National Institute of Justice
FY 10 Forensic DNA Research and Development Grant Program
Award Number 2011-DN-BX-K401

Final Technical Report

Characterization of X chromosomal short tandem repeat markers for forensic use

Principal Investigator
Toni M. Diegoli, Research Scientist
Armed Forces DNA Identification Laboratory/American Registry of Pathology
Contact information: toni.m.diegoli.ctr@mail.mil, 302-346-8519

Co-Principal Investigator
Michael D. Coble, Forensic Biologist
National Institute of Standards and Technology
Contact information: michael.coble@nist.gov, 301-975-4330

Submission Date: November 30, 2013

Recipient Organization
American Registry of Pathology
120 Old Camden Rd.
Suite A
Camden, DE 19934-5523

Project Period: April 1, 2011-November 30, 2013

Abstract

The use of X chromosomal short tandem repeat (STR) markers has been greatly increasing in the forensic setting over the last decade. The marker system offers the potential to provide information in addition to that obtained from autosomal STR systems currently used at crime laboratories and the courtroom, and in certain scenarios, markers on the X chromosome may be the only means of obtaining this information. Any investigated relationship situation where at least one female is involved may benefit from the use of X chromosomal STRs, which can be applied to cases of missing persons, criminal incest, immigration, deficiency paternity or other questioned relationships. In-depth characterization of the marker system is the first step in maximizing the power of this additional tool in the forensic arsenal.

In previous work performed at the Armed Forces DNA Identification Laboratory (AFDIL), two mini-X chromosomal STR multiplexes capable of amplifying 15 total markers (DXS6789, DXS7130, GATA31E08, GATA165B12, GATA172D05, DXS10147, DXS8378, DXS7132, DXS6803, HPRTB, DXS7423, and DXS9902, DXS7424, DXS101, and DXS6795) were developed. These assays employ techniques and instrumentation that is already in use in most laboratories for autosomal STR analysis. Developmental validation of the system was completed and allele frequencies have been recorded in a number of population groups. Therefore, a tool with which to further characterize this marker system already exists.

The 1991 report of the International Society for Forensic Genetics (ISFG) relating to the use of DNA polymorphisms in paternity testing has been used as a guideline for the evaluation of other marker systems such as autosomal STRs. Several requirements of this report remain unresolved for X STRs. First, mutation rates must be known in order to adequately handle possible mismatches attributable to mutational events. Second, questions of independent assortment and linkage disequilibrium must be addressed. Both criteria have been thoroughly investigated here through the study of relevant family groups and populations. Additionally, population data generated as a result of these studies is also reported, contributing to the continuing characterization and potential utility of these markers.

Through the study of 20,625 meioses in confirmed family trios or duos at the 15 X STR markers, eighteen mutations were observed across 7 of the 15 markers and in all three U.S. population groups (African American, U.S. Caucasian, U.S. Hispanic), resulting in an overall mutation rate of 8.73×10^{-4}. This mutation rate is similar to that reported for Y chromosomal and autosomal STRs as well as other published studies of X STRs. More than 50 three-generation pedigrees were examined for recombination events, resulting in rates that contradict the hypotheses of complete linkage within linkage groups and of free recombination between linkage groups.

Lastly, mixture interpretation is an important part of the forensic scientist's role in evaluating evidence from a crime scene, where mixed stains are common. The potential use of gonosomal STR markers to aid in the interpretation of such mixtures was investigated, and a multiplex was developed and characterized for this purpose. This assay correctly identified the sex and minimum number of contributors in all cases of artificial and theoretical mixtures tested as part of this study, and correctly assigned the actual number and sex of contributors 62% of the time. While it is clear from this initial development and characterization study that a gonosomal marker multiplex cannot solve all of the questions surrounding the interpretation of a mixed profile, there are benefits to its use in certain situations that justify continued study.

In conclusion, mutation rate studies, linkage analysis, and mixture evaluation were performed in order to further characterize the X chromosomal STR marker system for routine forensic use.

Table of Contents

Abbreviations

µL	Microliters
A	Adenine
AA	African American
AFDIL	Armed Forces DNA Identification Laboratory
BLAST	Basic local alignment search tool
BLAT	BLAST-like alignment tool
C	Cytosine
CEPH	Centre d'Etudes du Polymorphisme Humaine
cm	Centimeter
cM	Centimorgans
CN	U.S. Caucasian
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
EDTA	Ethylenediaminetetraacetic acid
FamID	Family identification number
FID	Father's identification number
F_{ST}	Fixation index
G	Guanine
H(exp)	Expected heterozygosity
H(obs)	Observed heterozygosity
Hisp	U.S. Hispanic
ISFG	International Society for Forensic Genetics
kV	Kilovolts
LOD	Logarithm of the odds
Mb	Megabases
MECI	Mean exclusion chance in trios involving daughters
MECII	Mean exclusion chance in father/daughter duos
mg	Milligram
MID	Mother's identification number
mL	Milliliter
mM	Millimolar
mtDNA	Mitochondrial DNA
NIGMS	National Institute of General Medical Sciences
^{o}C	Degrees Celcius
p (HWE)	P values of the exact test for Hardy-Weinberg equilibrium
PCR	Polymerase chain reaction
PDf	Power of discrimination in females
PDm	Power of discrimination in males
pg	Picogram
PIC	Polymorphism information content
PID	Patient identification number
Ref.	Reference
RF or Θ	Recombination fraction

RFU	Relative fluorescence unit
SRY	Sex-determining region of the Y chromosome
STR	Short tandem repeat
T	Thymine
TLE	Tris-low-EDTA buffer
T_m	Melting temperature
UCSC	University of California Santa Cruz
Z	LOD score

Executive Summary

Problem and purpose: The use of X chromosomal short tandem repeat (STR) markers has been greatly increasing in the forensic setting over the past decade. The marker system offers the potential to provide information in addition to that obtained from autosomal STR systems currently used at crime laboratories and the courtroom, and in certain scenarios, markers on the X chromosome may be the only means of obtaining this information. Any investigated relationship situation where at least one female is involved may benefit from the use of X chromosomal STRs, which can be applied to cases of missing persons, criminal incest, immigration, deficiency paternity or other questioned relationships. In-depth characterization of the marker system is the first step in maximizing the power of this additional tool in the forensic arsenal.

According to the 1991 report of the International Society for Forensic Genetics (ISFG; formerly ISFH (Haemogenetics)) relating to the use of DNA polymorphisms in paternity testing, mutation rates must be known in order to adequately address possible mismatches attributable to mutational events [1]. In order to enhance the practical application and interpretation of X STR markers, the rate of mutation should be determined through examination of a substantial number of meioses, and the dependence of mutation rates upon the origin, length, and structure of the allele should be investigated. Thus far, few studies have been published regarding the mutation rates of the commonly used markers, and data that have been published are typically based upon limited population groups and only a subset of the markers used by the community. Currently, there are no studies in the literature that investigate mutation rates among U.S. African American or U.S. Hispanic groups, for example.

Standard 8.1.3.2 of the U.S. DNA Advisory Board's Quality Assurance Standards for Forensic DNA Testing Laboratories states that laboratories shall establish and document match criteria on the basis of empirical data [2]. Another requirement of the 1991 ISFG report relating to the use of DNA polymorphisms is that questions of independent assortment and linkage disequilibrium be addressed [1]. For autosomal STRs, this ensures that the product rule can be used to multiply individual marker frequencies together to determine the overall rarity of a profile. It does not

preclude the use of linked markers, however. Y chromosomal STRs, for example, are linked to one another and are considered together as a group called a haplotype. Haplotype frequencies are measured directly from population data, and the counting method is used to determine the rarity of the profile. It follows that X chromosomal STRs may be a combination of the two techniques: the organization of several physically close markers into linkage groups, forming haplotypes with frequencies that could then be multiplied together once independent assortment of the groups was established.

Undoubtedly the use of multiple markers located on a single chromosome necessitates special consideration of the potential for linkage between closely situated markers. Potential linkage must be addressed in order to allow for the meaningful calculation of forensically relevant statistics. This study sought to answer the following basic questions towards a potential solution:

1. Is complete linkage observed between each of the markers within the proposed linkage groups?
2. Is the recombination rate equal to 0.5 (free recombination) between the proposed linkage groups?

Lastly, mixture interpretation is an important part of the forensic scientist's role in evaluating evidence from a crime scene, where mixed stains are common. However, routine analysis of mixtures continues to be a challenge, and varying solutions have been proposed and adopted for use [3-5]. One approach that has thus far only been partially explored is the use of markers on the sex chromosomes to aid in the mixture analysis process. Several studies reported the ability of Y STR markers to detect the male component in male-female mixtures from sexual assault cases [6,7]. Similarly, it has been suggested that X chromosomal markers may help to expose a female profile in male background, such as vaginal cells on a penis or female cells from male fingernail scrapings [8]. Due to their unique inheritance patterns, gonosomal markers hold the potential to supplement traditional mixture testing in certain specific situations.

Research design: Though commercial kits are available that probe a wide variety of genetic markers on both the Y chromosome and the autosomes, there is currently only one commercial

kit currently manufactured assaying markers on the X chromosome, the Investigator™ Argus X-12 kit (QIAGEN, Hilden, Germany). Due to patent and intellectual property issues between the United States and European STR kit manufacturers, the kit cannot be sold or marketed in the U.S. at this time. Most X chromosomal STR typing currently relies upon noncommercial multiplex assays that have been published, simultaneously amplifying 2 to 12 loci in a single reaction (see [9-21], for example). In previous work performed at the Armed Forces DNA Identification Laboratory (AFDIL), two mini-X chromosomal STR multiplexes capable of amplifying 15 total markers were developed used to generate several population databases [22,23]. These assays use techniques and instrumentation that is already present in most laboratories for autosomal STR analysis. Therefore, a tool with which to further characterize this marker system already exists.

For the mutation rate study, anonymous extracts from 958 families (parent-child duos and trios) were representative of the three major U.S. populations: African American, U.S. Caucasian, and U.S. Hispanic. Unrelated individuals were used to generate a U.S. population database described separately. In total, U.S. allele and haplotype frequencies were provided for 314 African American (108 males, 206 females), 434 U.S. Caucasian (165 males, 269 females), and 398 U.S. Hispanic (150 males, 248 females) individuals.

In order to confirm the findings of this study by comparison as well as compile a robust large-scale dataset for use by the forensic community, a literature review of published mutation rate studies available at the time of manuscript preparation was conducted. Data mining necessary to compile the included list of published mutation rates by marker was accomplished by recalculating the mutation rate within 33 independent studies [9,10,13,14,19,24-51] which were then combined for each marker to determine both the marker-specific and overall mutation rates, and confidence intervals were assigned to these pooled values.

Linkage analysis required the purchase of a set of commercially-available extracts representing multigenerational pedigrees from the National Institute of General Medical Sciences (NIGMS) repository. Donated by the Centre d'Etudes du Polymorphisme Humaine (CEPH), this collection

of extracts from lymphoblastic cell lines included families from Utah, France, Venezuela, and Amish country. Six additional U.S. Caucasian families previously typed at a subset of autosomal STR markers [52] and additional families suitable for linkage analysis from the mutation rate study were included as well. In total, 158 families were identified within this dataset for potential analyses.

Once appropriate sample sets were identified and typed, analyses were performed using two different methods. Using the classical method, each observed recombination event was noted and totaled for each of 14 marker pairs. This total was then divided by the total number of unambiguous meioses for that marker pair, producing the observed recombination rate, and corresponding logarithm of the odds (LOD) scores (Z) were calculated. The computer-based analysis method was performed through a collaborative relationship as described by Nothnagel, *et al.* [53] with modifications on the front-end specific to this dataset. This maximum likelihood approach takes mutation rate into consideration during the estimation of recombination rate.

In seeking to create and evaluate a mixture multiplex, a combination of markers located on both sex chromosomes was desired, including X and Y STRs as well as STRs within the X-Y homologous region, termed XY markers. A review of the literature resulted in a list of potential markers; selection was limited to markers for which there were published population genetic studies in order to exploit the collective knowledge of these established markers and their relevant characteristics, simplifying the process. Potential utility within the mixture multiplex was assessed according to the following criteria: (a) potential for small amplicon size; (b) large allele range with high degree of polymorphism; and (c) established use within the forensic community. Markers best matching these criteria were selected for inclusion into the multiplex, termed MIXplex, and organized according to amplicon size.

Extracts utilized in this study were commercially available, highly concentrated control DNAs for which the profiles at each of the markers chosen for inclusion in the multiplex were known since a large quantity of extract would be needed. PCR amplification, electrophoresis, and detection occurred in a manner similar to the protocol used for the X STR multiplexes, and

sensitivity testing was performed to evaluate the lower limits of the multiplex. Using these same six controls, 63 total mixtures of various sex and number of contributors were created and a subset was amplified using both the MIXplex and a commercial autosomal STR kit. Blind analyses focused on determination of the minimum number and sex of contributors.

Findings & conclusions: Through the study of 20,625 meioses in confirmed family trios or duos at 15 X STR markers (DXS6789, DXS7130, GATA31E08, GATA165B12, GATA172D05, DXS10147, DXS8378, DXS7132, DXS6803, HPRTB, DXS7423, and DXS9902, DXS7424, DXS101, and DXS6795), eighteen mutations were observed across 7 of the 15 markers and in all three U.S. population groups (African American, U.S. Caucasian, U.S. Hispanic), resulting in an overall mutation rate of 8.73×10^{-4}. Compared with the overall rate in this study, 14 published X STR studies reported higher mutation rates, but only two approached significance based upon confidence interval bounds: 4.76×10^{-3} [48] and 2.09×10^{-3} [30], both observed in German populations. The overall mutation rate for the combined study (published data plus this study) was the result of 102 mutations observed in 71,020 meioses at 36 markers. The combined overall rate of 1.44×10^{-3}, like the overall rate from the literature summary of 1.67×10^{-3}, was higher than that observed for the total U.S. dataset in this study (8.73×10^{-4}).

The overall rate for the African American population (1.71×10^{-4}) was the lowest overall rate, and was significantly smaller than both the overall rate from the literature summary (1.67×10^{-3}) as well as the combined overall rate (this study plus literature summary; 1.44×10^{-3}). The total number of observed meioses for the African American population was at least 15% less than that of either the U.S. Caucasian and U.S. Hispanic populations, potentially indicating that further study would be necessary to determine if this difference in mutation rate was authentic. In general, the observed marker-specific mutation rates within this study were similar across populations and markers as well as consistent with the overall rate (Table 6).

Two markers exhibited a mutation rate of zero in both this study and published studies: GATA165B12 and DXS10147. Both markers have a relatively small allele range; in U.S. populations, seven alleles were observed at GATA165B12 and eight at DXS10147 (see "U.S.

Population Database Creation" within "Results" section). Markers for which no mutation rate studies have been published (DXS7130 and DXS6795) exhibited no mutations in this study. Mutations were observed for the first time at DXS9902 and GATA172D05 in this study, likely owing to the larger number of total meioses examined here.

Within the combined dataset representing the pooled rates from this study and the literature summary, there were differences between the mutation rates of different markers (Tables 6 and 7). The difference between the mutation rate for DXS7132 (2.57×10^{-3}), which exhibited the largest marker-specific rate, and DXS101 (2.56×10^{-4}), GATA165B12 (zero mutations observed), and DXS7423 (5.01×10^{-4}) approached significance based on the confidence interval bounds. Additionally, the combined overall mutation rate (1.44×10^{-3}), though generally higher than many marker-specific rates in the combined dataset, yielded a confidence interval completely contained within those of the marker-specific rates for all but DXS101 and GATA165B12. Given the relatively large (>1300) number of meioses for all markers and the small confidence interval ranges, it is likely this difference is genuine for at least these two markers, and their true mutation rate is indeed lower than for other markers or for X STRs in general. Further study, however, could help confirm this hypothesis.

Examination of the progenal and parental genotypes of the 18 mutations observed in this study revealed that all mutations could be explained by a change of one repeat unit, which is consistent with the model of strand slippage during replication as the mechanism of microsatellite mutation [54]. In particular, repeat unit gains outnumbered losses in this study by approximately 2:1. This bias towards microsatellite expansion has been noted in other mutation rate studies [55-58], though both an excess of losses [59,60] as well as equal rates [41,61-64] have been noted by others.

Two U.S. Caucasian families exhibited two mutations each; one family displayed paternal mutations resulting in a loss of a repeat unit at both DXS9902 & DXS7132 while the other family showed a maternal mutation resulting in a gain of a repeat unit at DXS7132 and a paternal mutation resulting in a loss of a repeat unit at DXS8378. Individually, these three markers had

the highest mutation rates in this study and within the combined dataset. Though multiple mutational events within one family are rare, they should not be entirely unexpected. Based upon the upper confidence interval bounds of the two markers with the highest mutation rates (DXS9902 and DXS7132), two simultaneous mutations could be expected to occur approximately once in every 41,500 meioses.

While the number of trinucleotide markers investigated here was too small to make an accurate inference as to the impact of repeat size on mutation rate, previous studies of Y STR mutation rates have indicated a bias towards a higher mutation rate for longer repeat units [57,65]. Additionally though it has been previously noted for autosomal and Y STRs that mutations observed at microvariant and/or compound repeats appeared more common than at simple repeats [55,59,65], this study did not yield the same results. Separating the 15 markers into two groups based upon repeat structure ("simple" and "compound/complex" including microvariant), two sets of approximately equal numbers of markers and meioses were formed. Including data from published studies, almost identical rates for both types of repeats were revealed: 1.08×10^{-3} (CI: 6.8×10^{-4}-1.6×10^{-3}) for simple repeats and 1.01×10^{-3} (CI: 6.6×10^{-4}-1.5×10^{-3}) for compound/complex (including microvariant) repeats.

Despite examination of more than 2.2 times as many maternal as paternal meioses, the maternal mutation rate remained comparable to the overall observed rate and almost an order of magnitude smaller than the paternal rate. This trend was consistent across populations and overall, with the paternal rate reaching almost three times the maternal rate in the U.S. Caucasian and U.S. Hispanic populations. The overall mutation rate for maternal transfers was 4.22×10^{-4} while the mutation rate for paternal transfers was 1.71×10^{-3}. Though the confidence intervals for these two values overlap by a small margin, both rates fell outside of the 95% CI for the overall mutation rate for all meioses. In agreement with the results in this study, previous studies also found that paternal mutations are more frequent for both autosomal STRs [58-60,63,66,67] as well as X STRs [47,48]. Additionally, the higher paternal X STR mutation rate corroborates the idea that the mechanism of microsatellite mutation may be independent of recombination

[55,68] (which is absent within a paternally-inherited X chromosome) and explains the similarity in overall mutation rate for both the gonosomes and autosomes.

In order to continue to work towards routine use of X STRs for relationship testing in the forensic setting, datasets that combine results from multiple studies serve to maximize the information that can be concluded from individual profiles. As the largest X STR mutation rate study to date, and the only one to investigate U.S. populations, the total number of meioses available to the community for consideration has increased by over 40%. Combining the 20625 meioses from this study with those from consolidated published studies yielded a robust dataset of 71020 meioses for use by the forensic community.

Allele frequencies and forensic efficiency parameters calculated for each of the 15 markers in three population groups are shown in Table 11. In total, 160 alleles were observed across 15 markers, with 7-17 alleles at each marker. Marker DXS101 was the marker with the highest number of observed alleles (17) in the populations studied; DXS8378, GATA165B12 and DXS10147 were the markers with the lowest number (7). DXS101 also exhibited the highest observed heterozygosity values within all three populations (0.9466, 0.8513, and 0.8629 in African Americans, U.S. Caucasians, and U.S. Hispanics respectively) while the lowest heterozygosity values varied by population (0.6845 at DXS8378 in African Americans; 0.6691 at DXS6795 in U.S. Caucasians; 0.6532 at DXS9902 in U.S. Hispanics). The usefulness of certain markers was strongly dependent upon the population to which they were applied. For example, DXS6795 exhibited the second highest observed heterozygosity value (0.8105) in the U.S. Hispanic population, the lowest value (0.6691) in the U.S. Caucasian population, and a value (0.8107) at the midpoint of the 15 markers in the African American population. Overall, the forensic efficiency parameter values confirm the potential usefulness of these markers in certain specific kinship situations involving female offspring as well as identity testing.

Previous pairwise population comparisons with X STRs and U.S. populations revealed that the individual groups cannot be combined into one pooled database for forensic use and must instead be treated as three distinct databases [18,22], and this structure was maintained in this study.

The original publication describing the development of the multiplexes used here to amplify the 15 X STR markers included a population study of 349 African Americans, 268 U.S. Caucasians, and 245 U.S. Hispanics [22]. Of note, no significant differences between the three populations present in both studies were observed. Therefore, these databases could be combined to create a larger single database for each group. When combined, the U.S. populations exhibit similar forensic efficiency statistics with none of the marker-population combinations resulting in a significant deviation from Hardy-Weinberg equilibrium after the Bonferroni correction ($p < 0.0033$).

These data serve to greatly increase the amount of X STR information available for U.S. populations, for which only four previous studies exist [18,22,69,70], while simultaneously confirming the potential utility of the chosen markers for use in both kinship and identity testing. These databases provide the basis by which forensic scientists will calculate the statistical value of a match between two DNA profiles. Without this information, the use of X STRs is extremely limited, and the full potential of the marker system in forensic scenarios cannot be realized.

To begin the study of linkage, marker locations were determined based upon In Silico PCR BLAT searches [71] and organized along the chromosome. Of the 15 markers studied here, the four original linkage groups described by Szibor, *et al.* [30] contained the following markers: DXS8378 and DXS9902 in linkage group 1; DXS7132, DXS6789, DXS101, DXS7424, and GATA172D05 in linkage group 2; HPRTB in linkage group 3; and DXS7423 in linkage group 4. Additional markers (DXS6795, DXS6803, DXS7130, GATA165B12, GATA31E08, and DXS10147) included within each linkage group were hypothesized based upon location, the Forensic ChrX Research website [72], and linkage disequilibrium analysis.

Fifty families were analyzed using the classical method. Homozygous genotypes and mutations that rendered a marker uninformative for recombination were excluded, and recombination was defined as a change in source chromosome between two adjacent markers. The observed recombination rate varied from zero (marker pairs DXS7424-DXS101 and DXS10147-DXS743) to 0.21 (marker pair GATA31E08-DXS10147) within linkage groups and from 0.10 (border of

linkage groups 3 and 4) to 0.40 (border between linkage groups 1 and 2) between linkage groups. The highest recombination rate was observed for the pair of markers defining the boundary between linkage groups 1 and 2 (DXS6795 & DXS7132). All marker pairs, including those within linkage groups, exhibited a non-zero recombination rate except two: DXS7424-DXS101 in linkage group 2 and DXS10147-DXS7423 in linkage group 4. LOD scores, however, indicated linkage between all pairs except those at the boundary of linkage groups 1 and 2.

A slightly larger number (58) of families were analyzed using the computer-based method and three different starting values: distance-interpolated recombination rates, all recombination rates equal to 0.25, and four groups of unlinked markers. Despite these different starting points, the optimization converged at the same location in each case, indicating a robust optimization result. The recombination rates obtained with the computer-based method generally agreed with the values calculated manually, further indicating a robust computation. These values ranged from zero (marker pairs DXS7424-DXS101 and DXS10147-DXS743) to 0.2045 (marker pair GATA31E08-DXS10147) within linkage groups. Between linkage groups, the lowest rate occurred at the border of linkage groups 2 and 3 (0.0971) rather than between linkage groups 3 and 4 as for the classical analyses. The border between linkage groups 1 and 2 revealed the highest overall recombination rate for both the classical (0.40) and the computer-based (0.4462) analyses. As noted for the classical analyses, all marker pairs, including those within linkage groups, exhibited a non-zero recombination rate using the computer-based method except two: DXS7424-DXS101 in linkage group 2 and DXS10147-DXS7423 in linkage group 4.

Because mutations were ignored in the classical analyses, it was thought that the true recombination rate had likely been overestimated. However, when comparing the classical analyses to computer-based estimates performed using a maximum likelihood approach taking mutation rates into account, the values were found to be very similar. In all cases, the classically-calculated observed recombination rate fell within the 95% support intervals of the computer-based values except for marker pair DXS6789-DXS7424 where the rate of 0.03 fell just outside the lower limit of the 95% support interval (0.0350). Taken together, the results of

this combination of methods indicate a robust estimate of the recombination rate between these 14 marker pairs has been achieved.

The genetic distance values calculated as part of the computer-based analyses varied only marginally by starting value. The relative rate of recombination was generally positively correlated with physical distance between markers in this study: as the distance between the markers increased, the recombination rate increased. There was one notable exception to this trend. Marker pair GATA31E08-DXS10147 exhibited a much higher mutation rate (0.21 and 0.2045) than marker pairs with similar genetic distances separating them, as reflected by the genetic distance estimates. Further study is necessary to determine whether this result may indicate a true recombination "hot spot" or may be influenced by factors such as linkage disequilibrium.

The hypotheses of complete linkage within linkage groups and of free recombination between linkage groups were both contradicted by the results of this study. The ultimate goal, however, is to gain a better understanding of how potential linkage between this set of 15 X STR markers should direct likelihood calculations in kinship testing. These preliminary results indicate a need to delve even deeper with more comprehensive analyses. Additional calculations, when and if they are able to be performed, with additional families typed as part of this study will likely further strengthen the conclusions.

The potential use of gonosomal STR markers to aid in the interpretation of forensic mixtures was investigated, and a multiplex was developed and characterized for this purpose. After consideration of candidate gonosomal markers, several suboptimal primer pairs were excluded and the final multiplex consisted of three X STRs (DXS6795, DXS6789, and GATA31E08), two Y STRs (DYS393 and DYS438), one XY STR (DXYS267), and a portion of the sex determining region of the Y chromosome, SRY. Primer mix concentrations were adjusted empirically to balance peak heights within each multiplex, and tested on a panel of known control samples to ensure consistent quality and correct genotypes could be obtained before use on unknowns. Sensitivity testing using single-source samples revealed that full profiles could be obtained with

as little as 200 pg of input DNA, and heterozygous peak height ratios generally remained above 60% for recovered alleles. With duplicate or triplicate amplification of six two-person mixtures (four female-male, one male-male, and one female-female) at varying mixture ratios, complete profiles were reliably obtained for mixtures where the minor component was 20% or greater in most cases. This value coincided well with the single-source sensitivity results. Like the sensitivity results, the mixture testing also revealed that near-complete or complete profiles could also be obtained for most mixtures with the minor component at only 10%, or 100 pg input.

The design and potential of the MIXplex combined several key elements of mixture interpretation. Generally, reporting of mixed profiles centers on estimating the minimum number of contributors as well as attempting to assign a sex to the individual contributors in some two-person mixtures. The MIXplex correctly identified the sex and minimum number of contributors in all cases of artificial and theoretical mixtures tested as part of this study, and correctly assigned the actual number and sex of contributors 62% of the time. Currently, with autosomal STRs, sex can only be reliably assigned when both contributors are of the same gender, or the male contributor is the minor component of a male-female mixture [73]. Additional testing, such as a Y STR assay, is usually necessary to correctly infer and confirm these characteristics of a mixture, and the MIXplex offers an additional alternative. Corroboration of the suspected number and/or sex of contributors through this assay could direct future analysis, potentially saving time and money. Pre-screening samples thought to contain multiple contributors with this relatively inexpensive assay to 1.) confirm a mixture is present, and 2.) decide which assay, if any, would be most appropriate could eliminate uninformative testing altogether. Additionally, the MIXplex can clarify situations where the male allele at the amelogenin locus is not amplified due to a deletion on the Y chromosome [74,75] without complete Y STR typing. Moreover, profile subtraction, which is the elimination of alleles from a known contributor (usually female) to the mixture, is simplified in an assay where only four loci of seven markers are found on a female's chromosomes, while all markers are present within a male's chromosomes. Even when a male and a female contributor share alleles, there are an additional four markers at which the male alleles would be the only ones present.

While it is clear from this initial development and characterization study that a gonosomal marker multiplex cannot solve all of the questions surrounding the interpretation of a mixed profile, there are benefits to their use in certain situations that justify continued study. Additional optimization of this or a similar multiplex in combination with further characterization of assay parameters such as the reporting threshold and stutter ratios would be helpful to increase confidence in allele calls. Analysis of additional mixtures, both theoretical and actual, could illustrate both the strengths and the limitations of the current combination of markers, as well as suggest additional configurations that might aid interpretation even further. Casework mixtures should eventually be evaluated with a final assay in order to assess its performance and value in real-world settings.

Two of the key elements of mixture analysis rest in the initial determination of the minimum number of contributors to the mixed profile as well as the sex of these contributors [73], which the MIXplex helps to address. However, the same authors recognize that a standardized mixture interpretation protocol that will be appropriate for every mixed profile an analyst encounters is not feasible. Where the concept of a mixture multiplex made up of gonosomal markers fits into the overall forensic mixture interpretation scheme remains to be uncovered.

In conclusion, mutation rate studies, linkage analysis, and mixture evaluation were performed in order to further characterize the X chromosomal STR marker system for routine forensic use.

Main Body of Final Technical Report

Portions of text were taken directly from published manuscripts, or those submitted for publication and currently undergoing peer review.

Introduction

Over the past decade, X chromosomal short tandem repeat markers (X STRs) have been recognized as useful tools to supplement traditional kinship testing in the forensic setting due to their unique inheritance pattern and, correspondingly, the breadth of published literature on the

subject has expanded greatly in recent years. STR markers on the X chromosome may be useful in several forensic contexts. For instance, missing persons cases, such as those undertaken at the Armed Forces DNA Identification Laboratory (AFDIL), usually require the analysis of relatives due to a lack of direct reference material. Often times, mitochondrial DNA (mtDNA) typing can be used to address the potential for degraded or low quantity samples such as skeletal remains, particularly in closed populations and when a direct maternal reference is available, due to its relatively high copy number and protected location within the mitochondria of the cell. However, mtDNA is maternally inherited; therefore where maternal references are unavailable or where the unidentified individual matches one of the most common mtDNA haplotypes, mtDNA testing alone may be inadequate. In such cases, markers on the X chromosome may provide additional information [76,77], offering the potential to both augment traditional STR testing and mtDNA sequencing for human remains identification as well as differentiate pedigrees that would be otherwise indistinguishable with unlinked autosomal STRs [78].

For both the Y chromosome and the autosomes, commercial kits are available that probe a wide variety of genetic markers (see [79-83], for example). For the X chromosome, however, there is currently only one commercial kit: the Investigator™ Argus X-12 kit (QIAGEN, Hilden, Germany). This kit simultaneously amplifies and detects twelve X STRs (DXS8378, HPRTB, DXS7423, DXS7132, DXS10134, DXS10074, DXS10101, DXS10135, DXS10103, DXS10079, DXS10146, and DXS10148) plus amelogenin in four fluorescent dye channels [84]. Due to patent and intellectual property issues between the United States and European STR kit manufacturers, the kit cannot be sold or marketed in the U.S. at this time. It is believed that going forward, as more laboratories demand X STRs, additional X STR kits will be manufactured for the forensic community. Most X chromosomal STR typing currently relies upon noncommercial multiplex assays that have been published, simultaneously amplifying 2 to 12 loci in a single reaction (see [9-21], for example). Two such multiplexes capable of amplifying a total of 15 mini-X STRs have been developed at AFDIL and used to generate several population databases [22,23].

When specifically targeting DNA that is known to be degraded, short amplicon sizes are favored with the goal of recovering the maximum number of alleles [85]. Genotyping using reduced-size amplicons for the 13 core autosomal loci used in the United States produced an increase in the ability to recover information from compromised samples while maintaining concordant profiles [86], and additional mini-STR loci were characterized to further increase the information that could be obtained from degraded samples [87]. In an X STR study examining degraded samples, Asamura *et al.* demonstrated the success of two quadruplex reactions consisting of amplicons ranging from 76-169 base pairs (bp) in length [15], and reduced size amplicons were targeted for inclusion into the AFDIL multiplexes as well [22].

Mutation rates

According to the 1991 report of the International Society for Forensic Genetics (ISFG; formerly ISFH (Haemogenetics)) relating to the use of DNA polymorphisms in paternity testing, mutation rates must be known in order to adequately address possible mismatches attributable to mutational events [1]. In order to enhance the practical application and interpretation of X STR markers, the rate of mutation should be determined through examination of a substantial number of meioses, and the dependence of mutation rates upon the origin, length, and structure of the allele should be investigated.

Typical mutation rates for autosomal and Y chromosomal markers are in the range of ~1-5 x 10^{-3} [88]. Existing reports of X STR mutation rates generally fall within this range, though many are based upon a relatively small number of meiotic events (see Table 7 in the "Results" section for a summary of published X STR mutation rates). There is a need to investigate X STR mutation rates in populations outside of Europe in general and Germany specifically. Currently, there are no studies in the literature that investigate mutation rates among U.S. population groups, or from the continent of Australia, for example. Additionally, marker-specific mutation rates, along with the features of those markers that may affect the mutation rate, have previously been recognized as important for forensic and population genetic purposes [55,57,60]. For example, the two markers with the highest mutation rate in the literature are ARA and DXS8377, both of which contain trinucleotide repeats [89]. In addition, both have very long repeat stretches (on the order

of 10-50 repeat units) that are highly polymorphic (>20 alleles). The contribution of each of these characteristics, as well as the necessity of marker-specific rates, has not yet been investigated for X STR markers. A summary of pooled marker-specific mutation rates from the literature are noted in Table 7B (see "Results" section). Note that the ARA locus is no longer considered suitable for forensic use because it falls within the coding region of a gene in which a mutation would give rise to X-linked spinal and bulbar muscular atrophy [90].

Mutations typically occur as a result of strand slippage during DNA replication, and are the major mechanism responsible for the high degree of polymorphism seen in human microsatellites [54]. Single-step mutations (insertions or deletions of one repeat unit) are most common, affecting longer alleles more frequently than shorter ones [60]. Mutation rates can differ between males and females, with one estimate of the ratio of paternal to maternal mutations at 17:3 [60]. Mutation rates can also vary with population, as was demonstrated in several studies of Y chromosomal markers in which the African American population had a slightly higher mutation rate than other populations [61,62].

In this study, 15 commonly utilized X STR markers with wide variation in allele size, repeat structure, and polymorphism were typed in families from the largest U.S. population groups for which there exists corresponding frequency data: African American, U.S. Caucasian, and U.S. Hispanic. Overall and marker-specific mutation rates were determined as part of the validation of the X STR marker system for forensic use. Besides assessing the X STR mutation rate, this study sought to identify factors that might influence this rate, such as repeat type (i.e., trinucleotide or tetranucleotide), repeat motif (i.e. simple or complex), population, and parental origin of the allele.

Linkage

Standard 8.1.3.2 of the U.S. DNA Advisory Board's Quality Assurance Standards for Forensic DNA Testing Laboratories states that laboratories shall establish and document match criteria on the basis of empirical data [2]. Another requirement of the 1991 ISFG report relating to the use of DNA polymorphisms is that questions of independent assortment and linkage disequilibrium

be addressed [1]. For autosomal STRs, this ensures that the product rule can be used to multiply individual marker frequencies together to determine the overall rarity of a profile. It does not preclude the use of linked markers, however. Y chromosomal STRs, for example, are linked to one another and are considered together as a group called a haplotype. Haplotype frequencies are measured directly from population data, and the counting method is used to determine the rarity of the profile. It follows that X chromosomal STRs may be a combination of the two techniques: the organization of several physically close markers into linkage groups, forming haplotypes, whose frequencies could then be multiplied together once independent assortment of the groups was established.

From the ISFG report, it is clear that both linkage and linkage disequilibrium must be studied. Linkage refers to the co-segregation of closely located markers within a pedigree and can be measured by calculating the recombination fraction (RF) from family samples. A set of families satisfying the requirements of linkage study (multiple generations and offspring) have been established at the National Institute of General Medical Sciences (NIGMS) repository from lymphoblastic cell lines donated by the Centre d'Etudes du Polymorphisme Humaine (CEPH). The collection includes families from Utah, France, Venezuela, and Amish country, and is an important resource for the characterization of DNA polymorphisms and the construction of the human genetic map. In addition, any research effort that requires access to a common dataset can find value in the use of the CEPH reference families, as evidenced by the large amount of data that has already been collected from them and shared in a database made available to contributing researchers. Several reliable linkage maps of the human genome have resulted from such collaborations [91-93]. In fact, most forensic publications refer to the location of markers on the X chromosome according to the Marshfield map [93], which is based upon analysis of recombination rates in a subset of 8 CEPH families. The CEPH families were also used to create the Rutgers combined linkage-physical map of the human genome, which is a denser map incorporating both sequence-based positional information as well as recombination-based data [94]. This Rutgers map has been used to generate a consolidated list of physical and genetic distances between 39 commonly-used forensic X chromosomal markers [95], but since the entire set of 39 markers had not been directly measured, some genetic distances were interpolated.

Additionally, certain X chromosomal markers are missing from this analysis (DXS6795, for example). Therefore, while the physical map of the X chromosome may be well-established, further study of chosen markers to create a genetic map through family studies is required.

