The *GenePrint*® PowerPlex™ 2.1 System for the FBI Selection of Thirteen CODIS Core STR Loci and the Seven Standard STR Loci for ENFSI

*By James W. Schumm, Ph.D.**
The Bode Technology Group
jschumm@bodetech.com

INTRODUCTION

The era of databasing criminal populations is upon us. Using these databases will make it possible to link suspects to crime scenes and crime scenes to one another through comparisons of biological evidence. To achieve the ultimate benefit of this approach, legislation in a number of European nations, the United States and other countries has created centralized databases that will eventually include DNA profiles of millions of individuals, primarily convicted criminals. To achieve this end, all countries are moving to DNA typing using short tandem repeat (STR) polymorphisms, which provide a rapid, reliable and inexpensive method of analyzing and unambiguously digitizing data from large numbers of samples.

In recent years the selection of specific STR loci to comprise the core of these databases has produced a move toward standardization. In November 1997 (1–3), the Federal Bureau of Investigation (FBI) announced the selection of 13 STR loci to constitute the core of the United States national database, CODIS (Combined DNA Index System). At that time Promega Corporation, the first company to provide multiplex STR (e.g., the CTT Multiplex) and megaplex STR (e.g., the *GenePrint*® PowerPlex™ 1.1) systems, committed itself to providing rapid means to analyze the 13 CODIS core STR loci.

The *GenePrint*® PowerPlex™ 2.1 System has been developed for best performance with the Hitachi FMBIO® II Fluorescence Imaging System. The *GenePrint*® PowerPlex™ 2.1 System alone contains all the STR loci currently recommended by the European Network of Forensic Science Institutes (ENFSI) DNA Working Group. When used in combination with the *GenePrint*® PowerPlex™ 1.1 System, evaluation of the 13 CODIS core loci is achieved in two amplification reactions.

The GenePrint® *PowerPlex™ 2.1 System alone contains all the STR loci currently recommended by the ENFSI DNA Working Group. When used in combination with the* GenePrint® *PowerPlex™ 1.1 System, evaluation of the 13 CODIS core loci is achieved in two amplification reactions.*

THE *GENEPRINT*® POWERPLEX™ 2.1 MULTIPLEX SYSTEM—DESIGN

Beyond a commitment to high-quality systems, several key influences impacted the content and design of the *GenePrint*® PowerPlex™ 2.1 System. The first was the selection of particular loci by key groups responsible for standardization throughout the world. In particular, standards created in North America (the FBI's CODIS core loci) and in Europe (ENFSI-selected loci) have been incorporated into the system design as shown in Table 1. The *GenePrint*® PowerPlex™ 2.1 System contains all ENFSI STR loci. At the same time, complementation with the *GenePrint*® PowerPlex™ 1.1 System allows amplification of the thirteen CODIS STR loci in two reactions. Three STR loci (TPOX, TH01 and vWA) are shared between these two PowerPlex™ Systems, thus limiting the chance of undetected sample mix-up when performing the two reactions to approximately 1 in 1,000.

In addition to standard loci inclusion, the *GenePrint*® PowerPlex™ 2.1 System design was influenced by the requirement to avoid allele overlap between loci labeled with the same color. Several of the multiplex systems that are popular today were designed when limited numbers of database samples had been analyzed. As these systems became widely used, new alleles were identified that revealed overlap with alleles of other loci labeled with the same fluorescent dye. Obviously, this overlap can generate confusion or mistakes in allele calls when these instances go undetected during analysis. Promega examined preliminary allele distribution data

Standards created in North America (the FBI's CODIS core loci) and in Europe (ENFSI-selected loci) have been incorporated into the system design.

*Editor's Note: We wish to thank Jim Schumm for writing this article while employed by Promega Corporation. He is currently employed by The Bode Technology Group in Springfield, Virginia.

contained in the extensive databases of the Forensic Science Service (Rebecca Sparkes, personal communication). Taking into account the allele distributions for the loci in the *GenePrint*® PowerPlex™ 2.1 System, Promega has developed a tight-fitting design of loci in this system. This provides increased confidence that new alleles, observed in future work, will not migrate within the range of alleles from another locus.

