Bruce Budowle,[1] Ph.D.; Jill B. Smerick,[1] M.S.; Kathleen M. Keys,[1] B.S.; and Tamyra R. Moretti, Ph.D.

United States Population Data on the Multiplex Short Tandem Repeat Loci—HUMTHO1, TPOX, and CSF1PO—and the Variable Number Tandem Repeat Locus D1S80

REFERENCE: Budowle B, Smerick JB, Keys KM, Moretti TR. United States population data on the multiplex short tandem repeat loci—HUMTHO1, TPOX, and CSF1PO—and the variable number tandem repeat locus D1S80. J Forensic Sci 1997;42(5):846–849.

ABSTRACT: Allele frequencies for three tetrameric short tandem repeat (STR) loci HUMTHO1, TPOX, and CSF1PO and a variable number tandem repeat locus D1S80 were determined in United States Caucasian, African American, and Hispanic sample populations. All loci, except the TPOX locus in the Caucasian sample population, meet Hardy-Weinberg expectations. There is no evidence for association of alleles among the four loci. The allelic frequency data are similar to other comparable data within the same major population group.

KEYWORDS: forensic science; DNA typing, genetic markers, population genetics, polymerase chain reaction, variable number of tandem repeats, short tandem repeats, THO1, TPOX, CSF1PO, D1S80, United States, allele frequencies

Short tandem repeat (STR) loci are becoming more widely used for genetic characterization of forensic biological evidence (1–4). A commercially available kit enables multiplex PCR of three STR loci, HUMTHO1 (1,2,5), TPOX (6), and CSF1PO (7). Currently, there are little data on allele frequencies and genotype distributions in general United States populations for these STR loci. Another locus that is used widely for forensic analyses is the variable number of tandem repeat (VNTR) locus D1S80. This paper presents allele frequency data in African Americans, Caucasians, and Hispanics from the United States for these four loci. The data demonstrate that these loci can be useful for providing estimates of the frequency of a DNA profile in forensic identity testing.

Materials and Methods

Sample Preparation—Whole blood, obtained in EDTA vacutainer tubes by venipuncture from African American, Caucasian, and southwestern Hispanic individuals, was kindly provided by Dr. A. Eisenberg, University of North Texas Health Science Center, Fort Worth, Texas. The DNA was extracted by the phenol-chloroform method (8). The quantity of extracted DNA was estimated using the slot-blot procedure described by Waye et al. (9) and Budowle et al. (10).

[1]Forensic Science Research and Training Center, FBI Academy, Quantico, VA.

Received 1 July 1996; and in revised form 30 Sept. 1996; accepted 3 Dec. 1996.

STR Typing—The coamplification of HUMTHO1, TPOX, and CSF1PO was performed using the GenePrint STR system and 2–5 ng of template DNA according to the manufacturer's recommendations (Promega Corporation, Madison, WI), except that the PCR conditions were denaturation at 95°C for 30 s, primer annealing at 67°C for 30 s, and primer extension at 70°C for 30 s, for a total of 28 cycles. A Perkin Elmer GeneAmp PCR System 9600 thermal cycler was used for the PCR. Three μL of loading dye (10 mM NaOH, 95% formamide, 0.05% bromophenol blue, and 0.05% xylene cyanol FF) were mixed with 3 μL of PCR product. The samples were denatured for 2 min in a Perkin Elmer Model 480 DNA thermal cycler and 5 μL were loaded onto the cathodal end of the gel. A discontinuous denaturing polyacrylamide gel was used to separate the STR amplicons. The polyacrylamide gels (6%T, 2.%C; piperazine diacrylamide as the cross-linker; 31 cm long and 0.4 mm thick) contained 7M urea and 60 mM Tris-formate (with respect to the formate ion), pH 9.0. The gel was permitted to polymerize for a minimum of 1 h at ambient temperature. The gel was placed in a SA 32 apparatus (GIBCO-BRL, Gaithersburg, MD) and the electrode buffer was 90 mM Tris-borate (with respect to the borate ion), pH 8.3. Electrophoresis was performed initially at 80 W for approximately 5 min and then continued at 25 W at ambient temperature. The run was allowed to continue until the xylene cyanol tracking dye migrated to the top of the lower reservoir buffer (approximately 3 h). The gels were stained with silver according to the method of Budowle et al. (11).

D1S80 Typing—The D1S80 locus was typed according to the method described by Budowle et al. (12).

Statistical Analysis—Allele designations were determined by comparison of the sample fragments with those of the allelic ladders. The frequency of each allele for each locus 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. (2). Possible divergence from Hardy-Weinberg expectations (HWE) was tested by calculating the unbiased estimate of the expected homozygote/heterozygote frequencies (13–16), the likelihood ratio test (2,14,17), and the exact test (18), based on 2000 shuffling experiments. An interclass correlation criterion (19) for two-locus associations was used for detecting disequilibrium between the STR loci.

