CHAPTER 5

COMMONLY USED SHORT TANDEM REPEAT MARKERS AND COMMERCIAL KITS

> Ever since their discovery in the early 1980s, the ubiquitous occurrence of microsatellites – also referred to as short tandem repeats (STRs) or simple sequence repeats (SSRs)– has puzzled geneticists… [Understanding STRs] is important if we wish to understand how genomes are organized and why most genomes are filled with sequences other than genes.
>
> (Hans Ellegren 2004)

REPEATED DNA

Eukaryotic genomes are full of repeated DNA sequences (see Ellegren 2004). These repeated DNA sequences come in all types of 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.

These regions are often referred to as *satellite* DNA and may be found surrounding the chromosomal centromere. The term satellite arose due to the fact that frequently one or more minor satellite bands were seen in early experiments involving equilibrium density gradient centrifugation (Britten and Kohne 1968, Primrose 1998).

The core repeat unit for a medium length repeat, sometimes referred to as a *minisatellite* or a VNTR (variant number of tandem repeats), is in the range of approximately 10–100 bases in length (Tautz 1993, Chambers and MacAvoy 2000). The forensic DNA marker D1S80 is a minisatellite with a 16 bp repeat unit and contains alleles spanning the range of 16–41 repeat units (Kasai *et al.* 1990).

DNA regions with repeat units that are 2–6 bp in length are called *microsatellites*, simple sequence repeats (SSRs), or short tandem repeats (STRs). STRs have become popular DNA repeat markers because they are easily amplified by the polymerase chain reaction without the problems of differential amplification. This is due to the fact that 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.

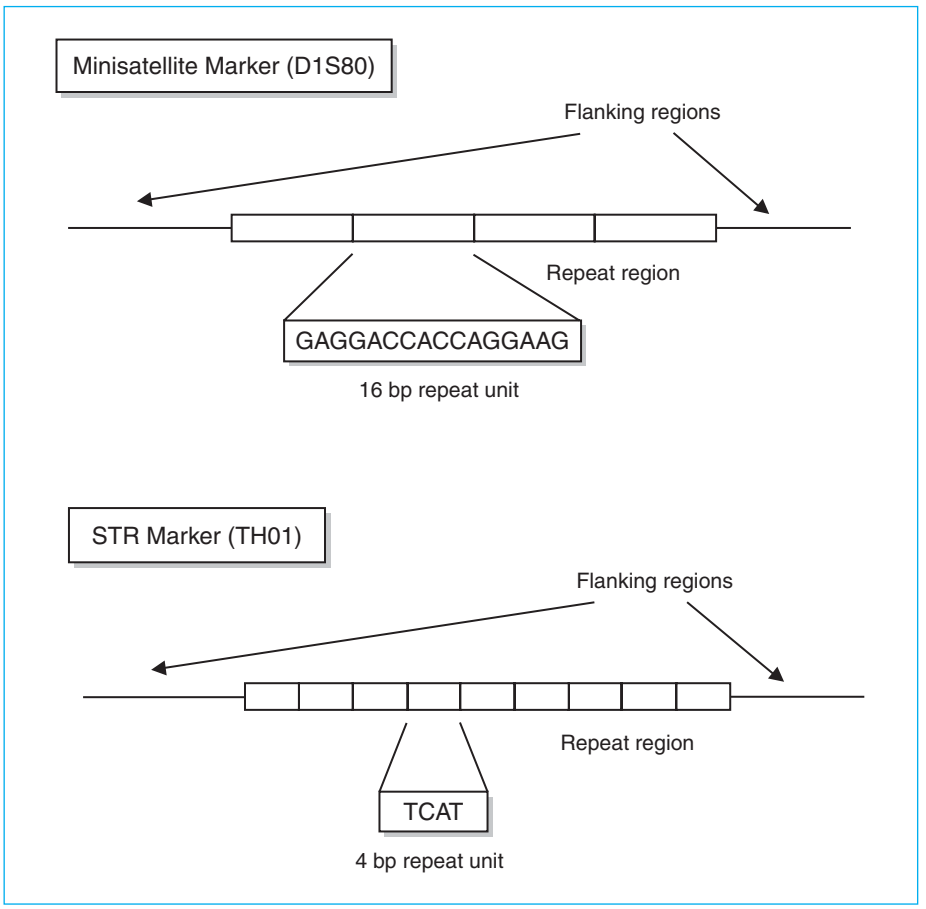


Figure 5.1
Schematic of minisatellite and microsatellite (STR) DNA markers. PCR primers are designed to target invariant flanking sequence regions. The number of tandem repeat units in the repeat regions varies among individuals making them useful markers for human identification.

Literally thousands of polymorphic microsatellites have been characterized in human DNA and there may be more than a million microsatellite loci present depending on how they are counted (Ellegren 2004). Regardless, microsatellites account for approximately 3% of the total human genome (International Human Genome Sequencing Consortium 2001). STR markers are scattered throughout the genome and occur on average every 10 000 nucleotides (Edwards *et al.* 1991, Collins *et al.* 2003, Subramanian *et al.* 2003). Computer searches of the recently available human genome reference sequence have attempted to comprehensively catalog the number and nature of STR markers in the genome (see Collins *et al.* 2003, Subramanian *et al.* 2003). A large number of STR markers have been characterized by academic and commercial laboratories for use in disease gene location studies. For example, the Marshfield Medical Research Foundation in Marshfield, Wisconsin (http://research.marshfieldclinic.org/genetics/) has gathered genotype data on over 8000 STRs that are scattered across the 23 pairs of human chromosomes

for the purpose of developing human genetic maps (Broman *et al.* 1998, Ghebranious *et al.* 2003).

ISOLATION AND TYPES OF STR MARKERS

In order to perform analysis on STR markers, the invariant flanking regions surrounding the repeats must be determined. Once the flanking sequences are known then PCR primers can be designed and the repeat region amplified for analysis. New STR markers are usually identified in one of two ways: (1) searching DNA sequence databases such as GenBank for regions with more than six or so contiguous repeat units (Weber and May 1989, Collins *et al.* 2003, Subramanian *et al.* 2003); or (2) performing molecular biology isolation methods (Edwards *et al.* 1991, Chambers and MacAvoy 2000).

STR repeat sequences are named by the length of the repeat unit. Dinucleotide repeats have two nucleotides repeated next to each other over and over again. Trinucleotides have three nucleotides in the repeat unit, tetranucleotides have four, pentanucleotides have five, and hexanucleotides have six repeat units in the core repeat. Theoretically, there are 4, 16, 64, 256, 1024, 4096 possible motifs for mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats, respectively (Jin *et al.* 1994). However, because microsatellites are tandemly repeated, some motifs are actually equivalent to others (D.N.A. Box 5.1). For reasons that will be discussed below, tetranucleotide repeats have become the most popular STR markers for human identification.

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 (Urquhart *et al.* 1994). *Complex hypervariable repeats* also exist with numerous non-consensus alleles that differ in both size and sequence and are therefore challenging to genotype reproducibly (Urquhart *et al.* 1993, Gill *et al.* 1994). This last category of STR markers is not as commonly used in forensic DNA typing due to difficulties with allele nomenclature and measurement variability between laboratories, although two commercial kits now include the complex hypervariable STR locus SE33, sometimes called ACTBP2 (Urquhart *et al.* 1993, Promega Corporation 2002, Applied Biosystems 2002).

Not all alleles for a STR locus contain complete repeat units. Even simple repeats can contain non-consensus alleles that fall in between alleles with full repeat units. *Microvariants* are alleles that contain incomplete repeat units. Perhaps the most common example of a microvariant is the allele 9.3 at the TH01 locus, which contains nine tetranucleotide repeats and one incomplete

D.N.A. Box 5.1
List of possible microsatellite motifs

Theoretically, there are 4, 16, 64, 256, 1024, 4096 possible motifs for mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats, respectively. However, because microsatellites are tandemly repeated, some motifs are actually equivalent to others. Two rules can be used to identify whether motif A is equivalent to motif B. Motif A is considered equivalent to motif B when (1) motif A is inversely complementary to motif B, or (2) motif A is different from motif B or the inversely complementary sequence of motif B by frameshift. For example, $(GAAA)_n$ is equivalent to $(AGAA)_n$ or $(AAGA)_n$, to $(AAAG)_n$ or $(TTTC)_n$, to $(TTCT)_n$ or $(TCTT)_n$, or to $(CTTT)_n$. In other words, the eight motifs are equivalent. Note that $(AGAG)_n$ is considered a dinucleotide repeat instead of a tetranucleotide motif (Jin *et al.* 1994).

Because of this equivalence in repeat motif structure there are only 2, 4, 10, 33, 102, and 350 possible motifs for mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats, respectively (see below).

Mononucleotide repeats (2):

A C

Dinucleotide repeats (4):

AC AG AT CG

Trinucleotide repeats (10):

AAC AAG AAT ACC ACG ACT AGC AGG ATC CCG

Tetranucleotide repeats (33):

AAAC	AAAG	AAAT	AACC	AACG	AACT	AAGC	AAGG	AAGT	AATC
AATG	AATT	ACAG	ACAT	ACCC	ACCG	ACCT	ACGC	ACGG	ACGT
ACTC	ACTG	**AGAT**	AGCC	AGCG	AGCT	AGGC	AGGG	ATCC	ATCG
ATGC	CCCG	CCGG							

AGAT or GATA motif is the most common for STR loci used by forensic scientists

Penta- (102) and hexanucleotide (350) repeats are not shown due to the sheer number of motifs possible.

Source:

Jin, L., Zhong, Y. and Chakraborty, R. (1994) The exact numbers of possible microsatellite motifs [letter]. *American Journal of Human Genetics*, 55, 582–583.

repeat of three nucleotides because the seventh repeat is missing a single adenine out of the normal AATG repeat unit (Puers *et al.* 1993).

DESIRABLE CHARACTERISTICS OF STRs USED IN FORENSIC DNA TYPING

For human identification purposes it is important to have DNA markers that exhibit the highest possible variation or a number of less polymorphic markers that can be combined in order to obtain the ability to discriminate between samples.

As will be discussed further in Chapter 7, forensic specimens are often challenging to PCR amplify because the DNA in the samples may be severely degraded (i.e., broken up into small pieces). Mixtures are prevalent as well in some forensic samples, such as those obtained from sexual assault cases containing biological material from both the perpetrator and victim.

The small size of STR alleles (~100–400 bp) compared to minisatellite VNTR alleles (~400–1000 bp) make the STR markers better candidates for use in forensic applications where degraded DNA is common. PCR amplification of degraded DNA samples can be better accomplished with smaller product sizes (see Chapter 7). Allelic dropout of larger alleles in minisatellite markers caused by preferential amplification of the smaller allele is also a significant problem with minisatellites (Tully *et al.* 1993). Furthermore, single base resolution of DNA fragments can be obtained more easily with sizes below 500 bp using denaturing polyacrylamide gel electrophoresis (see Chapter 12). Thus, for both biology and technology reasons the smaller STRs are advantageous compared to the larger minisatellite VNTRs.

Among the various types of STR systems, tetranucleotide repeats have become more popular than di- or trinucleotides. Penta- and hexanucleotide repeats are less common in the human genome but are being examined by some laboratories (Bacher *et al.* 1999). As will be discussed in Chapter 6, a biological phenomenon known as 'stutter' results when STR alleles are PCR amplified. Stutter products are amplicons that are typically one or more repeat units less in size than the true allele and arise during PCR because of strand slippage (Walsh *et al.* 1996). Depending on the STR locus, stutter products can be as large as 15% or more of the allele product quantity with tetranucleotide repeats. With di- and trinucleotides, the stutter percentage can be much higher (30% or more) making it difficult to interpret sample mixtures (see Chapter 7). In addition, the four base spread in alleles with tetranucleotides makes closely spaced heterozygotes easier to resolve with size-based electrophoretic separations (see Chapter 12) compared to alleles that could be two or three bases different in size with dinucleotides and trinucleotide markers, respectively.

