J Forensic Sci, March 2006, Vol. 51, No. 2
doi:10.1111/j.1556-4029.2006.00046.x
Available online at: www.blackwell-synergy.com

John M. Butler,[1] *Ph.D.*

Genetics and Genomics of Core Short Tandem Repeat Loci Used in Human Identity Testing

ABSTRACT: Over the past decade, the human identity testing community has settled on a set of core short tandem repeat (STR) loci that are widely used for DNA typing applications. A variety of commercial kits enable robust amplification of these core STR loci. A brief history is presented regarding the selection of core autosomal and Y-chromosomal STR markers. The physical location of each STR locus in the human genome is delineated and allele ranges and variants observed in human populations are summarized as are mutation rates observed from parentage testing. Internet resources for additional information on core STR loci are reviewed. Additional topics are also discussed, including potential linkage of STR loci to genetic disease-causing genes, probabilistic predictions of sample ethnicity, and desirable characteristics for additional STR loci that may be added in the future to the current core loci. These core STR loci, which form the basis for DNA databases worldwide, will continue to play an important role in forensic science for many years to come.

KEYWORDS: forensic science, DNA typing, short tandem repeat, mutation rate, CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D2S1338, D19S433, Penta D, Penta E, SE33, CODIS, national DNA databases, Y-STR, Y-chromosome, DYS19, DYS385, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, Y-GATA-H4

It has been almost a decade since the 13 genetic markers that form the core of the FBI Laboratory's Combined DNA Index System (CODIS) were selected in November 1997. Because of their use in the U.S. national DNA database (NDNAD) as well as other criminal justice databases around the world, these short tandem repeat (STR) loci dominate the genetic information that has been collected to date on human beings (1–3). In the U.S. and U.K. alone, more than 5 million profiles now exist in criminal justice DNA databases that contain information from these core loci or a subset (4,5). In addition, almost 1 million samples are run annually with core STR loci as part of parentage testing (6).

The 13 CODIS loci used in the U.S. are CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, and D21S11 (7). The U.K. and much of Europe utilize 10 core loci that include the additional markers D2S1338 and D19S433 along with eight overlapping loci FGA, TH01, VWA, D3S1358, D8S1179, D16S539, D18S51, and D21S11. These loci have become the common currency of data exchange for human identity testing both in forensic casework and paternity testing largely because of their ease of use in the form of commercial STR kits. Missing persons investigations and mass disaster victim identification typically also involve the same STR markers and kits (8,9).

This review article describes what has been learned over the past few years about these commonly used STR markers in terms of their population genetic variation and genomic locations. Their precise chromosomal information only recently became available with the completion of the Human Genome Project. The past few years have seen over a 1000 population studies performed—usually with data included in the FOR THE RECORD section of the *Journal of Forensic Sciences* or an Announcement of Population Data in *Forensic Science International*. Available Internet resources for further information on these commonly used STR markers are reviewed. In addition, controversial issues such as potential disease gene linkage and probabilistic predictions of sample ethnicity are discussed. Finally, commonly used Y-chromosome STR loci are briefly reviewed.

Historical Perspective on STR Marker Selection

STR markers were first described as effective tools for human identity testing in the early 1990s (10,11). The Forensic Science Service (FSS) began to aggressively search for new loci and study population variation with a number of STR candidates (12). The Royal Canadian Mounted Police (RCMP) also contributed to early efforts with STR typing (13) along with a number of European labs. The first FSS multiplex applied to forensic casework included the four loci TH01, VWA, FES/FPS, and F13A1 (14). A second generation multiplex (SGM) followed with the loci TH01, VWA, FGA, D8S1179, D18S51, and D21S11 (15). The U.K. NDNAD was launched in April 1995 utilizing the SGM loci and the amelogenin sex-typing test (16).

Seeing the promise of STR typing technology and the success being obtained in the U.K., the FBI Laboratory led U.S. efforts to establish core STR loci that would form the backbone of CODIS, the U.S. national database system. Fueled through funding provided by the Congressional DNA Identification Act of 1994, a community-wide STR Project was launched in April 1996 (7).

[1]National Institute of Standards and Technology, Gaithersburg, MD 20899-8311.

 Sources of Support: National Institute of Justice funded this work in part through interagency agreement 2003-IJ-R-029 with the NIST Office of Law Enforcement Standards.

Received 25 April 2005; and in revised form 31 Aug. 2005; accepted 1 Oct. 2005; published 9 Feb. 2006.

This project, which lasted for approximately 18 months, involved 22 DNA typing laboratories that collectively evaluated 17 candidate loci, which were available as commercial or preliminary kits from either Promega Corporation (Madison, WI) or Applied Biosystems (Foster City, CA). Performance studies and protocol evaluations were performed, population databases were established, and forensic validation was conducted on the various STR systems investigated. While early work with STRs involved detection on silver-stained polyacrylamide gels (17), the community has embraced fluorescence detection methods involving first gel electrophoresis (10,12,13) and then capillary electrophoresis with such instruments as the ABI 310 and ABI 3100 Genetic Analyzers (18). Over the years, the ABI 373 and 377 gel-based DNA sequencers have also played a significant role in forensic DNA typing (19).

For the STR Project, Promega Corporation provided F13A1, F13B, FES/FPS, and LPL as part of an "FFFL" multiplex and CSF1PO, TPOX, TH01, VWA, D16S539, D7S820, D13S317, and D5S818 as part of the PowerPlex kit (20). Applied Biosystems had the AmpFlSTR Blue kit consisting of D3S1358, VWA, and FGA and the AmpFlSTR Green I kit with TH01, TPOX, CSF1PO, and the sex-typing system amelogenin. AmpFlSTR Yellow multiplex with D5S818, D13S317, and D7S820 along with the AmpFlSTR Green II multiplex consisting of D8S1179, D21S11, D18S51, and amelogenin were also made available to participants in the STR evaluation project. Eventually AmpFlSTR Blue, Green I, and Yellow were combined to form the AmpFlSTR Profiler kit, and the Blue, Green II, and Yellow loci were eventually combined to create the AmpFlSTR Profiler Plus kit (21,22).

At the STR Project meeting held on November 13–14, 1997, the 13 STR loci were announced as the core CODIS markers required for the U.S. national database (7). In the late 1990s, Applied Biosystems began providing the Profiler Plus and COfiler kits to enable coverage of the 13 core loci for use on their instrument platforms (22). Promega Corporation developed the PowerPlex 2.1 kit to cover the additional loci not present in their PowerPlex 1.1 kit for use on the FMBIO detection platform (1,20,23).

Table 1 summarizes the various STR kits that have become available in the past decade. Since the turn of the century, new multiplex assays have been developed that amplify all 13 CODIS core loci in a single reaction. The PowerPlex 16 kit, which was released by the Promega Corporation in May 2000, amplifies the 13 core loci, amelogenin, and two pentanucleotide loci referred to as Penta D and Penta E (24). Applied Biosystems released their 16plex Identifiler kit in July 2001, which amplifies the 13 core loci, amelogenin, and two tetranucleotide loci D2S1338 and D19S433 (25).

The Penta loci were discovered and characterized by Promega scientists in an effort to find loci with high variability yet exhibiting low amounts of stutter product formation (26,27). Although Penta D and Penta E are not officially required loci for any NDNADs, they are considered as "core loci" for the purposes of this paper because of their presence in widely used commercial STR kits. The D2S1338 and D19S433 STR markers were identified in searches for new tetranucleotide loci in the late 1990s (28,29). The extra two STR loci in the PowerPlex 16 and Identifiler kits provide an increased power of discrimination and enable improved mixture interpretation (in the case of the low stutter penta loci) or increased overlap with European STR systems (in the case of D2S1338 and D19S433).

After it became available in 1999, the U.K. and much of Europe adopted a commercial STR kit from Applied Biosystems known as SGM Plus, which contains the original SGM loci and amelogenin plus D3S1358, D16S539, D2S1338, and D19S433 (30). When Germany established its NDNAD in 1998, the highly polymorphic STR locus SE33 (also known as ACTBP2) was included as a core locus because of its previous use in casework applications (31,32). Both Promega Corporation and Applied Biosystems (along with several German companies) now supply kits that include SE33 (Table 1).

TABLE 1—*Summary of available commercial STR kits that are commonly used.*

Kit Name	STR Loci Included	Random Match Probability with Author's Profile*
	Promega Corporation	
PowerPlex 1.1 and 1.2	CSF1PO, TPOX, TH01, VWA, D16S539, D13S317, D7S820, D5S818	7.4×10^{-10}
PowerPlex 2.1 (for Hitachi FMBIO users)	D3S1358, TH01, D21S11, D18S51, VWA, D8S1179, TPOX, FGA, Penta E	3.4×10^{-11}
PowerPlex ES	FGA, TH01, VWA, D3S1358, D8S1179, D18S51, D21S11, SE33, amelogenin	1.3×10^{-10}
PowerPlex 16	CSF1PO, FGA, TPOX, TH01, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, Penta D, Penta E, amelogenin	1.2×10^{-18}
PowerPlex 16 BIO (for Hitachi FMBIO users)	CSF1PO, FGA, TPOX, TH01, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, Penta D, Penta E, amelogenin	1.2×10^{-18}
	Applied Biosystems	
AmpFlSTR Blue	D3S1358, VWA, FGA	1.0×10^{-3}
AmpFlSTR Green I	Amelogenin, TH01, TPOX, CSF1PO	7.8×10^{-4}
AmpFlSTR Cofiler (CO)	D3S1358, D16S539, Amelogenin, TH01, TPOX, CSF1PO, D7S820	2.0×10^{-7}
AmpFlSTR Profiler Plus (Pro)	D3S1358, VWA, FGA, Amelogenin, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820	2.4×10^{-11}
AmpFlSTR Profiler Plus ID	D3S1358, VWA, FGA, Amelogenin, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820 (extra unlabeled D8-R primer)	2.4×10^{-11}
AmpFlSTR Profiler	D3S1358, VWA, FGA, Amelogenin, TH01, TPOX, CSF1PO, D5S818, D13S317, D7S820	9.0×10^{-11}
AmpFlSTR SGM Plus (SGM)	D3S1358, VWA, D16S539, D2S1338, Amelogenin, D8S1179, D21S11, D18S51, D19S433, TH01, FGA	4.5×10^{-13}
AmpFlSTR Sefiler (SE)	FGA, TH01, VWA, D3S1358, D8S1179, D16S539, D18S51, D21S11, D2S1338, D19S433, SE33, amelogenin	5.1×10^{-15}
AmpFlSTR Identifiler (ID)	CSF1PO, FGA, TPOX, TH01, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D2S1338, D19S433, amelogenin	7.2×10^{-19}

*Allele frequencies used for random match probability calculations (to unrelated individuals) from U.S. Caucasian population data associated with Butler et al. (37), Reid et al. (38), and Levadokou et al. (39). Subpopulation structure adjustments (θ corrections) were not made with these calculations (i.e., only p^2 and $2pq$ were used).

