Biology and Genetics of New Autosomal STR Loci Useful for Forensic DNA Analysis

J. M. Butler*, C. R. Hill
National Institute of Standards and Technology
Applied Genetics Group
Gaithersburg, Maryland
United States of America

TABLE OF CONTENTS

INTRODUCTION 16
I. CORE LOCI USED IN THE UNITED STATES AND EUROPE 16
II. COMMERCIAL KITS 18
III. AUTOSOMAL STR LOCI IN CURRENT COMMERCIAL KITS 18
IV. ADDITIONAL AUTOSOMAL STR LOCI 22
V. RELATIVE VARIABILITY OF AUTOSOMAL STR LOCI 22
VI. EXPANSION OF U.S. CORE LOCI 23
CONCLUSIONS 25
ACKNOWLEDGMENTS 25
REFERENCES 25
ABOUT THE AUTHORS 26

* Corresponding author: Dr. John M. Butler, Applied Genetics Group, National Institute of Standards and Technology, 100 Bureau Drive, Mail Stop 8314, Gaithersburg, MD 20899-8312; +1 301 975 4049 (voice); +1 301 975 8505 (fax); john.butler@ nist.gov.

1042-7201/24-01/Jan. 2012/15–26/$12.00 •

Biology and Genetics of New Autosomal STR Loci Useful for Forensic DNA Analysis

REFERENCE: Butler JM, Hill CR: Biology and genetics of new autosomal STR loci useful for forensic DNA analysis; *Forensic Sci Rev* 24:15; 2012.

ABSTRACT: Short tandem repeats (STRs) are regions of tandemly repeated DNA segments found throughout the human genome that vary in length (through insertion, deletion, or mutation) with a core repeated DNA sequence. Forensic laboratories commonly use tetranucleotide repeats, containing a four base pair (4-bp) repeat structure such as GATA. In 1997, the Federal Bureau of Investigation (FBI) Laboratory selected 13 STR loci that form the backbone of the U.S. national DNA database. Building on the European expansion in 2009, the FBI announced plans in April 2011 to expand the U.S. core loci to as many as 20 STRs to enable more global DNA data sharing. Commercial STR kits enable consistency in marker use and allele nomenclature between laboratories and help improve quality control. The STRBase website, maintained by the U.S. National Institute of Standards and Technology (NIST), contains helpful information on STR markers used in human identity testing.

KEY WORDS: Autosomal genetic markers, CODIS STRs, core loci, DNA typing, European Standard Set, expanded U.S. core loci, short tandem repeat (STR), STR kits.

INTRODUCTION

Eukaryotic genomes are full of repeated DNA sequences [15]. These repeated DNA sequences come in all sizes and are typically designated by the length of the core repeat unit and the number of contiguous repeat units or the overall length of the repeat region. Long repeat units may contain several hundred to several thousand bases in the core repeat.

DNA regions with repeat units that are 2 bp to 7 bp in length are called *microsatellites*, *simple sequence repeats* (SSRs), or most commonly *short tandem repeats* (STRs). STRs have become popular DNA repeat markers because they are easily amplified by the polymerase chain reaction (PCR) without the problems of differential amplification. This is because both alleles from a heterozygous individual are similar in size since the repeat size is small. The number of repeats in STR markers can be highly variable among individuals, which make these STRs effective for human identification purposes.

In the past two decades, a number of tetranucleotide STRs have been explored for application to human identification. The types of STR markers that have been sought have included short STRs for typing degraded DNA materials [12], STRs with low stuttering characteristics for analyzing mixtures [1], and male-specific Y chromosome STRs for analyzing male-female mixtures from sexual crimes [11,24]. In order to take advantage of the product rule and be able to combine the genetic information across multiple loci, autosomal STR markers used in forensic DNA typing are typically chosen from separate chromosomes or are widely spaced on the same chromosome to avoid any problems with linkage between the markers.

STR sequences not only vary in the length of the repeat unit and the number of repeats but also in the rigor with which they conform to an incremental repeat pattern. STRs are often divided into several categories based on the repeat pattern. *Simple repeats* contain units of identical length and sequence, *compound repeats* comprise two or more adjacent simple repeats, and *complex repeats* may contain several repeat blocks of variable unit length as well as variable intervening sequences [34]. Not all alleles for an STR locus contain complete repeat units. Even simple repeats can contain nonconsensus alleles that fall in between alleles with full repeat units. *Microvariants* are alleles that contain incomplete repeat units (e.g., the allele 9.3 at the TH01 locus [32]).

I. CORE LOCI USED IN THE UNITED STATES AND EUROPE

For DNA typing markers to be effective across a wide number of jurisdictions, a common set of standardized markers must be used. Since its selection by the Federal Bureau of Investigation (FBI) Laboratory in November 1997 [5,7], a core set of 13 STR loci has been required within the United States to upload DNA profiles to the national DNA database (NDIS, National DNA Index System). The 13 core U.S. loci currently used by the Combined DNA Index System (CODIS) software to enable DNA matches within the United States are CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, and D21S11. Only eight loci overlap with STR data gathered in the United Kingdom and most other European nations (**Table 1**).

Table 1. Comparison of STR loci present in kits used in the United States and Europe[a]

United States[b]				Europe[b]			
Extended[c]	**Powerplex 16**	**Identifiler**	**MiniFiler**	**ESX/ESI17**	**NGM (SElect)**	**SEfiler Plus**	**SGM Plus**
—	TPOX	TPOX	—	—	—	—	—
CSF1PO	CSF1PO	CSF1PO	CSF1PO	—	—	—	—
D5S818	D5S818	D5S818	—	—	—	—	—
D7S820	D7S820	D7S820	D7S820	—	—	—	—
D13S317	D13S317	D13S317	D13S317	—	—	—	—
FGA	*FGA*	*FGA*	FGA	*FGA*	*FGA*	*FGA*	*FGA*
vWA	*vWA*	*vWA*	—	*vWA*	*vWA*	*vWA*	*vWA*
D3S1358	*D3S1358*	*D3S1358*	—	*D3S1358*	*D3S1358*	*D3S1358*	*D3S1358*
D8S1179	*D8S1179*	*D8S1179*	—	*D8S1179*	*D8S1179*	*D8S1179*	*D8S1179*
D18S51	*D18S51*	*D18S51*	D18S51	*D18S51*	*D18S51*	*D18S51*	*D18S51*
D21S11	*D21S11*	*D21S11*	D21S11	*D21S11*	*D21S11*	*D21S11*	*D21S11*
TH01	*TH01*	*TH01*	—	*TH01*	*TH01*	*TH01*	*TH01*
D16S539	*D16S539*	*D16S539*	D16S539	*D16S539*	*D16S539*	*D16S539*	*D16S539*
D2S1338	—	D2S1338	D2S1338	D2S1338	D2S1338	D2S1338	D2S1338
D19S433	—	D19S433	—	D19S433	D19S433	D19S433	D19S433
D12S391	—	—	—	D12S391	D12S391	—	—
D1S1656	—	—	—	D1S1656	D1S1656	—	—
D2S441	—	—	—	D2S441	D2S441	—	—
D10S1248	—	—	—	D10S1248	D10S1248	—	—
—	—	—	—	D22S1045	D22S1045	—	—
—	—			SE33	(SE33)	SE33	—
—	Penta D	—	—	—	—	—	—
Penta E	Penta E	—	—	—	—	—	—
DYS391	—	—	—	—	—	—	—
Amelogenin	Amelogenin	Amelogenin	Amelogenin	Amelogenin	Amelogenin	Amelogenin	Amelogenin

