New Autosomal and Y-Chromosome STR Loci: Characterization and Potential Uses*

John M. Butler[1], Carolyn R. Hill[1], Amy E. Decker[1], Margaret C. Kline[1], Thomas M. Reid[2], and Peter M. Vallone[1]

[1]Biochemical Science Division, National Institute of Standards and Technology, Gaithersburg MD 20899
[2]DNA Diagnostics Center, Fairfield, OH 45014

 Sources of Support: National Institute of Justice (NIJ) funded this work in part through an interagency agreement 2003-IJ-R-029 with the NIST Office of Law Enforcement Standards. Corresponding Author: john.butler@nist.gov, Tel: 301-975-4049, Fax: 301-975-8505

For a copy of the presentation given, see http://www.cstl.nist.gov/biotech/strbase/pub_pres/Butler_Promega2007_NewSTRs.pdf

Abstract

Additional STR loci can be beneficial for a number of human identity, forensic casework, and DNA database applications. The marker selection and characterization process applied at NIST in developing these new loci and assays are described. New candidate STR markers were selected based on genomic position and measured heterozygosities and allele frequencies in major U.S. populations. For the 26 autosomal STR loci characterized, multiple miniSTR assays have been developed to aid degraded DNA analysis and a megaplex with 22 of the STRs and amelogenin has been created to enable rapid typing of reference samples. Concordance testing results from non-overlapping PCR primers has been performed to test for primer binding site mutations. Mutation rates have been measured with several hundred father/son pairs. STR repeat nomenclature was defined using sequenced alleles and internationally recognized guidelines. We also plan to update NIST SRM 2391b and 2395 certificates with typing information on these new loci to enable calibration of allele calls across laboratories worldwide.

Key Words: forensic DNA typing; STR; short tandem repeat; Y-STRs; NIST; multiplex assays; miniSTRs; degraded DNA; Y-chromosome

Introduction

A common set of core short tandem repeat (STR) markers form the basis of human identity testing performed worldwide and commercially available STR typing kits enable wide-spread use of these core loci (Butler 2006). While the current core loci are sufficient for general forensic matching of evidence to suspect, additional autosomal or Y-chromosomal STR loci can be beneficial or even necessary to address a variety of other human identity/relationship testing questions. At the U.S. National Institute of Standards and Technology (NIST), we are characterizing autosomal and Y-chromosome STR loci that have a number of potential uses.

In casework, additional information can be obtained from degraded DNA samples using miniSTR systems (Butler et al. 2003, Coble and Butler 2005). Another possible use for a limited set of supplemental loci is rapid screening of multiple crime scene samples to identify non-matching samples before applying an expensive commercial STR typing kit with a more complete set of core STR loci. For identity and relationship testing work, kinship analysis such as missing persons/mass disaster sample testing, complex paternity analysis, parentage testing with only one available parent, and immigration testing can benefit from additional genetic markers (Henke and Henke 2005). In fact, a recent article on immigration testing urged the use of 25 genetic loci to avoid erroneous conclusions when a limited number of reference samples are available (Karlson et al. 2007). More loci can help resolve relatives in growing national DNA databases to avoid adventitious matches. For example, although the U.K. National DNA Database started with 6 STR loci (SGM loci-Gill et al. 1996), it expanded to 10 STRs (SGM Plus kit—Cotton et al. 2000) a few years later, and a future pan-European database is expected to include more than 10 STRs (Gill et al. 2006a) including three miniSTRs developed at NIST (Coble and Butler 2005, Gill et al. 2006b).

The current core 13 autosomal STR loci utilized by the FBI Laboratory's CODIS (Combined DNA Index System) and crime laboratories throughout the United States and many other parts of the world were selected from a limited evaluation set of 17 autosomal STRs available from Applied

Biosystems and Promega Corporation in kit or prototype kit form back in 1996 and 1997 (Budowle et al. 1998; Butler 2006). The 13 CODIS core loci are CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, and D21S11. For more information on these STR loci, see http://www.cstl.nist.gov/biotech/strbase/coreSTRs.htm. Since the selection of these core loci, the Human Genome Project has been completed (International Human Genome Sequencing Consortium 2004) and now tens of thousands of STRs are known to exist in the human genome (Butler 2005).

Likewise, for historical reasons, a core set of Y-STR loci are also widely used. The so-called "minimal haplotype" was defined by European laboratories about ten years ago (Kayser et al. 1997, Schneider et al. 1998) and includes DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, and the multi-copy locus DYS385 a/b. In 2003, the Y-chromosome subcommittee of the FBI's Scientific Working Group on DNA Analysis Methods (SWGDAM) added DYS438 and DYS439 to the minimal haplotype so that 11 regions of the Y-chromosome are recommended for use in the United States (Butler 2003; SWGDAM 2004). Commercial Y-STR kits provide the SWGDAM recommended loci plus DYS437 in the case of PowerPlex Y (Promega Corporation, Madison, WI) or the SWGDAM recommended loci plus DYS437, DYS448, DYS456, DYS458, DYS635, and Y-GATA-H4 in the case of Yfiler (Applied Biosystems, Foster City, CA). The dye labels and PCR product size ranges for the loci present in these two widely used commercial Y-STR kits are shown in **Figure 1**. While the Y-STR kits amplify either 12 or 17 loci, over 400 Y-STRs have been located on the Y-chromosome in the past few years (Kayser et al. 2004; Hanson and Ballantyne 2006).

This article will address the work at NIST in characterizing new autosomal and Y-STR loci and also consider the impact of additional loci particularly with Y-STR haplotypes. Given that millions of genotypes now exist with the current autosomal core STR loci (and tens of thousands with core Y-STR loci), it is worth considering some desirable characteristics in potential supplemental STR loci. First, genomic position is important with autosomal STR loci so that adequate spacing from

other “new” loci as well as current STR loci can guarantee independence of inheritance and permit use of the product rule in estimating the rarity of a particular profile. Second, avoiding known disease genes or linkage to disease genes will help protect privacy concerns. Third, more variable genetic markers mean that fewer markers can be used to reach a desired rarity in a full profile and thus selecting markers with a higher heterozygosity (e.g., polymorphic content) is beneficial. Fourth, selecting loci with narrow allele size ranges makes differential amplification less likely and limits the electrophoretic “real-estate” consumed by a particular locus in a multiplex PCR assay. Finally, loci with “clean”, stable flanking regions enable primers to be designed immediately adjacent to the STR repeat region to aid creation of miniSTRs containing reduced sized amplicons. “Clean” flanking regions are DNA sequences without partial repeats or mononucleotide base runs that prevent selection of primers with sufficient annealing temperature to work effectively in our multiplex PCR assays.

The steps we have used in characterizing these new STR loci include (1) selection of the genetic loci from the literature, (2) design of PCR primers to enable genotyping, (3) examination of genetic variation in common U.S. populations, (4) establishment of allele nomenclature through sequencing several homozygote alleles, (5) creation of GeneMapper*ID* bins and panels for genotyping purposes, (6) construction of allelic ladders, (7) evaluation of random match probability with autosomal STR loci or assessment of the ability to separate common types with Y-STR loci, (8) performing mutation rate studies with father-son samples, (9) performing concordance studies with non-overlapping PCR primers to test for possible point mutations that would disrupt primer binding sites, (10) calibration of genotypes with NIST Standard Reference Materials, (11) working with collaborators or companies to evaluate performance of developed assays, and (12) publication of details on loci and assays to enable community use.

Materials and Methods

The autosomal STR work represents an extension of initial studies begun by Coble and Butler (2005) and continued by Hill *et al.* (2008). PCR primer sequences and other details regarding the autosomal 23plex assay will be the subject of a forthcoming publication. Portions of the Y-STR work have appeared previously in Schoske et al. (2004), Butler et al. (2006), Decker et al. (2007), and Decker et al. (2008). The ~660 U.S. population samples used here were first described in Butler et al. (2003a). See http://www.cstl.nist.gov/biotech/strbase/NISTpop.htm for further information on these samples. The father/son sample pairs were obtained from DNA Diagnostics Center (Fairfield, OH) and have been described in Decker et al. (2008). Microsoft Excel macros that have been developed by David Duewer of the NIST Analytical Chemistry Division to aid rapid sample comparisons have been or will be made available in the future at http://www.cstl.nist.gov/biotech/strbase/software.htm.

Results and Discussion

Autosomal STR Work

Selection of Loci

As described previously in Coble and Butler (2005), over 900 autosomal STR loci were examined from the literature in search of trinucleotide and tetranucleotide loci with narrow allele ranges (generally <24 bp), with no known microvariants, and moderate-to-high heterozygosities (generally >0.7) that were located on chromosomes or regions of chromosomes not occupied by current core or commonly used STR loci (Butler 2006). Our most stringent criteria was for loci to possess clean flanking regions where PCR primers could be designed very close to the STR repeat so that miniSTRs could be designed with amplicon sizes of <140 bp. The DNA sequence surrounding the STR repeat with PCR primer positions mapped are shown for three of the selected STR loci in **Figure 2**. A total of 32 autosomal STR loci passed these criteria and were further characterized in the laboratory. Six of these loci dropped out for various reasons that have

been explained in Hill et al. (2008). Chromosomal positions and primer sequences for the final 26 miniSTR loci are listed in **Table 1** and **Table 2**, respectively.

Characterization of Loci

The selected loci were characterized by examining the variation in ~660 U.S. population samples coming from African American, Caucasian, and Hispanic groups. Allele frequencies were determined (**Table 3**) and heterozygosity values calculated (**Table 4**). Over 17,000 genotypes were collected to measure these heterozygosities. Chromosomal positions were precisely defined (**Table 1**), allelic ladders constructed, and standard samples were sequenced and genotyped to provide reference repeat calibration (**Table 5**). We plan to include certified values for these 26 additional autosomal STR loci in future updates to the NIST Standard Reference Material SRM 2391b (PCR-based DNA Profiling Standard).

Using the most common allele frequency for each locus (regardless of population group) and assuming a homozygote at all loci, the theoretically most common type possible with these 26 unlinked autosomal STR loci is 4.5×10^{-20} (**Table 6**) assuming unrelated individuals and no subpopulation structure corrections. Thus, these 26 loci can be combined to create extremely rare STR profiles. Since these STR loci are on separate chromosomes or chromosomal regions compared to currently used core STRs (Butler 2006), the random match probabilities generated from loci in commercial STR kits should be able to be combined using the product rule with the random match probabilities from these additional 26 STR loci.

Megaplex Amplification

With the defined allele ranges characterized, a multiplex assay was developed that is capable of amplifying 22 of the 26 autosomal STRs and small amelogenin X-Y products (Haas-Rochholz and Weiler 1997) for sex-typing purposes. This 23plex, dubbed the “Autoplex” or “miniMegaplex”, uses 5-dye chemistry to keep all PCR products in the size range of 70 to 400 bp (**Figure 3**). This assay works well on reference samples with good quality DNA (typically ~1 ng is used with 30

cycles of PCR). Thus far, this Autoplex has been examined in over 1450 samples representing U.S. Caucasians, African Americans, Hispanics, and Asian individuals in order to perform concordance testing with previously developed miniSTR primers (Hill et al. 2008), to examine mutation rates, and to evaluate performance with extended family samples.

Concordance Studies

In order to compare the allele calls obtained with the multiple miniSTR assays and the overlapping 22 loci present in the Autoplex megaplex amplification, a concordance study was performed. Comparison of allele calls between the miniSTR assays and the Autoplex, which has different PCR primers, found full concordance in 99.80% of the 14,058 genotypes evaluated. In 639 samples containing full Autoplex profiles (22 loci x 639 samples = 14,058 genotypes), only 28 discordant types were observed compared to results obtained when using the miniSTR primers listed in **Table 2**. An example of allele dropout with one of the non-overlapping primers sets for the locus D5S2500 is shown in **Figure 4**. The primer binding site mutation causing this particular allele dropout has not yet been determined through DNA sequencing.

Our PCR amplification concordance rate of 99.80% is similar to the 99.74% concordance rate found in another study when evaluating allele calls from the Identifiler versus the MiniFiler commercial kits (Hill et al. 2007). Thus we conclude that our megaplex and miniplex assay PCR primers have been well-designed and exhibit very limited primer binding site mutations. We plan on sequencing the DNA samples showing differences with the two primer sets to confirm these null alleles. Roughly half of the allele dropout is from the megaplex primers suggesting that the flanking regions immediately next to the STR repeat where the miniSTR primers are located do not appear to have a higher level of mutation with these particular STR loci. Furthermore, we found that the amelogenin primers used (Haas-Rochholz and Weiler 1997) correctly called all male and female samples on over 1450 individuals examined. In fact, one African American sample that exhibits allele dropout for the amelogenin X allele in the Identifiler kit was detected using the amelogenin primers in our Autoplex assay (**Figure 5**).

Mutation Rate Studies

Mutation rates were measured in 395 father/son pairs (790 samples total) across the 22 loci in the Autoplex. With the 8690 possible allelic transfers (22 loci x 395 pairs), only 6 mutations were observed (**Table 7**). This total mutation rate of 0.069% is about 2-3 times less than the typical ~0.1-0.2% mutation rate observed with core STR in current use (see http://www.cstl.nist.gov/biotech/strbase/mutation.htm). The mutation rates with our new autosomal STR loci are likely lower due to the selection of loci for miniSTR applications that exhibited tighter allele ranges, more moderate heterozygosities, and more stable flanking regions for PCR primer annealing near the STR repeat region.

Extended Family Studies

In order to evaluate how well the additional STR loci can improve likelihood ratio calculations in missing persons cases or other forms of relationship testing, a set of extended family samples was tested (**Figure 6**). The 12 samples evaluated cover three generations so that siblings, cousins, uncle/nephew or aunt/niece, and grandparent/grandchild comparisons could be conducted as well as the much easier parent/child analysis. Immigration testing, missing persons or mass disaster kinship analysis involve testing reference samples from extended family members in order to help identify a putative relationship with an immigrant or an unidentified set of remains. **Table 8** illustrates the benefits of the additional 22 STR loci for the various comparisons. Likelihood ratios were calculated as previously described (Reid et al. 2004) using in-house allele frequency databases (e.g., Table 3). A mutation between the mother and the child at one of the 15 STRs present in Identifiler had in this case lowered the likelihood ratio to <1. With the particular alleles present from the additional 22 STRs, the likelihood ratio for this mother/child increased to over 5 million! Note that there was only a minor improvement in the likelihood ratio for longer distance multi-generational relationships (cousins and grandparents/grandchildren) as these types of questions cannot usually be solved with additional autosomal STRs and are best resolved in combination with lineage markers such as mitochondrial DNA or Y-STRs.