In the *GenePrint*® PowerPlex™ 2.1 System, the loci, Penta E, D18S51, D21S11, TH01 and D3S1358, are each labeled with fluorescein while the loci, FGA, TPOX, D8S1179 and vWA, are labeled with carboxy-tetramethylrhodamine. No alleles from one locus overlap with alleles of another system displayed in the same dye. In addition, all CODIS-associated or ENFSI-associated loci are contained below 375 bases except for some rare alleles of the FGA locus, and most loci are contained below 275 bases. This is of great importance when employing degraded sample materials sometimes encountered in forensic settings.

Table 1. Inclusion of Loci Selected as National/International Standards.

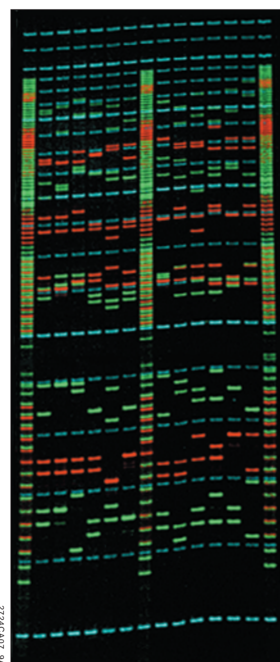
Locus	*GenePrint*® PowerPlex™ 1.1	*GenePrint*® PowerPlex™ 2.1
D16S539	✓	
D7S820	✓	
D13S317	✓	
D5S818	✓	
CSF1PO	✓	
TPOX	✓	✓
TH01	✓	✓
vWA	✓	✓
FGA		✓
D21S11		✓
D8S1179		✓
D18S51		✓
D3S1358		✓
Amelogenin*	✓	✓
Penta E		✓

Loci requested for use in either CODIS or the ENFSI DNA Working Group are indicated by the corresponding brackets.

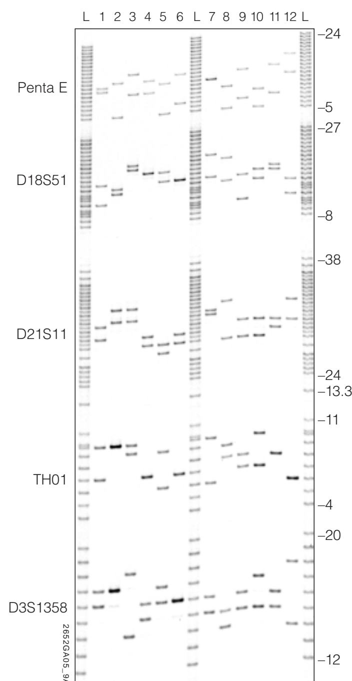
*Amelogenin is not a component of the *GenePrint*® PowerPlex™ 1.1 System but can be amplified along with this system.

Another driving force behind the design of the *GenePrint*® PowerPlex™ 2.1 System was Promega's commitment to provide the forensic and paternity communities with new high-quality loci. A four-year research and development effort produced a number of new loci that combine the benefits of many alleles and a high degree of polymorphism with a scarcity of unwanted character-

A.



B.



C.

