GenePrint™ STR* Multiplexes: Reliability, Flexibility and Throughput in Database and Casework-Compatible STR Analysis

By James W. Schumm, Ph.D.

™ The era of databasing criminal populations to link suspects to crime scenes, and crime scenes to one another, through comparison of biological evidence is upon us. To achieve the ultimate benefit of this approach, legislation in the United Kingdom, Canada, the United States and many other countries has created centralized databases which will eventually include DNA profiles of hundreds of thousands to millions of individuals, primarily convicted criminals. To this end, all countries are moving to the use of short tandem repeat (STR) polymorphisms, which provide a rapid, reliable and inexpensive method of analyzing and unambiguously digitizing data for large numbers of samples. This article describes Promega's effort to develop and improve methods to fit the needs of the digital information age while maintaining a flexible approach which allows large and small laboratories worldwide to apply the same genetic systems.

The amplification products of these multiplex systems may be detected using either the Hitachi FMBIO® or FMBIO® II Fluorescent Scanner, the ABI Model 373 or ABI PRISM™ 377 DNA Sequencer or the ABI Model 310 Capillary Electrophoresis Unit.

DNA markers which distinguish individuals from one another have been known since 1980. With the development of large numbers of VNTR (variable number tandem repeat) loci, a few of the most polymorphic markers became popular in forensic and paternity analyses. Despite the enormous discrimination power of VNTR systems, their use in databases is limited by their imprecision. Windows of standard deviation or the creation of bins is used to contend with this restriction. VNTRs are also limited in casework application because various markers require from 5 to 250ng of sample to assure reliable and reproducible success.

STR SELECTION CRITERIA

The use of properly selected STR loci can overcome the limitations of VNTR systems, producing digital results with 1ng or less of sample material. In our work, we have focused on STR selection criteria which provide highly discriminating markers without sacrificing quality of interpretation and digital storage. First, only STRs which demonstrate a high degree of variability within the population are selected. Second, the amplified products must be easily distinguished from one another. This means rejecting markers which contain frequent microvariants (i.e., alleles differing from one another by lengths shorter than the repeat length) as the closer and more random spacing of alleles is more difficult to interpret. Also, the prevalence of stutter bands (i.e., amplification artifacts which appear one or more repeat lengths above or below the true amplified allele) has led to the rejection of dinucleotide repeats as a class for these applications. In our work, only tetranucleotide repeats which display limited or no stutter have been selected.

The ability to amplify and detect very small amounts of DNA template (typically 1ng) is essential for forensic applications. In paternity analyses, the mutation rate of the markers chosen must be extremely low to avoid false exclusion of suspected fathers. In all cases, the reliability and reproducibility of the data must be irrefutable. Identification of the best markers for these applications is complicated by the fact that their desired traits are not fully compatible with one another. While it is possible to identify highly polymorphic markers with relatively low presence of stutter bands, there appears to be a correlation between a high degree of polymorphism, a tendency for microvariants and increased mutation frequency.

We have focused our attention on the development of tetranucleotide repeat loci which display few or no microvariants, minimal stutter bands, and have a relatively low mutation rate. The polymorphisms in our currently developed STR systems result almost exclusively from variation in the number of tetranucleotide repeats present at the locus, and not from insertion or deletion of one or two bases. This allows rapid and precise typing of easily amplified alleles ranging from 100 to 350 bases in length.

ALLELIC LADDERS

The discrete nature of alleles of these STR loci has also allowed the development of allelic ladders. Allelic ladders are composed of a collection of most or all of the amplified alleles found in the general population. These composites make ideal size markers because, in all STR loci we have developed to date except for vWA, the size markers and the amplified unknown alleles will contain not only the same size fragments, but the same sequence fragments. Thus, ladder components and unknowns co-migrate in gel electrophoresis regardless of the gel matrix or running buffer selected. In the same fashion, the addition of a fluorescein tag, often used as a reporter molecule in STR analysis, alters the ladder components and the amplified unknowns in an identical fashion. Thus, different laboratories using different separation techniques and different detection formats can compare their results with precision and reliability.

