TECHNICAL NOTE

Bruce Budowle,[1] *Ph.D. and Cynthia J. Sprecher,*[2] *B.S.*

Concordance Study on Population Database Samples Using the PowerPlex™ 16 Kit and AmpFℓSTR® Profiler Plus™ Kit and AmpFℓSTR® COfiler™ Kit*

REFERENCE: Budowle B, Sprecher CJ. Concordance study on population database samples using the PowerPlex™ 16 kit and AmpFℓSTR® Profiler Plus™ kit and AmpFℓSTR® COfiler™ kit. J Forensic Sci 2001;46(3):637–641.

ABSTRACT: Over 500 population database samples comprising African Americans, Bahamians, and Southwestern Hispanics were typed using the PowerPlex™ 16 and the Profiler Plus™ COfiler™ kits. There was only one sample in which a typing difference was observed. An FGA heterozygote profile was observed using the PowerPlex™ 16 primers, and a single allele FGA profile was observed using Profiler Plus primers. Thus, the extant data suggest that the primers used in the PowerPlex™ 16, Profiler Plus™, and COfiler™ kits are reliable for typing reference samples destined for use in CODIS. In addition, African American, Bahamian, and Southwestern Hispanic databases have been established for the STR loci Penta D and Penta E. Both loci are highly polymorphic. The application of the product rule is valid for estimating the rarity of a multiple loci profile consisting of these two and the 13 core STR loci.

KEYWORDS: forensic science, AmpFℓSTR® Profiler Plus™ Kit, AmpFℓSTR® COfiler™ Kit, PowerPlex™ 16 kit, GenePrint® PowerPlex™ 2.1 kit, CODIS, concordance, primers, STR, stochastic effects, allele drop-out, Penta D, Penta E, Hardy-Weinberg expectations

The 13 short tandem repeat (STR) loci CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, and D21S11 have been selected as the core loci for use in CODIS (1). Primers for each of the loci are designed to reside in invariant regions flanking the repeats of each STR locus. However, when a sufficiently large number of individuals are typed, a variant may be observed in the flanking region of the target sequence where a primer resides. If a mismatch in complementary binding is proximal to the 3′ end of the primer, no amplification of the target sequence may result (2–4). Profile comparisons within and between laboratories that use the same primer sets would be unaffected by potential allele drop-out. Although it is impossible to ensure that allele drop-out will never occur, it appears to be relatively rare for the current primer/STR systems (2–4); so allele drop-out is not a significant concern for profile comparisons even if different primer/STR systems are used. CODIS has recognized the following currently available systems/primers as valid: AmpFℓSTR® Profiler™ Kit, AmpFℓSTR® Profiler Plus™ Kit, and AmpFℓSTR® COfiler™ Kit (PE Biosystems, Foster City, CA) and the GenePrint® PowerPlex™ 1.1 kit, GenePrint® PowerPlex™ 1.2 kit, GenePrint® PowerPlex™ 2.1 kit, and 8 monoplex systems (Promega Corporation, Madison, WI).

Before using new primer sets for typing reference samples for CODIS, allele drop-out should be demonstrated to be low. New primer sets can be assessed by a concordance study. Concordance here is defined as a comparison of results obtained from typing reference database samples using primers from some of the above approved kits with results obtained using the proposed primer sets. If few or no typing discrepancies are observed, evidence is strong that there is no substantive allele drop-out. A new multiplex kit PowerPlex™ 16 kit (Promega Corporation, Madison, WI) enables simultaneous amplification of the 13 core CODIS STR loci, two additional STR loci (Penta D and Penta E), and the amelogenin locus has become available. This paper describes a concordance study on more than 500 population database samples comparing the results obtained with the PowerPlex™ 16 kit (Promega Corporation, Madison, WI) and those obtained using the AmpFℓSTR® Profiler Plus™ Kit and AmpFℓSTR® COfiler™ Kit (PE Biosystems, Foster City, CA). The data are also useful to support that the sequences of the primers for STR loci do not need to be known to demonstrate validity.