Y-Chromosome STR Work

Available Y-STR Data

For the past decade, a common set of Y-STR loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, and the multi-copy locus DYS385 a/b) referred to as the "minimal haplotype" have been widely used around the world. As of September 2007, the Y-Chromosome Haplotype Reference Database (http://www.YHRD.org) contained 50,867 sample results with the minimal haplotype loci from 464 different populations. YHRD contains 23,981 results with the extended SWGDAM-recommended haplotype (minimal haplotype loci + DYS438 and DYS439). The two primary companies providing Y-STR kits, Promega Corporation and Applied Biosystems, also have haplotype database websites containing information on the loci corresponding to their respective Y-STR kits. As of September 2007, there were 4004 12-locus PowerPlex Y haplotypes available at http://www.promega.com/techserv/tools/pplexy/ and 3561 17-locus Yfiler haplotypes available at http://www.appliedbiosystems.com/yfilerdatabase/. The Y-chromosome section of STRBase (http://www.cstl.nist.gov/biotech/strbase/y_strs.htm) contains a listing of links to these and other on-line haplotype databases. The National Institute of Justice has also funded Jack Ballantyne's group at the University of Central Florida to consolidate U.S. data covering the SWGDAM recommended loci. As of August 2007, 14,015 haplotypes have been compiled from various sources such as the PowerPlex Y and Yfiler databases. In addition, as will be discussed in more detail below, a number of population studies have been published in the literature from various world populations that are not yet captured in the on-line haplotype databases and are thus not searchable.

NIST Activities with Y-Chromosome Markers

Since 2000 the NIST Human Identity Project Team has been actively involved in Y-chromosome work and produced 22 publications (see http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm). In July 2003 a Human Y Chromosome Standard Reference Material (SRM 2395) was released (see http://www.cstl.nist.gov/biotech/strbase/srm2395.htm) that certifies 31 Y-STR and 42 Y-SNP

allele calls. In the past few years, we have beta-tested each of the commercial Y-STR kits and are actively sequencing variant Y-STR alleles as they are supplied to us by laboratories around the world (see http://www.cstl.nist.gov/biotech/strbase/STRseq.htm) . We have performed concordance studies with the primers contained in our in-house assays and those of commercial kits. We supplied approximately 20% of the 3561 Yfiler database (our 661 17-locus haplotypes are available at http://www.cstl.nist.gov/biotech/strbase/NISTpop.htm) and have examined mutation rates in 389 father/son pairs (Decker et al. 2008).

In addition to examining the various Y-STR kits, we have studied additional Y-STR loci and explored their potential benefits for the resolution of common haplotypes. Loci suspected to have high diversity values (e.g., Kayer et al. 2004) have been studied in our set of ~660 U.S. population samples representing Caucasians, African Americans, and Hispanics. Thus far, 37 Y-STRs (**Table 9**) have been examined in our full 665 U.S. samples (Butler et al. 2006, Decker et al. 2007) using previously published primers (**Table 10**) and small multiplexes (**Figure 7**). In addition, 92 Y-STRs have now been examined in a representative subset of our samples consisting of 32 Caucasians, 32 African Americans, and 31 Hispanics. Furthermore, we are working to define allele nomenclature for over 140 Y-STR loci to aid on-going genetic genealogy work.

Value of Additional Y-STR Loci

To address the question of potential advantages to typing additional loci beyond the minimal haplotype 9 loci, the SWGDAM 11 loci, the PowerPlex Y 12 loci, and the Yfiler 17 loci, the 37 Y-STR locus haplotypes reported previously (Butler et al. 2006, Decker et al. 2007) were examined to look for the number of unique and shared haplotypes (**Table 11**). The first column in Table 11 shows that 26 of the 656 samples examined possess a “most common type” 9-locus haplotype that is subdivided with additional loci. It is evident that the number of unique haplotypes increases as additional loci are used and that the number of samples sharing haplotypes is also reduced.

With the PowerPlex Y 12 loci, there are 505 unique haplotypes and the most common type is only shared in 12 individuals. The five additional loci present in Yfiler give rise to 626 unique types, 12 haplotypes that are shared twice, and two haplotypes that were observed three times (**Table 11**). Thus, 95% of the 17-locus Yfiler profiles in this NIST-generated U.S. population dataset are unique. The most common type that was shared in 12 individuals with only the 12 PowerPlex Y loci is now subdivided into 9 unique types and one shared by the remaining three individuals when the 17 Yfiler loci are utilized.

In many ways, the situation with Y-STRs is similar to that found with mitochondrial DNA where common types that exist (when only HV1 and HV2 sequence data are considered from the control region) can be subdivided with additional coding region sequence information (Coble et al. 2004). Looking across multiple studies, the best additional Y-STRs for resolving common 9-locus, 12-locus, or 17-locus haplotypes appear to include DYS449, DYS481, DYS570, and DYS576 (Decker et al. 2007; Hanson and Ballantyne 2007; Rodig et al. 2007). One recent study utilized 14 additional Y-STRs in a single multiplex to subdivide all 8 remaining shared haplotypes found following analysis of 17-locus Yfiler types from 572 U.S. Caucasians and African Americans (Hanson and Ballantyne 2007). The 14 Y-STRs used were DYS444, DYS446, DYS449, DYS459a/b, DYS481, DYS508, DYS522, DYS527a/b, DYS549, DYS552, DYS570, DYS576, DYS607, and DYS627.

When considering 12-locus versus 17-locus haplotypes, the five additional loci are useful in resolving many of the shared haplotypes in our NIST U.S. population sample set. Going beyond the 17 Yfiler loci, a few Y-STRs (e.g., DYS576 and DYS522) are more useful than others in sub-dividing the shared haplotypes. However, there are diminishing returns when typing additional loci beyond the 17 available in Yfiler. Not every new locus tested will be helpful. More importantly Y-STR testing in a forensic context will most likely be performed on evidence that is limited in quantity and quality and thus consuming more material to try and make the Y-STR haplotype

more unique is unwise especially since database sizes with additional Y-STRs are limited and estimating a meaningful haplotype frequency with the extended information will be difficult at best.

Yfiler Literature Data Analysis

To see if our NIST dataset containing 95% unique Yfiler profiles is similar to what is observed in other worldwide populations, we rigorously reviewed the forensic literature and also examined the full 3,561 Yfiler haplotype database (http://www.appliedbiosystems.com/yfilerdatabase/). As of late September 2007, a total of 29 Yfiler datasets are available in the literature. These literature results are summarized in **Table 12** according to the number of samples examined and the number of haplotypes observed.

A similar pattern that we previously observed in Table 11 appears to hold across all published studies. There are mostly unique types with about 5% of the 17-locus haplotypes shared by one or more samples usually as shared pairs or triplets. Combining the 29 published sets of data results in 6,893 samples with 6,514 haplotypes (discrimination capacity of 94.5%) of which 6,257 of these are only observed once (96.0% unique). Of course, as we do not have all of the 17-locus profiles and thus cannot compare across studies, it is possible that some of the Yfiler haplotypes may match between studies.

To explore this possibility on a smaller scale, we examined our 656 Yfiler types summarized in Table 11 and compared them with Yfiler profiles from a father/son study (Decker et al. 2008). Only the 389 sons were used in the analysis since the fathers should in principle have the same types as their sons with the exception of mutations (see next section). There was only a slight overlap between the two data sets mainly around the most common European haplotype. With the set of 1,045 NIST samples (656 US pop + 389 sons), there were 1,005 unique 17-locus Yfiler haplotypes, 15 haplotypes seen twice, 2 haplotypes seen three times, and 1 haplotype observed four times. Thus it appears that even in a combined data set, our number of unique haplotypes stays at around 95% (1005/1045 = 96.2%). With the combination of the 3,561 types in the Yfiler

database and the 6893 samples in the literature, a total of 10,454 Yfiler types have been reported worldwide of which ~95% of these complete 17-locus Yfiler haplotypes are unique.

Evaluation of Full 3561 Yfiler Database

An evaluation of the most common types with 9 loci, 11 loci, 12 loci and 17 loci in the 3561 Yfiler database can be seen in **Figure 8**. With the 9 locus minimal haplotype, 124 matching samples are observed (23 of 985 African Americans, 66 of 1276 Caucasians, 1 of 105 Filipinos, 28 of 597 Hispanics, 5 of 106 Native Americans, and 1 of 103 Vietnamese). By adding two additional loci with the 11 locus SWGDAM set, the most common type drops to 60 matching samples. With the 12 PowerPlex Y loci, 51 samples match. By going out to 17 Yfiler loci, only 10 matching samples remain (4 of 985 African Americans, 4 of 1276 Caucasians, and 2 of 597 Hispanics).

An examination of the full 3561 Yfiler database found 3189 unique profiles (89.6%), 132 haplotypes seen in two individuals, 14 seen in three individuals, 10 seen in four individuals, 1 seen in five individuals, 1 seen in 10 individuals (most common type described above), and 1 seen in 11 individuals. The haplotype shared by 11 individuals was a single step mutation at five different loci from the “most common type” shown in Figure 8D.

Y-STR Mutation Rates with 17 Yfiler Loci

A study of the 17 Yfiler loci across 389 father/son sample pairs with approximately 100 each from U.S. Caucasians, African Americans, Hispanics, and Asians found 24 differences between father and son (Decker et al. 2008). Of these mutations, 13 resulted in the gain of a repeat in the son while 11 resulted in the loss of a repeat. All were single repeat mutations except a two repeat loss at the GATA-H4 locus (allele 11 → 9). In addition, we observed two sample pairs that exhibited two mutations across the 17 Y-STRs examined. An African American father/son pair had mutations at DYS458 (allele 18 → 19) and DYS635 (allele 23 → 22) and an Asian father/son pair had mutations at DYS439 (allele 13 → 12) and GATA-H4 (allele 12 → 11). Both of these samples had paternity indices of >1,000,000 with 15 autosomal STRs when the mother was included

(Decker et al. 2008). We also observed four duplications, 1 triplication (DYS448 alleles 17.2, 19, 20), and 4 deletions that were seen in both father and son. **Table 13** shows a summary of Y-STR mutation rates from literature and NIST results for the 17 loci present in the Yfiler kit.

SRM 2395 Updates Planned

In order to enable calibration of Y-STR allele calls with many of these additional loci, we plan to add new information to the NIST SRM 2395 Human Y-Chromosome DNA Profiling Standard certificate for the following loci: DYS635 (to complete Yfiler loci), DYS449, DYS481, DYS570, DYS576, DYS492, DYS522, DYS532, DYS534, DYS572, DYS607, DYS652, DYS709, DYS710, DYS712, DYS715, and DYS717. Information will be available as well on the STRBase website at http://www.cstl.nist.gov/biotech/strbase/srm2395.htm. Many of these additional Y-STRs will enable genetic genealogy companies to support their work with calibrated results (see also http://www.cstl.nist.gov/biotech/strbase/pub_pres/GeneticGenealogy_Y-STR_nomenclature.pdf).

Conclusions

A total of 26 unlinked autosomal STR loci have been characterized across more than 600 samples. These loci have been developed into miniSTR assays with PCR products <140 bp in size to aid recovery of information from degraded DNA. In addition, a megaplex has been created to co-amplify 22 of these 26 loci and amelogenin in a single multiplex. Concordance studies, mutation rates, and population data have been collected using the 23plex assay. Allele nomenclatures have also been assigned for the NIST Standard Reference Material 2391b components enabling calibration of allele calls across laboratories worldwide. Thus far, three of the 26 autosomal STR loci—D10S1248, D2S441, and D22S1045—have been recommended for extending the core European loci (Gill et al. 2006b). Further information on these new loci will be made available on STRBase at http://www.cstl.nist.gov/biotech/strbase/newSTRs.htm.

Studies at NIST and worldwide have shown that approximately 95% of observed 17 locus Yfiler profiles are unique whereas only ~75% are unique when 12 locus PowerPlex Y profiles are extracted from the same datasets. Most of the remaining Yfiler haplotypes are shared by only two or three other samples in a studied dataset. While many of these shared haplotypes can be further subdivided with additional loci, it will likely be unpractical for forensic laboratories to conduct additional tests that provide diminishing returns. Future updates to the NIST SRM 2395 will include information on additional Y-STR loci though to aid other applications of Y-chromosome testing such as genetic genealogy.

Acknowledgments

We thank Mike Coble (now at the Armed Forces DNA Identification Laboratory) for his initial work with the autosomal miniSTR loci and Jan Redman and Rich Schoske for early work in preparing the U.S. population samples. Additionally we are grateful to Dave Duewer of the NIST Analytical Chemistry Division who created a number of Excel macros to enable rapid review of concordance sample evaluation and father-son allele sharing for detection of mutations.

References

Alves, C., Gomes, V., Prata, M. J., Amorim, A., and Gusmao, L. (2007) Population data for Y-chromosome haplotypes defined by 17 STRs (AmpFlSTR YFiler) in Portugal. *Forensic Sci. Int*. 171:250-255.

Barrot, C., Sanchez, C., Xifro, A., Ortega, M., Mas, J., Huguet, E., Corbella, J., Gene, M. (2007) Data for Y-chromosome haplotypes in Fang and Bubi populations from Bioko (Equatorial Guinea). *Forensic Sci. Int.* 168:e10-e12.

Berger, B., Lindinger, A., Niederstatter, H., Grubwieser, P., Parson, W. (2005) Y-STR typing of an Austrian population sample using a 17-loci multiplex PCR assay. *Int. J. Legal Med.* 119:241-246.

Budowle B, Moretti TR, Niezgoda SJ, Brown BL. CODIS and PCR-based short tandem repeat loci: law enforcement tools. Proceedings of the Second European Symposium on Human Identification (June 1998, Innsbruck, Austria). Madison, WI: Promega Corporation. 1998, 73-88; http://www.promega.com/geneticidproc/eusymp2proc/17.pdf

Butler, J.M. (2003) Recent developments in Y-short tandem repeat and Y-single nucleotide polymorphism analysis. *Forensic Sci. Rev.* 15:91-111.

Butler, J.M. (2005) Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers. 2nd Edition. Elsevier Academic Press; New York.