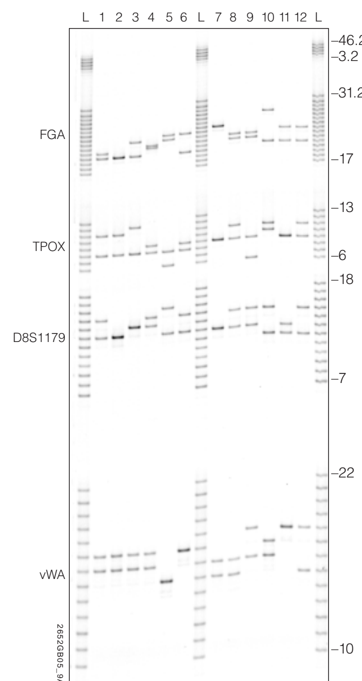


Figure 1. The *GenePrint*® PowerPlex™ 2.1 System. Twelve DNA samples (lanes 1–12) were amplified and are shown with allelic ladders (lanes L) for each of the nine loci contained in the *GenePrint*® PowerPlex™ 2.1 System. **Panel A:** Three-color display with fluorescein-labeled loci, Penta E, D18S51, D21S11, TH01 and D3S1358, shown in green; carboxy-tetramethylrhodamine-labeled loci, FGA, TPOX, D8S1179 and vWA, shown in red; and the carboxy-X-rhodamine-labeled ILS 600 fragments of the size marker shown in blue. **Panel B:** Scan using a 505nm filter, which reveals a black and white image of the fluorescein-labeled loci, Penta E, D18S51, D21S11, TH01 and D3S1358. **Panel C:** Scan using a 585nm filter, which reveals a black and white image of the TMR-labeled loci, FGA, TPOX, D8S1179 and vWA. In Panels B and C, each allelic ladder is labeled to the right with the number of copies of the repeated sequence contained within its corresponding largest and smallest alleles. All materials were separated using a 5% Long Ranger™ denaturing polyacrylamide gel and detected using the Hitachi FMBIO® II Fluorescence Imaging System.

istics (e.g., stutter bands and microvariants). These new loci are based upon pentanucleotide repeats rather than the traditional tetranucleotide repeat loci included in current multiplex systems. For a review of the discovery and the properties of pentanucleotide repeat loci, see reference 4. The first of these loci to be broadly distributed is the Penta E locus included in the *GenePrint*® PowerPlex™ 2.1 System. This locus has twenty alleles ranging from five to twenty-four repeat units with only a single known microvariant (allele 20.3) observed at an approximate frequency of one in eight hundred alleles.

THE *GENEPRINT*® POWERPLEX™ 2.1 MULTIPLEX SYSTEM—PERFORMANCE

Figure 1, Panel A, illustrates a three-color display of the *GenePrint*® PowerPlex™ 2.1 System. Five loci (Penta E, D18S51, D21S11, TH01 and D3S1358) are revealed with fluorescein (green), four loci (FGA, TPOX, D8S1179 and vWA) with carboxy-tetra-methylrhodamine (TMR, red), and the new Internal Lane Standard 600 (ILS 600), labeled with carboxy-X-rhodamine, is shown in blue. The scans of the fluorescein- and TMR-labeled loci are displayed separately in Panels B and C, respectively.

Note that nearly all the commonly existing alleles of each locus are included in the allelic ladders, even two-base variants of the D18S51 and D21S11 loci, and high molecular weight variants of the FGA locus (Table 2).

A new internal lane standard (ILS) has been constructed for use with this system. The Internal Lane Standard 600 (ILS 600) contains 22 fragments ranging in size from 60 bases to 600 bases. The ILS 600 has 16 fragments in common with the Fluorescent Ladder (CXR), 60–400 Bases (a component of the *GenePrint*® PowerPlex™ 1.1 System). In addition, it includes fragments of 425, 450, 475, 500, 550 and 600 bases. The design of this size marker provides each fragment essentially as a subset of the sequence of the next larger fragment. In this way, sequence variation does not disrupt the dependence of the relative migration on fragment length that is seen in other commercially available markers. The even distribution of fragments and ease of interpretation of the fragment sizes are demonstrated in Figure 2.

Figure 3 illustrates that as little as 0.1ng of DNA template is appropriately amplified at all loci using the very sensitive *GenePrint*® PowerPlex™ 2.1 System. Note that the most even distribution of allele intensities across loci is seen with 1ng of sample. This is the amount used to set our specifications with regard to allele balance.

An increase in the amount of template used, or increasing the number of cycles of thermal cycling during amplification, can produce an imbalanced profile with greater yield of smaller alleles relative to the larger alleles (data not shown). This phenomenon, observed with a variety of multiplex systems, is often seen with the use of FTA™ paper, which is suspected to contain large amounts of DNA even in small punches. Generally, this allele imbalance can be overcome by lowering the number of cycles in the thermal cycling program (for more information, see the GenePrint® *PowerPlex™ 2.1 System Technical Manual* #TMD011, reference 5).

THE *GENEPRINT*® POWERPLEX™ 2.1 MULTIPLEX SYSTEM—DISCRIMINATION POWER

Population statistics related to three of the *GenePrint*® PowerPlex™ 2.1 System loci, TH01, TPOX and vWA, have been previously published by Lins *et al.* (6). Preliminary analyses on the remaining loci have been developed as a collaboration between The Bode Technology Group and Promega Corporation. The power of discrimination, typical paternity index and typical power of exclusion are displayed for each of four different population groups for the *GenePrint*® PowerPlex™ 1.1 System alone (Table 3), *GenePrint*® PowerPlex™ 2.1 System alone (Table 4) and the *GenePrint*® PowerPlex™ 1.1 and 2.1 Systems together (Table 5).