Materials and Methods

Three sample populations were analyzed: African Americans ($N = 170$), Bahamians ($N = 160$), and Southwestern Hispanics ($N = 197$). Sample sources, preparation, and typing results have been reported (2). In an earlier study, the African American and Hispanic samples were amplified at the loci FGA, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D18S51, and D21S11 us-

[1] Senior scientist, Laboratory Division, FBI Academy, Quantico, VA.
[2] Senior scientist, Promega Corporation, 2800 Woods Hollow, Madison, WI.
*This is publication number 00-10 of the Laboratory Division of the Federal Bureau of Investigation. Names of commercial manufacturers are provided for identification only, and inclusion does not imply endorsement by the Federal Bureau of Investigation.

Received 16 June 2000; and in revised form 21 July 2000; accepted 21 July 2000.

ing the AmpFℓSTR® Profiler Plus™ Kit (PE Biosystems, Foster City, CA) (i.e., Profiler Plus kit) and at the loci CSF1PO, TPOX, TH01, D3S1358, D7S820, and D16S539 using the AmpFℓSTR® COfiler™ Kit (PE Biosystems, Foster City, CA) (i.e., COfiler kit). The loci D3S1358 and D7S820 were typed with both kits. The Bahamian samples were amplified at the loci CSF1PO, TPOX, TH01, vWA, D5S818, D7S820, D13S317, and D16S539 using the GenePrint® PowerPlex™ 1.1 kit (Promega Corporation, Madison, WI) (i.e., PowerPlex kit) and at the loci FGA, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D18S51, and D21S11 using the AmpFℓSTR® Profiler Plus™ Kit (PE Biosystems, Foster City, CA). An alpha kit format of the PowerPlex™ 16 kit (Promega Corporation, Madison, WI), which enables amplification of the 13 core STR loci, two pentanucleotide loci Penta D and Penta E, and the amelogenin locus, was used to type the previously analyzed samples. The PCR and typing were performed as previously described (2) or by following the manufacturer's recommendations.

Statistical Analysis: allele designations were determined by comparison of the sample fragments with those of the allelic ladders provided with each kit. At each locus, the frequency of each allele was calculated from the numbers of each genotype in the sample set (i.e., the gene count method). Unbiased estimates of expected heterozygosity were computed as described by Edwards et al. (5). Possible divergence from Hardy-Weinberg expectations (HWE) was tested by calculating the unbiased estimate of the expected homozygote/heterozygote frequencies (6–9) and also by performing an exact test (10), based on 2000 shuffling experiments. An interclass correlation criterion (11) for two-locus associations was used for detecting disequilibrium between the STR loci. The probability of discrimination (PD) and probability of exclusion (PE) was calculated according to Fisher (12).

A 2 × N contingency table exact test, based on 2000 shuffling experiments, was used to generate a G-statistic (13,14) to test for homogeneity between sample populations. The program was kindly provided by R. Chakraborty (University of Texas School of Biomedical Sciences, Houston, Texas).

Results and Discussion

Since different manufacturer's kits (Promega Corporation and PE Biosystems) were used for concordance typing, the primers used for PCR (although identifying the same locus) differ in their sequence. The comparison of typing results provides data on whether the frequency of allele drop-out (due to primer mismatch near the 3′ end) is higher in one kit relative to the other. Over 500 samples were analyzed in three population groups at all 13 STR loci, and typing discrepancies were not observed for any samples except two—one at the FGA locus (African American) and one at the D21S11 locus (Hispanic). An FGA heterozygote profile was observed using the PowerPlex™ 16 primers, and a single allele FGA profile was observed using Profiler Plus primers (Fig. 1). The

