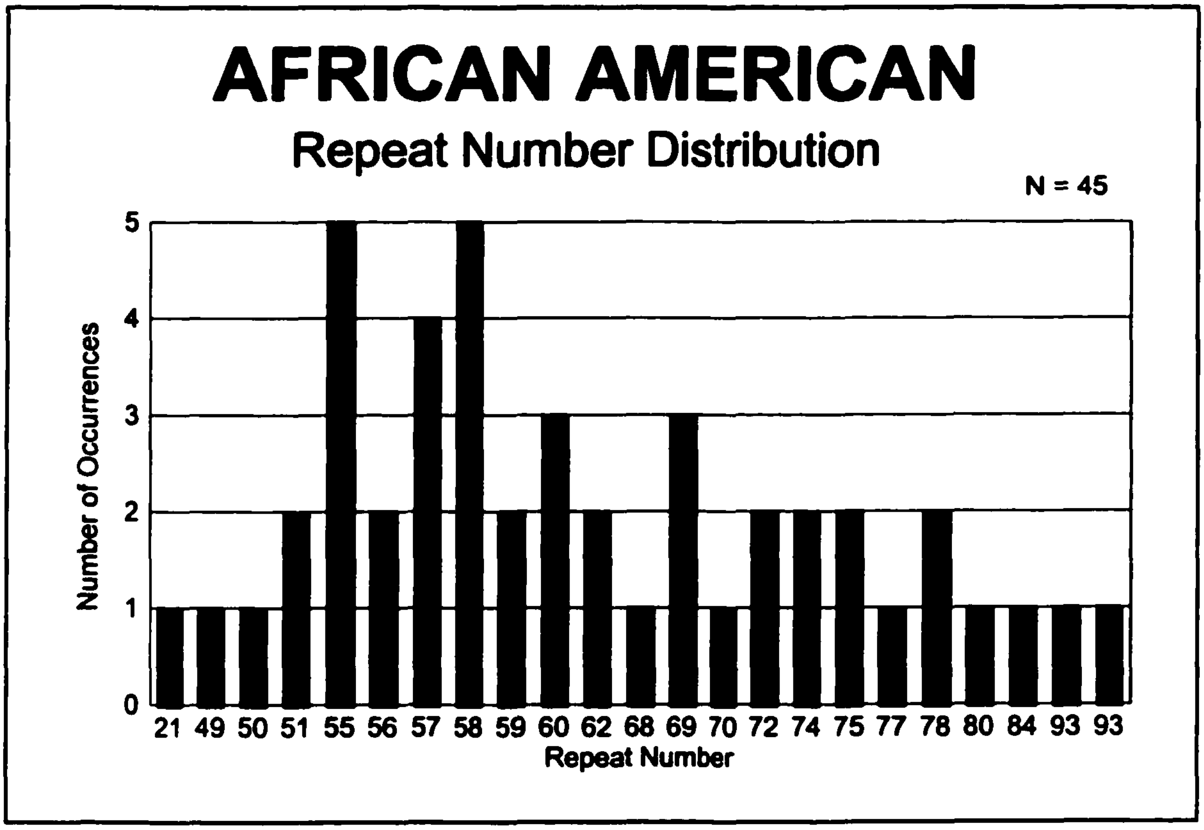


Figure 9b: African American samples displayed the highest range of repeat numbers, ranging between 21 and 93 repeats and a shared number of samples (n = 5) with the highest repeat frequency at 55 and 58 repeats.

Figure 9c: Graphic representation of repeat number diversity in New York City Asian populations.

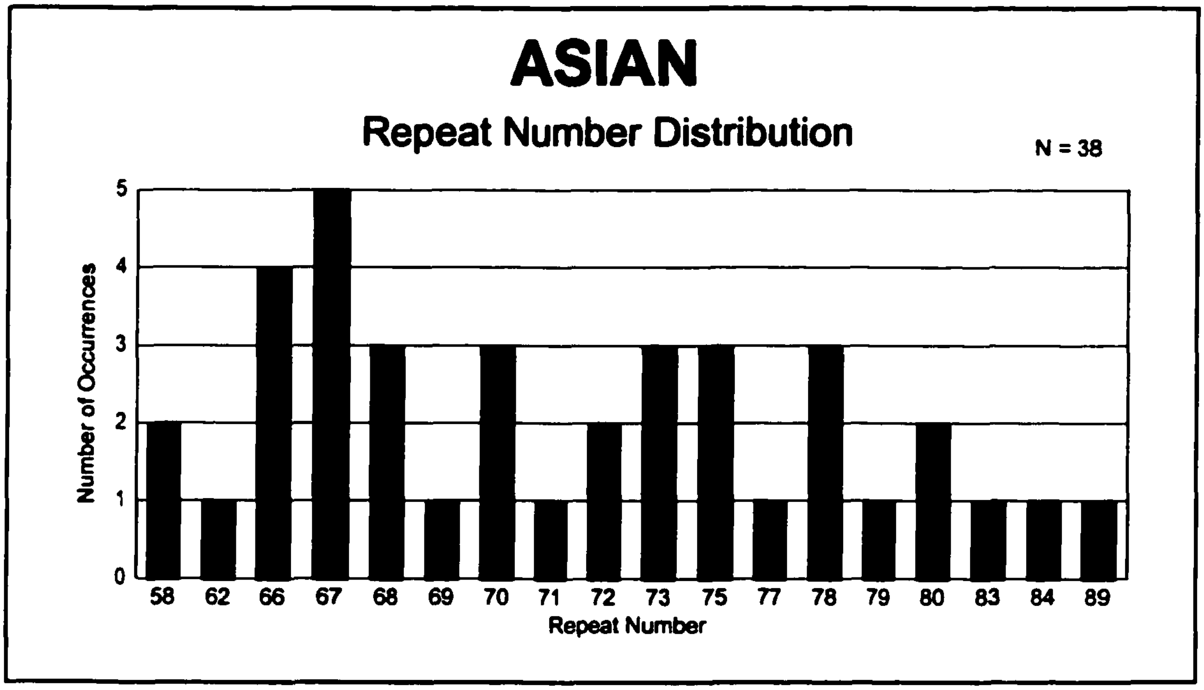


Figure 9c: Asian samples displayed the tightest range of repeat numbers with repeats between 58 and 89 repeats and 67 repeats being the most frequent at n = 5.

Figure 9d: Graphic representation of repeat number diversity in New York City Hispanic populations.