[a] *See* Figure 1 for information on additional kits including Sinofiler (a Chinese-specific kit), which contains D6S1043.
[b] The eight loci in italic are the current overlap between the U.S. and Europe.
[c] Extended U.S. core (proposed April 2011). The 14 loci in bold font show the expanded overlap with the new loci (*see* Ref. [17] for more information on the proposed U.S. extended core).

While there are over 20,000 tetranucleotide STR loci now known to exist in the human genome due to the Human Genome Project efforts completed after the 1997 selection of the 13 CODIS STRs, not many of these loci have extensive population data collected to date — nor have these markers been tested in multiplex PCR assays or with forensic DNA samples. The need to maintain connection to the legacy data for the original STRs, or at least a significant subset of previous data, limits the loci that can be considered for expanding marker sets [6,16].

In April 2009, the European Network of Forensic Science Institute (ENFSI) voted to adopt five additional STR loci (D12S391, D1S1656, D2S441, D10S1248, and D22S1045) to the already existing European Standard Set (ESS) of seven STRs (TH01, vWA, FGA, D8S1179, D18S51, D21S11, and D3S1358). These 12 EU core loci are typically accompanied by D16S539, D2S1338, and D19S433 when amplified with commercially available STR kits (**Figure 1**).

In April 2011, the FBI Laboratory proposed an expanded set of core STR loci for the United States in order to (a) reduce the likelihood of adventitious matches as the number of profiles stored in the U.S. national DNA database increases; (b) increase international compatibility to assist law enforcement data-sharing efforts; and (c) increase the discrimination power to assist missing-persons cases [17]. As noted in Table 1, the proposed required loci for an expanded U.S. core are CSF1PO, FGA, TH01, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D12S391, D1S1656, D2S441, D10S1248, D2S1338, D19S433, Penta E, and DYS391. In addition, the sex-typing marker amelogenin is included as a required locus.

II. COMMERCIAL KITS

Forensic DNA scientists rarely (and in the United States essentially never) use their own STR assays, opting to purchase quality-controlled commercial kits instead. These kits come with allelic ladders, positive control samples, and premixed reagents including the fluorescently labeled oligonucleotides (primers) that target the specific locations in the human genome to be PCR amplified. Table 1 includes some of the commonly used kits in the United States and Europe. Many of these kits enable simultaneous, multicolor fluorescence detection of 15 STRs and the sex-typing marker amelogenin in a single PCR reaction. For example, both Identifiler® and PowerPlex® 16 kits enable typing of the U.S. core 13 STRs plus two additional loci—Identifiler® provides D2S1338 and D19S433 and PowerPlex® 16 provides Penta D and Penta E (Table 1).

To meet the needs of the European community, new STR kits have been released since 2009 by Promega (PowerPlex® ESX and ESI 16 or 17), Applied Biosystems (AmpF*l*STR® NGM & NGM SElect), and Qiagen (ESSplex & ESSplex SE). These kits enable amplification of 15 or 16 STR loci with amelogenin using five-dye chemistry. **Figure 1** illustrates the dye labels and size positions for STR loci within commercial kits available as of September 2011.

Our group at the U.S. National Institute of Standards and Technology (NIST) has been involved in numerous concordance studies to help locate primer binding-site mutations that lead to discordance results between commercial kits using different PCR primers [20,22].

III. AUTOSOMAL STR LOCI IN CURRENT COMMERCIAL KITS

Each of the 24 autosomal STR loci used in common commercial kits has unique characteristics, either in terms of the number of alleles present, the type of repeat sequence, or the kinds of microvariants that have been observed. This section reviews some of the basic details on each of the core and commonly used STR loci that are present in commercial STR kits (**Table 2**).

CSF1PO is a simple tetranucleotide repeat found in the sixth intron of the c-fms proto-oncogene for the CSF-1 receptor on the long arm of chromosome 5. Common alleles contain an AGAT core repeat and range in size from 5 to 17 repeats. The PowerPlex® 16 PCR products for CSF1PO are 41 bp larger than those generated with AmpF*l*STR® kits. Mobility modifiers are included with CSF1PO in the Identifiler® kit to increase the apparent PCR product size by around 25 bp.

FGA is a compound tetranucleotide repeat found in the third intron of the human alpha fibrinogen locus on the long arm of chromosome 4. FGA has also been referred to in the literature as FIBRA or HUMFIBRA. The locus contains a CTTT repeat flanked on either side by degenerate repeats. Reported alleles range in size from 12.2 repeats to 51.2 repeats, spanning over 35 repeats. A 2-bp deletion, from the loss of a CT, in the region just prior to the core repeat motif is responsible for the x.2 microvariant alleles that are very prevalent in this STR system. PCR products from the PowerPlex® 16 STR kit are 112 bp larger than those generated with Applied Biosystems AmpF*l*STR® kits for equivalent alleles.

TH01 is a simple tetranucleotide repeat found in intron 1 of the tyrosine hydroxylase gene on the short arm of chromosome 11. The locus name arises from the initials for tyrosine hydroxylase and intron 1 (i.e., 01). The locus is sometimes incorrectly referred to as THO1, with a letter O instead of a zero. In the literature, TH01 has also been referred to as TC11 and HUMTH01. TH01 has a simple tetranucleotide sequence with a repeat motif of TCAT on the upper strand in the GenBank reference sequence with alleles ranging in size from 3 to 14 repeats. The repeat motif is commonly referenced as AATG, which is correct for the complementary (bottom) strand to the GenBank reference sequence. A common microvariant allele that exists in Caucasians contains a single-base deletion from allele 10 [32] and is designated allele 9.3 [2]. PCR products from the PowerPlex® 16 STR kit are 8 bp smaller than those generated with AmpF*l*STR® kits for equivalent alleles.