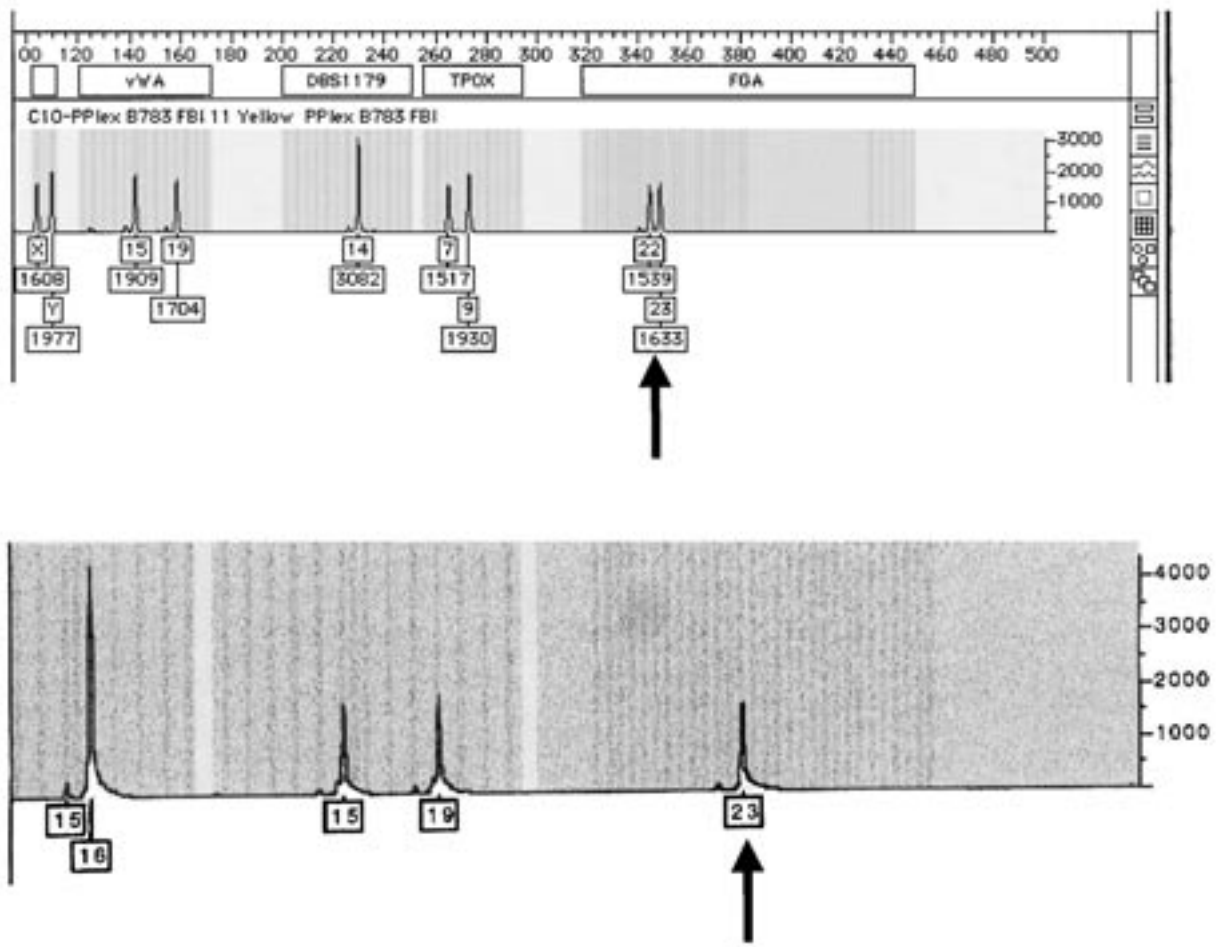

FIG. 1—*The upper panel displays some of the STR loci profiles of a sample generated using the PowerPlex™ 16 kit. The arrow points to the FGA profile (heterozygote). The bottom panel displays some of the STR profiles for the same sample generated using the Profiler Plus™ kit. The arrow points to the FGA profile (single peak).*

presence of the FGA null allele with Profiler Plus primers was confirmed by comparison of the peak height and peak area of the FGA single peak with those of the vWA and D3S1358 alleles generated using the Profiler Plus kit. Heterozygous allele peaks at one locus generally exhibit peak heights and peak areas similar to heterozygous allele peaks at other loci, labeled with the same fluorophore. Homozygous peak heights and peak areas generally are approximately twice those of heterozygous allele peaks. The FGA single peak's height (762 rfu) and area (8424 rfu) were similar to the heterozygous vWA peak heights (762 and 829 rfu) and peak area (8739 and 9336 rfu), and half that of the homozygous D3S1358 peak height (1860 rfu) and peak area (18 232 rfu) for the sample. In addition, when FGA primers described by Urquhart, et al. (15) were used to amplify the same sample, a heterozygous profile was observed (Fig. 2). However, one FGA allele was half the height of the other FGA allele (peak height 880 rfu vs. 1550 rfu), supporting the theory that there is a mismatch in complementary binding in one of the Urquhart, et al. (15) primers.