Butler, J.M. (2006) Genetics and genomics of core STR loci used in human identity testing. *J. Forensic Sci.* 51(2): 253-265.

Butler, J.M., Schoske, R., Vallone, P.M., Redman, J.W., Kline, M.C. (2003a) Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations. *J. Forensic Sci.* 48(4):908-911.

Butler, J.M., Shen, Y., McCord, B.R. (2003b) The development of reduced size STR amplicons as tools for analysis of degraded DNA. *J. Forensic Sci* 48(5) 1054-1064.

Butler, J.M., Decker, A.E., Vallone, P.M., Kline, M.C. (2006) Allele frequencies for 27 Y-STR loci with U.S. Caucasian, African American, and Hispanic samples. *Forensic Sci. Int.* 156:250-260.

Butler, J.M. and Coble, M.D. (2007) Authors' Response to Letter to Editor [regarding nomenclature for new miniSTR locus D10S1248]. *J. Forensic Sci.* 52(2): 494.; see http://www.cstl.nist.gov/biotech/strbase/miniSTR.htm#Nomenclature_Errata

Chang, Y.M., Perumal, R., Keat, P.Y., Kuehn, D.L.C. (2007) Haplotype diversity of 16 Y-chromosomal STRs in three main ethnic populations (Malays, Chinese and Indians) in Malaysia. *Forensic Sci. Int.* 167:70-76.

Coble, M.D., Just, R.S., O'Callaghan, J.E., Letmanyi, I.H., Peterson, C.T., Irwin, J.A., Parsons, T.J. (2004) Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. *Int. J. Legal Med.* 118: 137-146.

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

Cotton, E. A., Allsop, R. F., Guest, J. L., Frazier, R. R., Koumi, P., Callow, I. P., Seager1 A, and Sparkes, R. L. (2000) Validation of the AMPFlSTR® SGM plus system for use in forensic casework. *Forensic Sci. Int.* 112(2-3): 151-161.

Decker, A.E., Kline, M.C., Vallone, P.M., Butler, J.M. (2007) The impact of additional Y-STR loci on resolving common haplotypes and closely related individuals. *FSI Genetics* 1:215-217.

Decker, A.E., Kline, M.C., Redman, J.W., Reid, T.M., Butler, J.M. (2008) Analysis of mutations in father-son pairs with 17 Y-STR loci. *FSI Genetics (in press).*

Dixon, L.A., Dobbins, A.E., Pulker, H., Butler, J.M., Vallone, P.M., Coble, M.D., Parson, W., Berger, B., Grubweiser, P., Mogensen, H.S., Morling, N., Nielsen, K., Sanchez, J.J., Petkovski, E., Carracedo, A., Sanchez-Diz, P., Brion, M., Irwin, J.A., Just, R.S., Loreille, O., Parsons, T.J., Syndercombe-Court, D., Schmitter, H., Gill, P. (2006) Analysis of artificially degraded DNA using STRs and SNPs--results of a collaborative European (EDNAP) exercise. *Forensic Sci. Int.* 164: 33-44.

Domingues. P.M., Gusmao, L., Aparecida da Silva, D., Amorim, A., Pereira, R.W., de Carvalho, E.F. (2007) Sub-Saharan Africa descendants in Rio de Janeiro (Brazil): population and mutational data for 12 Y-STR loci. *Int. J. Legal Med.* 121:238-241.

Frigi, S., Pereira, F., Pereira, L., Yacoubi, B., Gusmao, L., Alves, C., Khodjet el Khil, H., Cherni, L., Amorim, A., El Gaaied, A. (2006) Data for Y-chromosome haplotypes defined by 17 STRs (AmpFlSTR Yfiler) in two Tunisian Berber communities. *Forensic Sci. Int.* 160:80-83.

Gill, P., Urquhart, A., Millican, E. S., Oldroyd, N. J., Watson, S., Sparkes, R., and Kimpton, C. P. (1996) A new method of STR interpretation using inferential logic--development of a criminal intelligence database. *Int. J. Legal Med.* 109: 14-22.

Gill, P., Fereday, L., Morling, N., Schneider, P.M. (2006a) The evolution of DNA databases--recommendations for new European loci. *Forensic Sci. Int.* 156:242-244

Gill, P., Fereday, L., Morling, N., Schneider, P.M. (2006b) Letter to editor -- New multiplexes for Europe-amendments and clarification of strategic development. *Forensic Sci Int.* 163:155-157.

Gusmão, L., Butler, J.M., Carracedo, A., Gill, P., Kayser, M., Mayr, W.R., Morling, N., Prinz, M., Roewer, L., Tyler-Smith, C., Schneider, P.M. (2006) DNA Commission of the International Society of Forensic Genetics (ISFG): An update of the recommendations on the use of Y-STRs in forensic analysis. *Forensic Sci. Int.* 157:187-197.

Gutierrez-Alarcon, A. B., Moguel-Torres, M., Leon-Jimenez, A. K., Cuellar-Nevarez, G. E., and Rangel-Villalobos, H. (2007) Allele and haplotype distribution for 16 Y-STRs (AmpFlSTR Y-filer kit) in the state of Chihuahua at North Center of Mexico. *Legal Med. (Tokyo)* 9(3): 154-157.

Haas-Rochholz, H., Weiler, G. (1997) Additional primer sets for an amelogenin gene PCR-based DNA-sex test. *Int. J. Legal Med.* 110(6): 312-315.

Hanson, E.K. and Ballantyne, J. (2006) Comprehensive annotated STR physical map of the human Y chromosome: Forensic implications. *Legal Med. (Tokyo)* 8(2):110-120.

Hanson, E.K. and Ballantyne, J. (2007) An ultra-high discrimination Y chromosome short tandem repeat multiplex DNA typing system. *PLoS ONE* 8:e688. Available at http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0000688

Hara, M., Kido, A., Takada, A., Adachi, N., and Saito, K. (2007) Genetic data for 16 Y-chromosomal STR loci in Japanese. *Legal Med (Tokyo)* 9(3): 161-170.

Henke, J., Henke, L. (2005) Which short tandem repeat polymorphisms are required for identification? Lessons from complicated kinship cases. *Croat Med J*. 46(4):593-597. Available at http://www.cmj.hr/2005/46/4/16100762.pdf

Hill, C.R., Kline, M.C., Mulero, J.J., Lagace, R.E., Chang, C.-W., Hennessy, L.K., Butler, J.M. (2007) Concordance study between the AmpFlSTR MiniFiler PCR Amplification Kit and conventional STR typing kits. *J. Forensic Sci.* 52(4): 870-873.

Hill, C.R., Kline, M.C., Coble, M.D., Butler, J.M. (2008) Characterization of 26 miniSTR loci for improved analysis of degraded DNA samples. *J. Forensic Sci., in press*. (Jan 2008 issue)

Huang, T.-Y., Hsu, Y.-T., Li, J.-M., Chung, J.-H., Shun, C.-T. (2007) Polymorphism of 17 Y-STR loci in Taiwan population. *Forensic Sci Int* (in press).

International Human Genome Sequencing Consortium. (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431:931-945.

Karlsson, A.O., Holmlund, G., Egeland, T., Mostad, P. (2007) DNA-testing for immigration cases: the risk of erroneous conclusions. *Forensic Sci. Int.* 172:144-149.

Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., Penacino, G., Perez-Lezaun, A., Piccinini, A., Prinz, M., Schmitt, C., Schneider, P. M., Szibor, R., Teifel-Greding, J., Weichhold, G. M., de Knijff, P., and Roewer, L. (1997) Evaluation of Y-chromosomal STRs: a multicenter study. *Int. J. Legal Med.* 110(3): 125-133.

Kayser, M., Kittler, R., Ralf, A., Hedman, M., Lee, A. C., Mohyuddin, A., Mehdi, S. Q., Rosser, Z., Stoneking, M., Jobling, M. A., Sajantila, A., and Tyler-Smith, C. (2004) A comprehensive survey of human Y-chromosomal microsatellites. *Am.J Hum Genet*. 74(6): 1183-1197.

Kline, M.C. (2006) NIST SRM Updates: Value-added to the Current Materials in SRM 2391b and SRM 2395. Poster at 17th International Symposium on Human Identification (Nashville, TN), October 10-12, 2006; available at http://www.cstl.nist.gov/biotech/strbase/pub_pres/Promega2006_Kline.pdf

Li, K.L., Kai, L., Yuming, M. (2007) Y chromosome STR haplotypes of Tibetan Living Tibet Lassa. *Forensic Sci. Int.* 172:79-83.

Liu, H.-M., Chen, P.-S., Chen, Y.-J., Lyou, J.-Y., Hu, H.-Y., Lin, J.-S., Tzeng, C.-H. (2007) Y-chromosome short tandem repeats analysis to complement paternal lineage study: a single institutional experience in Taiwan. *Transfusion* 47:918-926.

Mizuno, N., Nakahara, H., Sekiguchi., K., Yoshida, K., Nakano, M., Kasai, K. (2007) 16 Y chromosomal STR haplotypes in Japanese. *Forensic Sci. Int.* (in press).

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

Nicklas, J.A.and Buel, E. (2003) Development of an Alu-based, real-time PCR method for quantitation of human DNA in forensic samples. J Forensic Sci. 48: 936-944.

Onofri, V., Alessandrini, F., Turchi, C., Fraternale, B., Buscemi, L., Pesaresi, M., and Tagliabracci, A. (2007) Y-chromosome genetic structure in sub-Apennine populations of Central Italy by SNP and STR analysis. *Int. J. Legal Med*. 121:234-237.

Pereira, R. W., Monteiro, E. H., Hirschfeld, G. C., Wang, A. Y., and Grattapaglia, D. (2007) Haplotype diversity of 17 Y-chromosome STRs in Brazilians. *Forensic Sci. Int*. 171:226-236.

Reid, T.M., Wolf, C.A., Kraemer, C.M., Lee, S.C., Baird, M.L., Lee, R.F. (2004) Specificity of sibship determination using the ABI Identifiler multiplex system. *J. Forensic Sci.* 49(6): 1262-1264.

Rodig, H., Roewer, L., Gross, A., Richter, T., de Knijff, P., Kayser, M., Brabetz, W. (2007) Evaluation of haplotype discrimination capacity of 35 Y-chromosomal short tandem repeat loci. *Forensic Sci. Int.* (in press).

Roewer, L., Kruger, C., Willuweit, S., Nagy, M., Rodig, H., Kokshunova, L., Rothamel, T., Kravchenko, S., Jobling, M.A., Stoneking, M., Nasidze, I. (2007) Y-chromosomal STR haplotypes in Kalmyk population samples. *Forensic Sci. Int.* (in press).

Sanchez, C., Barrot, C., Xifro, A., Ortega, M., de Aranda, I. G., Huguet, E., Corbella, J., and Gene, M. (2007) Haplotype frequencies of 16 Y-chromosome STR loci in the Barcelona metropolitan area population using Y-Filer kit. *Forensic Sci. Int*. 172(2-3): 211-217.

Schneider, P.M., Meuser, S., Waiyawuth, W., Seo, Y., and Rittner, C. (1998) Tandem repeat structure of the duplicated Y-chromosomal STR locus DYS385 and frequency studies in the German and three Asian populations. *Forensic Sci.Int.* 97(1): 61-70.

Schoske, R., Vallone, P.M., Kline, M.C., Redman, J.W., Butler, J.M. (2004) 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: 107-121.

Soltyszewski, I., Pepinski, W., Spolnicka, M., Kartasinska, E., Konarzewska, M., Janica, J. (2007) Y-chromosomal haplotypes for the AmpFlSTR Yfiler PCR Amplification Kit in a population sample from Central Poland. *Forensic Sci. Int.* 168:61-67.

SWGDAM (2004) Report on the Current Activities of the Scientific Working Group on DNA Analysis Methods Y-STR Subcommittee. *Forensic Sci. Comm.* 6(3); available on-line at http://www.fbi.gov/hq/lab/fsc/backissu/july2004/standards/2004_03_standards03.htm

Turrina, S., Atzei, R., De Leo, D. (2006) Y-chromosomal STR haplotypes in a Northeast Italian population sample using 17plex loci PCR assay. *Int. J. Legal Med.* 120:56-59.

Vallone, P.M. and Butler, J.M. (2004) Y-SNP typing of U.S. African American and Caucasian samples using allele-specific hybridization and primer extension. *J. Forensic Sci.* 49(4): 723-732.

Veselinovic, I.S., Zgonjanin, D.M., Maletin, M.P., Stojkovic, O., Djurendic-Brenesel, M., Vukovic, R.M., Tasic, M.M. (2007) Allele frequencies and population data for 17 Y-chromosome STR loci in a Serbian population sample from Vojvodina province. *Forensic Sci Int* (in press).

Yan, J., Tang, H., Liu, Y., Jing, Y., Jiao, Z., Zhang, Q., Gao, J., Shang, L., Guo, H., Yu, J. (2007) Genetic polymorphisms of 17 Y-STRs haplotypes in Chinese Han population residing in Shandong province of China. *Legal Med.* 9:196-202.

Zhang, Q., Yan, J., Tang, H., Jiao, Z., Liu, Y. (2006) Genetic polymorphisms of 17 Y-STRs haplotypes in Tibetan ethnic minority group of China. *Legal Med.* 8:300-305.

Zhu, B., Shen, C., Xun, X., Yan, J., Deng, Y., Zhu, J., Liu, Y. (2007) Population genetic polymorphisms for 17 Y-chromosomal STRs haplotypes of Chinese Salar ethnic minority group. *Legal Med.* 9:203-209.

Zhu, B. Wu, Y., Shen, C., Yang, T., Deng, Y., Xun, X., Tian, Y., Yan, J., Li, T. (2007) Genetic analysis of 17 Y-chromosomal STRs haplotypes of Chinese Tibetan ethnic group residing in Qinghai province of China. *Forensic Sci. Int.* (in press).

Table 1. Information on 26 autosomal STR loci (Hill *et al.* 2008). Chromosomal locations based on May 2004 NCBI Build 35 of the human genome using BLAT (http://genome.ucsc.edu/cgi-bin/hgBlat).