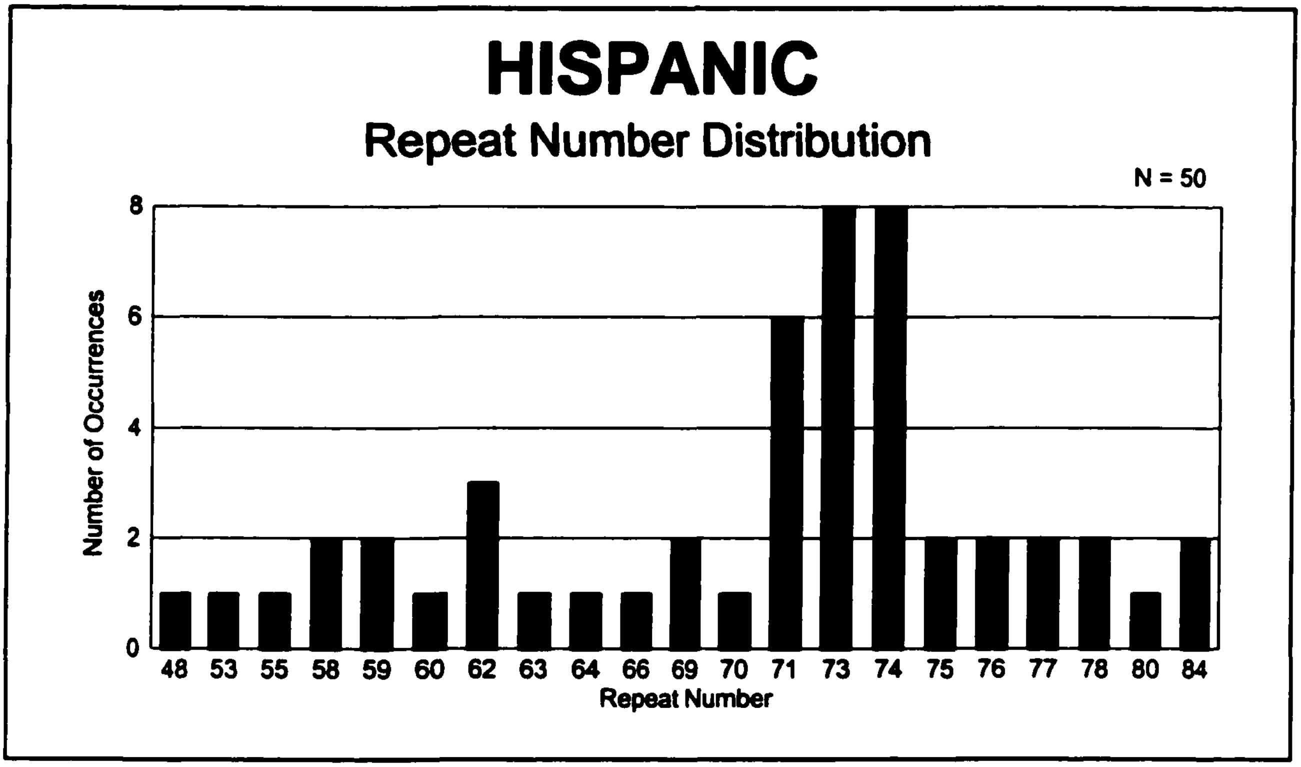


Figure 9d: The Hispanic population displayed repeat numbers between 48 and 84 repeats with the highest repeat frequency (n = 8) shared between 73 and 74 repeats.

Figure 10a: Combined New York City MSY1 repeat number diversity.

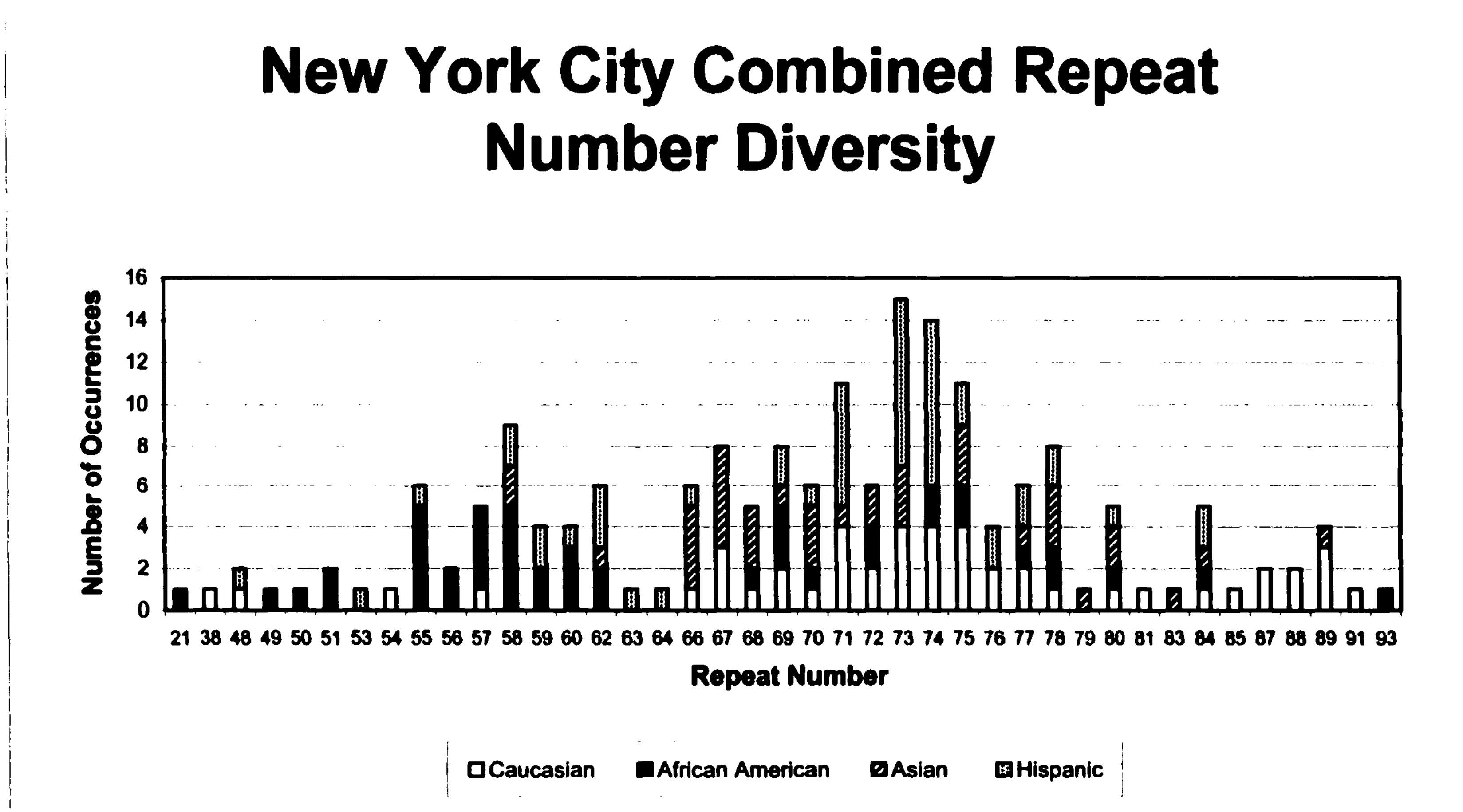


Figure 10a: The combined repeat number data from all four New York City populations displayed a repeat number range between 21 and 93 repeats for the 180 samples studied, with 60% residing within 66 and 78 repeats. African American population samples displayed a significant decrease in allele length as compared to Caucasian, Asian, and Hispanic populations, with a total of 23 samples (51.1% of the total number of African Americans studied) concentrated between 55 and 62 repeats.

Figure 10b: Combined New York City MSY1 repeat number diversity.