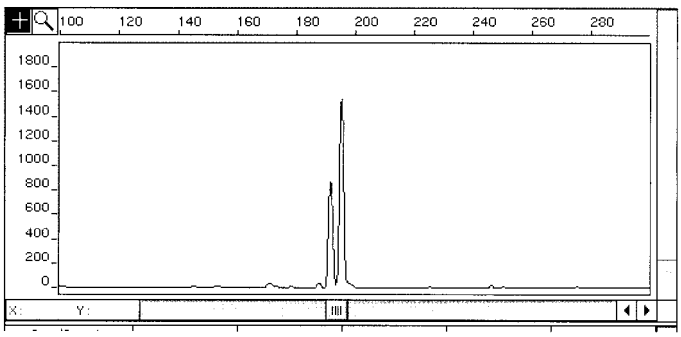

FIG. 2—*FGA profile of the same sample displayed in Fig. 1. The primers used to amplify the sample are those described by Urquhart, et al. (15). A heterozygous profile was observed and the peaks are unbalanced.*

In the other sample, allele drop-out was not due to a primer mismatch. But, the example is illustrative in presenting the different mechanisms that may be responsible for allele dropout, that is primer mismatch and stochastic affects (which should not be confused or misconstrued). At the D21S11 locus, a heterozygote profile was observed using the Profiler Plus primers and a single allele profile was observed using PowerPlex™ 16 primers (data not shown). However, the PowerPlex™ D21S11 peak was below the threshold for interpretation (200 rfu), such that stochastic effects may be the reason for not observing both alleles. Little sample was leftover, however another attempt was made to analyze the sample. The D21S11 type now appears as a heterozygote, consistent with the result obtained using the Profiler Plus kit (Fig. 3). The sample also had been analyzed previously using the PowerPlex™ 2.1 kit, which contains the same primers for the D21S11 locus as in the PowerPlex™ 16 kit, and a heterozygous profile was observed (16).