STR, short tandem repeat.

Thus, the creation of commercial STR kits has been historically driven by selection of loci that have become part of NDNADs. However, in some cases loci were selected for inclusion in databases based on which ones were available in commercial kits or already previously in use for forensic casework (e.g., SE33). It is also important to realize that patents play a role in the cost of STR kits and their commercial availability (33–36).

Locus Information

Information regarding the repeat structure and number of observed alleles for each core STR locus is available in Table 2. The first article describing each STR locus is also listed in Table 2 under the original reference column (40–55). Note that many of these loci were selected from genetic markers under evaluation by the Cooperative Human Linkage Center (CHLC) (see http://www.chlc.org).

The repeat motif for each STR marker is listed according to the International Society of Forensic Genetics (ISFG) recommendation that the repeat sequence motif be defined so that the first 5′-nucleotides on the GenBank forward strand define the repeat motif used (56). Observed allele ranges for each locus are also included in Table 2 along with PCR product sizes and dye labels for the various STR kits described in Table 1. It is important to remember that STR allele sizes are measured relative to an internal size standard during electrophoresis and, depending on the DNA strand that is dye labeled, may have a different apparent measured size than the actual DNA sequence (see (18)).

A detailed synopsis of each marker including the PCR product sizes generated with the various STR kits is available in Chapter 5 of *Forensic DNA Typing* (1). A full description of the allele range and number of alleles reported to date for each locus is contained in Appendix I of *Forensic DNA Typing* (1). Note that the most complex loci, D21S11 and SE33, contain a number of alleles with internal sequence variation that can only be fully appreciated through DNA sequence analysis of the STR repeat region. For example, Rolf et al. (57) found 102 different SE33 alleles upon sequencing a total of 33 different length variants.

Genomic Information

The Human Genome Project officially came to a successful completion in April 2003 with the announcement of a "finished" reference sequence of the human genome (58). However, the finished sequence continues to be refined and several compilations exist, which differ from one another. Using the BLAST-like alignment tool (BLAT) that is available at http://genome.ucsc.edu, each of the core STR loci has been located within the reference human genome sequence. Table 3 lists the 18 core loci in terms of their chromosomal locations. In addition, an evaluation of the physical position of these STR loci has been performed in the National Center for Biotechnology Information (NCBI) build 34 (July 2003) vs. NCBI build 35 (May 2004) versions of the human genome sequence. Reference sequences for the STR loci used for this BLAT search are available at http://www.cstl.nist.gov/biotech/strbase/seq_ref.htm. In some cases, the reverse complement of the GenBank accession sequence was used in order to have the forward strand possess the traditional repeat motif listed in Table 2.

With the exceptions noted below, the core loci are located on separate chromosomes and therefore expected to segregate independently of one another during meiosis. This independent segregation enables use of the product rule in estimating random match probabilities with DNA profiles generated from multiple STR loci (59). As can be seen in Table 3, CSF1PO and D5S818

TABLE 2—*Characteristics of alleles observed in 18 core STR loci. See Ref. 1, Appendix I for a more complete description of the number of alleles seen.*

Locus Name	First Reference	Repeat Motif ISFG format	Allele Range	Applied Biosystems STR Kits PCR Product Sizes (Dye Label)	Promega STR Kits PCR Product Sizes (Dye Label)	Number of Alleles Seen
CSF1PO	(40)	TAGA	5–16	CO: 276–320 bp (JOE) ID: 276–320 bp* (6FAM)	1.1: 287–331 bp (TMR) 16: 317–361 bp (JOE)	20
FGA	(41)	CTTT	12.2–51.2	Pro: 196–348 bp (5FAM) SGM/SE: 196–348 bp (NED) ID: 196–348 (PET)	2.1/16/ES: 308–464 bp (TMR)	80
TH01	(42)	TCAT	3–14	CO: 160–204 bp (JOE) SGM/SE: 160–204 bp (NED) ID: 160–204 bp (VIC)	1.1: 171–215 bp (TMR) 2.1/16/ES: 152–196 bp (FL)	20
TPOX	(43)	GAAT	4–16	CO: 209–257 bp (JOE) ID: 209–257 bp* (NED)	1.1: 216–264 bp (TMR) 2.1/16: 254–302 bp (TMR)	15
VWA	(44)	[TCTG][TCTA]	10–25	Pro/SGM: 152–212 bp (5FAM) ID: 152–212 bp (NED) SE: 152–212 bp (6FAM)	1.1/16/ES: 123–183 bp (TMR)	29
D3S1358	(45)	[TCTG][TCTA]	8–21	Pro/CO: 97–149 bp (5FAM) ID: 97–149 bp (VIC) SE: 97–149 bp (6FAM)	2.1/16/ES: 99–151 bp (FL)	25
D5S818	(46)	AGAT	7–18	Pro: 134–178 bp (NED) ID: 134–178 bp (PET)	1.1: 119–163 bp (FL) 16: 119–163 bp (JOE)	15
D7S820	(47)	GATA	5–16	Pro/CO: 253–297 bp (NED) ID: 253–297 bp (6FAM)	1.1: 211–255 bp (FL) 16: 211–255 bp (JOE)	30
D8S1179	(48)	[TCTA][TCTG]	7–20	Pro: 123–175 bp (JOE) ID: 123–175 bp (6FAM) SE: 123–175 bp (VIC)	2.1/16/ES: 203–255 bp (TMR)	15
D13S317	(49)	TATC	5–16	Pro: 193–237 bp (NED) ID: 193–237 bp* (VIC)	1.1: 157–201 bp (FL) 16: 157–201 bp (JOE)	17
D16S539	(50)	GATA	5–16	CO/SGM: 233–277 bp (5FAM) ID: 233–277 bp* (VIC) SE: 233–277 bp (6FAM)	1.1: 264–308 bp (FL) 16: 264–308 (JOE)	19
D18S51	(51)	AGAA	7–39.2	Pro/SGM: 264–394 (JOE) ID: 264–394 (NED) SE: 264–394 (PET)	2.1/16/ES: 286–416 bp (FL)	51
D21S11	(52)	Complex [TCTA][TCTG]	12–41.2	Pro/SGM: 138–256 bp (JOE) ID: 138–256 bp (6FAM) SE: 138–256 bp (PET)	2.1/16/ES: 155–273 bp (FL)	89
D2S1338	(53)	[TGCC][TTCC]	15–28	SGM: 289–341 bp (5FAM) ID: 289–341 bp* (VIC) SE: 289–341 bp (6FAM)	N/A	17
D19S433	(54)	AAGG	9–17.2	SGM/ID/SE: 106–140 bp (NED)	N/A	26
Penta D	(26)	AAAGA	2.2–17	N/A	16: 376–449 bp (JOE)	29
Penta E	(26)	AAAGA	5–24	N/A	2.1/16: 379–474 bp (FL)	33
SE33 (ACTBP2)	(55)	Complex AAAG	4.2–37	SE: 203–333 bp (VIC)	ES: 203–333 bp (JOE)	>100

Note that the more complex loci have internal variation, which can only be fully characterized with DNA sequence analysis. PCR product sizes are listed for the corresponding allele ranges and STR kits (see Table 1 for kit abbreviations) based on actual sequence length without any nontemplate addition, which may differ from observed size relative to an internal size standard. Five loci in the Identifiler STR kit are marked with an asterisk (*) since mobility modifiers have been added in order to adjust their apparent size during electrophoresis.

STR, short tandem repeat.

TABLE 3—*Genomic locations of core STR loci.*

Locus (UniSTS)	GenBank Accession (Allele Repeat #)	Chromosomal Location	Physical Position (July 2003; NCBI Build 34)	Physical Position (May 2004; NCBI Build 35)
TPOX (240638)	M68651 (11)	2p25.3 thyroid peroxidase, 10th intron	Chr 2 1.436 Mb	Chr 2 1.472 Mb
D2S1338 (30509)	AC010136 (20)	2q35	Chr 2 219.082 Mb	Chr 2 218.705 Mb
D3S1358 (148226)	AC099539 (16)	3p21.31	Chr 3 45.543 Mb	Chr 3 45.557 Mb
FGA (240635)	M64982 (21)	4q31.3 αfibrinogen, 3rd intron	Chr 4 156.086 Mb	Chr 4 155.866 Mb
D5S818 (54700)	AC008512 (11)	5q23.2	Chr 5 123.187 Mb	Chr 5 123.139 Mb
CSF1PO (156169)	X14720 (12)	5q33.1 c-fms proto-oncogene, 6th intron	Chr 5 149.484 Mb	Chr 5 149.436 Mb
SE33 (ACTBP2) (none reported)	V00481 (26.2)	6q14 β-actin related pseudogene	Chr 6 88.982 Mb	Chr 6 89.043 Mb
D7S820 (74895)	AC004848 (13)	7q21.11	Chr 7 83.401 Mb	Chr 7 83.433 Mb
D8S1179 (83408)	AF216671 (13)	8q24.13	Chr 8 125.863 Mb	Chr 8 125.976 Mb
TH01 (240639)	D00269 (9)	11p15.5 tyrosine hydroxylase, 1st intron	Chr 11 2.156 Mb	Chr 2 2.149 Mb
VWA (240640)	M25858 (18)	12p13.31 von Willebrand Factor, 40th intron	Chr 12 19.826 Mb	Chr 12 5.963 Mb
D13S317 (7734)	AL353628 (11)	13q31.1	Chr 13 80.520 Mb	Chr 13 81.620 Mb
Penta E (none reported)	AC027004 (5)	15q26.2	Chr 15 95.104 Mb	Chr 15 95.175 Mb
D16S539 (45590)	AC024591 (11)	16q24.1	Chr 16 86.168 Mb	Chr 16 84.944 Mb
D18S51 (44409)	AP001534 (18)	18q21.33	Chr 18 59.098 Mb	Chr 18 59.100 Mb
D19S433 (33588)	AC008507 (16)	19q12	Chr 19 35.109 Mb	Chr 19 35.109 Mb
D21S11 (240642)	AP000433 (29)	21q21.1	Chr 21 19.476 Mb	Chr 21 19.476 Mb
Penta D (none reported)	AP001752 (13)	21q22.3	Chr 21 43.912 Mb	Chr 21 43.880 Mb

Results with two different builds of the human genome are shown in order to illustrate that the physical position within the reference genome may shift slightly as new information becomes available. UniSTS is a comprehensive database of sequence tagged sites (STSs) available on the NCBI Web site: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi? db, units.