TPOX is a simple tetranucleotide repeat found in intron 10 of the human thyroid peroxidase gene near the very end of the short arm of chromosome 2. This STR locus possesses a simple AATG repeat with alleles ranging in size from 4 to 16 repeats. It is the least polymorphic of the 24 commonly used STR loci (see Table 4). PCR products from the PowerPlex® 16 STR kit are 45 bp larger than those generated with AmpF*l*STR® kits for equivalent alleles. Tri-allelic patterns are more prevalent in TPOX than in most other forensic STR markers (*see* STRBase, http://www.cstl.nist.gov/biotech/strbase/tri_tab.htm).

vWA is a compound tetranucleotide repeat found in intron 40 of the von Willebrand Factor gene on the short arm of chromosome 12. VWA has also been referred to in the literature as vWF and vWA. It possesses a TCTA repeat interspersed with a TCTG repeat with alleles ranging in size from 10 to 25 repeats. PCR products from the PowerPlex® 16 kit are 29 bp smaller than those generated with AmpF*l*STR® kits for equivalent alleles.

D3S1358 is a compound tetranucleotide repeat found on the short arm of chromosome 3. This locus possesses both TCTA and TCTG repeat units with alleles ranging in size from 6 to 26 repeats. PCR products from the PowerPlex® 16 kit are 2 bp larger than those generated with AmpF*l*STR® kits for equivalent alleles.

D5S818 is a simple tetranucleotide repeat found on the long arm of chromosome 5. The locus possesses AGAT repeat units with alleles ranging in size from 4 to 29 repeats. In both Promega and Applied Biosystems STR kits, D5S818 is one of the smaller-sized loci and as such should appear more often than some of the other loci in

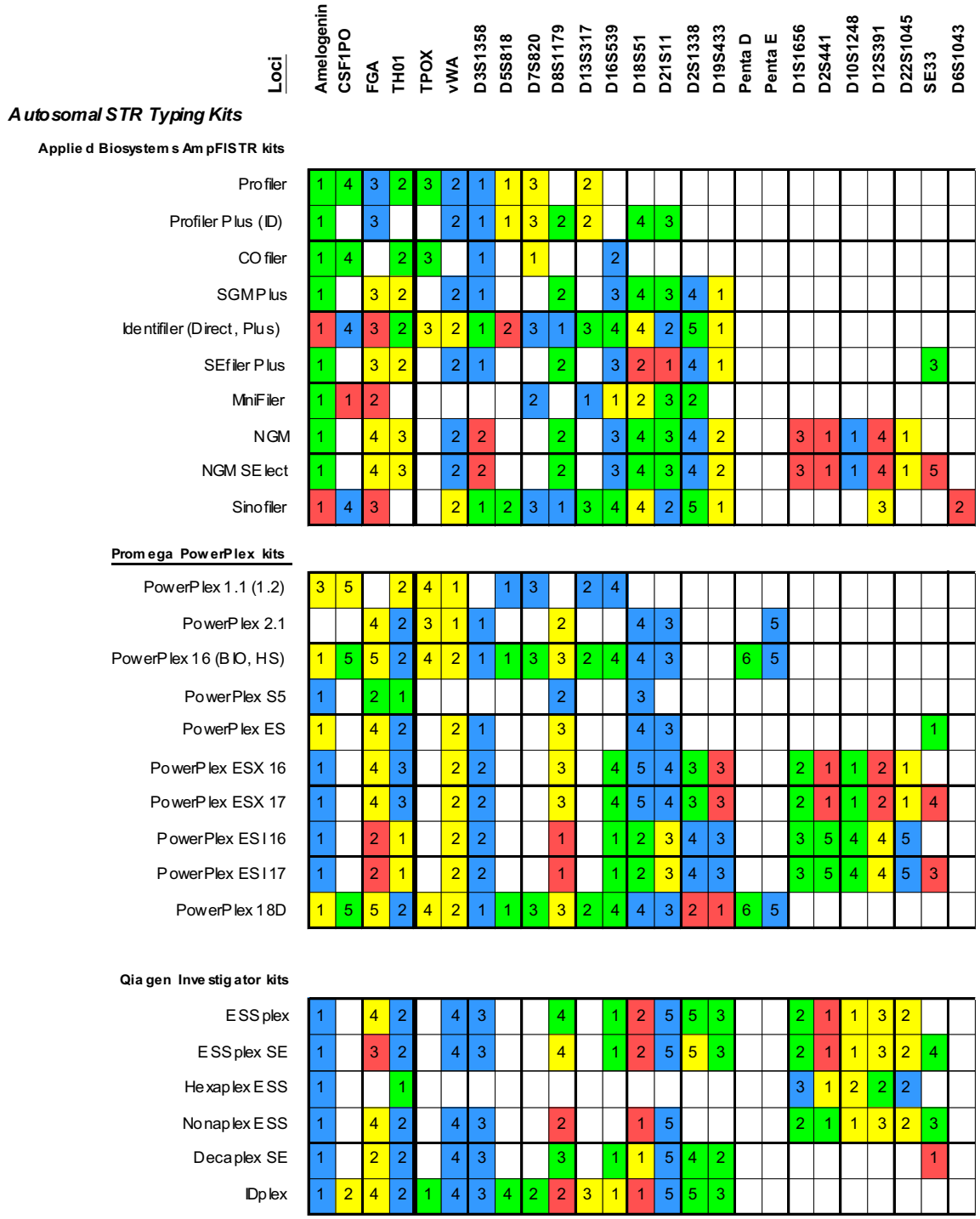

Figure 1. Commercially available STR typing kits with dye labels and size positions for loci present in the kits. (Reproduction with permission from Ref. [6]).

Table 2. Information on 24 commonly used autosomal STR loci present in commercial kits (Adapted from [6,7]; physical positions are from [33])