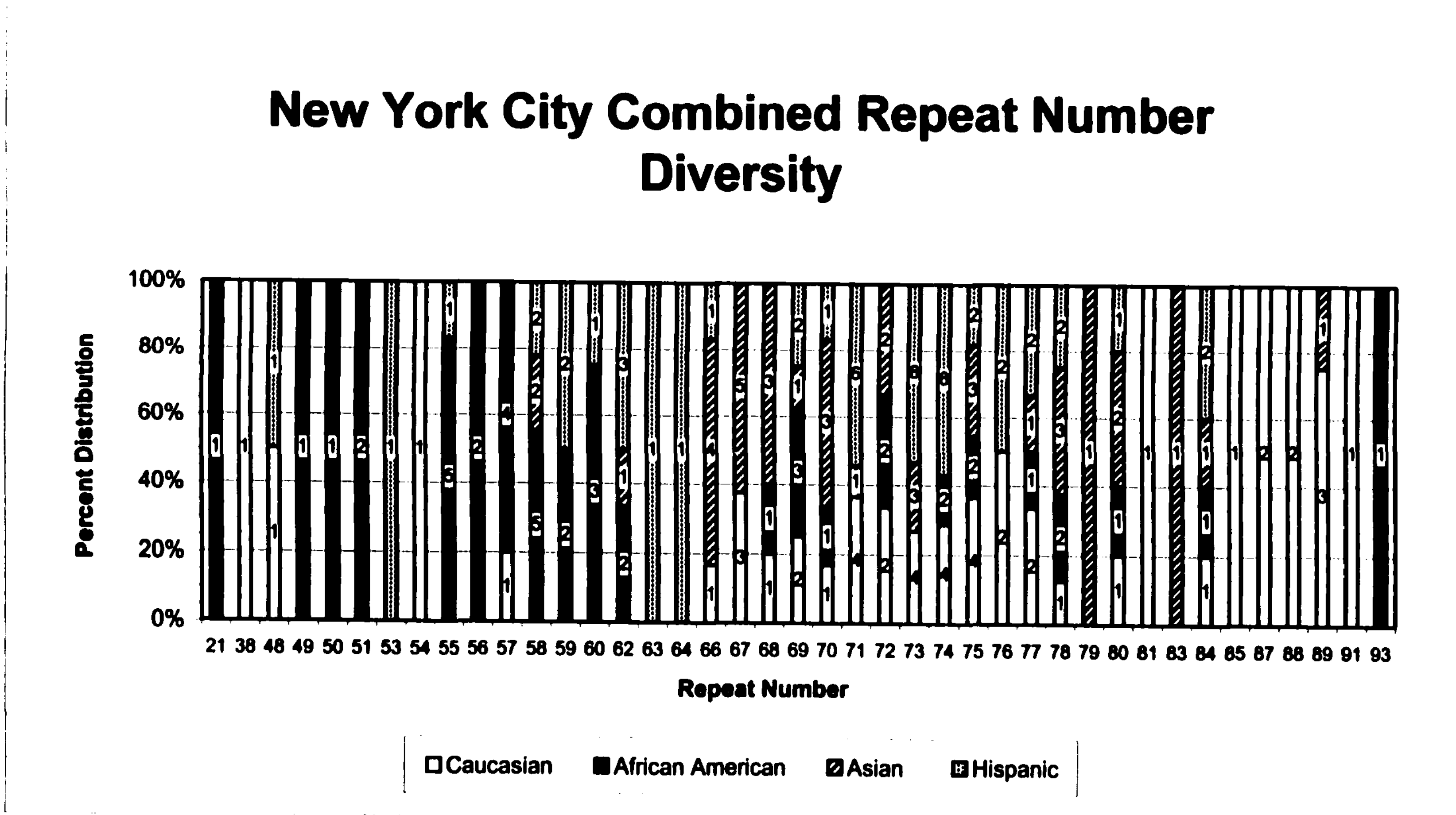


Figure 10b: Combined New York City MSY1 repeat number diversity represented graphically by percent distribution. The total number of samples for each population are displayed within each bar.

Figure 11a: Graphic representation of Caucasian modular structure distribution.

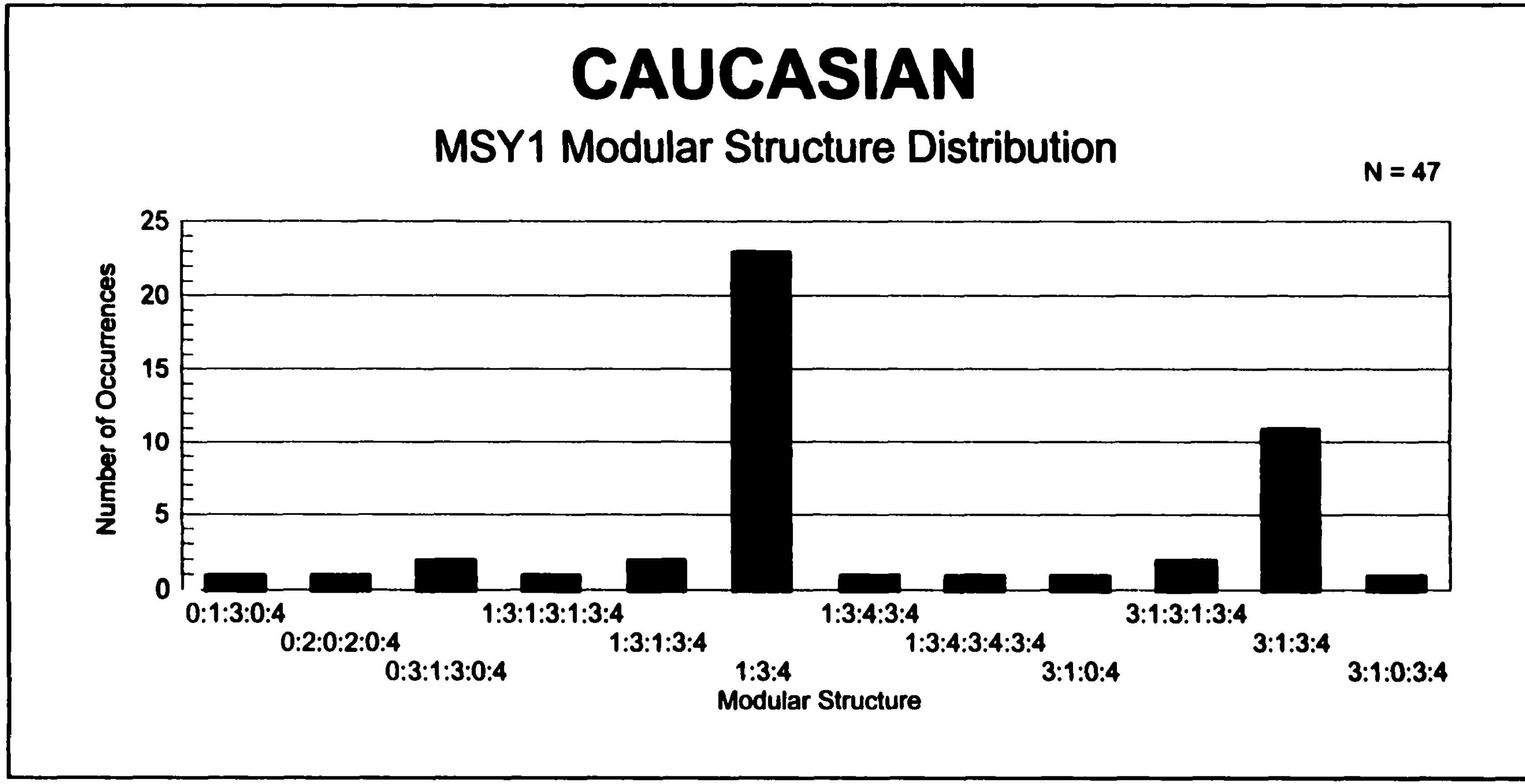


Figure 11a: The 1:3:4 modular structure was the most common in Caucasians with n = 23, a frequency of 48.9%. The 3:1:3:4 modular structure was the second most common structure with n = 11, a frequency of 23.4%. Caucasians were determined to be genetically similar to New York City Hispanics. Similar structural frequencies were observed in both populations.

Figure 11b: Graphic representation of African American modular structure distribution.