Locus	GenBank Accession (Allele Repeat #)	Chromosome	Location
D1GATA113	Z97987 (11)	Chr 1 7.377 Mb	1p36.23
D1S1627	AC093119 (13)	Chr 1 106.676 Mb	1p21.1
D1S1677 (NC02)	AL513307 (15)	Chr 1 160.747 Mb	1q23.3
D2S441 (NC02)	AC079112 (12)	Chr 2 68.214 Mb	2p14
D2S1776	AC009475 (10)	Chr 2 169.471 Mb	2q24.3
D3S3053	AC069259 (10)	Chr 3 173.234 Mb	3q26.31
D3S4529	AC117452 (12)	Chr 3 85.935 Mb	3p12.1
D4S2364 (NC02)	AC022317 (9)	Chr 4 93.976 Mb	4q22.3
D4S2408	AC110763 (9)	Chr 4 30.981 Mb	4p15.1
D5S2500	AC008791 (17)	Chr 5 58.735 Mb	5q11.2
D6S474	AL357514 (18)	Chr 6 112.986 Mb	6q21
D6S1017	AL035588 (10)	Chr 6 41.785 Mb	6p21.1
D8S1115	AC090739 (9)	Chr 8 42.656 Mb	8p11.21
D9S1122	AL161789 (12)	Chr 9 76.918 Mb	9q21.2
D9S2157	AL162417 (10)	Chr 9 133.065 Mb	9q34.2
D10S1248 (NC01)	AL391869 (13)	Chr 10 130.567 Mb	10q26.3
D10S1435	AL354747 (10)	Chr 10 2.233 Mb	10p15.3
D11S4463	AP002806 (13)	Chr 11 130.338 Mb	11q25
D12ATA63	AC009771 (14)	Chr 12 106.825 Mb	12q23.3
D14S1434 (NC01)	AL121612 (13)	Chr 14 93.298 Mb	14q32.13
D17S974	AC034303 (10)	Chr 17 10.459 Mb	17p13.1
D17S1301	AC016888 (11)	Chr 17 70.193 Mb	17q25.1
D18S853	AP005130 (11)	Chr 18 3.981 Mb	18p11.31
D20S482	AL121781 (14)	Chr 20 4.454 Mb	20p13
D20S1082	AL158015 (14)	Chr 20 53.299 Mb	20q13.2
D22S1045 (NC01)	AL022314 (14)	Chr 22 35.779 Mb	22q12.3

Table 2. MiniSTR primer sequences used (Hill *et al.* 2008). Bold, underlined "G" used for promoting full adenylation of PCR products. Underlined nucleotides indicate primer binding into the repeat region.

Locus	Forward Primer (5'dye labels shown)	Reverse Primer (extra G on 5'end --> +A)
D1GATA113	**[VIC]** - TCTTAGCCTAGATAGATACTTGCTTCC	**G**TCAACCTTTGAGGCTATAGGAA
D1S1627	**[VIC]** - CATGAGGTTTGCAAATACTATCTTAAC	**G**TTTTAATTTTCTCCAAATCTCCA
D1S1677 (NC02)	**[NED]** - TTCTGTTGGTATAGAGCAGTGTTT	**G**TGACAGGAAGGACGGAATG
D2S441 (NC02)	**[VIC]** - CTGTGGCTCATCTATGAAAACTT	**G**AAGTGGCTGTGGTGTTATGAT
D2S1776	**[FAM]** - TGAACACAGATGTTAAGTGTGTATATG	**G**TCTGAGGTGGACAGTTATGAAA
D3S3053	**[VIC]** - TCTTTGCTCTCATGAATAGATCAGT	**G**TTTGTGATAATGAACCCACTCAG
D3S4529	**[VIC]** - CCCAAAATTACTTGAGCCAAT	**G**AGACAAAATGAAGAAACAGACAG
D4S2364 (NC02)	**[FAM]** - CTAGGAGATCATGTGGGTATGATT	**G**CAGTGAATAAATGAAC<u>GAATGGA</u>
D4S2408	**[NED]** - AAGGTACATAACAGTTCAATAGAAAGC	**G**TGAAATGACTGAAAAATAGTAACCA
D5S2500	**[NED]** - CTGTTGGTACATAATA<u>GGTAGGTAGGT</u>	**G**TCGTGGGCCCCATA<u>AATC</u>
D6S474	**[NED] -** GGTTTTCCAAGAGATAGACCAATTA	**G**TCCTCTCATAAATCCCTACTCATATC
D6S1017	**[VIC]** - CCACCCGTCCATTTAGGC	**G**TGAAAAAGTAGATATAATGGTTGGT<u>G</u>
D8S1115	**[FAM]** - TCCACATCCTCACCAACAC	**G**CCTAGGAAGGCTACTGTCAA
D9S1122	**[VIC]** - GGGTATTTCAAGATAACTGTAGATAGG	**G**CTTCTGAAAGCTTCTAGTTTACC
D9S2157	**[FAM]** - CAAAGCGAGACTCTGTCTCAA	**G**AAAATGCTATCCTCTTTGGTATAAAT
D10S1248 (NC01)	**[FAM]** - TTAATGAATTGAACAAATGAGTGAG	**G**CAACTCTGGTTGTATTGTCTTCAT
D10S1435	**[FAM]** - TGTTATAATGCATTGAGTTTTATTCTG	**G**CCTGTCTCAAAAATAAAGAGATAGACA
D11S4463	**[FAM]** - TCTGGATTGATCTGTCTGTCC	**G**AATTAAATACCATCTGAGCACTGAA
D12ATA63	**[FAM]** - GAGCGAGACCCTGTCTCAAG	**G**GAAAAGACATAGGATAGCAATTT
D14S1434 (NC01)	**[VIC]** - TGTAATAACTCTACGA<u>CTGTCTGTCTG</u>	**G**AATAGGAGGTGGATGGATGG
D17S974	**[VIC]** - GCACCCAAAACTGAATGTCATA	**G**GTGAGAGTGAGACCCTGTC
D17S1301	**[FAM]** - AAGATGAAATTGCCATGTAAAAATA	**G**TGTGTATAACAAAATTCCTATGATGG
D18S853	**[NED]** - GCACATGTACCCTAAAACTTAAAAT	**G**TCAACCAAAACTCAACAAGTAGTAA
D20S482	**[FAM]** - CAGAGACACCGAACCAATAAGA	**G**CCACATGAATCAATTCCTATAATAAA
D20S1082	**[VIC]** - ACATGTATCCCAGAACTTAAAGTAAAC	**G**CAGAAGGGAAAATTGAAGCTG
D22S1045 (NC01)	**[NED]** - ATTTTCCCCGATGATAGTAGTCT	**G**CGAATGTATGATTGGCAATATTTTT

Table 3. Allele frequency information on 26 autosomal STR loci.

See http://www.cstl.nist.gov/biotech/strbase/NISTpopdata/Allele_Frequencies_for_26miniSTRs.pdf

TABLE - U.S. Caucasian, African American, and Hispanic allele frequencies for 26 new miniSTR loci

D1GATA113				
Allele	Total	Cauc.	Afr. Am.	Hisp.
7	0.1483	0.1475	0.1142	0.2122
8	0.0168		0.0394	0.0072
9	0.0183		0.0453	0.0036
10	0.0199	0.0057	0.0295	0.0288
11	0.1804	0.2069	0.1535	0.1799
12	0.5482	**0.5728**	**0.5551**	**0.4892**
13	0.0680	0.0670	0.0630	0.0791
N	654	261	254	139
H(ob)	0.6682	0.6322	0.6732	0.7266
P	0.5803	0.5971	0.9676	0.7490
PIC	0.6011	0.5550	0.6177	0.6316

D1S1627				
Allele	Total	Cauc.	Afr. Am.	Hisp.
10	0.1220	0.1355	0.1202	0.1000
11	0.1477	0.1660	0.1550	0.1000
12	0.0447	0.0153	0.0853	0.0250
13	0.3492	**0.3760**	**0.3178**	**0.3571**
14	0.3159	0.3015	0.2926	0.3857
15	0.0182	0.0057	0.0252	0.0286
16	0.0023		0.0039	0.0036
N	660	262	258	140
H(ob)	0.7455	0.7366	0.7829	0.6929
P	0.3532	0.5662	0.8207	0.0986
PIC	0.6967	0.6734	0.7310	0.6522

D1S1677				
Allele	Total	Cauc.	Afr. Am.	Hisp.
9	0.0008		0.0019	
10	0.0061	0.0076	0.0078	
11	0.0083	0.0076	0.0078	0.0107
12	0.1053	0.0817	0.1206	0.1214
13	0.2167	0.2395	0.1634	0.2714
14	0.3068	**0.3517**	0.2549	**0.3179**
15	0.2598	0.2357	**0.3152**	0.2036
16	0.0705	0.0551	0.0973	0.0500
17	0.0220	0.0171	0.0253	0.0250
18	0.0015		0.0039	
19	0.0023	0.0038	0.0019	
N	660	263	257	140
H(ob)	0.7455	0.7490	0.7432	0.7429
P	0.2724	0.7058	0.0191	0.5968
PIC	0.7398	0.7140	0.7527	0.7286

D2S441				
Allele	Total	Cauc.	Afr. Am.	Hisp.
9	0.0008	0.0019		
10	0.1833	0.2015	0.0837	0.3321
11	0.3508	**0.3498**	**0.3696**	**0.3179**
11.3	0.0523	0.0608	0.0525	0.0357
12	0.0932	0.0532	0.1712	0.0250
12.3	0.0038	0.0019	0.0058	0.0036
13	0.0303	0.0285	0.0409	0.0143
13.3	0.0008		0.0019	
14	0.2417	0.2452	0.2529	0.2143
14.3	0.0008		0.0019	
15	0.0409	0.0570	0.0195	0.0500
16	0.0008			0.0036
17	0.0008			0.0036
N	660	263	257	140
H(ob)	0.7742	0.7795	0.7977	0.7214
P	0.1966	0.7857	0.1520	0.0399
PIC	0.7382	0.7319	0.7235	0.6940

D2S1776				
Allele	Total	Cauc.	Afr. Am.	Hisp.
7	0.0015		0.0039	
8	0.0436	0.0766	0.0217	0.0216
9	0.1139	0.0881	0.1398	0.1151
10	0.1598	0.1782	0.1535	0.1367
11	0.3066	**0.3218**	**0.3209**	0.2518
12	0.2966	0.2644	0.2953	**0.3597**
13	0.0673	0.0651	0.0551	0.0935
14	0.0099	0.0057	0.0079	0.0216
15	0.0008		0.0020	
N	654	261	254	139
H(ob)	0.7630	0.8008	0.7402	0.7338
P	0.3490	0.3671	0.3763	0.2598
PIC	0.7389	0.7438	0.7262	0.732

D3S3053				
Allele	Total	Cauc.	Afr. Am.	Hisp.
7	0.0008	0.0019		
8	0.0185	0.0447	0.0020	
9	0.1726	0.2412	0.1992	
10	0.1279	0.1304	0.1892	0.0107
11	0.3945	**0.4105**	**0.4243**	**0.3107**
12	0.1680	0.1498	0.1713	0.1964
13	0.0470	0.0214	0.0139	0.1536
14	0.0455			0.2107
15	0.0223			0.1036
16	0.0023			0.0107
17	0.0008			0.0036
N	660	257	251	140
H(ob)	0.7381	0.7198	0.7131	0.8143
P	0.0000	0.5148	0.8156	0.4772
PIC	0.7359	0.6918	0.6698	0.7535

D3S4529				
Allele	Total	Cauc.	Afr. Am.	Hisp.
11	0.0008	0.0020		
12	0.0439	0.0402	0.0484	0.0429
13	0.2379	**0.3695**	0.0911	0.2643
14	0.2068	0.1145	**0.2907**	0.2214
15	0.2424	0.2209	0.2364	**0.2821**
16	0.2083	0.1968	0.2558	0.1500
17	0.0583	0.0542	0.0756	0.0393
18	0.0015	0.0020	0.0019	
N	660	249	258	140
H(ob)	0.7606	0.7229	0.7519	0.8286
P	0.2439	0.2428	0.5752	0.7017
PIC	0.7607	0.7226	0.7431	0.7394

D4S2364				
Allele	Total	Cauc.	Afr. Am.	Hisp.
7	0.0008	0.0019		
8	0.1742	0.1673	0.1459	0.2393
9	0.6265	**0.5494**	**0.7646**	**0.5179**
10	0.1970	0.2795	0.0895	0.2393
11	0.0015	0.0019		0.0036
N	660	263	257	140
H(ob)	0.5106	0.5513	0.3852	0.6643
P	0.0761	0.0227	0.7511	0.6312
PIC	0.4817	0.5236	0.3515	0.5493

D4S2408				
Allele	Total	Cauc.	Afr. Am.	Hisp.
7	0.0015		0.0039	
8	0.1904	0.2222	0.1417	0.2194
9	0.2791	**0.3161**	0.1870	**0.3777**
10	0.2301	0.2375	0.2441	0.1906
11	0.2378	0.1973	**0.3189**	0.1655
12	0.0596	0.0249	0.1024	0.0468
13	0.0015	0.0019	0.0020	
N	654	261	254	139
H(ob)	0.7217	0.7088	0.7520	0.6906
P	0.0116	0.1441	0.8213	0.1667
PIC	0.7346	0.7117	0.7373	0.7021

Table 3 (cont.)