Thus, to summarize, the advantages of using tetranucleotide STR loci in forensic DNA typing over VNTR minisatellites or di- and trinucleotide repeat STRs include:

- A narrow allele size range that permits multiplexing;
- A narrow allele size range that reduces allelic dropout from preferential amplification of smaller alleles;
- The capability of generating small PCR product sizes that benefit recovery of information from degraded DNA specimens; and
- Reduced stutter product formation compared to dinucleotide repeats that benefit the interpretation of sample mixtures.

In the past decade, a number of tetranucleotide STRs have been explored for application to human identification. The types of STR markers that have been sought out have included short STRs for typing degraded DNA materials, STRs with low stuttering characteristics for analyzing mixtures, and male-specific Y chromosome STRs for analyzing male-female mixtures from sexual crimes (Carracedo and Lareu 1998). The selection criteria for candidate STR loci in human identification applications include the following characteristics (Gill *et al.* 1996, Carracedo and Lareu 1998):

- High discriminating power, usually > 0.9, with observed heterozygosity > 70%;
- Separate chromosomal locations to ensure that closely linked loci are not chosen;
- Robustness and reproducibility of results when multiplexed with other markers;
- Low stutter characteristics;
- Low mutation rate; and
- Predicted length of alleles that fall in the range of 90–500 bp with smaller sizes better suited for analysis of degraded DNA samples.

In order to take advantage of the product rule, STR markers used in forensic DNA typing are typically chosen from separate chromosomes to avoid any problems with linkage between the markers (see Chapter 20).

A COMMON NOMENCLATURE FOR STR ALLELES

To aid in inter-laboratory reproducibility and comparisons of data, a common nomenclature has been developed in the forensic DNA community. DNA results cannot be effectively shared unless all parties are speaking the same language and referring to the same conditions. (It would do little good to describe the recipe for baking a cake in a language that is not understood by both the recipe giver and the chef. For example, if the recipe says to turn the oven on to 450 degrees Fahrenheit and the chef uses 450 Kelvin (~250°F), the results would be vastly different.) Likewise if one laboratory calls a sample 15 repeats at a particular STR locus and the same sample is designated 16 repeats by another laboratory, a match would not be considered, and the samples would be assumed to come from separate sources. As will be discussed in Chapter 18, the advent of national DNA databases with many laboratories worldwide contributing information has made it crucial to have internationally accepted nomenclature for designating STR alleles.

A repeat sequence is named by the structure (base composition) of the core repeat unit and the number of repeat units. However, because DNA has two strands, which may be used to designate the repeat unit for a particular STR marker, more than one choice is available and confusion can arise without a standard format. Also, where an individual starts counting the number of

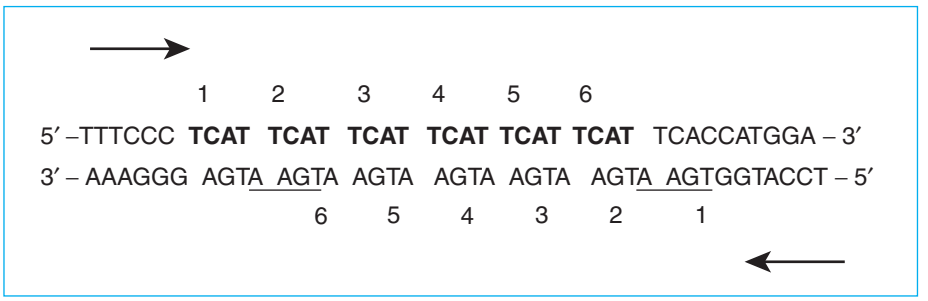


Figure 5.2
Example of the DNA sequence in a STR repeat region. Note that using the top strand versus the bottom strand results in different repeat motifs and starting positions. In this example, the top strand has six TCAT repeat units, while the bottom strand has six TGAA repeat units. Under ISFH recommendations (Bar et al. *1997), the top strand from GenBank should be used. Thus, this example would be described as having [TCAT] as the repeat motif. Repeat numbering, indicated above and below the sequence, proceeds in the 5′ to 3′ direction as illustrated by the arrows.*

repeats can also make a difference. With double-stranded DNA sequences being read in the 5′ to 3′ direction (see Chapter 2), the choice of the strand impacts the sequence designation. For example, the 'top' strand for an STR marker may be 5′-…$(GATA)_n$…-3′ while the 'bottom' strand for the same sequence would be 5′-…$(TATC)_n$ …-3′. Depending on the sequence surrounding the repeat region, the core repeat could be shifted relative to the other strand (Figure 5.2).

Recognizing the need for standardization in STR repeat nomenclature, a committee of forensic DNA scientists, known as the DNA Commission of the International Society of Forensic Haemogenetics (ISFH), issued guidelines for designating STR alleles in 1994 (Bar *et al.* 1994) and again in 1997 (Bar *et al.* 1997). The ISFH is now known as the International Society of Forensic Genetics (ISFG; see http://www.isfg.org/). The ISFG 1994 recommendations focused on allelic ladders and designation of alleles that contain partial repeat sequences. The ISFG 1997 guidelines discuss the sequence and repeat designation of STRs.

When reviewing the STR literature prior to 1997, an individual should keep in mind that repeat nomenclatures often differ from the ISFG 1997 guidelines. This fact can lead to some confusion if one is not careful. For example, early descriptions of the STR locus TH01 by the Forensic Science Service label the repeat TCAT (Kimpton *et al.* 1993) while Caskey and co-workers described the TH01 repeat as AATG (Edwards *et al.* 1991).

The latest ISFG recommendations are reviewed below (Bar *et al.* 1997):

Choice of the Strand

- For STRs within protein coding regions (as well as in the intron of the genes), the coding strand should be used. This would apply to STRs such as VWA (GenBank: M25716), TPOX (GenBank: M68651), and CSF1PO (GenBank: X14720).
- For repetitive sequences without any connection to protein coding genes like many of the D#S### loci, the sequence originally described in the literature of the first public database entry shall become the standard reference (and strand) for nomenclature. Examples here include D18S51 (GenBank: L18333) and D21S11 (GenBank: M84567).

- If the nomenclature is already established in the forensic field but not in accordance with the aforementioned guideline, the nomenclature shall be maintained to avoid unnecessary confusion. This recommendation applies to the continued use by some laboratories of the 'AATG repeat' strand for the STR marker TH01. The GenBank sequence for TH01 uses the coding strand and therefore contains the complementary 'TCAT repeat' instead.

Choice of the Motif and Allele Designation

- The repeat sequence motif should be defined so that the first 5′-nucleotides that can define a repeat motif are used. For example, 5′-GG TCA TCA TCA TGG-3′ could be seen as having 3 × TCA repeats or 3 × CAT repeats. However, under the recommendations of the ISFH committee only the first one (3 × TCA) is correct because it defines the first possible repeat motif.
- Designation of incomplete repeat motifs should include the number of complete repeats and, separated by a decimal point, the number of base pairs in the incomplete repeat. Examples of 'microvariants' with incomplete repeat units include allele 9.3 at the TH01 locus. TH01 allele 9.3 contains nine tetranucleotide AATG repeats and one incomplete ATG repeat of three nucleotides (Puers *et al.* 1993). Another microvariant example is allele 22.2 at the FGA locus, which contains 22 tetranucleotide repeats and one incomplete repeat with two nucleotides (Barber *et al.* 1996).
- Allelic ladders containing sequenced alleles that are named according to the recommendations listed above should be used as a reference for allele designation in unknown samples. Allelic ladders may be commercially obtained or prepared in house and should contain all common alleles.

ALLELIC LADDERS

An allelic ladder is an artificial mixture of the common alleles present in the human population for a particular STR marker (Sajantila *et al.* 1992). They are generated with the same primers as tested samples and thus provide a reference DNA size for each allele included in the ladder. Allelic ladders have been shown to be important for accurate genotype determinations (Smith 1995). These allelic ladders serve as a standard like a measuring stick for each STR locus. They are necessary to adjust for different sizing measurements obtained from different instruments and conditions used by various laboratories (see Chapters 14 and 15).

Allelic ladders are constructed by combining genomic DNA or locus-specific PCR products from multiple individuals in a population, which possess alleles that are representative of the variation for the particular STR marker (Sajantila *et al.* 1992, Baechtel *et al.* 1993). The samples are then co-amplified to produce an artificial sample containing the common alleles for the STR marker (Figure 5.3). Allele quantities are balanced by adjusting the input amount of each component so that the alleles are fairly equally represented in the ladder.

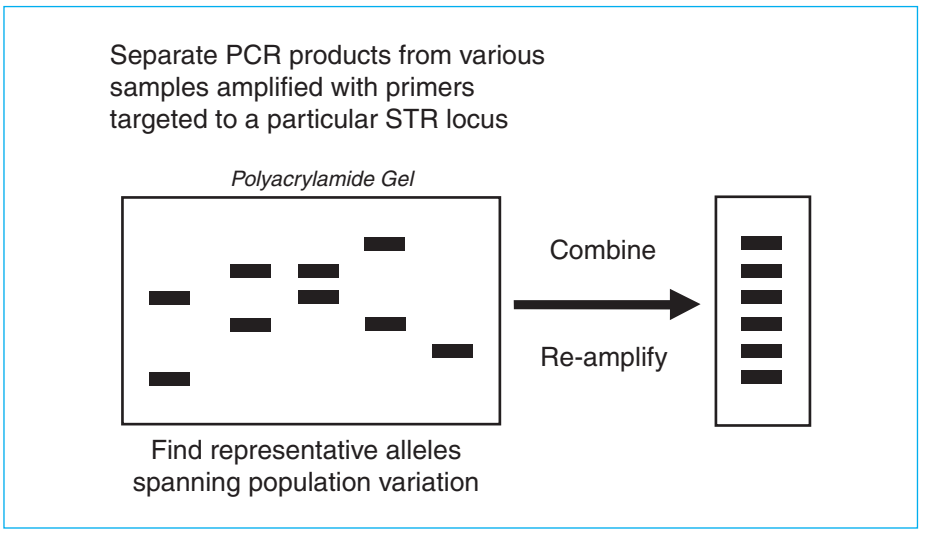


Figure 5.3

Principle of allelic ladder formation. STR alleles from a number of samples are separated on a polyacrylamide gel and compared to one another. Samples representing the common alleles for the locus are combined and re-amplified to generate an allelic ladder. Each allele in the allelic ladder is sequenced since it serves as the reference material for STR genotyping. Allelic ladders are included in commercially available STR kits.

For example, to produce a ladder containing five alleles with 6, 7, 8, 9, and 10 repeats, individual samples with genotypes of (6,8), (7,10), and (9,9) could be combined. Alternatively, the combination of genotypes could be (6,9), (7,8), and (10,10) or (6,6), (7,7), (8,8), (9,9), and (10,10).

Additional quantities of the same allelic ladder (second- and third-generation ladders) may be produced by simply diluting the original ladder 1/1000–1/1 000 000 parts with deionized water and then re-amplifying it using the same PCR primers (Baechtel *et al.* 1993). It is imperative that allelic ladders be generated with the same PCR primers as used to amplify unknown samples so that the allele 'rungs' on the ladder will accurately line up with that of the repeat number of the unknown sample when the unknown is compared to the ladder. As will be seen in the next section, commercial manufacturers now provide allelic ladders in their STR typing kits so that individual laboratories do not have to produce their own allelic ladders.

CHOICE OF MARKERS USED BY THE FORENSIC DNA TYPING COMMUNITY

For DNA typing markers to be effective across a wide number of jurisdictions, a common set of standardized markers must be used. The STR loci that are commonly used today were initially characterized and developed either in the laboratory of Dr. Thomas Caskey at the Baylor College of Medicine (Edwards *et al.* 1991, Hammond *et al.* 1994) or at the Forensic Science Service in England (Kimpton *et al.* 1993, Urquhart *et al.* 1994). The Promega Corporation (Madison, Wisconsin) initially commercialized many of the Caskey markers while Applied Biosystems (Foster City, California) picked up on the Forensic Science Service (FSS) STR loci as well as developing some new markers.