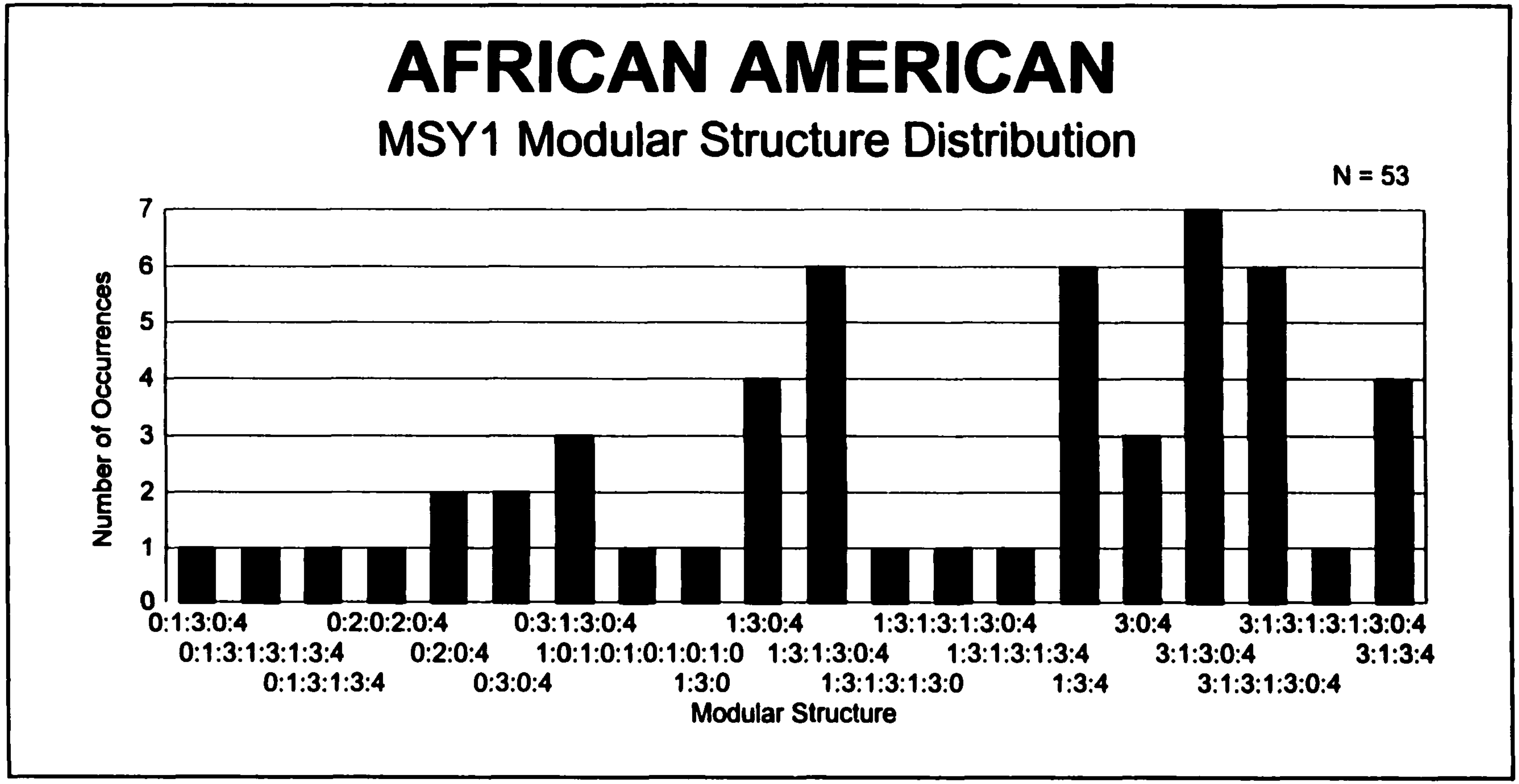


Figure 11b: Twenty modular structures were observed in a total of 53 individuals, making African Americans the most diverse population in New York City. This level of structural diversity is consistent to global MSY1 data from African populations in Jobling, *et, al.*, 1998a. A defined repeat structure was not observed in New York City African Americans, however, distinct repeat codes containing a high frequency of null alleles were observed; findings consistent to African populations (Jobling, *et. al.*, 1998b).

Figure 11c: Graphic representation of Asian modular structure distribution.

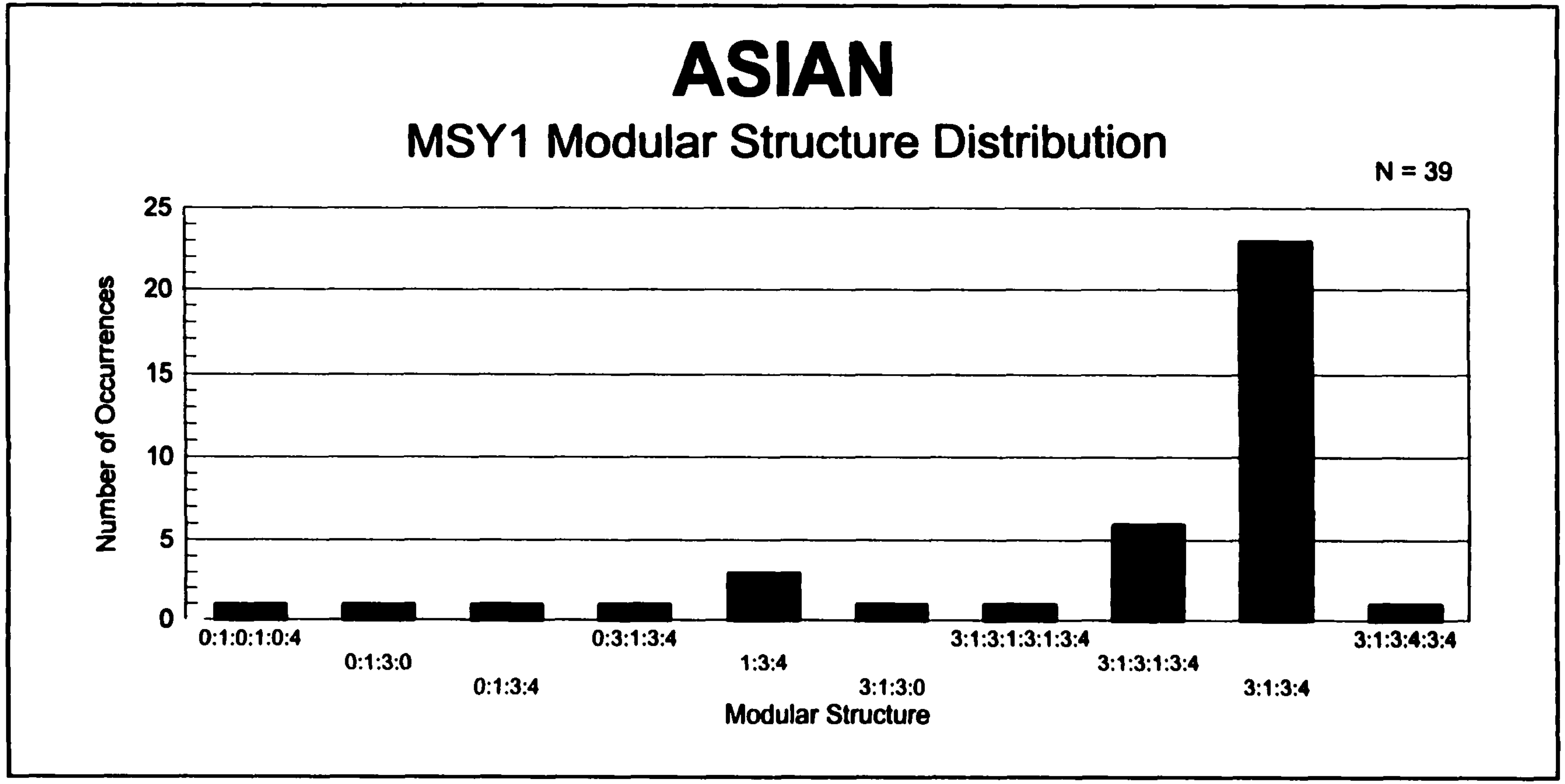


Figure 11c: The 3:1:3:4 modular structure was the most prevalent structure in Asian population samples with n = 23, a frequency of 60.0%. The second most common repeat structure was a 3:1:3:1:3:4 modular structure with n = 6, a frequency of 15.4%. The 3:1:3:1:3:4 repeat structure may be the result of a mutational extension of the major 3:1:3:4 structure (Jobling, *et al.*, 1998a).

Figure 11d: Graphic representation of Hispanic modular structure distribution.

