CROATIAN MEDICAL JOURNAL
CMJ
42(3):256-259,2001

FORENSIC SCIENCES

Establishing a Large DNA Data Bank Using the PowerPlex™ 1.1 and 2.1 Systems

Jeffrey D. Ban

Forensic Biology Section, Virginia Division of Forensic Science, Richmond, Va, USA

In the early 1990's, the importance of establishing a DNA Data Bank of convicted sex offender samples for comparison to unsolved cases became apparent to the Virginia Division of Forensic Science to help identify potential perpetrators. Ultimately, through the expansion of the data basing law to include all convicted offenders and juveniles convicted of a crime that would be considered a felony if tried as an adult, the Division has successfully used the DNA Data Bank to aid the law enforcement community in solving crimes where the victim was unable to identify the perpetrator. As the number of offender sample analyses has increased, in combination with the number of analyses of cases where a suspect could not be identified, the number of DNA Data Bank hits has also significantly increased. Initially, in 1997, when the Division converted its DNA Data Bank program from the restriction fragment length polymorphism technology to the short tandem repeat technology, one offender hit occurred on average for every 2,900 convicted offender samples that were entered into the Data Bank. However, by December 31, 2000, one DNA Data Bank hit occurred on average for every 700 samples entered into the Data Bank.

Key words: *databases, factual; DNA; DNA fingerprinting; forensic medicine; outsourced services; polymorphism, restriction fragment length; short tandem repeat; United States*

Establishment of the Division's STR Program

Initially, the Virginia Division of Forensic Science began to use the restriction fragment length polymorphism (RFLP) technology in the analysis of the convicted sex offender blood samples and a limited number of blood samples from offenders convicted of burglary in-house. However, due to the approximately 25,000 new blood samples received by the Division each year, by January of 1998 over 150,000 convicted offender samples had been received. During a 5-year period (1993-1997), the Division analyzed approximately 10,500 of these samples using the RFLP technology. These analyses resulted in 30 DNA Data Bank hits when the foreign DNA profile from the evidence samples was searched in the Combined DNA Index System (CODIS) against the convicted offender index comprised of the DNA profiles from convicted offenders, the forensic index comprised of foreign DNA profiles identified on the evidence from previously analyzed cases (suspect and non-suspect cases), or against both indices. However, with the advances in DNA technology, which allowed the analysis of smaller biological samples in shorter periods of time, it became apparent to the Division that typing both crime scene materials and analyzing convicted offender blood samples with short tandem repeat (STR) analysis offered significant advantages over the previously used RFLP technology. Therefore, in July 1997, after conducting validation studies, including sensitivity studies, and determining the optimum amplification and typing conditions as well as an upper stutter value for each locus (Table 1), the Division incorporated STRs and began to analyze convicted offender samples using the Promega *GenePrint*® PowerPlex™ 1.1 System kit, which includes the CSF1PO, TPOX, TH01, vWA, D16S539, D7S820, D13S317, and D5S818 loci (3) (Table 2).

In June 1998, the Division began to analyze evidentiary samples using the Promega *GenePrint*® PowerPlex™ 1.1 System kit and search the DNA profile foreign to the victim from both suspect and non-suspect cases in CODIS. Additionally, due to the large number of convicted offender samples that were backlogged and the Division's inability to analyze this high volume of samples with the current staff, in July 1998, the Division began to outsource these samples to The Bode Technology Group, Inc., located in Springfield, Virginia, USA. Approximately 70,000 samples per year were analyzed by the contract laboratory using the Promega *GenePrint*® PowerPlex™ 1.1 System kit.

Approximately a year after the Division began outsourcing the analysis of these samples and approximately six and a half months after the Division began analyzing crime scene samples using STRs from cases where a suspect could not be identified, the Division surpassed the number of DNA data bank hits it took five and a half years to make using the RFLP technology. By January 1, 2001, the Division had a total of

283 DNA data bank hits, 247 offender (Table 3) and 36 forensic hits. During the calendar year of 2000, the Division achieved 178 (160 offender and 18 forensic hits) of the 283 DNA data bank hits, averaging more than 3 hits per week (Fig. 1). To put this figure in perspective, as of January 2001 a total of 705 offender hits have been made in 24 states according to the Federal Bureau of Investigation; Virginia is credited with 247 of these hits, with the next closest state being Florida, which accounted for 129 offender hits.