D5S2500				
Allele	Total	Cauc.	Afr. Am.	Hisp.
14	0.2748	0.2943	**0.2819**	0.2250
15	0.0023		0.0058	
16				
17	0.3072	**0.3547**	0.2259	**0.3679**
18	0.2011	0.2302	0.1564	0.2286
19	0.0075	0.0075	0.0077	0.0071
20	0.1130	0.0057	0.2645	0.0357
21	0.0008		0.0019	
22	0.0023		0.0019	0.0071
23	0.0700	0.0811	0.0328	0.1179
24	0.0211	0.0264	0.0212	0.0107
N	664	265	259	140
H(ob)	0.7470	0.7472	0.7568	0.7286
P	0.0146	0.6897	0.6766	0.6364
PIC	0.7357	0.6789	0.7366	0.7059

D6S474				
Allele	Total	Cauc.	Afr. Am.	Hisp.
11	0.0031			0.0143
12				
13	0.0031			0.0107
14	0.2180	0.2704	**0.2749**	0.0214
15	0.2311	0.2101	0.2371	0.2607
16	0.2173	0.1420	0.1514	**0.4750**
17	0.2435	**0.2802**	0.2450	0.1679
18	0.0747	0.0895	0.0737	0.0500
19	0.0085	0.0078	0.0139	
20	0.0008		0.0020	
N	660	257	251	140
H(ob)	0.7596	0.8016	0.7649	0.6786
P	0.0006	0.2105	0.4592	0.6923
PIC	0.7524	0.7397	0.7441	0.6256

D6S1017				
Allele	Total	Cauc.	Afr. Am.	Hisp.
7	0.0881	0.0019	0.2124	0.0214
8	0.2440	0.2113	**0.2548**	0.2857
9	0.0505	0.0094	0.0792	0.0750
10	0.3780	**0.4792**	0.2510	**0.4214**
11	0.0324	0.0528	0.0135	0.0286
12	0.1642	0.2094	0.1429	0.1179
13	0.0399	0.0340	0.0425	0.0464
14	0.0030	0.0019	0.0039	0.0036
N	664	256	259	140
H(ob)	0.7395	0.6981	0.8069	0.6929
P	0.2462	0.2855	0.9903	0.7511
PIC	0.7237	0.6306	0.7683	0.6766

D8S1115				
Allele	Total	Cauc.	Afr. Am.	Hisp.
9	0.3464	0.1698	**0.5695**	0.2679
10	0.0075		0.0193	
11	0.0241		0.0579	0.0071
12				
13	0.0226	0.0038	0.0483	0.0107
14	0.0316	0.0189	0.0386	0.0429
15	0.0286	0.0585	0.0039	0.0179
16	0.3426	**0.5415**	0.0965	**0.4214**
17	0.1273	0.1415	0.1120	0.1286
18	0.0467	0.0396	0.0367	0.0786
19	0.0196	0.0226	0.0154	0.0214
20	0.0030	0.0038	0.0019	0.0036
N	664	265	259	140
H(ob)	0.6627	0.6604	0.6293	0.7286
P	0.0005	0.1869	0.8742	0.5776
PIC	0.7022	0.6182	0.6239	0.6866

D9S1122				
Allele	Total	Cauc.	Afr. Am.	Hisp.
9	0.0144	0.0096	0.0232	0.0071
10	0.0175	0.0173	0.0154	0.0214
11	0.1692	0.1808	0.1795	0.1286
12	0.3900	**0.4038**	**0.3707**	**0.4000**
13	0.3232	0.3212	0.3031	0.3643
14	0.0721	0.0635	0.0830	0.0679
15	0.0091	0.0038	0.0154	0.0071
16	0.0038		0.0077	0.0036
17	0.0008		0.0019	
N	659	260	259	140
H(ob)	0.7344	0.7423	0.7529	0.6857
P	0.7938	0.9674	0.4737	0.9547
PIC	0.6592	0.6430	0.6863	0.6302

D9S2157				
Allele	Total	Cauc.	Afr. Am.	Hisp.
7	0.0908	0.1065	0.0891	0.0643
8	0.0030	0.0057	0.0019	
9	0.0877	0.0875	0.0872	0.0893
9.1	0.0076	0.0057	0.0058	0.0143
10	0.0575	0.0095	0.1298	0.0143
11	0.2814	**0.2966**	**0.2558**	**0.3000**
12	0.0620	0.0570	0.0775	0.0429
13	0.1815	0.2567	0.0814	0.2250
14	0.1006	0.1065	0.0988	0.0929
15	0.0666	0.0437	0.0717	0.1000
16	0.0499	0.0152	0.0891	0.0429
17	0.0106	0.0095	0.0097	0.0143
18				
19	0.0008		0.0019	
N	660	263	258	140
H(ob)	0.8442	0.8403	0.8837	0.7786
P	0.6062	0.8568	0.3648	0.5926
PIC	0.8319	0.7866	0.8537	0.8044

D10S1248				
Allele	Total	Cauc.	Afr. Am.	Hisp.
8	0.0015		0.0039	
9	0.0015		0.0039	
10	0.0015		0.0039	
11	0.0144		0.0370	
12	0.0740	0.0358	0.1206	0.0607
13	0.2817	**0.3302**	0.2412	0.2643
14	0.2908	0.2887	**0.2782**	**0.3179**
15	0.2062	0.1925	0.2043	0.2357
16	0.1027	0.1245	0.0798	0.1036
17	0.0227	0.0245	0.0253	0.0143
18	0.0015	0.0019	0.0019	
19	0.0015	0.0019		0.0036
N	663	265	257	140
H(ob)	0.7915	0.7849	0.8249	0.7429
P	0.0789	0.8021	0.0456	0.2857
PIC	0.7424	0.7127	0.771	0.7191

D10S1435				
Allele	Total	Cauc.	Afr. Am.	Hisp.
5	0.0015		0.0039	
6				
7	0.0045	0.0113		
8				
9	0.0023	0.0038		0.0036
10	0.0053	0.0094		0.0071
11	0.1425	0.1830	0.1163	0.1143
12	0.3356	**0.3264**	**0.3391**	**0.3464**
13	0.2881	0.2623	0.3023	0.3107
13.3	0.0023			0.0107
14	0.1712	0.1830	0.1570	0.1750
14.3	0.0023	0.0038	0.0000	0.0036
15	0.0219	0.0113	0.0349	0.0179
16	0.0060	0.0019	0.0116	0.0036
17	0.0128	0.0019	0.0271	0.0071
18	0.0015	0.0019	0.0019	
19	0.0023		0.0058	
N	663	265	258	140
H(ob)	0.7662	0.7698	0.7984	0.7000
P	0.8561	0.6775	0.5520	0.7229
PIC	0.7143	0.7167	0.7148	0.6960

D11S4463				
Allele	Total	Cauc.	Afr. Am.	Hisp.
10	0.0015		0.0039	
11	0.0399	0.0170	0.0695	0.0286
12	0.0738	0.0509	0.0985	0.0714
13	0.3042	**0.3604**	0.2336	**0.3286**
14	0.3396	0.3528	**0.3340**	0.3250
15	0.1664	0.1642	0.1795	0.1464
16	0.0648	0.0491	0.0734	0.0786
17	0.0098	0.0057	0.0077	0.0214
N	664	265	259	140
H(ob)	0.7304	0.6755	0.7799	0.7429
P	0.7148	0.1991	0.6367	0.4182
PIC	0.7148	0.6643	0.7506	0.7145

Table 3 (cont.)

D12ATA63				
Allele	Total	Cauc.	Afr. Am.	Hisp.
10	0.0008			0.0036
11	0.0076	0.0019	0.0154	0.0036
12	0.1464	0.1385	0.1525	0.1500
13	0.1624	0.2154	0.1004	0.1786
14	0.0744	0.0173	0.1564	0.0286
15	0.2420	0.1615	**0.3340**	0.2214
16	0.0683	0.0577	0.0772	0.0714
17	0.2132	**0.2981**	0.1004	**0.2643**
18	0.0736	0.0981	0.0521	0.0679
19	0.0076	0.0096	0.0058	0.0071
20	0.0038	0.0019	0.0058	0.0036
N	659	260	259	140
H(ob)	0.8285	0.8423	0.7876	0.8786
P	0.0584	0.0585	0.3194	0.7136
PIC	0.8111	0.7798	0.7899	0.7912

D14S1434				
Allele	Total	Cauc.	Afr. Am.	Hisp.
9	0.0023		0.0039	0.0036
10	0.2069	0.1660	0.2607	0.1857
11	0.0287	0.0283	0.0253	0.0357
12	0.0536	0.0189	0.0953	0.0429
13	0.3172	**0.3849**	0.2393	0.3321
14	0.3716	0.3830	**0.3521**	**0.3857**
15	0.0144	0.0094	0.0195	0.0143
16	0.0045	0.0075	0.0039	
17	0.0008	0.0019		
N	663	265	257	140
H(ob)	0.6964	0.7208	0.6848	0.6714
P	0.4221	0.9052	0.1506	0.9944
PIC	0.6641	0.6157	0.6967	0.6504

D17S974				
Allele	Total	Cauc.	Afr. Am.	Hisp.
5	0.0015	0.0019	0.0019	
6	0.0008		0.0019	
7	0.0407	0.0245	0.0579	0.0393
8	0.0881	0.0434	0.1332	0.0893
9	0.3712	**0.4189**	**0.3224**	**0.3714**
10	0.2907	0.3283	0.2259	0.3393
11	0.1649	0.1415	0.2027	0.1393
12	0.0422	0.0415	0.0541	0.0214
N	664	265	259	140
H(ob)	0.7319	0.7019	0.7568	0.7429
P	0.9278	0.5242	0.6173	0.9528
PIC	0.6983	0.6408	0.7469	0.6705

D17S1301				
Allele	Total	Cauc.	Afr. Am.	Hisp.
9	0.0038	0.0057	0.0039	
10	0.0279	0.0396	0.0154	0.0286
11	0.2914	0.3679	0.1988	0.3179
12	0.4503	**0.4113**	**0.4981**	**0.4357**
13	0.1777	0.1340	0.2143	0.1929
14	0.0452	0.0396	0.0637	0.0214
15	0.0038	0.0019	0.0058	0.0036
N	664	265	259	140
H(ob)	0.6491	0.7170	0.6255	0.5643
P	0.6597	0.5632	0.7422	0.0338
PIC	0.6234	0.6155	0.6132	0.6098

D18S853				
Allele	Total	Cauc.	Afr. Am.	Hisp.
9	0.0008	0.0019		
10	0.0286	0.0321	0.0328	0.0143
11	0.3833	**0.5585**	0.1892	**0.4107**
12	0.1017	0.1000	0.1139	0.0821
13	0.1679	0.1038	0.2259	0.1821
14	0.2372	0.1755	**0.2954**	0.2464
15	0.0768	0.0283	0.1351	0.0607
16	0.0038		0.0077	0.0036
N	664	265	259	140
H(ob)	0.7108	0.6453	0.7722	0.7214
P	0.1395	0.6630	0.4394	0.9546
PIC	0.7155	0.5997	0.7630	0.6854

D20S482				
Allele	Total	Cauc.	Afr. Am.	Hisp.
9	0.0123	0.0136	0.0040	0.0250
10	0.0655	0.0019	0.0100	0.2821
11	0.0416	0.0214	0.0239	0.1107
12	0.1009	0.0214	0.0398	**0.3571**
13	0.1834	0.1965	0.1554	0.2107
14	0.3752	**0.4358**	**0.5139**	0.0143
15	0.1641	0.2140	0.2032	
16	0.0555	0.0914	0.0498	
17	0.0008	0.0019		
18				
19	0.0008	0.0019		
N	660	257	251	140
H(ob)	0.6918	0.6887	0.6733	0.7286
P	0.0000	0.1595	0.2887	0.2909
PIC	0.7523	0.6754	0.6260	0.69

D20S1082				
Allele	Total	Cauc.	Afr. Am.	Hisp.
8	0.0008		0.0019	
9				
10	0.0045		0.0116	
11	0.4639	**0.5604**	**0.3263**	**0.5357**
12	0.0633	0.0698	0.0695	0.0393
13	0.0489	0.0057	0.1100	0.0179
14	0.1642	0.1132	0.2413	0.1179
15	0.2003	0.1925	0.1815	0.2500
16	0.0512	0.0509	0.0579	0.0393
17	0.0030	0.0075		
N	664	265	259	140
H(ob)	0.6958	0.6528	0.7915	0.6000
P	0.0789	0.2827	0.5150	0.1146
PIC	0.6725	0.5908	0.7504	0.5851

D22S1045				
Allele	Total	Cauc.	Afr. Am.	Hisp.
8	0.0038		0.0097	
9				
10	0.0204		0.0428	0.0179
11	0.1193	0.1396	0.1304	0.0607
12	0.0317	0.0151	0.0564	0.0179
13	0.0076	0.0094	0.0039	0.0107
14	0.0597	0.0585	0.0798	0.0250
15	0.3293	0.3321	**0.2588**	**0.4536**
16	0.2832	**0.3623**	0.1868	0.3107
17	0.1337	0.0792	0.2101	0.0964
18	0.0091	0.0038	0.0156	0.0071
19	0.0023		0.0058	
N	663	265	257	140
H(ob)	0.7840	0.7849	0.8171	0.7214
P	0.0453	0.9191	0.2013	0.1133
PIC	0.7418	0.6853	0.8027	0.6348

N: sample size; H(ob): observed heterozygosity; *P*: Hardy-Weinberg equilibrium, exact test; PIC: polymorphism information content

Table 4. Additional information on 26 autosomal STR loci with listing order based on overall heterozygosity rank (Hill *et al.* 2008).

					Heterozygosity			
Locus	**Repeat Motif**	**Observed Allele Range**	**PCR Product Sizes (bp)***	**N**	**Af. Am.**	**Cau.**	**Hisp.**	**Overall**
D9S2157	ATA	7 - 19	71 - 107	661	0.884	0.840	0.779	0.844
D12ATA63	ATA	10 - 20	76 - 106	659	0.788	0.842	0.879	0.829
D10S1248 (NC01)	GGAA	10 - 20	83 - 123	663	0.825	0.785	0.743	0.792
D22S1045 (NC01)	TAA	6 - 17	76 - 109	663	0.817	0.785	0.721	0.784
D2S441 (NC02)	TCTA	9 - 17	78 - 110	660	0.798	0.780	0.721	0.774
D10S1435	GATA	4 - 18	82 - 139	663	0.798	0.770	0.700	0.766
D2S1776	GATA	6 - 14	127 - 161	654	0.740	0.801	0.734	0.763
D3S4529	GATA	12 - 19	111 - 139	660	0.752	0.723	0.829	0.761
D6S474	[AGAT][GATA]	11 - 18	107 - 136	648	0.765	0.802	0.679	0.761
D5S2500	[GGTA][GACA] [GATA][GATT]	14 - 24	85 - 126	664	0.757	0.747	0.729	0.747
D1S1627	ATT	10 - 16	81 - 100	660	0.783	0.737	0.693	0.746
D1S1677 (NC02)	GGAA	9 - 18	81 - 117	660	0.743	0.749	0.743	0.746
D6S1017	ATCC	6 - 13	81 - 110	664	0.807	0.698	0.693	0.740
D3S3053	GATA	8 - 14	84 - 108	648	0.713	0.724	0.814	0.739
D9S1122	GATA	9 - 17	93 - 125	659	0.753	0.742	0.686	0.734
D17S974	GATA	5 - 12	95 - 124	664	0.757	0.702	0.743	0.732
D11S4463	GATA	9 - 16	88 - 116	664	0.780	0.676	0.743	0.730
D4S2408	GATA	7 - 13	85 - 109	654	0.752	0.709	0.691	0.722
D18S853	TAA	9 - 16	82 - 104	664	0.772	0.645	0.721	0.711
D14S1434 (NC01)	[GATA][GACA]	13 - 20	70 - 98	663	0.685	0.721	0.650	0.696
D20S1082	ATA	8 - 17	73 - 101	664	0.792	0.653	0.600	0.696
D20S482	GATA	9 - 19	85 - 126	648	0.673	0.689	0.729	0.691
D1GATA113	GATA	7 - 13	81 - 105	654	0.673	0.632	0.727	0.668
D8S1115	AAT	9 - 20	63 - 96	664	0.629	0.660	0.729	0.663
D17S1301	TCTA	8 - 14	114 - 139	664	0.626	0.717	0.564	0.649
D4S2364 (NC02)	[GAAT][GGAT] [GAAT]	8 - 12	67 - 83	660	0.385	0.551	0.664	0.511

*Apparent size observed relative to GS500 LIZ size standard

Table 5. Genotypes with 26 autosomal STR loci on commonly used DNA positive control samples. See also http://www.cstl.nist.gov/biotech/strbase/miniSTR/miniSTR_NC_loci_types.htm.