Locus (UniSTS)[a,b]	Chromosomal location	Physical position (GRCh37 assembly)	GenBank accession (allele repeat #)	Category; repeat motif	Allele range[c]
D1S1656 (58809)	1q42	Chr 1 (230.905 Mb)	G07820 (15.3)	Compound; TAGA	8 to 20.3
TPOX (240638)	2p25.3 thyroid peroxidase, 10th intron	Chr 2 (1.493 Mb)	M68651 (11)	Simple; AATG	4 to 16
D2S441 (71306)	2p14	Chr 2 (68.239 Mb)	AC079112 (12)	Compound; TCTA/TCAA	8 to 17
D2S1338 (30509)	2q35	Chr 2 (218.879 Mb)	AC010136 (23)	Compound; TGCC/TTCC	10 to 31
D3S1358 (148226)	3p21.31	Chr 3 (45.582 Mb)	AC099539 (16)	Compound; TCTA/TCTG	6 to 26
FGA (240635)	4q31.3 alpha fibrinogen, 3rd intron	Chr 4 (155.509 Mb)	M64982 (21)	Compound; CTTT/TTCC	12.2 to 51.2
D5S818 (54700)	5q23.2	Chr 5 (123.111 Mb)	AC008512 (11)	Simple; AGAT	4 to 29
CSF1PO (156169)	5q33.1 c-fms proto-oncogene, 6th intron	Chr 5 (149.455 Mb)	X14720 (12)	Simple; AGAT	5 to 17
SE33 (ACTBP2) (none reported)	6q14 beta-actin related pseudogene	Chr 6 (88.987 Mb)	V00481 (26.2)	Complex; AAAG	3 to 49
D6S1043 (23182)	6q15	Chr 6 (92.450 Mb)	G08539 (11)	Compound; AGAT/AGAC	8 to 25
D7S820 (74895)	7q21.11	Chr 7 (83.789 Mb)	AC004848 (13)	Simple; GATA	5 to 16
D8S1179 (83408)	8q24.13	Chr 8 (125.907 Mb)	AF216671 (13)	Compound; TCTA/TCTG	6 to 20
D10S1248 (51457)	10q26.3	Chr 10 (131.093 Mb)	AL391869 (13)	Simple; GGAA	7 to 19
TH01 (240639)	11p15.5 tyrosine hydroxylase,1st intron	Chr 11 (2.192 Mb)	D00269 (9)	Simple; TCAT	3 to 14
vWA (240640)	12p13.31 von Willebrand Factor, 40th intron	Chr 12 (6.093 Mb)	M25858 (18)	Compound; TCTA/TCTG	10 to 25
D12S391 (2703)	12p13.2	Chr 12 (12.450 Mb)	G08921(20)	Compound; AGAT/AGAC	13 to 27.2
D13S317 (7734)	13q31.1	Chr 13 (82.692 Mb)	AL353628 (11)	Simple; TATC	5 to 17
Penta E (none reported)	15q26.2	Chr 15 (97.374 Mb)	AC027004 (5)	Simple; AAAGA	5 to 32
D16S539 (45590)	16q24.1	Chr. 16 (86.386 Mb)	AC024591 (11)	Simple; GATA	4 to 17
D18S51 (44409)	18q21.33	Chr 18 (60.949 Mb)	AP001534 (18)	Simple; AGAA	5.3 to 40
D19S433 (33588)	19q12	Chr 19 (30.416 Mb)	AC008507 (14)	Compound; AAGG/TAGG	5.2 to 20
D21S11 (240642)	21q21.1	Chr 21 (20.554 Mb)	AP000433 (29.1)	Complex; TCTA/TCTG	12 to 43.2
Penta D (none reported)	21q22.3	Chr 21 (45.056 Mb)	AP001752 (13)	Simple; AAAGA	1.1 to 19
D22S1045 (49680)	22q12.3	Chr 22 (37.536)	AL022314 (17)	Simple; ATT	7 to 20

[a] UniSTS is a comprehensive database of sequence-tagged sites (STSs) available on the NCBI website: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unists.
[b] The 13 CODIS core loci are highlighted in bold font.
[c] *See* Ref. [6], App. 1.

degraded DNA samples. Only a few rare microvariants have been reported at this STR marker and a G→T mutation has been reported 55 bp downstream of the repeat [16]. PCR products from the PowerPlex® 16 kit are 15 bp smaller than those generated with AmpF*l*STR® kits for equivalent alleles.

D7S820 is a simple tetranucleotide repeat found on the long arm of chromosome 7. The locus possesses primarily a GATA repeat with alleles ranging in size from 5 to 16 repeats. However, a number of microvariant alleles have been reported. In some cases, these x.1 and x.3 alleles likely result from a variation in the number of T nucleotides found in a poly(T) stretch that occurs 13 bases downstream of the core GATA repeat. Sequencing has revealed that "on-ladder" alleles contain nine tandem T's while x.3 alleles contain eight T's and x.1 alleles contain 10 T's [14]. PCR products from the PowerPlex® 16 STR kit are 42 bp smaller than those generated with AmpF*l*STR® kits for equivalent alleles.

D8S1179 is a compound tetranucleotide repeat found on chromosome 8. In early publications by the Forensic Science Service, D8S1179 is listed as D6S502 because of a labeling error in the Cooperative Human Linkage Center database from which this STR was chosen [3,30]. The locus consists primarily of alleles containing TCTA although a TCTG repeat unit enters the motif for all alleles larger than 13 repeats, usually at the second or third position from the 5'-end of the repeat region [3]. Alleles range in size from 6 to 20 repeats. PCR products from the PowerPlex® 16 kit are 80 bp larger than those generated with AmpF*l*STR® kits for equivalent alleles. AmpF*l*STR® Identifiler and Profiler Plus ID kits possess an extra, unlabeled D8S1179 reverse primer to prevent allele dropout in a small portion of Asian populations due to a mutation in the middle of the primer-binding site [28].

D13S317 is a simple TATC tetranucleotide repeat found on the long arm of chromosome 13. Common alleles contain between 5 and 17 repeat units. PCR products from the PowerPlex® 16 STR kit are 36 bp smaller than those generated with AmpF*l*STR® kits for equivalent alleles. A 4-bp deletion has been reported 24 bases downstream from the core TATC repeat that can impact allele designations with different primer sets [10,13].

D16S539 is a simple tetranucleotide repeat found on the long arm of chromosome 16. This locus possesses a core repeat unit of GATA ranging from 4 to 17 repeat units in length. PCR products from Promega STR kits are 12 bp larger than those generated with Applied Biosystems kits for equivalent alleles. In the Identifiler® kit, mobility modifiers are used to adjust observed size with D16S539. A point mutation (T→A) 38 bp downstream of the STR repeat impacts the reverse primers for both Applied Biosystems and Promega primer sets. Applied Biosystems added a redundant unlabeled primer in its COfiler®, SGM Plus®, and Identifiler® kits so that both possible alleles could be amplified [35]. On the other hand, Promega altered its D16S539 reverse-primer sequence between kits but kept the overall amplicon size the same [9,26,29].

D18S51 is a simple tetranucleotide AGAA repeat found on the long arm of chromosome 18. Alleles range in size from 5.3 to 40 repeats. A number of x.2 allele variants exist due to a 2-bp deletion from a loss of AG in the 3'-flanking region [3]. More than 70 alleles have been reported for D18S51, making it one of the more polymorphic of the commonly used loci. PCR products from the PowerPlex® 16 kit are 22 bp larger than those generated with AmpF*l*STR® kits for equivalent alleles.