ADVANTAGES OF USING STR SYSTEMS

The fact that STR analysis is based on the PCR process offers several additional advantages. First, the detection of small amounts of template (e.g., 1ng of DNA) has become routine. Second, there are a number of rapid purification methods which are compatible with PCR but which do not provide enough DNA of appropriate quality for use in Southern blot-based formats. Third, STR analysis is much faster than Southern blot-based protocols. Using STR analysis, results are obtained in 1-2 days while Southern-based typing methods may take 5-7 days to complete.

STR MULTIPLEXES

The major disadvantage in using STR systems is that those which display few or no microvariants and low mutation rates are not as polymorphic as the best of the VNTR markers. Thus, with STRs, there is a need to develop high throughput approaches to overcome this deficiency. The selection of individual STR loci, each with a limited size range of known alleles, means that several STR systems may be detected simultaneously in limited and well defined regions of the same lane on a gel. This offers the potential to develop STR multiplex systems. The multiplex systems we have developed for analysis using either semi-automated fluorescent detection methods or manual silver stain detection methods are listed in Table 1.

FLUORESCENT DETECTION

The PowerPlex™ System allows co-amplification of eight STR loci in a single reaction vessel, and the FFFL System allows co-amplification of four additional loci in a single reaction. Using the PowerPlex™ System in combination with the FFFL System allows the amplification of twelve STR loci in two multiplex reactions. The amplification products of these multiplex systems may be detected using either the Hitachi FMBIO® or FMBIO® II Fluorescent Scanner, the ABI Model 373 or ABI PRISM™ DNA Sequencer or the ABI Model 310 Capillary Electrophoresis Unit.

Figure 1 shows results obtained using the PowerPlex™ System and the Hitachi FMBIO® Fluorescent Scanner. In the PowerPlex™ System, four loci (D16S539, D7S820, D13S317 and D5S818) are labeled with fluorescein and displayed in green, and four loci (CSF1PO, TPOX, TH01 and vWA) are labeled with carboxy-tetramethylrhodamine (TMR) and displayed in red (Panel A). These two sets of loci are scanned separately by the instrument and may be displayed as two distinct black and white images (Panels B and C, respectively). The images shown in Figure 1 illustrate several features of each of the *GenePrint*™ Fluorescent STR Multiplexes. Alleles for loci detected using the same color are always spatially separated from neighboring systems. Each locus is provided with an allelic ladder which can be used as a visual size marker or analyzed using the sizing software of the instrument. Very few off-ladder alleles exist.

The Fluorescent Ladder (CXR), 60-400 Bases, is an additional size marker of unrelated sequence composed of 16 evenly spaced DNA fragments labeled with carboxy-X-rhodamine. When this marker is included in each gel lane, the instrument is capable of monitoring and correcting lane-to-lane migration differences of samples. Only one STR locus included in the PowerPlex™ System has shown microvariant alleles (TH01 allele 9.3). In systems that contain more microvariant alleles (e.g., FGA, D18S51 or D21S11), the presence of an additional size marker is more important to identify the small lane-to-lane migration differences which may occur.

Table 1. STR Loci Contained in Multiplex Systems Detected Using Silver Stain or Fluorescent Labels.

STR Multiplex System	CSF1PO	TH01	TPOX	vWA	F13A01	FESFPS	F13B	LPL	D5S818	D7S820	D13S317	D16S539
Silver STRs												
CTT	■	■	■									
FFv				■	■	■						
SilverSTR™ III Multiplex										■	■	■
Fluorescent STRs												
CTTv	■	■	■	■								
FFFL					■	■	■	■				
GammaSTR™ Multiplex									■	■	■	■
PowerPlex™ System	■	■	■	■					■	■	■	■