STR, short tandem repeat.

are both found on chromosome 5 and are separated by approximately 26.3 megabases (Mb). Likewise, Penta D and D21 are both located on chromosome 21 separated by approximately 24.4 Mb. However, the occurrence of loci on the same chromosome that are many millions of base pairs apart should not impact reliable use of the product rule as even loci less than a million bases apart can be shuffled separately because of recombination hot spots and patterns of linkage disequilibrium (60). To date, hundreds of population studies involving D5S818 and CSF1PO (see, e.g., the listing at http://www.cstl.nist.gov/biotech/strbase/population/PopSurvey.htm) conducted on unrelated individuals have failed to show any signs of significant linkage between these two loci.

The amelogenin locus that occurs on both the X and Y chromosomes and enables sex typing (61) was also located within the reference human genome sequence. AMELX is located on the X chromosome at 10.676 Mb (July 2003) and 11.075 Mb (May 2004). AMELY is located on the Y chromosome at 6.441 Mb (July 2003) and 6.781 Mb (May 2004).

In addition to the determination of physical locations of various DNA sequences, the cytogenetic map has been integrated with the human genome sequence to permit a more precise approximation of band locations (62). Table 3 contains the most up-to-date chromosomal band locations available for the core loci. For example, TPOX is found at 2p25.3 rather than the previously used span of "2p23-2pter."

Population Variation

Allele Range and Variants

STR typing is typically performed using size comparisons with standardized allelic ladders that possess the most common alleles, which have been sequenced to reveal the true number of repeats (63). Different STR kit manufacturers may supply allelic ladders with slightly different allele ranges. Note that in Table 2 the observed allele ranges for the core loci are listed based on a review of the published literature rather than available allelic ladders (see also Appendix I in Ref. (1)).

As more samples are run with STR loci, new alleles are constantly being discovered that do not size exactly with the ladder alleles. These "off-ladder" alleles can be variants with more or less of the core repeat unit than present in the common alleles found in the commercially available allelic ladder. Alternatively, these variant alleles may contain partial repeats or insertions/deletions in the flanking region close to the repeat.

A good example of an insertion/deletion event that creates off-ladder alleles is found in D7S820, which can contain 8, 9, or 10 adjacent T nucleotides starting 12 nucleotides downstream of the GATA repeat (64). This flanking region insertion/deletion gives rise to the 9.1, 9.3, 10.1, 10.3, etc., alleles observed in D7S820 (Table 4). In addition, new alleles can be discovered that occur outside the range defined by the commercially available allelic ladder. In many instances, these alleles are simply classified as greater than the largest allele (or smaller than the smallest allele) in the ladder rather than attempting to extrapolate to a predicted number of repeats. Table 4 contains a list of variant or "off-ladder" alleles that have been reported to the NIST STRBase Web site as of April 2005.

Triallelic patterns have been observed for many of the core STR loci and recorded on the NIST STRBase Web site (Table 5). Clayton et al. (65) have described possible reasons for triallelic patterns, which can occur as an imbalance in amounts between the three alleles (type 1) or equal amounts of all three alleles (type 2). A type 1 tri-allelic pattern imbalance is typically a situation where the sum of the peak heights for two of the alleles is approximately equivalent to the third allele (65). It is interesting to note that TPOX, which occurs closest to the tip of a chromosome (see Table 3), has the highest number of observed tri-allelic patterns—most

TABLE 4–*Variant or "off-ladder" alleles reported in STRBase for commonly used STR loci.*

STR Locus	Number Reported	Variant Alleles Reported as of April 2005
CSF1PO	11	5, 7.3, 8.3, 9.1, 9.3, 10.1, 10.2, 10.3, 11.1, 12.1, 16
FGA	69	12.2, 13.2, 14, 14.3, 15, 15.3, 16, 16.1, 16.2, "< 17", 17, 17.2, 18.2, 19.1, 19.2, 19.3, 20.1, 20.2, 20.3, 21.1, 21.2, 21.3, 22.1, 22.2, 22.3, 23.1, 23.2, 23.3, 24.1, 24.2, 24.3, 25.1, 25.2, 25.3, 26.1, 26.2, 26.3, 27.3, 29.2, 30.2, 31, 31.2, 32.1, 32.2, 33.1, 34.1, 34.2, 35.2, 41.1, 41.2, 42.1, 42.2, 43.1, 43.2, 44, 44.1, 44.2, 44.3, 45.1, 45.2, 46.1, 46.2, 47.2, 48.2, 49, 49.1, 49.2, 50.2, 50.3
TH01	7	4, 7.3, 8.3, 9.1, 10.3, 11, 13.3
TPOX	7	4, 5, 7.3, 13.1, 14, 15, 16
VWA	6	16.1, 18.3, 22, 23, 24, 25
D3S1358	18	8, 8.3, 9, 10, 11, 15.1, 15.2, 15.3, 16.2, 17.1, 17.2, 18.1, 18.2, 18.3, ">19", 20, 20.1, 21.1
D5S818	5	10.1, 11.1, 12.3, 17, 18
D7S820	22	5, 5.2, 6.3, 7.1, 7.3, 8.1, 8.2, 8.3, 9.1, 9.2, 9.3, 10.1, 10.3, 11.1, 11.3, 12.1, 12.2, 12.3, 13.1, 14.1, 15, 16
D8S1179	4	7, 15.3, 18, 20
D13S317	10	5, 6, 7, 7.1, 8.1, 11.1, 11.3, 13.3, 14.3, 16
D16S539	10	6, 7, 9.3, 11.3, 12.1, 12.2, 13.1, 13.3, 14.3, 16
D18S51	30	7, 8, 9, 11.2, 12.2, 12.3, 13.1, 13.3, 14.2, 15.1, 15.2, 16.1, 16.2, 16.3, 17.2, 17.3, 18.1, 18.2, 19.2, 20.1, 20.2, 21.2, 22.1, 22.2, 23.2, 24.2, 27, 28.1, 28.3, 40
D21S11	24	24.3, 25.1, 25.2, 25.3, 26.2, 27.1, 27.2, 28.1, 28.3, 29.1, 29.3, 30.3, 31.1, 31.3, 32.1, 33.1, 34.1, 34.3, 35.1, 36.1, 36.2, 37, 37.2, 39
Penta D	14	6, 6.4, 7.1, 7.4, 9.4, 10.3, 11.1, 11.2, 12.2, 12.4, 13.2, 13.4, 14.1, 14.4
Penta E	13	9.4, 11.4, 12.1, 12.2, 13.2, 14.4, 15.2, 15.4, 16.4, 17.4, 18.4, 19.4, 23.4
D2S1338	3	13, 23.2, 23.3
D19S433	11	6.2, 7, 8, "< 9", 11.1, 12.1, 13.2, 18, 18.2, 19.2, 20
SE33	0	None reported yet in STRBase

These 264 alleles were as of April 2005. For up-to-date information, see http://www.cstl.nist.gov/biotech/strbase/var_tab.htm. Many of these variant alleles have been seen more than once. Note that some of these alleles may be present in allelic ladders from commercial kits not used by laboratories reporting these variants.
STR, short tandem repeat.

of which are type 2 with equal intensity alleles (Table 5). Thus, it is possible that this section of chromosome 2 is more likely to be duplicated in some individuals for telomere maintenance to keep the end of the chromosome intact (66,67).

Characterizing a Variant Allele That Occurs Between Two Loci

Occasionally a variant allele can occur with a size between two loci in a multiplex STR electropherogram making it difficult to assign the allele to the appropriate locus without further characterization, such as individual locus amplification (Fig. 1). Unfortunately, some manufacturers only provide STR kits in multiplex format preventing easy single locus amplification with the same PCR primers. However, a different STR kit, which has the loci assembled in a different configuration in terms of size and dye label (see Table 2), can be used in some cases to effectively assign an unusual allele to the appropriate locus. Alternatively, single STR locus PCR amplification primers are available from Promega Corporation or can be synthesized based on locus-specific information recorded in the STR Fact Sheets on the NIST STRBase Web site.

There are several points of consideration that can be made in order to help ascertain to which locus an extremely off-ladder and interlocus allele belongs. First, if one of the loci contains two alleles and the other one only one allele within the common allele range, then it is likely that the interlocus allele belongs to the apparent homozygote. It is also worth checking if any new variant alleles have been reported previously by other labs (see Table 4).

In a situation such as is illustrated in Fig. 1, where the sample has a locus1 with only an allele "a" and locus2 only has an allele "c" with an allele "b" occurring between the two loci, the possible

TABLE 5—*A total of 62 tri-allelic patterns observed and reported on STRBase (http://www.cstl.nist.gov/biotech/strbase/tri_tab.htm).*

STR Locus	Number Reported	Tri-Allelic Patterns Reported as of April 2005
CSF1PO	2	9/11/12; 10/11/12
FGA	10	19/20/21; 19/22/23; 19/24/25; 20/21/22; 20/21/24; 20/23/24; 21/22/23; 21/25/26; 22/24/25; 22.2/23/23.2
TH01	1	7/8/9
TPOX	13	6/8/10; 6/9/10; 6/10/11; 6/10/12; 7/9/10; 7/10/11; 8/9/10; 8/10/11; 8/10/12; 8/11/12; 9/10/11; 9/10/12; 10/11/12
VWA	8	11/16/17; 12/18/19; 14/15/17; 14/15/18; 14/16/18; 14/17/18; 15/16/17; 18/19/20
D3S1358	4	15/16/17; 15/17/18; 16/17/19; 17/18/19
D5S818	2	10/11/12; 11/12/13
D7S820	2	8/9/12; 8/10/11
D8S1179	5	10/12/13; 10/12/15; 12/13/14; 12/13/15; 13/15/16
D13S317	3	8/11/12; 10/11/12; 10/12/13
D16S539	1	12/13/14
D18S51	7	12/13/15; 12/14/15; 12/16/17; 14/15/22; 15/16/20; 16/17/20; 19/22.2/23.2
D21S11	4	28/29/30; 28/30.2/31.2; 29/31/32; 30/30.2/31
Penta D	0	None reported yet in STRBase
Penta E	0	None reported yet in STRBase
D2S1338	0	None reported yet in STRBase
D19S433	0	None reported yet in STRBase
SE33	0	None reported yet in STRBase

Many of these tri-allelic patterns have been observed more than once.
STR, short tandem repeat.