Linkage disequilibrium measures the non-random association of two or more alleles that are not necessarily closely located on a chromosome, and is estimated from allele and haplotype frequencies. Statistical tests designed to indicate linkage disequilibrium across the genome can potentially highlight co-segregating markers on the X chromosome that may be physically or genetically linked. However, large sample sizes are required to obtain reliable estimates, and this measure alone cannot establish groups of markers that should be considered as haplotypes. While most population studies include a test for linkage disequilibrium, little scrutiny of linkage on the X chromosome has been performed. However, such information is necessary to the application of information obtained from such markers. Early linkage studies performed with 182 mother-multiple son constellations produced a map of the X chromosome that divided 16 X chromosomal STRs into four linkage groups [30]. The hypothesis was that alleles at linked markers combine to form haplotypes that could recombine during meiosis as "blocks." In the same study, linkage disequilibrium was estimated from a population of over 200 males and showed association between only two markers: DXS7424 and DXS101. No further confirmation of this proposed linkage situation was undertaken at this point. The Investigator Argus X-12 kit takes advantage of these four proposed linkage groups, analyzing haplotypes composed of three markers in each group [84]. Since this initial work, a number of different sets of physically close markers have been studied, demonstrating that alleles do indeed co-segregate as stable haplotypes [40,96,97], especially markers located around the centromere where recombination rates are reduced [98,99]. Still, little research on the linkage situation has been completed with the growing number of available markers across the entire chromosome.

In a study of the Argus X-8 kit, Tillmar *et al.* [100] were able to use observed haplotype frequencies to reveal linkage disequilibrium between markers within the same linkage group, but not between markers located in different linkage groups. The study showed that the paternity index would be significantly influenced if this observed linkage disequilibrium was not taken

into account. Additionally, 32 families were studied and the recombination fraction between linkage groups 3 and 4 was found to be approximately 25%, indicating non-random assortment. Of note, the mothers used in this study did not have confirmed haplotypes; that is, the gametic phase was unknown since no grandparents were included in the study. Therefore, two possibilities had to be considered for each child and accounted for in their analysis of linkage. Though it was shown that both linkage and linkage disequilibrium should be taken into account when using X chromosomal STRs, the limited number of informative meioses in these studied families did not allow a detailed picture of the linkage situation.

In another recent recombination report, three-generation pedigrees were analyzed at 39 X chromosomal STR markers [101]. Previous studies were confirmed, including a loose linkage between groups 3 and 4, which could be potentially misleading in certain cases of kinship analysis. The need for larger, collaborative recombination studies was emphasized by this study.

The Investigator™ Argus X-12 kit was designed with the specialized linkage situation on the X chromosome in mind. Because the chosen markers reside on a single chromosome, the initial four markers were chosen because they were physically far apart and presumably unlinked. Additional markers were chosen specifically based upon their reported linkage to the original markers, creating four linkage trios that could be treated as haplotypes: DXS8378-DXS10135-DXS10148, DXS7132-DXS10079-DXS10074, DXS10103-HPRTB-DXS10101, and DXS10146-DXS10134-DXS7423. This first trio was proposed for inclusion into the commercial multiplex after confirmation of heterogeneity in a German population and a small recombination study (89 informative meioses) in which the stability of the region containing the three proposed markers was assessed [40]. DXS10074 and DXS10079 were validated for forensic use through a study of their allele structure and recombination rate in a German population [96,102]. Though no recombination was observed during this study, less than 92 informative meioses were examined and linkage disequilibrium was established. In comparison, a study of two Brazilian populations revealed ambiguous results for DXS10079 and DXS10074 and a third marker within a 280-kb region of Xq12, where significant linkage disequilibrium was confirmed in the absence of an indication of significant linkage [103]. Evaluation of DXS10103 and DXS10101 for

acceptable heterozygosity and reliable amplification was performed as part of a study of the 133.14-133.45 Mb region of the X chromosome surrounding HPRTB [104]. Haplotype stability of the trio DXS10146-DXS10134-DXS7423 was assessed in a recombination study of less than 109 informative meioses [105]. Though linkage disequilibrium was not tested due to the small sample size, two crossing over events were observed between DXS10146 and DXS10135 out of 80 informative meioses. The authors still recommended these two markers for inclusion into the new commercial kit due to their high degree of polymorphism, pending a more accurate estimation of the genetic distance between them.

Undoubtedly the use of multiple markers located on a single chromosome necessitates special consideration of the potential for linkage between closely situated markers. Potential linkage must be addressed in order to allow for the meaningful calculation of forensically relevant statistics. This study sought to answer the following basic questions towards a potential solution:

3. Is complete linkage observed between each of the markers within the described linkage groups?
4. Is the recombination rate equal to 0.5 (free recombination) between the linkage groups?

Mixture multiplex

Mixture interpretation is an important part of the forensic scientist's role in evaluating evidence from a crime scene, where mixed stains can be common. However, routine analysis of mixtures continues to be a challenge, and varying solutions have been proposed and adopted for use [3-5]. One approach that has thus far only been partially explored is the use of markers on the sex chromosomes to aid in the mixture analysis process. Several studies reported the ability of Y STR markers to detect the male component in male-female mixtures from sexual assault cases [6,7]. Similarly, it has been suggested that X chromosomal markers may help to expose a female profile in male background, such as vaginal cells on a penis or female cells from male fingernail scrapings [8]. Due to their unique inheritance patterns, gonosomal markers hold the potential to supplement traditional mixture testing in certain specific situations.

To this end, a mixture multiplex has been created that combines markers from both the X and the Y chromosomes in an attempt to aid in the interpretation of such mixtures, providing clues as to the number of contributors and the sex of those contributors. By maximizing the information gained from these mixtures, the direction of further testing could potentially be influenced and optimized. The benefits and limitations of using gonosomal STRs in the evaluation of mixed evidence in the forensic setting were uncovered and discussed during this "proof-of-concept" study in the context of current and future efforts.

Methods

Mutation rate study and population database creation

Sample selection

Anonymous DNA extracts from non-excluded cases of disputed paternity originated from samples obtained from Analytical Genetic Testing Center (Denver, CO) and archived at the Department of Forensic Sciences, George Washington University (Washington, DC). These samples were previously typed during paternity testing at 3-5 SNP loci (ESD, GC, GM, KM, and PGM1) and eight autosomal STR loci using the Powerplex® 1.1 System (Promega Corporation), resulting in both an exclusion probability and a probability of paternity greater than 99% in all cases. Samples were provided by Dr. Moses Schanfield, and represented three major U.S. populations: African American, U.S. Caucasian, and U.S. Hispanic. Six additional U.S. Caucasian families previously typed at a subset of autosomal STR markers [52] were also used. In total, 958 families (parent-child duos and trios) were included in this study (Table 1). Note that the inheritance pattern of the X chromosome makes trios involving male children equivalent to mother-son pairs for the purposes of mutation rate analyses since the father does not contribute any alleles in this situation; the fathers were not typed in many of these cases.

Unrelated individuals from the mutation rate study were used to generate a U.S. population database described separately. U.S. allele and haplotype frequencies were provided for a total of 314 African American (108 males, 206 females), 434 U.S. Caucasian (165 males, 269 females), and 398 U.S. Hispanic (150 males, 248 females) individuals.

Table 1. Summary of samples examined during the mutation rate study and U.S. population database creation.

Number of:	African American	U.S. Caucasian	U.S. Hispanic	U.S. Total
Mother-son duos	147	182	161	490
Mother-daughter duos	11	15	14	40
Father-daughter duos	2	2	7	11
Mother-father-daughter trios	115	167	135	417
Total duos & trios	275	366	317	958
Total individuals typed	584	755	683	2022
Total unrelated individuals*	314	434	398	1146

*The population database is composed of these individuals only.

The use of these samples was reviewed and approved by the U.S. Army Medical Research and Materiel Command Institutional Review Board Office as well as at the institutions from which the samples originated.

PCR amplification and typing

Extracts were collected and processed in high-throughput, 96-well plate format with a witness present at the initial plate creation step (aliquoting of extracts). A subset of representative extracts was quantified using Quantifiler® Human DNA Quantification Kit (Applied Biosystems, Foster City, CA) and all extracts were normalized to approximately 1 ng/µL. Profiles were generated for 15 X STR markers using two mini-X STR multiplexes according to the protocol described in [22].

Analysis of data and quality control measures

Electrophoretic data were analyzed using GeneMapper® ID-X version 1.2 or 1.3 (Applied Biosystems) with custom bins and panels. New alleles were inferred based upon electrophoretic mobility and, when possible, sequenced to confirm repeat structure and length according to the protocol described in [22]; sequencing was also performed on all samples that exhibited null alleles. Electronic transfer of allele calls from GeneMapper® export files to a master file combining alleles from both multiplexes by sample was accomplished through the use of a custom macro in order to reduce the possibility of transcription errors. A single marker

(DXS9902) was included in both multiplexes and compared during analysis to ensure concordance for each sample.

Allele frequencies and forensic efficiency parameters were generated as described in [22] using PowerMarker version 3.25 [106] and the Forensic ChrX Research website version 2.0 [72]. Formulae required to calculate these parameters for both autosomal and X STRs are provided in Table 2. Pairwise linkage disequilibrium between the 15 markers was also tested using PowerMarker. Comparisons between similar published populations at overlapping markers were performed using Arelquin [107].

Table 2. Genetic formulae required for the calculation of forensic efficiency parameters.

Parameter	General formula	X STR formula (if different)
Mean exclusion chance in trios (MECI)*	$\sum_i fi^3(1-fi)^2 + \sum_i fi(1-fi)^3 + \sum_{i<j} fifj(fi+fj)(1-fi-fj)^2$	1. $\sum_i fi^3(1-fi) + \sum_i fi(1-fi)^2 + \sum_{i<j} fifj(fi+fj)(1-fi-fj)$ [108] 2. $1-\sum_i f_i^2 + \sum_i f_i^4 - \left(\sum_{i<j} f_i^2\right)^2$ [109]
Mean exclusion chance in father/daughter duos (MECII)	NA	$1-2\sum_i f_i^2 + \sum_i f_i^3$
Power of discrimination in females (PDf)	$1-2\left(\sum_i f_i^2\right)^2 + \sum_i f_i^4$	Same
Power of discrimination in males (PDm)	$1-2\left(\sum_i f_i^2\right)^2 + \sum_i f_i^4$	$1-\sum_i f_i^2$
Expected heterozygosity (H(exp))	$1-\sum_i f_i^2$	Same
Polymorphism information content (PIC)	$1-\sum_i f_i^2 - \left(\sum_i f_i^2\right)^2 + \sum_i f_i^4$	Same
Power of exclusion (PE)	$H^2(1-(1-H)H^2)$, H = H(exp)	Same
Paternity index (PI)	$1/2\sum_i f_i^2$	Same

*X STR formula applies only to trios involving daughters.

Mutations within families were identified by comparison to appropriate parent(s) using a spreadsheet program. Extracts from all families in which a potential mutation was discovered were sequenced at the affected marker in order to confirm profiles and rule out the presence of null alleles. Additionally in families with mutations, autosomal STR results were obtained using the AmpF*l*STR® Identifiler® PCR Amplification Kit (Applied Biosystems) according to the manufacturer's recommended protocol in order to confirm relationships. Custom software known as Laboratory Information Systems Applications (Future Technologies, Inc., Fairfax, Virginia) was used to determine the parentage index in each case using a one parent-one child calculation.

X STR mutation rates were calculated as the number of mutations divided by the total number of meioses. Confidence intervals (CI) were estimated using the exact binomial distribution [110] via spreadsheet formulas provided at http://statpages.org/confint.html. For the purposes of this paper, overall mutation rate refers to the general X STR mutation rate resulting from observed meioses at more than one marker, while the marker-specific rate describes the rate at only a single marker.

In order to confirm the findings of this study by comparison as well as compile a robust large-scale dataset for use by the forensic community, a literature review of published mutation rate studies available at the time of manuscript preparation was conducted. Data mining necessary to compile the included list of published and combined (this study plus published) mutation rates by marker was accomplished by recalculating the mutation rate within 33 independent studies [9,10,13,14,19,24-51] according to the described number of mutations and meioses broken down by marker. Individual studies (and, for the combined dataset, this study) were then combined for each marker to determine both the marker-specific and overall mutation rates, and confidence intervals were assigned to these pooled values. Studies in which there was any ambiguity as to the exact value of the parameters necessary for these calculations were excluded. A complete list of published mutation rate studies can be found in Table 7 (see "Results" section).

Linkage testing

Sample selection and typing

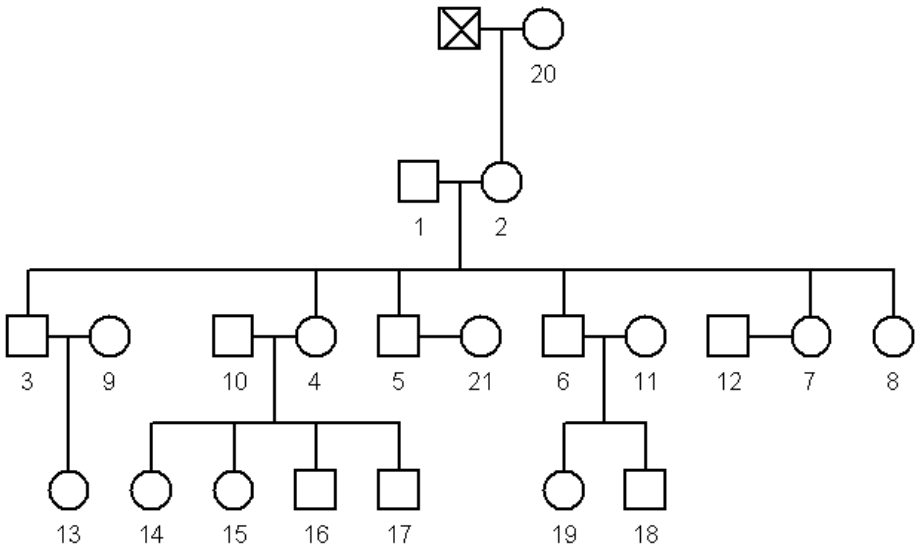
Before any test of linkage was initiated, the appropriate sample sets were identified for analyses. Such sets included the following family types:

Type I: three-generation families typically including a maternal grandfather, mother, and her child(ren); these families include, at a minimum
- A maternal grandfather, a mother, and her son
- A maternal grandfather, a mother, a father, and their daughter

Type II: two-generation families including a mother and two or more of her children; this type can include several different scenarios such as
- A mother and multiple sons
- A mother, a father, and two or more daughters
- A mother, a father, one son, and one daughter

Complex pedigrees including several generations and multiple offspring could therefore include multiple different families as defined above; see Figure 1 for an example.

Figure 1. Multiple family types present within one pedigree. This example of a pedigree used for linkage analysis (FamID 6) contains distinct sub-families of both type I and type II, as described in the accompanying table. Complex pedigrees such as this one were divided into their respective sub-families for analysis.



FamID	Family type	Individuals included	Description
6A	II	1, 2, 3, 4, 5, 6, 7, 8	Mother and father with multiple children
6B	II	6, 11, 18, 19	Mother and father with son and daughter
600	I	1, 2, 4, 10, 14, 15, 16, 17	Maternal grandparents, mother and father with multiple children

A set of commercially-available extracts from families satisfying these requirements was purchased from the National Institute of General Medical Sciences (NIGMS) repository. This collection of lymphoblastic cell lines donated by the Centre d'Etudes du Polymorphisme Humaine (CEPH) included families from Utah, France, Venezuela, and Amish country. Additionally, six U.S. Caucasian families previously typed at a subset of autosomal STR markers [52] and also used as part of the mutation rate study were included, as well as a subset of other families from the mutation rate study. In total, 158 families (54 Type I and 104 Type II) that were appropriate for linkage analysis according to the criteria defined above were identified within this dataset.

PCR amplification and typing occurred as described above for the mutation rate study.

Linkage analysis

Once appropriate sample sets were identified and typed, analyses were performed using two different methods. The first was manual and used only Type I families owing to the ability to unambiguously determine the source chromosome for each of the mother's X STR alleles via the maternal grandfather's profile. The general steps were as follows:

1. Use the maternal grandfather's profile to determine the paternal contribution to mother's profile and deduce maternal contribution.
2. Use father's profile to assign and remove the paternal contribution from analysis of daughters.
3. Compare maternal contribution (source chromosome) in each of the children to the mother's two chromosomes, looking specifically for instances where alleles of adjacent markers resulted from different chromosomes, therefore indicating a potential recombination event had occurred.

Each observed recombination event was noted and totaled for each marker pair. This total was then divided by the total number of unambiguous meioses for that marker pair, producing the observed recombination rate. Logarithm of the odds (LOD) scores (Z) were then calculated using the following formula:

$$Z = \log_{10}((1-\Theta)^{NR}\Theta^{R})/0.5^{NR+R})$$

where Θ is the recombination rate

NR represents the number of meiosis for which no recombination was observed

R represents the number of meiosis for which recombination was observed.

An example detailing this manual analysis is presented in Appendix A.

The second analysis method was performed as described by Nothnagel, *et al.* [53] with modifications on the front-end specific to this dataset. A PLINK format text file was created by assigning the following to each sample: family identification number (FamID), patient

identification number (PID), father's identification number (FID), mother's identification number (MID), sex (1 for male and 2 for female), phenotype (not used; coded zero for all samples) followed by 2 alleles for each of the 15 X STR markers. The second allele in male samples and any markers with missing allelic information were coded with a zero as a placeholder.

The design of the scripts used to perform likelihood calculations in this strategy limited the family structure to only those with male children; that is, type I families with a maternal grandfather, mother, and one or more sons, or a type II family with a mother and >1 son. Since the dataset included many families that did not conform to this structure, it was necessary to further modify the data before analysis to maximize the overall number of meiosis that could be included. This process was similar to that described for the second step of the classical analysis above; fathers' profiles were used to manually remove the paternal contribution from the daughter's profiles, effectively turning these daughters into sons. In doing this, several families required separation into multiple distinct families to allow a female profile to serve as both a daughter-turned-son as well as a mother. This was the case, for example, with the family shown in Figure 1, where individual 4 served as a mother in family 600 and a daughter-turned-son in family 6A. Also, any paternal mutations discovered during this process were catalogued.

Since this process of converting daughters to sons required manual manipulation of the data, additional quality control measures were incorporated to ensure the resulting dataset was error-free. To begin, all meioses were screened for mutations using a spreadsheet program. In particular, for daughters-turned-sons and their mothers, a mutation could indicate either a mistake in the process of removing the paternal contribution or a true maternal mutation. Mistakes were corrected by referring to the original unedited profiles, and true mutations were categorized as maternal, paternal, or unknown origin. Mutations with unknown origin discovered in daughters-turned-sons were considered paternal in order to allow inclusion in the linkage analysis. Since the linkage analysis scripts also had the limitation of being unable to tolerate mutations of >1 step or between maternal grandfathers and mothers, the loci in these cases were rendered uninformative for linkage by assigning either a 0,0 genotype in the affected

individuals or a homo- or hemizygous genotype to all individuals at that particular locus. Other mutations that did not have an impact on linkage analysis (between a maternal grandmother and mother, for example) or could be determined to be maternal were maintained in the dataset.

Lastly, duplications and triplications were reduced to the appropriate allele number by examining parental profiles and using a zero designation at any ambiguous or undeterminable loci. This file was submitted for analysis to the authors of the referenced publication, who made several modifications to the original scripts to accommodate this larger set of markers.

Mixture multiplex development and characterization

Selection of markers and primer design

A combination of markers located on both sex chromosomes was desired, including X and Y STRs as well as STRs within the X-Y homologous region, termed XY markers. A minimum of two markers of each type were targeted for inclusion into the final multiplex, along with the non-repetitive locus SRY for additional sex confirmation. A review of the literature resulted in a list of potential markers; selection was limited to markers for which there were published population genetic studies in order to exploit the collective knowledge of these established markers and their relevant characteristics, simplifying the process. Potential utility within the mixture multiplex was assessed according to the following criteria: (a) potential for small amplicon size; (b) large allele range with high degree of polymorphism; and (c) established use within the forensic community. Additionally, only single-copy Y STR markers were considered, and simple repeats were targeted in order to take advantage of the absence of known microvariants. These considerations were aimed towards simplifying interpretation when more than one individual was present. Markers best matching these criteria were selected for inclusion into the multiplex, termed MIXplex, and organized according to amplicon size.

Approximately 200 bp flanking either side of the repeat regions for the chosen markers were downloaded from the UCSC Genome Browser [111] using their BLAT In Silico PCR search [71] and the published primers. In many cases, published amplification primers were sufficient for incorporation into the multiplexes, but several markers required one or both primers be

redesigned. When necessary, primers were designed using the web-based program Primer3 [112], and selected primers were screened for use in multiplex reactions using the web-based algorithm AutoDimer [113].

One primer for each marker was labeled at the 5' end with a fluorescent dye, either 6FAM™, VIC®, NED™, or PET™ (Applied Biosystems). A tail was added to the complementary primer in the set at the 5' end in order to promote the complete adenylation of PCR products [114] and, in some cases, provide adequate spacing between amplicons in the multiplex. This tail was either GTTTCTT, ATT, or a single G.

Source of DNA samples

Extracts utilized in this study were generally commercially available, highly concentrated control DNAs for which the profiles at each of the markers chosen for inclusion in the multiplex were known (Table 3). One exception was an internally maintained female control DNA designated here as AFDIL-1. Since a large quantity of extract would be needed to optimize the multiplex and later generate and test multiple different mixtures, these control DNAs were a practical and reliable solution. All controls were given a sample ID that would allow rapid identification of sex with just a glance at the name (i.e. male 1, male 2, female 1, etc.) for simplicity during the development and testing process.

Table 3. Panel of control DNAs used for evaluation and quality control experiments.

Control DNA	Source	Part Number	Sex	Sample ID
Quantifiler® Human DNA Standard	Quantifiler® Human DNA Quantification Kit, Applied Biosystems	4343895	Male	Male 1
9948	Promega Corporation	DD2061	Male	Male 2
2800M	Promega Corporation	DD7101	Male	Male 3
9947a	Promega Corporation	DD1001	Female	Female 1
AFDIL-1	In-house	NA	Female	Female 2
K562	Promega Corporation	DD2011	Female	Female 3

Amplification, detection and analysis

Amplification was performed in a 10 µL reaction that consisted of 1x PCR Buffer II (Applied Biosystems), 1.5 units AmpliTaq Gold™ DNA polymerase (Applied Biosystems), 0.25 mM

dNTP Mix (Applied Biosystems), 0.15 mg/mL bovine serum albumin (Sigma-Aldrich, St. Louis, MO), 2 mM magnesium chloride solution (Applied Biosystems) and 2 µL of primer mix. Primer mix concentrations were adjusted empirically to balance peak heights within each multiplex and individual concentrations are listed in Table 20 (see "Results" section). Each preparation of primer mix was tested on a panel of known control samples to ensure consistent quality and correct genotypes could be obtained before use on unknowns.

Annealing temperatures of 55, 60, and 64 ^{o}C were evaluated on a GeneAmp® 9700 (Applied Biosystems). Final thermal cycling parameters were as follows:

- 96 ^{o}C for 10 minutes
- 28 cycles: 94 ^{o}C for 1 minute, 55 ^{o}C for 1 minute, 72 ^{o}C for 1 minute
- 60 ^{o}C for 45 minutes
- 4 ^{o}C soak

Samples were prepared for capillary electrophoresis by adding 1 µL amplified product to 8.7 µL Hi-Di™ formamide (Applied Biosystems) and 0.3 µL LIZ-500 size standard (Applied Biosystems). Samples were injected at 3.0 kV for 10 s and run using a 36 cm array and POP-6™ polymer on a 3130xl Genetic Analyzer (Applied Biosystems). Data were analyzed using Genemapper ID version 3.2 or ID-X version 1.1, 1.2, or 1.3 (Applied Biosystems) with custom bins and panels.

When necessary to confirm repeat structure and length, allele sequencing was performed according to the protocol described in [22] using the primers given in Table 4.

Table 4. Unlabeled primers used for sequencing purposes.

Marker Name	Forward Primer Sequence (5'-3')	Reverse Primer Sequence (5'-3')
DYS393	TTGTAGTTATGTTTTATTTGTCATTCC	AAATGTTACAAAAAGAATGGCCTA
DYS438	TGATGCAAGAAAGATTCACTGAT	AGGAGAATCGCTTGAACCTG
DXS6795	TTCATGCTGTTGCTTTCCAG	CCATCCCCTAAACCTCTCAT
DXS6789	TCAAGCTTGCAGACAGCCTA	TCGAAAAGATAGCCAATCACTG
GATA31E08	AGCAAGGGGAGAAGGCTAGA	TCAGCTGACAGAGCACAGAGA

Sensitivity testing was performed to evaluate the lower limits of the multiplex with single-source samples. The initial concentration of two female and one male sample was determined using the Quantifiler® Human DNA Quantification Kit (Applied Biosystems). Samples were then serially diluted with Tris-low-EDTA buffer (TLE; 10 mM Tris, 0.1 mM EDTA, pH 8.0) to form the following dilution series: 1000, 500, 200, 100, 50, and 25 pg/µL. One microliter of each concentration was tested in triplicate to determine the minimum quantity of input DNA required to reliably obtain full profiles for both male and female samples.

Mixture creation and analysis

Six control DNAs (Table 3) were quantified using the Quantifiler® Human DNA Quantification Kit and normalized to 1 ng/µl using TLE buffer. A subset of two male (Quantifiler® standard, 2800M) and two female (AFDIL-1, K562) samples were chosen to create 4 male-female mixtures, 1 male-male mixture, and 1 female-female mixture at the following ratios: 0:100, 10:90, 20:80, 30:70, 40:60, 50:50, 60:40, 70:30, 80:20, 90:10, 100:0. One microliter of each mixture at each ratio was amplified in duplicate or triplicate with the MIXplex and once with the autosomal STR kit AmpF*l*STR® Identifiler® (Applied Biosystems) according to the manufacturer's recommended protocol.

In additional to the creation of these artificial mixtures used to test the performance of the MIXplex assay, the profiles of all six controls were used to generate 63 theoretical mixture profiles for all possible combinations of male(s)-female(s), multiple males, and multiple females. These profiles were then coded with generic numeric identifiers and randomized in order to mask the sex and number of contributors to the profile, and used to test the theoretical ability of the MIXplex compared to Identifiler®. Table 5 lists both the artificial and theoretical mixtures and their components.

Table 5. Naming conventions and components of mixtures created for evaluation of MIXplex. Samples/mixtures in bold were prepared and amplified while the others' profiles were theoretically determined based on known profiles of the individual contributors.

Sample identifier	Profile identifier	Quantifiler® standard (male)	9948 (male)	2800M (male)	9947a (female)	AFDIL-1 (female)	K562 (female)
female 1	**123**				**X**		
female 2	**113**					**X**	
female 3	**152**						**X**
male 1	**106**	**X**					
male 2	**133**		**X**				
male 3	**138**			**X**			
mixture 1-A1	136	X			X		
mixture 1-A2	**118**	**X**				**X**	
mixture 1-A3	**135**	**X**					**X**
mixture 1-A4	149		X		X		
mixture 1-A5	119		X			X	
mixture 1-A6	137		X				X
mixture 1-A7	109			X	X		
mixture 1-A8	**139**			**X**		**X**	
mixture 1-A9	**148**			**X**			**X**
mixture 1-F1	160				X	X	
mixture 1-F2	163				X		X
mixture 1-F3	**107**					**X**	**X**
mixture 1-M1	108	X	X				
mixture 1-M2	**140**	**X**		**X**			
mixture 1-M3	121		X	X			
mixture 3-A1	141	X	X		X		
mixture 3-A2	120	X	X			X	
mixture 3-A3	147	X	X				X
mixture 3-A4	146	X		X	X		
mixture 3-A5	145	X		X		X	
mixture 3-A6	142	X		X			X
mixture 3-A7	110		X	X	X		
mixture 3-A8	122		X	X		X	
mixture 3-A9	134		X	X			X
mixture 3-B1	143	X			X	X	
mixture 3-B2	101		X		X	X	
mixture 3-B3	132			X	X	X	
mixture 3-B4	111	X			X		X
mixture 3-B5	131		X		X		X
mixture 3-B6	117			X	X		X
mixture 3-B7	124	X				X	X
mixture 3-B8	112		X			X	X
mixture 3-B9	103			X		X	X

Table continues on next page.

Sample identifier	Profile identifier	Quantifiler® standard (male)	9948 (male)	2800M (male)	9947a (female)	AFDIL-1 (female)	K562 (female)
mixture 3-F	144				X	X	X
mixture 3-M	125	X	X	X			
mixture 4-A1	102	X	X	X	X		
mixture 4-A2	151	X	X	X		X	
mixture 4-A3	157	X	X	X			X
mixture 4-B1	150	X			X	X	X
mixture 4-B2	114		X		X	X	X
mixture 4-B3	126			X	X	X	X
mixture 4-C1	128	X	X		X	X	
mixture 4-C2	116	X	X		X		X
mixture 4-C3	127	X	X			X	X
mixture 4-C4	156	X		X	X	X	
mixture 4-C5	153	X		X	X		X
mixture 4-C6	104	X		X		X	X
mixture 4-C7	129		X	X	X	X	
mixture 4-C8	158		X	X	X		X
mixture 4-C9	161		X	X		X	X
mixture 5A	105	X	X	X	X	X	
mixture 5B	130	X	X	X	X		X
mixture 5C	155	X	X	X		X	X
mixture 5D	159	X	X		X	X	X
mixture 5E	162	X		X	X	X	X
mixture 5F	154		X	X	X	X	X
mixture 6P	115	X	X	X	X	X	X

Results

Mutation rate study

Overall mutation rates

A total of 20,625 meioses in confirmed family trios or duos were analyzed at 15 X STR markers (DXS6789, DXS7130, GATA31E08, GATA165B12, GATA172D05, DXS10147, DXS8378, DXS7132, DXS6803, HPRTB, DXS7423, and DXS9902, DXS7424, DXS101, and DXS6795). Eighteen mutations were observed across 7 of the 15 markers and in all three U.S. population groups, resulting in an overall mutation rate of 8.73×10^{-4} (Table 6). The probability of paternity or maternity for the families showing mutation(s) were all $\geq 99.99\%$ and allele sequencing

results confirmed profiles (data not shown). Compared with the overall rate in this study, 14 published X STR studies reported higher mutation rates, but only two approached significance based upon confidence interval bounds: 4.76 x 10^{-3} [48] and 2.09 x 10^{-3} [30], both observed in German populations (Table 7). None of the studies with smaller overall mutation rates were significantly different from this study's overall rate. Many studies that exhibited low overall mutation rate values also relied upon a smaller number of meioses (hundreds rather than thousands), and the uncertainty of the resultant rate was reflected in larger confidence intervals. One notable exception was the study relying upon only 180 meioses in a Korean population that observed the highest overall mutation rate (5.56 x 10^{-3}) [32]. The need for larger sample sizes was underscored by these data, and therefore mutation rates taking into account published data as well as data collected as part of this study were calculated.

Table 6A-C. X STR mutation rates from this study (A & B), a literature summary (C), and the combined datasets (C).

A.

	African American				U.S. Caucasian			
Marker	**Mutations**	**Meioses**	**Mutation rate (x10^{-3})**	**95% CI (x10^{-3})**	**Mutations**	**Meioses**	**Mutation rate (x10^{-3})**	**95% CI (x10^{-3})**
DXS8378	0	390	0.00	0-9.4	3	533	5.63	1.2-16.4
DXS9902	0	390	0.00	0-9.4	1	533	1.88	0.05-10.4
DXS6795	0	390	0.00	0-9.4	0	533	0.00	0-6.9
DXS7132	1	390	2.56	0.06-14.2	3	533	5.63	1.2-16.4
DXS6803	0	390	0.00	0-9.4	0	533	0.00	0-6.9
DXS6789	0	390	0.00	0-9.4	0	533	0.00	0-6.9
DXS7424	0	390	0.00	0-9.4	0	533	0.00	0-6.9
DXS101	0	390	0.00	0-9.4	0	533	0.00	0-6.9
GATA172D05	0	390	0.00	0-9.4	1	533	1.88	0.05-10.4
DXS7130	0	390	0.00	0-9.4	0	533	0.00	0-6.9
GATA165B12	0	390	0.00	0-9.4	0	533	0.00	0-6.9
HPRTB	0	390	0.00	0-9.4	1	533	1.88	0.05-10.4
GATA31E08	0	390	0.00	0-9.4	0	533	0.00	0-6.9
DXS10147	0	390	0.00	0-9.4	0	533	0.00	0-6.9
DXS7423	0	390	0.00	0-9.4	0	533	0.00	0-6.9
Overall	**1**	**5850**	**0.17**	**0.004-0.95**	**9**	**7995**	**1.13**	**0.51-2.1**

Table continues on next page.

B.

	U.S. Hispanic				This study (U.S. total)			
Marker	**Mutations**	**Meioses**	**Mutation rate (x10^{-3})**	**95% CI (x10^{-3})**	**Mutations**	**Meioses**	**Mutation rate (x10^{-3})**	**95% CI (x10^{-3})**
DXS8378	1	452	2.21	0.06-12.3	4	1375	2.91	0.79-7.4
DXS9902	3	452	6.64	1.4-19.3	4	1375	2.91	0.79-7.4
DXS6795	0	452	0.00	0-8.1	0	1375	0.00	0-2.7
DXS7132	0	452	0.00	0-8.1	4	1375	2.91	0.79-7.4
DXS6803	1	452	2.21	0.06-12.3	1	1375	0.73	0.02-4.1
DXS6789	2	452	4.42	0.54-15.9	2	1375	1.45	0.18-5.2
DXS7424	0	452	0.00	0-8.1	0	1375	0.00	0-2.7
DXS101	0	452	0.00	0-8.1	0	1375	0.00	0-2.7
GATA172D05	0	452	0.00	0-8.1	1	1375	0.73	0.02-4.1
DXS7130	0	452	0.00	0-8.1	0	1375	0.00	0-2.7
GATA165B12	0	452	0.00	0-8.1	0	1375	0.00	0-2.7
HPRTB	1	452	2.21	0.06-12.3	2	1375	1.45	0.18-5.2
GATA31E08	0	452	0.00	0-8.1	0	1375	0.00	0-2.7
DXS10147	0	452	0.00	0-8.1	0	1375	0.00	0-2.7
DXS7423	0	452	0.00	0-8.1	0	1375	0.00	0-2.7
Other								
Overall	**8**	**6780**	**1.18**	**0.51-2.3**	**18**	**20625**	**0.87**	**0.52-1.4**

C.

	Literature summary				Combined			
Marker	**Mutations**	**Meioses**	**Mutation rate (x10^{-3})**	**95% CI (x10^{-3})**	**Mutations**	**Meioses**	**Mutation rate (x10^{-3})**	**95% CI (x10^{-3})**
DXS8378	3	2982	1.01	0.21-2.9	7	4357	1.61	0.65-3.3
DXS9902	0	458	0.00	0-8.0	4	1833	2.18	0.59-5.6
DXS6795	*na*	*na*	*na*	*na*	0	1375	0.00	0-2.7
DXS7132	10	4064	2.46	1.2-4.5	14	5439	2.57	1.4-4.3
DXS6803	2	1015	1.97	0.24-7.1	3	2390	1.26	0.26-3.7
DXS6789	3	3478	0.86	0.18-2.5	5	4853	1.03	0.33-2.4
DXS7424	2	1805	1.11	0.13-4.0	2	3180	0.63	0.08-2.3
DXS101	1	2534	0.39	0.01-2.2	1	3909	0.26	0.01-1.42
GATA172D05	0	876	0.00	0-4.2	1	2251	0.44	0.01-2.5
DXS7130	*na*	*na*	*na*	*na*	0	1375	0.00	0-2.7
GATA165B12	0	958	0.00	0-3.8	0	2333	0.00	0-1.6
HPRTB	6	3627	1.65	0.61-3.6	8	5002	1.60	0.69-3.2
GATA31E08	1	1127	0.89	0.02-4.9	1	2502	0.40	0.01-2.2
DXS10147	0	54	0.00	0-66.0	0	1429	0.00	0-2.6
DXS7423	2	2614	0.77	0.09-2.8	2	3989	0.50	0.06-1.8
Other	54	24803	2.18	1.6-2.8	54	24803	2.18	1.6-2.8
Overall	**84**	**50395**	**1.67**	**1.3-2.1**	**102**	**71020**	**1.44**	**1.2-1.7**

Table 7. Published X STR mutation rates. A. Overall mutation rates. The overall mutation rates from 33 published studies were compiled and 95% confidence intervals (CI) calculated. Populations with more than one published mutation rate study were combined to provide population-specific rates (bold). B. Pooled marker-specific X STR mutation rates from published studies with 95% confidence intervals. Markers investigated as part of this study are shown in bold. Ref.: reference; CI: confidence interval.