Today both Applied Biosystems and the Promega Corporation have STR kits that address the needs of the DNA typing community and cover a common set of STR loci. The availability of STR kits that permit robust multiplex amplification of eight or more STR markers has truly revolutionized forensic DNA. Matching probabilities that exceed one in a billion are possible in a single amplification with 1 ng (or less) of DNA sample. Just as impressive is the fact that results can be obtained today in only a few hours compared to the weeks that restriction fragment length polymorphism (RFLP) methods took just a few years ago.

One of the first STR multiplexes to be developed was a quadruplex created by the Forensic Science Service that comprised the four loci TH01, FES/FPS, VWA, and F13A1 (Kimpton *et al.* 1994). This so-called 'first-generation multiplex' had a matching probability of approximately 1 in 10 000. The FSS followed with a second-generation multiplex (SGM) made up of six polymorphic STRs and a gender identification marker (Gill *et al.* 1996, Sparkes *et al.* 1996). The six STRs in SGM are TH01, VWA, FGA, D8S1179, D18S51, and D21S11 and provide a matching probability of approximately 1 in 50 million. The gender identification marker amelogenin will be described in more detail later in this chapter.

The first commercial STR kit capable of multiplex amplification became available from the Promega Corporation in 1994 for silver stain analysis. This kit consisted of the STR loci CSF1PO, TPOX, and TH01 and is often referred to as the 'CTT' triplex using the first letter in each locus. The CTT triplex only had a matching probability of ~1 in 500 but was still widely used in the United States in the mid-1990s as it was the first available STR multiplex kit and could be performed with a fairly low start-up cost.

THE 13 CODIS STR LOCI

In the United States, utilization of STRs initially lagged behind that of Europe, especially the efforts of the Forensic Science Service in the United Kingdom. However, beginning in 1996, the FBI Laboratory sponsored a community-wide forensic science effort to establish core STR loci for inclusion within the national DNA database known as CODIS (Combined DNA Index System). Chapter 18 covers CODIS and DNA databases in more detail. This STR Project beginning in April 1996 and concluding in November 1997 involved 22 DNA typing laboratories and the evaluation of 17 candidate STR loci. The evaluated STR loci were CSF1PO, F13A01, F13B, FES/FPS, FGA, LPL, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, and D21S11.

At the STR Project meeting on 13–14 November 1997, 13 core STR loci were chosen to be the basis of the future CODIS national DNA database (Budowle *et al.* 1998). The 13 CODIS core loci are CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, and D21S11. Table 5.1 lists the original references in the literature for these 13 STRs. When all 13 CODIS core loci are tested, the average random match probability is rarer than one in a

trillion among unrelated individuals (Chakraborty *et al.* 1999). The genetics section of this book provides more information on the calculation of random match probability and evaluation of the 13 CODIS STRs in various populations (see Chapter 20).

The three most polymorphic markers are FGA, D18S51, and D21S11, while TPOX shows the least variation between individuals. A summary of information on the 13 STRs is contained in Table 5.2, which describes the chromosomal

Table 5.1

Original reference describing each of the 13 CODIS STR loci and the gender identification marker amelogenin.

Locus Name	Reference
CSF1PO	Hammond, H.A., Jin, L., Zhong, Y., Caskey, C.T. and Chakraborty, R. (1994) Evaluation of 13 short tandem repeat loci for use in personal identification applications. *American Journal of Human Genetics*, 55, 175–189.
FGA	Mills, K.A., Even, D. and Murray, J.C. (1992) Tetranucleotide repeat polymorphism at the human alpha fibrinogen locus (FGA). *Human Molecular Genetics*, 1, 779.
TH01	Polymeropoulos, M.H., Xiao, H., Rath, D.S. and Merril, C.R. (1991) Tetranucleotide repeat polymorphism at the human tyrosine hydroxylase gene (TH). *Nucleic Acids Research*, 19, 3753.
TPOX	Anker, R., Steinbrueck, T. and Donis-Keller, H. (1992) Tetranucleotide repeat polymorphism at the human thyroid peroxidase (hTPO) locus. *Human Molecular Genetics*, 1, 137.
VWA	Kimpton, C.P., Walton, A. and Gill, P. (1992) A further tetranucleotide repeat polymorphism in the vWF gene. *Human Molecular Genetics,* 1, 287.
D3S1358	Li, H., Schmidt, L., Wei, M.-H., Hustad, T., Lerman, M.I., Zbar, B. and Tory, K. (1993) Three tetranucleotide polymorphisms for loci: D3S1352, D3S1358, D3S1359. *Human Molecular Genetics*, 2, 1327.
D5S818	Cooperative Human Linkage Center GATA3F03.512
D7S820	Cooperative Human Linkage Center GATA3F01.511
D8S1179	Cooperative Human Linkage Center GATA7G07.37564
D13S317	Cooperative Human Linkage Center GATA7G10.415
D16S539	Cooperative Human Linkage Center GATA11C06.715
D18S51	Staub, R.E., Speer, M.C., Luo, Y., Rojas, K., Overhauser, J., Otto, L. and Gilliam, T.C. (1993) A microsatellite genetic linkage map of human chromosome 18. *Genomics,* 15, 48–56.
D21S11	Sharma, V. and Litt, M. (1992) Tetranucleotide repeat polymorphism at the D21S11 locus. *Human Molecular Genetics*, 1, 67.
Amelogenin	Sullivan, K.M., Mannucci, A., Kimpton, C.P. and Gill, P. (1993) A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. *BioTechniques*, 15, 637–641.

Cooperative Human Linkage Center information is available via the Internet: http://www.chlc.org

Locus Name	Chromosomal Location	Physical Position [a]	Repeat Motif ISFG Format [b]	GenBank Accession [c]	GenBank Allele	Allele Range [d]	Number of Alleles Seen [e]
CSF1PO	5q33.1 c-fms proto-oncogene, 6th intron	Chr 5 149.484 Mb	TAGA	X14720	12	5–16	20
FGA	4q31.3 alpha fibrinogen, 3rd intron	Chr 4 156.086 Mb	CTTT	M64982	21	12.2–51.2	80
TH01	11p15.5 tyrosine hydroxylase, 1st intron	Chr 11 2.156 Mb	TCAT	D00269	9	3–14	20
TPOX	2p25.3 thyroid peroxidase, 10th intron	Chr 2 1.436 Mb	GAAT	M68651	11	4–16	15
VWA	12p13.31 von Willebrand Factor, 40th intron	Chr 12 19.826 Mb	[TCTG][TCTA]	M25858	18	10–25	28
D3S1358	3p21.31	Chr 3 45.543 Mb	[TCTG][TCTA]	NT_005997	18	8–21	24
D5S818	5q23.2	Chr 5 123.187 Mb	AGAT	G08446	11	7–18	15
D7S820	7q21.11	Chr 7 83.401 Mb	GATA	G08616	12	5–16	30
D8S1179	8q24.13	Chr 8 125.863 Mb	[TCTA][TCTG]	G08710	12	7–20	17
D13S317	13q31.1	Chr 13 80.52 Mb	TATC	G09017	13	5–16	17
D16S539	16q24.1	Chr 16 86.168 Mb	GATA	G07925	11	5–16	19
D18S51	18q21.33	Chr 18 59.098 Mb	AGAA	L18333	13	7–39.2	51
D21S11	21q21.1	Chr 21 19.476 Mb	Complex [TCTA][TCTG]	AP000433	29	12–41.2	82

[a]Physical positions and chromosomal locations determined on July 2003 human genome reference sequence (NCBI build 34) using hgBLAT (http://genome.ucsc.edu).
[b]The DNA Commission of the International Society of Forensic Genetics (ISFG) has published several papers encouraging standardization in STR allele nomenclature (see Bar *et al.* 1994, 1997). STR repeats should be called on the strand sequence originally described in the first public database entry using the first 5′-nucleotides that can define a repeat motif.
[c]GenBank sequence information for a particular STR locus may be accessed at (http://www.ncbi.nlm.nih.gov/GenBank) by entering the accession number shown here. Reference sequences are also available at http://www.cstl.nist.gov/biotech/strbase/seq_ref.htm.
[d]Numbers in this column refer to the number of repeat units present in the alleles. More detail on alleles that have been observed and their PCR products with commercially available STR kits may be found in Appendix I.
[e]See Appendix I.

location, the repeat motif, allele range, and GenBank accession number where the DNA sequence for a reference allele may be found. The chromosomal locations for these STRs have been updated on the recently completed human genome reference sequence. We have included detailed allele sequence information and PCR product sizes with commercially available STR kits in Appendix I.

Using the previously described classification scheme for categorizing STR repeat motifs (Urquhart *et al.* 1994), the 13 CODIS core STR loci may be divided up into four categories:

1. Simple repeats consisting of one repeating sequence: TPOX, CSF1PO, D5S818, D13S317, D16S539;
2. Simple repeats with non-consensus alleles (e.g., 9.3): TH01, D18S51, D7S820;
3. Compound repeats with non-consensus alleles: VWA, FGA, D3S1358, D8S1179;
4. Complex repeats: D21S11.

Table 5.2 (facing) Summary information on the 13 CODIS core STR loci.

COMMERCIALLY AVAILABLE STR KITS

A number of kits are available for single or multiplex PCR amplification of STR markers used in DNA typing. Two primary vendors for STR kits used by the forensic DNA community exist: the Promega Corporation located in Madison, Wisconsin, and Applied Biosystems located in Foster City, California. These companies have expended a great deal of effort over the past decade to bring STR markers to forensic scientists in kit form. More recently in Europe, companies such as Serac (Bad Homburg, Germany) and Biotype (Dresden, Germany) have begun offering commercial STR kits.

The technology has evolved quickly in the late 1990s for more sensitive, rapid, and accurate measurements of STR alleles. At the same time, the number of STRs that can be simultaneously amplified has increased from three or four with silver-stained systems to over 15 STRs using multiple-color fluorescent tags (see Chapter 13). A list of commercially available STR multiplexes and when they were released as products is shown in Table 5.3.

Table 5.3 (below) Information on commercially available STR multiplexes (fluorescently-labeled).

Name	Source	Release Date	STR Loci Included
TH01, TPOX, CSF1PO monoplexes (silver stain)	Promega	Feb 1993	TH01, TPOX, CSF1PO
AmpF/STR® Blue	Applied Biosystems	Oct 1996	D3S1358, VWA, FGA
AmpF/STR® Green I	Applied Biosystems	Jan 1997	Amelogenin, TH01, TPOX, CSF1PO
CTTv	Promega	Jan 1997	CSF1PO, TPOX, TH01, VWA (vWF)

Name	Source	Release Date	STR Loci Included
FFFL	Promega	Jan 1997	F13A1, FES/FPS, F13B, LPL
GammaSTR	Promega	Jan 1997	D16S539, D13S317, D7S820, D5S818
PowerPlex™ (version 1.1 and 1.2 later)	Promega	Jan 1997 Sept 1998	CSF1PO, TPOX, TH01, VWA, D16S539, D13S317, D7S820, D5S818
AmpF/STR® Profiler™	Applied Biosystems	May 1997	D3S1358, VWA, FGA, Amelogenin, TH01, TPOX, CSF1PO, D5S818, D13S317, D7S820
AmpF/STR® Profiler Plus™	Applied Biosystems	Dec 1997	D3S1358, VWA, FGA, Amelogenin, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820
AmpF/STR® COfiler™	Applied Biosystems	May 1998	D3S1358, D16S539, Amelogenin, TH01, TPOX, CSF1PO, D7S820
AmpF/STR® SGM Plus™	Applied Biosystems	Feb 1999	D3S1358, VWA, D16S539, D2S1338, Amelogenin, D8S1179, D21S11, D18S51, D19S433, TH01, FGA
PowerPlex® 2.1 (for Hitachi FMBIO users)	Promega	June 1999	D3S1358, TH01, D21S11, D18S51, VWA, D8S1179, TPOX, FGA, Penta E
PowerPlex® 16	Promega	May 2000	CSF1PO, FGA, TPOX, TH01, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, Penta D, Penta E, amelogenin
PowerPlex® 16 BIO (for Hitachi FMBIO users)	Promega	May 2001	CSF1PO, FGA, TPOX, TH01, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, Penta D, Penta E, amelogenin
AmpF/STR® Identifiler™	Applied Biosystems	July 2001	CSF1PO, FGA, TPOX, TH01, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D2S1338, D19S433, amelogenin
AmpF/STR® Profiler Plus™ ID (extra unlabeled D8-R primer)	Applied Biosystems	Sept 2001	D3S1358, VWA, FGA, Amelogenin, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820
PowerPlex® ES	Promega	Mar 2002	FGA, TH01, VWA, D3S1358, D8S1179, D18S51, D21S11, SE33, amelogenin
AmpF/STR® SEfiler™	Applied Biosystems	Sept 2002	FGA, TH01, VWA, D3S1358, D8S1179, D16S539, D18S51, D21S11, D2S1338, D19S433, SE33, amelogenin