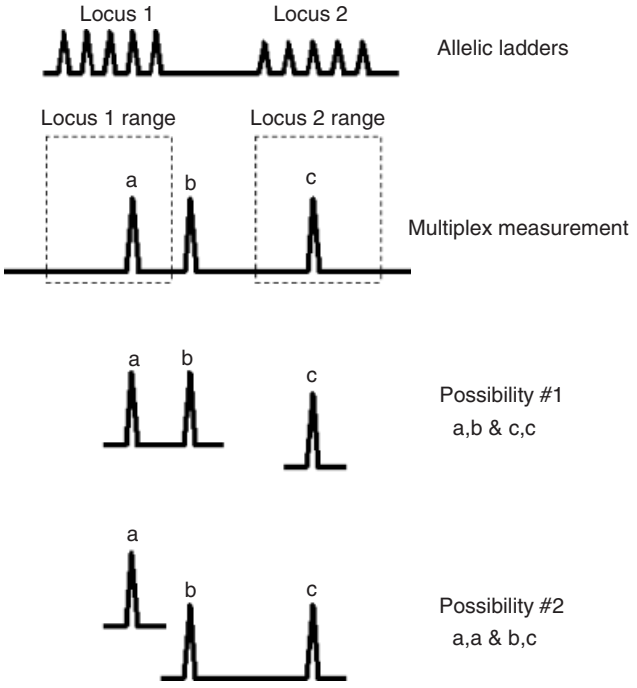

FIG. 1—*Illustration of an interlocus allele observed in a measurement involving multiplex amplification where it becomes difficult to assign allele "b" to locus 1 or locus 2.*

genotypes are as follows: locus1 (a,b) and locus2 (c,c) or locus1 (a,a) and locus2 (b,c). Heterozygosities of the two loci in question can be considered to predict which locus is more likely to be a heterozygote and possess two different alleles. For example, if a green-colored peak occurs between D16S539 and D2S1338 in the Identifiler kit and only a single allele is observed in each of the D16 and D2 normal allele ranges, then the interlocus allele more likely belongs to D2S1338 because D2 has a higher heterozygosity. The STRBase, variant allele section (or Table 4) can be examined to see if any other laboratories have observed extremely large D16 or extremely small D2 alleles. In this example, no large D16S539 alleles have been reported in STRBase, whereas several D2S1338 allele 13 observations have been noted. Finally, genotype frequencies can be examined to see if a locus1 or a locus2 homozygote is more common. For example, in an Identifiler genotype database (e.g., http://www.cstl.nist.gov/biotech/strbase/NISTpopdata/JFS2003IDresults.xls) a D16S539 11,11 homozygote occurs 10.7% of the time while a D2S1338 20,20 homozygote occurs only 0.92% of the time. Thus, it is more likely that the interlocus allele is a D2S1338 allele 13 rather than a D16S539 allele 17. While these considerations can help advise a laboratory on the best way to proceed with associating interlocus off-ladder alleles, it is recommended that final confirmation be performed with single locus amplification for each of the two adjacent STR loci.

Null Alleles with Commercial STR Kits

Sequence variation does occur in the flanking regions surrounding STR loci, and some PCR primers have been noted to be impacted by a primer binding site mutation, which can lead to allele dropout. For example, allele dropout at the VWA locus with the Applied Biosystems kits was reported (68) and ascribed to a point mutation near the 3'end of the forward PCR primer used (69). Potential null alleles resulting from allele dropout can often be predicted through statistical evaluation of STR typing data via comparison of the observed number of homozygotes to those expected based on Hardy-Weinberg equilibrium (70,71).

Because of the fact that different assays or commercial STR kits have primers that anneal to different flanking region sequences around a particular STR locus, concordance studies are conducted to detect possible null alleles. An examination of over 2000 samples comparing the PowerPlex 16 kit to the Profiler Plus and COfiler kit results found 22 examples of allele dropout because of a primer mismatch at seven of the 13 core STR loci in common (72,73). In addition, mutations under primer-binding sites have impacted the detection of D5S818 (74), D16S539 (75), and D18S51 (76) alleles with various PCR primer sets. The use of an extra or "degenerate" primer to account for possible sequence variation under a primer-binding site has been done with VWA (77), D16S539 (21), and D8S1179 (78) in some STR kits.

Mutation Rates

In situations where a direct comparison between evidence and a suspect is being made, mutation rates are not important. However, with comparisons between relatives in parentage testing and kinship analysis, such as may be applied in mass disaster victim identification, mutational events can play a significant role (79). Table 6 summarizes mutation rate data collected by the American Association of Blood Banks (AABB) as part of their 2003 annual report. These data come from several paternity testing laboratories. Not surprisingly, the loci with the highest mutation rate, e.g., SE33, FGA, D18S51, are the most polymorphic and possess the highest number of alleles (see Table 2). An exception to this observation is the complex repeat STR locus D21S11 where internal sequence variation may go undetected in size-based separations.

Population Studies

The literature contains over 1000 papers with information on STR allele frequencies observed in various population groups from around the world. An attempt to encapsulate many of these studies into a helpful list based on the commercial STR kits from which the data were generated has been made by Brian Burritt of the San Diego Police Department. As of early 2005, this list contains 365 population studies based on 183 literature references. This information has been made available on the internet at http://www.cstl.nist.gov/biotech/strbase/population/PopSurvey.htm.

In addition, Brian Burritt has developed a Microsoft Excel-based program called OmniPop that permits calculation of a user-inputted profile's frequency using allele frequencies from 166 published population surveys. OmniPop can be downloaded at http://www.cstl.nist.gov/biotech/strbase/population/OmniPop150.4.2.xls.

While most population studies include only 100–150 samples (see (80)), a few reported data sets have included thousands of individuals (81,82). A widely used population set is that published by Budowle et al. (83). Allele frequencies between small- and large-sized population databases (for the same or similar population group) rarely differ significantly for common alleles. Large data sets typically identify a greater number of rare alleles as more individuals in a population are included in the analysis. These rare alleles can be reliably accounted for through use of a minimum allele frequency as recommended by the National Research Council report (59).

TABLE 6—*Summary of apparent mutations observed at core STR loci in the course of parentage testing.*

STR System	Maternal Meioses (%)	Paternal Meioses (%)	Number from Either	Total Number of Mutations	Mutation Rate (%)
CSF1PO	95/304,307 (0.03)	982/643,118 (0.15)	410	1487/947,425	0.16
FGA	205/408,230 (0.05)	2210/692,776 (0.32)	710	3125/1,101,006	0.28
TH01	31/327,172 (0.009)	41/452,382 (0.009)	28	100/779,554	0.01
TPOX	18/400,061 (0.004)	54/457,420 (0.012)	28	100/857,481	0.01
VWA	184/564,398 (0.03)	1482/873,547 (0.17)	814	2480/1,437,945	0.17
D3S1358	60/405,452 (0.015)	713/558,836 (0.13)	379	1152/964,288	0.12
D5S818	111/451,736 (0.025)	763/655,603 (0.12)	385	1259/1,107,339	0.11
D7S820	59/440,562 (0.013)	745/644,743 (0.12)	285	1089/1,085,305	0.10
D8S1179	96/409,869 (0.02)	779/489,968 (0.16)	364	1239/899,837	0.14
D13S317	192/482,136 (0.04)	881/621,146 (0.14)	485	1558/1,103,282	0.14
D16S539	129/467,774 (0.03)	540/494,465 (0.11)	372	1041/962,239	0.11
D18S51	186/296,244 (0.06)	1094/494,098 (0.22)	466	1746/790,342	0.22
D21S11	464/435,388 (0.11)	772/526,708 (0.15)	580	1816/962,096	0.19
Penta D	12/18,701 (0.06)	21/22,501 (0.09)	24	57/41,202	0.14
Penta E	29/44,311 (0.065)	75/55,719 (0.135)	59	163/100,030	0.16
D2S1338	15/72,830 (0.021)	157/152,310 (0.10)	90	262/225,140	0.12
D19S433	38/70,001 (0.05)	78/103,489 (0.075)	71	187/173,490	0.11
SE33 (ACTBP2)	0/330 (<0.30)	330/51,610 (0.64)	None reported	330/51,940	0.64

Includes compilation of multiple years from American Association of Blood Banks (AABB) 2003 annual report (see http://www.aabb.org/About_the_AABB/Stds_and_Accred/ptannrpt03.pdf, Appendix 2); see also http://www.cstl.nist.gov/biotech/strbase/mutation.htm.

Web Resources

A number of Internet resources regarding STR markers and their use in human identity testing applications are listed in Table 7. More information on some of these resources is described below.

STRBase

One of the most comprehensive and widely used Internet resource on core STR loci involved in human identity testing is the National Institute of Standards and Technology Short Tandem Repeat Internet Database, which is commonly referred to as STRBase (http://www.cstl.nist.gov/biotech/strbase/). This site was created in 1997 by John Butler and Dennis Reeder (84) and has been described by Ruitberg et al. (85). New information is regularly added including variant alleles, triallelic patterns, and addresses for scientists working with STRs. In the past year, new sections of STRBase have been created to describe ongoing efforts with miniSTRs, validation procedures, single nucleotide polymorphisms of forensic interest, Y-chromosome markers and databases, and population data summaries.

Profile Frequency Estimates

Calculations for the rarity of a particular STR profile using core STR loci may be performed over the Internet using several different Web sites. The European Network of Forensic Science Institutes (ENFSI) has sponsored a site that enables different calculations and frequency estimates of an inputted STR profile against 24 different European populations using the SGM Plus kit loci (see http://www.str-base.org/calc.php).

STR profile frequency estimates can also be calculated using Canadian population databases generated by the Centre for Forensic Sciences and the RCMP (along with FBI Caucasian, African American, and Apache databases). This Web site can be accessed at http://www.csfs.ca/pplus/profiler.htm and an index of the available databases can be found at http://www.csfs.ca/databases/index.htm. The FBI raw STR data are publicly available for download at http://www.fbi.gov/hq/lab/fsc/backissu/july1999/dnaloci.txt.