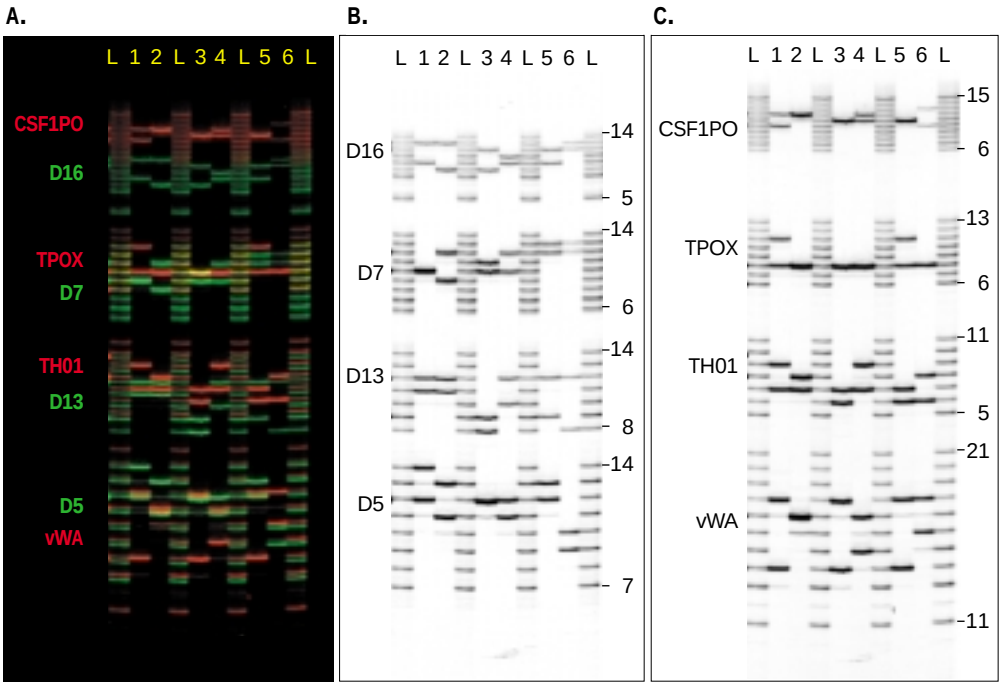

Figure 1. The *GenePrint*™ PowerPlex™ Fluorescent STR System (Fluorescein/TMR) (Hitachi). Six DNA samples were amplified (Lanes 1-6) and are shown with allelic ladders for the corresponding system (Lanes L). Panel A: Two-color image of all eight STR systems which were amplified simultaneously and detected using the Hitachi FMBIO® Fluorescent Scanner. The amplified products of the fluorescein-labeled loci (D16S539, D7S820, D13S317 and D5S818) are shown in green, while the TMR-labeled loci (CSF1PO, TPOX, TH01 and vWA) are shown in red. Panel B: A scan using a 505nm filter, which reveals the corresponding black and white image of the fluorescein-labeled loci (D16S539, D7S820, D13S317 and D5S818). Panel C: A scan using a 625nm filter, which reveals a black and white image of the TMR-labeled loci (CSF1PO, TPOX, TH01 and vWA). In Panels B and C, each allelic ladder is labeled to its right with the number of copies of the repeated sequence contained within its corresponding largest and smallest alleles. All materials were separated using a 4% denaturing polyacrylamide gel.

Differences in the properties of the lasers contained in the Hitachi and ABI instruments require that a separate configuration of the PowerPlex™ System be used with each instrument. Figure 2 shows results obtained using the PowerPlex™ System and the ABI PRISM™ 377 DNA Sequencer. In this case, the four fluorescein-labeled loci (D16S539, D7S820, D13S317 and D5S818) are displayed in purple, the TMR-labeled loci (CSF1PO, TPOX, TH01 and vWA) are shown in black and the 16-fragment Fluorescent Ladder (CXR), 60-400 Bases, is displayed in red. Users have reported that this configuration of the PowerPlex™ System also allows three-color detection analysis using the ABI Model 373 DNA Sequencer and the ABI Model 310 Capillary Electrophoresis Unit, or two-color analysis (without the Fluorescent Ladder (CXR), 60-400 Bases) using the GenomyxSC™ DNA Sequencing System or the Molecular Dynamics FluorImager™ 595 Fluorescent Scanner.

The fluorescein-labeled FFFL Multiplex (Table 1) contains the STR loci F13A01, FESFPS, F13B and LPL. This system is compatible with all of the instruments mentioned above and also with the Molecular Dynamics FluorImager™ 575 and SI Fluorescent Scanners. The two additional fluorescein-labeled quadriplex systems described in Table 1 (i.e., the CTTv and GammaSTR™ Fluorescent STR Multiplexes) are also compatible with all of these instruments (Figure 3).