A. Overall mutation rates.

Ref.	Population	Markers (N)	Mutations (N)	Meioses (N)	Mutation rate (x 10^{-3})	95% CI (x 10^{-3})
[41]	Argentina	7	1	1015	0.99	0.02-5.5
[25]	Austria & Germany	3	0	834	0.00	0-4.4
[23]	China	2	0	312	0.00	0-11.7
[32]	China	4	0	424	0.00	0-8.7
[38]	China	5	8	4295	1.86	0.8-3.7
[46]	China	15	13	11850	1.10	0.6-1.9
	China combined		**21**	**16881**	**1.24**	**0.77-1.9**
[39]	Columbia	10	4	1460	2.74	0.7-7.0
[16]	Germany	1	2	580	3.45	0.42-12.4
[17]	Germany	1	0	340	0.00	0-10.8
[18]	Germany	1	0	404	0.00	0-9.1
[19]	Germany	1	0	300	0.00	0-12.2
[22]	Germany	4	0	372	0.00	0-9.9
[24]	Germany	16	16	7658	2.09	1.2-3.4
[29]	Germany	10	0	500	0.00	0-7.3
[36]	Germany	8	1	2800	0.36	0.01-2.0
[37]	Germany	3	3	1029	2.92	0.6-8.5
[45]	Germany	8	8	1680	4.76	2.1-9.4
	Germany combined		**30**	**15663**	**1.29**	**1.2-2.7**
[43]	Ghana	11	0	198	0.00	0-18.5
[44]	Greenland, Denmark, Somalia	12	20	6156	3.25	2.0-5.0
[35]	Hungary	4	1	768	1.30	0.03-7.2
[27]	Italy	3	0	240	0.00	0-15.2
[34]	Italy	12	0	1080	0.00	0-3.4
	Italy combined		**0**	**1320**	**0.00**	**0-2.8**
[48]	Japan	12	0	648	0.00	0-5.7
[26]	Korea	5	1	180	5.56	0.14-30.1
[40]	Pakistan	13	0	1300	0.00	0-2.8
[42]	Pakistan	5	0	840	0.00	0-4.4
	Pakistan combined		**0**	**2140**	**0.00**	**0-1.7**
[30]	Philippines	5	1	445	2.25	0.06-12.5
[28]	Poland	4	0	320	0.00	0-11.5
[31]	Poland	4	0	264	0.00	0-13.9
[33]	Poland	4	0	600	0.00	0-6.1
	Poland combined		**0**	**1184**	**0.00**	**0-3.1**
[20]	Spain	2	1	214	4.67	0.12-25.8
[21]	Spain	5	0	125	0.00	0-29.1
[47]	Spain	6	4	1164	3.44	0.9-8.8
	Spain combined		**5**	**1503**	**3.33**	**1.1-7.8**
		Totals	**84**	**50395**	**1.67**	**1.3-2.1**

B. Marker-specific mutation rates.

Marker	Mutations (N)	Meioses (N)	Pooled mutation rate (x 10^{-3})	95% CI (x 10^{-3})	Reference
ARA	4	673	5.94	1.6-15.1	[9,13,30,32]
DXS8377	10	1702	5.88	2.8-10.8	[9,10,13,14,28,30-32,42,43,46]
DXS10135	8	1416	5.65	2.4-11.1	[39,40,47,48]
DXS10079	7	1497	4.68	1.9-9.6	[47,49,50]
DXS10103	2	513	3.90	0.47-14.0	[47]
DXS10146	2	513	3.90	0.47-14.0	[47]
DXS10134	4	1127	3.55	0.97-9.1	[39,47,48,51]
DXS10148	3	856	3.50	0.72-10.2	[40,47]
DXS10075	3	984	3.05	0.63-8.9	[49,50]
DXS7132	**10**	**4064**	**2.46**	**1.2-4.5**	[19,30,33-39,41-43,46-49,51]
DXS10074	5	2111	2.37	0.77-5.5	[39,47-51]
DXS6803	**2**	**1015**	**1.97**	**0.24-7.1**	[29,41]
DXS6809	4	2133	1.88	0.51-4.8	[19,41,42,49-51]
HPRTB	**6**	**3627**	**1.65**	**0.61-3.6**	[9,10,13,14,19,24,30,32,34-39,42-44,46-48]
DXS7424	**2**	**1805**	**1.11**	**0.13-4.0**	[13,19,27,30,44,46,49]
DXS8378	**3**	**2982**	**1.01**	**0.21-2.9**	[13,19,30,34-40,42-44,46-48,51]
DXS10101	1	1073	0.93	0.02-5.2	[39,47,48]
GATA31E08	**1**	**1127**	**0.89**	**0.02-4.9**	[19,43,44,49,51]
DXS6789	**3**	**3478**	**0.86**	**0.18-2.5**	[14,19,26,30,41-43,49-51]
DXS7423	**2**	**2614**	**0.77**	**0.09-2.8**	[9,19,28,30,34-39,42-44,46-48,51]
DXS9898	1	1936	0.52	0.01-2.9	[13,30,42,44,46,49,51]
DXS101	**1**	**2534**	**0.39**	**0.01-2.2**	[9,10,13,14,19,25,30-32,42,44,46,49]
DXS10011	0	50	0.00	0-71.1	[13]
DXS10147	**0**	**54**	**0.00**	**0-66.0**	[51]
DXS6810	0	100	0.00	0-36.2	[43]
DXS6793	0	100	0.00	0-36.2	[43]
DXS6797	0	168	0.00	0-21.7	[45]
DXS9902	**0**	**458**	**0.00**	**0-8.0**	[30,43,51]
DXS6807	0	598	0.00	0-6.2	[13,19,30,46]
GATA172D05	**0**	**876**	**0.00**	**0-4.2**	[19,30,32,33,42,43,51]
DXS9895	0	917	0.00	0-4.0	[29,30]
GATA165B12	**0**	**958**	**0.00**	**0-3.8**	[45,49]
DXS6801	0	1084	0.00	0-3.4	[43,49,50]
DXS7133	0	1459	0.00	0-2.5	[13,19,30,33,45,46,49]
DXS6800	0	1694	0.00	0-2.2	[30,31,45,46,49]
DXS981	0	2099	0.00	0-1.8	[10,14,41,43,45,49]

The overall mutation rate for the combined study (published data plus this study) was the result of 102 mutations observed in 71,020 meioses at 36 markers (Table 6). The combined overall rate of 1.44 x 10^{-3}, like the overall rate from the literature summary of 1.67 x 10^{-3}, was higher than that observed for the total U.S. dataset in this study (8.73 x 10^{-4}). This difference can be

explained by considering that additional markers were included in the calculation of the higher two rates that were not included in the total U.S. dataset. When the three markers with the highest marker-specific mutation rates in the literature summary dataset and the combined dataset (ARA, DXS8377, and DXS10135) were removed, the overall mutation rates became 1.33 x 10^{-3} (95% CI: 1.0-1.7 x 10^{-3}) and 1.19 x 10^{-3} (95% CI: 0.94-1.5 x 10^{-3}) respectively. These values further decrease to 1.17 x 10^{-3} (95% CI: 0.79-1.7 x 10^{-3}) for the literature summary dataset and 1.04 x 10^{-3} (95% CI: 0.77-1.4 x 10^{-3}) for the combined dataset if only the 15 markers used in this study are considered. These values are more similar to the overall rate observed in this study, and illustrate the variation of mutation rate with STR marker.

Population-specific mutation rates

The overall rate for the African American population (1.71 x 10^{-4}) was the lowest overall rate, and was significantly smaller than both the overall rate from the literature summary (1.67 x 10^{-3}) as well as the combined overall rate (this study plus literature summary; 1.44 x 10^{-3}). The total number of observed meioses for the African American population was at least 15% less than that of either the U.S. Caucasian and U.S. Hispanic populations, potentially indicating that further study would be necessary to determine if this difference in mutation rate was authentic.

Considering the significance of the difference between overall rates for the African American population compared to the literature summary and combined study, populations from the literature summary with more than one published mutation rate study were combined to obtain one overall mutation rate per population (Table 7A). The overall rates for populations from China [23,32,38,46], Germany [16-19,22,24,29,36,37,45], Italy [27,34], Pakistan [40,42], Poland [28,31,33], and Spain [20,21] showed no significant differences based upon confidence interval bounds from each other, the literature summary, the combined study, or the U.S. Caucasian and U.S. Hispanic populations. However, the overall rates for the published German and Spanish populations were significantly larger than that of the African American population studied here.

Mutations at certain markers appeared to be population-specific in this study as well. For example, mutations at DXS7132 were only observed in the African American and U.S.

Caucasian populations while mutations at DXS6789 occurred only in U.S. Hispanics. Confidence intervals for each marker-specific rate in all three populations demonstrated substantial overlap, however. For the individual U.S. population groups, more meioses will need to be observed before the true marker-specific and/or population-specific mutation rates can be assessed.

Marker-specific mutation rates

In general, the observed marker-specific mutation rates within this study were similar across populations and markers as well as consistent with the overall rate (Table 6). There were three instances, however, where the difference between the marker-specific rate and the overall rate approached significance based upon overlap of the confidence intervals: markers DXS8378 and DXS7132 (each 5.63×10^{-3}) in the U.S. Caucasian population and marker DXS9902 (6.64×10^{-3}) in the U.S. Hispanic population, which exhibited mutation rates higher than the overall rate observed in this study (8.73×10^{-4}). However, these rates were not significantly different from the relevant population-specific overall rates (1.13×10^{-3} for U.S. Caucasians and 1.18×10^{-3} for U.S. Hispanics), and in the collective U.S. population, the marker-specific rates for DXS8378, DXS7132, and DXS9902 agreed with the overall rate.

Two markers exhibited a mutation rate of zero in both this study and published studies: GATA165B12 and DXS10147. Both markers have a relatively small allele range; in U.S. populations, seven alleles were observed at GATA165B12 and eight at DXS10147 (see "U.S. Population Database Creation" within "Results" section). Markers for which no mutation rate studies have been published (DXS7130 and DXS6795) exhibited no mutations in this study. Mutations were observed for the first time at DXS9902 and GATA172D05 in this study, likely owing to the larger number of total meioses examined here.

Within the combined dataset representing the pooled rates from this study and the literature summary, there were differences between the mutation rates of different markers (Tables 6 and 7B). The difference between the mutation rate for DXS7132 (2.57×10^{-3}), which exhibited the largest marker-specific rate, and DXS101 (2.56×10^{-4}), GATA165B12 (zero mutations

observed), and DXS7423 (5.01×10^{-4}) approached significance based on the confidence interval bounds. Additionally, the combined overall mutation rate (1.44×10^{-3}), though generally higher than many marker-specific rates in the combined dataset, yielded a confidence interval completely contained within those of the marker-specific rates for all but DXS101 and GATA165B12. Given the relatively large (>1300) number of meioses for all markers and the small confidence interval ranges, it is likely this difference is genuine for at least these two markers, and their true mutation rates are indeed lower than for other markers or for X STRs in general. Further study, however, could help confirm this hypothesis.

Characterization of observed mutations

Characteristics of the 18 mutations observed in this study are summarized in Table 8. Examination of the progenal and parental genotypes revealed that all observed mutations could be explained by a change of one repeat unit. Both gains and losses of repeat units were noted in the observed mutations, with gains outnumbering losses by approximately 2:1. For approximately half of the mutations, the progenitor allele was the most frequent for that particular marker and population, reflecting the greater opportunity for observing mutation. Additionally, since in most cases the intermediate alleles tended to be the most frequent, mutations involving the largest or the smallest alleles were not observed in this dataset. Because the opportunity to observe mutations in these extremes was so much lower than for the intermediate-sized alleles, no attempt to correlate mutation rate with progenitor allele size was made in this study.

Table 8. Characteristics of the 18 observed mutations. One-step mutations and single mutational events were assumed for assigning "Origin" and "Result," as other types of mutations are less common. Paternal genotypes are not listed in the table for families with sons.

Marker	Population	Origin	Maternal genotype	Paternal genotype	Progenal genotype	Result
DXS8378	U.S. Caucasian	Maternal	11,12	--	10	Loss
DXS8378	U.S. Caucasian	Unknown	10,12	10	10,**11**	Unknown
DXS8378	U.S. Caucasian*	Paternal	10,10	12	10,**11**	Loss
DXS8378	U.S. Hispanic	Maternal	11,12	--	13	Gain
DXS9902	U.S. Caucasian**	Paternal	9,10	10	9,**9**	Loss
DXS9902	U.S. Hispanic	Paternal	10,10	11	10,**12**	Gain
DXS9902	U.S. Hispanic	Unknown	10,11	11	11,**12**	Gain
DXS9902	U.S. Hispanic	Maternal	11,12	10	**10**,10	Loss
DXS7132	African American	Paternal	14,16	14	**15**,16	Gain
DXS7132	U.S. Caucasian*	Maternal	13,14	--	**15**,16	Gain
DXS7132	U.S. Caucasian	Paternal	13,14	15	14,**16**	Gain
DXS7132	U.S. Caucasian**	Paternal	15,15	14	**13**,15	Loss
DXS6803	U.S. Hispanic	Maternal	12,12.3	13.3	**13.3**,13.3	Gain
DXS6789	U.S. Hispanic	Paternal	20,23	22	20,**23**	Gain
DXS6789	U.S. Hispanic	Paternal	20,23	23	20,**22**	Loss
GATA172D05	U.S. Caucasian	Unknown	6,10	10	10,**11**	Gain
HPRTB	U.S. Caucasian	Maternal	11,12	--	13	Gain
HPRTB	U.S. Hispanic	Paternal	14,14	12	**13**,14	Gain

*Two mutations observed within a single family.
**Two mutations observed within a second single family.

Two U.S. Caucasian families exhibited two mutations each; one family displayed paternal mutations resulting in a loss of a repeat unit at both DXS9902 & DXS7132 while the other family showed a maternal mutation resulting in a gain of a repeat unit at DXS7132 and a paternal mutation resulting in a loss of a repeat unit at DXS8378. Individually, these three markers had the highest mutation rates in this study and within the combined dataset. Though multiple mutational events within one family are rare, they should not be entirely unexpected. Based upon the upper confidence interval bounds of the two markers with the highest mutation rates (DXS9902 and DXS7132), two simultaneous mutations could be expected to occur approximately once in every 41,500 meioses.

Mutation rate and repeat structure

To examine whether the mutation rates for the 12 tetranucleotide (DXS6789, DXS7130, GATA31E08, GATA165B12, GATA172D05, DXS10147, DXS8378, DXS7132, DXS6803, HPRTB, DXS7423, and DXS9902) and 3 trinucleotide (DXS7424, DXS101, and DXS6795) X STR markers used in this study might vary according to repeat unit size, tetranucleotide and trinucleotide mutation rates were calculated (Table 9). Though the population-specific rates appeared similar for both types of repeats, the rates for the total U.S. and combined datasets showed values approaching significance (based on confidence interval bounds) that were higher for tetranucleotide repeats (1.09×10^{-3} and 1.19×10^{-3} respectively) compared with trinucleotide repeats (either no mutations observed at all, or 3.54×10^{-4} respectively). The relatively small number of markers with trinucleotide repeats (3) compared to those with tertranucleotide repeats (12) may begin to explain this difference, and further study of trinucleotide markers would be necessary to confirm this trend for X STRs.

Table 9. X STR mutation rate by repeat length and type. The 15 markers used to determine the overall mutation rate in this study were grouped either according to repeat length (tetra- or trinucleotide) or repeat type (simple or compound/complex), and corresponding mutation rates and 95% confidence intervals (CI) were calculated. Compound/complex repeat type included markers with microvariants. Markers with tetranucleotide repeat motifs included DXS6789, DXS7130, GATA31E08, GATA165B12, GATA172D05, DXS10147, DXS8378, DXS7132, DXS6803, HPRTB, DXS7423, and DXS9902. Markers with trinucleotide repeat motifs included DXS7424, DXS6795, and DXS101. Markers with simple repeat motifs were DXS7424, GATA165B12, GATA172D05, DXS10147, DXS8378, HPRTB, and DXS9902. Markers with compound/complex repeat motifs (including microvariants) were DXS6789, DXS7130, GATA31E08, DXS101, DXS6795, DXS7132, DXS6803, and DXS7423.

		African American				U.S. Caucasian			
Repeat	**Markers (N)**	**Mutations**	**Meioses**	**Mutation rate ($x10^{-3}$)**	**95% CI ($x10^{-3}$)**	**Mutations**	**Meioses**	**Mutation rate ($x10^{-3}$)**	**95% CI ($x10^{-3}$)**
Tetranucleotide	12	1	4680	0.21	0.01-1.2	9	6396	1.41	0.64-2.7
Trinucleotide	3	0	1170	0.00	0-3.1	0	1599	0.00	0-2.3
Simple	7	0	2730	0.00	0-1.3	6	3731	1.61	0.6-3.5
Compound/ Complex	8	1	3120	0.32	0.01-1.8	3	4264	0.70	0.15-2.0

Table continues on next page.

		U.S. Hispanic			
Repeat	**Markers (N)**	**Mutations**	**Meioses**	**Mutation rate ($x10^{-3}$)**	**95% CI ($x10^{-3}$)**
Tetranucleotide	12	8	5424	1.47	0.64-2.9
Trinucleotide	3	0	1356	0.00	0-2.7
Simple	7	5	3164	1.58	0.51-3.7
Compound/ Complex	8	3	3616	0.83	0.17-2.4

		This study (U.S. total)				Combined			
Repeat	**Markers (N)**	**Mutations**	**Meioses**	**Mutation rate ($x10^{-3}$)**	**95% CI ($x10^{-3}$)**	**Mutations**	**Meioses**	**Mutation rate ($x10^{-3}$)**	**95% CI ($x10^{-3}$)**
Tetranucleotide	12	18	16500	1.09	0.65-1.7	45	37753	1.19	0.87-1.6
Trinucleotide	3	0	4125	0.00	0-0.89	3	8464	0.35	0.07-1.0
Simple	7	11	9625	1.14	0.57-2.0	22	20385	1.08	0.68-1.6
Compound/ Complex	8	7	11000	0.64	0.26-1.3	26	25832	1.01	0.66-1.5

Separating the 15 markers into two groups based upon repeat structure ("simple" and "compound/complex" including microvariant), two sets of approximately equal numbers of markers and meioses were formed (Table 9). Mutation rates appeared slightly higher for simple repeats than for compound/complex repeats in the U.S. Caucasian, U.S. Hispanic, and total U.S. populations. However, the addition of data from published studies, considerably increasing the size of the dataset, revealed almost identical rates for both types of repeats: 1.08×10^{-3} (CI: 6.8×10^{-4}-1.6×10^{-3}) for simple repeats and 1.01×10^{-3} (CI: 6.6×10^{-4}-1.5×10^{-3}) for compound/complex (including microvariant) repeats.

Maternal versus paternal mutation rate

Paternal mutations outnumbered maternal mutations, resulting in different mutation rates depending upon allele origin (Table 10). Despite examination of more than 2.2 times as many maternal as paternal meioses, the maternal mutation rate remained comparable to the overall observed rate and almost an order of magnitude smaller than the paternal rate. This trend was consistent across populations and overall, with the paternal rate reaching almost three times the maternal rate in the U.S. Caucasian and U.S. Hispanic populations. The overall mutation rate for

maternal transfers was $4.22x10^{-4}$ while the mutation rate for paternal transfers was $1.71x10^{-3}$. Though the confidence intervals for these two values overlap by a small margin, both rates fell outside of the 95% CI for the overall mutation rate for all meioses ($8.73x10^{-4}$; Table 6).

Table 10. X STR mutation rate by origin of mutation. Maternal and paternal mutation rates for each of the studied populations are shown.

	African American				U.S. Caucasian			
Origin	**Mutations**	**Meioses**	**Mutation rate ($x10^{-3}$)**	**95% CI ($x10^{-3}$)**	**Mutations**	**Meioses**	**Mutation rate ($x10^{-3}$)**	**95% CI ($x10^{-3}$)**
Maternal	0	4095	0.00	0-0.9	3	5458	0.55	0.1-1.6
Paternal	1	1755	0.57	0.01-3.2	4	2533	1.58	0.4-4.0

	U.S. Hispanic				This study (U.S. total)			
Origin	**Mutations**	**Meioses**	**Mutation rate ($x10^{-3}$)**	**95% CI ($x10^{-3}$)**	**Mutations**	**Meioses**	**Mutation rate ($x10^{-3}$)**	**95% CI ($x10^{-3}$)**
Maternal	3	4649	0.65	0.1-1.9	6	14202	0.42	0.2-0.9
Paternal	4	2129	1.88	0.5-4.8	9	6417	1.40	0.6-2.7

Population database creation

Population genetic parameters

Allele frequencies and forensic efficiency parameters calculated for each of the 15 markers in three population groups are shown in Table 11. In total, 160 alleles were observed across 15 markers, with 7-17 alleles at each marker. Marker DXS101 was the marker with the highest number of observed alleles (17) in the populations studied; DXS8378, GATA165B12 and DXS10147 were the markers with the lowest number (7). DXS101 also exhibited the highest observed heterozygosity values within all three populations (0.9466, 0.8513, and 0.8629 in African Americans, U.S. Caucasians, and U.S. Hispanics respectively) while the lowest heterozygosity values varied by population (0.6845 at DXS8378 in African Americans; 0.6691 at DXS6795 in U.S. Caucasians; 0.6532 at DXS9902 in U.S. Hispanics). The usefulness of certain markers was strongly dependent upon the population to which they were applied. For example, DXS6795 exhibited the second highest observed heterozygosity value (0.8105) in the U.S. Hispanic population, the lowest value (0.6691) in the U.S. Caucasian population, and a value

(0.8107) at the midpoint of the 15 markers in the African American population. Overall, the forensic efficiency parameter values confirm the potential usefulness of these markers in certain specific kinship situations involving female offspring as well as identity testing.

Table 11. Allele frequencies and summary statistics for 15 X-chromosomal STR markers in 3 U.S. population groups. AA: African American, CN: U.S. Caucasian, Hisp: U.S. Hispanic, N: number of alleles, H(exp): expected heterozygosity, H(obs): observed heterozygosity, PIC: polymorphism information content, PDf: power of discrimination in females, PDm: power of discrimination in males, MECI: mean exclusion chance in trios involving daughter, MECII: mean exclusion chance in father/daughter duos, p (HWE): p value of the exact test for Hardy-Weinberg equilibrium.

	DXS6795			DXS9902			DXS8378			
	AA	**CN**	**Hisp**	**AA**	**CN**	**Hisp**	**AA**	**CN**	**Hisp**	
N	520	703	646	520	703	646	520	703	646	N
6										6
7										7
8				0.0615		0.0046		0.0085		8
8.3										8.3
9	0.1212	0.2859	0.1300	0.0692	0.0498	0.0108	0.0115	0.0185	0.0170	9
9.3										9.3
10	0.3038	0.0327	0.1068	0.3135	0.3172	0.3622	0.2635	0.3286	0.4443	10
10.3						0.0015				10.3
11	0.1577	0.4651	0.2415	0.3231	0.3613	0.4241	0.3712	0.3556	0.3313	11
11.1				0.0077	0.0327	0.0279				11.1
11.3										11.3
12	0.0827	0.0327	0.1548	0.2096	0.2205	0.1563	0.3115	0.2518	0.1811	12
12.1					0.0028	0.0031				12.1
12.3										12.3
13	0.0846	0.1792	0.3297	0.0135	0.0156	0.0062	0.0423	0.0356	0.0248	13
13.3										13.3
14	0.0442	0.0014	0.0217	0.0019		0.0031		0.0014	0.0015	14
14.3										14.3
15	0.1865	0.0028	0.0139							15
15.3										15.3
16	0.0135		0.0015							16
16.3										16.3
17	0.0058									17
PIC	0.7944	0.6118	0.7479	0.7019	0.6655	0.6005	0.6333	0.6424	0.5950	PIC
H(exp)	0.8172	0.6677	0.7800	0.7446	0.7164	0.6635	0.6938	0.7005	0.6591	H(exp)
H(obs)	0.8107	0.6691	0.8105	0.7621	0.7138	0.6532	0.6845	0.6952	0.6653	H(obs)
PDf	0.9438	0.8336	0.9194	0.8920	0.8687	0.8238	0.8457	0.8523	0.8197	PDf
PDm	0.8172	0.6677	0.7800	0.7446	0.7164	0.6635	0.6938	0.7005	0.6591	PDm
MECI	0.7944	0.6118	0.7479	0.7019	0.6655	0.6005	0.6333	0.6424	0.5950	MECI
MECII	0.6759	0.4652	0.6171	0.5635	0.5228	0.4547	0.4874	0.4974	0.4483	MECII
p(HWE)	0.7551	0.0707	0.3724	0.8238	0.0879	0.8353	0.9894	0.7500	0.9458	p(HWE)

Table continues on next page.

	GATA172D05			DXS7132			DXS6803			
	AA	**CN**	**Hisp**	**AA**	**CN**	**Hisp**	**AA**	**CN**	**Hisp**	
N	520	703	646	520	704	646	520	703	646	N
6	0.2038	0.1636	0.1192							6
7	0.0692		0.0031				0.0077			7
8	0.1750	0.1807	0.1455				0.0173			8
8.3										8.3
9	0.2923	0.0469	0.0588				0.0327		0.0031	9
9.3								0.0014		9.3
10	0.1269	0.2774	0.2817			0.0015	0.1058	0.0299	0.0170	10
10.3										10.3
11	0.1019	0.2134	0.2988	0.0212	0.0099	0.0015	0.3673	0.2418	0.2771	11
11.1										11.1
11.3							0.0058	0.0171	0.0588	11.3
12	0.0288	0.1181	0.0913	0.1038	0.0810	0.1022	0.2404	0.2646	0.3142	12
12.1										12.1
12.3							0.0827	0.1323	0.1656	12.3
13	0.0019		0.0015	0.2577	0.2827	0.2492	0.0788	0.1536	0.0666	13
13.3							0.0269	0.1366	0.0898	13.3
14				0.3269	0.3608	0.3514	0.0269	0.0043	0.0062	14
14.3							0.0038	0.0128	0.0015	14.3
15				0.2269	0.2045	0.2353	0.0038			15
15.3								0.0057		15.3
16				0.0423	0.0497	0.0387				16
16.3				0.0058		0.0015				16.3
17				0.0135	0.0099	0.0124				17
17.3						0.0015				17.3
18				0.0019	0.0014	0.0046				18
18.3										18.3
19										19
20										20
20.3										20.3
21										21
22										22
23										23
24										24
25										25
26										26
27										27
28										28
29										29
30										30
31										31
PIC	0.7847	0.7728	0.7526	0.7240	0.6958	0.7052	0.7535	0.7839	0.7492	PIC
H(exp)	0.8103	0.8020	0.7842	0.7620	0.7389	0.7469	0.7801	0.8104	0.7808	H(exp)
H(obs)	0.8447	0.7955	0.7500	0.8301	0.7621	0.6976	0.8204	0.7770	0.7782	H(obs)
PDf	0.9384	0.9316	0.9218	0.9053	0.8887	0.8942	0.9251	0.9376	0.9204	PDf
PDm	0.8103	0.8020	0.7842	0.7620	0.7389	0.7469	0.7801	0.8104	0.7808	PDm
MECI	0.7847	0.7728	0.7526	0.7240	0.6958	0.7052	0.7535	0.7839	0.7492	MECI
MECII	0.6628	0.6471	0.6231	0.5890	0.5565	0.5669	0.6260	0.6619	0.6196	MECII
p(HWE)	0.2128	0.6158	0.5084	0.2895	0.6342	0.2715	0.9786	0.5050	0.7068	p(HWE)

Table continues on next page.

	DXS6789			GATA165B12			DXS7130			
	AA	**CN**	**Hisp**	**AA**	**CN**	**Hisp**	**AA**	**CN**	**Hisp**	
N	520	703	646	520	703	646	520	703	646	N
6										6
7				0.0038						7
8				0.0404	0.0043	0.0077				8
8.3										8.3
9				0.1038	0.3030	0.2430	0.0038			9
9.3										9.3
10				0.3269	0.3428	0.4830	0.0269	0.0043	0.0046	10
10.3										10.3
11				0.4154	0.3172	0.2446	0.0673	0.0413	0.0588	11
11.1										11.1
11.3										11.3
12				0.1000	0.0327	0.0217	0.2154	0.1067	0.1966	12
12.1										12.1
12.3										12.3
13				0.0096			0.1673	0.0356	0.0542	13
13.3							0.0212	0.0640	0.0248	13.3
14	0.0038	0.0043	0.0093				0.0404	0.0057	0.0201	14
14.3							0.1615	0.2048	0.1687	14.3
15	0.2423	0.0384	0.0341				0.0038		0.0031	15
15.3							0.2288	0.3642	0.3560	15.3
16	0.1154	0.0142	0.0325							16
16.3							0.0558	0.1494	0.0882	16.3
17	0.0038	0.0014	0.0046				0.0019			17
17.3							0.0058	0.0228	0.0232	17.3
18	0.0135	0.0014	0.0031							18
18.3									0.0015	18.3
19	0.0692	0.0256	0.0341							19
20	0.1962	0.3912	0.3994							20
20.3								0.0014		20.3
21	0.2173	0.2717	0.3111							21
22	0.0981	0.1479	0.1099							22
23	0.0365	0.0868	0.0495							23
24	0.0038	0.0171	0.0124							24
25										25
26										26
27										27
28										28
29										29
30										30
31										31
PIC	0.8036	0.7035	0.6851	0.6483	0.6246	0.5844	0.8164	0.7575	0.7649	PIC
H(exp)	0.8263	0.7411	0.7255	0.6981	0.6890	0.6473	0.8367	0.7841	0.7904	H(exp)
H(obs)	0.8155	0.7286	0.6855	0.7136	0.7175	0.7097	0.8155	0.7881	0.7702	H(obs)
PDf	0.9471	0.8953	0.8843	0.8591	0.8389	0.8127	0.9530	0.9268	0.9306	PDf
PDm	0.8263	0.7411	0.7255	0.6981	0.6890	0.6473	0.8367	0.7841	0.7904	PDm
MECI	0.8036	0.7035	0.6851	0.6483	0.6246	0.5844	0.8164	0.7575	0.7649	MECI
MECII	0.6875	0.5661	0.5464	0.5049	0.4780	0.4363	0.7048	0.6300	0.6394	MECII
p(HWE)	0.1127	0.3011	0.0790	0.4856	0.1170	0.2379	0.8648	0.4266	**0.0229**	p(HWE)

Table continues on next page.

	DXS101			DXS7424			HPRTB			
	AA	**CN**	**Hisp**	**AA**	**CN**	**Hisp**	**AA**	**CN**	**Hisp**	
N	520	703	646	520	703	646	520	703	646	N
6										6
7										7
8								0.0014		8
8.3										8.3
9				0.0019			0.0385	0.0043		9
9.3										9.3
10				0.0077	0.0043	0.0046	0.0154	0.0057	0.0015	10
10.3										10.3
11				0.0808	0.0057	0.0046	0.0904	0.1294	0.0851	11
11.1										11.1
11.3										11.3
12				0.0731	0.0398	0.0464	0.2712	0.3272	0.2817	12
12.1										12.1
12.3										12.3
13				0.2250	0.0626	0.1084	0.2577	0.3229	0.3762	13
13.3										13.3
14				0.2750	0.2119	0.2121	0.2212	0.1394	0.1842	14
14.3										14.3
15		0.0242	0.0170	0.1692	0.2603	0.1950	0.0923	0.0569	0.0526	15
15.3										15.3
16	0.0019	0.0028	0.0015	0.1192	0.2703	0.2678	0.0135	0.0114	0.0186	16
16.3										16.3
17	0.0058	0.0043		0.0365	0.1238	0.1053		0.0014		17
17.3										17.3
18	0.0327	0.0612	0.0697	0.0096	0.0156	0.0511				18
18.3										18.3
19	0.1000	0.0569	0.0449	0.0019	0.0057	0.0031				19
20	0.0558	0.0171	0.0155			0.0015				20
20.3										20.3
21	0.1692	0.0270	0.0263							21
22	0.0635	0.0199	0.0124							22
23	0.0635	0.0669	0.0666							23
24	0.0923	0.2048	0.2399							24
25	0.0827	0.1821	0.1811							25
26	0.1038	0.1252	0.1904							26
27	0.1442	0.1010	0.0820							27
28	0.0423	0.0640	0.0402							28
29	0.0250	0.0256	0.0077							29
30	0.0135	0.0128	0.0015							30
31	0.0038	0.0043	0.0031							31
PIC	0.8911	0.8696	0.8363	0.7936	0.7625	0.7932	0.7619	0.7091	0.6921	PIC
H(exp)	0.8995	0.8806	0.8523	0.8176	0.7931	0.8176	0.7926	0.7491	0.7348	H(exp)
H(obs)	0.9466	0.8513	0.8629	0.8544	0.8253	0.8024	0.7767	0.7881	0.6895	H(obs)
PDf	0.9815	0.9748	0.9622	0.9428	0.9266	0.9424	0.9263	0.8971	0.8869	PDf
PDm	0.8995	0.8806	0.8523	0.8176	0.7931	0.8176	0.7926	0.7491	0.7348	PDm
MECI	0.8911	0.8696	0.8363	0.7936	0.7625	0.7932	0.7619	0.7091	0.6921	MECI
MECII	0.8112	0.7799	0.7326	0.6748	0.6354	0.6741	0.6346	0.5719	0.5523	MECII
p(HWE)	0.6232	0.3965	0.9997	0.8320	0.3977	0.1601	**0.0083**	**0.0393**	0.7030	p(HWE)

Table continues on next page.

	DXS7423			DXS10147			GATA31E08			
	AA	**CN**	**Hisp**	**AA**	**CN**	**Hisp**	**AA**	**CN**	**Hisp**	
N	520	703	646	520	703	646	520	703	646	N
6				0.1385	0.2376	0.3452	0.0019			6
7				0.2788	0.0413	0.0604	0.0212		0.0077	7
8	0.0077			0.3615	0.2959	0.3947	0.0250	0.0014	0.0062	8
8.3										8.3
9				0.1808	0.4139	0.1889	0.1635	0.1878	0.1130	9
9.3										9.3
10				0.0365	0.0114	0.0077	0.1596	0.0284	0.0325	10
10.3										10.3
11		0.0014		0.0019		0.0031	0.0635	0.2020	0.1811	11
11.1										11.1
11.3										11.3
12	0.0038		0.0015				0.2635	0.2119	0.3808	12
12.1										12.1
12.3										12.3
13	0.1115	0.1024	0.0325	0.0019			0.2135	0.2518	0.2121	13
13.3										13.3
14	0.4500	0.3001	0.3003				0.0692	0.0982	0.0557	14
14.3										14.3
15	0.3212	0.3997	0.4830				0.0192	0.0156	0.0108	15
15.3										15.3
16	0.0981	0.1664	0.0991					0.0014		16
16.3										16.3
17	0.0077	0.0284	0.0836					0.0014		17
17.3										17.3
18		0.0014								18
18.3										18.3
19										19
20										20
20.3										20.3
21										21
22										22
23										23
24										24
25										25
26										26
27										27
28										28
29										29
30										30
31										31
PIC	0.6165	0.6623	0.6049	0.6945	0.6225	0.6266	0.7994	0.7759	0.7268	PIC
H(exp)	0.6721	0.7112	0.6587	0.7384	0.6829	0.6856	0.8226	0.8049	0.7601	H(exp)
H(obs)	0.7184	0.6952	0.6976	0.7816	0.6691	0.7097	0.8107	0.8216	0.7258	H(obs)
PDf	0.8369	0.8677	0.8297	0.8877	0.8390	0.8421	0.9453	0.9329	0.9092	PDf
PDm	0.6721	0.7112	0.6587	0.7384	0.6829	0.6856	0.8226	0.8049	0.7601	PDm
MECI	0.6165	0.6623	0.6049	0.6945	0.6225	0.6266	0.7994	0.7759	0.7268	MECI
MECII	0.4709	0.5190	0.4587	0.5543	0.4760	0.4809	0.6822	0.6512	0.5925	MECII
p(HWE)	0.9471	0.1218	0.7420	0.0637	0.8893	0.5185	0.2097	0.6188	0.3914	p(HWE)

Rare and atypical alleles

Twelve alleles not previously observed in other U.S. populations were noted at 8 markers in this study (Table 12): 9.3, 15, and 15.3 at DXS6803; 17 and 20.2 at DXS7130; 17.3 at DXS7132; 11 and 18 at DXS7423; 10.3 at DXS9902; 13 at DXS10147; 7 at GATA165B12; and 17 at HPRTB. A subset of 6 of these alleles was sequenced to confirm repeat structure. At DXS6803, the sequenced alleles 9.3 and 15.3 conformed to the published repeat structure [22,115] and were observed only in the U.S. Caucasian population. Previously, the 9.3 allele has been observed in Han population of China [41,116] while the 15.3 allele was observed in populations from Croatia [117] and Japan [118]. The 15 allele at DXS6803 had not been previously reported at the time of publication in any world-wide population, and was observed here in two African American samples. Both the 17 and 20.2 alleles newly observed at DXS7130 in U.S. populations exhibited novel repeat structures. The typical repeat motif in non-microvariant DXS7130 alleles is TATC [22,115], but the 17 allele included both this standard TATC as well as one AATC. This allele was observed only once in the African American population, but had been previously observed in populations from the Brazilian Amazon [20], Japanese immigrants residing in Brazil [119], and northwestern China [120]. While microvariant alleles at DXS7130 typically exhibit a $(TATC)_5$-ATC-$(TATC)_x$ pattern [22,115], the 20.2 allele contained an additional partial repeat (ATC). The 20.2 allele was not previously reported at the time of publication, and was observed here in a single U.S. Caucasian sample. Both the 17.3 allele at DXS7132 and the 11 and 18 alleles at DXS7423 were observed previously in non-U.S. populations. Observed here in a single U.S. Hispanic sample, the 17.3 allele at DXS7132 had been observed one time each in populations from Northern Portugal [121], Galicia (Spain) [121], and Nicaragua [122]. The presence of both the 11 and 18 alleles at DXS7423 had been reported in many global populations including those from Europe [121,123,124], Asia [125,126], northern Africa [127], and South America [121,128,129], though each was observed in only one U.S. Caucasian sample in this study. The 10.3 allele at DXS9902 and 13 allele at DXS10147 had not previously been observed in any population at the time of publication. Both the 7 allele at GATA165B12 and 17 allele at HPRTB were consistent with published simple repeat structures [89,130,131] and were observed for the first time in U.S. populations in the U.S. Hispanic population. Previously, the 7 allele at

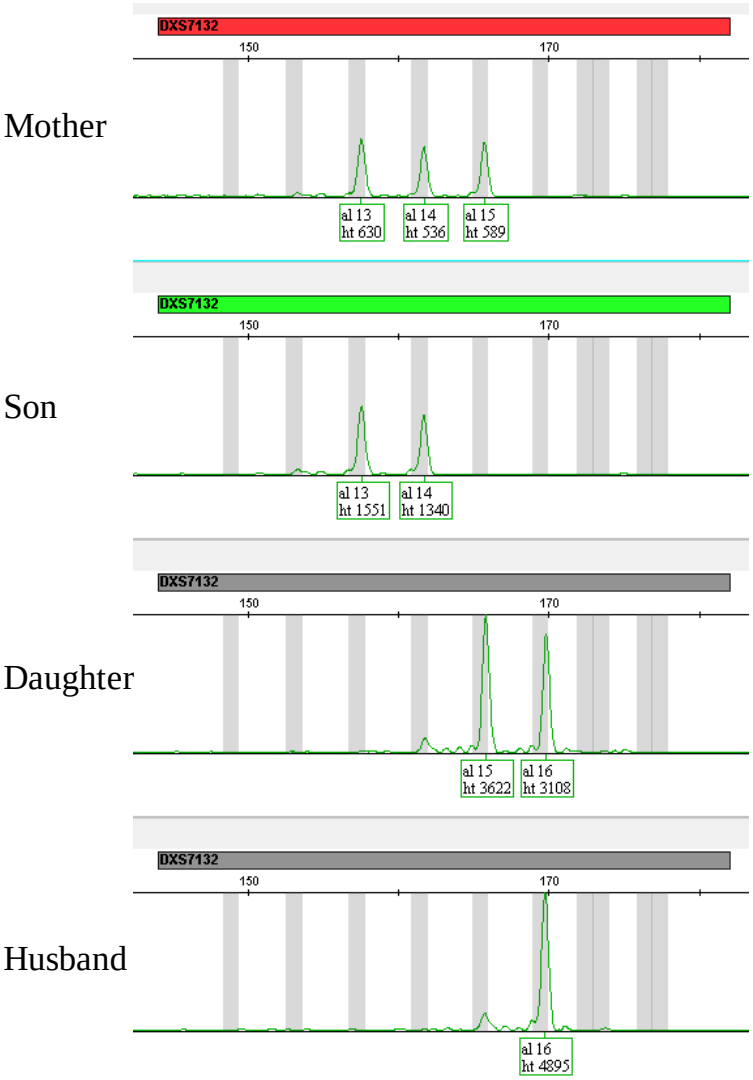
GATA165B12 had been noted in several Chinese populations [49,126,132] while the 17 allele at HPRTB had been widely detected throughout the world.