Potential Linkage to Disease Genes

It is important to keep in mind that even though medical genetic researchers claim to have shown linkage between a particular disease gene and a core STR marker, these types of findings are often tentative and should not prevent the continued use of the STR locus in question. In fact, many times these linkage "findings" can later be proven false with further studies, such as with TH01 (86,87). To date there has only been a single call to remove an infrequently used STR marker from future consideration in human

TABLE 7—*Web resources regarding STR markers and forensic DNA typing.*

Short Tandem Repeat Internet Database (STRBase) with details on STR loci http://www.cstl.nist.gov/biotech/strbase
STR profile frequency calculations with SGM Plus loci http://www.str-base.org/index.php
STR profile frequency calculations with Profiler Plus and COfiler loci http://www.csfs.ca/pplus/profiler.htm
The Distribution of the Human DNA-PCR Polymorphisms http://www.uni-duesseldorf.de/WWW/MedFak/Serology/database.html
Y-Chromosome Haplotype Reference Database (YHRD) http://www.yhrd.org/index.html
Progress in Forensic Genetics 9 and 10 (Conference Proceedings of the ISFG) http://www.ics-publishing.com/periodicals/ics
Conference Proceedings of the International Symposium on Human Identification http://www.promega.com/geneticidproc/
Denver District Attorney's Office with DNA court case summaries http://www.denverda.org
FBI's Combined DNA Index System (CODIS) http://www.fbi.gov/hq/lab/codis/index1.htm
Forensic Science Service http://www.forensic.gov.uk
International Society of Forensic Genetics (ISFG) http://www.isfg.org
European DNA Profiling Group (EDNAP) http://www.isfg.org/ednap/ednap.htm
European Network of Forensic Science Institutes (ENFSI) DNA Working Group http://www.enfsi.org/ewg/dnawg

identity testing (88). The X-chromosome STR locus HumARA (11) is a CAG repeat located in a coding region (androgen receptor gene, exon 1) that has been directly linked to several genetic diseases (see (88)). It is probably worthwhile to reiterate that none of the 18 core loci shown in Table 2 and widely used in human identity testing are located in a gene coding region (i.e., exon) or are trinucleotide repeats, which can be prone to expansions that cause genetic defects (89).

An STR profile is simply a string of numbers that provides a unique genetic identifier to a tested sample. Yet because this information ultimately may be linked back to an individual, privacy concerns have been raised as to whether or not predisposition to a genetic disease can be ascertained from the presence of a particular STR allele. In some jurisdictions, there is a perceived problem with using genetic loci that are linked in some form to a genetic disease. Regions of the human genome are being explored with microsatellite (i.e., STR) markers to ascertain disease gene locations through linkage as demonstrated with family studies of effected individuals (90). Colin Kimpton et al. (91) and coworkers from the European DNA Profiling Group recognized early on in the application of STRs for human identity testing that "it is likely that many or possibly most STRs will eventually be shown to be useful in following a genetic disease or other genetic trait within a family and therefore this possibility must be recognized at the outset of the use of such systems."

Indeed, a number of the core STR loci described in this review have been reported to be useful in tracking various genetic diseases through loss of heterozygosity or allelic imbalance. For example, D8S1179 was used to localize a gene connected to Meckel–Gruber syndrome, which is the most common monogenic cause of neural tube defects (92). Another study employing 401 STR markers showed that D8S1179 was the most closely examined locus associated with the gene responsible for urinary microalbuminuria, which impairs kidney function and can lead to an elevated risk for cardiovascular disease (93).

The reason that suspected linkages are even reported in the first place for some of the core STR loci is that many of them are utilized in genome-wide scans in searches for disease-causing genes. For example, the Marshfield panel of more than 400 STRs (Weber set 10) that are spaced across the human genome includes TPOX, D7S820, D8S1179, D13S317, D16S539, and D19S433 (90). It is important to keep in mind that many of the early selections for candidate STR loci by the FSS (12) and by Promega Corporation (94) came from CHLC loci (http://www.chlc.org) that form the basis for genome scans used today for genetic linkage studies. Thus, many of the core STR loci in current use have a common origin to loci widely used for human disease gene linkage analysis studies.

One core STR locus that has gotten a bad reputation over the years for supposed linkage to genetic diseases is TH01, which occurs in the first intron of the tyrosine hydroxylase gene (see Table 3). Allele associations with particular TH01 alleles have been noted for individuals with schizophrenic (95,96) and bipolar disorders (97). However, other researchers failed to confirm these associations (98,99). Likewise, a reported association between TH01 alleles 9.3 and 10 with hypertension (100) was not found with further testing (86). A recent study claims that individuals possessing TH01 allele 7 have less nicotine dependence and are less likely to smoke in a dependent manner, although the data are far from definitive (101).

Trisomy-21, otherwise known as Down's syndrome, can often be detected by the presence of three alleles in any polymorphic marker found on chromosome 21 (102). Certainly, the core STR locus D21S11 qualifies as a useful test for trisomy-21 (103). Likewise, trisomy-18 (Edwards' syndrome) assessment from prenatal samples has been performed with D18S51 (104). In addition, loss of heterozygosity or extreme allelic imbalance is also considered to demonstrate linkage to cancer in some instances (105–107).

Probabilistic Predictions of Sample Ethnicity

Information regarding the probable ethnicity of an unknown offender has the potential to assist investigators in narrowing their search for the true perpetrator, provided that the information is reliable. Since early in the use of DNA typing, efforts have been made to infer ethnic origin from DNA profiles (108). The approach that is generally taken is to examine alleles present in the evidentiary profile and compare them with allele frequencies found in various population data sets. Likelihood ratios can then be created based on competing hypotheses (i.e., that the profile could have come from one population vs another).

Of course this approach requires a number of assumptions, including that the population data sets are representative of individuals coming from a particular ethnic background (109). While any population database with individuals of self-declared ethnicity cannot be regarded as "ethnically pure" and therefore poor calibrators of ethnic origin, efforts have been made to provide a probabilistic prediction with commonly used STR loci (110–112).

Studies involving hundreds of STR loci have found that there are STRs that are more likely to have drastic allele frequency differences between various population groups (see (113,114)). However, it is important to keep in mind that ambiguity is introduced by the relatively high rate of mutation with STR loci (see Table 6), which makes it challenging to separate alleles that are identical by state from those identical by descent (115). Typically single nucleotide polymorphisms (SNPs) or *Alu* insertion elements are more likely to be used for estimating ethnic origin because of their lower mutation rate and the likelihood that a particular allele becomes fixed in a certain population (116–118).

Additional STRs Beyond the Current Core Loci

The efforts of the Human Genome Project have increased knowledge regarding the human genome, and hence there are many more STR loci available now than there were 10 years ago. In fact, more than 20,000 tetranucleotide STR loci have been characterized in the human genome (119) and there may be more than a million STR loci present depending on how they are counted (120). STR sequences account for approximately 3% of the total human genome (121). Yet as noted in the historical perspective at the beginning of this article, even if the initial set of STR loci screened was not substantial, an effective DNA database could only be constructed by generating genotypes with a common set of genetic markers. The current core loci have played and will continue to play a vital role in human identity testing. Commercial STR kits exist, which have further increased the use of these STR loci.

With the fact that millions of DNA samples have now been examined across the core STR loci discussed in this article, it is perhaps worth taking a brief retrospective evaluation and asking the question, are these loci the best available? And if not, what characteristics would be beneficial to future applications in human identity testing? With 20/20 hindsight, are there characteristics for or lessons learned that could be applied in developing additional loci to complement current STR systems (and possibly become part of core loci of the future)?

TABLE 8—*Characteristics of core Y-chromosome STR loci.*

STR Marker	Position (Mb)	Repeat Motif	Allele Range	Mutation Rate (%)	STR Diversity
DYS393	3.17	AGAT	8–17	0.05	0.363
DYS19	10.12	TAGA	10–19	0.20	0.498
DYS391	12.54	TCTA	6–14	0.40	0.552
DYS439	12.95	AGAT	8–15	0.38	0.639
DYS389I/II	13.05	[TCTG] [TCTA]	9–17/24–34	0.20, 0.31	0.538/0.675
DYS438	13.38	TTTTC	6–14	0.09	0.594
DYS390	15.71	[TCTA] [TCTG]	17–28	0.32	0.701
DYS385 a/b	19.19, 19.23	GAAA	7–28	0.23	0.838
DYS392	20.97	TAT	6–20	0.05	0.596

Positions in megabases (Mb) along the Y-chromosome were determined with NCBI build 35 (May 2004) using BLAT. Allele ranges represent the full range of alleles reported in the literature. Mutation rates summarized from YHRD (http://www.yhrd.org; accessed 6 April 2005). The listed STR diversity values are calculated from 244 U.S. Caucasian males (see ref. (134)) and can be helpful in ranking the relative informativeness of the loci.

STR, short tandem repeat.

Certainly for most applications in human identity testing, where a high degree of polymorphism in a marker is advantageous, it is desirable to have loci with better allele frequency distributions than TPOX and TH01. The most common alleles for these two loci can occur at frequencies of greater than 60% in some populations. However, as noted in Table 6, these less polymorphic loci have lower mutation rates, which can make them more useful in some parentage testing situations. Thus, because of different needs, not all human identity testing applications may desire the same characteristics or select the same core STR marker set.

Simple repeat loci are desirable over highly complex loci, such as D21S11 and SE33, with internal sequence variation that can potentially add ambiguity to results and that can only be fully characterized through sequence analysis (rather than PCR product size measurements). However, it should be noted that because of matches at additional loci being tested, it is highly unlikely that a case (i.e., suspect to evidence match) would ever be impacted by potential internal sequence variation at a complex locus such as D21S11. Thus, in practice, forensic DNA testing does not require the sequencing of specific STR alleles to confirm a length-based match discovered at a single complex locus during multiplex STR analysis.

STR loci with a large allele span, such as FGA that possesses alleles spreading across almost 40 repeat units or 160 bp (see Table 2), consume a great deal of potential electrophoretic real estate in STR multiplexes. Two or three moderately polymorphic STR loci on separate chromosomes would be more powerful when the product rule was applied and would easily fit into the same PCR product space. In addition, if a higher molecular weight FGA allele is present in a sample, it undoubtedly will not be amplified as well as a companion lower molecular weight allele. This allele imbalance could even result in allele dropout, particularly in DNA examined from environmentally traumatized samples.

A number of studies have shown what is theoretically predicted—that DNA types can be recovered more effectively from degraded DNA samples when the PCR products are smaller (122–124). Therefore, future loci for consideration in forensic casework applications should contain a more compact allele range and be able to be amplified as small PCR products (125). Unfortunately, core loci such as FGA cannot be made much smaller because of their enormous allele range (125). STR loci that are sufficiently polymorphic and possess a smaller size range do exist and are beginning to be characterized (126).

As new assays that incorporate desirable STR markers (e.g., (126)) are developed, they may still meet some resistance by those who wish to maintain consistency to legacy data in national databases that already contain millions of DNA profiles generated with the previously established core loci. However, it is possible to attach information from additional markers to current STR tests (127); much like the FSS did in the late 1990s, as they added four new STR loci (D2S1338, D3S1358, D16S539, and D19S433) when the U.K. NDNAD went from the six STRs of SGM to the 10 STRs of SGM Plus. As has been noted, mass disaster investigations, which do not rely on large databases constructed over time in many different laboratories, may be more amenable to adopting new loci and assays (128). Recently, there has been a recommendation to adopt three new miniSTR loci (D10S1248, D14S1434, and D22S1045) as part of the standard European loci (127).