Although the Promega *GenePrint*® PowerPlex™ 1.1 System routinely provides sufficient discrimination power (Table 4) for identifying a single individual when a DNA profile is searched in CODIS from evidence samples containing a single foreign donor's DNA, this is not always the case when a sample contains a mixture of body fluids from two or more unknown individuals (6). Therefore, to provide better discrimination power and reduce the number of adventitious hits when searching a DNA profile from an evidence sample consisting of a mixture of DNA from two or more donors, in January 2000, the Division began to analyze convicted offender samples using the Promega *GenePrint*® PowerPlex™ 2.1 System kit (Table 4), which includes the Penta E, D18S51, D21S11, TH01, D3S1358, FGA, TPOX, D8S1179 and vWA loci (Table 5)(6). In May 2000, the Division began to use the Promega *GenePrint*® PowerPlex™ 2.1 System kit in the anylysis of evidentiary samples, when an evidence sample contained a mixture of two or more unknown individuals where no primary donor could be identified, and a large number of adventitious hits were made by solely the *GenePrint*® PowerPlex™ 1.1 System. The *GenePrint*® PowerPlex™ 2.1 System was also used when the samples appeared to have had locus dropout at the higher molecular weight loci on the *GenePrint*® PowerPlex™ 1.1 System due to the quality of the DNA sample when received into the laboratory or a limited amount of sample present. Before the Division implemented the *GenePrint*® PowerPlex™ 2.1 System for analyzing evidence samples, upper stutter values for each locus had been initially established by use of non-mixture samples. Subsequently, these stutter values were further evaluated with mixed samples containing DNA from two individuals of known types to establish the final upper stutter values used by the Virginia Division of Forensic Science (Table 6).

Table 1. Upper stutter values for the *GenePrint*® PowerPlex™ 1.1 system[a]

Locus	% Stutter	Locus	% Stutter
CSF1PO	11.0	D16S539	12.0
TPOX	8.0	D7S820	9.0
TH01	5.0	D13S317	9.0
vWA	14.0	D5S818	11.0

[a]For each locus, the stutter band optical densities (OD) were averaged, the preceding allele optical densities were averaged, and the stutter OD average/allele OD average percentage was calculated. The standard deviation (SD) of this percentage was calculated and then brought out to three standard deviations (3SD). This standard deviation was added to the stutter OD average/allele OD average percentage to obtain a cutoff percentage with a 99% confidence interval. These values were then rounded up to the nearest whole number.

DNA Data Bank Staff

The Division currently employs a DNA data bank supervisor and 2 full time DNA data bank exam-

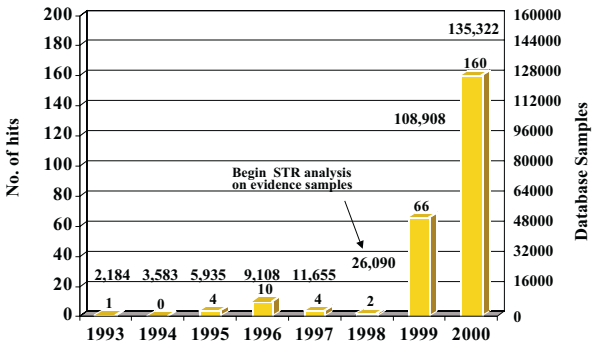

Figure 1. Number of hits vs samples analyzed. Initially the Virginia Division of Forensic Science began to analyze convicted offender samples using the restriction fragment length polymorphism (RFLP) technology. In 1997, the Division converted to the short tandem repeat (STR) technology for convicted offender samples and in 1998 for crime scene samples. As the number of convicted offender samples analyzed increased in conjunction with the analysis of non-suspect cases, the number of DNA Data Bank hits increased. The number of DNA Data Bank hits specified above reflects the number of offender hits.