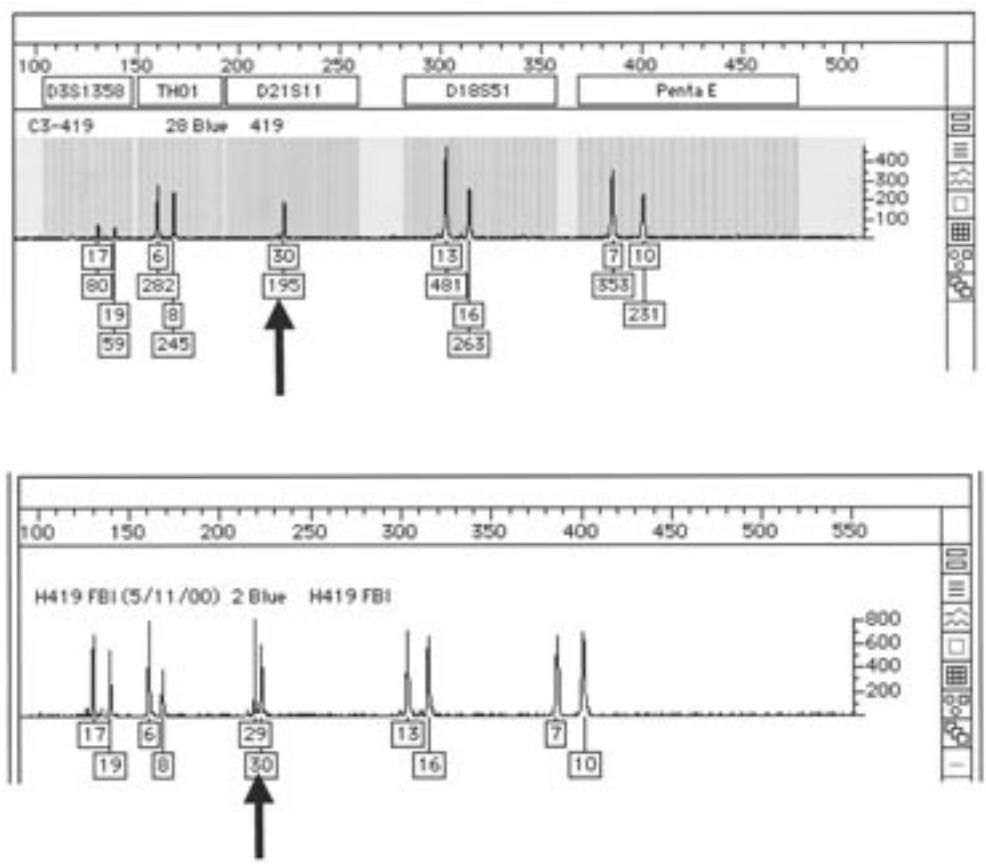

FIG. 3—*The upper panel displays some of the STR loci profiles of a sample generated using the PowerPlex™ 16 kit. The peak heights of the loci are sufficiently low that stochastic effects during PCR should be considered. The arrow points to the D21S11 profile (single peak). The bottom panel displays some of the STR profiles for the same sample generated using the same PowerPlex™ 16 kit. The peak heights in the second analysis are higher than in the above panel. The arrow points to the D21S11 profile (heterozygote). Loss of one of the D21S11 peaks (i.e., an allele) in the upper panel is due to stochastic effects during PCR.*

Thus, the data support that the allele drop-out initially observed in this sample at the D21S11 locus was not due to the primer sequence, but the result of stochastic affects during PCR. It is important to appreciate that stochastic affects may cause allele drop-out for low quantity samples during PCR. A minimum stochastic interpretation threshold should be implemented (17).

The data support that allele drop-out is rare for primer sets found in the PowerPlex™ 16 and Profiler Plus and COfiler kits. Reference sample data generated by the kits described in this study should be reliable for entry into CODIS and generally should be compatible for comparison/searching purposes. However, a mismatch at only one locus (in which an allele is in common between two profiles) of a 13-locus profile should be considered a potential inclusion for investigative purposes.

In this study, we tested the primer sets for allele drop-out for databasing/population data samples. It should be noted that the evaluation was carried out with an alpha test format PowerPlex 16 kit. The final version of the kit contains the same primers, thus the impact of the concordance study here still holds; but the concentration of some of the reagent components has been changed. It is important to establish (for any STR kit) the minimum amount of DNA to use for PCR amplification and/or to establish a minimum stochastic threshold based on peak height or area. To obtain typeable results sufficient to avoid stochastic affects during PCR, a minimum of 0.5 ng of template DNA was required (although 1 to 2 ng are preferable). A stochastic threshold should be determined empirically within the laboratory. Thus, before implementing the final version of the kit for reference sample analysis, it is incumbent upon the laboratory to perform an in-house sensitivity of detection experiment and establish a minimum threshold for interpretation. Our study did not perform any validation studies for forensic samples. Before implementation for casework, appropriate validation studies must be performed to determine acceptable working parameters and that reliable results can be obtained when analyzing forensic samples.

TABLE 1—*Observed allele frequencies for the Penta D locus.*

Allele	African Americans (N = 197)	Bahamians (N = 176)	Southwestern Hispanics (N = 202)
2.2	0.119	0.128	0.015
3.2	0.008	0.020	0.000
5	0.043	0.057	0.002
6	0.005	0.000	0.000
6.3	0.003	0.000	0.000
7	0.028	0.034	0.003
8	0.129	0.102	0.020
9	0.157	0.136	0.198
10	0.122	0.116	0.210
11	0.135	0.170	0.141
12	0.129	0.114	0.136
13	0.081	0.097	0.208
14	0.028	0.014	0.062
15	0.010	0.011	0.002
16	0.000	0.000	0.002
17	0.0254	0.000	0.000
Obs. homozyg.	13.7%	11.4%	16.3%
Exp. homozyg.	11.3%	11.3%	16.8%
P value (homozyg)	0.295	0.974	0.873
Exact test	0.501	0.042	0.637
PD	0.973	0.968	0.945
PE	0.765	0.766	0.660