Y-Chromosome STR Loci

Although the primary focus of this review is on autosomal STR loci that are widely used for human identity testing, Y chromosome STR loci are growing in popularity and are briefly considered here. The Y chromosome is found only in males, and therefore genetic markers along the Y chromosome can be specific to the male portion of a male–female DNA mixture such as is common in sexual assault cases. Y chromosome markers can also be useful in missing persons investigations, some paternity testing scenarios, historical investigations, and genetic genealogy, because of the fact that most of the Y chromosome (barring mutation) is passed from father to son without changes.

A core set of Y-chromosome STR (Y-STR) loci is widely used in laboratories worldwide for human identity testing and genetic genealogy (129). The minimal haplotype loci (MHL) were selected in the late 1990s from a meager set of available Y-STRs (130,131). The MHL include DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, and the polymorphic, multi-copy marker DYS385. In 2003, the Y-chromosome subcommittee of the Scientific Working Group on DNA Analysis Methods (SWGDAM) recommended two additional Y-STRs named DYS438 and DYS439 for inclusion in the U.S. minimal haplotype (132).

Table 8 contains information on these Y-STR loci, including their chromosomal location, allele ranges, and mutation rates. Alleles observed with Y-STR markers are concatenated to form a haplotype for each examined DNA sample. Y-STR results from individual loci cannot be combined with the product rule, because the core Y-STR loci are all on the nonrecombining portion of the Y chromosome. To date, almost 200 studies have been conducted to examine Y-STR haplotype variation, including one with 2443 male individuals from five North American population groups

TABLE 9—*A summary of locus configuration in Y-chromosome STR typing kits from Promega Corporation and Applied Biosystems.*

PowerPlex Y (kit released in September 2003 by Promega Corporation)			
Dye Label	Locus	Alleles in Ladder	PCR product sizes (Based on DNA Sequence; ref. (136)) (bp)
FL	DYS391	6,8–13	90–118
	DYS389I	10–15	148–168
	DYS439	8–15	203–231
	DYS389II	24–34	256–296
JOE	DYS438	8–12	101–121
	DYS437	13–17	183–199
	DYS19	10–19	232–268
	DYS392	7–18	294–327
TMR	DYS393	8–16	104–136
	DYS390	18–27	191–227
	DYS385 a/b	7–25	243–315
Yfiler (kit released in December 2004 by Applied Biosystems)			
Dye Label	Locus	Alleles in Ladder	PCR product sizes (relative to GS500 LIZ size standard; ref (137)) (bp)
6-FAM	DYS456	13–18	105–124
	DYS389I	10–15	143–165
	DYS390	18–27	192–228
	DYS389II	24–34	253–294
VIC	DYS458	14–20	131–156
	DYS19	10–19	176–211
	DYS385 a/b	7–25	243–318
NED	DYS393	8–16	100–132
	DYS391	7–13	151–176
	DYS439	8–15	198–225
	DYS635	20–26	246–271
	DYS392	7–18	291–327
PET	GATA H4	8–13	122–142
	DYS437	13–17	183–198
	DYS438	8–13	224–249
	DYS448	17–24	280–325

STR, short tandem repeat.

(133). Additional Y-STR markers are also being examined beyond the core loci in order to determine the value of expanding haplotypes generated in the future (134,135).

A number of online databases exist, which permit a comparison of a Y-STR haplotype to those haplotypes already observed in various populations (for a summary of databases, see http://www.cstl.nist.gov/biotech/strbase/y_strs.htm). The largest of these databases is the Y-chromosome haplotype reference database (YHRD; http://www.yhrd.org), which contains over 28,000 haplotypes run with the minimal haplotype loci. Commercial Y-STR kits are now available that amplify the entire set of core Y-STR loci in a single, robust multiplex assay (Table 9). These kits can produce male-specific amplification even in the presence of more than a 1000-fold excess of female DNA (138).

Conclusions

STR markers have become important tools for human identity testing and will continue to be widely used for many years because of their high degree of variability, ease of use in multiplex amplification formats, and implementation in NDNADs (139). Utilization of a uniform set of core STR loci provides the capability for national and international sharing of criminal DNA profiles.

The core loci currently employed in human identity testing have demonstrated their usefulness in aiding the resolution of numerous criminal and parentage testing cases over the past dozen years. Robust commercial STR kits permit reliable amplification of these core loci from small amounts of starting DNA template. Resulting STR profiles enable high powers of discrimination to be achieved among both related and unrelated individuals.

Acknowledgments

I gratefully acknowledge funding from the National Institute of Justice and the support of an incredible research team at NIST, including Margaret Kline, Peter Vallone, and Michael Coble. In particular, the diligent past and present efforts of Christian Ruitberg and Janette Redman for inputting a comprehensive collection of over 2300 STR articles into our Reference Manager database has made the work of creating this review article much easier. The careful review and thoughtful comments of an anonymous reviewer also helped improve this article.

References

1. Butler JM. Forensic DNA typing: biology, technology, and genetics of STR markers. 2nd ed. New York: Elsevier; 2005.
2. Gill P. Role of short tandem repeat DNA in forensic casework in the UK—past, present, and future perspectives. BioTechniques 2002;32:366–72.
3. Jobling MA, Gill P. Encoded evidence: DNA in forensic analysis. Nat Rev Genet 2004;5:739–51.
4. FBI's Combined DNA Index System (CODIS) Homepage; http://www.fbi.gov/hq/lab/codis/index1.htm
5. Forensic Science Service Homepage; http://www.forensic.gov.uk
6. American Association of Blood Banks. Annual Report Summary for Testing in 2003 prepared by the Parentage Testing Standards Program Unit, October 2004. Available at http://www.aabb.org/About_the_AABB/Stds_and_Accred/ptannrpt03.pdf
7. 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, Innsbruck, Austria, June 1998. Madison, WI: Promega Corporation. 1998; 73-88; http://www.promega.com/geneticidproc/eusymp2proc/17.pdf
8. Ballantyne J. Mass disaster genetics. Nat Genet 1997;15:329–31.
9. International Commission on Missing Persons Homepage; http://www.ic-mp.org
10. Edwards A, Civitello A, Hammond HA, Caskey CT. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 1991;49:746–56.
11. Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 1992;12:241–53.
12. Kimpton CP, Gill P, Walton A, Urquhart A, Millican ES, Adams M. Automated DNA profiling employing multiplex amplification of short tandem repeat loci. PCR Meth Appl 1993;3:13–22.
13. Frégeau CJ, Fourney RM. DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. BioTechniques 1993;15:100–19.
14. Kimpton CP, Fisher D, Watson S, Adams M, Urquhart A, Lygo J, et al. Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. Int J Legal Med 1994;106:302–11.
15. Kimpton CP, Oldroyd NJ, Watson SK, Frazier RRE, Johnson PE, Millican ES, et al. Validation of highly discriminating multiplex short tandem repeat amplification systems for individual identification. Electrophoresis 1996;17:1283–93.
16. Werrett DJ. The national DNA database. Forensic Sci Int 1997;88:33–42.
17. Lins AM, Sprecher CJ, Puers C, Schumm JW. Multiplex sets for the amplification of polymorphic short tandem repeat loci—silver stain and fluorescence detection. BioTechniques 1996;20:882–9.
18. Butler JM, Buel E, Crivellente F, McCord BR. Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. Electrophoresis 2004;25:1397–412.
19. Frégeau CJ, Bowen KL, Fourney RM. Validation of highly polymorphic fluorescent multiplex short tandem repeat systems using two generations of DNA sequencers. J Forensic Sci 1999;44(1):133–66.
20. Lins AM, Micka KA, Sprecher CJ, Taylor JA, Bacher JW, Rabbach D, et al. Development and population study of an eight-locus short tandem repeat (STR) multiplex system. J Forensic Sci 1998;43:1168–80.