Table 12. Novel alleles observed in U.S. populations. Twelve alleles never previously observed in U.S. populations at the time of publication were noted in this study. Six of these alleles were sequenced and the repeat structure is presented below. N: number of times allele observed in population.

Marker	Allele	Repeat structure (if sequenced)	Population	N
DXS6803	9.3	$(TCTA)_8$-TCA-TCTA	U.S. Caucasian	1
	15	*not sequenced*	African American	2
	15.3	$(TCTA)_{14}$-TCA-(TCTA)	U.S. Caucasian	2
DXS7130	17	$(TATC)_4$-AATC-$(TATC)_{12}$	African American	1
	20.2	$(TATC)_5$-ATC-$(TATC)_4$-ATC-$(TATC)_{10}$	U.S. Caucasian	1
DXS7132	17.3	*not sequenced*	U.S. Hispanic	1
DXS7423	11	*not sequenced*	U.S. Caucasian	1
	18	*not sequenced*	U.S. Caucasian	1
DXS9902	10.3	*not sequenced*	U.S. Hispanic	1
DXS10147	13	*not sequenced*	African American	1
GATA165B12	7	$(AGAT)_7$	African American	1
HPRTB	17	$(ATCT)_{17}$	African American	1

A triallelic pattern was observed in one U.S. Caucasian sample at marker DXS7132 (Figure 2). Because the three alleles were of even signal intensity, this example was characterized as a Type 2 pattern, which was found to be less frequent than Type 1 triallelic patterns that consist of alleles of unequal signal strength [133,134]. Ancillary testing of this individual's husband, daughter and son revealed transmission of the 13 and 14 allele as a unit from the mother to the son, while the 15 allele was transmitted alone from the mother to the daughter. This inheritance pattern supports the hypothesis that Type 2 triallelic patterns are due to localized duplication events affecting the individual's germ cells [133] rather than X chromosomal aneuploidy, which in this case would be accompanied by the observation of an additional allele at the other 14 markers.

Figure 2. Observed triallelic pattern at DXS7132. A single U.S. Caucasian sample (mother) in the studied populations revealed a Type 2 triallelic pattern. Ancillary testing of additional family members (son, daughter, husband) revealed an inheritance pattern consistent with localized duplication of the region containing the STR repeat. The 13 and 14 alleles were passed as a unit from the mother to her son, whose profile would typically exhibit only one X STR allele per marker. The 15 allele was passed separately to the daughter, whose 16 allele was inherited from her father (husband).



Null alleles

In addition to new or rare alleles, several instances of primer binding site mutations resulting in null alleles were observed in this dataset at 7 markers: DXS101, DXS6795, DXS6803,

DXS7130, DXS7132, GATA172D05, and HPRTB (Table 13). One null allele was observed at marker DXS101 in a U.S. Caucasian sample at a frequency of 0.0014. Sequencing revealed a C→T transition located 6 base pairs (bp) from the 3' end of the reverse primer binding site which resulted in complete suppression of amplification. Two null 11 alleles at DXS6803 were found to be caused by a C→T transition under the reverse primer binding site located 15 bp from the 3' end. In both cases, residual amplification product could be observed under the detection threshold, indicating partial binding of the primer could still occur. This particular null allele was present only in the U.S. Hispanic population (at a frequency of 0.0031), suggesting a possible population-specificity to this particular polymorphism associated with an 11 allele. At DXS6795, a T→C transition 15 bp from the 3' end of the forward primer resulted in partially null alleles 6 times in the African American population (frequency of 0.0115) and 2 times in the U.S. Hispanic population (frequency of 0.0031). An additional null allele at DXS6795 was observed in one U.S. Hispanic sample due to an A→G transition 2 bp from the 3' end of the forward primer binding site. Six instances of reduced amplification efficiency resulting in a partially null allele were observed at marker DXS7130 due to a G→A transition in the reverse primer binding site 11 bp from the 3' end. This polymorphism was observed in combination with a 12, 13, or 14 allele with an overall frequency of 0.0115 in the African American population, making this base change more common than certain rare alleles at this locus (9, 15, 17, and 17.3). Again, the possible population-specificity of the polymorphism was highlighted by its presence within the African American population only. One U.S. Hispanic sample exhibited a partially null 17 allele at DXS7132 due to a C→A transversion at the 5'-most end (20 bp from the 3' end) of the reverse primer binding site; this was the only observed transversion. At GATA172D05, a null allele was observed at one U.S. Caucasian sample resulting from a G→A transition 7 bp from the 3' end of the reverse primer binding site. This mutation has been described previously in the U.S. Hispanic population [18,22], but this is the first instance of its observation in a U.S. Caucasian. Lastly, two partially null 12 alleles at HPRTB were observed in the U.S. Hispanic population. A previously noted AG deletion in the flanking region [18,135] fell under the reverse primer binding site used in this study 14 and 15 bp from the 3' end. Since many of these primer sets are shared with other published X STR assays, knowledge of the frequency of observed null alleles and the populations in which they occur can aid in

interpretation. The frequency and diversity of these polymorphisms suggest the additional discriminatory power that may be available should routine sequencing of STR repeats become feasible for forensic laboratories in the future.

Table 13. Null alleles observed in U.S. populations. Ten null alleles observed in three U.S. populations are the result of primer binding site mutations. N: number of times null allele observed in population. *Position refers to the number of base pairs from the 3' end of the indicated primer.

Marker	Suppression of amplification	Base change	Position*	Primer orientation	Allele	Population	N
DXS101	Complete	C→T	6	Reverse	23	U.S. Caucasian	1
DXS6795	Partial	T→C	15	Forward	10	African American	6
						U.S. Hispanic	2
	Complete	A→G	2	Forward	11	U.S. Hispanic	1
DXS6803	Partial	C→T	15	Reverse	11	U.S. Hispanic	2
DXS7130	Partial	G→A	11	Reverse	12	African American	2
					13	African American	3
					14	African American	1
DXS7132	Partial	C→A	20	Reverse	17	U.S. Hispanic	1
GATA172D05	Complete	G→A	7	Reverse	12	U.S. Caucasian	1
HPRTB	Partial	AGdel	14-15	Reverse	12	U.S. Hispanic	2

Hardy-Weinberg equilibrium

Departures from Hardy-Weinberg equilibrium (indicated by a p-value for the exact test that is less than 0.05; shown in bold in Table 11) occurred in 3 of the 45 marker-population combinations: HPRTB in the African American population ($p = 0.0084$) and U.S. Caucasian population ($p = 0.0420$), and DXS7130 in the U.S. Hispanic population ($p = 0.0229$). After application of the Bonferroni correction, however, none of these values remain significant ($p < 0.0033$).

Comparisons to published populations

Previous pairwise population comparisons with X STRs and U.S. populations revealed that the individual groups cannot be combined into one pooled database for forensic use and must instead

be treated as three distinct databases [18,22]. Pairwise F_{ST} value comparisons for populations from sample set C revealed that all populations were significantly different from one another (indicated by a corresponding p value < 0.05) at 13 of the 15 markers (Table 14). The U.S. Caucasian and U.S. Hispanic populations did not differ significantly at marker DXS6789, however, and all three populations were similar to one another at marker DXS7132. Given these results, the three populations were determined to be disctinct with respect to the multiplexes applied here, and were maintained separately in this study.

Table 14. Pairwise F_{ST} values (below diagonal) and corresponding p values with standard deviations (above diagonal) comparing populations from sample set C by marker. P values indicative of populations that are not significantly different from one another ($p > 0.5$) are bolded. AA: African American; CN: U.S. Caucasian, Hisp: U.S. Hispanic.

Marker	Population	AA	CN	Hisp
DXS101	AA	*	$<10^{-5}$	$<10^{-5}$
	CN	0.02692	*	0.00822±0.0009
	Hisp	0.03828	0.00266	*
DXS6789	AA	*	$<10^{-5}$	$<10^{-5}$
	CN	0.05889	*	**0.06405±0.0024**
	Hisp	0.06051	0.00187	*
DXS6795	AA	*	$<10^{-5}$	$<10^{-5}$
	CN	0.14077	*	$<10^{-5}$
	Hisp	0.08007	0.07417	*
DXS6803	AA	*	$<10^{-5}$	$<10^{-5}$
	CN	0.02543	*	$<10^{-5}$
	Hisp	0.02143	0.00889	*
DXS7130	AA	*	$<10^{-5}$	$<10^{-5}$
	CN	0.03579	*	0.00010±0.0001
	Hisp	0.0175	0.00834	*
DXS7132	AA	*	**0.28740±0.0053**	**0.76349±0.0043**
	CN	0.00034	*	**0.20325±0.0042**
	Hisp	-0.00095	0.00071	*
DXS7423	AA	*	$<10^{-5}$	$<10^{-5}$
	CN	0.02218	*	$<10^{-5}$
	Hisp	0.04198	0.01248	*
DXS7424	AA	*	$<10^{-5}$	$<10^{-5}$
	CN	0.04349	*	0.00713±0.0009
	Hisp	0.02987	0.00347	*
DXS8378	AA	*	0.02198±0.0016	$<10^{-5}$
	CN	0.00419	*	0.00010±0.0001
	Hisp	0.03518	0.01241	*
DXS9902	AA	*	0.04762±0.0022	$<10^{-5}$
	CN	0.00271	*	0.00218±0.0005
	Hisp	0.01415	0.00696	*
DXS10147	AA	*	$<10^{-5}$	$<10^{-5}$
	CN	0.0801	*	$<10^{-5}$
	Hisp	0.05958	0.04885	*
GATA31E08	AA	*	$<10^{-5}$	$<10^{-5}$
	CN	0.02413	*	$<10^{-5}$
	Hisp	0.02734	0.02225	*
GATA165B12	AA	*	$<10^{-5}$	$<10^{-5}$
	CN	0.03689	*	$<10^{-5}$
	Hisp	0.05484	0.0195	*
GATA172D05	AA	*	$<10^{-5}$	$<10^{-5}$
	CN	0.06217	*	0.00149±0.0004
	Hisp	0.07581	0.00565	*
HPRTB	AA	*	0.00020±0.0001	$<10^{-5}$
	CN	0.01001	*	0.00703±0.0009
	Hisp	0.0105	0.00453	*

Similarity to other published populations was determined using pairwise F_{ST} comparisons and corresponding p values (Table 15). Though few studies addressed the same populations investigated here, and none covered all 15 markers, a subset of seven markers (DXS8378, HPRTB, GATA172D05, DXS7423, DXS7132, DXS101, and DXS6789) overlapped with this study for the African American and U.S. Hispanic populations [18]. In general, equivalent populations were found to be similar to one another at overlapping markers with the exception of the U.S. Hispanic populations at markers DXS7423, DXS8378, and GATA172D05. For two of these markers (DXS7423 and DXS8378), the U.S. Hispanic population from Gomez, *et al.* [18] was more similar to the African American or U.S. Caucasian populations from this study. Other unexpected associations involving the U.S. Hispanic populations were observed, underscoring the variability of self-described U.S. Hispanic populations.

Table 15. P values (with standard deviations) corresponding to pairwise F_{ST} values (not shown) comparing studied U.S. populations (top) to published populations (left) at overlapping markers. P values indicative of populations that are not significantly different from one another ($p > 0.5$) are bolded. Values at the intersection of equivalent populations are shaded in gray. AA: African American; CN: U.S. Caucasian, Hisp: U.S. Hispanic.

Marker	Ref.	Population	AA	CN	Hisp
DXS101	[18]	AA	**0.13523±0.0036**	$<10^{-5}$	$<10^{-5}$
		Hisp	$<10^{-5}$	**0.56509±0.0043**	**0.18701±0.0033**
	[136]	AA	**0.12048±0.0033**	$<10^{-5}$	$<10^{-5}$
		CN	$<10^{-5}$	**0.09197±0.0027**	**0.25047±0.0042**
DXS6789	[18]	AA	**0.38303±0.0049**	$<10^{-5}$	$<10^{-5}$
		Hisp	$<10^{-5}$	**0.06118±0.0024**	**0.05702±0.0022**
DXS7132	[18]	AA	**0.57321±0.0050**	**0.87556±0.0036**	**0.44451±0.0053**
		Hisp	**0.13276±0.0031**	**0.06178±0.0025**	**0.16276±0.0038**
DXS7423	[18]	AA	**0.12979±0.0032**	$<10^{-5}$	$<10^{-5}$
		Hisp	**0.11019±0.0031**	0.03435±0.0020	0.01406±0.0010
DXS8378	[18]	AA	**0.45728±0.0039**	0.01921±0.0012	$<10^{-5}$
		Hisp	**0.07940±0.0026**	**0.27265±0.0048**	0.01406±0.0011
GATA172D05	[18]	AA	**0.40263±0.0043**	$<10^{-5}$	$<10^{-5}$
		Hisp	0.00010±0.0001	0.00673±0.0008	0.00050±0.0002
HPRTB	[18]	AA	**0.73240±0.0044**	0.01436±0.0011	0.00653±0.0008
		Hisp	**0.45322±0.0042**	**0.24275±0.0046**	**0.49203±0.0048**
	[137]	AA	**0.43550±0.0051**	0.00050±0.0002	0.02812±0.0018
		CN	0.00139±0.0004	**0.68498±0.0048**	0.00347±0.0006
		Hisp	0.00990±0.0011	0.04594±0.0019	**0.75171±0.0040**

The original publication describing the development of the multiplexes used here to amplify the 15 X STR markers included a population study of 349 African Americans, 268 U.S. Caucasians, and 245 U.S. Hispanics [22]. Of note, no significant differences between the three populations present in both studies were observed. Therefore, these databases could be combined to create a larger single database for each group (Table 16). When combined, the U.S. populations exhibit similar forensic efficiency statistics with none of the marker-population combinations resulting in a significant deviation from Hardy-Weinberg equilibrium after the Bonferroni correction ($p < 0.0033$).

Table 16. Combined allele frequencies and summary statistics for 15 X STR markers in 3 U.S. population groups from two studies. AA: African American, CN: U.S. Caucasian, Hisp: U.S. Hispanic, N: number of alleles, H(exp): expected heterozygosity, H(obs): observed heterozygosity, PIC: polymorphism information content, PDf: power of discrimination in females, PDm: power of discrimination in males, MECI: mean exclusion chance in trios involving daughter, MECII: mean exclusion chance in father/daughter duos, p (HWE): p value of the exact test for Hardy-Weinberg equilibrium.

	DXS6795			DXS9902			DXS8378			
	AA	**CN**	**Hisp**	**AA**	**CN**	**Hisp**	**AA**	**CN**	**Hisp**	
N	1043	1117	1013	1043	1117	1013	1043	1117	1013	N
5										5
6	0.0010									6
7				0.0010						7
8	0.0010			0.0470		0.0039	0.0048	0.0063	0.0010	8
8.3										8.3
9	0.1189	0.2919	0.1491	0.0719	0.0439	0.0168	0.0115	0.0143	0.0138	9
9.3										9.3
10	0.2857	0.0269	0.1076	0.2991	0.3241	0.3662	0.2723	0.3384	0.4215	10
10.1						0.0020				10.1
10.3						0.0010				10.3
11	0.1764	0.4673	0.2349	0.3461	0.3706	0.3978	0.3653	0.3518	0.3258	11
11.1				0.0086	0.0286	0.0296				11.1
11.3										11.3
12	0.0853	0.0313	0.1520	0.2167	0.2184	0.1728	0.3087	0.2543	0.2132	12
12.1					0.0018	0.0039				12.1
12.3										12.3
13	0.0911	0.1791	0.3189	0.0086	0.0125	0.0039	0.0345	0.0322	0.0227	13
13.3										13.3
14	0.0479	0.0009	0.0207	0.0010		0.0020	0.0029	0.0027	0.0020	14
14.3										14.3
15	0.1783	0.0027	0.0158							15
15.3										15.3
16	0.0105		0.0010							16
16.3										16.3
17	0.0038									17
PIC	0.8011	0.6048	0.7541	0.6913	0.6533	0.6159	0.6356	0.6358	0.6060	PIC
H(exp)	0.8233	0.6627	0.7855	0.7363	0.7070	0.6766	0.6958	0.6958	0.6700	H(exp)
H(obs)	0.8263	0.6337	0.8135	0.7316	0.6892	0.6838	0.6605	0.7084	0.6838	H(obs)
PDf	0.9465	0.8284	0.9226	0.8855	0.8605	0.8347	0.8473	0.8475	0.8271	PDf
PDm	0.8233	0.6627	0.7855	0.7363	0.7070	0.6766	0.6958	0.6958	0.6700	PDm
MECI	0.8011	0.6048	0.7541	0.6913	0.6533	0.6159	0.6356	0.6358	0.6060	MECI
MECII	0.6843	0.4580	0.6245	0.5514	0.5095	0.4704	0.4899	0.4903	0.4592	MECII
p(HWE)	0.8655	0.0984	0.2196	0.6772	**0.0315**	0.5959	0.7691	0.7213	0.7444	p(HWE)

Table continues on next page.

	GATA172D05			DXS7132			DXS6803			
	AA	**CN**	**Hisp**	**AA**	**CN**	**Hisp**	**AA**	**CN**	**Hisp**	
N	1043	1117	1013	1043	1118	1013	1043	1117	1013	N
5										5
6	0.1946	0.1692	0.1165							6
7	0.0537	0.0009	0.0049				0.0048		0.0020	7
8	0.1764	0.1737	0.1412				0.0134			8
8.3										8.3
9	0.2848	0.0483	0.0632				0.0288	0.0009	0.0020	9
9.3								0.0009		9.3
10	0.1438	0.2847	0.2932			0.0010	0.1246	0.0349	0.0197	10
10.1										10.1
10.3									0.0010	10.3
11	0.1055	0.2095	0.2813	0.0163	0.0116	0.0030	0.3672	0.2498	0.2695	11
11.1										11.1
11.3							0.0077	0.0143	0.0582	11.3
12	0.0403	0.1110	0.0977	0.1026	0.0868	0.0948	0.2368	0.2641	0.3268	12
12.1										12.1
12.3							0.0757	0.1235	0.1530	12.3
13	0.0010	0.0027	0.0020	0.2579	0.2925	0.2586	0.0796	0.1441	0.0652	13
13.3							0.0316	0.1432	0.0977	13.3
14				0.3490	0.3605	0.3406	0.0249	0.0063	0.0039	14
14.3							0.0019	0.0143	0.0010	14.3
15				0.2224	0.1852	0.2409	0.0019			15
15.3								0.0036		15.3
16				0.0364	0.0546	0.0444	0.0010			16
16.3				0.0038		0.0010				16.3
17				0.0086	0.0072	0.0118				17
17.3						0.0010				17.3
18				0.0029	0.0018	0.0030				18
PIC	0.7884	0.7726	0.7572	0.7093	0.6971	0.7058	0.7597	0.7833	0.7483	PIC
H(exp)	0.8136	0.8016	0.7878	0.7500	0.7395	0.7480	0.7845	0.8096	0.7796	H(exp)
H(obs)	0.8500	0.7807	0.7676	0.8026	0.7735	0.7081	0.7974	0.7904	0.773	H(obs)
PDf	0.9401	0.9317	0.9243	0.8968	0.8898	0.8943	0.9287	0.9374	0.9202	PDf
PDm	0.8136	0.8016	0.7878	0.7500	0.7395	0.7480	0.7845	0.8096	0.7796	PDm
MECI	0.7884	0.7726	0.7572	0.7093	0.6971	0.7058	0.7597	0.7833	0.7483	MECI
MECII	0.6675	0.6470	0.6287	0.5718	0.5581	0.5677	0.6343	0.6612	0.6186	MECII
p(HWE)	0.2151	0.1429	0.8982	0.3650	0.6347	0.2813	0.9855	0.6405	0.3303	p(HWE)

Table continues on next page.

	DXS6789			GATA165B12			DXS7130			
	AA	**CN**	**Hisp**	**AA**	**CN**	**Hisp**	**AA**	**CN**	**Hisp**	
N	1043	1117	1013	1043	1117	1013	1043	1117	1013	N
5										5
6										6
7				0.0019						7
8				0.0393	0.0072	0.0099				8
8.3										8.3
9				0.1371	0.3071	0.2438	0.0058	0.0009		9
9.3										9.3
10				0.3289	0.3321	0.4709	0.0249	0.0027	0.0039	10
10.1										10.1
10.3										10.3
11				0.3873	0.3232	0.2488	0.0719	0.0394	0.0582	11
11.1										11.1
11.3										11.3
12				0.0988	0.0304	0.0267	0.2042	0.1047	0.2024	12
12.1										12.1
12.3										12.3
13				0.0067			0.1582	0.0439	0.0632	13
13.3							0.0221	0.0564	0.0227	13.3
14	0.0048	0.0036	0.0079				0.0422	0.0098	0.0188	14
14.3							0.1755	0.2122	0.1885	14.3
15	0.2272	0.0421	0.0434				0.0019		0.0020	15
15.3							0.2253	0.3554	0.3475	15.3
16	0.1151	0.0161	0.0434				0.0010			16
16.3							0.0585	0.1477	0.0711	16.3
17	0.0077	0.0027	0.0049				0.0010			17
17.3							0.0067	0.0260	0.0207	17.3
18	0.0153	0.0009	0.0030							18
18.3							0.0010		0.0010	18.3
19	0.0604	0.0242	0.0346							19
20	0.2013	0.3787	0.3998							20
20.3								0.0009		20.3
21	0.2119	0.2892	0.2912							21
22	0.1035	0.1459	0.1135							22
23	0.0460	0.0806	0.0464							23
24	0.0058	0.0161	0.0109							24
25	0.0010		0.0010							25
PIC	0.8116	0.7040	0.6979	0.6632	0.6258	0.5944	0.8206	0.7623	0.7620	PIC
H(exp)	0.8329	0.7423	0.7351	0.7117	0.6900	0.6561	0.8402	0.7884	0.7890	H(exp)
H(obs)	0.8421	0.7133	0.6973	0.7158	0.7012	0.6784	0.8158	0.7807	0.7568	H(obs)
PDf	0.9508	0.8953	0.8926	0.8684	0.8397	0.8200	0.9549	0.9291	0.9285	PDf
PDm	0.8329	0.7423	0.7351	0.7117	0.6900	0.6561	0.8402	0.7884	0.7890	PDm
MECI	0.8116	0.7040	0.6979	0.6632	0.6258	0.5944	0.8206	0.7623	0.7620	MECI
MECII	0.6981	0.5669	0.5607	0.5206	0.4793	0.4466	0.7104	0.6360	0.6358	MECII
p(HWE)	0.1885	0.0993	0.4868	0.8493	0.3855	0.9952	0.5784	0.3077	**0.0463**	p(HWE)

Table continues on next page.

	DXS101			DXS7424			HPRTB			
	AA	**CN**	**Hisp**	**AA**	**CN**	**Hisp**	**AA**	**CN**	**Hisp**	
N	1043	1117	1013	1043	1117	1013	1043	1117	1013	N
7								0.0009		7
8							0.0010	0.0009		8
8.3										8.3
9				0.0019	0.0009		0.0384	0.0036	0.0010	9
9.3										9.3
10				0.0096	0.0045	0.0059	0.0144	0.0045	0.0049	10
10.1										10.1
10.3										10.3
11				0.0700	0.0054	0.0039	0.0940	0.1244	0.0809	11
11.1										11.1
11.3										11.3
12				0.0642	0.0421	0.0365	0.2838	0.3321	0.2725	12
12.1										12.1
12.3										12.3
13				0.2205	0.0609	0.1145	0.2733	0.3214	0.3830	13
13.3										13.3
14	0.0010			0.2483	0.2077	0.2004	0.1975	0.1477	0.1895	14
14.3										14.3
15	0.0019	0.0304	0.0138	0.1812	0.2578	0.2093	0.0825	0.0528	0.0513	15
15.3										15.3
16	0.0029	0.0018	0.0010	0.1457	0.2614	0.2715	0.0153	0.0107	0.0168	16
16.3										16.3
17	0.0038	0.0045		0.0431	0.1289	0.1115		0.0009		17
17.3										17.3
18	0.0479	0.0743	0.0652	0.0125	0.0242	0.0415				18
18.3										18.3
19	0.0872	0.0528	0.0385	0.0019	0.0054	0.0039				19
20	0.0652	0.0197	0.0188	0.0010	0.0009	0.0010				20
20.3										20.3
21	0.1419	0.0278	0.0286							21
22	0.0652	0.0143	0.0217							22
23	0.0709	0.0671	0.0642							23
24	0.0978	0.1996	0.2438							24
25	0.0901	0.1817	0.1728							25
26	0.1007	0.1334	0.1984							26
27	0.1304	0.0877	0.0790							27
28	0.0518	0.0591	0.0365							28
29	0.0240	0.0269	0.0099							29
30	0.0115	0.0152	0.0049							30
31	0.0038	0.0036	0.0020							31
33	0.0019		0.0010							33
PIC	0.9008	0.8713	0.8361	0.8015	0.7703	0.7882	0.7571	0.7053	0.6911	PIC
H(exp)	0.9082	0.8821	0.8520	0.8245	0.7993	0.8137	0.7882	0.7462	0.7337	H(exp)
H(obs)	0.9263	0.8506	0.8595	0.8500	0.8361	0.8000	0.7842	0.7976	0.7216	H(obs)
PDf	0.9842	0.9753	0.9622	0.9462	0.9307	0.9399	0.9240	0.8947	0.8865	PDf
PDm	0.9082	0.8821	0.8520	0.8245	0.7993	0.8137	0.7882	0.7462	0.7337	PDm
MECI	0.9008	0.8713	0.8361	0.8015	0.7703	0.7882	0.7571	0.7053	0.6911	MECI
MECII	0.8259	0.7823	0.7326	0.6848	0.6450	0.6676	0.6288	0.5675	0.5512	MECII
p(HWE)	0.3703	**0.0498**	0.9489	0.6025	0.1476	**0.0377**	0.6834	**0.0403**	0.7875	p(HWE)

Table continues on next page.

	DXS7423			DXS10147			GATA31E08			
	AA	**CN**	**Hisp**	**AA**	**CN**	**Hisp**	**AA**	**CN**	**Hisp**	
N	1043	1117	1013	1043	1117	1013	1043	1117	1013	N
5				0.0010						5
6				0.1342	0.2310	0.344	0.0010			6
7				0.2800	0.0403	0.060	0.0211		0.0049	7
8	0.0077			0.3624	0.2972	0.414	0.0268	0.0018	0.0049	8
8.3										8.3
9				0.1774	0.4136	0.171	0.1515	0.1808	0.1362	9
9.3										9.3
10				0.0403	0.0161	0.010	0.1534	0.0251	0.0523	10
10.1										10.1
10.3										10.3
11		0.0009		0.0038	0.0018	0.002	0.0623	0.2095	0.1807	11
11.1										11.1
11.3										11.3
12	0.0048		0.0010				0.2560	0.2095	0.3583	12
12.1										12.1
12.3										12.3
13	0.0968	0.1003	0.0316	0.0010			0.2378	0.2551	0.2024	13
13.3										13.3
14	0.4660	0.3187	0.3001				0.0757	0.1030	0.0474	14
14.3										14.3
15	0.3241	0.3930	0.4817				0.0134	0.0134	0.0128	15
15.3										15.3
16	0.0872	0.1513	0.0977				0.0010	0.0009		16
16.3										16.3
17	0.0134	0.0349	0.0879					0.0009		17
17.3										17.3
18		0.0009								18
18.3										18.3
19										19
20										20
20.3										20.3
21										21
22										22
23										23
24										24
25										25
26										26
27										27
28										28
29										29
30										30
31										31
33										33
PIC	0.6037	0.6604	0.6059	0.6958	0.6263	0.6184	0.7969	0.7735	0.7429	PIC
H(exp)	0.6606	0.7098	0.6596	0.7391	0.6854	0.6780	0.8205	0.8030	0.7743	H(exp)
H(obs)	0.7053	0.6867	0.6892	0.7474	0.6578	0.6919	0.8263	0.8241	0.7703	H(obs)
PDf	0.8279	0.8664	0.8304	0.8886	0.8419	0.8367	0.9442	0.9317	0.9177	PDf
PDm	0.6606	0.7098	0.6596	0.7391	0.6854	0.6780	0.8205	0.8030	0.7743	PDm
MECI	0.6037	0.6604	0.6059	0.6958	0.6263	0.6184	0.7969	0.7735	0.7429	MECI
MECII	0.4579	0.5172	0.4597	0.5559	0.4801	0.4726	0.6790	0.6481	0.6115	MECII
p(HWE)	0.9073	0.0936	0.8123	0.1265	0.3951	0.3083	0.3373	0.5504	0.8781	p(HWE)

Linkage disequilibrium

A test of multi-locus linkage disequilibrium in both the population described in this study as well as the combined population revealed three marker pairs with significant p values after the Bonferroni correction ($p < 0.0011$) in both populations (Table 17). One of these pairs (DXS6789-GATA165B12) was a set of markers that is not adjacent on the X chromosome, and therefore the association may instead be due to sampling effects. However, the pairs of DXS7423-DXS10147 and DXS10147-GATA31E08 both had p values equal to zero in both populations and are adjacent on the chromosome within the originally proposed 4th linkage group [8,30]. These results confirmed a previous study that observed linkage disequilibrium between markers DXS10147 and DXS7423 [105] and indicated that these markers should likely be considered as a haplotype for statistical purposes. Nine additional significant associations (after Bonferroni correction) were noted in the combined population; two of these were between adjacent markers: DXS6803-DXS6789, which are within ~9 Mb of each other, and DXS101-DXS7424, for which previously reported linkage disequilibrium exists [27,103]. For reference, haplotype frequencies for markers included in each of the originally proposed linkage groups 1, 2, and 4 are presented in Appendix B (linkage group 3 is represented by HPRTB alone in this set of markers).

Table 17. Multi-locus exact test for linkage disequilibrium results for original and combined populations. P values for marker pairs with significant p values ($p < 0.05$) in at least one of the populations are shown. Bold values are those that remain significant after the Bonferroni correction ($p < 0.0011$).

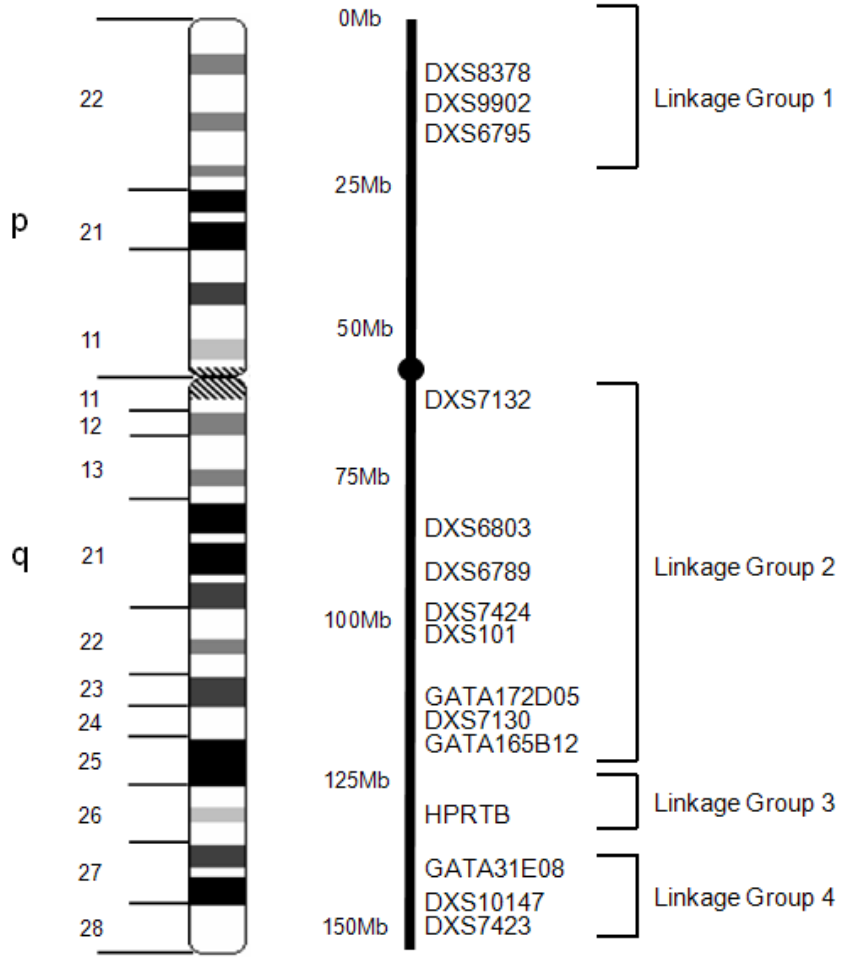
Locus combination	Sample set C	Combined population
DXS6795-DXS9902	0.0510	0.0030
DXS6795-DXS8378	0.0510	0.0160
DXS9902-GATA172D05	0.0030	**$<10^{-4}$**
DXS8378-GATA172D05	0.0510	0.0140
GATA172D05-DXS6803	0.0510	0.0070
DXS7132-DXS6789	0.0330	0.0510
DXS6803-DXS6789	0.0510	**$<10^{-4}$**
DXS6803-GATA165B12	0.0510	**$<10^{-4}$**
DXS6789-GATA165B12	**$<10^{-4}$**	**$<10^{-4}$**
DXS6789-DXS7130	0.0060	**$<10^{-4}$**
GATA165B12-DXS7130	0.0510	0.0120
GATA165B12-DXS101	0.0060	**$<10^{-4}$**
DXS7130-DXS101	0.0050	**$<10^{-4}$**
DXS7130-DXS7424	0.0510	0.0420
DXS101-DXS7424	0.0020	**$<10^{-4}$**
DXS7424-HPRTB	0.0510	**0.0010**
DXS7424-DXS7423	0.0510	**0.0010**
HPRTB-DXS7423	0.0130	0.0090
HPRTB-DXS10147	**0.0010**	0.0070
DXS7423-DXS10147	**$<10^{-4}$**	**$<10^{-4}$**
DXS10147-GATA31E08	**$<10^{-4}$**	**$<10^{-4}$**

Linkage analysis

Physical location of studied loci

Marker locations were determined based upon In Silico PCR BLAT searches [71] and organized along the chromosome in Figure 3. Of the 15 markers studies here, the four original linkage groups described by Szibor, *et al.* [30] contained the following markers: DXS8378 and DXS9902 in linkage group 1; DXS7132, DXS6789, DXS101, DXS7424, and GATA172D05 in linkage group 2; HPRTB in linkage group 3; and DXS7423 in linkage group 4. Additional markers (DXS6795, DXS6803, DXS7130, GATA165B12, GATA31E08, and DXS10147) included within each linkage group were hypothesized based upon location, the Forensic ChrX Research website [72], and linkage disequilibrium analysis.