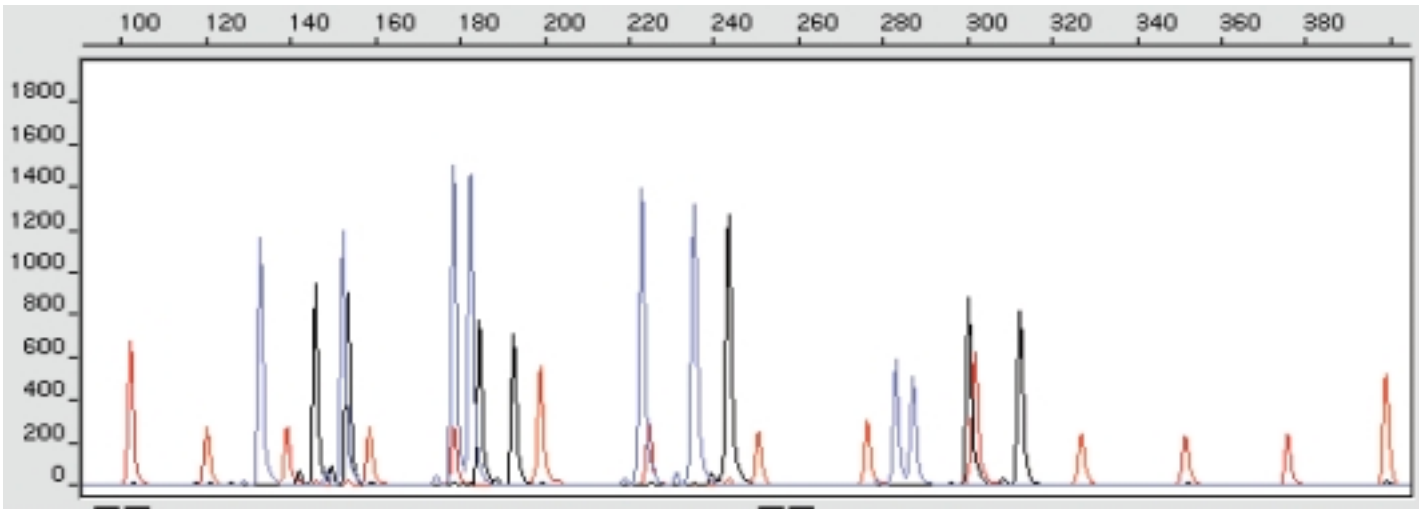

Figure 2. The *GenePrint*™ PowerPlex™ Fluorescent STR System (Fluorescein/TMR) (ABI). The electropherogram of a DNA sample co-amplified at eight loci is shown. The amplified products of the fluorescein-labeled loci (D16S539, D7S820, D13S317 and D5S818) are shown as purple peaks, while the TMR-labeled loci (CSF1PO, TPOX, TH01 and vWA) are shown as black peaks. The Fluorescent Ladder (CXR), 60-400 Bases, is displayed as red peaks. All materials were separated using a 4% denaturing polyacrylamide gel and detected with the ABI PRISM™ 377 DNA Sequencer.

SILVER DETECTION

Six of the twelve loci included in the PowerPlex™ and FFFL Systems are currently available in multiplexes that may be analyzed using silver stain analysis (Table 1). These are contained in the CTT Multiplex (CSF1PO, TPOX, TH01) and the FFv Multiplex (F13A01, FESFPS, vWA). A third multiplex system, the SilverSTR™ III Multiplex (D16S539, D7S820, D13S317), is being developed. Use of these nine loci in silver detection analyses offers significant discriminating power to laboratories that are not able to afford the capital investment required to purchase an instrument for fluorescent detection formats. With these silver detection systems, nine of the same loci that are available for high throughput screening will also be available in multiplex format for low cost, lower throughput needs. While the silver stain approach is slightly more cumbersome and time consuming than fluorescence detection systems, only a gel rig, a power supply and a few plastic tubs are necessary to set up the detection method. Figure 4 illustrates typical results obtained using the CTT Multiplex.