	Standard DNA Template Genotypes			
Locus	**9947A**	**9948**	**ABI 007**	**K562**
D1GATA113	11,12	7,12	12,12	11,12
D1S1627	13,14	11,13	11,14	10,14
D1S1677 (NC02)	13,14	13,14	13,13	13,14
D2S441 (NC02)	10,14	11,12	14,15	10,14
D2S1776	10,10	10,12	8,10	11,11
D3S3053	9,11	9,12	9,9	12,12
D3S4529	13,13	12,12	13,13	14,14
D4S2364 (NC02)	9,10	9,10	9,10	9,9
D4S2408	9,10	10,10	10,11	10,11
D5S2500	14,23	14,17	17,18	14,14
D6S474	14,18	17,17	14,14	15,18
D6S1017	9,10	8,8	10,10	8,11
D8S1115	9,18	15,17	15,17	16,16
D9S1122	12,13	12,15	12,12	10,14,15
D9S2157	7,13	7,11	13,13	13,13
D10S1248 (NC01)	13,15	12,15	12,15	12,12
D10S1435	10,11	12,13	11,13	10,12
D11S4463	12,13	12,14	14,14	13,14
D12ATA63	13,13	13,18	13,17	17,17
D14S1434 (NC01)	11,13	13,14	11,14	10,10
D17S974	7,10	10,11	9,10	8,8
D17S1301	12,12	11,12	12,13	11,12
D18S853	11,14	11,11	11,11	12,15
D20S482	14,15	13,14	14,15	15,15
D20S1082	11,14	11,15	12,14	11,11
D22S1045 (NC01)	11,14	16,18	11,16	16,16

Table 6. Most common type possible with these 26 autosomal STR loci based on allele frequency data from three major U.S. population groups (Caucasian, African American, or Hispanic) found in Table 3. Based on assuming the most common allele from any of the population groups and the highly unlikely possibility that all loci are homozygous.

Locus	Allele	Group	Most common frequency (p)	p^2
D1GATA113	12	C	0.5728	0.3281
D1S1627	13	C	0.376	0.1414
D1S1677	14	C	0.3517	0.1237
D2S441	11	AA	0.3696	0.1366
D2S1776	12	H	0.3597	0.1294
D3S3053	11	AA	0.4243	0.1800
D3S4529	13	C	0.3695	0.1365
D4S2364	9	AA	0.7646	0.5846
D4S2408	9	H	0.3777	0.1427
D5S2500	17	H	0.3679	0.1354
D6S474	16	C	0.475	0.2256
D6S1017	10	C	0.4792	0.2296
D8S1115	9	AA	0.5695	0.3243
D9S1122	12	C	0.4038	0.1631
D9S2157	11	H	0.3	0.0900
D10S1248	13	C	0.3302	0.1090
D10S1435	12	H	0.3464	0.1200
D11S4463	13	C	0.3604	0.1299
D12ATA63	15	AA	0.334	0.1116
D14S1434	14	H	0.3857	0.1488
D17S974	9	C	0.4189	0.1755
D17S1301	12	AA	0.4981	0.2481
D18S853	11	C	0.5585	0.3119
D20S482	14	AA	0.5139	0.2641
D20S1082	11	C	0.5604	0.3140
D22S1045	15	H	0.4536	0.2058
			product	4.53E-20

Table 7. Mutation observed in 22 STR loci examined across 395 father/son pairs previously confirmed via Identifiler and Yfiler testing (Decker et al. 2008).

Ethnicity	Locus	Allele (father)	Allele (child)
African American	D6S474	14,16	17,17
Caucasian	D2S1776	13,13	12,14
Hispanic	D11S4463	14,15	13,16
Asian	D10S1435	11,14	12,13
Asian	D3S4529	13,13	15,15
Asian	D20S482	12,15	13,14

Table 8. Comparison of likelihood ratios with 15 STRs (Identifiler kit) versus 37 STRs (Identifiler plus 22 Autoplex loci) to evaluate various potential relationships.

Relationship Examined	**15 STRs** (Identifiler, ID15)	ID15 **+ Autoplex 22** STRs = **37 loci** (A37)
Mother/Child* (*with single mutation)	0.214	**5,200,000**
Siblings	477	**113,000**
Uncle/Nephew	824	**247,000**
Cousins	0.45	2.25
Grandparents/ Grandchildren	0.53	1.42

Table 9. Information on 37 Y-STR loci used (Decker et al. 2007). The 17 Yfiler kit loci are shown in bold font.

Locus	GenBank Accession	Reference Allele	Amplicon size (bp)	Size range (bp)	Allele range	Y-position
DYS19	AC017019	[TAGA]$_3$tagg[TAGA]$_{12}$ = 15	252	232-268	10-19	10.132
DYS385a/b	AC022486	(GAAA)$_{11}$	369	353-425	7-25	19.261
DYS389I/II	AC004617	[TCTG]$_5$[TCTA]$_{12}$…[TCTG]$_3$[TCTA]$_9$	247, 367	239-259, 347-387	10-15, 24-34	13.122
DYS390	AC011289	[TCTG]$_8$[TCTA]$_{11}$[TCTG]$_1$[TCTA]$_4$ = 24	215	191-227	18-27	15.784
DYS391	AC011302	(TCTA)$_{11}$	287	271-295	7-13	12.613
DYS392	AC011745	(TAT)$_{13}$	254	236-269	7-18	21.043
DYS393	AC006152	(AGAT)$_{12}$	119	103-135	8-16	3.191
DYS437	AC002992	(TCTA)$_{10}$[TCTG]$_2$(TCTA)$_4$ = 16	192	180-196	13-17	12.977
DYS438	AC002531	(TTTTC)$_{10}$	221	211-236	8-13	13.376
DYS439	AC002992	(AGAT)$_{13}$	252	232-260	8-15	13.025
DYS444	AC007043	(ATAG)$_{14}$	308	292-316	10-16	17.736
DYS446	AC006152	(TCTCT)$_{14}$	308	278-343	8-21	3.192
DYS448	AC025227	[AGAGAT]$_{11}$N$_{42}$[AGAGAT]$_8$ = 19	294	282-324	17-24	22.775
DYS449	AC051663	(TTTC)$_{15}$...(TTTC)$_{14}$ = 29	355	335-387	24-37	8.278
DYS456	AC010106	(AGAT)$_{15}$	149	141-161	13-18	4.331
DYS458	AC010902	(GAAA)$_{16}$	123	115-139	14-20	7.928
DYS463	AC007275	(AAAGG)$_7$(AAGGG)$_{15}$(AAGGA)$_2$ = 24	254	219-274	17-28	7.704
DYS485	AC009233	(TTA)$_{16}$	278	260-284	10-18	20.559
DYS495	AC004474	(AAT)$_{15}$	212	203-221	12-18	13.521
DYS505	AC012078	(TCCT)$_{12}$	174	162-186	9-15	3.701
DYS508	AC006462	(TATC)$_{11}$	177	165-193	8-15	16.303
DYS520	AC007275	(ATAG)$_{10}$(ATAC)$_{10}$ = 20	179	171-203	18-26	7.790
DYS522	AC007247	(GATA)$_{10}$	352	344-380	8-17	7.476
DYS532	AC016991	(CTTT)$_{14}$	479	459-491	9-17	8.439
DYS533	AC053516	(ATCT)$_{12}$	214	202-222	9-14	16.903
DYS534	AC053516	(CTTT)$_{15}$	208	188-228	10-20	16.903
DYS540	AC010135	(TTAT)$_{12}$	266	258-274	10-14	17.075
DYS556	AC011745	(AATA)$_{11}$	211	203-219	9-13	21.011
DYS557	AC007876	(TTTC)$_{16}$	196	176-220	11-22	21.644
DYS570	AC012068	(TTTC)$_{17}$	256	236-280	12-23	6.921
DYS576	AC010104	(AAAG)$_{17}$	191	175-207	13-21	7.113
DYS594	AC010137	(AAATA)$_{10}$	264	259-284	9-14	20.116
DYS635	AC004772	[TCTA]$_4$(TGTA)$_2$[TCTA]$_2$(TGTA)$_2$ [TCTA]$_2$(TGTA)$_2$[TCTA]$_9$ = 23	176	152-192	17-27	12.890
DYS643	AC007007	(CTTTT)$_{11}$	145	125-165	7-15	15.936
Y-GATA-H4	AC011751	(TAGA)$_{12}$	368	352-372	8-13	17.253

Table 10. PCR primer sequences (Butler et al. 2006) used for the additional 20 Y-STRs beyond Yfiler (see Table 9) that may be used in five separate multiplex assays as illustrated in **Figure 7**. The gtttctt "PIG-tails" were added to the 5'-end of the reverse primer to promote full adenylation.

Locus	Dye	Forward Primer Sequence	Reverse Primer Sequence
DYS444	NED	TCTAAGGGATCCAAAGGCAGAA	gtttcttGTGTGAACCATTTGGCATGTTTA
DYS446	FAM	TATTTTCAGTCTTGTCCTGTC	gtttcttAAATGTATGGCCAACATAGCAAAACCA
DYS449	NED	CCTGGAAGTGGAGTTTGCTGT	gtttcttTGGAGTCTCTCAAGCCTGTTCTA
DYS463	FAM	AATTCTAGGTTTGAGCAAAGACA	gtttcttATGAGGTTGTGTGACTTGACTG
DYS485	NED	CCTGGGTGACAAGAGTTATACTCT	gtttcttGCAGACTTCGCCACTACATAAT
DYS495	FAM	AGCAAACTTTGAAGCCAGAAAG	gtttcttCTTGGGCAACAGAGCGAGA
DYS505	FAM	TCTGGCGAAGTAACCCAAAC	gtttcttTCGAGTCAGTTCACCAGAAGG
DYS508	VIC	ACAATGGCAATCCCAAATTC	gtttcttGAACAAATAAGGTGGGATGGAT
DYS520	NED	AACAGCCTGCCCAACATAGT	gtttcttACCATCATGCCCTGCAATA
DYS522	VIC	CCTTTGAAATCATTCATAATGC	gtttcttTCATAAACAGAGGGTTCTGG
DYS532	NED	TTGGTTTTATGCCTTTCACT	gtttcttTAGGTGACAGAGCAGGATTC
DYS533	VIC	CATCTAACATCTTTGTCATCTACC	gtttcttTGATCAGTTCTTAACTCAACCA
DYS534	NED	CATCTACCCAACATCCATCTA	gtttcttGACAAAGATGTTAGATGAATAGACA
DYS540	NED	GACCGTGTACTCTGGCCAAT	gtttcttCAGGAGGCTAGCTCAGGAGA
DYS556	FAM	TGCTGTCACATCACCAATGA	gtttcttTTTGGTTGCTGAAGCATTGA
DYS557	VIC	TTTTCTGTGCCAAGCCTACA	gtttcttTCTAATGCACCTTGAGGGATG
DYS570	NED	GAACTGTCTACAATGGCTCACG	gtttcttTCAGCATAGTCAAGAAACCAGACA
DYS576	FAM	TTGGGCTGAGGAGTTCAATC	gtttcttGGCAGTCTCATTTCCTGGAG
DYS594	VIC	GATGTGCCTAATGCCACAGA	gtttcttCCCTGGTGTTAATCGTGTCC
DYS643	FAM	AAGCCATGCCTGGTTAAACT	gtttcttTGTAACCAAACACCACCCATT

Table 11. Numbers of unique (blue font) and shared haplotypes observed with various combinations of Y-STR loci across 656 U.S. population samples that are part of the Yfiler haplotype database. Data used for this analysis available at http://www.cstl.nist.gov/biotech/strbase/NISTpop.htm**.**

# times haplotype observed	MHL	SWGDAM	PPY	Yfiler	ALL 37
1	**429**	**486**	**505**	**626**	**652**
2	34	33	34	**12**	**2**
3	13	10	14	**2**	.
4	4	6	3	.	.
5	3	1	2	.	.
6	1	1	.	.	.
7	1	2	1	.	.
8	1	.	.	.	.
9	2	.	.	.	.
10	.	1	.	.	.
11	1	.	.	.	.
12	.	.	**1**	.	.
13	1	.	.	.	.
14	.	.	.	.	.
15	.	**1**	.	.	.
16	.	.	.	.	.
17	.	.	.	.	.
18	.	.	.	.	.
19	.	.	.	.	.
20	.	.	.	.	.
21	.	.	.	.	.
22	.	.	.	.	.
23	.	.	.	.	.
24	.	.	.	.	.
25	.	.	.	.	.
26	**1**	.	.	.	.
HD	0.996644	0.998529	0.999064	0.999916	0.999991
DC	0.748476	0.824695	0.853659	0.97561	0.996951
# HT	491	541	560	640	654

Table 12. Summary of 29 Yfiler population studies published in the literature as of September 2007. Multiple haplotypes in Mizuno et al. (2007): 43 in 2; 12 in 3; 3 in 4; 1 in 5; 1 in 7; 2 in 8; 1 in 9; 1 in 22.