Figure 3. Physical location of 15 X STR markers and four proposed linkage groups on the chromosome.



Recombination rate assessment

Of the Type I families identified within the dataset as appropriate for linkage analysis, 50 families were analyzed using the classical method. Homozygous genotypes and mutations that rendered a marker uninformative for recombination were excluded, and recombination was defined as a change in source chromosome between two adjacent markers. The recombination rate and LOD scores based upon observed instances of recombination are shown in Table 18.

Logarithm of the odds (LOD) scores above 3 are generally considered strongly indicative of linkage; however, an LOD of 2 was used initially to identify the four linkage groups [30].

Table 18. Observed recombination rate between 14 pairs of X STR markers. LG: linkage group, LOD: logarithm of the odds.

Location	X STR marker pair	Total number of informative meioses	Instances of observed recombination	Observed recombination rate	LOD scores (Z)
Linkage group 1	DXS8378-DXS9902	187	21	0.11	27.76
	DXS9902-DXS6795	194	26	0.13	25.21
Border LG1-LG2	DXS6795-DXS7132	158	63	0.40	1.42
Linkage group 2	DXS7132-DXS6803	186	23	0.12	25.77
	DXS6803-DXS6789	182	15	0.08	32.29
	DXS6789-DXS7424	172	6	0.03	40.47
	DXS7424-DXS101	179	0	0	53.88
	DXS101-GATA172D05	194	15	0.08	35.47
	GATA172D05-DXS7130	149	14	0.09	24.69
	DXS7130-GATA165B12	91	4	0.04	20.27
Border LG2-LG3	GATA165B12-HPRTB	141	16	0.11	20.79
Border LG3-LG4	HPRTB-GATA31E08	200	19	0.10	32.94
Linkage group 4	GATA31E08-DXS10147	177	37	0.21	13.87
	DXS10147-DXS7423	134	0	0	40.34

The observed recombination rate varied from zero (marker pairs DXS7424-DXS101 and DXS10147-DXS743) to 0.21 (marker pair GATA31E08-DXS10147) within linkage groups and from 0.10 (border of linkage groups 3 and 4) to 0.40 (border between linkage groups 1 and 2) between linkage groups. The highest recombination rate was observed for the pair of markers defining the boundary between linkage groups 1 and 2; however, free recombination ($\Theta = 0.5$) was never observed. All marker pairs, including those within linkage groups, exhibited a non-zero recombination rate except two: DXS7424-DXS101 in linkage group 2 and DXS10147-DXS7423 in linkage group 4. LOD scores, however, indicated linkage between all pairs except those at the boundary of linkage groups 1 and 2.

The computer-based analyses undertaken with this dataset required the modification of scripts previously designed to accommodate only 12 X STR markers [53]. Unforeseen complications with the ability to perform the necessary computations with available computer systems prevented the analyses of the Type II families at the time of writing. Instead a slightly larger number (58) of Type I families were analyzed using three different starting values: distance-interpolated recombination rates, all recombination rates equal to 0.25, and four groups of unlinked markers. Despite these different starting points, the optimization converged at the same location in each case, indicating a robust optimization result; the values for the recombination rates as well as their 95% support intervals are given in Table 19.

Table 19. Maximum likelihood estimates of recombination rate between 14 pairs of X STR markers. The computer-based analysis described by Nothnagel, *et al.* [53] was performed on 58 Type I families using three different starting values: distance-interpolated recombination rates, all recombination rates equal to 0.25, and four groups of unlinked markers. LG: linkage group.

Location	X STR marker pair	Recombination rate	95% support interval
Linkage Group 1	DXS8378-DXS9902	0.1114	(0.0714-0.1514)
	DXS9902-DXS6795	0.1223	(0.0795-0.1650)
Border LG1-LG2	DXS6795-DXS7132	0.4462	(0.3865-0.5060)
Linkage Group 2	DXS7132-DXS6803	0.0964	(0.0587-0.1341)
	DXS6803-DXS6789	0.0935	(0.0578-0.1292)
	DXS6789-DXS7424	0.0663	(0.0350-0.0977)
	DXS7424-DXS101	0.0000	(NaN-NaN)
	DXS101-GATA172D05	0.0924	(0.0592-0.1256)
	GATA172D05-DXS7130	0.0736	(0.0411-0.1062)
	DXS7130-GATA165B12	0.0484	(0.0149-0.0819)
Border LG2-LG3	GATA165B12-HPRTB	0.0971	(0.0589-0.1353)
Border LG3-LG4	HPRTB-GATA31E08	0.1092	(0.0696-0.1488)
Linkage Group 4	GATA31E08-DXS10147	0.2045	(0.1569-0.2521)
	DXS10147-DXS7423	0.0000	(NaN-NaN)

The recombination rates obtained with the computer-based method generally agreed with the values calculated manually, further indicating a robust computation. These values ranged from zero (marker pairs DXS7424-DXS101 and DXS10147-DXS743) to 0.2045 (marker pair GATA31E08-DXS10147) within linkage groups. Between linkage groups, the lowest rate

occurred at the border of linkage groups 2 and 3 (0.0971) rather than between linkage groups 3 and 4 as for the classical analyses. The border between linkage groups 1 and 2 revealed the highest overall recombination rate for both the manual (0.40) and the computer-based analyses (0.4462), including a 95% support interval that supports free recombination ($\Theta = 0.5$). As noted in the manual analyses, all marker pairs, including those within linkage groups, exhibited a non-zero recombination rate using the computer-based method except two: DXS7424-DXS101 in linkage group 2 and DXS10147-DXS7423 in linkage group 4.

Because mutations were ignored in the manual analyses, it was thought that the true recombination rate had likely been overestimated. In the family described in Appendix A, regarding the 13 allele at DXS7132 in individual 5 as a 14→13 mutational event rather than recombination between DXS7132 and the markers above and below would eliminate 2 of 18 observed events, for example. However, when comparing the manual analyses to computer-based estimates performed using a maximum likelihood approach taking mutation rates into account, the values were found to be very similar. In all cases, the manually-calculated observed recombination rate fell within the 95% support intervals of the computer-based values except for marker pair DXS6789-DXS7424 where the rate of 0.03 fell just outside the lower limit of the 95% support interval (0.0350). Taken together, the results of this combination of methods indicate a robust estimate of the recombination rate between these 14 marker pairs has been achieved.

The genetic distance values calculated as part of the computer-based analyses varied only marginally by starting value, and are given for each marker pair in Table 20. The relative rate of recombination was generally positively correlated with physical distance between markers in this study: as the distance between the markers increased, the recombination rate increased. There was one notable exception to this trend. Marker pair GATA31E08-DXS10147 exhibited a much higher mutation rate (0.21 and 0.2045) than marker pairs with similar genetic distances separating them, as reflected by the genetic distance estimates. Further study is necessary to determine whether this result may indicate a true recombination "hot spot" or may be influenced by factors such as linkage disequilibrium sampling bias, and/or population substructure.

However, the results obtained with this set of markers also agreed with five published recombination studies [100,101,138-140] with regards to recombination between linkage groups 3 & 4 even though their borders were defined by different markers in each. Using a mapping function to compare observed to expected recombination rates for this pair based on physical distance, the expected rates in the published studies were lower than the observed rates. In this study, this phenomenon is reflected in a genetic distance calculation that is almost 1.7 times greater than that of the physical distance. These concordant results may indicate the presence of a region of locally enhanced recombination that may also include marker pair GATA31E08-DXS10147.

Table 20. Genetic distances between 14 X STR marker pairs given for three different starting value schemes. LG: linkage group; Mb: megabases; cM: centimorgans; theta: recombination rate.

			Starting value scheme (cM)		
Location	**X STR marker pair**	**Physical distance (Mb)**	**Distance-interpolated**	**Theta equals 0.25**	**4 unlinked groups**
Linkage Group 1	DXS8378-DXS9902	5.90	11.3310	11.3310	11.3314
	DXS9902-DXS6795	7.92	12.4783	12.4782	12.4781
Border LG1-LG2	DXS6795-DXS7132	41.42	71.6852	71.6863	71.6929
Linkage Group 2	DXS7132-DXS6803	21.75	9.7662	9.7664	9.7663
	DXS6803-DXS6789	9.02	9.4599	9.4602	9.4599
	DXS6789-DXS7424	5.17	6.6694	6.6696	6.6698
	DXS7424-DXS101	0.79	0.0000	0.0000	0.0000
	DXS101-GATA172D05	11.76	9.3463	9.3462	9.3481
	GATA172D05-DXS7130	5.02	7.4173	7.4175	7.4159
	DXS7130-GATA165B12	2.62	4.8550	4.8552	4.8555
Border LG2-LG3	GATA165B12-HPRTB	12.74	9.8320	9.8321	9.8313
Border LG3-LG4	HPRTB-GATA31E08	6.62	11.1001	11.1000	11.0994
Linkage Group 4	GATA31E08-DXS10147	9.35	21.7218	21.7221	21.7211
	DXS10147-DXS7423	0.05	0.0000	0.0000	0.0000

Mixture Multiplex

Assay development

A list of a subset of the markers that were considered for inclusion in the MIXplex, along with details of those that were chosen, appears in Table 21. Other markers were considered but excluded due to some combination of undesirable factors including but not limited to: a complex

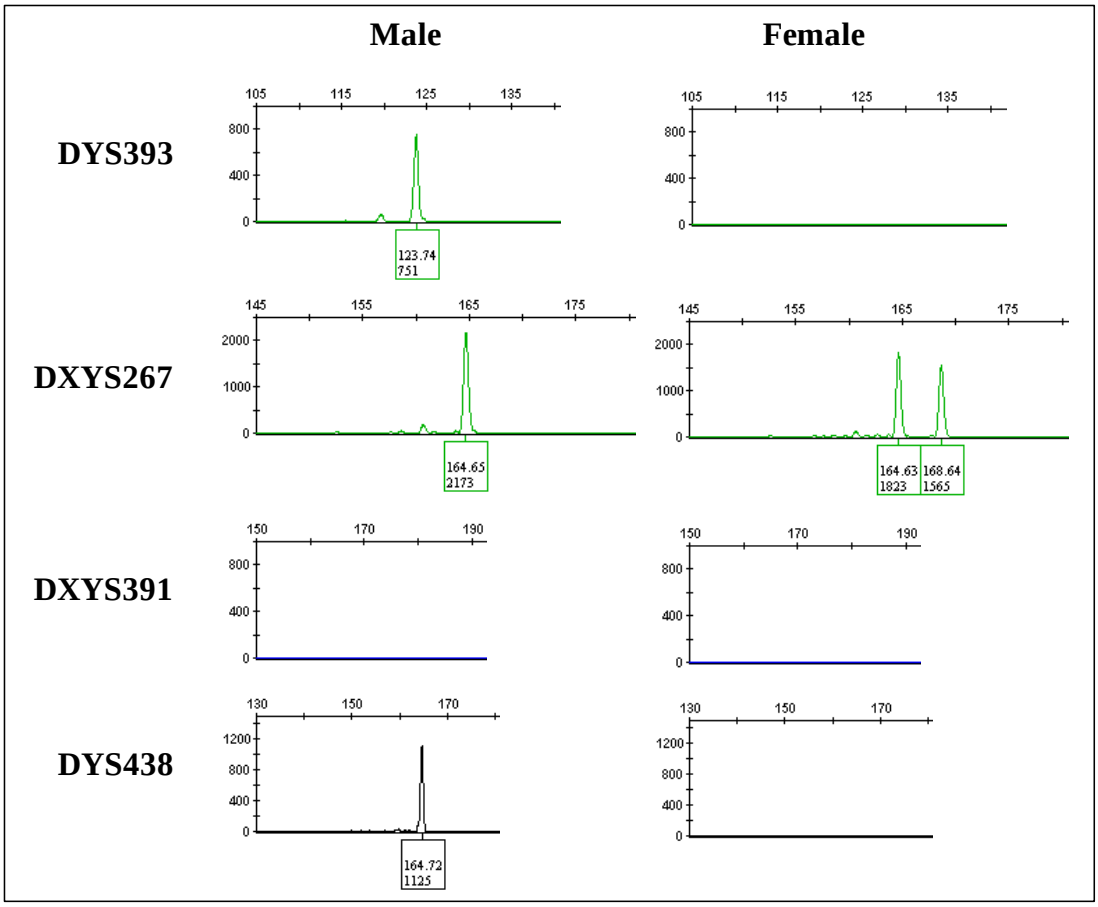
repeat structure (DYS390, DYS635); published amplicons that were too large or too small (DYS19, DYS390, DYS439, DYS635, DYS392, Y-GATA-H4, DYS448); and/or unresolved allele nomenclature issues (DYS439, Y-GATA-H4).

Table 21. List of potential STR markers for inclusion in a mixture multiplex. In addition to the markers characterised for inclusion in the mini-X STR multiplexes, Y and XY markers were needed for the mixture multiplex. Below are markers that were considered, and their selection-relevant characteristics. F: forward; R: reverse; x: number of repeats; bp: base pairs; T_m: melting temperature; Ref: reference.

Marker	Repeat motif	Amplicon size (bp)	Amplification Primer (5'→3')		T_m (°C)	Ref.
DYS393	$AGAT_x$	107-139	F	GTGGTCTTCTACTTGTGTCAATAC	54.7	[141]
			R	AACTCAAGTCCAAAAAATGAGG	57.4	
DXYS267	TATA-$GATA_x$-GACA-GATA	147-179	F	GTGGTCTTCTACTTGTGTCAATAC	54.7	[142]
			R	CTAAATAAAAGTCATATCAGCTGC	53.1	
DXYS391	$TCTA_x$	151-207	F	TTCATTCAATCATACACCCATATC	57.8	[143]
			R	GGAATAAAATCTCCCTGGTTG	57.5	
DYS438	$TTTTC_x$	133-173	F	TGGGGAATAGTTGAACGGTAA	59.3	[144]
			R	GGAGGTTGTGGTGAGTCGAG	60.7	
DYS437	$TCTA_x$-$(TCTG)_2$-$(TCTA)_4$	181-197	F	GACTATGGGCGTGAGTGCAT	61.1	[145]
			R	AGACCCTGTCATTCACAGATGA	59.6	
DYS458	$GAAA_x$	132-160	F	GCAACAGGAATGAAACTCCAAT	60.4	[146]
			R	GTTCTGGCATTACAAGCATGAG	57.5	
DYS391	$TCTA_x$	147-179	F	CTATTCATTCAATCATACACCCATAT	57.5	[141]
			R	ACATAGCCAAATATCTCCTGGG	59.4	
DYS456	$AGAT_x$	137-161	F	GGACCTTGTGATAATGTAAGATA	52.8	[147]
			R	CCCATCAACTCAGCCCAAAAC	63.9	

Singleplex amplifications were performed first to evaluate primer selection for successful amplification, complete adenylation, and peak migration (Figure 4). At this stage, primer sets were redesigned when necessary to address practical issues. The first iteration of the MIXplex included GATA172D05 and DXYS391 in the blue channel, DYS393 and DXYS267 in the green channel, and SRY, GATA31E08, and DYS438 in the yellow channel. All primer sets exhibited expected amplification for both males and females except DXYS391.

Figure 4. Singleplex testing results for mixture multiplex markers. Candidate primer sets that had not been previously evaluated during the development of the X STR multiplexes were tested in singleplex with two male and three female samples. Here, representative male and female profiles at each marker are shown.

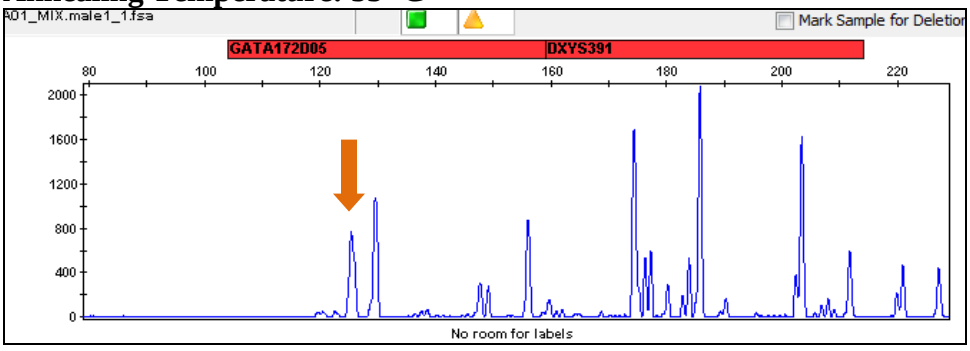


When the primer sets were combined for multiplex amplification, extraneous peaks appeared in the 6-FAM™ (blue) channel, obscuring the true allele peaks for the samples regardless of annealing temperature (Figure 5). A series of multiplex primer mixes was created that omitted one primer set at a time in order to attempt to uncover which primer(s) might be responsible through process of elimination. Amplification revealed that when the primer set for either GATA172D05 or DYS438 were removed from the multiplex primer mix, the extraneous peaks disappeared (data not shown). Because these extraneous peaks were seen only in the 6-FAM™ (blue) channel, the forward (labeled) GATA172D05 primer and the reverse (unlabeled) DYS438

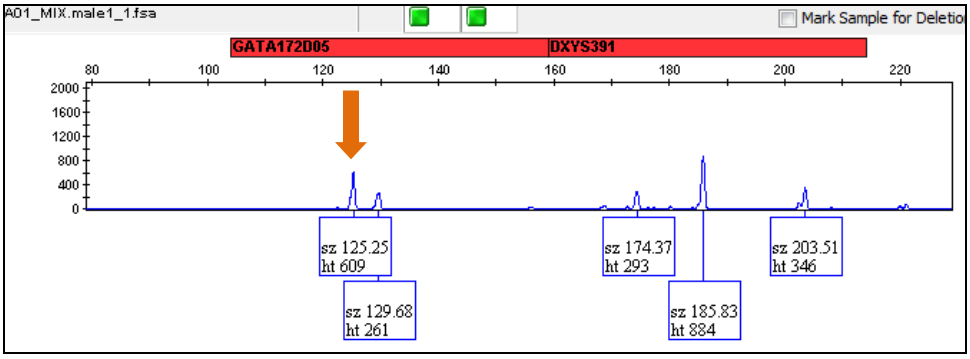
primer were likely binding elsewhere in the genome to produce the additional amplicons. A BLAST search using these two primers revealed at least 10 additional amplicons between 70 and 250 bp.

Figure 5. First version of a mixture multiplex amplified at both 55 °C (top) and 60 °C (bottom). The 6-FAM™ (blue) channel exhibits multiple non-specific peaks at both annealing temperatures used. Authentic GATA172D05 allele peak in each electropherogram is indicated with an arrow. All other peaks (green and yellow channels; not shown) appeared as expected.

Annealing Temperature: 55 °C



Annealing Temperature: 60 °C

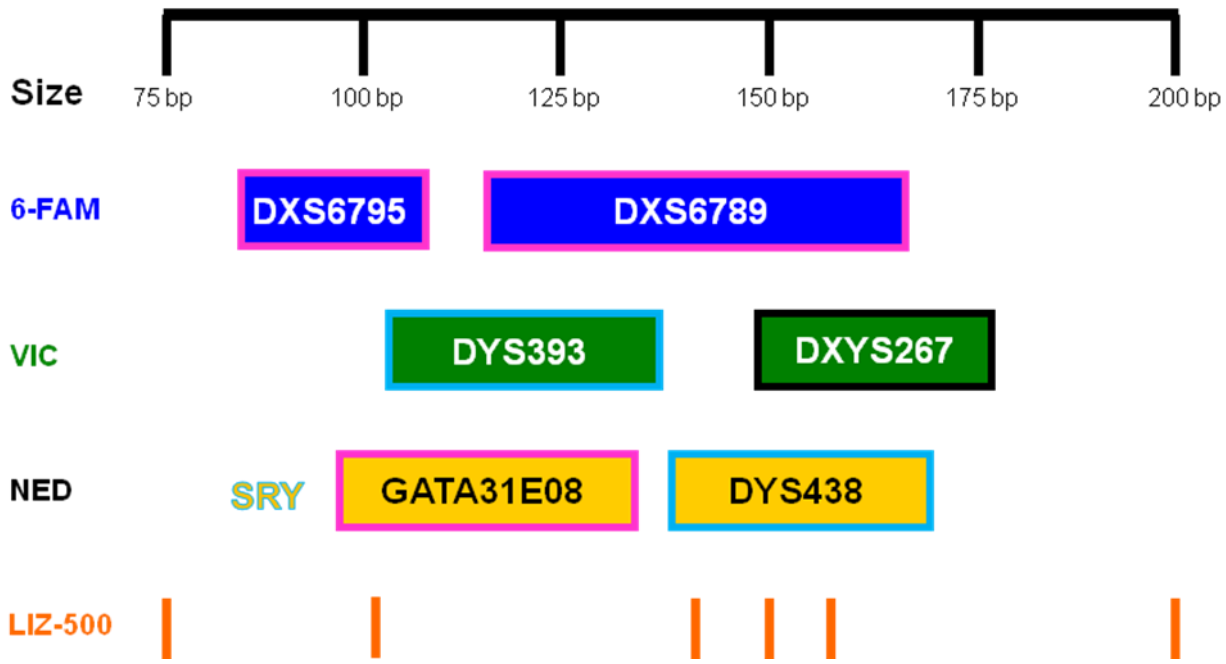


As a result of the initial poor performance of the DXYS391 primer set and the artifacts generated as a result of the GATA172D05 forward primer, both markers were subsequently replaced with the X STR markers DXS6795 and DXS6789, which were known to be robust and polymorphic based upon previous X STR multiplex development work [22]. The organization of the final MIXplex is show in Figure 6A. Further optimization aimed at balancing the inter-locus peak

heights was undertaken, and example electropherograms depicting the final multiplex used for the purposes of this study are shown in Figure 6B. Though additional improvements, such as decreasing the rate of incomplete adenylation, would be necessary before routine use in a casework laboratory, the multiplex as shown was sufficient to test the principle of gonosomal markers to aid in mixture interpretation. Primer sequences and concentrations in the primer mix are detailed in Table 22.

Figure 6. Multiplex organization of the finalized MIXplex. A. The electrophoretic position and known amplicon size range is shown. Markers with pink borders are X STRs, markers with blue borders are Y chromosomal markers, and the one with a black border is an XY marker. B. Example electropherograms of male, female, and male-female mixed profiles generated using the final mixture multiplex.

A.



B.

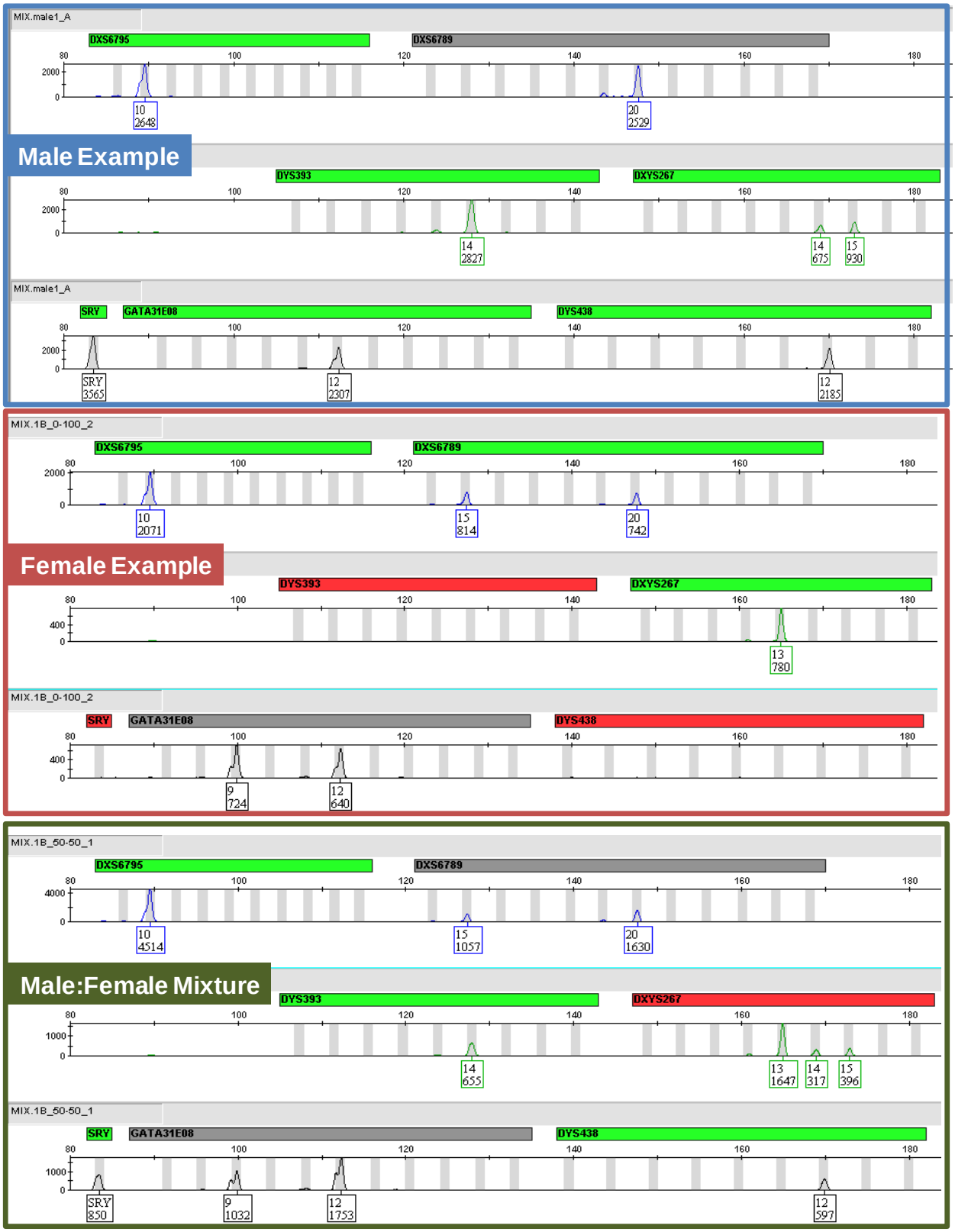


Table 22. Final primer sequences and concentrations used in MIXplex assay. Bases in bold are tails added to the primer sequence to promote complete adenylation of the amplicon or to improve electrophoretic separation of adjacent amplicons. Amplicon size ranges include tails. Conc.: concentration; bp: base pairs.

Marker Name		Primer Sequence (5'-3')	Ref.	Final Conc. (µM)	Amplicon Size (bp)
DXS6795	F	6FAM-TGACATGGCTTTCTTTACAATTAC	[22]	0.8	81-114
	R	**G**CCATGTTACATAAACAAGGAGTTATG	[22]	0.8	
DXS6789	F	6FAM-CCTCGTGATCATGTAAGTTGG	[15]	1.2	124-168
	R	**ATT**CAGAACCAATAGGAGATAGATGGT	[15]	1.2	
DYS393/DXYS267	F	VIC-GTGGTCTTCTACTTGTGTCAATAC	[142]	1.6	
DYS393	R	**G**AACTCAAGTCCAAAAAATGAGG	[142]	2.0	108-140
DXYS267	R	**G**CTAAATAAAGTCATATCAGCTGC	[142][a]	0.1	148-180
SRY	F	NED-AAAAATTGGCGATTAAGTCAAA	[22]	0.8	86
	R	**G**TTGACTACTTGCCCTGCTGA	[22]	0.8	
GATA31E08	F	NED-CAGAGCTGGTGATGATAGATGA	[15]	2.0	99-143
	R	**ATT**CTCACTTTTATGTGTGTATGTATCTCC	[15]	2.0	
DYS438	F	NED-TGGGGAATAGTTGAACGGTAA	[145]	1.2	140-180
	R	**GTTTCTT**GGAGGTTGTGGTGAGTCGAG	[144]	1.2	

[a]Basic primer sequence obtained from publication with underlined base modified in this study to match GenBank sequence.

Allele sequencing

In anticipation of a need to sequence certain alleles to confirm repeat number and structure, two new sequencing primer pairs were designed for markers DYS393 and DYS438. The primer sequences are shown in Table 23 along with previously designed sequencing primers for the three X markers [22].

Table 23. Unlabelled primers used for sequencing purposes.

Marker Name	Forward Primer Sequence (5'→3')	Reverse Primer Sequence (5'→3')
DYS393	TTGTAGTTATGTTTTATTTGTCATTCC	AAATGTTACAAAAAGAATGGCCTA
DYS438	TGATGCAAGAAAGATTCACTGAT	AGGAGAATCGCTTGAACCTG
DXS6795	TTCATGCTGTTGCTTTCCAG	CCATCCCCTAAACCTCTCAT
DXS6789	TCAAGCTTGCAGACAGCCTA	TCGAAAAGATAGCCAATCACTG
GATA31E08	AGCAAGGGGAGAAGGCTAGA	TCAGCTGACAGAGCACAGAGA

Sensitivity testing

Full single-source MIXplex profiles were reliably obtained with as little as 200 pg of input DNA (Table 24). Complete loss of an allele despite triplicate amplification was seen in two samples (one male and one female) at one marker (DXYS267 and GATA321E08 respectively) with 50 pg of input, and multiple alleles were lost from profiles amplified from only 25 pg of input. Sensitivity results did not vary by sex of the sample, as might be expected due to the chromosomal copy number differences between the samples. Similarly, the entirely homozygous sample female 3 did not show increased sensitivity compared to the other two samples. Heterozygous peak height ratios remained above 60% for recovered alleles in all replicates (37) except two, both at DXYS267: one replicate of male 3 and one replicate of female 2 (both 52%). Marker DXYS267 appeared to be the least robust amplicon, as it was the first marker to exhibit loss of recovery in all three samples. In contrast, marker DXS6795 may be the most robust amplicon, with complete loss of amplification for only one allele (9) in one sample (male 3).

Table 24. MIXplex sensitivity testing results. Each single-source sample was amplified in triplicate at each of 6 different input quantities (54 amplifications total). Thresholds of 100 RFU for male or heterozygous alleles and 200 RFU for homozygous alleles were used to define the presence of a peak. Triplicate results were combined for the purposes of this table, with green boxes representing alleles that were above threshold in all three amplifications, yellow boxes representing alleles that were present in only 1 or 2 amplifications, and red boxes indicating that no alleles were above threshold. Allele calls are designated for each sample within the boxes.

Quantity (pg)	Sample	DXS6795	DXS6789	DXS6789	DYS393	DXYS267	DXYS267	SRY	GATA31E08	GATA31E08	DYS438
1000	Male 3	9	21		13	12	13	SRY	14		9
	Female 2	10	15	20		13			9	12	
	Female 3	9	21			15			11		
500	Male 3	9	21		13	12	13	SRY	14		9
	Female 2	10	15	20		13			9	12	
	Female 3	9	21			15			11		
200	Male 3	9	21		13	12	13	SRY	14		9
	Female 2	10	15	20		13			9	12	
	Female 3	9	21			15			11		
100	Male 3	9	21		13	12	13	SRY	14		9
	Female 2	10	15	20		13			9	12	
	Female 3	9	21			15			11		
50	Male 3	9	21		13	12	13	SRY	14		9
	Female 2	10	15	20		13			9	12	
	Female 3	9	21			15			11		
25	Male 3	9	21		13	12	13	SRY	14		9
	Female 2	10	15	20		13			9	12	
	Female 3	9	21			15			11		

Mixture analysis: artificial mixtures

The results of duplicate or triplicate amplification of six two-person mixtures (four female-male, one male-male, and one female-female) at varying mixture ratios is shown in Table 25. Complete profiles (defined as containing all expected alleles >100 RFU) were reliably obtained for mixtures where the minor component was 20% or greater in most cases. This value coincided well with the single-source sensitivity results for the MIXplex, which showed that reliable profiles were generated with just 200 pg input. Like the sensitivity results, the mixture testing also revealed that near-complete or complete profiles could also be obtained for most

mixtures with the minor component at only 10%, or 100 pg input. Several notable exceptions did exist, however. The minor component of mixture 1-A2 (Table 25A), 1-A3 (Table 25B), and 1-M2 (Table 25E) was completely undetected for the 10:90 (but not the 90:10) amplifications. In each case, the PCR reaction itself appeared to be suboptimal, as indicated by low peak heights and/or missing alleles for the major component (at 900 pg). It is likely that re-amplification may generate more robust profiles for both the major and minor components which more accurately reflect the sensitivity of the MIXplex.

When allele peak heights were divided into two categories, above or below 1000 RFU (shaded blue or not shaded blue respectively in Table 25 below), the pattern highlighted the minor component in more extreme mixture ratios (10:90, 20:80, 30:70, 70:30, 80:20, 90:10) and heterozygous markers in the relatively even mixture ratios (40:60, 50:50, 60:40), as expected. Little bias towards sex of the contributor or marker type was seen. Therefore, input DNA concentration was found to have the largest effect on allele recovery with the MIXplex, as would be a desired characteristic of an assay aimed at aiding mixture interpretation.

Table 25A-F. Artificial mixture testing results. Six two-person mixtures (four female-male, one male-male, and one female-female) were amplified in duplicate or triplicate both neat and at nine different mixture ratios. Expected profiles for the individual components as well as the mixtures are shown above each table, and RFU values for each peak greater than the 100 RFU reporting threshold were recorded within each box. Blue boxes highlight values >1000 RFU and uncoloured boxes contain values between 100 and 1000 RFU. Orange boxes highlight missing alleles (<100 RFU). Male 1 or M1: Quantifiler® Human DNA Standard; male 3 or M3: 2800M; female 2 or F2: AFDIL-1; female 3: K562.