D21S11 is a complex tetranucleotide repeat found on the long arm of chromosome 21. A variable number of TCTA and TCTG repeat blocks surround a constant 43-bp section made up of the sequence {$[TCTA]_3$ TA $[TCTA]_3$ TCA $[TCTA]_2$ TCCA TA}. Alleles range in size from 12 to 43.2 repeats. The x.2 microvariant alleles arise primarily from a 2-bp (TA) insertion on the 3'-end of the repeat region [27]. PCR products from the PowerPlex® 16 kit are 17 bp larger than those generated with AmpF*l*STR® kits for equivalent alleles. D21S11 is far more polymorphic than can be easily detected with size-based length separations. Fine differences in the D21S1 allele structures can only be determined by DNA sequencing since so many of the alleles have the same length but different internal sequence structure because some of the repeat units are switched around. For example, there are four different alleles designated as 30 repeats, which are indistinguishable by size-based methods alone [6].

D2S1338 is a compound tetranucleotide repeat found on the long arm of chromosome 2. Alleles ranging from 10 to 31 repeats have been observed. D2S1338 has a high heterozygosity and is present in the SGM Plus®, Identifiler®, SEfiler Plus™, MiniFiler™, NGM™, and NGM SElect™ kits from Applied Biosystems and the PowerPlex® ESI 16/17, PowerPlex® ESX 16/17, and PowerPlex® 18D kits from Promega.

D19S433 is a compound tetranucleotide repeat located on chromosome 19 with observed alleles ranging from 5.2 to 20 repeats. The x.2 alleles are due to an AG deletion prior to the core AAGG repeat [18]. D19S433 is present in the SGM Plus®, Identifiler®, SEfiler Plus™, NGM™, and NGM SElect™ kits from Applied Biosystems and the PowerPlex® ESI 16/17, PowerPlex® ESX 16/17, and PowerPlex® 18D kits from Promega.

Penta D is a pentanucleotide repeat found on chromosome 21 about 25 Mb from D21S11. Alleles ranging from 1.1 to 19 repeats have been observed although some of the shorter alleles are likely due to flanking region deletions [25]. Penta D is present in the PowerPlex® 16 and PowerPlex® 18D kits.

Penta E is a pentanucleotide repeat with very low stutter-product formation that is located on the long arm of chromosome 15 with alleles ranging from 5 to 32 repeats. Penta E is highly polymorphic (see Table 4) and is present in the PowerPlex® 16 and PowerPlex® 18D kits.

D1S1656 is a tetranucleotide repeat found on the long arm of chromosome 1 with alleles ranging from 8 to 20.3 repeats. The x.3 alleles arise from a TGA insertion typically after four full TAGA repeats. It is part of the extended European Standard Set and is present in NGM™ and NGM SElect™ kits from Applied Biosystems, the PowerPlex® ESI and ESX Systems from Promega, and ESSplex and ESSplex SE kits from Qiagen.

D2S441 is a tetranucleotide repeat located on the short arm of chromosome 2 more than 60 Mb from TPOX. It can be amplified as a miniSTR and works well on degraded DNA samples [12]. A compound repeat motif with TCTA and TCAA sequences can range from 8 to 17 repeats. Some x.3 alleles have been observed as well as same-size, different-sequence alleles [31]. It is part of the extended European Standard Set and is present in NGM™ and NGM SElect™ kits from Applied Biosystems, the PowerPlex® ESI and ESX Systems from Promega, and ESSplex and ESSplex SE kits from Qiagen.

D10S1248 is a simple tetranucleotide repeat found on the long arm of chromosome 10 and possesses 7 to 19 GGAA repeats. It can be amplified as a miniSTR and works well on degraded DNA samples [12]. It is part of the extended European Standard Set and is present in NGM™ and NGM SElect™ kits from Applied Biosystems, the PowerPlex® ESI and ESX Systems from Promega, and ESSplex and ESSplex SE kits from Qiagen.

D12S391 is a highly polymorphic compound tetranucleotide found on the short arm of chromosome 12 only 6.3 megabases from vWA. It possesses over 50 different alleles ranging from 13 to 27.2 repeats in length. A number of same-size, different-sequence alleles have been identified through sequence analysis [27,31]. It is part of the extended European Standard Set and is present in NGM™ and NGM SElect™ kits from Applied Biosystems, the PowerPlex® ESI and ESX Systems from Promega, and ESSplex and ESSplex SE kits from Qiagen.

D22S1045 is a simple trinucleotide repeat found on chromosome 22 with alleles ranging from 7 to 20 ATT repeats. While it is not as polymorphic as most of the other 24 core and common STR loci (see Table 4), it can be amplified as a miniSTR and works well on degraded DNA samples [12]. It is part of the extended European Standard Set and is present in NGM™ and NGM SElect™ kits from Applied Biosystems, the PowerPlex® ESI and ESX Systems from Promega, and ESSplex and ESSplex SE kits from Qiagen.

D6S1043 is a compound tetranucleotide repeat with alleles ranging from 8 to 25 AGAT or AGAC repeats. Some x.2 and x.3 alleles have been reported in population studies. D6S1043 is part of the Sinofiler™ kit and has been used to date almost exclusively in Chinese and other Asian population studies. D6S1043 is located less than 4 Mb from SE33 on the long arm of chromosome 6.

SE33 is the most variable STR locus studied to date. It is located on the long arm of chromosome 6 and contains a core AAAG repeat structure. Appendix 1 in *Advanced Topics in Forensic DNA Typing: Methodology* [6] describes 178 observed alleles ranging from 3 to 49 repeats. For example, sequence analysis has revealed 15 different 29.2 alleles possessing a variety of internal sequence combinations. SE33 is a core locus for the German national DNA database and, with its growing availability in the NGM SElect™ and PowerPlex® ESI 17 and ESX 17 Systems, is being adopted by other laboratories around Europe.

IV. ADDITIONAL AUTOSOMAL STR LOCI

To improve results with challenging DNA samples, a set of 26 autosomal STR loci was characterized in our laboratory at NIST [21]. These loci have also been developed into a 26plex multiplex amplification for rapid examination of reference samples [19]. **Table 3** contains information on these 26 STR loci, which were also selected to be genetically well-spaced from the 24 commercially available loci shown in Table 2.