Table 2. The *GenePrint*® PowerPlex™ 1.1 system locus-specific information[a] (3)

Locus	Repeat sequence 5′-3′	Chromosome location	Size range of allelic ladder (bp)	Alleles present in allelic ladder	Fluorescent label
D16S539	AGAT	16q24-qter	264-304	5,8-15	fluorescein[b]
D7S820	AGAT	7q11.21-22	215-247	6-14	fluorescein
D13S317	AGAT	13q22-q31	165-197	7-15	fluorescein
D5S818	AGAT	5q23.3-32	119-151	7-15	fluorescein
CSF1PO	AGAT	5q33.3-34	291-327	6-15	TMR[c]
TPOX	AATG	2p25-1pter	224-252	6-13	TMR
TH01	AATG	11p15.5	179-203	5-11	TMR
vWA	TCTA	12p12-pter	127-167	11,13-21	TMR

[a]All repeat sequences were defined using the recommendation of the DNA Commission of the International Society of Forensic Haemogenetics (ISFH): 1) for STR loci within coding genes, the coding strand shall be used and the repeat sequence motif defined using the first possible 5′ nucleotide of the repeat motif; and 2) for STR loci not associated with a coding gene, the first database entry or original literature description shall be used (4,5).
[b]Fluorescein is detected at a wavelength of 505 nm.
[c]TMR – carboxy-tetramethylrhodamine is detected at a wavelength of 585 nm.

iners, who conduct STR analysis on some of the convicted offender blood samples submitted to the Division, as well as review the analytical data that is returned from the contract laboratory on a bi- monthly basis. In addition, the DNA data bank also employs 8 part-time staff who dry down the blood samples onto stain cards for permanent storage, file the samples for easy retrieval when ready for analysis, generate lists of samples that require analysis, locate, pull, and package the convicted offender samples for distribution to the contractor, prepare chain of custody documentation associated with each shipment of samples and inventory, and re-file the samples upon return from the contractor. Once the convicted offender samples have been analyzed by the *GenePrint*® PowerPlex™ 1.1 and 2.1 Systems, the DNA data bank staff also upload into CODIS the approximate 3,000 to 4,000 STR profiles that are returned from the contractor on a monthly basis.

Table 3. Previous criminal convictions of offenders identified[a]

Previous criminal convictions	%
Sex crime	7
Homicide	3
Wound/assoult	7
Burglary/robbery	42
Drugs	10
Forgery/uttering	4
Miscellaneous	27

[a]In 1989, the Virginia Division of Forensic Science initially received blood samples from convicted sex offenders. However, the law was expanded to included all felony convictions in 1998, after a study demonstrated that criminals elevated to more serious crimes from their original criminal conviction after released from prison. Historically, more than half of the DNA data bank hits that have been made on sexual assault cases have identified an individual originally convicted of burglary/robbery. The percentages represent individuals identified as a result of a hit to the convicted offender DNA data bank.

Storage of Data Bank Samples

The blood samples that are collected from all convicted felony offenders in Virginia are drawn by medical personnel at the local or regional jails, the Virginia Department of Correction facilities, or at the local health department if the offender is not serving time in a jail or correctional facility. Subsequently, the blood samples are delivered to the Virginia Division of Forensic Science, where each sample is deposited on a stain card and allowed to dry for long term storage. Due to the stability of the DNA, once

Table 6. Upper stutter values for the *GenePrint*® PowerPlex™ 2.1 system[a]

Locus	% Stutter	Locus	% Stutter
Penta E	2.0[b]	D3S1358	10.0
D18S51	9.0	FGA	9.0
D21S11	10.0	D8S1179	8.0

[a]For each locus, the stutter band optical densities (OD) were averaged, the preceding allele optical densities were averaged, and the stutter OD average/allele OD average percentage was calculated. The standard deviation (SD) of this percentage was calculated and then brought out to three standard deviations (3SD). This standard deviation was added to the stutter OD average/allele OD average percentage to obtain a cutoff percentage with a 99% confidence interval. These values were then rounded up to the nearest whole number.
[b]No stutter was observed during the validation studies. Therefore, the percent stutter specified is based upon recommendations of the manufacturer reported in the *GenePrint*® PowerPlex™ 2.1 System Technical Manual (6).