POWER OF DISCRIMINATION

Use of these STR multiplex systems provides extremely powerful discrimination. Preliminary development of population statistics for the twelve-locus combination of the PowerPlex™ and FFFL Systems (Table 2), the eight-locus PowerPlex™ System alone (Table 3), or the nine-locus combination of the three silver stain-compatible STR

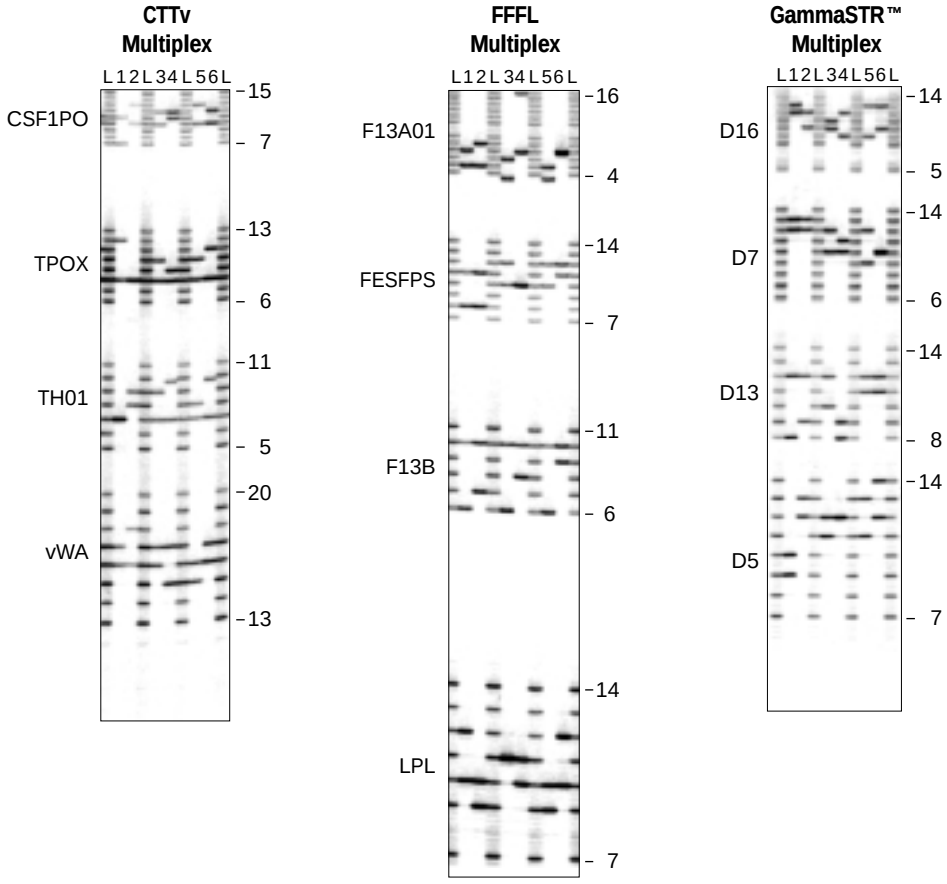

Figure 3. *GenePrint*™ fluorescein-labeled STR Multiplex Systems. In each panel, six DNA samples have been amplified (Lanes 1-6) and are shown along with allelic ladders for the corresponding system (Lanes L). DNA samples amplified using the CTTv, FFFL and GammaSTR™ Multiplex systems are shown. Each allelic ladder is labeled to its right with the number of copies of the repeated sequence contained within its corresponding largest and smallest alleles. All materials were separated using a 4% denaturing polyacrylamide gel and detected using the Molecular Dynamics FluorImager™ SI Fluorescent Scanner.