Table 5.3 (Continued)

The adoption of the 13 core loci for CODIS in the United States has led to development of STR multiplexes that cover these markers. At the turn of the century, two PCR reactions were required to obtain information from all the 13 STRs: either PowerPlex® 1.1 and PowerPlex® 2.1 or Profiler Plus™ and COfiler™ (see Table 5.3). As an internal check to reduce the possibility of mixing up samples, both manufacturers included overlapping loci in their kits that should produce concordant data between samples amplified from the same biological material. The Profiler Plus™ and COfiler™ kits have the loci

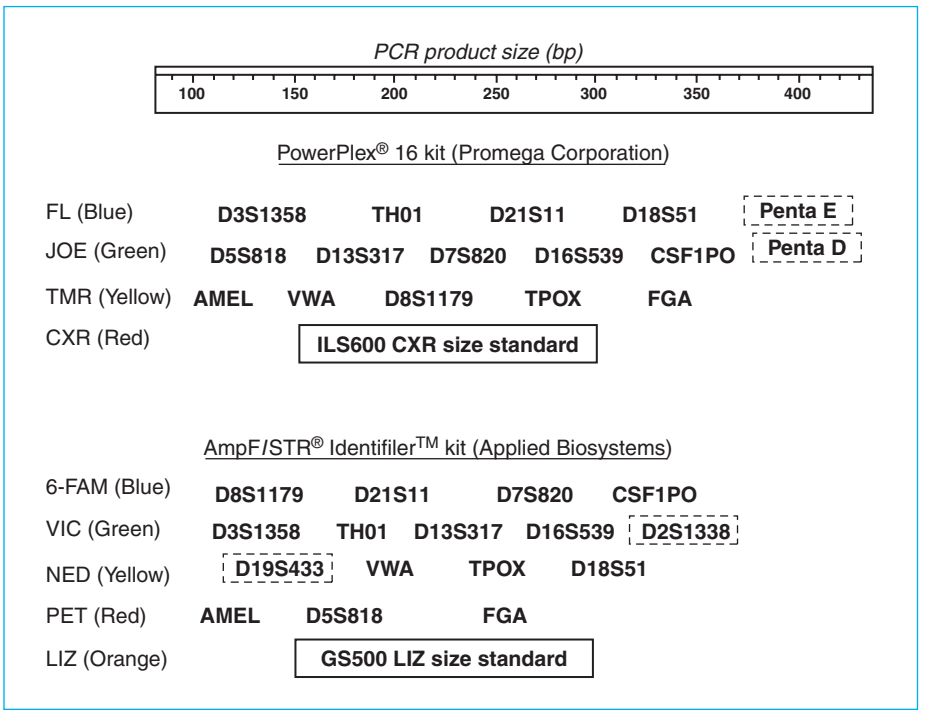


Figure 5.4

Commercially available STR kit solutions for a single amplification of the 13 CODIS core loci. General size ranges and dye-labeling strategies are indicated. The PowerPlex 16 kit uses four dyes while the Identifiler kit uses five dyes. Loci with dotted boxes are additional loci specific to each kit.

D3S1358 and D7S820 in common while the PowerPlex® 1.1 and PowerPlex® 2.1 have the loci TH01, TPOX, and VWA in common.

Since 2000, both Promega and Applied Biosystems have marketed multiplex PCR reactions that permit co-amplification of all 13 STRs in a single reaction along with the amelogenin sex-typing marker and two additional STR loci (Figure 5.4). Electropherograms with size separated PCR products for Promega's PowerPlex® 16 (Figure 5.5) can be viewed as color separated panels of loci or as an overlay of all colors. The allelic ladders for the Applied Biosystems' AmpF*l*STR® Identifiler™ kits are displayed in Figure 5.6.

As will be discussed in the Technology section, two primary methods are used in modern forensic DNA laboratories to separate and detect fluorescently labeled STR alleles. Some PowerPlex kits have been balanced to work with the Hitachi FMBIO II scanner while PowerPlex® 16, Identifiler™, Profiler Plus™, and COfiler™ reactions are typically analyzed on an ABI Prism 310 or 3100 Genetic Analyzer capillary electrophoresis system (see Chapter 14).

Commercial manufacturers of STR kits have spent a great deal of research effort defining which markers would be included in each kit as well as verifying if primer pairs are compatible and will work well in combination with each other during multiplex PCR conditions (Wallin *et al.* 2002, Krenke *et al.* 2002). Promega has published and patented their PCR primer sequences (Masibay *et al.* 2000, Krenke *et al.* 2002) whereas Applied Biosystems have kept their primer

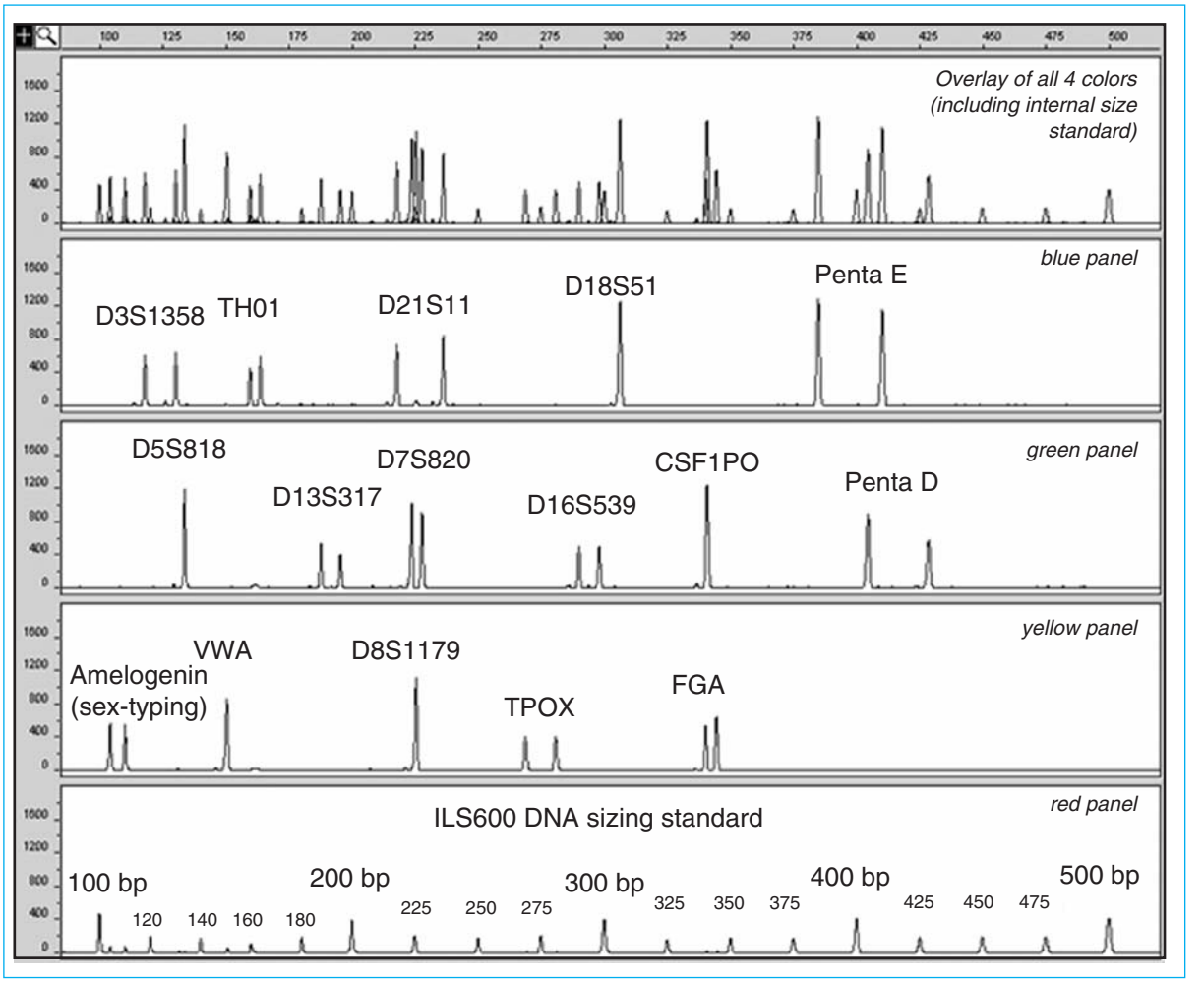


Figure 5.5
PowerPlex® 16 result from 1 ng genomic DNA.

sequences proprietary although some information has been revealed regarding the use of degenerate primers (see Chapter 6). The issue over failure to disclose kit primer sequences impacted several court cases early on in the legal acceptance of STR technology but appears to have been resolved now (D.N.A. Box 5.2).

Most laboratories do not have the time or resources to design primers, optimize PCR multiplexes, and quality control primer synthesis. The convenience of using ready-made kits is also augmented by the fact that widely used primer sets and conditions allow improved opportunities for sharing data between laboratories without fear of possible null alleles (see Chapter 6). Available STR multiplex sets vary based on which STR loci are included, the fluorescent dye combinations, the DNA strand that is labeled, allelic ladders present in kits, and most importantly, the primer sequences utilized for PCR amplification. It is important to keep in mind that commercially available kits quickly dictate which STRs will be used by the vast majority of forensic laboratories.

D.N.A. Box 5.2
Disclosure of STR kit primer sequences

During the early adoption of STR typing technology in U.S. court systems, three cases ruled that DNA results would not be permissible as evidence because the commercial STR kit PCR primer sequences and developmental validation studies were not public information. These cases were People *v.* Bokin (San Francisco, California, May 1999), People *v*. Shreck (Boulder, Colorado, April 2000), and State *v.* Pfenning (Grand Isle, Vermont, Apr 2000).

Shortly after the Pfenning case, the Promega Corporation made the decision to publish their STR kit primer sequences (see news in *Nature* 27 July 2000 issue, volume 406, p. 336) and have done so since (Masibay *et al.* 2001, Krenke *et al.* 2002), along with obtaining several patents in the area of multiplex amplification of STR loci.

Applied Biosystems has repeatedly refused to release the primer sequences present in their STR kits claiming that this information is proprietary. The company is concerned that they would lose revenue if generic brand products were produced by other entities using the revealed primer information. However, in at least 16 cases, the primer sequences for the ProfilerPlus™ and COfiler™ kits have been supplied by Applied Biosystems under a protective court order. Numerous publications since 2000 have demonstrated the reliable use of Applied Biosystems STR kits including detailed validation studies (see Holt *et al.* 2002).

The arguments that not enough information exists to support the reliable use of commercial STR kits whose every component is not public knowledge have fallen by the wayside as millions of DNA profiles have been reliably generated with these kits in the past few years.

For further information:

Masibay, A. *et al.* (2000) Promega Corporation reveals primer sequences in its testing kits [letter]. *Journal of Forensic Sciences*, 45, 1360–1362.