A. Mixture testing: female-male mixture 1-A2

	DXS6795	DXS6789	DXS6789	DYS393	DXYS267	DXYS267	DXYS267	SRY	GATA31E08	GATA31E08	DYS438
female 2	**10**	**15**	**20**		**13**				**9**	**12**	
male 1	**10**		**20**	**14**		**14**	**15**	**SRY**		**12**	**12**
female 2:male 1	**10**	**15**	**20**	**14**	**13**	**14**	**15**	**SRY**	**9**	**12**	**12**
0:100	1320	320	334		529				372	387	
	2071	814	742		780				724	640	
	1880	513	586		889				597	581	
10:90	1015	230	257		317				279	279	
	384	115	132		113				117	113	
	186										
20:80	8624	3145	4123	432	3798	190	212	554	2604	3085	383
	6835	2303	2604	511	2408	242	162	489	1504	1761	321
	3405	827	726	100	831			169	650	686	123
30:70	5067	1743	2066	557	2282	211	292	487	1708	1634	474
	5053	1320	1937	344	1652	169	175	358	1391	1602	331
	2443	792	916	203	842			206	691	649	158
40:60	5780	2113	2973	605	2257	251	312	604	1772	2175	634
	3758	1255	1586	259	1197	113	125	412	848	1254	296
	5271	1774	2081	509	1634	255	178	677	1600	1837	371
50:50	4514	1057	1630	655	1647	317	396	850	1032	1753	597
	4711	1354	1651	445	1421	210	322	606	839	1570	594
	2453	626	1036	394	587	180	166	412	577	673	390
60:40	3673	743	1563	647	871	283	399	888	867	1368	712
	3585	823	1747	916	1049	406	320	754	851	1335	618
	4971	1285	1818	928	1279	478	401	700	524	2002	838
70:30	2613	537	1199	702	650	324	387	846	489	1158	687
	2571	677	1249	758	575	334	378	746	533	1117	655
	2601	538	1119	615	573	278	267	625	408	845	544
80:20	2350	504	1458	1157	475	478	420	1288	456	1058	646
	1984	442	1213	688	398	297	313	1045	332	915	574
	3950	452	2128	1125	754	578	677	1454	549	1344	1074
90:10	1724		922	829	185	331	480	874	135	1005	676
	2064	235	1143	1055	211	466	450	1021	144	834	594
	1836	120	985	686	206	331	417	1061	250	882	809
100:0	2466		1477	1823		768	867	2089		1126	1859
	1874		1392	1064		478	478	1158		902	1100
	2187		1614	1263		542	631	1571		975	1399

B. Mixture testing: female-male mixture 1-A3

	DXS6795	DXS6795	DXS6789	DXS6789	DYS393	DXYS267	DXYS267	SRY	GATA31E08	GATA31E08	DYS438
female 3	**9**			**21**			**15**			**13**	
male 1		**10**	**20**		**14**	**14**	**15**	**SRY**	**12**		**12**
female 3:male 1	**9**	**10**	**20**	**21**	**14**	**14**	**15**	**SRY**	**12**	**13**	**12**
0:100	908			536			322			485	
	2647			1697			957			1608	
	1078			652			359			576	
10:90	143										
	616			297			138			307	
	816			426			196			448	
20:80	5500	561	678	3368	282	341	2410	362	412	2686	425
	4173	231	476	2601	213	224	1497	254	249	1573	231
	3838	324	498	2026	165	185	1341	128	277	1629	235
30:70	5334	902	922	3123	596	493	2422	394	688	2680	725
	5573	874	940	3497	417	333	1820	797	801	2999	883
	3614	508	616	1690	318	200	915	344	314	1352	266
40:60	3782	831	804	2502	588	335	1467	797	431	1660	515
	4200	1050	1136	2326	609	351	1273	575	610	1652	554
	4006	1043	825	2044	563	295	1440	738	569	1935	482
50:50	2380	666	814	1132	442	273	1034	625	605	1159	615
	3262	866	874	1495	703	390	1191	840	770	1352	594
	2274	737	581	1355	382	210	828	517	467	962	410
60:40	4390	2548	1427	2162	1255	730	1816	1312	1391	1656	1271
	2311	1148	1303	1963	1096	547	1317	738	946	1443	850
	2661	1252	1105	1492	848	437	926	869	632	1332	806
70:30	1672	1051	907	988	528	282	736	733	584	548	494
	1568	1166	807	669	596	278	832	905	546	599	772
	1585	1424	1031	1107	731	318	628	876	776	788	609
80:20	1534	1616	1540	699	1284	615	1191	1389	939	951	1257
	1088	1571	904	732	1077	547	766	999	579	394	902
	1147	1510	930	547	1136	491	749	969	952	485	786
90:10	515	1228	1083	411	720	359	817	1073	714	249	925
	825	1719	951	472	1085	520	788	1051	990	231	949
	771	1427	1148	341	768	374	660	1631	1029	318	1146
100:0		2247	1586		1393	643	597	1614	1104		1199
		2008	1259		1162	516	570	1505	980		1130
		1853	1518		1371	582	576	1112	1154		906

C. Mixture testing: female-male mixture 1-A8

	DXS6795	DXS6795	DXS6789	DXS6789	DXS6789	DYS393	DXYS267	DXYS267	SRY	GATA31E08	GATA31E08	GATA31E08	DYS438
F2		**10**	**15**	**20**				**13**		**9**	**12**		
M3	**9**				**21**	**13**	**12**	**13**	**SRY**			**14**	**9**
F2:M3	**9**	**10**	**15**	**20**	**21**	**13**	**12**	**13**	**SRY**	**9**	**12**	**14**	**9**
0:100		6612	2625	2429				3025		2051	1764		
		6301	2411	2206				2642		1472	1499		
		8140	2663	2398				4191		2387	2044		
10:90	291	3787	1227	1463	187	216	285	1971	285	1229	1232	207	232
	400	4785	1752	1476	306	154	224	1931	112	1098	1211	134	103
	557	4563	2028	1730	172	279	314	2397	239	1352	1540	187	169
20:80	836	5079	2013	1583	697	339	369	2178	633	1101	1251	362	382
	703	4388	1569	1430	437	245	470	2060	426	1191	972	262	313
	1077	5941	1966	2093	464	527	414	2934	550	1392	1490	367	385
30:70	1091	3607	1201	1199	512	569	548	1978	804	1334	1991	584	552
	623	3499	1383	1472	553	444	442	1631	551	920	901	314	569
	775	3648	1264	1096	599	353	582	2231	691	1117	878	390	505
40:60	1833	4535	1987	2003	1140	937	878	2262	848	1250	1347	1003	735
	1049	3628	1312	1278	884	531	543	1460	693	809	1055	613	625
	1791	5108	1898	1997	1150	1031	896	3053	1371	1486	1381	716	737
50:50	1246	2746	839	1078	753	1041	709	1837	962	928	866	969	854
	1593	2594	1136	975	1018	773	545	1327	846	601	695	774	677
	1753	2917	1126	1113	895	1091	686	2254	1234	821	800	800	921
60:40	2452	2885	777	1000	1301	1277	822	1754	1071	516	639	1257	1075
	1898	2349	751	1036	907	1206	738	1488	1262	586	568	876	758
	2560	3070	1070	886	1585	1088	1295	1979	1465	937	602	1256	968
70:30	1686	1161	481	504	987	930	858	1117	1210	460	522	922	921
	1867	1729	608	727	1098	1001	648	1103	1163	379	385	887	779
	2310	1787	534	757	1425	1100	932	1468	1174	480	410	1236	922
80:20	3205	1733	552	608	1913	1577	1060	1413	1660	379	499	1471	1520
	2430	1651	354	577	1341	1142	950	994	1479	348	423	1133	1091
	2838	1612	550	752	2026	1452	1217	1482	1740	376	382	1427	1297
90:10	2935	460	188	452	1855	1538	1166	1066	2182		244	1697	1712
	2551	748	109	379	1568	1168	905	810	1454	157	109	1177	1211
	2466	589	248	392	1629	1250	1180	871	1527		244	1289	1423
100:0	3840				2769	2291	1384	1035	2506			1863	1810
	3425				2400	1486	1261	763	2021			1332	1573
	3758				2578	2003	1458	939	2242			1679	1814

D. Mixture testing: female-male mixture 1-A9

	DXS6795	DXS6789	DYS393	DXYS267	DXYS267	DXYS267	SRY	GATA31E08	GATA31E08	DYS438
female 3	**9**	**21**				**15**		**13**		
male 3	**9**	**21**	**13**	**12**	**13**		**SRY**		**14**	**9**
female 3:male 3	**9**	**21**	**13**	**12**	**13**	**15**	**SRY**	**13**	**14**	**9**
0:100	6397	3941				2715		2907		
	7003	3415				2043		2689		
	6412	4133				2059		3208		
10:90	4315	2820	191	236	123	1618	207	1764	208	221
	5253	2948	266	191	127	1569	355	1953	233	272
	5934	2686	189	158	114	1533	312	2042	333	199
20:80	5288	3441	356	355	182	1655	519	2115	295	538
	5191	3298	433	386	231	1421	559	1932	342	288
	5157	3130	424	213	229	1488	306	1637	284	338
30:70	5070	2989	432	433	266	1519	778	1258	526	599
	5279	3142	566	480	345	1393	757	1521	591	513
	4810	3205	528	449	261	1119	852	1522	473	561
40:60	6371	4254	867	732	454	1308	1375	1660	841	778
	5036	3258	749	578	445	1226	1210	1757	801	735
	5528	2717	701	640	353	1333	1000	1823	882	779
50:50	4461	2630	922	782	470	840	1013	1301	834	961
	5351	2903	1150	799	605	1286	1178	1631	899	1101
	5031	2938	967	864	465	811	1026	1404	995	1197
60:40	5189	3242	1431	1156	879	1030	1824	1444	1296	1114
	4831	3101	1214	775	666	785	1614	1155	1310	1065
	4198	2993	1354	819	669	594	1239	1040	1269	1075
70:30	3954	2520	1043	885	623	538	1545	855	1228	1328
	4780	2732	1330	1078	812	525	1753	947	1228	1466
	4048	2331	1318	889	647	412	1986	984	1232	1428
80:20	4318	2708	1520	1192	807	465	2242	592	1517	1432
	4314	2706	1420	1129	788	368	2293	574	1198	1425
	2949	1961	949	993	694	340	1495	393	1123	992
90:10	3911	2376	1447	1359	726	221	1865	370	1641	1768
	3986	2316	1808	1507	981	348	2548	368	1700	2018
	4189	2295	1868	1106	948	227	1971	417	1623	1823
100:0	4168	2594	1925	1690	1022		2953		2208	2255
	3874	2159	1866	1235	988		2614		1953	1857
	3840	2225	2003	1299	900		2387		1701	1716

E. Mixture testing: male-male mixture 1-M2

	DXS6795	DXS6795	DXS6789	DXS6789	DYS393	DYS393	DXYS267	DXYS267	DXYS267	DXYS267	SRY	GATA31E08	GATA31E08	DYS438	DYS438
M1		**10**	**20**			**14**			**14**	**15**		**12**			**12**
M3	**9**			**21**	**13**		**12**	**13**			**SRY**		**14**	**9**	
M3:M1	**9**	**10**	**20**	**21**	**13**	**14**	**12**	**13**	**14**	**15**	**SRY**	**12**	**14**	**9**	**12**
0:100	694			450	424		182	141			519		418	422	
	581			371	418		236	161			582		378	384	
10:90	250			156	139						209		143	173	
	420			233	246		130				368		262	258	
20:80	4651	937	866	3396	3058	367	2151	1200	168	102	3733	317	2659	2919	508
	3986	479	731	2469	2387	347	1418	875	147	164	3056	432	2390	2381	261
30:70	3757	812	1357	2645	2997	535	1542	1187	225	238	3355	532	2352	2895	827
	3547	882	1060	2758	2393	839	1527	869	362	339	2988	436	2047	2539	599
40:60	2416	1217	743	1683	1492	451	997	572	184	237	2429	400	1677	1470	650
	2157	660	954	1290	1744	518	834	669	209	232	1863	380	1401	1650	519
50:50	2179	1031	1275	1720	1266	650	699	524	324	483	1852	629	1138	989	765
	2347	992	863	1365	1601	662	784	607	325	352	2153	546	963	916	865
60:40	1471	1050	1222	1296	1124	1074	626	431	410	382	1812	642	990	1133	7965
	1597	1263	878	1013	1249	736	532	469	326	380	1821	597	948	961	852
70:30	1517	1594	1890	1192	1375	2247	333	264	425	337	2113	1178	851	769	1078
	844	878	730	578	448	534	302	179	229	255	1282	638	703	440	519
80:20	556	1983	1156	236	265	1011	121		413	479	1364	1038	117	204	816
	735	846	846	367	487	837	217	157	329	237	1480	655	206	374	645
90:10		1715	1335			1111			458	627	1447	877			1263
	307	1626	999	210	271	821			366	405	1427	982	160	194	898
100:0		1998	887			1087			449	412	1008	1096			1014

F. Mixture testing: female-female mixture 1-F3

	DXS6795	DXS6795	DXS6789	DXS6789	DXS6789	DXYS267	DXYS267	GATA31E08	GATA31E08	GATA31E08
female 2		**10**	**15**	**20**		**13**		**9**	**12**	
female 3	**9**				**21**		**15**			**13**
female 3:female 2	**9**	**10**	**15**	**20**	**21**	**13**	**15**	**9**	**12**	**13**
0:100	7011				4161		2596			3497
	5026				2666		1723			2737
10:90	4544	479	253	452	2470	364	1291	263	312	1888
	4055	710	181	569	2503	305	1568	195	376	1993
20:80	4742	1613	541	723	2319	595	1359	439	490	1793
	4335	1131	416	512	2244	577	1459	453	342	2138
30:70	3799	1849	623	862	1966	697	1076	521	505	1587
	3652	1944	815	702	2133	996	1364	482	590	1666
40:60	4465	2785	1209	1472	2264	1486	1398	835	1205	1320
	3005	2523	1186	1272	1788	1260	1041	976	911	1548
50:50	2990	3022	973	1439	1531	1005	956	644	822	1208
	3000	3283	1095	1454	1704	1253	846	813	967	1504
60:40	2402	4283	1457	1636	1502	1315	827	982	1122	1221
	1875	3417	1156	1083	1039	1228	672	780	889	1073
70:30	1794	4365	1533	1212	922	1678	573	1212	1381	764
	1599	3674	1171	1448	781	1560	761	989	1096	703
80:20	1462	5453	2162	2291	944	2215	382	1574	1521	643
	1117	4549	1449	1426	491	2074	268	1428	1204	384
90:10	561	4712	1826	1730	212	2079	237	1175	1329	191
	638	4845	1734	1343	257	1737	225	1304	1189	229
100:0		7605	2844	2534		3198		1585	1828	
		6197	2705	2500		2695		2206	2330	

The same mixtures and mixture ratios were amplified with the autosomal STR kit Identifiler® (Applied Biosystems) and analyzed for the ability to detect the presence of a mixture. Based upon the criteria of observing at least two markers with >2 alleles above 100 RFU, all mixtures with a minor component 20% or greater as well as most mixtures with a minor component of 10% or greater, were correctly identified as mixtures, as expected (data not shown). Assignment of the sex of the contributors, however, was more complicated due to the normal variation in peak height seen with the single sex-typing marker, amelogenin.

Mixture analysis: theoretical mixtures

Sixty-three theoretical MIXplex profiles were analyzed to determine the minimum number and sex of contributors without knowledge of the profile source(s). During this process, a general method and sequence of analysis was determined to be optimal and is summarized in Table 26. To begin, the presence or absence of the SRY peak is noted. Its presence indicates that at least one male contributed to the profile. Its absence, however, does not necessarily preclude a male contribution. In order to account for the possibility of an SRY null allele, the presence or absence of alleles at the two Y chromosomal markers DYS393 and DYS438 must be noted. If the SRY peak is missing, but at least one peak is present at a Y marker, then it can be assumed there is a male component to the mixture. Thus far, an SRY null allele has never been observed by the authors during the typing of over 3000 samples, or in the literature at this time.

Table 26. Interpretation method used in this study to determine the minimum number and sex of contributors for MIXplex profiles in the absence of peak height information.

Order	Marker Type	Inference(s) made
1	SRY	Determines if there is a male component present
2	DYS	1. Confirms presence or absence of male component of mixture 2. Maximum number of alleles at one marker = minimum number of males
3	DXS	(Maximum number of alleles at one marker – minimum number of males)/2 = minimum number of females*
4	DXYS	1. Confirms conclusions regarding the minimum number of total contributors 2. Confirms minimum number of male contributors**

*Must round up to nearest whole number.
**Must see all DYS393 alleles captured in DXYS267 profile.

Once the presence of a male component has been established, the Y STR markers are used to determine the minimum number of male contributors to the mixture. The largest number of detectable alleles present at a Y STR marker in the multiplex represents the minimum number of males present in the mixture. Additional males may have contributed to the profile but are not represented due to allele sharing.

After establishing the minimum number of male contributors or the absence of a male component, the X STR markers are examined to determine the possible presence of a female

component to the mixture. The previously-determined minimum number of males in the mixture is subtracted from the number of alleles at the X STR marker with the maximum number of alleles and divided in half. After rounding up to the nearest whole number, the result is the minimum number of female contributors to the profile. Again, additional females may be present but are masked by allele sharing.

Lastly, the XY homologous marker is used to confirm inferences made in the previous three steps concerning sex and minimum number of contributors. The number of alleles can be used to confirm the minimum number of contributors in the same way as with autosomal profiles; there must be at least one allele present for each suspected contributor (male or female). Additionally, since the DXYS primer pair is simply amplifying a larger flanking region surrounding DYS393, all of the DYS393 alleles must also be represented in the DXYS267 profile.

In contrast, the interpretation logic used for the interpretation of theoretical mixtures with Identifiler® profiles in the absence of peak height information is relatively simplistic (Table 27). Since there are only two marker types present in this assay, inferences are limited to the presence or absence of a minimum of one male contributor, and a determination of the minimum number of contributors. No additional conclusions regarding the sex or the number of each sex of contributors is possible.

Table 27. Interpretation method used in this study to determine the minimum number of contributors for Identifiler® profiles in the absence of peak height information.

Order	Marker type	Inference(s) made
1	Amelogenin	Determine if there is a male component present
2	Remaining autosomal markers	Maximum number of alleles at one marker ÷ 2 = minimum number of contributors*

*Must round up to nearest whole number.

Once established, these analysis methods were used on all 63 theoretical mixtures to define the sex and minimum number of the suspected contributors. These results were subsequently compared to the true contributors, and the difference was noted (Table 28). For ease of comparison, Identifiler® differences were designated as female when both male(s) and female(s) contributors were not inferred, and a loss of sex information despite correct the number of contributors is captured as a difference of "1 sex."

Table 28. Blind analysis of the MIXplex profiles of 63 theoretical mixtures. The actual combination indicates the true sex and number of contributors used to create the theoretical profile. The minimum combination describes the number and sex of contributors determined according the analysis methods above (Tables 23 & 24) for both the MIXplex and Identifiler® kit. The difference between the minimum and actual combination is shown in the last column. Complete loss of information is highlighted in red text, while loss of either contributor number information or sex determination losses were noted in blue and green text, respectively. "Sex" in the difference column designates a loss of contributor sex determination though the number of contributors was captured correctly.

Profile identifier	Actual combination	MIXplex minimum combination	MIXplex difference	Identifiler® minimum combination	Identifiler® difference
101	1 male/2 females	1 male/2 females		2 total, 1 male	1 female + 1 sex
102	3 males/1 female	3 males/1 female		3 total, 1 male	1 female + 2 sex
103	1 male/2 females	1 male/2 females		3 total, 1 male	1 female + 2 sex
104	2 males/2 females	2 males/1 female	1 female	4 total, 1 male	3 sex
105	3 males/2 females	3 males/1 female	1 female	4 total, 1 male	1 female + 3 sex
106	1 male	1 male		1 male	
107	2 females	2 females		2 females	
108	2 males	2 males		2 total, 1 male	1 male
109	1 male/1 female	1 male/1 female		2 total, 1 male	1 female
110	2 males/1 female	2 males/1 female		2 total, 1 male	1 female + 1 sex
111	1 male/2 females	1 male/2 females		2 total, 1 male	1 female + 1 sex
112	1 male/2 females	1 male/1 female	1 female	2 total, 1 male	1 female + 1 sex
113	1 female	1 female		1 female	
114	1 male/3 females	1 male/2 females	1 female	3 total, 1 male	1 female + 2 sex
115	3 males/3 females	3 males/1 female	2 females	4 total, 1 male	2 females + 3 sex
116	2 males/2 females	2 males/2 females		3 total, 1 male	1 female + 2 sex
117	1 male/2 females	1 male/1 female	1 female	3 total, 1 male	2 sex
118	1 male/1 female	1 male/1 female		2 total, 1 male	1 sex
119	1 male/1 female	1 male/1 female		2 total, 1 male	1 sex
120	2 males/1 female	2 males	1 female	3 total, 1 male	2 sex
121	2 males	2 males		2 total, 1 male	1 sex
122	2 males/1 female	2 males/1 female		3 total, 1 male	2 sex
123	1 female	1 female		1 female	

Table continues on next page

Profile identifier	Actual combination	MIXplex minimum combination	MIXplex difference	Identifiler® minimum combination	Identifiler® difference
124	1 male/2 females	1 male/1 female	1 female	3 total, 1 male	2 sex
125	3 males	3 males		3 total, 1 male	2 sex
126	1 male/3 females	1 male/2 females	1 female	3 total, 1 male	1 female + 2 sex
127	2 males/2 females	2 males/1 female	1 female	3 total, 1 male	1 female + 2 sex
128	2 males/2 females	2 males/1 female	1 female	3 total, 1 male	1 female + 2 sex
129	2 males/2 females	2 males/2 females		3 total, 1 male	1 female + 2 sex
130	3 males/2 females	3 males/1 female	1 female	4 total, 1 male	1 female + 3 sex
131	1 male/2 females	1 male/2 females		2 total, 1 male	1 female + 1 sex
132	1 male/2 females	1 male/2 females		3 total, 1 male	2 sex
133	1 male	1 male		1 male	
134	2 males/1 female	2 males/1 female		3 total, 1 male	2 sex
135	1 male/1 female	1 male/1 female		2 total, 1 male	1 sex
136	1 male/1 female	1 male/1 female		2 total, 1 male	1 sex
137	1 male/1 female	1 male/1 female		2 total, 1 male	1 sex
138	1 male	1 male		1 male	
139	1 male/1 female	1 male/1 female		2 total, 1 male	1 sex
140	2 males	2 males		2 total, 1 male	1 sex
141	2 males/1 females	2 males/1 females		3 total, 1 male	2 sex
142	2 males/1 females	2 males/1 females		3 total, 1 male	2 sex
143	1 male/2 females	1 male/2 females		3 total, 1 male	2 sex
144	3 females	2 females	1 female	3 females	
145	2 males/1 female	2 males/1 female		3 total, 1 male	2 sex
146	2 males/1 female	2 males/1 female		3 total, 1 male	2 sex
147	2 males/1 female	2 males/1 female		3 total, 1 male	2 sex
148	1 male/1 female	1 male/1 female		2 total, 1 male	1 sex
149	1 male/1 female	1 male/1 female		2 total, 1 male	1 sex
150	1 male/3 females	1 male/2 females	1 female	3 total, 1 male	1 female + 2 sex
151	3 males/1 female	3 males	1 female	4 total, 1 male	3 sex
152	1 female	1 female		1 female	
153	2 males/2 females	2 males/1 female	1 female	3 total, 1 male	1 female + 2 sex
154	2 males/3 females	2 males/2 females	1 female	3 total, 1 male	2 females + 2 sex
155	3 males/2 females	3 males/1 female	1 female	4 total, 1 male	1 female + 3 sex
156	2 males/2 females	2 males/1 female	1 female	3 total, 1 male	1 female + 2 sex
157	3 males/1 female	3 males	1 female	4 total, 1 male	3 sex
158	2 males/2 females	2 males/1 female	1 female	3 total, 1 male	1 female + 2 sex
159	2 males/3 females	2 males/2 females	1 female	4 total, 1 male	1 female + 3 sex
160	2 females	2 females		2 females	
161	2 males/2 females	2 males/1 female	1 female	3 total, 1 male	1 female + 2 sex
162	2 males/3 females	2 males/2 females	1 female	4 total, 1 male	1 female + 3 sex
163	2 females	2 females		2 females	

Using the MIXplex, in no instance was the determination of the minimum number or sex of contributors incorrect; these basic parameters were correctly inferred for all 63 theoretical mixtures of varying sex and number of contributors without reliance on peak heights. Additionally, for this particular set of theoretical mixtures, the absolute number of male contributors was estimated correctly in every case. Upon closer inspection, this outcome was

possible due to the particular DYS alleles present in the male profiles used in this study, which lacked allele sharing at both Y STR markers at the same time. It is possible for male contributors to share the same Y STR alleles at both DYS393 and DYS438, decreasing the minimum number of male contributors inferred. Though Identifiler® analysis allowed correct inference of the minimum number of contributors also, the MIXplex analysis was able to determine this number with more accuracy. Additionally, the only sex information that could be inferred using Identifiler® was the presence or absence of a male.

The maximum difference between the minimum and the actual number and sex of contributors using the MIXplex was 2 females, which occurred with the only 6-person mixture. Identifiler® analysis also differed by these 2 females, and was additionally unable to determine the sex of three contributors to this mixture. For both assays, it is likely that as the total number of contributors increases, the difference between the minimum and the actual number of contributors will increase due to a finite number of possible alleles and the frequency of the most common alleles. A difference between the minimum and the actual contributor combination of 1 female was observed 23 times, or at a frequency of 0.365, using the MIXplex. Consequently, these results indicated that the correct sex and number of contributors was obtained from 62% of the MIXplex profiles from this set of samples. On the other hand, the correct determination of these parameters was only possible for 16% of the profiles generated with the Identifiler® kit. Of note, 6 of the correctly identified profiles (60% of the 16%) were single-source samples.

Conclusions

The use of X chromosomal STRs is greatly increasing in the global forensic DNA typing community. Although standard autosomal STRs are predominately used to generate extremely high random match probabilities for evaluating the significance of a match between a suspect and evidentiary material, there are niche situations where other marker systems are useful to the forensic scientist. For example, Y STRs have been utilized widely in cases where spermatozoa are absent during sexual assault evidence screening. Mitochondrial DNA testing has found an

important role in forensic DNA testing when the quantity and quality of nuclear DNA is compromised.

Depending upon the scenario, X chromosomal STRs can provide the investigator with additional information beyond that of autosomal STRs. In missing persons investigations, X STRs may provide more information than autosomal STRs. In criminal incest investigations, the use of X chromosomal markers would be potentially more informative than any other marker system; Y chromosomal STRs and mtDNA analysis are both useless in this case. Regardless of the situation in which X chromosomal STRs are utilized, knowledge of the mutation rates of the individual markers is essential to the characterization of the marker system as a whole. Mutations can impact conclusions drawn from statistical calculations for any marker system; the mutation rate studies undertaken here will ensure maximum utility of X chromosomal markers in situations such as those described above.

The implementation of the routine use of X STRs should present no major hurdles for the criminal justice system and/or laboratory practice for a number of reasons. First, X STRs would have no difficulty being accepted into the judical system given the universal use of autosomal STRs. Prosecutors are now familiar with the science of STRs and their application; the difference to address is that all of the markers are on one chromosome instead of multiple chromosomes. However, the criminal justice system is already familiar with the concept of linked markers combining to form haplotypes through the acceptance of marker systems such as Y chromosomal STRs and mtDNA. Finally, the generation of statistical information for X chromosomal STRs is based upon that of autosomal STR evidence – each haplotype block frequency is multiplied together to generate one random match probability. It should be noted that before X STR markers are to be used in the criminal justice system, a clearly defined knowledge of the boundaries of each linkage group must be determined. The foundation for this knowledge starts with the linkage analysis performed as part of this study.

Mutation rates

The overall mutation rate for X STR markers observed in this study was similar to that reported for autosomal STRs [59,63,67,148] and Y STRs [55-57,61,62,64,65]. The overall mutation rates from published X STR studies (Table 7) also demonstrated general similarity with previously reported Y and autosomal rates. Since the mechanism of mutation is likely the same for each system, this result was expected. The trend observed in this study towards lower mutation rates in African American populations is in contrast to that observed for Y chromosomal STRs, where a higher mutation rate in African Americans compared with U.S. Caucasians or U.S. Hispanics was detected [61,62].

Through sequencing of regions flanking an STR mutational event and analysis of haplotypes in families with multiple children, it has been demonstrated that one-step mutations are far more frequent than mutations involving a change of more than one repeat unit or a partial repeat unit [59,60,66]. Therefore, when assigning "Origin" and "Result" for each observed inconsistency, one-step mutations and single mutational events were favored, as was consistent with the approach of other mutation rate studies [58,60,63]. As a consequence, multi-step mutations that have been artificially ignored may actually exist in this study, but the overall mutation rate should remain unchanged. For example, the maternal mutational event at DXS9902 in which the obligate paternal allele contained 10 repeats was characterized as resulting from the mutation of the mother's 11 allele into a 10 allele (a loss of one repeat unit; Table 8). However, it is also possible that the maternal 12 allele could be the progenitor of the daughter's 10 allele, requiring a two-step mutational event. Of note, the assignment of "Origin" and "Result" in this case would not be changed, and the scenario nevertheless contributes one mutation to the overall rate. In total and including this example, five mutations (the aforementioned example, the mutation of unknown origin at DXS9902, and the three mutations involving mother-son duos) observed in this study could be alternatively explained by a two-step mutation without requiring an additional mutational event. In all cases, the designation for "Origin" and "Result" as well as the overall number of observed mutations would remain unchanged if a two-step mutation were considered.

All mutations observed in this study could be explained by a change in one repeat unit, which is consistent with the model of strand slippage during replication as the mechanism of microsatellite mutation [54]. In particular, repeat unit gains outnumbered losses in this study. This bias towards microsatellite expansion has been noted in other mutation rate studies [55-58], though both an excess of losses [59,60] as well as equal rates [41,61-64] have been noted by others.

While the number of trinucleotide markers investigated here was too small to make an accurate inference as to the impact of repeat size on mutation rate, previous studies of Y STR mutation rates have indicated a bias towards a higher mutation rate for longer repeat units [57,65]. Additionally though it has been previously noted for autosomal and Y STRs that mutations observed at microvariant and/or compound repeats appeared more common than at simple repeats [55,59,65], this study did not yield the same results.

In agreement with the results of this study, previous studies also found that paternal mutations are more frequent for both autosomal STRs [58-60,63,66,67] as well as X STRs [47,48]. Additionally, the higher paternal X STR mutation rate corroborates the idea that the mechanism of microsatellite mutation may be independent of recombination [55,68] (which is absent within a paternally-inherited X chromosome) and explains the similarity in overall mutation rate for both the gonosomes and autosomes.

Both overall and marker-specific mutation rates were determined for 15 commonly used X STR markers, contributing to the array of data on X STR mutation rates in particular and to X STRs as a useful tool in the forensic arsenal in general. In order to continue to work towards routine use of X STRs for relationship testing in the forensic setting, datasets that combine results from multiple studies serve to maximize the information that can be concluded from individual profiles. As the largest X STR mutation rate study to date, and the only one to investigate U.S. populations, the total number of meioses available to the community for consideration has increased by over 40%. Combining the 20,625 meioses from this study with those from

consolidated published studies yielded a robust dataset of 71,020 meioses for use by the forensic community.

Population database creation

These data serve to greatly increase the amount of X STR information available for U.S. populations, for which only four previous studies exist [18,22,69,70], while simultaneously confirming the potential utility of the chosen markers for use in both kinship and identity testing. These databases provide the basis by which forensic scientists will calculate the statistical value of a match between two DNA profiles. Without this information, the use of X STRs is extremely limited, and the full potential of the marker system in forensic scenarios cannot be realized.

Linkage testing

Because mutations were ignored in the classical analyses, it was thought that the true recombination rate had likely been overestimated. In the family described in Appendix A, regarding the 13 allele at DXS7132 in individual 5 as a 14→13 mutational event rather than recombination between DXS7132 and the markers above and below would eliminate 2 of 18 observed events, for example. However, when comparing the manual analyses to computer-based estimates performed using a maximum likelihood approach taking mutation rates into account, the values were found to be very similar. In all cases, the manually-calculated observed recombination rate fell within the 95% support intervals of the computer-based values except for marker pair DXS6789-DXS7424 where the rate of 0.03 fell just outside the lower limit of the 95% support interval (0.0350). Taken together, the results of this combination of methods indicate a robust estimate of the recombination rate between these 14 marker pairs has been achieved.

The hypotheses of complete linkage within linkage groups and of free recombination between linkage groups were both contradicted by the results of this study. The ultimate goal, however, is to gain a better understanding of how potential linkage between this set of 15 X STR markers should direct likelihood calculations in kinship testing. These preliminary results indicate a need to delve even deeper with more comprehensive analyses. Additional calculations, when and if

they are able to be performed, with the remaining 104 Type II families will likely further strengthen the conclusions in this study.

Mixture multiplex evaluation

The design and potential of the MIXplex combined several key elements of mixture interpretation. Generally, reporting of mixed profiles centers on estimating the minimum number of contributors as well as attempting to assign a sex to the individual contributors in some two-person mixtures. The MIXplex correctly identified the sex and minimum number of contributors in all cases of artificial and theoretical mixtures tested as part of this study, and correctly assigned the actual number and sex of contributors 62% of the time. Currently, with autosomal STRs, sex can only be reliably assigned when both contributors are of the same gender, or the male contributor is the minor component of a male-female mixture [73]. Additional testing, such as a Y STR assay, is usually necessary to correctly infer and confirm these characteristics of a mixture, and the MIXplex offers an additional alternative. Corroboration of the suspected number and/or sex of contributors through this assay could direct future analysis, potentially saving time and money. Pre-screening samples thought to contain multiple contributors with this relatively inexpensive assay to 1.) confirm a mixture is present, and 2.) decide which assay, if any, would be most appropriate could eliminate uninformative testing altogether. Additionally, the MIXplex can clarify situations where the male allele at the amelogenin locus is not amplified due to a deletion on the Y chromosome [74,75] without complete Y STR typing. Moreover, profile subtraction, which is the elimination of alleles from a known contributor (usually female) to the mixture, is simplified in an assay where only four loci of seven markers are found on a female's chromosomes, while all markers are present within a male's chromosomes. Even when a male and a female contributor share alleles, there are an additional four markers at which the male alleles would be the only ones present.

While it is clear from this initial development and characterization study that a gonosomal marker multiplex cannot solve all of the questions surrounding the interpretation of a mixed profile, there are benefits to their use in certain situations that justify continued study. Additional optimization of this or a similar multiplex in combination with further

characterization of assay parameters such as the reporting threshold and stutter ratios would be helpful to increase confidence in allele calls. Analysis of additional mixtures, both theoretical and actual, could illustrate both the strengths and the limitations of the current combination of markers, as well as suggest additional configurations that might aid interpretation even further. Casework mixtures should eventually be evaluated with a final assay in order to assess its performance and value in real-world settings.

Two of the key elements of mixture analysis rest in the initial determination of the minimum number of contributors to the mixed profile as well as the sex of these contributors [73], which the MIXplex helped to address. However, the same authors recognize that a standardized mixture interpretation protocol that will be appropriate for every mixed profile an analyst encounters is not feasible. Where the concept of a MIXplex, or an improved version of it, fits into the overall forensic mixture interpretation scheme remains to be uncovered.

References

[1] 1991 report concerning recommendations of the DNA Commission of the International Society for Forensic Haemogenetics relating to the use of DNA polymorphisms, Forensic Sci. Int. 52 (1992) 125-130.

[2] DNA Advisory Board, Quality assurance standards for forensic DNA testing laboratories, Forensic Sci. Comm. 2 (2000).

[3] M.W. Perlin, M.M. Legler, C.E. Spencer, J.L. Smith, W.P. Allan, J.L. Belrose, et al., Validating TrueAllele(R) DNA mixture interpretation, J.Forensic Sci. 56 (2011) 1430-1447.

[4] B. Bekaert, A. Van Geystelen, N. Vanderheyden, M.H. Larmuseau, R. Decorte, Automating a combined composite-consensus method to generate DNA profiles from low and high template mixture samples, Forensic.Sci.Int.Genet. 6 (2012) 588-593.

[5] A. Biedermann, S. Bozza, K. Konis, F. Taroni, Inference about the number of contributors to a DNA mixture: Comparative analyses of a Bayesian network approach and the maximum allele count method, Forensic Sci. Int. Genet. 6 (2012) 689-696.

[6] N. Cerri, U. Ricci, I. Sani, A. Verzeletti, F. De Ferrari, Mixed stains from sexual assault cases: autosomal or Y-chromosome short tandem repeats? Croat. Med. J. 44 (2003) 289-292.

[7] W. Parson, H. Niederstatter, A. Brandstatter, B. Berger, Improved specificity of Y-STR typing in DNA mixture samples, Int. J. Legal Med. 117 (2003) 109-114.

[8] R. Szibor, X-chromosomal markers: past, present and future, Forensic Sci. Int. Genet. 1 (2007) 93-99.

[9] M.T. Zarrabeitia, T. Amigo, C. Sanudo, A. Zarrabeitia, D. Gonzalez-Lamuno, J.A. Riancho, A new pentaplex system to study short tandem repeat markers of forensic interest on X chromosome, Forensic Sci. Int. 129 (2002) 85-89.

[10] D. Athanasiadou, B. Stradmann-Bellinghausen, C. Rittner, K.W. Alt, P.M. Schneider, Development of a quadruplex PCR system for the genetic analysis of X-chromosomal STR loci, Int. Congr. Ser. 1239 (2003) 311-314.

[11] L. Henke, M. Dulmer, J. Henke, A pentaplex PCR assay for the genetic analysis of ChrX short tandem repeat (STR) loci, Int. Congr. Ser. 1261 (2004) 266-268.

[12] C. Bini, S. Ceccardi, G. Ferri, S. Pelotti, M. Alu, E. Roncaglia, et al., Development of a heptaplex PCR system to analyse X-chromosome STR loci from five Italian population samples. A collaborative study, Int. Congr. Ser. 1261 (2004) 272-274.

[13] M. Poetsch, H. Petersmann, A. Repenning, E. Lignitz, Development of two pentaplex systems with X-chromosomal STR loci and their allele frequencies in a northeast German population, Forensic Sci. Int. 155 (2005) 71-76.

[14] K.A. Tabbada, M.C. De Ungria, L.P. Faustino, D. Athanasiadou, B. Stradmann-Bellinghausen, P.M. Schneider, Development of a pentaplex X-chromosomal short tandem repeat typing system and population genetic studies, Forensic Sci. Int. 154 (2005) 173-180.

[15] H. Asamura, H. Sakai, K. Kobayashi, M. Ota, H. Fukushima, MiniX-STR multiplex system population study in Japan and application to degraded DNA analysis, Int. J. Legal Med. 120 (2006) 174-181.

[16] C. Robino, A. Giolitti, S. Gino, C. Torre, Development of two multiplex PCR systems for the analysis of 12 X-chromosomal STR loci in a northwestern Italian population sample, Int. J. Legal Med. 120 (2006) 315-318.

[17] H. Asamura, H. Sakai, M. Ota, H. Fukushima, Japanese population data for eight X-STR loci using two new quadruplex systems, Int. J. Legal Med. 120 (2006) 303-309.

[18] I. Gomes, M. Prinz, R. Pereira, C. Meyers, R.S. Mikulasovich, A. Amorim, et al., Genetic analysis of three US population groups using an X-chromosomal STR decaplex, Int. J. Legal Med. 121 (2007) 198-203.

[19] S. Turrina, R. Atzei, G. Filippini, D. De Leo, Development and forensic validation of a new multiplex PCR assay with 12 X-chromosomal short tandem repeats, Forensic.Sci.Int.Genet. 1 (2007) 201-204.