TABLE 2—*Observed allele frequencies for the Penta E locus.*

Allele	African Americans (N = 197)	Bahamians (N = 176)	Southwestern Hispanics (N = 202)
5	0.124	0.105	0.035
6	0.003	0.000	0.002
7	0.102	0.111	0.089
8	0.165	0.210	0.015
9	0.043	0.051	0.010
10	0.046	0.037	0.037
11	0.084	0.043	0.074
12	0.107	0.159	0.191
13	0.122	0.091	0.106
14	0.058	0.060	0.067
15	0.053	0.043	0.104
16	0.020	0.045	0.116
17	0.038	0.037	0.042
18	0.015	0.006	0.040
19	0.005	0.000	0.037
20	0.008	0.003	0.015
21	0.008	0.000	0.010
22	0.000	0.000	0.005
23	0.000	0.000	0.003
24	0.000	0.000	0.003
Obs. homozyg.	10.2%	16.5%	13.4%
Exp. homozyg.	9.6%	11.3%	9.6%
P value (homozyg.)	0.806	0.031	0.066
Exact test	0.918	0.026	0.239
PD	0.979	0.972	0.979
PE	0.801	0.769	0.804

Because the PowerPlex™ 16 kit contains primers for two STR loci in addition to the 13 core CODIS loci, an ancillary benefit of this concordance study is that population data were generated on two pentanucleotide STR loci. The frequency distributions of observed alleles for the Penta D and Penta E STR loci in the three sampled populations are shown in Tables 1 and 2. The observed and expected homozygosities, exact test for departures from HWE, PD, and PE are also provided. Both loci are highly polymorphic in the three populations. In fact, the Penta E locus has a higher discriminating power than any of the 13 core STR loci, and the Penta D locus is equal in power of discrimination to the most polymorphic core loci (2). The observed Penta E heterozygosity ranges from 83.5% (Bahamians) to 89.8% (African Americans). These observations are similar to other population data (18) for this locus. The observed Penta D heterozygosity ranges from 83.7% (Hispanics) to 86.3% (Bahamians).

There was little evidence for departure from HWE in any of the populations. Based on the exact test, the only departure from expectations was observed in the Bahamian sample population (Penta D, $p = 0.042$; Penta E, $p = 0.026$). The departure was significant, but not highly significant. After Bonferroni correction (19), these observations are not significant. The majority of the departures from HWE are due to genotypes consisting of rare alleles (e.g., those observed fewer than five times). Rare alleles generally have no consequence for estimating genotype frequencies, because rare allele frequencies are replaced by a minimum allele frequency (20,21).

An inter-class correlation test analysis was performed to detect any correlations between alleles at any of the pair-wise compar-

isons of the Penta D and Penta E loci and with any of the other PCR-based markers previously described for the three populations groups (2,22,23). Thus, for African Americans and Hispanics 22 loci were evaluated (i.e., 231 pair-wise comparisons performed) and for Bahamians 15 loci were tested (i.e., 105 pair-wise comparisons performed). The number of significant departures is 11 for African Americans (i.e., 4.8% of the observations), 6 for Bahamians (i.e., 5.7% of the observations), and 8 for Hispanics (i.e., 3.5% of the observations). The overall data are consistent with the expectations of independence between pairs of loci in any of the sample populations. Thus, the assumption of independence is valid for estimating a multiple-locus profile frequency.

The allele frequencies for the Penta D and Penta E loci were compared (using a $2 \times N$ contingency table exact test) between the three population databases and at the Penta E locus for African Americans, Caucasians, and Hispanics from Pennsylvania as well (personal communication, C. Tomsey). Significant differences in allele frequencies between major population databases in the present study were observed at the loci (data not shown). However, even though the Bahamians showed some initial departure from HWE, the distribution of alleles at the Penta D locus was similar between African Americans/Bahamians ($p = 0.531$). At the Penta E locus, the allele distributions were similar for Pennsylvania Caucasians/Pennsylvania Hispanics ($p = 0.419$), Pennsylvania African Americans/Pennsylvania Hispanics ($p = 0.137$), Pennsylvania African Americans/Bahamians ($p = 0.283$), Pennsylvania African Americans/African Americans ($p = 0.541$), and Pennsylvania Hispanics/Southwestern Hispanics ($p = 0.566$).