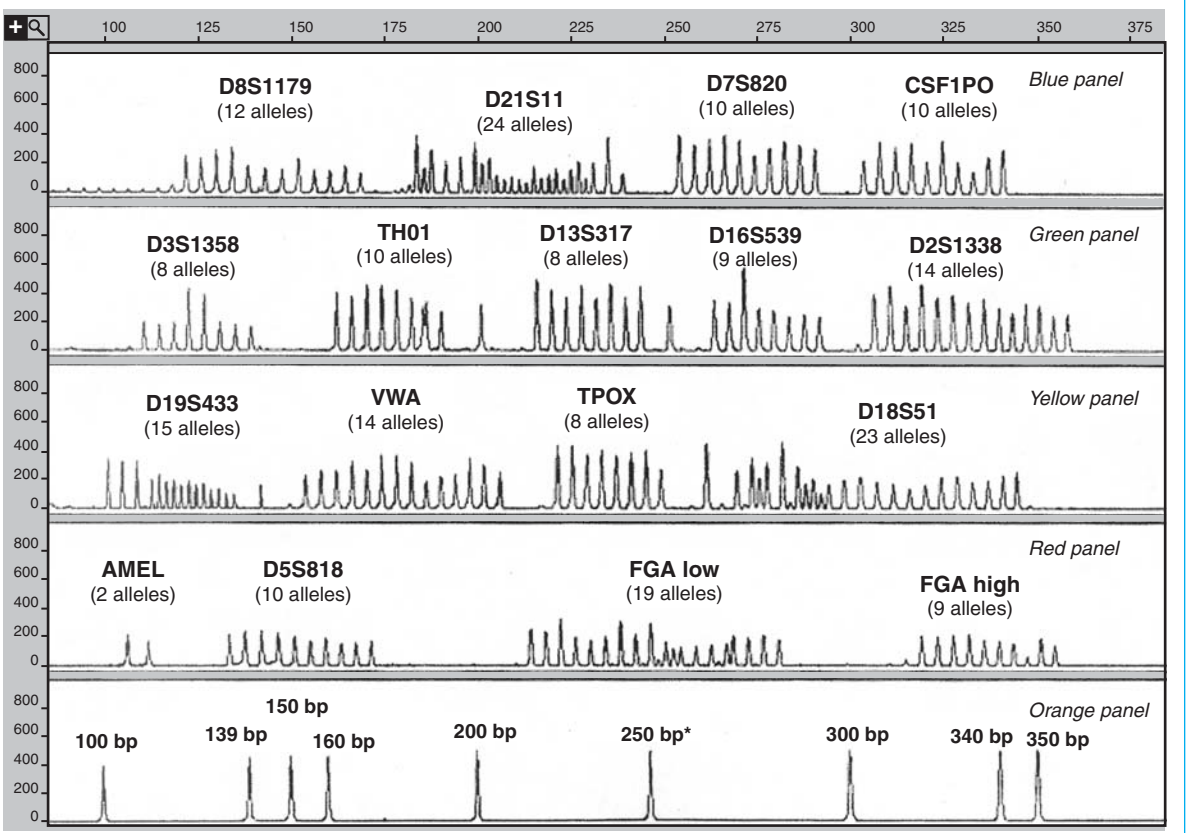
Krenke, BE. *et al.* (2002) Validation of a 16-locus fluorescent multiplex system. *Journal of Forensic Science*, 47, 773–785.

Holt, CL. *et al.* (2002) TWGDAM validation of AmpFlSTR PCR amplification kits for forensic DNA casework. *Journal of Forensic Science*, 47, 66–96.

http://www.scientific.org/archive/archive.html
http://www.denverda.org/html_website/denver_da/DNA_resources.html
http://www.denverda.org/legalResource/AB Sequence case list.pdf

COMMERCIAL ALLELIC LADDERS

Each manufacturer of STR kits provides allelic ladders that may be used for accurate genotyping. It is important to note that kits from the Promega Corporation and Applied Biosystems for comparable STR markers often contain different alleles in their allelic ladders. For example, the PowerPlex® 1.1 kit from Promega contains alleles 7–15 in its D5S818 allelic ladder while the Profiler Plus™ kit from Applied Biosystems contains alleles 7–16 in its D5S818 allelic ladder. By having an allele present in the ladder, a laboratory can be more confident of a call from



LIZ-labeled GS500 DNA sizing standard

Figure 5.6
AmpFlSTR® Identifiler™ allelic ladders (Applied Biosystems). A total of 205 alleles are included in this set of allelic ladders used for genotyping a multiplex PCR reaction involving 15 STR loci and the amelogenin sex-typing test.

an unknown sample that is being analyzed. In the D5S818 example listed here, one would be more confident typing an observed allele 16 when using the Applied Biosystems kit than the Promega kit because the D5S818 allelic ladder has an allele 16 in the ABI kit. The alleles present in the two sources of commercially available multiplex STR kits are reviewed and contrasted in Table 5.4.

Some of the more recent kits come with an amazing number of alleles in their ladders. For example, the Identifiler™ kit from Applied Biosystems contains 205 alleles (Table 5.4; Figure 5.6). Putting together and mass-producing such a large set of alleles is an impressive feat. The Promega PowerPlex® 16 kit has 209 alleles in its allelic ladders.

AMPFLSTR® IDENTIFILER™ KIT INNOVATIONS

Table 5.4 (facing)

Applied Biosystems introduced two new technologies with their AmpFlSTR® Identifiler™ kit when it was released in 2001. The first, and most obvious,

Table 5.4

Comparison of represented alleles in commercially available STR allelic ladders

	Promega Corporation STR Kits								Applied Biosystems AmpFlSTR Kits									
Loci/Kit	*PP1.1 Alleles*	*#*	*PP2.1 Alleles*	*#*	*PP16 Alleles*	*#*	*PP ES Alleles*	*#*	*ProfilerPlus Alleles*	*#*	*COfiler Alleles*	*#*	*SGM Plus Alleles*	*#*	*Identifiler Alleles*	*#*	*SEfiler Alleles*	*#*
CSF1PO	6–15	10			6–15	10					6–15	10			6–15	10		
FGA			17–46.2	19	16–46.2	28	16–46.2	28	17–30	14			17–51.2	28	17–51.2	28	17–51.2	28
TH01	5–11	7	4–13.3	10	4–13.3	10	4–13.3	10			5–9.3,10	7	4–13.3	10	4–13.3	10	4–13.3	10
TPOX	6–13	8	6–13	8	6–13	8					6–13	8			6–13	8		
VWA	10–22	13	10–22	13	10–22	13	10–22	13	11–21	11			11–24	14	11–24	14	11–24	14
D3S1358			12–20	9	12–20	9	12–20	9	12–19	8	12–19	8	12–19	8	12–19	8	12–19	8
D5S818	7–15	9			7–16	10			7–16	10					7–16	10		
D7S820	6–14	9			6–14	9			6–15	10	6–15	10			6–15	10		
D8S1179			7–18	12	7–18	12	7–18	12	8–19	12			8–19	12	8–19	12	8–19	12
D13S317	7–15	9			7–15	9			8–15	8					8–15	8		
D16S539	5,8–15	9			5,8–15	9					5,8–15	9	5,8–15	9	5,8–15	9	5,8–15	9
D18S51			8–27	22	8–27	22	8–27	22	9–26	21			7,9–27	23	7,9–27	23	7,9–27	23
D21S11			24–38	24	24–38	24	24–38	24	24.2–38	22			24–38	24	24–38	24	24–38	24
D2S1338													15–28	14	15–28	14	15–28	14
D19S433													9–17.2	15	9–17.2	15	9–17.2	15
Penta D					2.2–17	14												
Penta E			5–24	20	5–24	20												
SE33							4.2–37	35									4.2–37	35
Amelogenin	X,Y	2			X,Y	2	X,Y	2	X,Y	2	X,Y	2	X,Y	2	X,Y	2	X,Y	2
Total Alleles		76		137		209		155		118		54		159		205		194

involves the use of 5-dye detection systems (see Chapter 13) where four different dyes (6FAM™, VIC™, NED™, and PET™) are used to label the PCR products rather than the traditional three dyes (5FAM, JOE, NED or FL, JOE, TMR) as used with the previous AmpFlSTR or PowerPlex kits. A one dye detection channel is always used for an internal size standard to correlate electrophoretic mobilities to an apparent PCR product size (see Chapter 15). Thus, the fifth dye (LIZ™) in 5-dye detection and the fourth dye (ROX or CXR) in 4-dye detection are used for labeling the internal size standard. The extra dye channel for labeling PCR products enables smaller PCR products to be generated and placed in a separate dye channel rather than extending the size range for amplicons within the three previously available dye channels.

The second technology introduced with the Identifiler™ kit involves mobility modifying non-nucleotide linkers (Applied Biosystems 2001). The mobility modifier is composed of hexaethyleneoxide (HEO) that imparts a shift of approximately 2.5 nucleotides with each additional HEO unit (Grossman *et al.* 1994). This non-nucleotide linker is synthesized into the 5′-end of the PCR primer so that when the PCR product is created it contains these extra molecules on one end (Figure 5.7). By incorporating non-nucleotide linkers, mobilities for amplified alleles from one member of a pair of closely spaced STR loci can be shifted relative to the other. Thus, overlapping size ranges can be prevented (Figure 5.8).

The primary reason for introducing mobility modifiers is to permit continued use of the same PCR primers for amplifying STR loci and still have optimal inter-locus spacing within the various color channels. For example, if the loci D7S820 and CSF1PO, which are labeled with two different fluorophores in the COfiler kit and therefore do not interfere with one another, were labeled with the same colored fluorescent label (e.g., 6FAM) as they are in the Identifiler STR kit, the allelic ladder products would have overlapped by ~13 bp (Figure 5.8). To prevent this overlap in allele size ranges, either PCR primer binding sites must be altered to change the overall size of the PCR product or mobility

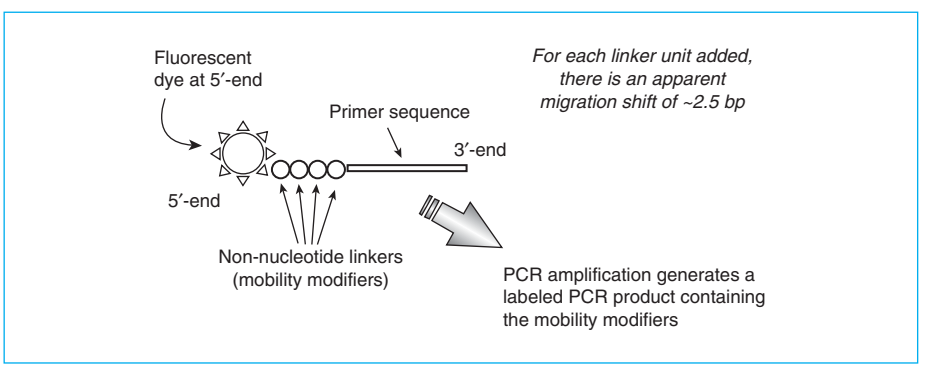


Figure 5.7
Illustration of mobility modifiers used in Applied Biosystems' Identifiler STR kit. Non-nucleotide linkers are synthesized into the primer between the fluorescent dye and 5′-end of the primer sequence. During PCR amplification, the dye and linker are incorporated into the amplicon. With the added non-nucleotide linker, the mobility of the generated STR allele will be shifted to a larger apparent size during electrophoresis. This shift of STR alleles for a particular locus then enables optimal inter-locus spacing for STR loci labeled with the same fluorescent dye without having to alter the PCR primer binding positions (see Figure 5.8).