CONCLUSIONS

The move to global standardization of STR analysis has accelerated. The selection of 13 STR core loci by the FBI for use in the United States national database, CODIS, and the identification of seven STR loci from this set plus the Amelogenin locus as recommended standards from the ENFSI DNA Working Group are major steps towards this goal. The development of systems such as the *GenePrint*® PowerPlex™ 1.1 and 2.1 Systems supports this process of database development.

Table 2. Characteristics of Loci Contained in the *GenePrint*® PowerPlex™ 2.1 System.

STR Locus	Label	Size Range of Allelic Ladder Components (bases)	Repeat Number of Allelic Ladder Components	Alleles not in Allelic Ladder (Frequency > 0.001)	Alleles not in Allelic Ladder (Frequency > 0.002)
Penta E	FL	379–474	5–24	20.3	
D18S51	FL	290–366	8–10, 10.2, 11–13, 13.2, 14–27		
D21S11	FL	203–259	24, 24.2, 25, 25.2, 26–28, 28.2, 29, 29.2, 30, 30.2, 31, 31.2, 32, 32.2, 33, 33.2, 34, 34.2, 35, 35.2, 36–38		
TH01	FL	156–195	4–9, 9.3, 10–11, 13.3		
D3S1358	FL	115–147	12–20		
FGA	TMR	326–444	17–30, 31.2, 43.2, 44.2, 45.2, 46.2	18.2, 19.2, 22.2 23.2, 24.2, 25.2	22.2, 23.2
TPOX	TMR	262–290	6–13		
D8S1179	TMR	203–247	7–18		
vWA	TMR	123–171	10–22		

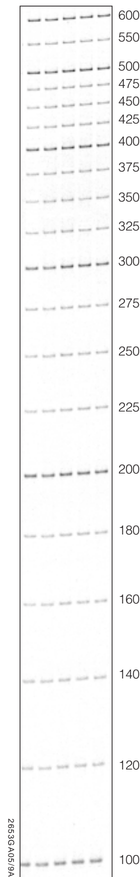


Figure 2. The Internal Lane Standard 600. The Internal Lane Standard 600 contains fragments ranging from 60 to 600 bases in length. It was mixed with the samples shown in Figure 1 before loading the gel. Following separation, the Internal Lane Standard 600 was detected using a 650nm filter with the Hitachi FMBIO® II Fluorescence Imaging System. Fragments smaller than 100 bases are not shown on this gel. Fragment sizes are shown to the right of the gel. The 100, 200, 300, 400, 500 and 600 base fragments display double the intensity of the others.