V. RELATIVE VARIABILITY OF AUTOSOMAL STR LOCI

Table 4 includes a summary of U.S. population results from the 47 autosomal STR loci discussed in this article. The loci are listed in rank order based on their probability of identity and heterozygosity. These results were obtained on a subset of the NIST U.S. population samples [10] involving 249 African American, 162 Caucasian, 139 Hispanic, and 2 Asian samples. A number of the newly available STR loci, such as D12S391 and D1S1656, provide a better probability of identity than widely used loci such as D18S51 and FGA. For each locus in Table 4, calculations are provided for observed heterozygosity (H), polymorphism information content (PIC), and the probability of identity across the combined U.S. population data set. The "genotypes observed" results column can also be a useful metric to locus performance. For example, a comparison of D1S1656 and Penta D in Table 4 is instructive. These two loci both have 15 alleles

Table 3. Information on NIST 26 miniSTR loci [21]

STR locus[a]	Chromosomal location	Physical position[b] (GRCh37 assembly)	GenBank accession (allele repeat #)	Category; repeat motif	Allele range
D1GATA113	1p36.23	Chr 1 (7.443 Mb)	Z97987 (11)	Simple; GATA	7 to 13
D1S1627	1p21.1	Chr 1 (106.964 Mb)	AC093119 (13)	Simple; ATT	10 to 16
D1S1677	1q23.3	Chr 1 (163.560 Mb)	AL513307 (15)	Simple; TTCC	9 to 19
D2S441	2p14	Chr 2 (68.239 Mb)	AC079112 (12)	Compound; TCTA/TCAA	8 to 17
D2S1776	2q24.3	Chr 2 (169.145 Mb)	AC009475 (11)	Simple; AGAT	7 to 15
D3S4529	3p12.1	Chr 3 (85.852 Mb)	AC117452 (13)	Simple; ATCT	11 to 18
D3S3053	3q26.31	Chr 3 (171.751 Mb)	AC069259 (9)	Simple; TATC	7 to 13
D4S2408	4p15.1	Chr 4 (31.304 Mb)	AC110763 (9)	Simple; ATCT	7 to 13
D4S2364	4q22.3	Chr 4 (93.517 Mb)	AC022317 (9)	Simple; ATTC	7 to 11
D5S2500	5q11.2	Chr 5 (58.699 Mb)	AC008791 (17)	Compound; GATA/GATT	14 to 24
D6S1017	6p21.1	Chr 6 (41.677 Mb)	AL035588 (10)	Simple; ATCC	7 to 14
D6S474	6q21	Chr 6 (112.879 Mb)	AL357514 (17)	Complex; GATA/GACA	13 to 20
D8S1115	8p11.21	Chr 8 (42.536 Mb)	AC090739 (9)	Simple; ATT	9 to 20
D9S1122	9q21.2	Chr 9 (79.689 Mb)	AL161789 (12)	Simple; TAGA	9 to 17
D9S2157	9q34.2	Chr 9 (136.035 Mb)	AL162417 (10)	Simple; ATA	7 to 19
D10S1435	10p15.3	Chr 10 (2.243 Mb)	AL354747 (11)	Simple; TATC	5 to 19
D10S1248	10q26.3	Chr 10 (131.093 Mb)	AL391869 (13)	Simple; GGAA	7 to 19
D11S4463	11q25	Chr 11 (130.872 Mb)	AP002806 (14)	Simple; TATC	10 to 17
D12ATA63	12q23.3	Chr 12 (108.322 Mb)	AC009771 (13)	Compound; TAA/CAA	10 to 20
D14S1434	14q32.13	Chr 14 (95.308 Mb)	AL121612 (13)	Complex; CTGT/CTAT	9 to 17
D17S974	17p13.1	Chr 17 (10.519 Mb)	AC034303 (11)	Simple; CTAT	5 to 12
D17S1301	17q25.1	Chr 17 (72.681 Mb)	AC016888 (12)	Simple; AGAT	9 to 15
D18S853	18p11.31	Chr 18 (3.990 Mb)	AP005130 (14)	Simple; ATA	9 to 16
D20S482	20p13	Chr 20 (4.506 Mb)	AL121781 (14)	Simple; AGAT	9 to 19
D20S1082	20q13.2	Chr 20 (53.866 Mb)	AL158015 (14)	Simple; ATA	8 to 17
D22S1045	22q12.3	Chr 22 (37.536 Mb)	AL022314 (17)	Simple; ATT	7 to 20

[a] The three loci in bold font are part of the European standard set included in Table 2.
[b] Physical positions are from [33].

observed, yet D1S1656 has 82 genotypes observed while Penta D only has 68 genotypes observed. The greater number of genotypes formed with the different combinations of alleles in D1S1656 leads to better heterozygosity (0.8804 compared to 0.8786) and probability of identity (0.0229 vs. 0.0336) values. Furthermore, additional genotype combinations mean that D1S1656 will likely be more useful than Penta D for detecting contributors in DNA mixtures. Further analyses with these samples are available in previous publications [20,21].

VI. EXPANSION OF U.S. CORE LOCI

As noted earlier and illustrated in Table 1, the FBI-sponsored CODIS Core Loci Working Group [17] has proposed expanding the required U.S. core loci from the current 13 to 19 autosomal STRs, of which 14 overlap with those commonly used in European STR typing kits, such as NGM™ and PowerPlex® ESX 16 (see Figure 1). In addition, the sex-typing marker amelogenin has been included as a required locus and a Y-chromosome marker DYS391 has been proposed to help confirm sex-typing

Table 4. Results for 47 autosomal STRs with 552 U.S. population samples

STR locus[a]	Alleles observed	Genotypes observed	H(obs)	PIC(exp)	Prob. of identity
SE33	46	229	0.9275	0.9401	0.0079
Penta E	19	111	0.8714	0.9025	0.0167
D2S1338	13	66	0.8822	0.8849	0.0212
D1S1656	15	82	0.8804	0.8830	0.0229
D12S391	24	97	0.8804	0.8758	0.0248
D18S51	19	78	0.8750	0.8734	0.0257
FGA	23	75	0.8859	0.8630	0.0297
Penta D	15	68	0.8786	0.8531	0.0336
D6S1043	23	88	0.8641	0.8460	0.0348
D21S11	22	72	0.8442	0.8349	0.0390
D9S2157	12	56	0.8424	0.8327	0.0395
D12ATA63	11	40	0.8261	0.8142	0.0489
D19S433	16	69	0.8025	0.7901	0.0572
D8S1179	11	42	0.7736	0.7920	0.0581
vWA	10	33	0.8243	0.7878	0.0613
D7S820	11	29	0.8043	0.7674	0.0741
D3S4529	7	22	0.7663	0.7583	0.0753
D16S539	7	26	0.7772	0.7601	0.0756
TH01	8	22	0.7500	0.7499	0.0795
D10S1248	12	34	0.7971	0.7496	0.0820
D6S474	9	27	0.7899	0.7480	0.0837
D1S1677	11	33	0.7554	0.7428	0.0841
D4S2408	7	18	0.7101	0.7348	0.0850
D5S2505	10	31	0.7464	0.7409	0.0853
D22S1045	11	39	0.7717	0.7444	0.0854
D2S1776	9	29	0.7500	0.7357	0.0858
D13S317	8	29	0.7464	0.7432	0.0862
D2S441	12	35	0.7772	0.7416	0.0873
D18S853	8	26	0.7192	0.7263	0.0901
D6S1017	8	28	0.7518	0.7308	0.0907
D3S1358	10	24	0.7699	0.7335	0.0910
D11S4463	8	27	0.7409	0.7187	0.0949
D8S1115	11	43	0.6649	0.7021	0.0966
D10S1435	17	41	0.7591	0.7154	0.0990
CSF1PO	10	30	0.7446	0.7092	0.1022
D17S974	8	21	0.7355	0.7046	0.1065
D5S818	9	32	0.7192	0.6977	0.1096
D1S1627	7	23	0.7373	0.6977	0.1100
D3S3053	7	20	0.7174	0.6882	0.1143
D20S1082	9	28	0.7174	0.6865	0.1190
TPOX	8	26	0.7156	0.6793	0.1226
D14S1434	8	21	0.6902	0.6701	0.1261
D20S482	10	27	0.6703	0.6389	0.1439
D9S1122	9	26	0.7301	0.6560	0.1446
D17S1301	7	19	0.6377	0.6175	0.1590
D1GATA113	7	22	0.6612	0.5999	0.1697
D4S2364	5	9	0.5000	0.4733	0.2791