21. Wallin JM, Holt CL, Lazaruk KD, Nguyen TH, Walsh PS. Constructing universal multiplex PCR systems for comparative genotyping. J Forensic Sci 2002;47:52–65.
22. Holt CL, Buoncristiani M, Wallin JM, Nguyen T, Lazaruk KD, Walsh PS. TWGDAM validation of AmpFlSTR PCR amplification kits for forensic DNA casework. J Forensic Sci 2002;47:66–96.
23. Levedakou EN, Freeman DA, Budzynski MJ, Early BE, Damaso RC, Pollard AM, et al. Characterization and validation studies of powerPlex 2.1, a nine-locus short tandem repeat (STR) multiplex system and penta D monoplex. J Forensic Sci 2002;47:757–72.
24. Krenke BE, Tereba A, Anderson SJ, Buel E, Culhane S, Finis CJ, et al. Validation of a 16-locus fluorescent multiplex system. J Forensic Sci 2002;47:773–85.
25. Collins PJ, Hennessy LK, Leibelt CS, Roby RK, Reeder DJ, Foxall PA. Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpFlSTR Identifiler PCR Amplification Kit. J Forensic Sci 2004;49:1265–77.
26. Bacher J, Schumm JW. Development of highly polymorphic pentanucleotide tandem repeat loci with low stutter. Profiles DNA 1998;2(2): 3–6; http://www.promega.com/profiles/202/ProfilesinDNA_202_03.pdf
27. Schumm JW, Bacher JW. Materials and methods for identifying and analyzing intermediate tandem repeat DNA markers. U.S. Patent 6,238,863 (issued May 29, 2001).
28. Watson S, Kelsey Z, Webb R, Evans J, Gill P. The development of a third generation STR multiplex system (TGM). In: Olaisen B, Brinkmann B, Lincoln PJ, editors. Progress in forensic genetics 7: proceedings of the 17th international ISFH congress, Oslo, Sept. 2–6, 1997. Amsterdam: Elsevier, 1998;192–4.
29. Lareu M-V, Barral S, Salas A, Rodriguez M, Pestoni C, Carracedo A. Further exploration of new STRs of interest for forensic genetic analysis. In: Olaisen B, Brinkmann B, Lincoln PJ, editors. Progress in forensic genetics 7. Proceedings of the 17th International ISFH Congress, Oslo, Sept. 2–6, 1997. Amsterdam: Elsevier, 1998;198–200.
30. Cotton EA, Allsop RF, Guest JL, Frazier RR, Koumi P, Callow IP, et al. Validation of the AMPFlSTR SGM plus system for use in forensic casework. Forensic Sci Int 2000;112:151–61.
31. Wiegand P, Budowle B, Rand S, Brinkmann B. Forensic validation of the STR systems SE33 and TC11. Int J Legal Med 1993;105:315–20.
32. Schneider HR, Rand S, Schmitter H, Weichhold G. ACTBP2—nomenclature recommendations of GEDNAP. Int J Legal Med 1998;111: 97–100.
33. Caskey CT, Edwards AO. DNA typing with short tandem repeat polymorphisms and identification of polymorphic short tandem repeats. U.S. Patent 5,364,759 (issued Nov. 15, 1994).
34. Jäckle H, Tautz D. Process for analyzing length polymorphisms in DNA regions. U.S. Patent RE37,984 (reissued Feb. 11, 2003).
35. Schumm JW, Sprecher CJ. Multiplex amplification of short tandem repeat loci. U.S. Patent 6,221,598 (issued April 24, 2001).
36. Schumm JW, Sprecher CJ. Multiplex amplification of short tandem repeat loci. U.S. Patent 6,479,235 (issued Nov. 12, 2002).
37. Butler JM, Schoske R, Vallone PM, Redman JW, Kline MC. Allele frequencies for 15 autosomal STR loci on US. Caucasian, African American, and Hispanic populations. J Forensic Sci 2003;48:908–11.
38. Reid TM, Ingala DA, Kraemer CM, Dage WM, Dieckhoner C, Fortman J, et al. Distribution of HUMACTBP2 (SE33) alleles in three North American populations. J Forensic Sci 2003;48:1422–3.
39. Levadokou EN, Freeman DA, Budzynski MJ, Early BE, McElfresh KC, Schumm JW, et al. Allele frequencies for fourteen STR loci of the PowerPlex 1.1 and 2.1 multiplex systems and Penta D locus in Caucasians, African-Americans, Hispanics, and other populations of the United States of America and Brazil. J Forensic Sci 2001;46:736–61.
40. Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R. Evaluation of 13 short tandem repeat loci for use in personal identification applications. Am J Hum Genet 1994;55:175–89.
41. Mills KA, Even D, Murray JC. Tetranucleotide repeat polymorphism at the human alpha fibrinogen locus (FGA). Hum Mol Genet 1992; 1:779.
42. Polymeropoulos MH, Xiao H, Rath DS, Merril CR. Tetranucleotide repeat polymorphism at the human tyrosine hydroxylase gene (TH). Nucleic Acids Res 1991;19:3753.
43. Anker R, Steinbrueck T, Donis-Keller H. Tetranucleotide repeat polymorphism at the human thyroid peroxidase (hTPO) locus. Hum Mol Genet 1992;1:137.
44. Kimpton CP, Walton A, Gill P. A further tetranucleotide repeat polymorphism in the vWF gene. Hum Mol Genet 1992;1:287.
45. Li H, Schmidt L, Wei M-H, Hustad T, Lerman MI, Zbar B, et al. Three tetranucleotide polymorphisms for loci: D3S1352, D3S1358, D3S1359. Hum Mol Genet 1993;2:1327.
46. Cooperative Human Linkage Center (http://www.chlc.org) GATA3F03.512
47. Cooperative Human Linkage Center (http://www.chlc.org) GATA3F01.511
48. Cooperative Human Linkage Center (http://www.chlc.org) GATA7G07. 37564
49. Cooperative Human Linkage Center (http://www.chlc.org) GATA7G10.415
50. Cooperative Human Linkage Center (http://www.chlc.org) GATA11C06. 715
51. Straub RE, Speer MC, Luo Y, Rojas K, Overhauser J, Ott J, et al. A microsatellite genetic linkage map of human chromosome 18. Genomics 1993; 15:48–56.
52. Sharma V, Litt M. Tetranucleotide repeat polymorphism at the D21S11 locus. Hum Mol Genet 1992;1:67.
53. Cooperative Human Linkage Center (http://www.chlc.org) GGAA3A09. 31762
54. Cooperative Human Linkage Center (http://www.chlc.org) GGAA2A03. 135
55. Polymeropoulos MH, Rath DS, Xiao H, Merril CR. Tetranucleotide repeat polymorphism at the human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2). Nucleic Acids Res 1992;20:1432.
56. Bar W, Brinkmann B, Budowle B, Carracedo A, Gill P, Lincoln P, et al. DNA recommendations: further report of the DNA commission of the ISFH regarding the use of short tandem repeat systems. Int J Legal Med 1997;110:175–6.
57. Rolf B, Schürenkamp M, Junge A, Brinkmann B. Sequence polymorphism at the tetranucleotide repeat of the human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2) locus. Int J Legal Med 1997;110:69–72.
58. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature 2004;431:931–45.
59. National Research Council. The evaluation of forensic DNA evidence. Washington, DC: National Academy Press; 1996.
60. Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 2002;3:299–309.
61. Sullivan KM, Mannucci A, Kimpton CP, Gill P. A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X–Y homologous gene amelogenin. BioTechniques 1993;15:637–41.
62. Furey TS, Haussler D. Integration of the cytogenetic map with the draft human genome sequence. Hum Mol Genet 2003;12:1037–44.
63. Griffiths RAL, Barber MD, Johnson PE, Gillbard SM, Haywood MD, Smith CD, et al. New reference allelic ladders to improve allelic designation in a multiplex STR system. Int J Legal Med 1998;111:267–72.
64. Egyed B, Furedi S, Angyal M, Boutrand L, Vandenberghe A, Woller J, et al. Analysis of eight STR loci in two Hungarian populations. Forensic Sci Int 2000;113:25–7.
65. Clayton TM, Guest JL, Urquhart AJ, Gill PD. A genetic basis for anomalous band patterns encountered during DNA STR profiling. J Forensic Sci 2004;49:1207–14.
66. Chakhparonian M, Wellinger RJ. Telomere maintenance and DNA replication: how closely are these two connected? Trends Genet 2003;19(8): 439–46.
67. Louis EJ, Vershinin AV. Chromosome ends: different sequences may provide conserved functions. BioEssays 2005;27:685–97.
68. Kline MC, Jenkins B, Rodgers S. Non-amplification of a vWA allele. J Forensic Sci 2003;43:250–1.
69. Walsh S. Commentary on Kline MC, Jenkins B, Rogers S, Non-amplification of a vWA allele. J Forensic Sci 1998;43(1):250.
70. Chakraborty R, De Andrade M, Daiger SP, Budowle B. Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. Ann Hum Genet 1992;56:45–57.
71. Budowle B, Defenbaugh DA, Keys KM. Genetic variation at nine short tandem repeat loci in Chamorros and Filipinos from Guam. Legal Med 2000;2:26–30.
72. Budowle B, Masibay A, Anderson SJ, Barna C, Biega L, Brenneke S, et al. STR primer concordance study. Forensic Sci Int 2001;124:47–54.
73. Budowle B, Sprecher CJ. Concordance study on population database samples using the PowerPlex 16 kit and AmpFlSTR Profiler Plus kit and AmpFlSTR COfiler kit. J Forensic Sci 2001;46:637–41.
74. Alves C, Gusmao L, Pereira L, Amorim A. Multiplex STR genotyping: comparison study, population data and new sequence information. Progress in Forensic Genetics 9 2003; ICS 1239:131–5.
75. Nelson MS, Levedakou EN, Matthews JR, Early BE, Freeman DA, Kuhn CA, et al. Detection of a primer-binding site polymorphism for the STR locus D16S539 using the Powerplex 1.1 system and validation of a