[20] E.M. Ribeiro Rodrigues, F.P. Leite, M.H. Hutz, J. Palha Tde, A.K. Ribeiro dos Santos, S.E. dos Santos, A multiplex PCR for 11 X chromosome STR markers and population data from a Brazilian Amazon Region, Forensic Sci. Int. Genet. 2 (2008) 154-158.

[21] L. Gusmao, C. Alves, P. Sanchez-Diz, M.T. Zarrabeitia, M.A. Abovich, I. Aragon, et al., Results of the GEP-ISFG collaborative study on an X-STR Decaplex, Forensic Sci. Int. Genet. Suppl. Ser. 1 (2008) 677-679.

[22] T.M. Diegoli, M.D. Coble, Development and characterization of two mini-X chromosomal short tandem repeat multiplexes, Forensic Sci. Int. Genet. 5 (2011) 415-421.

[23] T.M. Diegoli, L. Kovacevic, N. Pojskic, M.D. Coble, D. Marjanovic, Population study of fourteen X chromosomal short tandem repeat loci in a population from Bosnia and Herzegovina, Forensic Sci. Int. Genet. 5 (2011) 350-351.

[24] R. Szibor, S. Lautsch, I. Plate, N. Beck, Population data on the X chromosome short tandem repeat locus HumHPRTB in two regions of Germany, J. Forensic Sci. 45 (2000) 231-233.

[25] J. Edelmann, R. Szibor, DXS101: a highly polymorphic X-linked STR, Int. J. Legal Med. 114 (2001) 301-304.

[26] S. Hering, E. Kuhlisch, R. Szibor, Development of the X-linked tetrameric microsatellite marker HumDXS6789 for forensic purposes, Forensic Sci. Int. 119 (2001) 42-46.

[27] J. Edelmann, S. Hering, E. Kuhlisch, R. Szibor, Validation of the STR DXS7424 and the linkage situation on the X-chromosome, Forensic Sci. Int. 125 (2002) 217-222.

[28] M.T. Zarrabeitia, T. Amigo, C. Sanudo, M.M. de Pancorbo, J.A. Riancho, Sequence structure and population data of two X-linked markers: DXS7423 and DXS8377, Int. J. Legal Med. 116 (2002) 368-371.

[29] D. Huang, Q. Yang, C. Yu, R. Yang, Development of the X-linked tetrameric microsatellite markers HumDXS6803 and HumDXS9895 for forensic purpose, Forensic Sci. Int. 133 (2003) 246-249.

[30] R. Szibor, M. Krawczak, S. Hering, J. Edelmann, E. Kuhlisch, D. Krause, Use of X-linked markers for forensic purposes, Int. J. Legal Med. 117 (2003) 67-74.

[31] P. Wiegand, B. Berger, J. Edelmann, W. Parson, Population genetic comparisons of three X-chromosomal STRs, Int. J. Legal Med. 117 (2003) 62-65.

[32] K.J. Shin, B.K. Kwon, S.S. Lee, J.E. Yoo, M.J. Park, U. Chung, et al., Five highly informative X-chromosomal STRs in Koreans, Int. J. Legal Med. 118 (2004) 37-40.

[33] S. Turrina, D. De Leo, Population genetic comparisons of three X-chromosomal STRs (DXS7132, DXS7133 and GATA172D05) in North and South Italy, Int. Congr. Ser. 1261 (2004) 302-304.

[34] W. Pepinski, M. Skawronska, A. Niemcunowicz-Janica, E. Koc-Zorawska, J. Janica, I. Soltyszewski, Polymorphism of four X-chromosomal STRs in a Polish population sample, Forensic Sci. Int. 151 (2005) 93-95.

[35] W. Pepinski, A. Niemcunowicz-Janica, M. Skawronska, J.R. Janica, E. Koc-Zorawska, J. Janica, et al., Polymorphism of four X-chromosomal STRs in a religious minority of Old Believers residing in northeastern Poland, Int. Congr. Ser. 1288 (2006) 307-309.

[36] W.M. Tang, K.Y. To, Four X-chromosomal STRs and their allele frequencies in a Chinese population, Forensic Sci. Int. 162 (2006) 64-65.

[37] W. Pepinski, A. Niemcunowicz-Janica, M. Skawronska, E. Koc-Zorawska, J. Janica, J. Berent, et al., X-chromosomal polymorphism data for the ethnic minority of Polish Tatars and the religious minority of Old Believers residing in northeastern Poland, Forensic Sci. Int. Genet. 1 (2007) 212-214.

[38] A. Zalan, A. Volgyi, M. Jung, O. Peterman, H. Pamjav, Hungarian population data of four X-linked markers: DXS8378, DXS7132, HPRTB, and DXS7423, Int. J. Legal Med. 121 (2007) 74-77.

[39] D. Becker, H. Rodig, C. Augustin, J. Edelmann, F. Gotz, S. Hering, et al., Population genetic evaluation of eight X-chromosomal short tandem repeat loci using Mentype Argus X-8 PCR amplification kit, Forensic Sci. Int. Genet. 2 (2008) 69-74.

[40] T. Hundertmark, S. Hering, J. Edelmann, C. Augustin, I. Plate, R. Szibor, The STR cluster DXS10148-DXS8378-DXS10135 provides a powerful tool for X-chromosomal haplotyping at Xp22, Int. J. Legal Med. 122 (2008) 489-492.

[41] Q.L. Liu, D.J. Lv, X.L. Wu, H.Y. Sun, X.Y. Wu, H.L. Lu, Development of a five ChX STRs loci typing system, Int. J. Legal Med. 122 (2008) 261-265.

[42] A. Pico, A. Castillo, C. Vargas, A. Amorim, L. Gusmao, Genetic profile characterization and segregation analysis of 10 X-STRs in a sample from Santander, Colombia, Int. J. Legal Med. 122 (2008) 347-351.

[43] M.A. Tariq, O. Ullah, S.A. Riazuddin, S. Riazuddin, Allele frequency distribution of 13 X-chromosomal STR loci in Pakistani population, Int. J. Legal Med. 122 (2008) 525-528.

[44] L.A. Glesmann, P.F. Martina, R.L. Vidal, C.I. Catanesi, Mutation rate of 7 X-STRs of common use in population genetics, J. Basic Appl. Genet. 20 (2009) 37-41.

[45] A. Nadeem, M.E. Babar, M. Hussain, M.A. Tahir, Development of pentaplex PCR and genetic analysis of X chromosomal STRs in Punjabi population of Pakistan, Mol. Biol. Rep. 36 (2009) 1671-1675.

[46] M. Poetsch, D. El-Mostaqim, F. Tschentscher, E.N. Browne, C. Timmann, R.D. Horstmann, et al., Allele frequencies of 11 X-chromosomal loci in a population sample from Ghana, Int. J. Legal Med. 123 (2009) 81-83.

[47] C. Tomas, V. Pereira, N. Morling, Analysis of 12 X-STRs in Greenlanders, Danes and Somalis using Argus X-12, Int. J. Legal Med. 126 (2012) 121-128.

[48] S. Tetzlaff, R. Wegener, I. Lindner, Population genetic investigation of eight X-chromosomal short tandem repeat loci from a northeast German sample, Forensic Sci. Int. Genet. 6 (2012) e155-e156.

[49] Q.L. Liu, D.J. Lu, L. Quan, Y.F. Chen, M. Shen, H. Zhao, Development of multiplex PCR system with 15 X-STR loci and genetic analysis in three nationality populations from China, Electrophoresis. 33 (2012) 1299-1305.

[50] M. Castaneda, V. Mijares, J.A. Riancho, M.T. Zarrabeitia, Haplotypic blocks of X-linked STRs for forensic cases: study of recombination and mutation rates, J. Forensic Sci. 57 (2012) 192-195.

[51] T. Nishi, A. Kurosu, Y. Sugano, J. Kaminiwa, Y. Sekine, S. Yano, et al., Application of a novel multiplex polymerase chain reaction system for 12 X-chromosomal short tandem repeats to a Japanese population study, Leg. Med. (Tokyo). 15 (2013) 43-46.

[52] K.L. O'Connor, A.O. Tillmar, Effect of linkage between vWA and D12S391 in kinship analysis, Forensic Sci. Int. Genet. 6 (2012) 840-844.

[53] M. Nothnagel, R. Szibor, O. Vollrath, C. Augustin, J. Edelmann, M. Geppert, et al., Collaborative genetic mapping of 12 forensic short tandem repeat (STR) loci on the human X chromosome, Forensic Sci. Int. Genet. 6 (2012) 778-784.

[54] G. Levinson, G.A. Gutman, Slipped-strand mispairing: a major mechanism for DNA sequence evolution, Mol. Biol. Evol. 4 (1987) 203-221.

[55] M. Kayser, L. Roewer, M. Hedman, L. Henke, J. Henke, S. Brauer, et al., Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs, Am. J. Hum. Genet. 66 (2000) 1580-1588.

[56] L. Gusmao, P. Sanchez-Diz, F. Calafell, P. Martin, C.A. Alonso, F. Alvarez-Fernandez, et al., Mutation rates at Y chromosome specific microsatellites, Hum. Mutat. 26 (2005) 520-528.

[57] B.M. Dupuy, M. Stenersen, T. Egeland, B. Olaisen, Y-chromosomal microsatellite mutation rates: differences in mutation rate between and within loci, Hum. Mutat. 23 (2004) 117-124.

[58] J.L. Weber, C. Wong, Mutation of human short tandem repeats, Hum. Mol. Genet. 2 (1993) 1123-1128.

[59] E.M. Dauber, A. Kratzer, F. Neuhuber, W. Parson, M. Klintschar, W. Bar, et al., Germline mutations of STR-alleles include multi-step mutations as defined by sequencing of repeat and flanking regions, Forensic Sci. Int. Genet. 6 (2012) 381-386.

[60] B. Brinkmann, M. Klintschar, F. Neuhuber, J. Huhne, B. Rolf, Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat, Am. J. Hum. Genet. 62 (1998) 1408-1415.

[61] A.E. Decker, M.C. Kline, J.W. Redman, T.M. Reid, J.M. Butler, Analysis of mutations in father-son pairs with 17 Y-STR loci, Forensic Sci. Int. Genet. 2 (2008) e31-e35.

[62] J. Ge, B. Budowle, X.G. Aranda, J.V. Planz, A.J. Eisenberg, R. Chakraborty, Mutation rates at Y chromosome short tandem repeats in Texas populations, Forensic Sci. Int. Genet. 3 (2009) 179-184.

[63] E. Santos de Andrade, A.V. Gomes, G. Raposo, L. Mauricio-da-Silva, R. dos Santos Silva, Mutation rates at 14 loci in the population from Perambuco Northeast Brazil, Forensic Sci. Int. Genet. Suppl. Ser. 3 (2009) e141-e143.

[64] M. Goedbloed, M. Vermeulen, R.N. Fang, M. Lembring, A. Wollstein, K. Ballantyne, et al., Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFlSTR Yfiler PCR amplification kit, Int. J. Legal Med. 123 (2009) 471-482.

[65] K.N. Ballantyne, M. Goedbloed, R. Fang, O. Schaap, O. Lao, A. Wollstein, et al., Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications, Am. J. Hum. Genet. 87 (2010) 341-353.

[66] M. Muller, U. Sibbing, C. Hohoff, B. Brinkmann, Haplotype-assisted characterization of germline mutations at short tandem repeat loci, Int. J. Legal Med. 124 (2010) 177-182.

[67] AABB Relationship Testing Program Unit, Annual Report Summary for Testing in 2008, 2008,.

[68] H. Ellegren, Microsatellites: simple sequences with complex evolution, Nat. Rev. Genet. 5 (2004) 435-445.

[69] A. Edwards, A. Civitello, H.A. Hammond, C.T. Caskey, DNA typing and genetic mapping with trimeric and tetrameric tandem repeats, Am. J. Hum. Genet. 49 (1991) 746-756.

[70] T.M. Diegoli, A. Linacre, P.M. Vallone, J.M. Butler, M.D. Coble, Allele frequency distribution of twelve X-chromosomal short tandem repeat markers in four U.S. population groups, Forensic Sci. Int. Genet. Suppl. Ser. 3 (2011) e481-e483.

[71] W.J. Kent, BLAT--the BLAST-like alignment tool, Genome Res. 12 (2002) 656-664.

[72] R. Szibor, S. Hering, J. Edelmann, A new Web site compiling forensic chromosome X research is now online, Int. J. Legal Med. 120 (2006) 252-254.

[73] B. Budowle, A.J. Onorato, T.F. Callaghan, A. Della Manna, A.M. Gross, R.A. Guerrieri, et al., Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework, J. Forensic Sci. 54 (2009) 810-821.

[74] V.K. Kashyap, S. Sahoo, T. Sitalaximi, R. Trivedi, Deletions in the Y-derived amelogenin gene fragment in the Indian population, BMC Med. Genet. 7 (2006) 37.

[75] M. Steinlechner, B. Berger, H. Niederstatter, W. Parson, Rare failures in the amelogenin sex test, Int. J. Legal Med. 116 (2002) 117-120.

[76] D.A. Silva, F.S.N. Manta, M. Desiderio, C. Tavares, E.F. de Carvalho, Paternity testing involving human remains identification and putative half sister: Usefulness of an X-hexaplex STR markers, Forensic Sci. Int. Genet. Suppl. Ser. 2 (2009) 230-231.

[77] K.S. Andreaggi, R.S. Just, T.M. Diegoli, O. Loreille, L.M. Stagnitto, J.A. Irwin, et al., Application of modified STR amplification protocols to commingled remains from the USS Oklahoma. Presented at 62nd Annual Meeting of the American Academy of Forensic Sciences, Seattle, WA, 2010.

[78] N. Pinto, L. Gusmao, A. Amorim, X-chromosome markers in kinship testing: a generalisation of the IBD approach identifying situations where their contribution is crucial, Forensic Sci. Int. Genet. 5 (2011) 27-32.

[79] R. Sparkes, C. Kimpton, S. Watson, N. Oldroyd, T. Clayton, L. Barnett, et al., The validation of a 7-locus multiplex STR test for use in forensic casework. (I). Mixtures, ageing, degradation and species studies, Int. J. Legal Med. 109 (1996) 186-194.

[80] B.E. Krenke, L. Viculis, M.L. Richard, M. Prinz, S.C. Milne, C. Ladd, et al., Validation of a male-specific, 12-locus fluorescent short tandem repeat (STR) multiplex, Forensic Sci. Int. 148 (2005) 1-14.

[81] B.E. Krenke, A. Tereba, S.J. Anderson, E. Buel, S. Culhane, C.J. Finis, et al., Validation of a 16-locus fluorescent multiplex system, J. Forensic Sci. 47 (2002) 773-785.

[82] J.J. Mulero, C.W. Chang, L.M. Calandro, R.L. Green, Y. Li, C.L. Johnson, et al., Development and validation of the AmpFlSTR Yfiler PCR amplification kit: a male specific, single amplification 17 Y-STR multiplex system, J. Forensic Sci. 51 (2006) 64-75.

[83] J.M. Wallin, C.L. Holt, K.D. Lazaruk, T.H. Nguyen, P.S. Walsh, Constructing universal multiplex PCR systems for comparative genotyping, J. Forensic Sci. 47 (2002) 52-65.

[84] Qiagen, Investigator Argus X-12 Handbook, (2010).

[85] P. Wiegand, M. Kleiber, Less is more--length reduction of STR amplicons using redesigned primers, Int. J. Legal Med. 114 (2001) 285-287.

[86] J.M. Butler, Y. Shen, B.R. McCord, The development of reduced size STR amplicons as tools for analysis of degraded DNA, J. Forensic Sci. 48 (2003) 1054-1064.

[87] M.D. Coble, J.M. Butler, Characterization of new miniSTR loci to aid analysis of degraded DNA, J. Forensic Sci. 50 (2005) 43-53.

[88] P.M. Schneider, Scientific standards for studies in forensic genetics, Forensic Sci. Int. 165 (2007) 238-243.

[89] J. Edelmann, S. Hering, M. Michael, R. Lessig, D. Deischel, G. Meier-Sundhausen, et al., 16 X-chromosome STR loci frequency data from a German population, Forensic Sci. Int. 124 (2001) 215-218.

[90] A.R. La Spada, E.M. Wilson, D.B. Lubahn, A.E. Harding, K.H. Fischbeck, Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy, Nature. 352 (1991) 77-79.

[91] A comprehensive genetic linkage map of the human genome. NIH/CEPH Collaborative Mapping Group, Science. 258 (1992) 67-86.

[92] J. Weissenbach, G. Gyapay, C. Dib, A. Vignal, J. Morissette, P. Millasseau, et al., A second-generation linkage map of the human genome, Nature. 359 (1992) 794-801.

[93] K.W. Broman, J.C. Murray, V.C. Sheffield, R.L. White, J.L. Weber, Comprehensive human genetic maps: individual and sex-specific variation in recombination, Am. J. Hum. Genet. 63 (1998) 861-869.

[94] T.C. Matise, F. Chen, W. Chen, F.M. De La Vega, M. Hansen, C. He, et al., A second-generation combined linkage physical map of the human genome, Genome Res. 17 (2007) 1783-1786.

[95] F.B. Machado, E. Medina-Acosta, Genetic map of human X-linked microsatellites used in forensic practice, Forensic Sci. Int. Genet. 3 (2009) 202-204.

[96] S. Hering, C. Augustin, J. Edelmann, M. Heidel, J. Dressler, H. Rodig, et al., DXS10079, DXS10074 and DXS10075 are STRs located within a 280-kb region of Xq12 and provide stable haplotypes useful for complex kinship cases, Int. J. Legal Med. 120 (2006) 337-345.

[97] R. Szibor, S. Hering, E. Kuhlisch, I. Plate, S. Demberger, M. Krawczak, et al., Haplotyping of STR cluster DXS6801-DXS6809-DXS6789 on Xq21 provides a powerful tool for kinship testing, Int. J. Legal Med. 119 (2005) 363-369.

[98] J. Edelmann, S. Hering, C. Augustin, S. Kalis, R. Szibor, Validation of six closely linked STRs located in the chromosome X centromere region, Int. J. Legal Med. 124 (2010) 83-87.

[99] J. Edelmann, S. Hering, C. Augustin, U. Immel, R. Szibor, Chromosome X centromere region—Haplotype frequencies for different populations, Forensic Sci. Int. Genet. Suppl. Ser. 2 (2009) 398-399.

[100] A.O. Tillmar, P. Mostad, T. Egeland, B. Lindblom, G. Holmlund, K. Montelius, Analysis of linkage and linkage disequilibrium for eight X-STR markers, Forensic Sci. Int. Genet. 3 (2008) 37-41.

[101] S. Hering, J. Edelmann, C. Augustin, E. Kuhlisch, R. Szibor, X chromosomal recombination--a family study analysing 39 STR markers in German three-generation pedigrees, Int. J. Legal Med. 124 (2010) 483-491.

[102] S. Hering, C. Augstin, J. Edelmann, M. Heidel, J. Drebler, R. Szibor, A cluster of six closely linked STR-markers: Recombination analysis in a 3.6-Mb region at Xq12-13.1, Int. Congr. Ser. 1288 (2006) 289-291.

[103] I.H. Ferreira da Silva, A.G. Barbosa, D.A. Azevedo, P. Sanchez-Diz, L. Gusmao, C.C. Tavares, et al., An X-chromosome pentaplex in two linkage groups: haplotype data in Alagoas and Rio de Janeiro populations from Brazil, Forensic Sci. Int. Genet. 4 (2010) e95-e100.

[104] H. Rodig, F. Kloep, L. Weissbach, C. Augustin, J. Edelmann, S. Hering, et al., Evaluation of seven X-chromosomal short tandem repeat loci located within the Xq26 region, Forensic Sci. Int. Genet. 4 (2010) 194-199.

[105] J. Edelmann, S. Hering, C. Augustin, R. Szibor, Characterisation of the STR markers DXS10146, DXS10134 and DXS10147 located within a 79.1 kb region at Xq28, Forensic Sci. Int. Genet. 2 (2008) 41-46.

[106] K. Liu, S.V. Muse, PowerMarker: an integrated analysis environment for genetic marker analysis, Bioinformatics. 21 (2005) 2128-2129.

[107] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0): an integrated software package for population genetics data analysis, Evol. Bioinform Online. 1 (2005) 47-50.

[108] T. Kishida, W. Wang, M. Fukuda, Y. Tamaki, Duplex PCR of the Y-27H39 and HPRT loci with reference to Japanese population data on the HPRT locus, Nihon Hoigaku Zasshi. 51 (1997) 67-69.

[109] D. Desmarais, Y. Zhong, R. Chakraborty, C. Perreault, L. Busque, Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA), J. Forensic Sci. 43 (1998) 1046-1049.

[110] C.J. Clopper, E.S. Pearson, The use of confidence or fiducial limits illustrated in the case of the binomial, Biometrika. 26 (1934) 404-413.

[111] W.J. Kent, C.W. Sugnet, T.S. Furey, K.M. Roskin, T.H. Pringle, A.M. Zahler, et al., The human genome browser at UCSC, Genome Res. 12 (2002) 996-1006.

[112] S. Rozen, H.J. Skaletsky, Primer3 on the WWW for general users and for biologist programmers, in: Krawetz S., Misener S. (Eds.), Bioinformatics Methods and Protocols: Methods in Molecular Biology, Humana Press, Totowa, NJ, 2000, pp. 365-386.

[113] P.M. Vallone, J.M. Butler, AutoDimer: a screening tool for primer-dimer and hairpin structures, BioTechniques. 37 (2004) 226-231.

[114] M.J. Brownstein, J.D. Carpten, J.R. Smith, Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping, BioTechniques. 20 (1996) 1004-6, 1008-10.

[115] J. Edelmann, R. Szibor, The X-linked STRs DXS7130 and DXS6803, Forensic Sci. Int. 136 (2003) 73-75.

[116] Q.L. Liu, D.J. Lu, J.Z. Zhu, H.L. Lu, Y.M. Luo, Q. Fang, Fluorescent multiplex amplification of three X-STR loci, Yi Chuan Xue Bao. 33 (2006) 1053-1059.

[117] K. Crkvenac Gornik, Z. Grubic, K. Stingl, I. Tonkovic Durisevic, D. Begovic, Application of microsatellite loci on the chromosome X for rapid prenatal detection of the chromosome X numerical abnormalities, Croat.Med.J. 52 (2011) 392-395.

[118] Y. Nakamura, K. Minaguchi, Sixteen X-chromosomal STRs in two octaplex PCRs in Japanese population and development of 15-locus multiplex PCR system, Int. J. Legal Med. 124 (2010) 405-414.

[119] E.M. Rodrigues, J. Palha Tde, A.K. Santos, S.E. Dos Santos, Genetic data of twelve X-STRs in a Japanese immigrant population resident in Brazil, Forensic.Sci.Int.Genet. 4 (2010) e57-8.

[120] R. Sun, Y. Zhu, F. Zhu, W. Kuang, J. Feng, Z. Tian, et al., Genetic polymorphisms of 10 X-STR among four ethnic populations in northwest of China, Mol. Biol. Rep. 39 (2012) 4077-4081.

[121] L. Gusmao, P. Sanchez-Diz, C. Alves, I. Gomes, M.T. Zarrabeitia, M. Abovich, et al., A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: data on 15 Iberian and Latin American populations, Int. J. Legal Med. 123 (2009) 227-234.

[122] C. Nunez, M. Baeta, J.M. Aznar, C. Sosa, Y. Casalod, M. Bolea, et al., Genetic diversity of 10 X chromosome STRs in an admixed population of Nicaragua, Forensic Sci. Int. Genet. 7 (2013) e95-e96.

[123] N. Cerri, A. Verzeletti, F. Gasparini, B. Bandera, F. De Ferrari, Population data for four X-chromosomal STR loci in a population sample from Brescia (nothern Italy), Int. Congr. Ser. 1288 (2006) 286-288.

[124] M.J. Illescas, J.M. Aznar, S. Cardoso, A. Lopez-Oceja, D. Gamarra, J.F. Sanchez-Romera, et al., Genetic characterization of ten X-STRs in a population from the Spanish Levant, Forensic Sci. Int. Genet. 6 (2012) e180-e181.

[125] J. Tie, S. Uchigasaki, S. Oshida, Genetic polymorphisms of eight X-chromosomal STR loci in the population of Japanese, Forensic Sci. Int. Genet. 4 (2010) e105-e108.

[126] C. Li, T. Ma, S. Zhao, S. Zhang, J. Xu, Z. Zhao, et al., Development of 11 X-STR loci typing system and genetic analysis in Tibetan and Northern Han populations from China, Int. J. Legal Med. 125 (2011) 753-756.

[127] M. Poetsch, A. Knop, D. El-Mostaqim, N. Rakotomavo, N. von Wurmb-Schwark, Allele frequencies of 11 X-chromosomal loci of two population samples from Africa, Int. J. Legal Med. 125 (2011) 307-314.

[128] C. Bobillo, A. Sala, L. Gusmao, D. Corach, Genetic analysis of 10 X-STRs in Argentinian population, Forensic Sci. Int. Genet. 5 (2011) e14-e16.

[129] E.M. Ribeiro-Rodrigues, T.D. Palha, E.A. Bittencourt, A. Ribeiro-Dos-Santos, S. Santos, Extensive survey of 12 X-STRs reveals genetic heterogeneity among Brazilian populations, Int. J. Legal Med. 125 (2011) 445-452.

[130] S.H. Shin, J.S. Yu, S.W. Park, G.S. Min, K.W. Chung, Genetic analysis of 18 X-linked short tandem repeat markers in Korean population, Forensic Sci. Int. 147 (2005) 35-41.

[131] R. Szibor, J. Edelmann, S. Hering, I. Gomes, L. Gusmao, Nomenclature discrepancies in the HPRTB short tandem repeat, Int. J. Legal Med. 123 (2009) 185-186.

[132] W. Wu, H. Hao, Q. Liu, Y. Su, X. Zheng, D. Lu, Allele frequencies of seven X-linked STR loci in Chinese Han population from Zhejiang Province, Forensic Sci. Int. Genet. 4 (2009) e41-e42.

[133] T.M. Clayton, J.L. Guest, A.J. Urquhart, P.D. Gill, A genetic basis for anomalous band patterns encountered during DNA STR profiling, J. Forensic Sci. 49 (2004) 1207-1214.

[134] R.L. Huel, L. Basic, K. Madacki-Todorovic, L. Smajlovic, I. Eminovic, I. Berbic, et al., Variant alleles, triallelic patterns, and point mutations observed in nuclear short tandem repeat typing of populations in Bosnia and Serbia, Croat. Med. J. 48 (2007) 494-502.

[135] G. Mertens, M. Gielis, N. Mommers, A. Mularoni, J. Lamartine, H. Heylen, et al., Mutation of the repeat number of the HPRTB locus and structure of rare intermediate alleles, Int. J. Legal Med. 112 (1999) 192-194.

[136] R.C. Allen, J.W. Belmont, Trinucleotide repeat polymorphism at DXS101, Hum. Mol. Genet. 2 (1993) 1508.

[137] A. Edwards, H.A. Hammond, L. Jin, C.T. Caskey, R. Chakraborty, Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups, Genomics. 12 (1992) 241-253.

[138] S. Inturri, S. Menegon, A. Amoroso, C. Torre, C. Robino, Linkage and linkage disequilibrium analysis of X-STRs in Italian families, Forensic Sci. Int. Genet. 5 (2011) 152-154.

[139] H. Pamjav, R. Kugler, A. Zalan, A. Volgyi, Z. Straky, P. Endredy, et al., X chromosomal recombination study in three-generation families in Hungary, Forensic Sci. Int. Genet. 6 (2012) e95-e96.

[140] Q.L. Liu, Z.D. Li, C.T. Li, H. Zhao, Y.D. Wu, Q. Li, et al., X chromosomal recombination-a family study analyzing 26 X-STR Loci in Chinese Han 3-generation pedigrees, Electrophoresis. 34 (2013) 3016-3022.

[141] M. Kayser, A. Caglia, D. Corach, N. Fretwell, C. Gehrig, G. Graziosi, et al., Evaluation of Y-chromosomal STRs: a multicenter study, Int. J. Legal Med. 110 (1997) 125-33, 141-9.

[142] B.M. Dupuy, T. Gedde-Dahl, B. Olaisen, DXYS267: DYS393 and its X chromosome counterpart, Forensic Sci. Int. 112 (2000) 111-121.

[143] D.R. Carvalho-Silva, S.D. Pena, Molecular characterization and population study of an X chromosome homolog of the Y-linked microsatellite DYS391, Gene. 247 (2000) 233-240.

[144] C.M. Ruitberg, D.J. Reeder, J.M. Butler, STRBase: a short tandem repeat DNA database for the human identity testing community, Nucleic Acids Res. 29 (2001) 320-322.

[145] Q. Ayub, A. Mohyuddin, R. Qamar, K. Mazhar, T. Zerjal, S.Q. Mehdi, et al., Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information, Nucleic Acids Res. 28 (2000) e8.

[146] R. Schoske, P.M. Vallone, M.C. Kline, J.W. Redman, J.M. Butler, High-throughput Y-STR typing of U.S. populations with 27 regions of the Y chromosome using two multiplex PCR assays, Forensic Sci. Int. 139 (2004) 107-121.

[147] A.J. Redd, A.B. Agellon, V.A. Kearney, V.A. Contreras, T. Karafet, H. Park, et al., Forensic value of 14 novel STRs on the human Y chromosome, Forensic Sci. Int. 130 (2002) 97-111.

[148] AABB Parentage Testing Standards Program Unit, Annual Report Summary for Testing in 2003, 2004.

Dissemination of Research Findings

Publications

2013 Toni M. Diegoli, Adrian Linacre, Moses S. Schanfield, Michael D. Coble, Mutation rates of 15 X chromosomal short tandem repeat markers, ***submitted to Forensic Science International: Genetics***.

2013 Toni M. Diegoli, Adrian Linacre, Michael D. Coble, Population genetic data for 15 X chromosomal short tandem repeat markers in three U.S. populations, Forensic Science International Genetics 8 (2014) 64-67. Available in advance online via Science Direct at http://dx.doi.org/10.1016/j.fsigen.2013.07.008. *Accepted for publication 11Jul2013.*

2013 Toni M. Diegoli, Adrian Linacre, Michael D. Coble, Characterization of X chromosomal short tandem repeat markers for forensic use, Forensic Sci. Int. Gene. Suppl. (2013), http://dx.doi.org/10.1016/j.fsigss.2013.10.074.

2013 Toni M. Diegoli, Adrian Linacre, Michael D. Coble, A gonosomal marker multiplex to aid in mixture interpretation, Forensic Sci. Int. Gene. Suppl. (2013), http://dx.doi.org/10.1016/j.fsigss.2013.10.095.

2013 Toni M. Diegoli, Chapter 5: Forensic X Chromosomal Short Tandem Repeat Typing, in: Moses Schanfield and Dragan Primorac, (Eds.), Forensic DNA Applications, CRC Press, Boca Raton, Florida, ***submitted February 2013 for publication late 2013/early 2014***.

Presentations

2014 Kimberly S. Andreaggi and Toni M. Diegoli (faculty presenters and organizers) with Timothy P. McMahon and Suni M. Edson (chairs), Development of Emerging DNA

Technologies for Identification: Expanding the Capabilities of a Missing Persons Laboratory (full-day workshop, *accepted*). 66th Annual Scientific Meeting of the American Academy of Forensic Sciences, Seattle, WA, February 17-22, 2014. Including Toni M. Diegoli, "**Characterization of X STRs for Forensic Use**" and "Next Generation Sequencing for the Personnel Accounting Mission" (oral presentations).

2013 Toni M. Diegoli, Michael Nothnagel, Adrian Linacre, Michael Krawczak, Michael D. Coble, A recombination study of 15 X chromosomal short tandem repeat markers using multigenerational family pedigrees (oral presentation). 25th World Congress of the International Society for Forensic Genetics, Melbourne, Victoria, Australia, September 2-7, 2013.

2013 Toni M. Diegoli, Adrian Linacre, Michael D. Coble, MIXplex: can a gonosomal marker multiplex aid in mixture analysis? (poster presentation). 25th World Congress of the International Society for Forensic Genetics, Melbourne, Victoria, Australia, September 2-7, 2013.

2013 Toni M. Diegoli, Adrian Linacre, Michael D. Coble, Characterization of X chromosomal short tandem repeat markers for forensic use: a project overview (poster presentation). 25th World Congress of the International Society for Forensic Genetics, Melbourne, Victoria, Australia, September 2-7, 2013.

2013 Moses S. Schanfield, Daniele Tiesma, Toni Diegoli, Michael Coble, and Michael Crawford. Preliminary report on the anthropology of 15 X STR loci (oral presentation). 8th International Society for Applied Biological Sciences Conference on Forensic, Anthropologic and Medical Genetics, Split, Croatia, June 24-28, 2013.

2013 Moses S. Schanfield, Daniele Tiesma, Toni Diegoli, and Michael Coble. Preliminary report on the anthropology of 15 X STR loci (oral presentation). 82nd Annual Meeting of the American Association of Physical Anthropologists, Knoxville, TN, April 9-13, 2013.

2013 Toni M. Diegoli and Michael D. Coble. A Study of Recombination between 15 X Chromosomal Short Tandem Repeat Markers in Multigenerational Family Pedigrees (poster presentation). 65th Annual Meeting, American Academy of Forensic Sciences, Washington, DC.

2012 Toni M. Diegoli and Michael D. Coble. A Study of Recombination between 15 X Chromosomal Short Tandem Repeat Markers in Multigenerational Family Pedigrees: Preliminary Results (poster presentation). 23rd International Symposium on Human Identification, Nashville, TN, October 16, 2012.

2012 Toni M. Diegoli, Adrian Linacre, and Michael D. Coble. Characterization of X chromosomal short tandem repeat markers for forensic use (oral presentation; invited speaker). 2012 Green Mountain DNA Conference, Burlington, VT, August 3, 2012.

2012 Toni M. Diegoli and Michael D. Coble. Characterization of X chromosomal short tandem repeat markers for forensic use (poster presentation). DNA Research & Development Poster Session, 2012 National Institute of Justice Conference, Arlington, VA, June 19, 2012.

2011 Toni M. Diegoli, Adrian Linacre, Michael D. Coble. Mutation rates of 15 X-chromosomal short tandem repeat markers: Final results of a large-scale study (poster presentation). 24th World Congress of the International Society for Forensic Genetics, Vienna, Austria.

2011 Moses S. Schanfield, Michael Coble, Toni Diegoli. Anthropological usefulness of 15 X chromosome STR loci across four linkage groups (oral presentation). 80th Annual Meeting of the American Association of Physical Anthropologists, Minneapolis, MN.

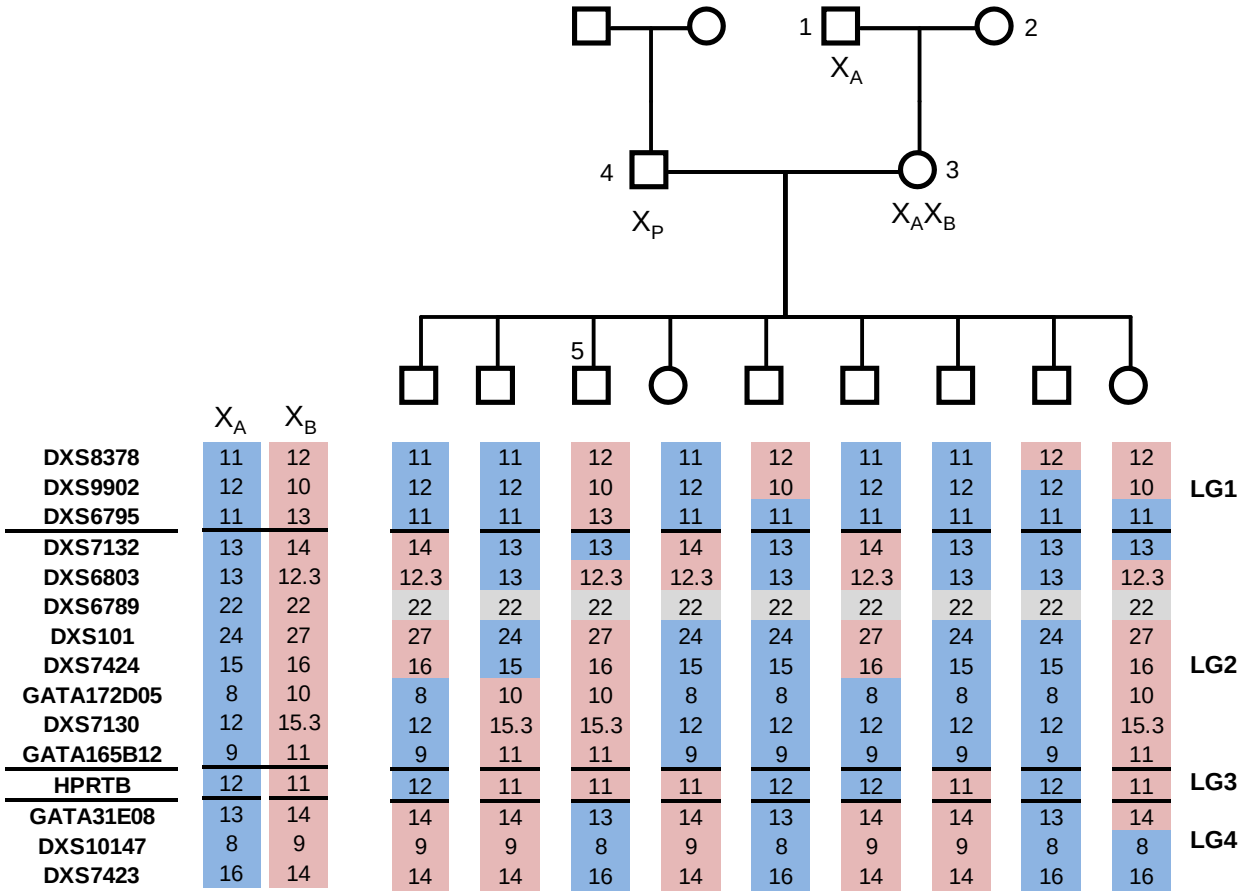
2010 Toni M. Diegoli, Adrian Linacre, and Michael D. Coble. "Analysis of mutation rates in family trio samples for 15 X-chromosomal short tandem repeats" (poster presentation). 20th International Symposium on the Forensic Sciences of the Australian and New Zealand Forensic Science Society, Sydney, Australia.