Population Tested w/ Yfiler loci	Number Tested (N)	# Haplotypes (h)	# Unique HT	# Haplotypes in Multiples	Haplotype Diversity (HD)	Discrimination Capacity (h/N)	Reference
U.S. (ABI)	3561						Mulero et al. (2006)
U.S. Caucasian, African Am., Hispanic, Asian sons (NIST)	389	389	389	--	1.0000	1.0000	Decker et al. (2008)
U.S. Caucasian (VA, SD, AL)	191	190	189	1 in 2		0.9948	Hanson and Ballantyne (2007)
African American (VA, SD, AL)	381	378	375	3 in 2		0.9921	Hanson and Ballantyne (2007)
Chinese Han (Shandong province)	131	129	127	2 in 2	0.9998	0.9847	Yan et al. (2007)
Chinese Salar (Qinghai province)	133	123	117	3 in 2; 2 in 3; 1 in 4	0.9983	0.9248	Zhu et al. (2007)
Chinese Tibetan ethnic minority (Tibet)	112	112	112	--	1.0000	1.0000	Zhang et al. (2006)
Chinese Tibetan ethnic minority (Tibet)	167	163	159	4 in 2	0.9998	0.9760	Zhu et al. (2007)
Taiwanese	200	192	176	8 in 2	0.9996	0.9600	Huang et al. (2007)
Tirol (Austria)	135	130	125	5 in 2	0.9994	0.9630	Berger et al. (2005)
Sub-Saharan Africans (Rio de Janeiro, Brazil)	135	133	131	2 in 2	0.9998	0.9852	Domingues et al. (2007)
Taiwanese	104	90	82	6 in 2; 1 in 3; 1 in 7		0.8654	Liu et al. (2007)
Japanese (Hokkaido, Honshu, Shikoku, Kyushu)	1079	950	886	*See caption*	0.9992	0.8804	Mizuno et al. (2007)
Serbian (Vojvodina province)	185	176	168	7 in 2; 1 in 3	0.9994	0.9514	Veselinovic et al. (2007)
Kalmykian (Elista, Russia)	99	75	64	8 in 2; 1 in 3; 1 in 6; 1 in 10	0.9860	0.7576	Roewer et al. (2007)
Barcelona metropolitan area (Catalonia, NE Spain)	247	247	247	--	1.0000	1.0000	Sanchez et al. (2007)
Chinese Tibetan (Lassa area of Tibet)	107	106	105	1 in 2	0.9998	0.9907	Li et al. (2007)
Brazilians (five geopolitical regions of Brazil)	500	481	466	11 in 2; 4 in 3	0.9998	0.9620	Pereira et al. (2007)
Portugese	250	231	213	17 in 2; 1 in 3	0.9994	0.9240	Alves et al. (2007)
Polish (central Poland)	255	252	249	3 in 2	0.9999	0.9882	Soltyszewski et al. (2007)
Italian (NE Italy)	155	153	151	2 in 2	0.9997	0.9871	Turrina et al. (2006)
Fang (Bioko Island, Equatorial Guinea)	110	101	94	6 in 2; 1 in 4	0.9980	0.9182	Barrot et al. (2007)
Bubi (Bioko Island, Equatorial Guinea)	133	102	87	7 in 2; 4 in 3; 1 in 4; 2 in 5; 1 in 6	0.9932	0.7669	Barrot et al. (2007)
North African Berber (Takrouna & Sejenane, Tunisia)	66	47	40	4 in 2; 1 in 3; 1 in 7; 1 in 8		0.7121	Frigi et al. (2006)

Malaysian Malays	334	327	321	5 in 2; 1 in 3	0.9999	0.9790	Chang et al. (2007)
Malaysian Chinese	331	313	295	18 in 2	0.9997	0.9456	Chang et al. (2007)
Malaysian Indians	315	306	297	9 in 2	0.9998	0.9714	Chang et al. (2007)
Sub-Apennine (central Italy)	162	155	149	5 in 2; 1 in 3	0.9994	0.9568	Onofri et al. (2007)
Mexican (Chihuahua, North Central Mexico)	326	315	306	7 in 2; 2 in 3	0.9997	0.9663	Gutierrez-Alarcon et al. (2007)
Japanese	161	153	146	6 in 2; 1 in 3	0.9994	0.9503	Hara et al. (2007)
	6893	**6514**	**6257**		**0.9997**	**0.9610**	

Table 13. Mutation rates with 17 Y-STR loci in Yfiler based on NIST and literature results (Decker et al. 2008).

Yfiler kit loci	**Literature Summary**			**NIST Results**			
Locus	Mutations	# Meioses	Mutation Rate	Mutations	# Meioses	Mutation Rate	TOTAL
DYS19	22	9241	0.238%	1	389	0.257%	0.239%
DYS389I	14	7445	0.188%	5	389	1.285%	0.243%
DYS389II	22	7432	0.296%	6	389	1.542%	0.358%
DYS390	21	8723	0.241%	1	389	0.257%	0.241%
DYS391	25	8672	0.288%	0	389	<0.003%	0.276%
DYS392	5	8636	0.058%	0	389	<0.003%	0.055%
DYS393	6	7425	0.081%	0	389	<0.003%	0.077%
DYS385 a/b	30	13765	0.218%	0	389	<0.003%	0.212%
DYS438	2	4075	0.049%	0	389	<0.003%	0.045%
DYS439	22	4052	0.543%	5	389	1.285%	0.608%
DYS437	6	3971	0.151%	0	389	<0.003%	0.138%
DYS448	1	557	0.180%	0	389	<0.003%	0.106%
DYS456	4	557	0.718%	1	389	0.257%	0.529%
DYS458	6	557	1.077%	4	389	1.028%	1.057%
DYS635	6	1430	0.420%	3	389	0.771%	0.495%
GATA-H4	4	1593	0.251%	3	389	0.771%	0.353%

Figure 1. Schematic representation of dye labels and PCR product size ranges for Y-STR loci present in the PowerPlex Y (Promega Corporation, Madison, WI) and Yfiler (Applied Biosystems, Foster City, CA) kits. Note that all 12 loci of PowerPlex Y are included within the 17 loci present in the Yfiler kit with the additional 5 Y-STR loci being boxed.

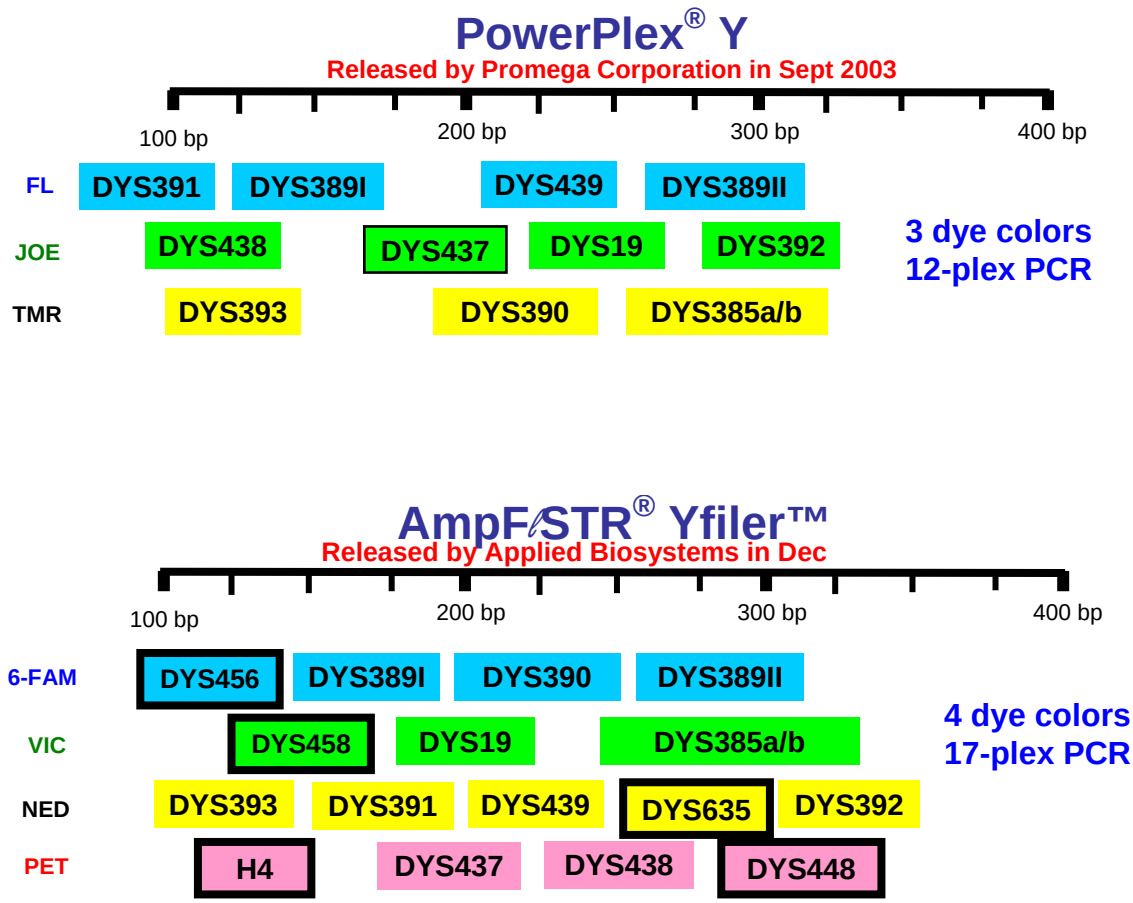
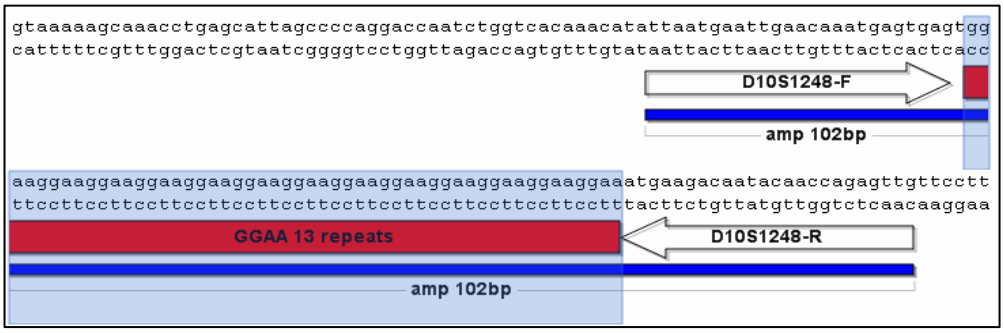


Figure 2. Illustration of PCR primer positions and repeat structure and nomenclature for three of the autosomal STR loci D10S1248, D2S441, and D22S1045.

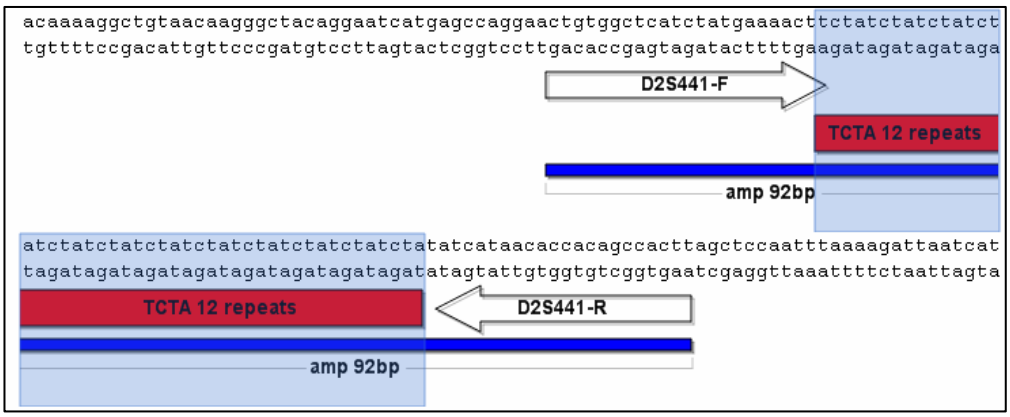
D10S1248

GenBank accession AL391869; positions 136,773..136,874



D2S441

GenBank accession AC079112; positions 85,324..85,415



D22S1045

GenBank accession AL022314; positions 92,943..93,047

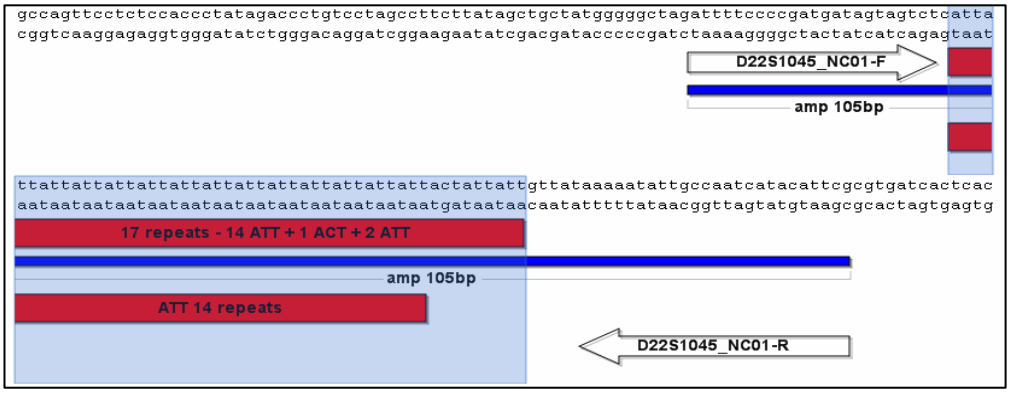


Figure 3. Results from a 23plex (Autoplex) amplification including 22 autosomal STRs and amelogenin for sex-typing.