Table 4. Power of discrimination of the *GenePrint*® PowerPlex™ Systems in various populations

	Power of discrimination		
STR System	African-American	Caucasian-American	Hispanic-American
PowerPlex™ 1.1 System (8 STR loci)	0.9982125	0.9968853	0.9973337
PowerPlex™ 2.1 System (9 STR loci)	0.9999219	0.9999242	0.9997134
PowerPlex™ 1.1 System plus PowerPlex™ 2.1 System (14 STR loci)	0.9999988	0.9999982	0.9999951

Table 5. The *GenePrint*® PowerPlex™ 2.1 system locus specific information[a] (6)

Locus	Repeat sequence 5′-3′	Chromosome location	Size range of allelic ladder (bp)	Alleles present in allelic ladder	Fluorescent label
Penta E	AAAGA	15q	379-474	5-24	fluorescein[b]
D18S51	AGAA	18q21.3	290-366	8-10, 10.2, 11-13, 13.2, 14-27	fluorescein
D21S11	TCTA	21q11-21q21	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	fluorescein
TH01	AATG	11p15.5	156-195	4-9, 9.3, 10-11, 13.3	fluorescein
D3S1358	TCTA	3p	115-147	12-20	fluorescein
FGA	TTTC	4q28	326-444	17-30, 31.2, 43.2, 44.2, 45.2, 46.2	TMR[c]
TPOX	AATG	2p23-2pter	262-290	6-13	TMR
D8S1179	TCTA	8	203-247	7-18	TMR
vWA	TCTA	12p12-pter	123-171	10-22	TMR

[a]All repeat sequences were defined according to the recommendation of the DNA Commission of the International Society of Forensic Haemogenetics (ISFH): 1) for STR loci within coding genes, the coding strand shall be used and the repeat sequence motif defined using the first possible 5′ nucleotide of the repeat motif; and 2) for STR loci not associated with a coding gene, the first database entry or original literature description shall be used (4,5).
[b]Fluorescein is detected at a wavelength of 505 nm.
[c]TMR - Carboxy-tetramethylrhodamine is detected at a wavelength of 585 nm.

the samples are dried, each sample is stored in a humidity-controlled room at room temperature within an individual envelope containing a unique bar coded DNA number. The envelopes are then stored in numbered boxes. A particular sample can be located based on the box number and the unique sample number by querying the sample in the DNA data bank sample tracking computer.

Conclusion

Thus far, the Division's approach toward establishing and effectively using a large DNA data bank and the *GenePrint*® PowerPlex™ 1.1 and 2.1 Systems has proven to be quite successful. As the number of convicted offender blood samples analyzed increases in combination with the analysis of non-suspect cases, the number of DNA data bank hits has increased four-fold. Undoubtedly, the Division will continue to increase the number of DNA data bank hits at an even greater rate in the future, as the DNA Data Bank continues to grow due to the use of the STR technology.

Acknowledgments

The author would like to thank Paul B. Ferrara, Ph.D., Virginia Division of Forensic Science Division Director, and Deanne F. Dabbs, Forensic Biology Section Program Manager, for their valuable assistance in the reviewing process of this manuscript.

References

1 Code of Virginia, 1950. Code of Virginia Sect. 19.2-310.2. Charlottesville (VA): Matthew Bender and Company, Inc.; 2000.

2 Code of Virginia 1950. Code of Virginia Sect. 16.1-299.1. Charlottesville (VA): Matthew Bender and Company, Inc., 2000.

3 Promega Corporation. GenePrint® PowerPlex™ 1.1 System Technical Manual 1999.

4 Bär W, Brinkmann B, Budowle B, Carracedo A, Gill P, Lincoln P, et al. DNA recommendations: further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. Int J Legal Med 1997;110:175-6.

5 Gill P, Brinkmann B, d'Aloja E, Andersen J, Bär W, Carracedo A, et al. Considerations from the European DNA profiling group (EDNAP) concerning STR nomenclature. Forensic Sci Int 1997;87:185-92.

6 Promega Corporation. GenePrint® PowerPlex™ 2.1 System Technical Manual 1999.

Received: March 21, 2001
Accepted: April 10, 2001

Correspondence to:
Jeffrey D. Ban
Virginia Division of Forensic Science
700 North Fifth Street
Richmond, Virginia 23219, USA
Jban@dfs.state.va.usa