Acknowledgements

The authors would like to thank Minh Nguyen and the National Institute of Justice; Dr. Moses Schanfield, Dr. Peter Vallone, Dr. John Butler, and Erica Butts for providing sample extracts and associated quantification data; Drs. Michael Nothnagel and Michael Krawczak for the computer-based linkage analysis; Dr. Adrian Linacre for strategic oversight and manuscript review; Patty Czarnecki for assistance in organizing and aliquoting sample extracts; Jessica Saunier for bioinformatics assistance; and Dr. Jodi Irwin, James Canik, Col. Louis Finelli, Dr. Timothy McMahon, and Lt. Col. Laura Regan for logistical and administrative support. The opinions or assertions presented here are the private views of the authors and should not be construed as official or as reflecting the views of the Department of Defense, its branches, the US Army Medical Research and Materiel Command or the Armed Forces Medical Examiner System.

Appendix A. Example of manual recombination assessment using a multigenerational pedigree: methods & challenges.

In this example, a three-generation pedigree is used to determine the phase of each X chromosome and uncover potential recombination events between adjacent markers and linkage groups.



	X_A	X_B										
DXS8378	11	12	11	11	12	11	12	11	11	12	12	
DXS9902	12	10	12	12	10	12	10	12	12	12	10	LG1
DXS6795	11	13	11	11	13	11	11	11	11	11	11	
DXS7132	13	14	14	13	13	14	13	14	13	13	13	
DXS6803	13	12.3	12.3	13	12.3	12.3	13	12.3	13	13	12.3	
DXS6789	22	22	22	22	22	22	22	22	22	22	22	
DXS101	24	27	27	24	27	24	24	27	24	24	27	
DXS7424	15	16	16	15	16	15	15	16	15	15	16	LG2
GATA172D05	8	10	8	10	10	8	8	8	8	8	10	
DXS7130	12	15.3	12	15.3	15.3	12	12	12	12	12	15.3	
GATA165B12	9	11	9	11	11	9	9	9	9	9	11	
HPRTB	12	11	12	11	11	11	12	12	11	12	11	LG3
GATA31E08	13	14	14	14	13	14	13	14	14	13	14	
DXS10147	8	9	9	9	8	9	8	9	9	8	8	LG4
DXS7423	16	14	14	14	16	14	16	14	14	16	16	

Figure A-1. Maternally-inherited X STR haplotypes for each of the nine children are shown in this figure beside the source haplotypes (X_A and X_B) present in the mother (individual 3). Alleles originating from each of the source haplotypes are color-coded and separated into the four linkage groups. Ambiguous source haplotypes (homozygous markers) are shown in gray. X_A: X STR haplotype passed from maternal grandfather (individual 1) in its entirety to mother (individual 3). X_B: X STR haplotype present in mother (individual 3) resulting from recombination of the two X STR haplotypes present in the maternal grandmother (individual 2). X_P: X STR haplotype present in father (individual 4).

Methods

- The profile of the maternal grandfather (individual) 1 is used to determine the source chromosome (X_A and X_B) for each allele present in the mother (individual 3).
- The profile of the father (individual 4) is used to separate the maternally-inherited alleles present in daughters through process of elimination.
- Maternally-inherited X STR haplotypes of children are compared to source haplotypes (X_A and X_B) from the mother (individual 3) and then used to infer potential recombination events between markers.

Challenges

1. Homozygous markers are uninformative. In this example, individual 3 possesses a homozygous 22 allele at marker DXS6789. It is therefore impossible to tell from which source haplotype (X_A or X_B) an offspring's 22 allele originated. Without this knowledge, recombination between DXS6789 and the two bordering markers (DXS6803 & GATA165B12) cannot be assessed. Though only one marker is homozygous in this example, 290 marker pairs (41%) were rendered uninformative in this study due to homozygous genotypes.

2. Mutations must be considered in addition to recombination. In this example, the 13 allele at DXS7132 in the profile of individual 3 was inherited from her father (individual 1). The same allele in individual 5, however, could have been inherited from X_A through two recombination events (between DXS6795 & DXS7132 and between DXS7132 & DXS6803) OR from X_B through a 14→13 mutation. Incorporating information on the rate of mutation could aid in an understanding which scenario is more likely.

3. There is a lack of software tools available. Given the unique inheritance pattern of the X chromosome, software tools designed to accommodate autosomal markers are not necessarily helpful for markers on the X chromosome. An ideal tool must be able to simultaneously accommodate both haploid (male) and diploid (female) genotypes.

Appendix B. Observed haplotypes for U.S. populations. A. Observed haplotypes for proposed linkage group 1. B. Observed haplotypes for proposed linkage group 2. C. Observed haplotypes for proposed linkage group 4. AA: African American, AS: U.S. Asian, CN: U.S. Caucasian, Hisp: U.S. Hispanic, N: number of samples.

A.

DXS8378	DXS9902	DXS6795	AA N=108	CN N=165	Hisp N=150	Overall N=423
8	10	11		1		1
9	9	11		1		1
9	10	9		1		1
9	10	11		1		1
9	11	9		1		1
9	11	13			1	1
9	11	15	1			1
9	11.1	12			1	1
10	8	10	1			1
10	8	11	1			1
10	8	13	1			1
10	9	9		1		1
10	9	11		2		2
10	9	12	1			1
10	9	15	1			1
10	10	9	1	4	7	12
10	10	10	2	1	4	7
10	10	11	2	12	6	20
10	10	12			4	4
10	10	13	2	5	9	16
10	10	14	1			1
10	10	15	2			2
10	11	9	2	7	2	11
10	11	10	2	1	4	7
10	11	11	2	9	8	19
10	11	12	1	1	6	8
10	11	13		4	7	11
10	11	14			1	1
10	11	15	2			2
10	11.1	9	1		1	2
10	11.1	11		1		1
10	11.1	13		1		1
10	12	9		5	1	6
10	12	10			2	2
10	12	11	1	8	1	10
10	12	12		1	2	3
10	12	13		3	2	5
10	12	15	2			2
10	12	16	1			1
10	13	10	1			1
11	8	15	1			1

DXS8378	DXS9902	DXS6795	AA N=108	CN N=165	Hisp N=150	Overall N=423
11	9	10	1			1
11	9	11		1		1
11	9	12	1			1
11	9	13			1	1
11	10	9	1	6	3	10
11	10	10	4		1	5
11	10	11	2	8	3	13
11	10	12	2	1	1	4
11	10	13	1	3	10	14
11	10	14	1		1	2
11	10	15	2	1	1	4
11	11	9	3	6	2	11
11	11	10	2	3	2	7
11	11	11		12	7	19
11	11	12	1	1		2
11	11	13	1	3	7	11
11	11	14			1	1
11	11	15	1			1
11	11.1	9			2	2
11	11.1	13	1	1	1	3
11	12	9		1	1	2
11	12	10	7		2	9
11	12	11	1	4	5	10
11	12	13		1	1	2
11	12	15	2			2
11	13	11	2			2
12	8	9	1			1
12	8	10	1			1
12	9	10	1			1
12	9	13		2		2
12	10	9	1	7	2	10
12	10	10	1		1	2
12	10	11	2	2	3	7
12	10	12	2		4	6
12	10	13	1	2	4	7
12	10	15	4			4
12	11	9	1	6		7
12	11	10	5			5
12	11	11	4	4	6	14
12	11	12	1		1	2
12	11	13	1	1	2	4
12	11	14	1			1
12	11	16	1			1
12	11.1	9		1		1
12	12	9	2	3	1	6
12	12	10	1	1		2
12	12	11	1	3	2	6
12	12	12		1	1	2

DXS8378	DXS9902	DXS6795	AA N=108	CN N=165	Hisp N=150	Overall N=423
12	12	15	3			3
13	8	14	1			1
13	10	9	1	1		2
13	10	11	1	1		2
13	10	13		1		1
13	11	10			1	1
13	11	11		1		1
13	11	13		1		1
13	11.1	9			1	1
13	11.1	11		1		1
13	12	9		1		1
13	12	13	1	1		2
13	12	15	1			1
14	11.1	11		1		1
Total number of haplotypes:			67	58	51	103
Total number of unique haplotypes:			42	32	21	43
Count of most common haplotype:			7	12	10	20

B.

DXS7132	DXS6803	DXS6789	DXS7424	DXS101	GATA172D05	DXS7130	GATA165B12	AA N=108	CN N=165	Hisp N=150	Overall N=423
11	11	15	16	28	7	12	10	1			1
11	11	16	16	24	9	15.3	10	1			1
11	12.3	21	15	26	6	14.3	10	1			1
12	10	20	11	26	6	15.3	12	1			1
12	10	21	16	15	8	13.3	10		1		1
12	11	15	14	18	9	12	11	1			1
12	11	15	14	22	11	12	9	1			1
12	11	16	16	24	6	14.3	11	1			1
12	11	19	11	21	9	12	11	1			1
12	11	19	16	18	11	14.3	10			1	1
12	11	19	16	28	11	11	10			1	1
12	11	19	16	28	11	14.3	11			1	1
12	11	20	16	28	6	12	9			1	1
12	11	21	15	26	10	16.3	11	1			1
12	11	22	14	26	9	14.3	11	1			1
12	11.3	21	14	23	11	14	10			1	1
12	12	14	15	25	11	13.3	9		1		1
12	12	15	14	28	8	15.3	11	1			1
12	12	20	12	26	10	15.3	9			1	1
12	12	20	12	28	6	10	10	1			1
12	12	20	13	26	9	15.3	11	1			1

DXS7132	DXS6803	DXS6789	DXS7424	DXS101	GATA172D05	DXS7130	GATA165B12	AA N=108	CN N=165	Hisp N=150	Overall N=423
12	12	20	13	30	12	14	11	1			1
12	12	20	16	19	10	16.3	10			1	1
12	12	21	15	23	8	16.3	11			1	1
12	12	22	15	23	10	14.3	11			1	1
12	12	22	16	26	9	12	10	1			1
12	12	23	14	21	9	11	10	1			1
12	12.3	21	16	24	8	16.3	10	1			1
12	12.3	22	14	26	10	12	10			1	1
12	13	20	16	28	12	15.3	12		1		1
12	13	21	12	26	11	15.3	10		1		1
12	13	21	16	24	12	14.3	9		1		1
12	13.3	20	16	25	6	12	10		1		1
12	13.3	21	14	24	11	12	10			1	1
12	13.3	21	16	21	9	14.3	10			1	1
12	14	15	15	26	11	12	12	1			1
12	15.3	20	16	24	6	15.3	10		1		1
12	15.3	20	16	25	11	15.3	11		1		1
13	10	16	13	27	9	11	10	1			1
13	10	20	13	27	9	15.3	11	1			1
13	10	21	14	25	9	14.3	10	1			1
13	11	14	16	24	8	14.3	10			1	1
13	11	15	13	19	11	12	11	1			1
13	11	15	13	29	8	13.3	11	1			1
13	11	16	14	29	10	15.3	11	1			1
13	11	16	15	21	11	14.3	11	1			1
13	11	20	12	24	9	15.3	9			1	1
13	11	20	15	24	8	12	10			1	1
13	11	20	16	21	10	16.3	9			1	1
13	11	20	16	22	12	16.3	10		1		1
13	11	20	16	23	6	15.3	10		1		1
13	11	20	16	25	12	15.3	10			1	1
13	11	20	17	18	12	16.3	10		1		1
13	11	21	14	24	6	14.3	10			1	1
13	11	21	15	24	9	13	10			1	1
13	11	21	15	26	10	15.3	12			1	1
13	11	21	16	19	7	12	11	1			1
13	11	21	16	23	10	12	10			1	1
13	11	22	13	26	8	15.3	9	1			1
13	11	22	14	21	11	16.3	12	1			1
13	11	22	14	28	10	14.3	11		1		1
13	11	22	15	21	7	14.3	9	1			1
13	11	22	15	24	10	15.3	9		1		1
13	11	22	16	25	11	13	10		1		1
13	11	22	16	28	12	13	10			1	1

DXS7132	DXS6803	DXS6789	DXS7424	DXS101	GATA172D05	DXS7130	GATA165B12	AA N=108	CN N=165	Hisp N=150	Overall N=423
13	11	22	17	24	8	16.3	10		1		1
13	11.3	16	17	26	11	15.3	11			1	1
13	11.3	20	14	24	11	16.3	10			1	1
13	11.3	20	14	26	6	12	10			1	1
13	11.3	23	13	26	10	16.3	9			1	1
13	12	15	13	29	11	14.3	12	1			1
13	12	15	14	24	8	12	10			1	1
13	12	19	14	24	12	12	10		1		1
13	12	20	13	25	10	12	10			1	1
13	12	20	14	24	10	15.3	11		1		1
13	12	20	14	25	12	12	9			1	1
13	12	20	15	18	10	14.3	10			1	1
13	12	20	15	21	11	14.3	11		1		1
13	12	20	15	24	8	14.3	11		1		1
13	12	20	15	26	6	11	9		1		1
13	12	20	15	27	10	15.3	10		1		1
13	12	20	16	19	10	15.3	11			1	1
13	12	20	16	25	11	12	10			1	1
13	12	20	16	25	12	12	10			1	1
13	12	20	17	20	12	12	11			1	1
13	12	21	13	19	9	13	11	1			1
13	12	21	13	20	8	15.3	11	1			1
13	12	21	13	24	10	15.3	9			1	1
13	12	21	14	21	6	14.3	11	1			1
13	12	21	14	24	8	15.3	10		1		1
13	12	21	15	23	6	15.3	9		1		1
13	12	21	16	25	11	14.3	11		1		1
13	12	21	17	24	11	12	11		1		1
13	12	21	17	25	10	12	10			1	1
13	12	21	17	26	11	14.3	11			1	1
13	12	22	12	19	9	12	10	1			1
13	12	22	13	26	11	15.3	9			1	1
13	12	22	15	27	11	15.3	9		1		1
13	12	22	16	19	11	10	10		1		1
13	12	22	17	25	12	15.3	10		1		1
13	12	23	15	26	11	14.3	12		1		1
13	12	23	16	25	6	14.3	10		1		1
13	12.3	16	13	25	11	12	10			1	1
13	12.3	19	16	18	8	15.3	9			1	1
13	12.3	20	13	26	6	12	9			1	1
13	12.3	20	14	22	8	15.3	11			1	1
13	12.3	20	14	28	10	16.3	9		1		1
13	12.3	20	16	24	8	16.3	11		1		1
13	12.3	20	18	24	6	16.3	10			1	1

DXS7132	DXS6803	DXS6789	DXS7424	DXS101	GATA172D05	DXS7130	GATA165B12	AA N=108	CN N=165	Hisp N=150	Overall N=423
13	12.3	21	15	27	11	12	10		1		1
13	12.3	21	17	25	6	16.3	11			1	1
13	12.3	22	14	23	8	16.3	11			1	1
13	12.3	23	15	24	10	14.3	9		1		1
13	12.3	23	16	20	8	12	11		1		1
13	13	20	16	26	8	15.3	10		1		1
13	13	20	17	25	12	13.3	10		1		1
13	13	21	12	28	10	15.3	9		1		1
13	13	21	14	18	9	15.3	11	1			1
13	13	21	14	20	10	12	10	1			1
13	13	21	15	24	10	15.3	9		1		1
13	13	21	16	24	6	13	9		1		1
13	13	21	16	26	9	15.3	10		1		1
13	13	22	17	24	9	13.3	12		1		1
13	13	23	16	26	6	15.3	9		1		1
13	13.3	15	17	18	12	15.3	10		1		1
13	13.3	20	15	25	11	16.3	11		1		1
13	13.3	20	16	19	11	14.3	9			1	1
13	13.3	21	13	26	10	12	10			1	1
13	13.3	21	15	23	6	15.3	11		1		1
13	13.3	21	15	26	10	15.3	10			1	1
13	13.3	21	16	15	11	15.3	9		1		1
13	13.3	21	18	24	10	13	11			1	1
13	13.3	22	14	26	10	15.3	11		1		1
13	13.3	22	15	25	9	16.3	9		1		1
13	13.3	23	12	24	10	14.3	10			1	1
13	13.3	23	16	24	10	14.3	11		1		1
13	14	20	13	23	10	9	12	1			1
13	14.3	22	17	24	8	15.3	10		1		1
14	8	19	11	21	9	13	10	1			1
14	9	20	14	25	10	15.3	9			1	1
14	9	21	16	26	10	15.3	11	1			1
14	10	15	11	21	9	13	10	1			1
14	10	18	11	21	6	15.3	10	1			1
14	10	20	16	24	10	12	8		1		1
14	10	21	14	24	10	13.3	9		1		1
14	10	21	15	18	6	14.3	9		1		1
14	10	21	17	21	7	14	11	1			1
14	10	23	14	27	11	15.3	10			1	1
14	11	15	12	15	10	15.3	11		1		1
14	11	15	14	25	8	13.3	10		1		1
14	11	15	14	25	9	12	9	1			1
14	11	15	15	21	10	14.3	10	1			1
14	11	15	15	23	9	15.3	10	1			1

DXS7132	DXS6803	DXS6789	DXS7424	DXS101	GATA172D05	DXS7130	GATA165B12	AA N=108	CN N=165	Hisp N=150	Overall N=423
14	11	15	16	22	9	12	11	1			1
14	11	15	17	18	12	13.3	11		1		1
14	11	16	13	25	9	12	10	1			1
14	11	16	13	27	6	15.3	12	1			1
14	11	16	15	27	9	13	11	1			1
14	11	16	17	26	9	15.3	8	1			1
14	11	19	13	26	11	16.3	10			1	1
14	11	19	14	25	9	12	11		1		1
14	11	20	13	26	8	12	10			1	1
14	11	20	14	22	11	14.3	11	1			1
14	11	20	14	24	8	14.3	10			1	1
14	11	20	14	26	8	13	10	1			1
14	11	20	14	27	6	11	11		1		1
14	11	20	16	19	11	12	11	1			1
14	11	20	16	25	10	10	10		1		1
14	11	20	16	25	11	13	9	1			1
14	11	20	16	26	6	15.3	10	1			1
14	11	20	17	19	10	12	11		1		1
14	11	21	13	19	6	13	11	1			1
14	11	21	13	23	10	14.3	10	1			1
14	11	21	14	18	11	14	10			1	1
14	11	21	14	22	7	14.3	10	1			1
14	11	21	14	24	11	12	9			1	1
14	11	21	14	25	9	15.3	10			1	1
14	11	21	14	27	12	13	10		1		1
14	11	21	15	21	8	13	11	1			1
14	11	21	15	25	10	16.3	10		1		1
14	11	21	16	24	12	16.3	9		1		1
14	11	21	16	27	10	16.3	8	1			1
14	11	21	18	24	11	11	10			1	1
14	11	21	19	25	11	15.3	11			1	1
14	11	22	16	18	11	15.3	10	1			1
14	11	22	16	26	8	16.3	9		1		1
14	11.3	20	14	23	8	13	10	1			1
14	11.3	20	14	26	12	15.3	10			1	1
14	11.3	20	16	25	12	15.3	10			1	1
14	11.3	20	16	27	10	15.3	11		1		1
14	11.3	20	17	19	8	16.3	10			1	1
14	11.3	20	17	24	11	13.3	10		1		1
14	11.3	21	14	25	12	12	9		1		1
14	11.3	21	15	27	12	17.3	11		1		1
14	11.3	21	16	25	6	14.3	10		1		1
14	11.3	21	18	24	10	14.3	11		1		1
14	11.3	21	18	24	11	13	10			1	1

DXS7132	DXS6803	DXS6789	DXS7424	DXS101	GATA172D05	DXS7130	GATA165B12	AA N=108	CN N=165	Hisp N=150	Overall N=423
14	11.3	22	17	24	11	15.3	11			1	1
14	12	15	11	23	8	14	11	1			1
14	12	15	13	27	10	11	11	1			1
14	12	15	15	25	6	16.3	9		1		1
14	12	15	16	19	6	14.3	10		1		1
14	12	20	13	19	7	10	10	1			1
14	12	20	14	23	10	15.3	9			1	1
14	12	20	14	24	11	12	9			1	1
14	12	20	15	24	10	15.3	11		1		1
14	12	20	15	25	11	13	11			1	1
14	12	20	15	26	8	16.3	10		1		1
14	12	20	16	25	10	13.3	9		1		1
14	12	20	16	26	10	12	10			1	1
14	12	21	13	27	11	16.3	11		1		1
14	12	21	14	24	11	14.3	11			1	1
14	12	21	14	26	11	15.3	9		1		1
14	12	21	14	31	10	12	10		1		1
14	12	21	15	18	11	14.3	12	1			1
14	12	21	15	26	8	15.3	9		1		1
14	12	21	15	27	12	15.3	9			1	1
14	12	21	16	24	11	14.3	10			1	1
14	12	21	16	24	11	16.3	11			1	1
14	12	21	16	25	10	16.3	10			1	1
14	12	22	11	21	9	12	12	1			1
14	12	22	14	27	10	14.3	11		1		1
14	12	22	15	28	10	15.3	11			1	1
14	12	22	15	29	11	14.3	10			1	1
14	12	22	16	18	6	15.3	11		1		1
14	12	22	16	24	6	13	10			1	1
14	12	22	16	24	11	13	10			1	1
14	12	22	16	25	10	13	11		1		1
14	12	22	17	20	10	15.3	10		1		1
14	12	22	17	24	11	14.3	11		1		1
14	12	22	18	26	10	15.3	11			1	1
14	12	23	14	26	6	12	11			1	1
14	12	23	14	26	12	14.3	9		1		1
14	12	24	14	24	10	15.3	10			1	1
14	12.3	16	13	27	10	12	11	1			1
14	12.3	20	14	26	8	14.3	9			1	1
14	12.3	20	15	26	12	14.3	11			1	1
14	12.3	20	16	18	10	14.3	11			1	1
14	12.3	20	16	24	6	12	10	1			1
14	12.3	20	16	24	12	15.3	10	1			1
14	12.3	20	16	26	6	12	10		1		1

DXS7132	DXS6803	DXS6789	DXS7424	DXS101	GATA172D05	DXS7130	GATA165B12	AA N=108	CN N=165	Hisp N=150	Overall N=423
14	12.3	20	16	26	8	15.3	10			1	1
14	12.3	20	17	18	8	15.3	9			1	1
14	12.3	21	14	27	6	13	10		1		1
14	12.3	21	14	28	11	12	10		1		1
14	12.3	21	15	24	6	14.3	10		1		1
14	12.3	21	15	25	12	16.3	11			1	1
14	12.3	21	16	20	6	15.3	10			1	1
14	12.3	22	12	24	10	15.3	10			1	1
14	12.3	22	16	25	8	15.3	11		1		1
14	12.3	22	17	18	10	15.3	10			1	1
14	12.3	22	17	23	8	15.3	11		1		1
14	12.3	23	16	24	10	14.3	9			1	1
14	12.3	24	15	24	6	12	9		1		1
14	13	15	13	25	7	13	12	1			1
14	13	16	11	27	6	13	11	1			1
14	13	16	17	18	10	14.3	10			1	1
14	13	20	13	25	6	15.3	11		1		1
14	13	20	15	24	11	12	9		1		1
14	13	20	15	25	6	15.3	11		1		1
14	13	20	15	25	10	14.3	11			1	1
14	13	20	15	26	8	16.3	10		1		1
14	13	20	16	15	11	11	10		1		1
14	13	20	16	15	12	15.3	10			1	1
14	13	20	16	21	8	14.3	9		1		1
14	13	20	16	25	8	16.3	11		1		1
14	13	20	17	18	11	15.3	10			1	1
14	13	20	17	24	10	15.3	10		1		1
14	13	21	11	21	9	11	10	1			1
14	13	21	13	25	11	12	10	1			1
14	13	21	13	27	10	16.3	10		1		1
14	13	21	15	24	6	14.3	9		1		1
14	13	21	15	25	11	15.3	10		1		1
14	13	21	17	18	12	16.3	12		1		1
14	13	21	18	25	8	15.3	9			1	1
14	13	22	14	19	8	13	10		1		1
14	13	22	17	18	10	13.3	10		1		1
14	13	23	14	28	6	14.3	11			1	1
14	13	23	17	28	11	15.3	11		1		1
14	13.3	19	13	25	8	14	12		1		1
14	13.3	20	12	30	12	15.3	10		1		1
14	13.3	20	14	27	10	15.3	10		1		1
14	13.3	20	15	23	10	15.3	11		1		1
14	13.3	20	15	26	10	11	9		1		1
14	13.3	20	16	27	12	15.3	10			1	1

DXS7132	DXS6803	DXS6789	DXS7424	DXS101	GATA172D05	DXS7130	GATA165B12	AA N=108	CN N=165	Hisp N=150	Overall N=423
14	13.3	21	13	25	11	15.3	10		1		1
14	13.3	21	14	26	12	15.3	11		1		1
14	13.3	22	14	29	11	16.3	9		1		1
14	13.3	22	16	30	6	15.3	11		1		1
14	14.3	21	14	24	6	15.3	10			1	1
14	15.3	20	13	25	10	12	11		1		1
15	8	18	15	20	9	11	11	1			1
15	9	15	13	19	7	11	11	1			1
15	9	21	13	26	9	10	12	1			1
15	9.3	20	13	26	8	12	10		1		1
15	10	16	13	21	9	12	11	1			1
15	10	16	14	16	11	15.3	10	1			1
15	10	20	14	24	8	13.3	9		1		1
15	10	20	16	18	11	16.3	11		1		1
15	10	23	13	19	9	13	11	1			1
15	10	23	13	19	11	11	10			1	1
15	11	14	15	24	11	15.3	9		1		1
15	11	15	11	23	10	16.3	10			1	1
15	11	15	13	19	10	13	10			1	1
15	11	15	14	19	11	11	11	1			1
15	11	16	14	20	9	11	11	1			1
15	11	19	9	27	9	14	10	1			1
15	11	19	16	24	10	14.3	9			1	1
15	11	19	18	25	10	15.3	10		1		1
15	11	20	13	24	10	15.3	11			1	1
15	11	20	13	25	11	14.3	11			1	1
15	11	20	13	25	11	15.3	9		1		1
15	11	20	15	27	11	13.3	11		1		1
15	11	20	16	25	10	15.3	11	1			1
15	11	20	17	18	12	11	11		1		1
15	11	20	17	24	8	12	9			1	1
15	11	21	11	23	12	15.3	10		1		1
15	11	21	12	25	8	15.3	9	1			1
15	11	21	13	25	6	15.3	10			1	1
15	11	21	13	26	12	12	9			1	1
15	11	21	13	27	8	12	11	1			1
15	11	21	13	27	12	15.3	10	1			1
15	11	21	13	28	10	13	10	1			1
15	11	21	15	27	12	14.3	11		1		1
15	11	21	16	18	8	15.3	9		1		1
15	11	21	16	23	10	14.3	11			1	1
15	11	21	17	23	10	13.3	11			1	1
15	11	21	18	23	9	15.3	10			1	1
15	11	21	18	24	12	12	11			1	1

DXS7132	DXS6803	DXS6789	DXS7424	DXS101	GATA172D05	DXS7130	GATA165B12	AA N=108	CN N=165	Hisp N=150	Overall N=423
15	11	22	13	19	9	15.3	11	1			1
15	11	22	13	25	10	15.3	11			1	1
15	11	22	14	25	12	15.3	9	1			1
15	11	22	16	22	10	15.3	9		1		1
15	11	22	16	22	11	15.3	9			1	1
15	11	23	16	28	11	12	10			1	1
15	11	23	17	27	12	12	10			1	1
15	11	24	14	31	6	15.3	10			1	1
15	11.3	20	16	21	9	15.3	9			1	1
15	11.3	22	16	24	7	15	10			1	1
15	12	14	15	26	10	14.3	10			1	1
15	12	15	13	24	6	13	10	1			1
15	12	15	13	29	10	10	8	1			1
15	12	15	14	27	9	16.3	11	1			1
15	12	15	18	24	10	15.3	10			1	1
15	12	16	13	19	7	15.3	12	1			1
15	12	16	16	25	11	11	9			1	1
15	12	20	13	18	8	15.3	10		1		1
15	12	20	13	25	10	15.3	9			1	1
15	12	20	14	24	10	12	10			1	1
15	12	20	14	26	10	15.3	10			1	1
15	12	20	14	27	9	12	12	1			1
15	12	20	15	27	6	12	11		1		1
15	12	20	16	25	10	15.3	10		1		1
15	12	20	16	25	11	14.3	10			1	1
15	12	20	16	25	12	15.3	9		1		1
15	12	21	14	21	6	12	10	1			1
15	12	21	14	24	8	14.3	12		1		1
15	12	21	15	21	9	17	10	1			1
15	12	21	16	22	9	17.3	11	1			1
15	12	21	16	24	12	16.3	9		1		1
15	12	22	13	24	10	15.3	10			1	1
15	12	22	15	26	11	15.3	10			1	1
15	12	22	16	27	10	15.3	10		1		1
15	12	23	14	26	11	15.3	9		1		1
15	12	24	17	25	8	15.3	10		1		1
15	12.3	16	13	25	11	12	10			1	1
15	12.3	19	14	24	11	16.3	10		1		1
15	12.3	19	17	24	11	15.3	9			1	1
15	12.3	20	10	26	12	15.3	11			1	1
15	12.3	20	13	25	8	15.3	10		1		1
15	12.3	20	14	24	11	14.3	10	1			1
15	12.3	20	14	24	11	15.3	10			1	1
15	12.3	20	16	22	10	14.3	11		1		1

DXS7132	DXS6803	DXS6789	DXS7424	DXS101	GATA172D05	DXS7130	GATA165B12	AA N=108	CN N=165	Hisp N=150	Overall N=423
15	12.3	20	16	26	11	11	11			1	1
15	12.3	21	13	24	10	14.3	10			1	1
15	12.3	21	14	28	11	16.3	10		1		1
15	12.3	21	16	18	11	15.3	11			1	1
15	12.3	21	16	25	11	15.3	10		1		1
15	12.3	22	15	26	10	14.3	10			1	1
15	12.3	24	14	27	10	14.3	11		1		1
15	13	16	11	27	8	17.3	11	1			1
15	13	20	16	24	8	14.3	11			1	1
15	13	20	16	24	8	15.3	10		1		1
15	13	20	17	25	10	15.3	10		1		1
15	13	21	15	26	12	15.3	11		1		1
15	13	21	16	25	12	15.3	10		1		1
15	13	22	14	22	11	13.3	11		1		1
15	13	22	15	26	6	15.3	9			1	1
15	13.3	17	14	28	11	14.3	10			1	1
15	13.3	20	10	26	8	14	11			1	1
15	13.3	20	16	18	6	12	9			1	1
15	13.3	21	14	28	8	14.3	10		1		1
15	13.3	21	15	27	10	15.3	10			1	1
15	13.3	22	12	26	10	17.3	9		1		1
15	13.3	23	15	23	6	15.3	10		1		1
15	14	20	11	21	6	12	9	1			1
16	10	22	16	27	11	17.3	9		1		1
16	11	15	16	29	8	15.3	11	1			1
16	11	20	14	24	10	15.3	10		1		1
16	11	20	15	22	10	14	10			1	1
16	11	21	14	28	9	15.3	11		1		1
16	11	23	16	19	6	12	10	1			1
16	11.3	21	17	23	8	14.3	11			1	1
16	12	15	14	29	10	14.3	10		1		1
16	12	16	17	28	8	14.3	11			1	1
16	12	20	14	25	12	15.3	10		1		1
16	12	20	16	24	6	10	9			1	1
16	12	20	16	25	10	14.3	10			1	1
16	12.3	16	14	24	8	14.3	11	1			1
16	12.3	20	17	25	10	15.3	11		1		1
16	13	20	14	29	11	15.3	11		1		1
16	13	22	13	21	9	10	12	1			1
16	13	22	15	24	8	15.3	11		1		1
16	13.3	21	14	27	11	15.3	10		1		1
16	13.3	21	15	24	12	16.3	11	1			1
17	11	15	12	18	9	13	11	1			1
17	11	20	15	20	10	16.3	11			1	1

DXS7132	DXS6803	DXS6789	DXS7424	DXS101	GATA172D05	DXS7130	GATA165B12	AA N=108	CN N=165	Hisp N=150	Overall N=423
17	11	23	12	25	10	12	11		1		1
17	12	20	14	25	11	16.3	9			1	1
17	12.3	20	13	25	10	15.3	10			1	1
17	13	21	11	23	8	12	10	1			1
17	13.3	21	15	19	8	16.3	11		1		1
18	12	15	14	24	10	16.3	10		1		1
Total number of unique haplotypes:								108	165	150	423
Total number of singletons:								108	165	150	423
Most common haplotype count:								1	1	1	1

C.

GATA31E08	DXS10147	DXS7423	AA N=108	CN N=165	Hisp N=150	Overall N=423
7	7	16	1			1
7	8	15	1			1
7	9	14	1			1
8	6	17			1	1
8	7	15	1			1
8	8	15	1			1
8	9	14	1			1
9	6	14		2	2	4
9	6	15	1	9	6	16
9	6	16		1		1
9	7	14	4	1		5
9	7	15	4		1	5
9	7	17		1		1
9	8	13			1	1
9	8	14	4	2	5	11
9	8	15	1	6	4	11
9	8	16	2	1		3
9	8	17			1	1
9	9	13		3	1	4
9	9	14	1	3		4
9	9	15	3	3	4	10
9	9	16	1	3	1	5
9	9	17		1	1	2
10	6	15			3	3
10	7	14	1			1
10	7	15	1			1
10	7	16		1		1
10	8	14	7	1	1	9
10	8	15	4			4

GATA31E08	DXS10147	DXS7423	AA N=108	CN N=165	Hisp N=150	Overall N=423
10	8	17			2	2
10	9	13			1	1
10	9	14	1			1
10	9	15	1	1		2
10	9	16			1	1
10	10	16	1			1
11	6	14	2	2		4
11	6	15	1	5	4	10
11	7	14			1	1
11	7	15	2		1	3
11	7	16		1		1
11	7	17		1		1
11	8	14	2	5	6	13
11	8	15	1	3	2	6
11	8	16		2	1	3
11	8	17			2	2
11	9	13		2		2
11	9	14	1	4	1	6
11	9	15		5	2	7
11	9	16		1	2	3
12	6	13	2			2
12	6	14	2	1	2	5
12	6	15	1	5	16	22
12	6	16			3	3
12	6	17			1	1
12	7	14	4		2	6
12	7	15	1		2	3
12	7	17		2		2
12	8	14	4	4	11	19
12	8	15	4	7	9	20
12	8	16		1	1	2
12	8	17			1	1
12	9	13		4	2	6
12	9	14	2	5	2	9
12	9	15		4	1	5
12	9	16	1	3	1	5
12	9	17			1	1
12	10	15	1			1
12	10	16	1		1	2
13	6	14	1	1	1	3
13	6	15	1	1	6	8
13	6	16		1	1	2
13	6	17		1		1
13	7	14	5			5
13	7	15	3		1	4
13	7	16		1		1
13	7	17		1	1	2
13	8	13			1	1

GATA31E08	DXS10147	DXS7423	AA N=108	CN N=165	Hisp N=150	Overall N=423
13	8	14	6	5	3	14
13	8	15	3	4	7	14
13	8	16	1	4		5
13	8	17			2	2
13	9	13		5	1	6
13	9	14	3	6	2	11
13	9	15		9	2	11
13	9	16		2		2
13	10	15	1	1		2
13	10	16	1			1
14	6	15		1	2	3
14	7	14	2			2
14	8	13		1		1
14	8	14	2	3	2	7
14	8	15		4	1	5
14	8	16		1		1
14	8	17			1	1
14	9	13	2	2		4
14	9	14			1	1
14	9	15		2	1	3
15	6	14		1		1
15	7	15	1			1
15	9	13		1		1
15	9	14		1		1
15	9	15	1			1
Total number of haplotypes:			54	61	61	102
Total number of unique haplotypes:			30	26	31	38
Count of most common haplotype:			7	9	16	22