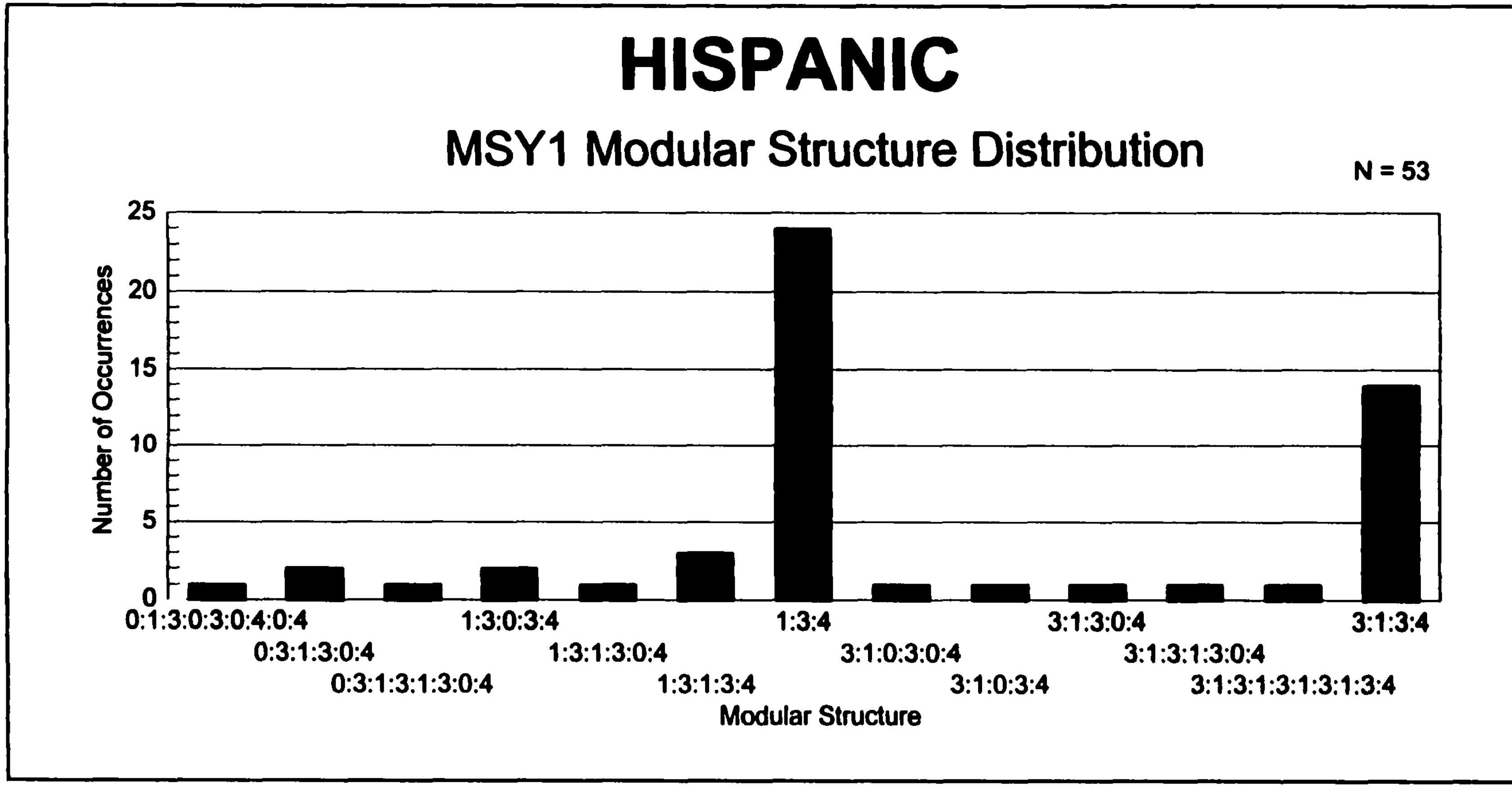


Figure 11d: The lowest degree of modular structure diversity was observed in New York City Hispanics, results consistent with microsatellite data (Zawacki, 2001). The 1:3:4 modular structure was the most common in Hispanics with n = 14, a frequency of 26.4%. The 3:1:3:4 modular structure was the second most common structure with n = 14, a frequency of 26.4%. As stated in Figure 11a, Hispanics show strikingly similar genetic characteristics to New York City Caucasians.

Figure 12: Combined distribution of MSY1 repeat structure in New York City Caucasian, African American, Asian, and Hispanic populations.

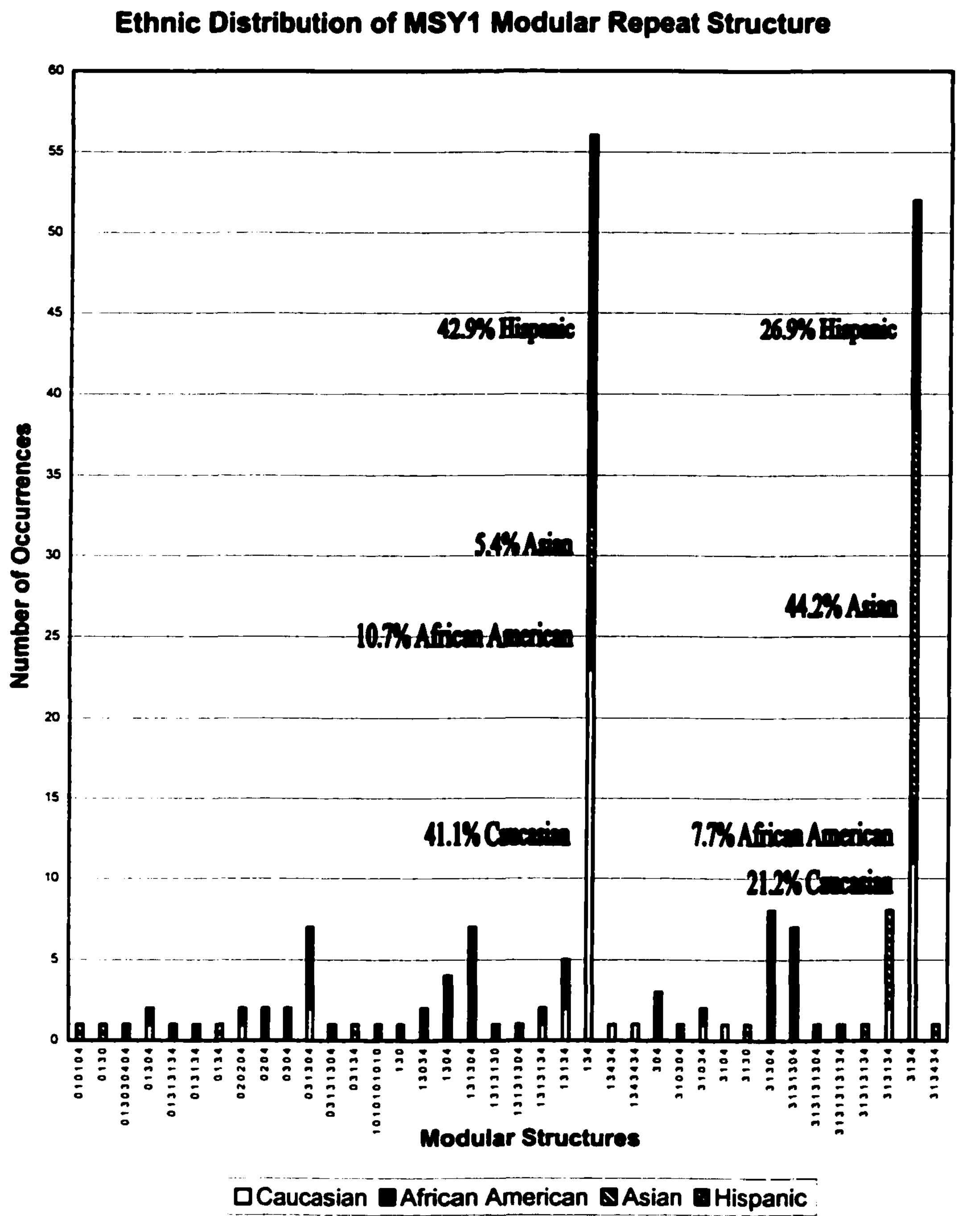


Figure 12: The 1:3:4 and 3:1:3:4 modular repeats remain the most common structures observed in New York City when data from all four populations are combined. The 1:3:4: structure is observed in 29.2% and the 3:1:3:4 structure was observed in 27.1% of the total number of modular repeat types.

Figure 13: Examination of the 3' repeat block in Caucasian, African American, Asian, and Hispanic populations.

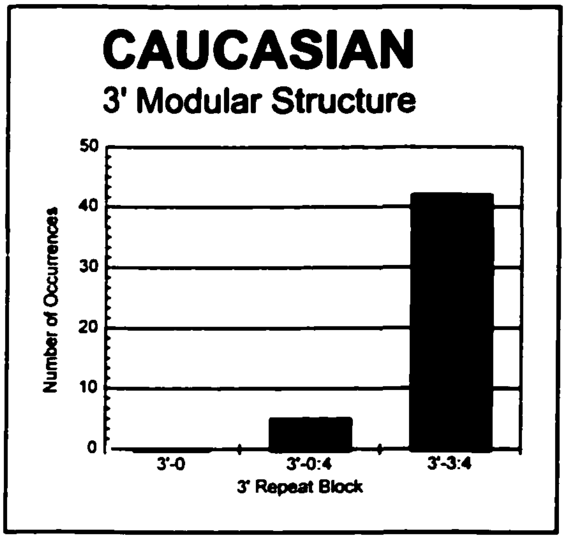


Figure 13a

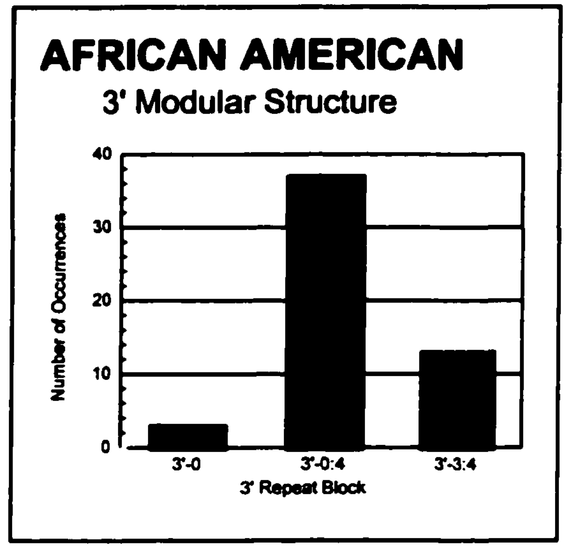


Figure 13b:

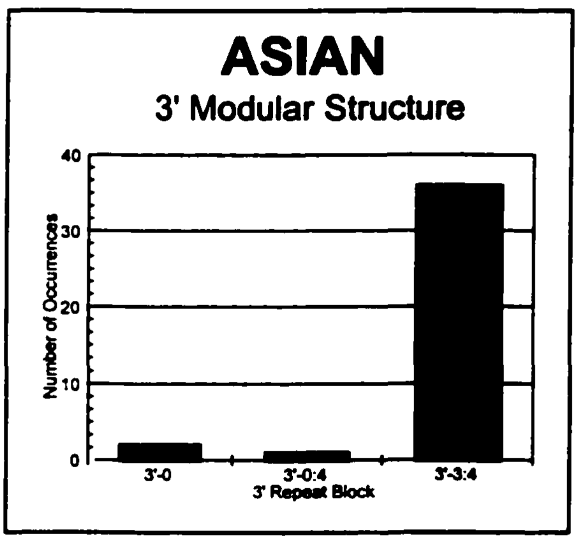


Figure 13c

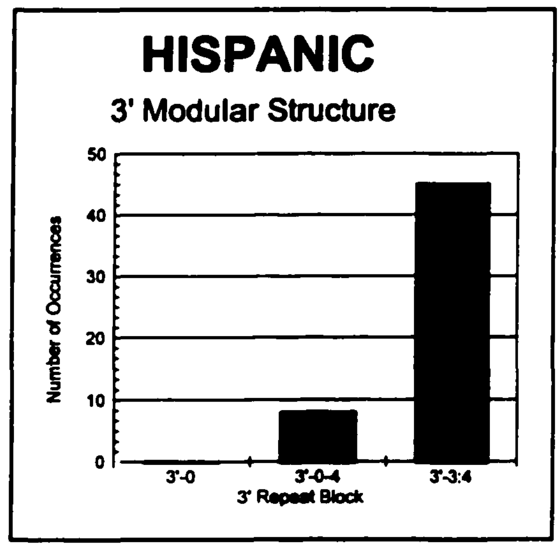


Figure 13d

Figure 13: Figures 14a-d represent the graphical examination of the 3' repeat block in all four populations. The 3' repeat block was broken into three main structures: 3'-0, 3'-0:4, and 3'-3:4. The 3'-0 represent repeat codes which end in null alleles and the 3'-0:4 represents repeat codes which contain a string of null alleles prior to the final type-4 repeat block. 36 (69.2%) African American samples contained the 3'-0:4 repeat block. This structure was only present in 5 Caucasians, 1 Asian, and 8 Hispanics, a total of 14 within the three populations combined. Of the 50 combined samples grouped within the 3'-0:4 repeat block, African American samples accounted for 72% of the total.

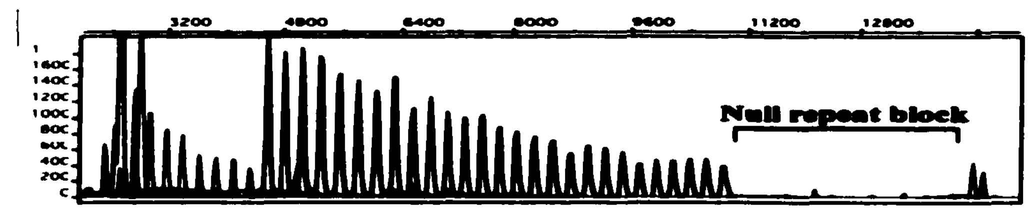


Figure 14a: Electropherogram view of AA16 forward MVR-PCR.

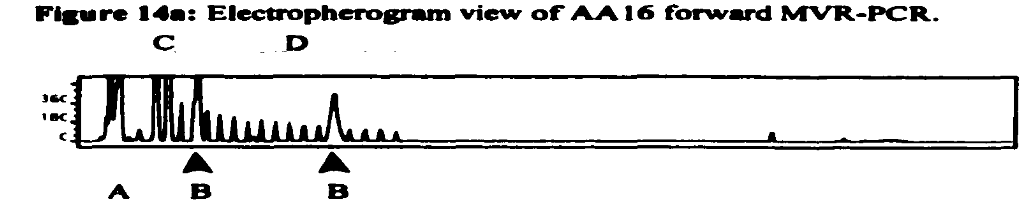


Figure 14b: Electropherogram view of AA16 Tag 4R reverse reaction. Label A represents dye primer peaks while label B represents two dye artifacts which are associated with the Rox labeled Tag-4R primers. Label C represents the two final 4 repeats of the digital code. The electropherogram was zoomed to 540 fluorescent units to display the decreased signal of the 4 subtype (label D) repeats following the last 3 repeat.

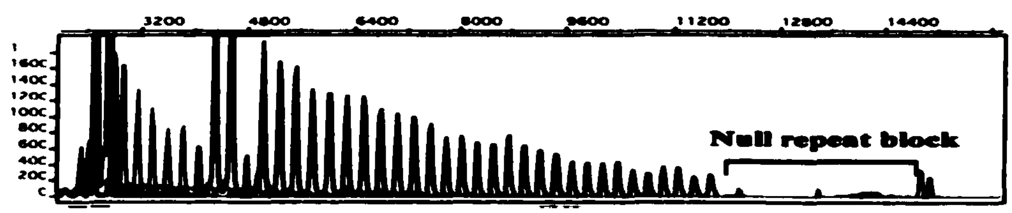


Figure 14c: Electropherogram view of AA17 forward reaction.

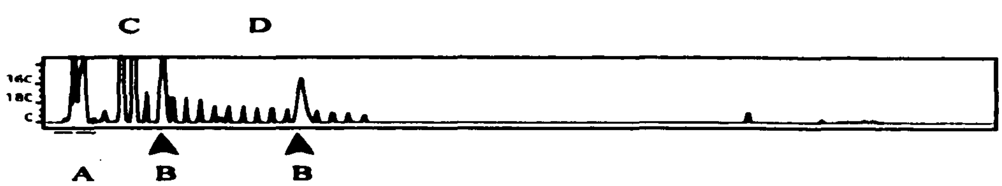


Figure 14d: Electropherogram view of AA17 Tag 4R reverse MVR-PCR. Labels A-D are expressed as in Figure 14b above.

Figure 14e: Gel image of AA16-17. Forward and reverse MSY1 MVR-PCR results represented.

Figure 14: Electropherogram representation of the 3'-0:4 null repeat block prevalent in the New York City African American population. Forward (Figure 14a and c) and reverse (Figure 14b and d) mapping reactions are displayed. The 3'-0:4 null repeat block is captured within the bracket in Figures 14a and c. The corresponding gel slice view (Figure 14e) has been included to display the absence of fluorescent fragments in the 3' block prior to the final two type-4 repeats.

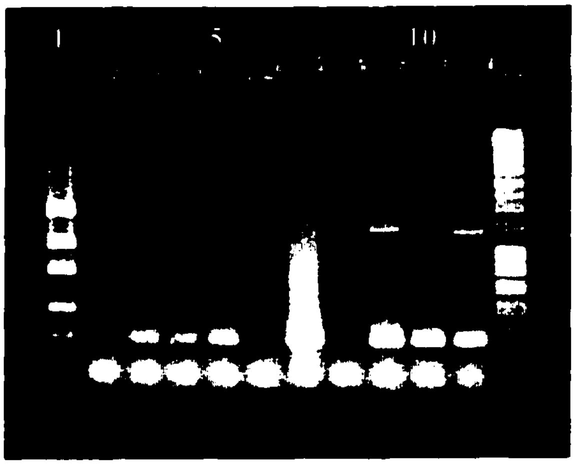


Figure 15a: Product gel containing non-probative casework samples N1-N8. Lane 3-7 contain forensic unknown samples N1-N5 respectively, including the reagent blank in lane 2. Lanes 9-11 contain exemplar samples N6-N8 respectively, including the reagent blank in lane 8. MSY1 amplified fragments were only present in two of the five mixed samples examined (N2-N3).