degenerate primer to correct for the polymorphism. J Forensic Sci 2002; 47:345–9.
76. Clayton TM, Hill SM, Denton LA, Watson SK, Urquhart AJ. Primer binding site mutations affecting the typing of STR loci contained within the AMPFlSTR SGM Plus kit. Forensic Sci Int 2004;139:255–9.
77. Lazaruk K, Wallin J, Holt C, Nguyen T, Walsh PS. Sequence variation in humans and other primates at six short tandem repeat loci used in forensic identity testing. Forensic Sci Int 2001;119:1–10.
78. Leibelt C, Budowle B, Collins P, Daoudi Y, Moretti T, Nunn G, et al Identification of a D8S1179 primer binding site mutation and the validation of a primer designed to recover null alleles. Forensic Sci Int 2003;133:220–7.
79. Leclair B, Frégeau CJ, Bowen KL, Fourney RM. Enhanced kinship analysis and STR-based DNA typing for human identification in mass fatality incidents: the Swissair flight 111 disaster. J Forensic Sci 2004;49(5): 939–53.
80. Chakraborty R. Sample size requirements for addressing the population genetic issues of forensic use of DNA typing. Hum Biol 1992;64:141–59.
81. Einum DD, Scarpetta MA. Genetic analysis of large data sets of North American Black, Caucasian, and Hispanic populations at 13 CODIS STR loci. J Forensic Sci 2004;49:1121–3.
82. Weir BS, Bagdonavicius A, Blair B, Eckhoff C, Pearman C, Stringer P, et al. Allele frequency data for profiler plus loci in Australia. J Forensic Sci 2004;49:1121–3.
83. Budowle B, Shea B, Niezgoda S, Chakraborty R. CODIS STR loci data from 41 sample populations. J Forensic Sci 2001;46:453–89.
84. Butler JM, Ruitberg CM, Reeder DJ. STRBase: a short tandem repeat DNA internet-accessible database. Proceedings from the Eighth International Symposium on Human Identification, Scottsdale, AZ, Sept. 1997. Madison, WI: Promega Corporation, 1998:38–47. Available at http://www.promega.com/geneticidproc/ussymp8proc/13.html
85. Ruitberg CM, Reeder DJ, Butler JM. STRBase: a short tandem repeat DNA database for the human identity testing community. Nucleic Acids Res 2001;29:320–2.
86. Klintschar M, Immel U, Stiller D, Kleiber M. HumTH01 and blood pressure. An obstacle for forensic application? In: Doutremepuich C, Morling N, editors. Progress in forensic genetics 10, Vol. 1261. Amsterdam, The Netherlands: International Congress Series, 2004;589–91.
87. Klintschar M, Immel UD, Stiller D, Kleiber M. TH01, a tetrameric short tandem repeat locus in the tyrosine hydroxylase gene: association with myocardial hypertrophy and death from myocardial infarction? Dis Mark 2005;21:9–13.
88. Szibor R, Hering S, Edelmann J. Letter to the editor: the HumARA genotype is linked to spinal and bulbar muscular dystrophy and some further disease risks and should no longer be used as a DNA marker for forensic purposes. Int J Legal Med 2005;119:179–80.
89. Tan EC, Lai PS. Molecular diagnosis of neurogenetic disorders involving trinucleotide repeat expansions. Expert Rev Mol Diagn 2005;5:101–9.
90. Ghebranious N, Vaske D, Yu A, Zhao C, Marth G, Weber JL. STRP screening sets for the human genome at 5 cm density. BMC Genom 2003; 4(6):1–10.
91. Kimpton CP, Gill P, d'Aloja E, Andersen JF, Bar W, Holgersson S, et al. Report on the second EDNAP collaborative STR exercise. Forensic Sci Int 1995;71:137–52.
92. Morgan NV, Gissen P, Sharif SM, Baumber L, Sutherland J, Kelly DA, et al. A novel locus for Meckel–Gruber syndrome, MKS3, maps to chromosome 8q24. Hum Genet 2002;111:456–61.
93. Fox CS, Yang Q, Guo CY, Cupples LA, Wilson PW, Levy D, et al. Genome-wide linkage analysis to urinary microalbuminuria in a community-based sample: the Framingham Heart Study. Kidney Int 2005;67: 70–4.
94. Schumm JW, Lins AM, Micka KA, Sprecher CJ, Rabbach DR, Bacher JW. Automated fluorescent detection of STR multiplexes—development of the *GenePrint* PowerPlex and FFFL multiplexes for forensic and paternity applications. Proceedings from the Seventh International Symposium on Human Identification, Scottsdale, AZ, Oct. 1996. Madison, WI: Promega Corporation, 1997;70–8. Available at http://www.promega.com/geneticidproc/ussymp7proc/0711.html
95. Meloni R, Laurent C, Campion D, Ben Hadjali B, Thibaut F, Dollfus S, et al. A rare allele of a microsatellite located in the tyrosine hydroxylase gene found in schizophrenic patients. C R Acad Sci III 1995;318: 803–9.
96. Thibaut F, Ribeyre JM, Dourmap N, Meloni R, Laurent C, Campion D, et al. Association of DNA polymorphism in the first intron of the tyrosine hydroxylase gene with disturbances of the catecholaminergic system in schizophrenia. Schizophr Res 1997;23:259–64.
97. Meloni R, Leboyer M, Bellivier F, Barbe B, Samolyk D, Allilaire JF, et al. Association of manic-depressive illness with tyrosine hydroxylase microsatellite marker. Lancet 1995;345:932.
98. Burgert E, Crocq MA, Bausch E, Macher JP, Morris-Rosendahl DJ. No association between the tyrosine hydroxylase microsatellite marker HUMTH01 and schizophrenia or bipolar I disorder. Psychiatr Genet 1998;8:45–8.
99. McQuillin A, Lawrence J, Curtis D, Kalsi G, Smyth C, Hannesdottir S, et al. Adjacent genetic markers on chromosome 11p15.5 at or near the tyrosine hydroxylase locus that show population linkage disequilibrium with each other do not show allelic association with bipolar affective disorder. Psychol Med 1999;29:1449–54.
100. Sharma P, Hingorani A, Jia H, Ashby M, Hopper R, Clayton D, et al. Positive association of tyrosine hydroxylase microsatellite marker to essential hypertension. Hypertension 1998;32(4):676–82.
101. Anney RJ, Olsson CA, Lotfi-Miri M, Patton GC, Williamson R. Nicotine dependence in a prospective population-based study of adolescents: the protective role of a functional tyrosine hydroxylase polymorphism. Pharmacogenetics 2004;14:73–81.
102. Liou JD, Chu DC, Cheng PJ, Chang SD, Sun CF, et al. Human chromosome 21-specific DNA markers are useful in prenatal detection of Down syndrome. Ann Clin Lab Sci 2004;34:319–23.
103. Pertl B, Yau SC, Sherlock J, Davies AF, Mathew CG, Adinolfi M. Rapid molecular method for prenatal detection of Down's syndrome. Lancet 1994;343:1197–8.
104. Yoon HR, Park YS, Kim YK. Rapid prenatal detection of down and Edwards syndromes by fluorescent polymerase chain reaction with short tandem repeat markers. Yonsei Med J 2002;43:557–66.
105. Rubocki RJ, Duffy KJ, Shepard KL, McCue BJ, Shepherd SJ, Wisecarver JL. Loss of heterozygosity detected in a short tandem repeat (STR) locus commonly used for human DNA identification. J Forensic Sci 2000;45:1087–9.
106. Poetsch M, Petersmann A, Woenckhaus C, Protzel C, Dittberner T, Lignitz E, et al. Evaluation of allelic alterations in short tandem repeats in different kinds of solid tumors—possible pitfalls in forensic casework. Forensic Sci Int 2004;145:1–6.
107. Vauhkonen H, Hedman M, Vauhkonen M, Kataja M, Sipponen P, Sajantila A. Evaluation of gastrointestinal cancer tissues as a source of genetic information for forensic investigations by using STRs. Forensic Sci Int 2004;139:159–67.
108. Evett IW, Pinchin R, Buffery C. An investigation of the feasibility of inferring ethnic origin from DNA profiles. J Forensic Sci Soc 1992; 32:301–6.
109. Walsh SJ, Triggs CM, Curran JM, Cullen JR, Buckleton JS. Evidence in support of self-declaration as a sampling method for the formation of sub-population DNA databases. J Forensic Sci 2003;48:1091–3.
110. Klintschar M, Al-Hammadi N, Ebner A, Reichenpfader B. Is it possible to determine the ethnic origin of Caucasian individuals using short tandem repeat loci? Prog Forensic Genet 8 2000;ICS 1193: 100–2.
111. Lowe AL, Urquhart A, Foreman LA, Evett IW. Inferring ethnic origin by means of an STR profile. Forensic Sci Int 2001;119:17–22.
112. Rowold DJ, Herrera RJ. Inferring recent human phylogenies using forensic STR technology. Forensic Sci Int 2003;133:260–5.
113. Smith MW, Lautenberger JA, Shin HD, Chretien JP, Shrestha S, Gilbert DA, et al. Markers for mapping by admixture linkage disequilibrium in African American and Hispanic populations. Am J Hum Genet 2001;69:1080–94.
114. Rosenberg NA, Li LM, Ward R, Pritchard JK. Informativeness of genetic markers for inference of ancestry. Am J Hum Genet 2003; 73(6):1402–22.
115. de Knijff P. Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. Am J Hum Genet 2000;67:1055–61.
116. Frudakis T, Venkateswarlu K, Thomas MJ, Gaskin Z, Ginjupalli S, Gunturi S, et al. A classifier for the SNP-based inference of ancestry. J Forensic Sci 2003;48:771–82.
117. Bamshad MJ, Wooding S, Watkins WS, Ostler CT, Batzer MA, Jorde LB. Human population genetic structure and inference of group membership. Am J Hum Genet 2003;72:578–89.
118. Bamshad M, Wooding S, Salisbury BA, Stephens JC. Deconstructing the relationship between genetics and race. Nat Rev Genet 2004;5: 598–609.
119. Collins JR, Stephens RM, Gold B, Long B, Dean M, Burt SK. An exhaustive DNA micro-satellite map of the human genome using high performance computing. Genomics 2003;82:10–19.

120. Ellegren H. Microsatellites: simple sequences with complex evolution. Nat Rev Genet 2004;5:435–45.
121. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860–921.
122. Whitaker JP, Clayton TM, Urquhart AJ, Millican ES, Downes TJ, Kimpton CP, et al. Short tandem repeat typing of bodies from a mass disaster: high success rate and characteristic amplification patterns in highly degraded samples. BioTechniques 1995;18:670–7.
123. Wiegand P, Kleiber M. Less is more—length reduction of STR amplicons using redesigned primers. Int J Legal Med 2001;114: 285–7.
124. Gill P, Curran J, Elliot K. A graphical simulation model of the entire DNA process associated with the analysis of short tandem repeat loci. Nucleic Acids Res 2005;33:632–43.
125. Butler JM, Shen Y, McCord BR. The development of reduced size STR amplicons as tools for analysis of degraded DNA. J Forensic Sci 2003;48(5):1054–64.
126. Coble MD, Butler JM. Characterization of new miniSTR loci to aid analysis of degraded DNA. J Forensic Sci 2005;50:43–53.
127. Gill P, Fereday L, Morling N, Schneider PM. The evolution of DNA databases—recommendations for new European STR loci. Forensic Sci Int 2005, in press.
128. Gill P, Werrett DJ, Budowle B, Gurrier R. An assessment of whether SNPs will replace STRs in national DNA databases—joint considerations of the DNA working group of the European network of forensic science institutes (ENFSI) and the scientific working group on DNA analysis methods (SWGDAM). Sci Justice 2004;44(1):51–3.
129. Butler JM. Recent developments in Y-short tandem repeat and Y-single nucleotide polymorphism analysis. Forensic Sci Rev 2003;15(2):91–111.
130. Kayser M, Caglia A, Corach D, Fretwell N, Gehrig C, Graziosi G, et al. Evaluation of Y-chromosomal STRs: a multicenter study. Int J Legal Med 1997;110:125–33.
131. Schneider PM, Meuser S, Waiyawuth W, Seo Y, Rittner C. Tandem repeat structure of the duplicated Y-chromosomal STR locus DYS385 and frequency studies in the German and three Asian populations. Forensic Sci Int 1998;97:61–70.
132. SWGDAM Y-STR subcommittee. Report on the current activities of the scientific working group on DNA analysis methods (SWGDAM) Y-STR subcommittee. Forensic Sci Comm 2004;6(3); available at http://www.fbi.gov/hq/lab/fsc/backissu/july2004/standards/2004_03_standards03.htm
133. Budowle B, Adamowicz M, Aranda XG, Barna C, Chakraborty R, Cheswick D, et al. Twelve short tandem repeat loci Y chromosome haplotypes: genetic analysis on populations residing in North America. Forensic Sci Int 2005;150:1–15.
134. Schoske R, Vallone PM, Kline MC, Redman JW, Butler JM. High-throughput Y-STR typing of U.S. populations with 27 regions of the Y chromosome using two multiplex PCR assays. Forensic Sci Int 2004;139:107–21.
135. Butler JM, Decker AE, Vallone PM, Kline MC. Allele frequencies for 27 Y-STR loci with U.S. Caucasian, African American, and Hispanic samples. Forensic Sci Int 2005, in press.
136. Promega Corporation. PowerPlex Y system technical manual, part# TMD018; available from http://www.promega.com/tbs/tmd018/tmd018.pdf
137. Applied Biosystems. AmpFlSTR Yfiler PCR amplification kit user's manual; part# 4358101; available from http://www.appliedbiosystems.com/
138. Krenke BE, Viculis L, Richard ML, Prinz M, Milne SC, Ladd C, et al. Validation of a male-specific, 12-locus fluorescent short tandem repeat (STR) multiplex. Forensic Sci Int 2005;148:1–14.
139. National Institute of Justice. The future of forensic DNA testing: predictions of the research and development working group. Washington, DC, Nov 2000; available at http://www.ojp.usdoj.gov/nij/pubs-sum/183697.htm

Additional information and reprint requests:
John M. Butler, Ph.D.
Biotechnology Division
National Institute of Standards and Technology
100 Bureau Dr., Mail Stop 8311
Gaithersburg, MD 20899-8311
E-mail: john.butler@nist.gov