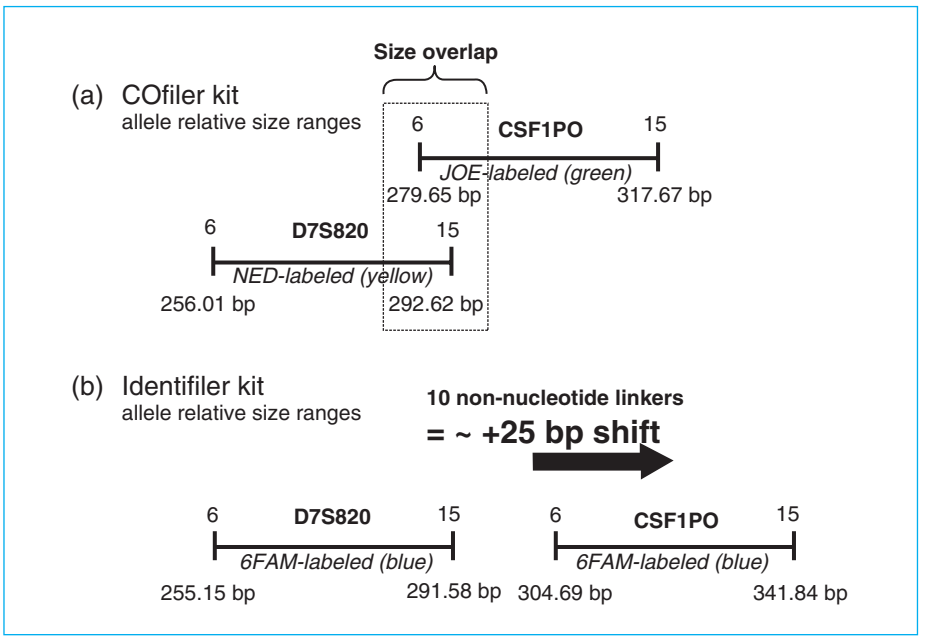


Figure 5.8

Illustration of how non-nucleotide linkers attached to CSF1PO PCR products in the Identifiler STR kit help with inter-locus spacing between D7S820 and CSF1PO. In the COfiler kit (a), CSF1PO and D7S820 are labeled with different colored fluorescent labels and thus do not interfere with one another. However, in the Identifiler kit (b), both D7S820 and CSF1PO are labeled with the same dye and would therefore have overlapping STR alleles unless primer positions were changed or mobility modifiers were used. A ~25 bp shift of the CSF1PO PCR products is accomplished by the addition of 10 non-nucleotide linkers. PCR product sizes for allelic ladder ranges displayed here are from the COfiler and Identifiler kit user's manuals. Note that sizes for D7S820 alleles do not match exactly because different dye labels are used with both the PCR products and the internal size standard thus impacting their relative mobilities.

modifiers can be introduced to shift the apparent molecular weight of the larger PCR product to an even larger size. In the case of the Identifiler™ kit, the locus CSF1PO was shifted by approximately 25 bp – most likely through the addition of 10 HEO non-nucleotide linkers to the 5′-end of the labeled PCR primer. Non-nucleotide linkers are also present on four other loci in the Identifiler™ kit: D2S1338, D13S317, D16S539, and TPOX.

Promega has changed primer sequences for a few of the loci between PowerPlex versions (see Masibay *et al.* 2000, Butler *et al.* 2001, Krenke *et al.* 2002). For example, between the PowerPlex® 1.1 and PowerPlex® 16 kits, the CSF1PO primer positions were drastically altered in order to achieve a 30 bp shift in PCR product size between the two kits (Figure 5.9). This primer change and subsequent PCR product shift was instituted so that CSF1PO and D16S539 loci could be labeled with the same dye in the PowerPlex® 16 kit. Note that if the original CSF1PO primers had been kept, there would have been a 13 bp overlap between D16S539 allele 15 (304 bp) and CSF1PO allele 6 (291 bp) making these systems incompatible in the same dye color without altering the PCR product size (i.e., primer positions) for one of them.

As will be discussed in Chapter 6, different primer positions have the potential to lead to allele dropout if a primer binding site mutation impacts one of the primer pairs. Hence concordance studies are needed between various STR kits to assess the level of potential allele dropout (Budowle *et al.* 2001). On the other hand, Applied Biosystems has maintained the same primers over time

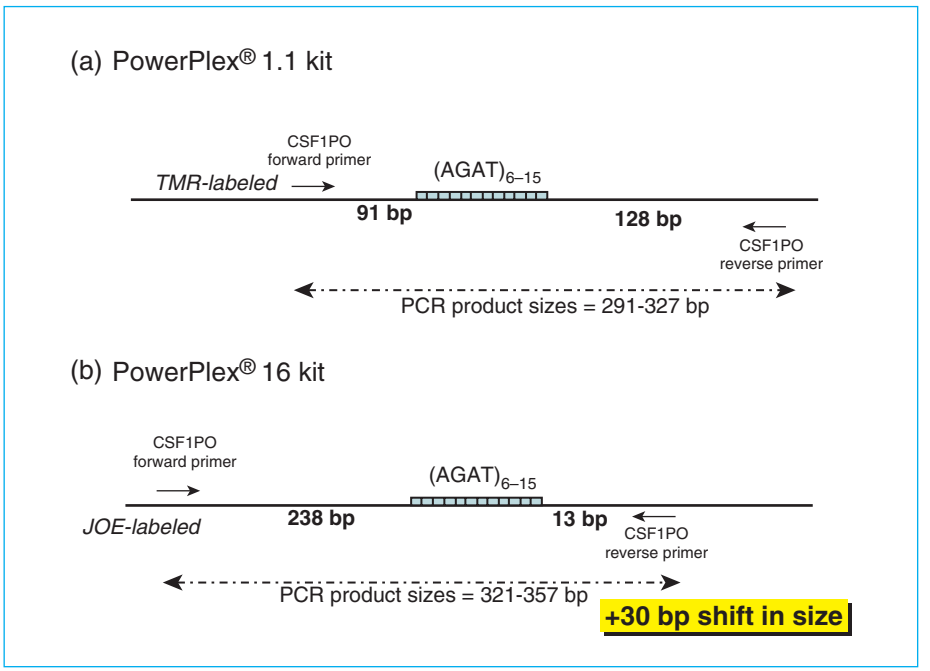


Figure 5.9

Variation in CSF1PO primer positions between (a) PowerPlex 1.1 and (b) PowerPlex 16 STR kits. The base pair (bp) numbers in bold indicate the distance between the repeat region and 3′-end of the pertinent primer. The overall PCR product size for CSF1PO is shifted +30 bp with the primer changes from PowerPlex 1.1 to PowerPlex 16.

and through their various AmpF*l*STR® kits (Holt *et al.* 2002) by introducing 5-dye chemistry and mobility modifiers for products that would normally overlap with one another (see Figure 5.8).

DETAILS ON ALLELES PRESENT IN THE 13 CODIS STR LOCI

Each of the 13 core STR loci 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 13 core STR loci. We have included in Appendix I a detailed summary of the alleles that have been reported as of June 2004 for the 13 core STR loci along with their expected sizes using various kits that are available from Promega or Applied Biosystems. The size difference in the PCR products produced by the different STR kits is important because a large difference is more likely to lead to null alleles when comparing results between two kits (see Chapter 6).

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 6–15 repeats. An allele 16 has also been reported (Margolis-Nunno *et al.* 2001)

as have several *x*.1 and *x*.3 variant alleles (see Appendix I). PCR products from Promega's PowerPlex® 1.1 STR kit are 11 bp larger than those generated with Applied Biosystems kits for equivalent alleles. Since PowerPlex 16 adds 30 bp to the size of CSF1PO relative to PowerPlex® 1.1 (see Figure 5.9), then PowerPlex® 16 PCR products for CSF1PO are 41 bp larger than those generated with AmpF*l*STR® kits.

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. The spread in allele sizes is larger for FGA than any of the other core STR loci. 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 Promega's PowerPlex® 2.1 and 16 STR kits are 112 bp larger than those generated with Applied Biosystems AmpF*l*STR® kits for equivalent alleles. This size difference between these two primer sets is the largest of any of the 13 core loci. So far a total of 80 different FGA alleles have been reported (see Appendix I) making it one of the most polymorphic loci used in human identity testing.

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 an '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. 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 and is designed allele 9.3. Other *x*.3 alleles have been reported such as 8.3, 10.3, and 13.3 (Griffiths *et al.* 1998). TH01 has probably been the most studied of the 13 core loci with over 1000 population studies reported in the literature using this DNA marker. PCR products from Promega's PowerPlex® 1.1 STR kit are 11 bp larger than those generated with Applied Biosystems AmpF*l*STR® kits for equivalent alleles. PowerPlex® 2.1 STR kits produce amplicons that are 19 bp smaller than PowerPlex® 1.1. The PowerPlex® 2.1 and PowerPlex® 16 STR kits contain the same PCR primers for TH01.

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. TPOX has also been referred to in the literature as hTPO. This STR locus

possesses a simple AATG repeat and is the least polymorphic of the 13 core loci. PCR products from Promega's PowerPlex™ 1.1 STR kit are 7 bp larger than those generated with Applied Biosystems AmpFlSTR® kits for equivalent alleles. PowerPlex® 2.1 STR kits produce amplicons that are 38 bp larger in size relative to PowerPlex® 1.1. The PowerPlex® 2.1 and PowerPlex® 16 STR kits contain the same PCR primers for TPOX. Tri-allelic (three banded) patterns are more prevalent in TPOX than any other forensic STR marker (see Chapter 6).

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. The VWA marker targeted by STR multiplex kits is only one of three repeats present in that region of the von Willebrand Factor. The other two have not been found to be as polymorphic (Kimpton *et al.* 1992). PCR products from Promega's PowerPlex® 1.1 STR kit are 29 bp smaller than those generated with Applied Biosystems AmpFlSTR® kits for equivalent alleles. The PowerPlex® 1.1 and PowerPlex® 2.1 STR kits overlap at three STRs including VWA. Both kits produce amplicons that are equivalent in size for VWA alleles. The PowerPlex® 2.1 and PowerPlex® 16 STR kits contain the same PCR primers for VWA.

D3S1358 is a compound tetranucleotide repeat found on the short arm of chromosome 3. This locus possesses both AGAT and AGAC repeat units (Mornhinweg *et al.* 1998). The D3 marker is common to Applied Biosystems AmpFlSTR® multiplexes Blue™, Profiler™, Profiler Plus™, COfiler™, SGM Plus™, SEfiler™, and Identifiler™. PCR products from Promega's PowerPlex® 2.1 STR kit are 2 bp larger than those generated with Applied Biosystems kits for equivalent alleles. The PowerPlex® 2.1 and PowerPlex® 16 STR kits contain the same PCR primers for D3S1358.

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 7–16 repeats. In both Promega and Applied Biosystems STR kits, D5S818 is one of the smaller sized loci and as such should appear more than some of the other loci in degraded DNA samples. Only a few rare microvariants have been reported at this STR marker. PCR products from Promega's PowerPlex® 1.1 STR kit are 15 bp smaller than those generated with Applied Biosystems kits for equivalent alleles and PowerPlex® 16 retains the original PowerPlex® 1.1 primers.

D7S820 is a simple tetranucleotide repeat found on the long arm of chromosome 7. The locus possesses primarily a GATA repeat. However, a number of new D7 microvariant alleles have been reported recently (see Appendix I). These *x*.1 and *x*.3 alleles likely result due to 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 (Egyed *et al.* 2000). PCR products from Promega's PowerPlex® 1.1 STR kit are 42 bp smaller than those generated with Applied Biosystems 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 (Oldroyd *et al.* 1995, Barber and Parkin 1996). 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 (Barber and Parkin 1996). PCR products from Promega's PowerPlex® 2.1 and PowerPlex® 16 STR kits are 80 bp larger than those generated with Applied Biosystems 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 Asian populations due to a mutation in the middle of the primer-binding site (Leibelt *et al.* 2003).

D13S317 is a simple TATC tetranucleotide repeat found on the long arm of chromosome 13. Common alleles contain between 7–15 repeat units although alleles 5, 6, and 16 have been reported (see Appendix I). PCR products from Promega's PowerPlex® 1.1 STR kit are 36 bp smaller than those generated with Applied Biosystems 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 calls with different primer sets (Butler *et al.* 2003, Drábek *et al.* 2004). PowerPlex® 16 primers, while generating the same size amplicons as the original PowerPlex® 1.1 primers, have been shifted to avoid this 4 bp deletion that is present in some African-American samples.

D16S539 is a simple tetranucleotide repeat found on the long arm of chromosome 16. Nine common alleles exist that possess a core repeat unit of GATA. These include an allele with five repeats and consecutive alleles ranging from 8–15 repeat units in length. PCR products from Promega STR kits are 31 bp larger than those generated with Applied Biosystems kits for equivalent alleles. 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 an extra or 'degenerate' unlabeled primer in their COfiler™, SGM Plus™, and Identifiler™ kits so that both possible alleles could be amplified (Wallin *et al.* 2002). On the other hand, Promega altered their D16S539 reverse primer sequence between kits but kept the overall amplicon size the same (Butler *et al.* 2001). The 3′-end of the PowerPlex® 1.1 reverse primer was lengthened by five nucleotides to create the PowerPlex® 16 reverse primer and thus move the primer mismatch caused by this mutation further into the primer to prevent allele dropout (Nelson *et al.* 2001, Krenke *et al.* 2002).