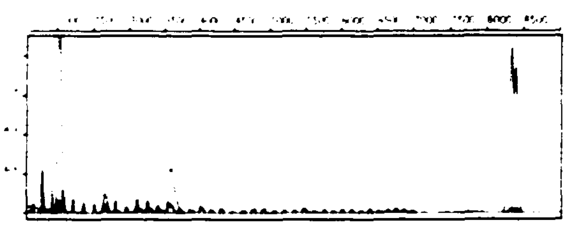


Figure 15b: Electropherogram representing lane 4 mixed sample N2 forward MVR-PCR mapping result.

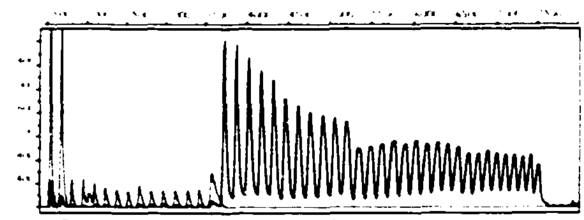


Figure 15c: Electropherogram representing lane 4 Mixed sample N2 reverse MVR-PCR mapping result.

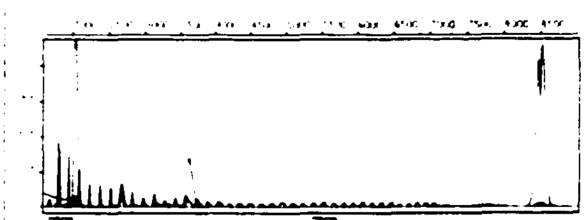


Figure 15d: Electropherogram representing lane 5 mixed sample N3 forward MVR-PCR mapping result.

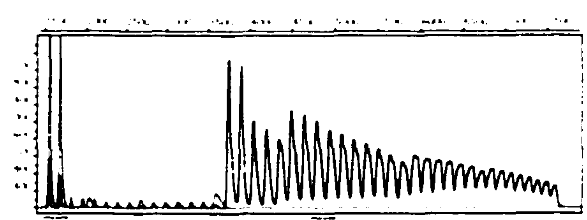


Figure 15e: Electropherogram representing lane 5 Mixed sample N3 reverse MVR-PCR mapping result.

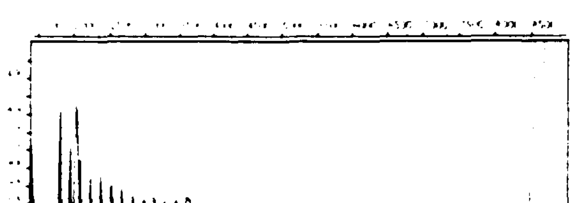


Figure 15f: Electropherogram representing lane 9 exemplar N6 forward MVR-PCR mapping result.

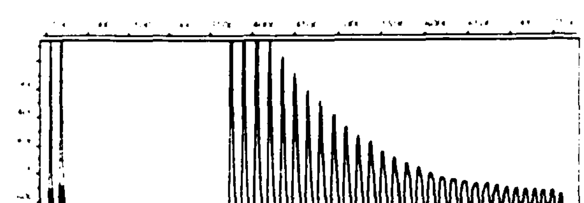


Figure 15g: Electropherogram representing lane 9 exemplar N6 reverse MVR-PCR mapping result.

Figure 15: Results of the non-probative MSY1 mixed sample case study. Of the known samples, only the included exemplar N6 is represented in the schematic. The forward reactions (Figures 15b, 15d, and 15f) all contain a distinct coding result. The type1 and type 3 repeats prior to the 3'-0:4 block code with decreased efficiency. This decrease in fluorescent intensity is consistent to additional base substitutions within the allele. The decrease in coding is not present in the reverse reactions (figures 15c, 15e, and 15g). This repeat code characteristic adds another level of discrimination to this particular inclusion.

Section VIII: Bibliography

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J., Staden, R., Young, I.G. (1981). Sequence and Organization of the Human Mitochondrial Genome. *Nature*, 290:457-464.

Armour, J.A.L., Jeffreys, A.J. (1992). Biology and Applications of Human Minisatellite Loci. *Current Opinion in Genetics and Development*, 2:850-856.

Armour, J.A.L., Anttinenn, T., May, C.A., Vega, E.E., Sajantila, A., Kidd, J.R., Kidd, K.K., Bertranpetit, J., Paabo, S., Jeffreys, A.J. (1996). Minisatellite Diversity Supports a Recent African Origin for Modern Humans. *Nature Genetics*, 13:154-160.

Cann, R.L., Stoneking, M., Wilson, M. (1987). Mitochondrial DNA and Human Evolution. *Nature*, 325:31-36.

Chakraborty, R., Stivers, D.N., Su, B., Zhong, Y., Budowle, B. (1999). The Utility of Short Tandem Repeat Loci Beyond Human Identifcation: Implications for Development of New DNA Typing Systems. *Electrophoresis*, 20:1682-1696.

Ciminelli, B.M., Pompei, F., Malaspina, P., Hammer, M., Persichetti, F., Palena, A., Anagnou, N., Guanti, G., Jodice, C., Terrenato, L., Novelletto, A. (1995). Recurrent Simple Tandem Repeat Mutations During Human Y-Chromosome Radiation in Caucasian Subpopulations. *Molecular Evolution*, 41:966-973.

de Knijff, P., Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M.A., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., Penacino, G., Perez-Lezaun, A., Piccinini, A., Prinz, M., Schmitt, C., Schneider, P.M., Szibor, R., Teifel-Greding, J., Weichhold, G., Roewer, L. (1997). Chromosome Y Microsatellites: Population Genetic and Evolutionary Aspects. *International Journal of Legal Medicine*, 110:134-140.

Dupuy, B.M., Andreassen, R., Flones A.G., Tomassen, K., Egeland, T., Brion, M., Carracedo, A., Olaisen, B. (2001). Y-Chromosome Variation in a Norwegian Population Sample. *Forensic Science International*, 177:163-173.

Elmozino, M., Prinz, M. Y-Miniml Haplotype Nanoplex Procedures (http://ystr.charite.de/text/locus_pcr2_3.html).

European YHRD (http://ystr.charite.de/index_gr.html).

Federal Bureau of Investigation (FBI), DNA Analysis Unit II. Mitochondrial DNA Analysis Protocol, (Current Protocol).

Foster, E.A., Jobling, M.A., Tayor, P.G., Donnelly, P., de Knijff, P., Mieremet, R., Zerjal, T., Tyler-Smith, C. (1998). Jefferson Fathered Slave's Last Child. *Nature*, 396:27-28.

Grunenwald, H. (2001) FailSafe™ PCR: A New System For Reliable and Consistent Amplification of Both Routine and Challenging Templates. *EPICENTRE Forum*, 7(2):9-11.

Gusmao, L., Gonzalez-Neira, A., Pestoni, C., Brion, M., Lareu, M.V., Carracedo, A. (1999). Robustness of the Y STRs DYS19, DYS389 I AND II, DYS390 and DYS393: Optimization of a PCR Pentaplex. *Forensic Science International*, 106:163-172.

Hammer, M.F. (1995). A Recent Common Ancestry for Human Y Chromosomes. *Nature*, 378:376-378.

Hammer, M.F., Zegura, S.L. (1996). The Role of the Y Chromosome in Human Evolutionary Studies. *Evolutionary Anthropology*, 5:116-134.

Hammer, M.F., Spurdle, A.B., Karafet, T., Bonner, M.R., Wood, E.T., Novelletto, A., Malaspina, P., Mitchell, R.J., Horai, S., Jenkins, T., Zegura, S.L. (1997). The Geographic Distribution of Human Y Chromosome Variation. *Genetics*, 145:787-805.

Hammer, M.F., Karafet, T., Rasanayagam, A., Wood, E.T., Altheide, T.K., Jenkins, T., Griffiths, R.C., Templeton, A.R., Zegura, S.L. (1998). Out of Africa and Back Again: Nested Cladistic Analysis of Human Y Chromosome Variation. *Molecular Biology of Evolution*, 15(4):427-441.

Harding, R.M., Fullerton, S.M., Griffiths, R.C., Bond, J., Cox, M.J., Schneider, J.A., Moulin, D.S., Clegg, J.B. (1997). Archaic African and Asian Lineages in the Genetic Ancestry of Modern Humans. *American Journal of Human Genetics*, 60:772-789.

Henke, J., Henke, L., Chatthopadhayay, P., Kayser, M., Dulmer, M., Cleef, S., Poche, H., Faelske-Zech, H. (2001). Applications of Y-Chromosomal STR Haplotypes to Forensic Genetics. *Croatian Medical Journal*, 42(3):292-297.