[a] The 13 CODIS core loci are highlighted in bold font.

results in the event of an amelogenin Y deletion. The primary advantage of using DYS391 is that it is located more than 7 Mb from amelogenin. Moreover, DYS391 is a fairly stable locus possessing few duplications or null alleles with a relatively narrow allele range [8].

In the April 2011 list of proposed new core loci, four loci were listed as "recommended" (Section B) rather than "required" (Section A) [17]. The Section B loci include TPOX, D22S1045, SE33, and Penta D. TPOX is the least variable of the 13 CODIS core loci (see Table 4) and thus first to be considered for removal from the required list without significant impact to overall DNA profile information. D22S1045 is a trinucleotide repeat with accompanying higher stutter products (both N-3 and N+3 stutter) and one of the least variable loci of the expanded European Standard Set (see Table 4). SE33 is a valuable locus in terms of variability but requires a significant amount of "electrophoretic real estate" due to its wide allele range. SE33 also has a much higher mutation rate than other autosomal STR loci and could therefore adversely impact kinship associations in missing-persons cases. Of the two pentanucleotide loci, Penta D is not as polymorphic as Penta E (see Table 4).

While it would be nice to have future STR kits that include all 24 commonly used loci shown in Table 2, there is limited electrophoretic space in creating STR kits with nonoverlapping PCR products that are less than 500 bp in size using the five-dye channels available in current instrumentation. Perhaps with future kits and six-dye instrument capabilities, such as available with the ABI 3500 and 3500xL Genetic Analyzers, all desired STR loci can be incorporated into a single kit.

In the past, INTERPOL has adopted the European Standard Set as its core set of loci. If future STR typing kits are created that are capable of analyzing all of the European loci as a subset of the expanded U.S. core, then there may very well be the possibility of a global autosomal STR panel. Only time will tell if robust kits can be developed to accomplish this feat.

U.S. efforts to expand the number of core loci will likely follow patterns established and lessons learned with the expansion of the European Standard Set [8]. Following initial announcements of new proposed European loci [16], companies produced prototype kits for evaluation. The final expanded set of loci was decided after kit availability and data review. It is important that following the final selection of expanded loci, population data be gathered and software implemented before required compliance with the new loci. In the ongoing U.S. locus expansion effort, as with the 2005–2011 European effort to go from 7 to 12 core loci, data-driven decisions will be made with improved casework capabilities as a priority.

CONCLUSIONS

A growing list of publications describing the application of STR loci to forensic DNA typing has exceeded 3,500 references (see NIST STRBase website, http://www.cstl.nist.gov/biotech/strbase). STR markers have become important tools for human identity testing. Commercially available STR kits are now widely used in forensic and paternity-testing laboratories. The adoption and expansion of core STR loci in national DNA databases around the world ensure that these STR markers will be used for many years to come.

ACKNOWLEDGMENTS

Support of the NIST Human Identity Project Team, including Margaret Kline, Peter Vallone, David Duewer, Mike Coble, Erica Butts, and Kristen O'Connor, is greatly appreciated. Collaborations with Applied Biosystems, Promega Corporation, and Qiagen have enabled the collection of concordance data between many of the commercial kits. Forensic DNA research conducted at NIST is supported by an interagency agreement between the National Institute of Justice and the NIST Office of Law Enforcement Standards. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice. Certain commercial equipment, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by NIST, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

REFERENCES

1. Bacher J, Schumm JW: Development of highly polymorphic pentanucleotide tandem repeat loci with low stutter; *Profiles in DNA* (Promega) 2(2):3; 1998.
2. Bar W, Brinkmann B, Lincoln P, Mayr WR, Rossi U: DNA recommendations — 1994 report concerning further recommendations of the DNA Commission of the ISFH regarding PCR-based polymorphisms in STR (short tandem repeat) systems; *Int J Leg Med* 107:159; 1994.
3. Barber MD, Parkin BH: Sequence analysis and allelic designation of the two short tandem repeat loci D18S51 and D8S1179; *Int J Leg Med* 109:62; 1996.
4. Brinkmann B, Meyer E, Junge A: Complex mutational events at the HumD21S11 locus; *Hum Genet* 98:60; 1996.
5. Budowle B, Moretti TR, Niezgoda SJ, Brown BL: CODIS and PCR-based short tandem repeat loci: Law enforcement tools; *Proceedings — The Second European Symposium on Human Identification*; Innsbruck, Austria; pp 73–88; June 1998; http://www.promega.com/products/pm/genetic-identity/ishi-conference-proceedings/2nd-eshi-oral-presentations/ (Accessed December 2011).
6. Butler JM: *Advanced Topics in Forensic DNA Typing: Methodology*; Elsevier/Academic Press: San Diego, CA; 2012.
7. Butler JM: Genetics and genomics of core short tandem repeat loci used in human identity testing; *J Forensic Sci* 51:253; 2006.
8. Butler JM: NIST update; Presentation — The 17th National CODIS Conference; Jacksville, FL; November 2011); http://www.cstl.nist.gov/strbase/pub_pres/NIST-Update-CODIS2011.pdf (Accessed December 2011).
9. Butler JM, Devaney JM, Marino MA, Vallone PM: Quality control of PCR primers used in multiplex STR amplification reactions; *Forensic Sci Int* 119:87; 2001.
10. Butler JM, Schoske R, Vallone PM, Redman JW, Kline MC: Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations; *J Forensic Sci* 48:908; 2003.
11. Carracedo A, Lareu MV: Development of new STRs for forensic casework: criteria for selection, sequencing & population data and forensic validation. *Proceedings — the Ninth International Symposium on Human Identification*; Orlando, CA; pp 89–107; October 1998; http://www.promega.com/products/pm/genetic-identity/ishi-conference-proceedings/9th-ishi-oral-presentations/ (Accessed December 2011).
12. Coble MD, Butler JM: Characterization of new miniSTR loci to aid analysis of degraded DNA; *J Forensic Sci* 50:43; 2005.
13. Drabek J, Chung DT, Butler JM, McCord BR: Concordance study between Miniplex assays and a commercial STR typing kit; *J Forensic Sci* 49:859; 2004.
14. Egyed B, Furedi S, Angyal M, Boutrand L, Vandenberghe A, Woller J, Padar Z: Analysis of eight STR loci in two Hungarian populations; *Forensic Sci Int* 113: 25; 2000.
15. Ellegren H: Microsatellites: simple sequences with complex evolution; *Nat Rev Genet* 5:435; 2004.
16. Gill P, Fereday L, Morling N, Schneider PM: The evolution of DNA databases-Recommendations for new European STR loci; *Forensic Sci Int* 156:242; 2006.
17. Hares DR: Expanding the CODIS core loci in the United States. *Forensic Sci Int Genet* 6:e52; 2012.
18. Heinrich M, Felske-Zech H, Brinkmann B, Hohoff C: Characterisation of variant alleles in the STR systems D2S1338, D3S1358 and D19S433; *Int J Legal Med* 119:310; 2005.
19. Hill CR, Butler JM, Vallone PM: A 26plex Autosomal STR Assay to Aid Human Identity Testing; *J Forensic Sci* 54:1008; 2009.
20. Hill CR, Duewer DL, Kline MC, Sprecher CJ, McLaren RS, Rabbach DR, Krenke BE, Ensenberger MG, Fulmer PM, Storts DR, Butler JM: Concordance and population studies along with stutter and peak height ratio analysis for the PowerPlex (R) ESX 17 and ESI 17 Systems; *Forensic Sci Int Genet* 5:269; 2011.
21. Hill CR, Kline MC, Coble MD, Butler JM: Characterization of 26 MiniSTR Loci for Improved Analysis of Degraded DNA Samples; *J Forensic Sci* 53:73; 2008.
22. Hill CR, Kline MC, Duewer DL, Butler JM: Strategies for concordance testing; *Profiles in DNA (Promega)*; 2010; http://www.promega.com/resources/articles/profiles-in-