D18S51 is a simple tetranucleotide repeat found on the long arm of chromosome 18. It has a repeat motif of AGAA. A number of *x*.2 allele variants exist due to a 2 bp deletion from a loss of AG in the 3′-flanking region (Barber and Parkin 1996). More than 50 alleles have been reported for D18S51 making it one of the more polymorphic of the 13 core loci. PCR products from Promega's PowerPlex® 2.1 STR kit are 22 bp larger than those generated with Applied Biosystems AmpF*l*STR® kits for equivalent alleles. The PowerPlex® 2.1 and PowerPlex® 16 STR kits contain the same PCR primers for D18S51.

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}. The *x*.2 microvariant alleles arise primarily from a 2 bp (TA) insertion on the 3′-end of the repeat region (Brinkmann *et al.* 1996). PCR products from Promega's PowerPlex® 2.1 STR kit are 17 bp larger than those generated with Applied Biosystems AmpF*l*STR® kits for equivalent alleles. The PowerPlex® 2.1 and PowerPlex® 16 STR kits contain the same PCR primers for D21S11.

Early papers in the literature by the Forensic Science Service had alleles named based on the dinucleotide subunit CV, where the V represents either an A, T, or G (Urquhart *et al.* 1994), while other authors adopted a different allele naming scheme based on the primary tetranucleotide repeat (Moller *et al.* 1994). As outlined in the European DNA Profiling Group inter-laboratory study on D21S11 (Gill *et al.* 1997), a simple formula can be used to convert the Urquhart (U) designation into the Moller (M) equivalent:

$$M = \tfrac{1}{2} \times (U - 5) \tag{5.1}$$

Today most laboratories use the Moller allele notation since it fits the ISFG allele designation recommendation (Bar *et al.* 1997).

D21S11 is far more polymorphic than can be easily detected with sized-based length separations. A careful search of the literature has revealed more than 80 reported alleles, many of which are the same length (see Appendix I). Fine differences in the D21S11 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 (Appendix I).

The three most polymorphic of the 13 loci are D21S11, FGA, and D18S51. These loci contain numerous microvariant alleles that are being uncovered as more and more samples are examined around the world.

ADDITIONAL STR LOCI COMMONLY USED

The 13 core loci used within the United States for CODIS are effective DNA markers for human identification and will most likely continue to be used for some time. However, these 13 markers are by no means the only STRs that have been evaluated or used by forensic labs around the world. Dozens of other markers have been used, some quite extensively (Table 5.5).

Table 5.5 (below) Some additional STR markers used in the forensic DNA community. STR markers in bold are part of commonly used multiplex kits. For information on Y chromosome STRs, see Tables 9.2 and 9.5.

Locus Name	Chromosomal Location	GenBank Accession	Repeat ISFG format	Allele Range	Amplicon Size Range	Reference
ARA	Xcen–q13	M21748	CAG	14–32	255–315 bp	Hammond *et al*. (1994)
APOAI1	11q23–qter	J00048	AAAG	Complex	263–291 bp	Dupuy and Olaisen (1997)
ACTBP2	6	V00481	AAAG	4.2–37	198–325 bp	Dupuy and Olaisen (1997)
CD4	12p12–pter	M86525	TTTTC	6–16	125–175 bp	Hammond *et al*. (1994)
CYAR04	15q21.1	M30798	AAAT	5–12	173–201 bp	Hammond *et al*. (1994)
F13A01	6p24.3–25.1	M21986	GAAA	3.2–16	281–331 bp	Hammond *et al*. (1994)
F13B	1q31–q32.1	M64554	TTTA	6–12	169–193 bp	Promega
FABP	4q28–31	M18079	ATT	10–15	199–220 bp	Hammond *et al*. (1994)
FES/FPS	15q25–qter	X06292	ATTT	7–14	222–250 bp	Hammond *et al*. (1994)
HPRTB	Xq26.1	M26434	TCTA	6–17	259–303 bp	Hammond *et al*. (1994)
LPL	8p22	D83550	TTTA	7–14	105–133 bp	Promega
Penta D	21q	AP001752	AAAGA	2.2–17	376–449 bp	**PowerPlex 16**
Penta E	15q	AC027004	AAAGA	5–24	379–474 bp	**PowerPlex 16**
PLA2A1	12q23–qter	M22970	AAT		118–139 bp	Hammond *et al*. (1994)
RENA4	1q32	M10151	ACAG		255–275 bp	Hammond *et al*. (1994)
D1S1656	1pter–qter	G07820	(TAGA) (TAGG)	9–19.3	125–168 bp	Wiegand *et al*. (1999)
D2S1242	2pter–qter	L17825	(GAAA) (GAAG)	10–18	141–175 bp	Reichenpfader *et al.* (1999)
D2S1338	2q35–37.1	G08202	(TGCC) (TTCC)	15–28	289–341 bp	**SGM Plus, Identifiler**
D3S1359	3p	AA306290	TCTA	11–25.3	196–255 bp	Poltl *et al*. (1998)
D3S1744	3q24	G08246	GATA	14–22	150–182 bp	Lifecodes

Locus Name	Chromosomal Location	GenBank Accession	Repeat ISFG format	Allele Range	Amplicon Size Range	Reference
D6S477	6pter–qter	G08543	TCTA	13.2–22	206–240 bp	Carracedo and Lareu (1998)
D7S809	7pter–qter	X73290	(AGGA) (AGGC)	9 alleles	241–289 bp	Tamaki *et al*. (1996)
D8S347	8q22.3–24.3	L12268	AGAT	16–28	340–388 bp	Poltl *et al*. (1997)
D8S639	8p21–p11	L24797	(AGAT) (AGGT)	20–33.3	316–371 bp	Seidl *et al*. (1999)
D9S302	9q31–33	G08746	ATCT	17 alleles	255–353 bp	Carracedo and Lareu (1998)
D10S2325	10pter–qter	G08790	TCTTA	6–17	113–168 bp	Wiegand *et al*. (1999)
D11S488	11q24.1–25	L04732	(AAAG) (GAAG)	26–41	242–302 bp	Seidl *et al*. (1999)
D11S554	11p11.2–12	M87277	AAAG	Complex	176–286 bp	Dupuy and Olaisen (1997)
D12S391	12	G08921	(AGAT) (AGAC)	15–26	209–253 bp	Lareu *et al*. (1996)
D12S1090	12q12	Not found	GATA	9–33	212–306 bp	Lifecodes
D18S535	18pter–qter	G07985	GATA	9–16	130–158 bp	Wiegand *et al*. (1999)
D18S849	18q12–q21	G07992	GATA	9–20	93–133 bp	Lifecodes
D19S433	19q12–13.1	G08036	AAGG	9–17.2	106–140 bp	**SGM Plus, Identifiler**
D20S161	20pter–qter	L16405	TAGA	14–22	156–187 bp	Hou *et al*. (1999)
D22S683	22pter–qter	G08086	(TA) (TATC)	12–21.2	168–206 bp	Carracedo and Lareu (1998)
DXS6807	Xpter–p22.2	G09662	GATA	11–17	251–275 bp	Edelmann and Szibor (1999)

Table 5.5 (Continued)

Applied Biosystems has created the AmpF*l*STR® SGM Plus™ kit that co-amplifies 10 STR loci including two new STRs: D19S433 and D2S1338. With the adoption of the SGM Plus kit by the Forensic Science Service and much of Europe, the amount of population data on the STR loci D19S433 and D2S1338 will continue to grow. These two loci are also part of the Identifiler 16plex STR kit. Likewise, the Promega Corporation has included two pentanucleotide STR loci, Penta E and Penta D, in their GenePrint® PowerPlex® 2.1 and PowerPlex® 16 kits. Because these markers are included in the STR multiplexes in conjunction with the 13 core loci for developing DNA databases, they will become more prevalent as the number of samples in the databases grows.

Owing to the fact that the German national DNA database requires analysis of the complex hypervariable STR locus SE33, Promega created the PowerPlex®

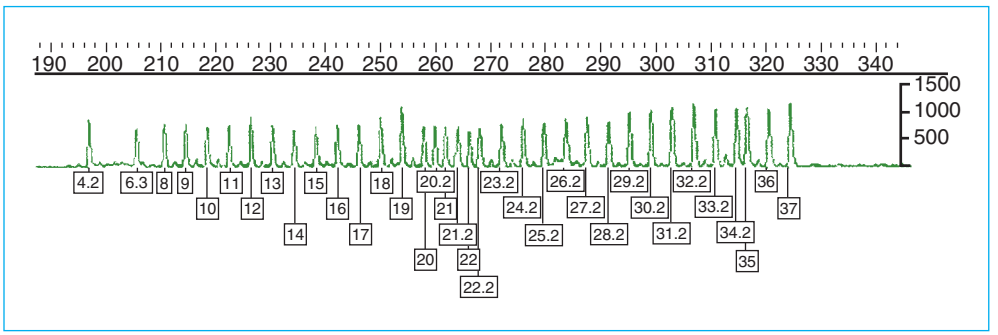


Figure 5.10
SE33 (ACTBP2) allelic ladder from PowerPlex ES kit produced by the Promega Corporation.

ES kit in 2001 and Applied Biosystems released the SEfiler™ kit in late 2002 to provide SE33 in a commercial kit form. The PowerPlex® ES allelic ladder for SE33 contains 35 alleles, which demonstrates this STR marker's variability (Figure 5.10).

Promega also has a multiplex commonly referred to as FFFL, which is used by many laboratories in South America to amplify the four STRs F13A01, F13B, FES/FPS, and LPL. Table 5.5 includes a listing of these markers as well as many others that have appeared in the literature along with useful information such as the GenBank accession number, references, and size ranges with a reported set of PCR primers. STR markers on the Y chromosome are described in Chapter 9. These Y-STRs are becoming increasingly popular due to their ability to aid sexual assault investigations through male-specific amplification.

GENDER IDENTIFICATION WITH AMELOGENIN

The ability to designate whether a sample originated from a male or a female source is useful in sexual assault cases, where distinguishing between the victim and the perpetrator's evidence is important. Likewise, missing persons and mass disaster investigations can benefit from gender identification of the remains. Over the years a number of gender identification assays have been demonstrated using PCR methods (Sullivan *et al.* 1993, Eng *et al.* 1994, Reynolds and Varlaro 1996). By far the most popular method for sex-typing today is the amelogenin system as it can be performed in conjunction with STR analysis.