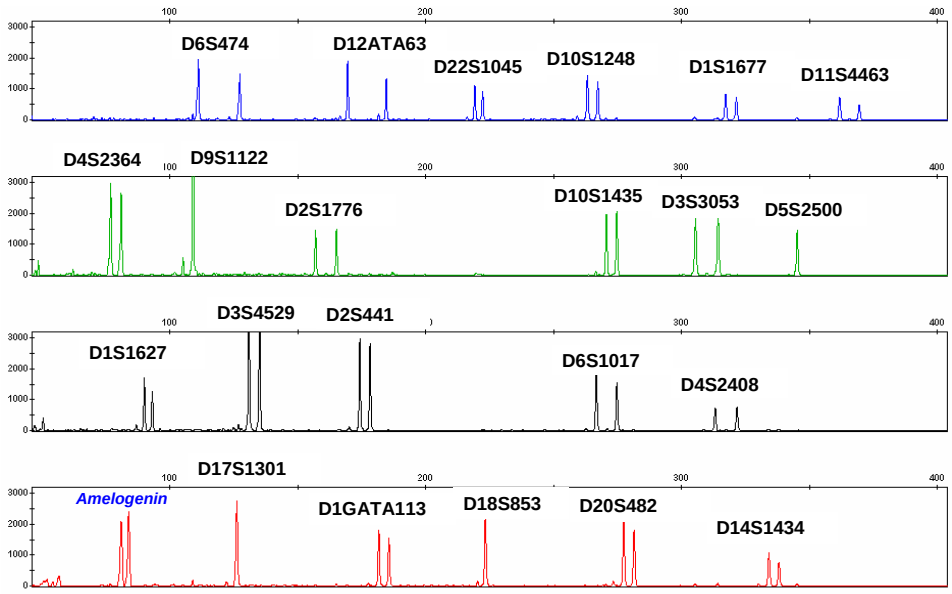
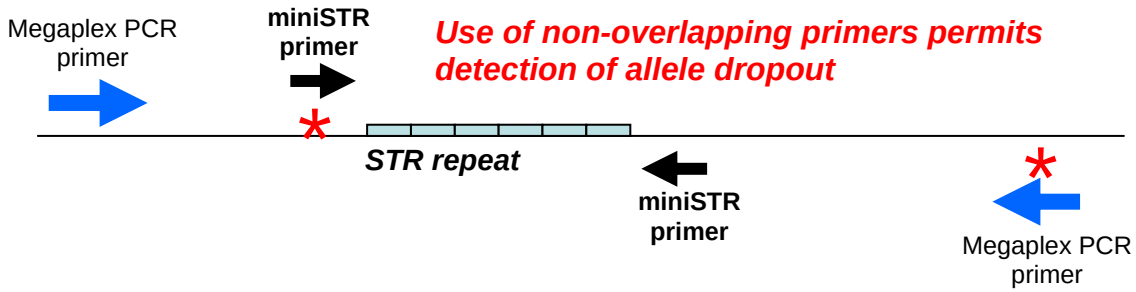


Figure 4. (Top) Illustration of how non-overlapping PCR primers can be used to detect allele dropout (null alleles). (Bottom) Discordant example showing loss of allele 23 in D5S2500 with the 23plex.



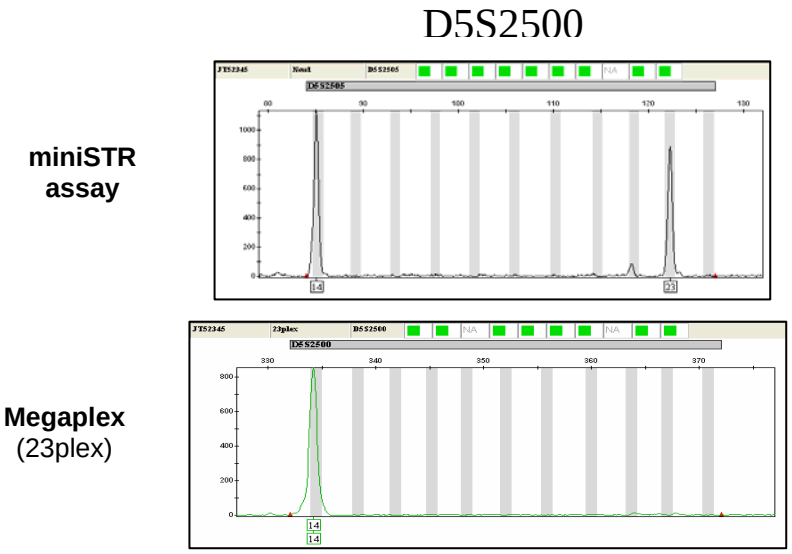


Figure 5. An African American sample that shows loss of the X amelogenin allele due to a point mutation in the Identifiler primer binding site which is detected in the Autoplex (NIST 23plex).

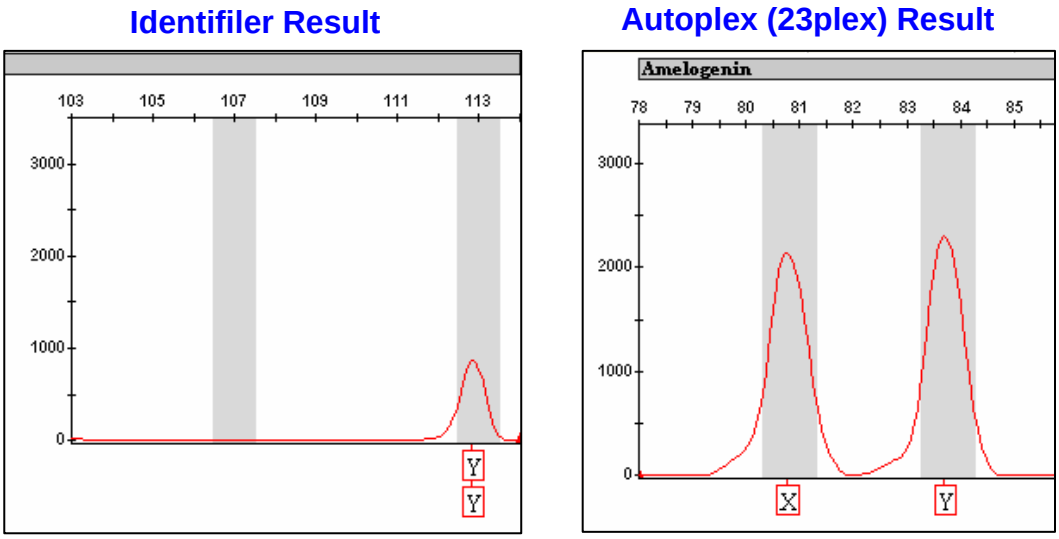


Figure 6. Pedigree showing extended family samples evaluated with Identifiler 15 STR loci results (ID15) compared to Identifiler plus the 22 STRs present in the Autoplex (A37). Likelihood ratios for the 15 vs 37 STR comparison are shown.

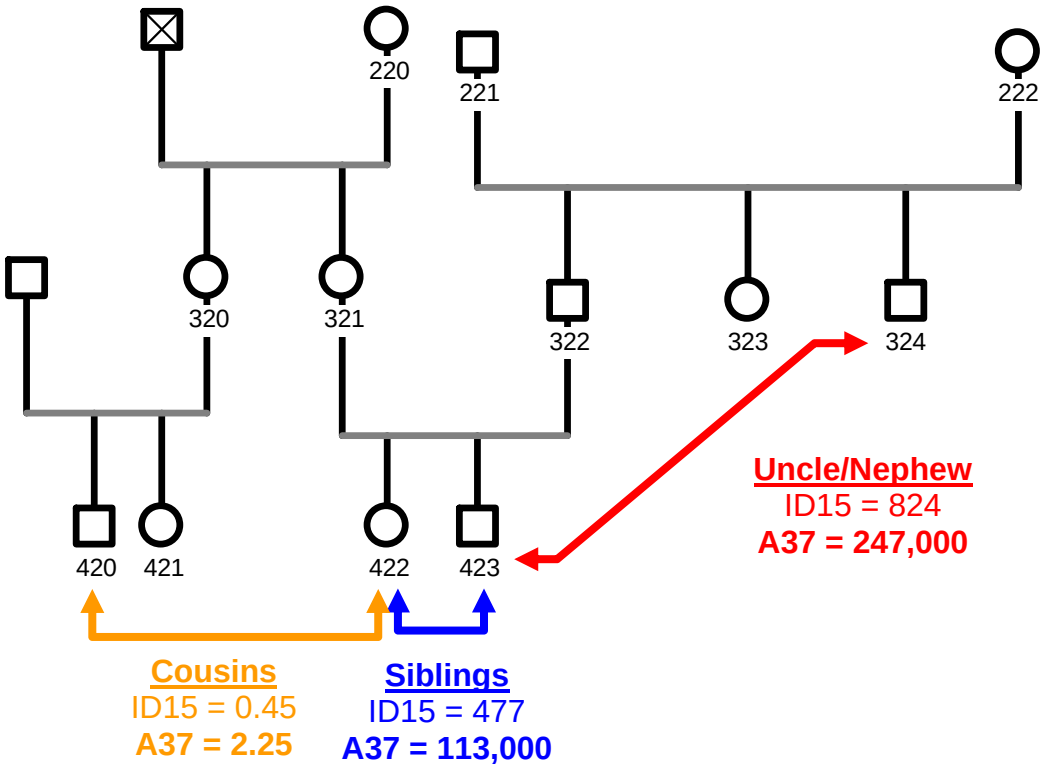


Figure 7. Five multiplex in-house assays used to type 20 additional Y-STRs across the NIST U.S. population samples.

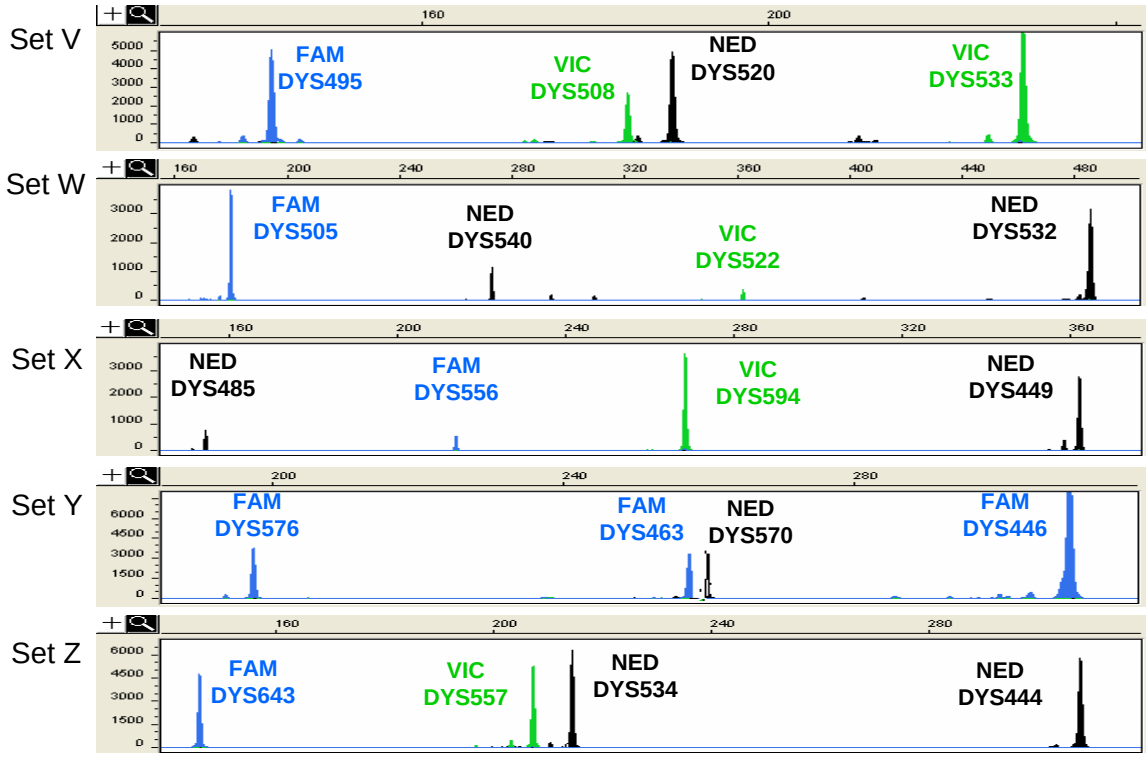


Figure 8. Frequency of the most common haplotype within the Yfiler 3561 dataset using different combinations of Y-STR loci.

(A) Minimal haplotype

Population	# Haplotypes	# Haplotypes (with Selected Alleles)	Frequency
African American	985	23	0.0234
Asian	330	0	0
Caucasian	1276	66	0.0517
Filipino	105	1	0.0095
Hispanic	597	28	0.0469
Native American	106	5	0.0472
Sub-saharan African	59	0	0
Vietnamese	103	1	0.0097
All	**3561**	**124**	**0.0348**

Most common type
9 locus minHT

DYS19 – 14
DYS389I – 13
DYS389II – 29
DYS390 – 24
DYS391 – 11
DYS392 – 13
DYS393 – 13
DYS385 a/b – 11,14

(B) SWGDAM recommended loci

Population	# Haplotypes	# Haplotypes (with Selected Alleles)	Frequency
African American	985	11	0.0112
Asian	330	0	0
Caucasian	1276	32	0.0251
Filipino	105	1	0.0095
Hispanic	597	13	0.0218
Native American	106	2	0.0189
Sub-saharan African	59	0	0
Vietnamese	103	1	0.0097
All	**3561**	**60**	**0.0168**

Most common type
11 locus SWGDAM

DYS19 – 14
DYS389I – 13
DYS389II – 29
DYS390 – 24
DYS391 – 11
DYS392 – 13
DYS393 – 13
DYS385 a/b – 11,14
DYS438 -- 12
DYS439 -- 12

(C) PowerPlex Y loci

Population	# Haplotypes	# Haplotypes (with Selected Alleles)	Frequency
African American	985	11	0.0112
Asian	330	0	0
Caucasian	1276	28	0.0219
Filipino	105	0	0
Hispanic	597	9	0.0151
Native American	106	2	0.0189
Sub-saharan African	59	0	0
Vietnamese	103	1	0.0097
All	**3561**	**51**	**0.0143**

Most common type
12 locus PP Y

DYS19 – 14
DYS389I – 13
DYS389II – 29
DYS390 – 24
DYS391 – 11
DYS392 – 13
DYS393 – 13
DYS385 a/b – 11,14
DYS438 -- 12
DYS439 -- 12
DYS437 -- 15

(D) Yfiler loci

Population	# Haplotypes	# Haplotypes (with Selected Alleles)	Frequency
African American	985	4	0.0041
Asian	330	0	0
Caucasian	1276	4	0.0031
Filipino	105	0	0
Hispanic	597	2	0.0034
Native American	106	0	0
Sub-saharan African	59	0	0
Vietnamese	103	0	0
All	**3561**	**10**	**0.0028**

Most common type
17 locus Yfiler

DYS19 – 14
DYS389I – 13
DYS389II – 29
DYS390 – 24
DYS391 – 11
DYS392 – 13
DYS393 – 13
DYS385 a/b – 11,14
DYS438 -- 12
DYS439 -- 12
DYS437 -- 15
DYS456 -- 16
DYS458 -- 17
DYS635 -- 23
DYS448 -- 19
GATA-H4 -- 12