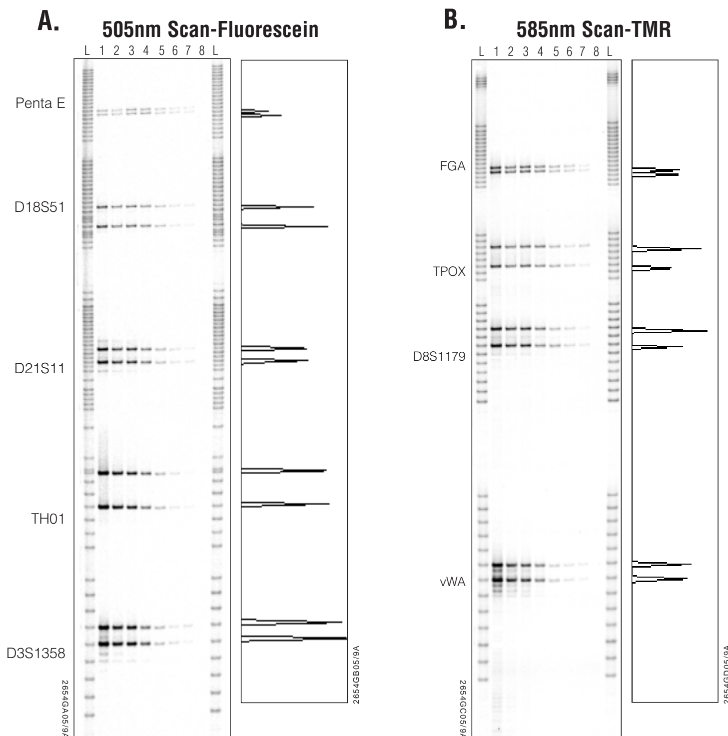


Figure 3. Amplification of various amounts of template with the *GenePrint*® PowerPlex™ 2.1 System. The *GenePrint*® PowerPlex™ 2.1 System was used to amplify 10ng, 5ng, 2ng, 1ng, 0.5ng, 0.2ng, 0.1ng of a human DNA template. Products were separated in lanes 1–7, respectively. The fluorescein-labeled loci of the PowerPlex™ 2.1 System (Penta E, D18S51, D21S11, TH01 and D3S1358) are shown in the 505nm scan (Panel A). The TMR-labeled loci of the PowerPlex™ 2.1 System (FGA, TPOX, D8S1179 and vWA) are shown in the 585nm scan. A lane trace of lane 4 (1ng template) is shown to the right of each panel.

Table 3. Population Statistics Using the *GenePrint*® PowerPlex™ 1.1 System.

	African-American	Caucasian-American	Hispanic-American	Asian-American
Power of Discrimination	1 in 2.74×10^8	1 in 1.14×10^8	1 in 1.45×10^8	1 in 1.32×10^8
Typical Paternity Index	498	260	319	471
Typical Power of Exclusion	0.9982125	0.9968853	0.9973337	0.9981793

Table 4. Population Statistics Using the *GenePrint*® PowerPlex™ 2.1 System.

	African-American	Caucasian-American	Hispanic-American	Asian-American
Power of Discrimination	1 in 3.0×10^{11}	1 in 8.46×10^{10}	1 in 1.02×10^{11}	1 in 1.52×10^{11}
Typical Paternity Index	13,130	13,199	3,250	41,800
Typical Power of Exclusion	0.9999219	0.9999242	0.9997134	0.9999759

Table 5. Population Statistics Using the *GenePrint*® PowerPlex™ 1.1 System and PowerPlex™ 2.1 System.

	African-American	Caucasian-American	Hispanic-American	Asian-American
Power of Discrimination	1 in 4.19×10^{16}	1 in 1.07×10^{16}	1 in 1.59×10^{16}	1 in 2.17×10^{16}
Typical Paternity Index	6.11×10^5	4.08×10^5	1.63×10^5	2.02×10^6
Typical Power of Exclusion	0.9999988	0.99999982	0.9999951	0.9999995

ACKNOWLEDGMENTS

I wish to thank Cindy Sprecher for developing some of the figures for this work and Laura Hennes and Paraj Mandrekar for their part in evaluation of the preliminary population data generated in collaboration with Kevin McElfresh, Deborah DiPierro and Lisa Barnes of The Bode Technology Group, Inc. I also wish to thank Rebecca Sparkes of the Forensic Science Service for sharing preliminary data on the allele distributions of the loci contained within the *GenePrint*® PowerPlex™ 1.1 and 2.1 Systems from the FSS database.

REFERENCES

1. Budowle, B. (1997) Studies for selecting core STR loci for CODIS. Presented at the *DNA Forensics: Science, Evidence, and Future Prospects.* McLean, VA, November 17–18, 1997.
2. Moretti, T. (1998) The CODIS STR Project: evaluation of fluorescent multiplex STR systems. Presented at the *American Academy of Forensic Sciences Annual Meeting*. San Francisco, CA, February 9–14, 1998.
3. Budowle, B. (1998) Examples of STR population databases for CODIS and for casework. Presented at the *9th Symposium on Human Identification*. Orlando, FL, October 7–10, 1998.
4. Bacher, J. and Schumm, J. (1998) *Profiles in DNA* **2**(2), 3.
5. GenePrint® *PowerPlex™ 2.1 System Technical Manual* #TMD011, Promega Corporation.
6. Lins, A.M. *et al.* (1998) *Journal of Forensic Sciences* **43**, 1178.

GenePrint is a trademark of Promega Corporation and is registered with the U.S. Patent and Trademark Office. PowerPlex is a trademark of Promega Corporation.

FMBIO is a registered trademark of Hitachi Software Engineering Company, Ltd. FTA is a trademark of Flinders Technologies Pty, Ltd. Long Ranger is a trademark of FMC Corporation.