Amelogenin is a gene that codes for proteins found in tooth enamel. The British Forensic Science Service was the first to describe the particular PCR primer sets that are used so prevalently in forensic DNA laboratories today (Sullivan *et al.* 1993). These primers flank a 6 bp deletion within intron 1 of the amelogenin gene on the X homologue (Figure 5.11). PCR amplification of this area with their primers results in 106 bp and 112 bp amplicons from the X and

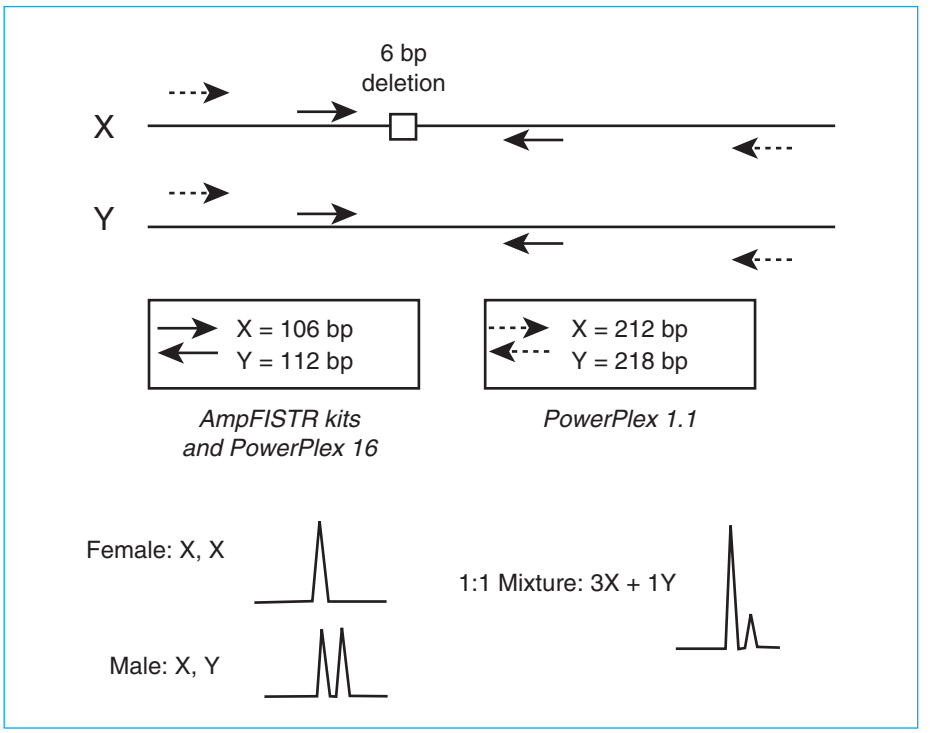


Figure 5.11

Schematic of the amelogenin sex-typing assay. The X and Y chromosomes contain a high degree of sequence homology at the amelogenin locus. The primer sets depicted here target a 6 bp deletion that is present only on the X chromosome. The presence of a single peak indicates that the sample comes from a female while two peaks identifies the sample's source as male. The primers to amplify the 106/112 bp fragments are used in the AmpFlSTR kits while the PowerPlex 1.1 kit uses the larger primer set.

Y chromosomes, respectively. Primers, which yield a 212 bp X-specific amplicon and a 218 bp Y-specific product by bracketing the same 6 bp deletion, were also described in the original amelogenin paper (Sullivan *et al.* 1993) and have been used in conjunction with the D1S80 VNTR system (Budowle *et al.* 1996).

An advantage with the above approach, i.e., using a single primer set to amplify both chromosomes, is that the X chromosome product itself plays a role as a positive control. This PCR-based assay is extremely sensitive. Mannucci and co-workers were able to detect as little as 20 pg (~3 diploid copies) as well as sample mixtures where female DNA was in 100-fold excess of male DNA (Mannucci *et al.* 1994).

Other regions of the amelogenin gene have size differences between the X and Y homologues and may be exploited for sex-typing purposes. For example, Eng and co-workers (1994) used a single set of primers that generated a 977 bp product for the X chromosome and a 788 bp fragment for the Y chromosome. In this case, a 189 bp deletion in the Y relative to the X chromosome was used to differentiate the two chromosomes.

A careful study found that 19 regions of absolute homology, ranging in size from 22–80 bp, exist between the human amelogenin X and Y genes that can be used to design a variety of primer sets (Haas-Rochholz and Weiler 1997). Thus, by spanning various deletions of the X and/or Y chromosome, it is possible to

generate PCR products from the X and Y homologues that differ in size and contain size ranges that can be integrated into future multiplex STR amplifications.

ANOMALOUS AMELOGENIN RESULTS

While amelogenin is an effective method for sex-typing biological samples in most cases, the results are not full proof. A rare deletion of the amelogenin gene on the Y chromosome can cause the Y chromosome amplicon to be absent (Santos *et al.* 1998). In such a case, a male sample would falsely appear as a female. It appears that this deletion of the Y chromosome amelogenin region is more common in Indian populations (Thangaraj *et al.* 2002) than those of European or African origins. A study of almost 30 000 males in the Austrian National DNA database revealed that only six individuals lacked the amelogenin Y-amplicon (Steinlechner *et al.* 2002). These individuals were verified to be male with Y-STRs and amplification of the SRY region (see Chapter 9).

Amelogenin X allele dropout has also been observed in males. In this case only the amelogenin Y-amplicon is present (Shewale *et al.* 2000). This phenomenon was observed only three times out of almost 7000 males examined and likely results from a rare polymorphism in the primer binding sites for the amelogenin primers used in commercial STR kits. A different set of amelogenin primers targeting the same 6 bp deletion on the X chromosome amplified both the X and Y alleles of amelogenin (Shewale *et al.* 2000).

STRBASE: A DYNAMIC SOURCE OF INFORMATION ON STR MARKERS

The rapid growth of the human identification applications for STR loci insures that static written materials, such as this book, will quickly become out-of-date. New alleles are constantly being discovered (including 'off-ladder' microvariant alleles), additional STR markers are being developed, and population data increases with each month of published journals. Indeed, a growing list of publications describing the application of STR loci to forensic DNA typing has exceeded 2000 references.

The growth of the World Wide Web now permits dynamic sources of information to be widely available. Several years ago a web site was created to enable forensic scientists to keep abreast with the rapidly evolving field of DNA typing. In anticipation of the impact of STR markers on DNA typing and the need for a common source of information that could evolve as the process improved, an internet-accessible informational database was created in early 1997. STRBase was officially launched in July 1997 and is maintained by the DNA Technologies Group of the National Institute of Standards and Technology (Butler *et al.* 1997, Ruitberg *et al.* 2001). STRBase may be reached via the World Wide Web using

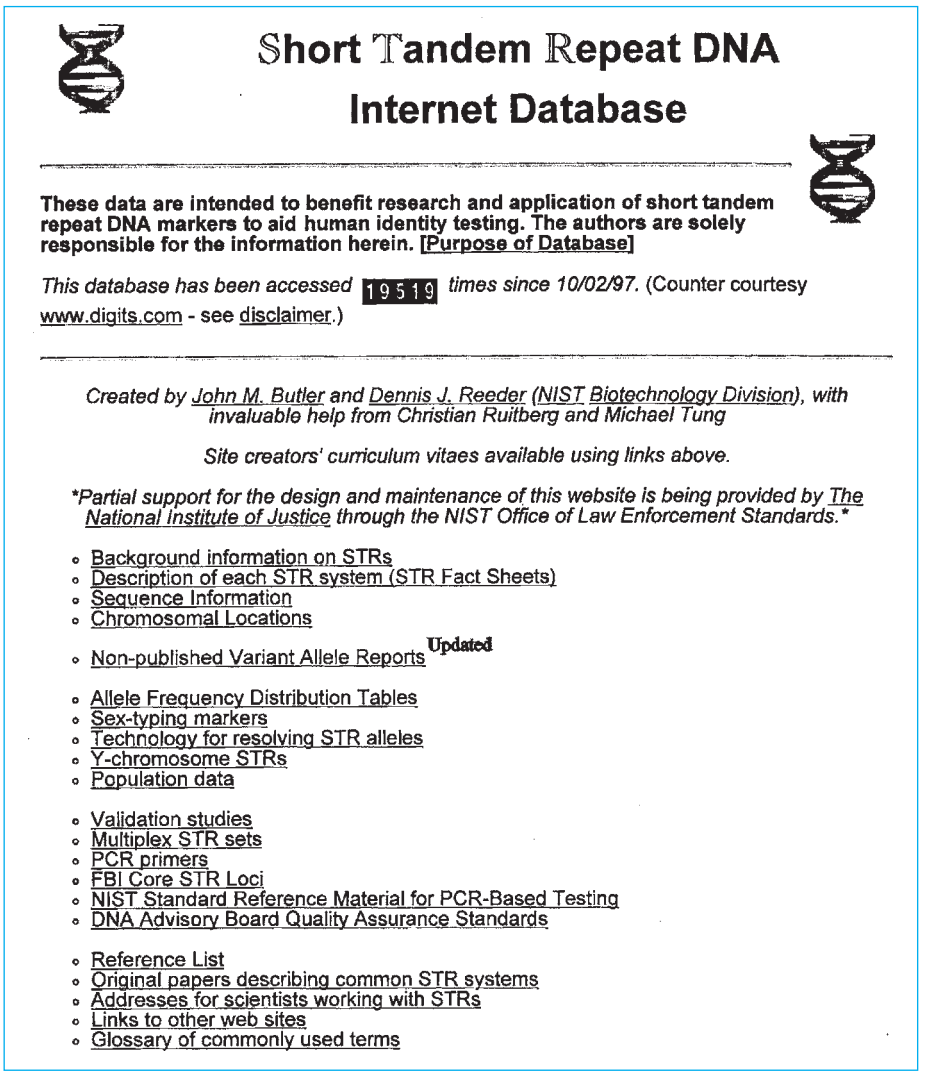


Figure 5.12
Homepage for STRBase, an internet-accessible database of information on STR markers used in forensic DNA typing. STRBase may be accessed via the URL: http://www.cstl.nist.gov/biotech/strbase/ and contains among other things a comprehensive listing of all papers relating to STR typing for human identity testing purposes now numbering over 2000 references.

the following URL: http://www.cstl.nist.gov/biotech/strbase. The home page for STRBase is shown in Figure 5.12.

STRBase contains a number of useful elements. Continually updated information includes the listing of references related to STRs and DNA typing (over 2000 references), addresses for scientists working in the field, and new microvariant or 'off-ladder' STR alleles. Other information that is updated less frequently includes STR fact sheets (with allele information similar to Appendix I), links to other web pages, a review of technology used for DNA typing as well as published primer sequence information, and population data for STR markers.

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 of the 13 CODIS core loci for the U.S. national

DNA database ensures that these STR markers will be used for many years to come. However, as we will see in the next two chapters, results from STR markers require careful interpretation in order to be effective tools for law enforcement.

REFERENCES AND ADDITIONAL READING

Applied Biosystems (2001) AmpF/STR® Identifiler™ PCR Amplification Kit User's Manual. Foster City, CA.

Applied Biosystems (2002) AmpF/STR® SEfiler™ PCR Amplification Kit User's Manual. Foster City, CA.

Baechtel, F.S., Smerick, J.B., Presley, K.W. and Budowle, B. (1993) *Journal of Forensic Sciences*, 38, 1176–1182.

Bar, W., Brinkmann, B., Lincoln, P., Mayr, W.R. and Rossi, U. (1994) *International Journal of Legal Medicine*, 107, 159–160.

Bar, W., Brinkmann, B., Budowle, B., Carracedo, A., Gill, P., Lincoln, P., Mayr, W.R. and Olaisen, B. (1997) *International Journal of Legal Medicine*, 110, 175–176.

Barber, M.D., McKeown, B.J. and Parkin, B.H. (1996) *International Journal of Legal Medicine*, 108, 180–185.

Barber, M.D. and Parkin, B.H. (1996) *International Journal of Legal Medicine*, 109, 62–65.

Bacher, J.W., Hennes, L.F., Gu, T., Tereba, A., Micka, K.A., Sprecher, C.J., Lins, A.M., Amiott, E.A., Rabbach, D.R., Taylor, J.A., Helms, C., Donis-Keller, H. and Schumm, J.W. (1999) *Proceedings of the Ninth International Symposium on Human Identification,* pp. 24–37. Madison, Wisconsin: Promega Corporation.

Brinkmann, B., Meyer, E. and Junge, A. (1996) *Human Genetics*, 98, 60–64.

Britten, R.J. and Kohne, D.E. (1968) *Science*, 161, 529–540.

Broman, K.W., Murray, J.C., Sheffield, V.C., White, R.L. and Weber, J.L. (1998) *American Journal of Human Genetics*, 63, 861–869.

Budowle, B., Moretti, T.R., Niezgoda, S.J. and Brown, B.L. (1998) *Proceedings of the Second European Symposium on Human Identification*, pp. 73–88. Madison, Wisconsin: Promega Corporation.

Budowle, B., Masibay, A., Anderson, S.J., Barna, C., Biega, L., Brenneke, S., Brown, B.L., Cramer, J., DeGroot, G.A., Douglas, D., Duceman, B., Eastman, A., Giles, R., Hamill, J., Haase, D.J., Janssen, D.W., Kupferschmid, T.D., Lawton, T., Lemire, C., Llewellyn, B., Moretti, T., Neves, J., Palaski, C., Schueler, S., Sgueglia, J., Sprecher, C., Tomsey, C. and Yet, D. (2001) *Forensic Science International*, 124, 47–54.