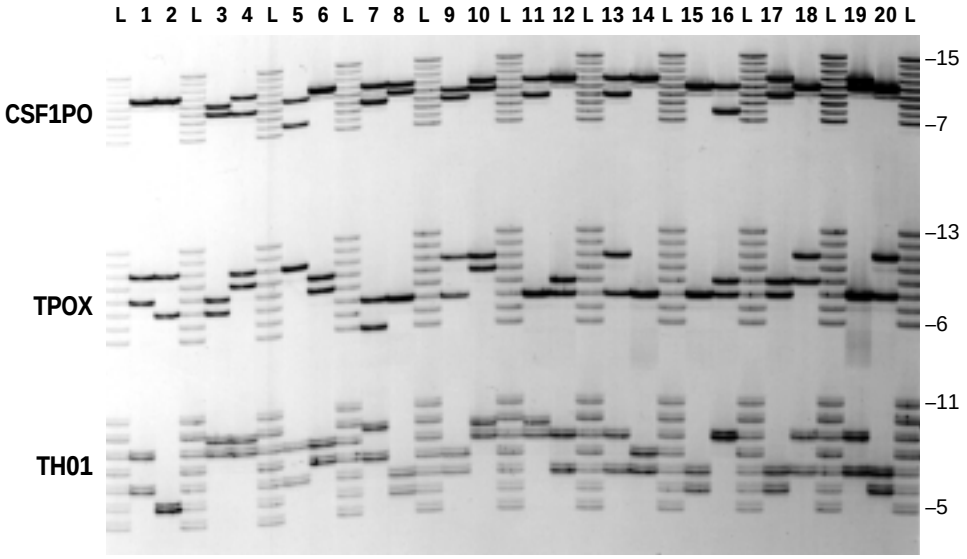

Figure 4. Silver stain detection of the *GenePrint*™ CTT Multiplex System. Twenty DNA samples were amplified and are shown (Lanes 1-20) along with allelic ladders for the corresponding systems (Lanes L). Regions of the gel containing amplified alleles of the CSF1PO, TPOX and TH01 loci are labeled to the left, with the largest and smallest alleles of the corresponding allelic ladders labeled to the right. All materials were separated using a 4% denaturing polyacrylamide gel and detected using the DNA Silver Staining System. The doublets observed at the TH01 locus result from silver staining of opposing strands of the amplified product which are separated in this denaturing gel system due to their sequence differences. With fluorescent systems, only one strand of the amplified product is fluorescently labeled, generating a single visible band (Figure 1).

Multiplexes (Table 4) are illustrated. These data are based on analysis of more than two hundred individuals from each of three major racial and ethnic groups present in the United States: African-Americans, Caucasian-Americans and Hispanic-Americans.

As an example of the enormous discriminating ability of these systems, consider the combined twelve-locus system (i.e., the PowerPlex™ and FFFL Systems) evaluated in the African-American population. The matching probability has been calculated at nearly 1 in 3 trillion. For paternity applications, these two systems used in combination provide typical paternity indices of greater than 2,500 in each racial group, and power of exclusion of greater than 0.9997, making these systems a competitive alternative to VNTRs or any other available system in terms of discriminating power.

Table 2. Population Statistics for the Twelve-Locus Combined PowerPlex™ and FFFL Systems.

	African-American	Caucasian-American	Hispanic-American
Matching Probability	1 in 2.91×10^{12}	1 in 1.78×10^{11}	1 in 2.37×10^{11}
Typical Paternity Index	6,691	5,605	2,617
Typical Power of Exclusion	.99989	.99983	.99973

Table 3. Population Statistics for the Eight-Locus Combined PowerPlex™ System.

	African-American	Caucasian-American	Hispanic-American
Matching Probability	1 in 2.61×10^{8}	1 in 1.18×10^{8}	1 in 1.45×10^{8}
Typical Paternity Index	403	354	319
Typical Power of Exclusion	.9979	.9976	.9973

Table 4. Population Statistics for the Nine-Locus Combined Silver Stain Systems.

	African-American	Caucasian-American	Hispanic-American
Matching Probability	1 in 4.93×10^{9}	1 in 1.05×10^{9}	1 in 1.83×10^{9}
Typical Paternity Index	987	723	556
Typical Power of Exclusion	.9992	.9989	.9986

SUMMARY

The *GenePrint*™ STR Multiplexes described in this article have been selected and developed because they display a minimum of genetic or amplification artifacts and elicit low mutation rates. This has allowed development of reliable, easily applied systems. Loci containing microvariant alleles are nearly absent from the *GenePrint*™ STR Multiplexes. The two loci which do contain microvariants (i.e., TH01 and F13A01) are well characterized and easily included in the analysis. The inclusion of allelic ladders with each system provides a rapid and accurate method of allele determination. The development of both silver stain and fluorescent detection methods for the same STR systems provides universal application across laboratories with different levels of funding and sophistication. Whatever approach for detection is selected, there are enough multiplex systems available in the *GenePrint*™ product line to provide adequate discrimination for all routine forensic and paternity applications.