Hopkins, B., William, N.J., Webb, M.B.T., Debenham, P.G., Jeffreys, A.J. (1994). The Use of Minisatellite Variant Repeat-Polymerase Chain Reaction (MVR-PCR) to Determine the Source of Saliva on a Used Postage Stamp. *Journal of Forensic Science*, 39(2);526-531.

Hurles, M.E., Irven, C., Nicholson, J., Taylor, P.G., Santos, F.R., Loughlin, J., Jobling, M.A., Sykes, B.C. (1998). European Y-Chromosomal Lineages in Polynesians: A Contrast to the Population Structure Revealed by mtDNA. *American Journal of Human Genetics*, 63:1793-1806.

Jeffreys, A.J., Neumann, R., Wilson, V. (1990). A Repeat Unit Sequence Variation in Minisatellites: A Novel Source of DNA Polymorphism for Studying Variation and Mutation by Single Molecule Analysis. *Cell*, 60:473-485.

Jeffreys, A.J., MacLeod, A., Tamaki, K., Neil, D.L., Monckton, D.G. (1991). Minisatellite Repeat Coding as a Digital Approach to DNA Typing. *Nature*, 354:204-209.

Jobling, M.A., Tyler-Smith. (1995). C. Father and Sons: The Y Chromosome and Human Evolution. *Trends in Genetics*, 11(11):449-446.

Jobling, M.A., Pandya, A., Tyler-Smith, C. (1997). The Y Chromosome in Forensic Analysis and Paternity Testing. *International Journal of Legal Medicine*, 110:118-124.

Jobling, M.A. (1994). A Survey of Long-range DNA Polymorphisms on the Human Y Chromosome. *Human Molecular Genetics*, 3(1):107-114.

Jobling, M.A., Bouzerki, N., Taylor, P.G. (1998a). Hypervariable DNA Digital Codes for Human Paternal Lineages: MVR-PCR at the Y-specific Minisatellite, MSY1 (DYF155S1). *Human Molecular Genetics*, 7(4):643-653.

Jobling, M.A., Bouzerki, N., Taylor, P.G., Hammer, M.F. (1998b). Novel Mutation Process in the Evolution of a Haploid Minisatellite, MSY1: Array Homogenization without Homogenization. *Human Molecular Genetics*, 7(4):655-659.

Jobling, M.A., Heyer, E., Dieltjes, P., de Knijff, P. (1999). Y-chromosome-specific Microsatellite Mutation Rates Re-examined Using a Minisatellite, MSY1. *Human Molecular Genetics*, 8(11):2117-2120.

Jorde, L.B., Rogers, A.R., Bamshad, M., Watlins, W.S., Krakowiak, P., Sung, S., Kere, J., Harpending, H.C. (1997). Microsatellite Diversity and the Demographic History of Modern Humans. *Proceedings of the National Academy of Science*, 94:3100-3103.

Jorde, L.B., Watkins, W.S., Bamshad, M.J., Dixon, M.E., Ricker, C.E., Seielstad, M.T., Batzer, M.A. (2000). The Distribution of Human Genetic Diversity: A Comparison of Mitochondrial, Autosomal, and Y-Chromosome Data. *American Journal of Human Genetics*, 66:979-988.

Kayser, M., Tyler-Smith, C., Jobling, M.A., Sajantila, A. (2002a). A Systematic Search for New Polymorphic Microsatellite Loci on the Human Y Chromosome. III International Forensic Y-User Workshop, November 7-9, 2002, Porto, Portugal.

Kayser, M., Brauer, S., Willuweit, S., Schüdlich, H., Batzer, M.A., Zawacki, J., Prinz, M., Roewer, L., Stoneking, M. (2002b) Online Y-chromosomal Short Tandem Repeat Haplotype Database (YHRD) for U.S. Populations. *Journal of Forensic Science*, 47(3):513-519.

Malaspina, P., Persichetti, F., Novelletto, A., Iodice, C., Terrenato, L., Wolfe, J., Ferraro, M., Prantera, G. (1990). The Human Y Chromosome Shows a Low Level of Polymorphism. *Annals of Human Genetics*, 54:297-305.

National Research Council (NRC). (1996). The Evaluation of Forensic DNA Evidence. National Academy Press, Washington, D.C.

Nei, M., Roychoudhury, A.K. (1974). Sampling Variences of Heterozygosity and Genetic Distance. *Genetics*, 76:379-390.

Nei, M. (1978). Estimation of Average Heterozygosity and Genetic Distance from a Small Number of Individuals. *Genetics*, 89:583-590.

New York City Office of the Chief Medical Examiner, Department of Forensic Biology (OCMEa). Protocols for Forensic Mitochondrial DNA Sequencing, v1.0, (Draft Protocol).

New York City Office of the Chief Medical Examiner, Department of Forensic Biology (OCMEb). Protocols for Forensic STR Analysis, v7.1, (Current Protocol).

New York City Alliance Against Sexual Assault (http://www.nycagainstrape.org/index.html).

Perkin Elmer Corporation (1996). QuantiBlot Human DNA Detection Kit Package Insert.

Reynolds, R., Sensabaugh, G., Blake., E. (1991). Analysis of Genetic Markers in Forensic Samples Using the Polymerase Chain Reaction. *Annals of Chemistry*, 63:(1):2-15.

Roewer, L., de Knijff, P., Kayser, M. (http://www.promega.com/geneticidproc/esusymp3proc/03.pdf).

Roewer, L., Arnemann, J., Spurr, N.K., Grzeschik, K.H., Epplen, J.T. (1992). Simple Repeat Sequences on the Human Y Chromosome Are Equally Polymorphic as their Autosomal Counterparts. *Human Genetics*, 89:389-394.

Roewer, L., Kayser, M., Dieltjes, P., Nagy, M., Bakker, E., Krawczak, M., de Knijff, P. (1996). Analysis of Molecular Variance (AMOVA) of Y-chromosome-specific Microsatellites in two closely related human populations. *Human Molecular Genetics*, 5(7):1029-1033.

Schneider, S., Roessli, D., and Excoffier (2000). Arlequin ver. 2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland. (http://lgb.unige.ch/arlequin/).

Seielstad, M., Bakele, E., Ibrahim, M.., Toure, M. (1999). A View of Modern Human Origins from Y Chromosome Microsatellite Variation. *Genome Research*, 9:558-567.

Stoneking, M. (1993). DNA and Recent Human Evolution. *Evolutionary Anthropology*, 2:60-73.

Tamaki, K., Monckton, D.G., MacLeod, A., Neil, D.L., Allen, M., Jeffreys, A.J. (1992). Minisatellite Variant Repeat (MVR) Mapping: Analysis of "Null" Repeat Units at D1S8. *Human Molecular Genetics*, 1(6):401-406.

Tamaki, K., Monckton, D.G., MacLeod, A., Allen, M., Jeffreys, A.J. (1993). Four State MVR-PCR: Increased Discrimination of Digital DNA Typing by Simultaneous Analysis of Two Polymorphic Sites Within Minisatellite Variant Repeats at D1S8. *Human Molecular Genetics*, 2(10):1629-1632.

Technical Information on Y-Chromosome Testing. (http://www.freepages.geneology.rootsweb.com/~allpoms/genetics1b.html).

Thangaraj, K., Ramana, G.V., Singh, H,. (1999). Y-Chromosome and Mitochondrial DNA Polymorphisms in Indian Populations. *Electrophoresis*, 20:1743-1747.

Vigilant, L., Stoneking, M. Harpending, H., Hawkes, K., Wilson, A.C. (1991). African Populations and the Evolution of Human Mitochondrial DNA. *Science*, 253:1503-1507.

Walsh, P.S., Metzger, D.A., Higuchi, R. (1991). Chelex 100 as a Medium for Simple Extraction of DNA for PCR-based Typing from Forensic Material. *Biotechniques*, 10(4):506-513.

Werner, P. Analysis of Molecular Variance (AMOVA). (http://online.sfsu.edu/~efc/ classes/biol710/amova/amova.htm).

Zawacki, J.E. (2001). Analysis of Four New York City Populations Using a Y Chromosomal STR Multiplex Supplemented by a Characterization of Novel and Intermediate Alleles. Masters of Science in Forensic Science, John Jay College of Criminal Justice of the City University of New York.