In conclusion, over 500 samples were typed and allele drop-out was observed rarely using primers from either manufacturer's kit. Although allele drop-out can never be entirely eliminated, the extant data suggest that the primers used in the PowerPlex™ 16, Profiler Plus™, and COfiler™ kits are reliable for typing reference samples destined for use in CODIS. Furthermore, the data support that the sequences of the primers for STR loci do not need to be known to demonstrate validity. In addition, African American, Bahamian, and Southwestern Hispanic databases have been established for the loci Penta D and Penta E. Both loci are highly polymorphic. The application of the product rule is valid for estimating the rarity of a multiple loci profile consisting of these two and the 13 core STR loci.

References

1. Budowle B, Moretti TR, Niezgoda SJ, Brown BL. CODIS and PCR-based short tandem repeat loci: law enforcement tools. In: Second European Symposium on Human Identification 1998, Promega Corporation, Madison, WI, 1998;73–88.
2. Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. J Forensic Sci 1999;44:1277–86.
3. Kline MC, Jenkins B, Rodgers S. Non-amplification of a vWA allele. J Forensic Sci 1998;43:250.
4. Walsh S. Commentary on Kline MC, Jenkins B., Rodgers S. Non-amplification of a vWA allele. J Forensic Sci 1998;43:1103–4.
5. Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R. Genetic variation at five trimeric and tetrameric repeat loci in four human population groups. Genomics 1992;12:241–53.
6. Chakraborty R, Smouse PE, Neel JV. Population amalgamation and genetic variation: observations on artificially agglomerated tribal populations of Central and South America. Amer J Hum Genet 1988;43: 709–25.
7. Chakraborty R, Fornage M, Guegue R, Boerwinkle E. Population genetics of hypervariable loci: analysis of PCR based VNTR polymorphism within a population. In: Burke T, Dolf G, Jeffreys AJ, Wolff R, editors, DNA fingerprinting: approaches and applications, Birkhauser Verlag, Berlin, 1991;127–43.
8. Nei M, Roychoudhury AK. Sampling variances of heterozygosity and genetic distance. Genetics 1974;76:379–90.
9. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 1978;89:583–90.
10. Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 1992;48:361–72.
11. Karlin S, Cameron EC, Williams PT. Sibling and parent-offspring correlation estimation with variable family size. Proc Natl Acad Sci 1981;78:2664–8.
12. Fisher RA. Standard calculations for evaluating a blood group system. Heredity 1951;5:95–102.
13. Roff DA, Bentzen P. The statistical analysis of mitochondrial DNA polymorphisms: χ^2 and the problem of small samples. Mol Biol Evol 1989;6:539–45.
14. Lewontin RC, Felsenstein J. The robustness of homogeneity tests in 2 X N tables. Biometrics 1965;21:19–33.
15. Urquhart A, Oldroyd NJ, Kimpton CP, Gill P. Highly discriminating heptaplex short tandem repeat PCR system for forensic identification. BioTechniques 1995;18(1):116–21.
16. Budowle B. STR allele concordance between different primer sets—a brief summary. Profiles in DNA 2000;3(3):10–1.
17. Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Budowle B. Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. J Forensic Sci (in press).
18. Garofano L, Pizzamiglio M, Donato F, Biondo F, Rossetti M. Budowle B. Italian population data on two new short tandem repeat loci: D2S1338 and Penta E. Forens Sci Int 1999;105:131–6.
19. Weir BS., Multiple tests. In: Genetic data analysis, Sinauer Associates, Inc., Sunderland, MA, 1990;109–10.
20. Budowle B, Monson KL, Chakraborty R. Estimating minimum allele frequencies for DNA profile frequency estimates for PCR-based loci. Int J Leg Med 1996;108:173–6.
21. National Research Council II Report. The evaluation of forensic evidence. National Academy Press, Washington, D.C., 1996.
22. Budowle B, Baechtel FS, Smerick JB, Presley KW, Giusti AM, Parsons G, et al. D1S80 population data in African Americans, Caucasians, Southeastern Hispanics, Southwestern Hispanics, and Orientals. J Forensic Sci 1995;40(1):38–44.
23. Budowle B, Koons BW, Moretti TR. Subtyping of the HLA-DQA1 Locus and Independence Testing with PM and STR/VNTR Loci. J Forensic Sci 1998;43:657–60.

Additional information and reprint requests:
Bruce Budowle, Ph.D.
FSRTC
FBI Academy
Quantico, VA 22135