dna/2010/strategies-for-concordance-testing/ (Accessed December 2011).
23. Jiang W, Kline M, Hu P, Wang Y: Identification of dual false indirect exclusions on the D5S818 and FGA loci; *Leg Med (Tokyo)* 13:30; 2011.
24. Kayser M, Kittler R, Ralf A, Hedman M, Lee AC, Mohyuddin A, Mehdi SQ, Rosser Z, Stoneking M, Jobling MA, Sajantila A, Tyler-Smith C: A comprehensive survey of human Y-chromosomal microsatellites; *Am J Hum Genet* 74:1183; 2004.
25. Kline MC, Hill CR, Decker AE, Butler JM: STR sequence analysis for characterizing normal, variant, and null alleles; *Forensic Sci Int Genet* 5:329; 2011.
26. Krenke BE, Tereba A, Anderson SJ, Buel E, Culhane S, Finis CJ, Tomsey CS, Zachetti JM, Masibay A, Rabbach DR, Amiott EA, Sprecher CJ: Validation of a 16-locus fluorescent multiplex system; *J Forensic Sci* 47:773; 2002.
27. Lareu MV, Pestoni MC, Barros F, Salas A, Carracedo A: Sequence variation of a hypervariable short tandem repeat at the D12S391 locus; *Gene* 182:151; 1996.
28. Leibelt C, Budowle B, Collins P, Daoudi Y, Moretti T, Nunn G, Reeder D, Roby R: Identification of a D8S1179 primer binding site mutation and the validation of a primer designed to recover null alleles; *Forensic Sci Int* 133:220; 2003.
29. Masibay A, Mozer TJ, Sprecher C: Promega Corporation reveals primer sequences in its testing kits [letter]; *J Forensic Sci* 45:1360; 2000.
30. Oldroyd NJ, Urquhart AJ, Kimpton CP, Millican ES, Watson SK, Downes T, Gill PD: A highly discriminating octoplex short tandem repeat polymerase chain reaction system suitable for human individual identification; *Electrophoresis* 16:334; 1995.
31. Phillips C, Fernandez-Formoso L, Garcia-Magarinos M, Porras L, Tvedebrink T, Amigo J, Fondevila M, Gomez Tato A, Alvarez-Dios J, Freire-Aradas A, Gomez-Carballa A, Mosquera-Miguel A, Carracedo A, Lareu MV: Analysis of global variability in 15 established and 5 new European Standard Set (ESS) STRs using the CEPH human genome diversity panel; *Forensic Sci Int Genet* 5:155; 2011.
32. Puers C, Hammond HA, Jin L, Caskey CT, Schumm JW: Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUMTH01 [AATG]n and reassignment of alleles in population analysis by using a locus-specific allelic ladder; *Am J Hum Genet* 53:953; 1993.
33. Thanakiatkrai P, Welch L: Evaluation of nucleosome forming potentials (NFPs) of forensically important STRs; *Forensic Sci Int Genet* 5:285; 2011.
34. Urquhart A, Kimpton CP, Downes TJ, Gill P: Variation in short tandem repeat sequences — A survey of twelve microsatellite loci for use as forensic identification markers; *Int J Leg Med* 107:13; 1994.
35. Wallin JM, Holt CL, Lazaruk KD, Nguyen TH, Walsh PS: Constructing universal multiplex PCR systems for comparative genotyping; *J Forensic Sci* 47:52; 2002.

ABOUT THE AUTHORS

J. M. Butler
C. R. Hill

John M. Butler holds a B.S. degree in chemistry from Brigham Young University (Provo, UT) and a Ph.D. in analytical chemistry from the University of Virginia (Charlottesville, VA). He is currently a NIST Fellow and Applied Genetics Group Leader at the National Institute of Standards and Technology (Gaithersburg, MD). Dr. Butler serves on the Scientific Working Group for DNA Analysis Methods (SWGDAM) as chair of the Mixture Interpretation Committee. He is a member of the Department of Defense DNA Quality Assurance Committee, the Virginia Department of Forensic Sciences Science Advisory Committee, and the FBI's CODIS Core Loci Working Group. Dr. Butler is the author of *Forensic DNA Typing* – a textbook now in its third edition that is used worldwide to train people in this important field. He also created and maintains the NIST STRBase website (http://www.cstl.nist.gov/strbase) with useful information on short tandem repeat markers used in forensic DNA analysis.

Carolyn R. (Becky) Hill holds a B.S. degree in biology from the University of Virginia (Charlottesville, VA) and a master's degree in molecular biology from George Mason University (Fairfax, VA). She is a research biologist in the Applied Genetics Group at the National Institute of Standards and Technology (Gaithersburg, MD). Her research focuses on developing new PCR assays and performing STR kit concordance studies using U.S. population samples.