TABLE 1—*HUMTHO1 allele frequencies in United States sample populations.*

Allele	African American* (N = 198)§	Caucasian† (N = 197)	Hispanic‡ (N = 206)
5	0.000	0.000	0.002
6	0.111	0.221	0.228
7	0.434	0.175	0.340
8	0.189	0.127	0.083
8.3	0.000	0.003	0.000
9	0.146	0.168	0.100
9.3	0.106	0.305	0.245
10	0.013	0.003	0.002

*Observed homozygosity = 0.298; expected homozygosity (unbiased) = 0.268; HWE − homozygosity test (p = 0.339), likelihood ratio test (p = 0.117), and exact test (p = 0.076).

†Observed homozygosity = 0.279; expected homozygosity (unbiased) = 0.214; HWE − homozygosity test (p = 0.027), likelihood ratio test (p = 0.451), and exact test (p = 0.367).

‡Observed homozygosity = 0.228; expected homozygosity (unbiased) = 0.243; HWE − homozygosity test (p = 0.631), liklihood ratio test (p = 0.932), and exact test (p = 0.940).

§N = number of individuals.

A 2 × N contingency table exact test was used to generate a G-statistic (1000 shuffling experiments) (20,21) 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

The distributions of observed allelic frequencies for HUMTHO1, TPOX, CSF1PO, and D1S80 are shown in Tables 1–4. All alleles differed in size by one repeat unit (i.e., 4 base pairs for the STR loci and 16 base pairs for the D1S80 locus) for all loci, except for the HUMTHO1 allele 9.3 (and the single observation of an 8.3 HUMTHO1 allele). The 9.3 and 8.3 alleles are 1 base pair smaller in size than the 10 and 9 alleles, respectively (22,23). The observed heterozygosities for the loci HUMTHO1, TPOX, CSF1PO, and D1S80 are 70.2%, 78.3%, 79.8%, and 87.6% for the African American sample, are 72.1%, 66.5%, 76.7%, and 80.2% for the Caucasian sample, and are 77.2%, 58.3%, 69.9%, and 83.3% for the Hispanic sample, respectively (Tables 1–3). The high heterozygosities substantiate the utility of these loci for forensic discrimination.

TABLE 2—*TPOX allele frequencies in United States sample populations.*

Allele	African American* (N = 198)§	Caucasian† (N = 197)	Hispanic‡ (N = 206)
6	0.086	0.000	0.005
7	0.023	0.000	0.005
8	0.369	0.546	0.556
9	0.179	0.124	0.034
10	0.091	0.036	0.034
11	0.230	0.256	0.269
12	0.023	0.038	0.095
13	0.000	0.000	0.002

*Observed homozygosity = 0.217; expected homozygosity (unbiased) = 0.236; HWE − homozygosity test (p = 0.541), likelihood ratio test (p = 0.506), and exact test (p = 0.632).

†Observed homozygosity = 0.335; expected homozygosity (unbiased) = 0.381; HWE − homozygosity test (p = 0.193), likelihood ratio test (p = 0.021), and exact test (p = 0.031).

‡Observed homozygosity = 0.418; expected homozygosity (unbiased) = 0.391; HWE − homozygosity test (p = 0.443). likelihood ratio test (p = 0.605), and exact test (p = 0.466).

§N = number of individuals.

TABLE 3—*CSF1PO allele frequencies in United States sample populations.*

Allele	African American* (N = 198)§	Caucasian† (N = 197)	Hispanic‡ (N = 206)
6	0.003	0.000	0.000
7	0.043	0.003	0.002
8	0.086	0.005	0.000
9	0.030	0.020	0.007
10	0.278	0.259	0.250
11	0.199	0.302	0.269
12	0.301	0.320	0.391
13	0.053	0.071	0.066
14	0.008	0.015	0.010
15	0.000	0.005	0.005

*Observed homozygosity = 0.202; expected homozygosity (unbiased) = 0.218; HWE − homozygosity test (p = 0.579), likelihood ratio test (p = 0.927), and exact test (p = 0.993).

†Observed homozygosity = 0.234; expected homozygosity (unbiased) = 0.264; HWE − homozygosity test (p = 0.325), likelihood ratio test (p = 0.680), and exact test (p = 0.570).

‡Observed homozygosity = 0.301; expected homozygosity (unbiased) = 0.291; HWE − homozygosity test (p = 0.742), likelihood ratio test (p = 0.099), and exact test (p = 0.201).

§N = number of individuals.

The TPOX locus in Caucasians is the only example of a deviation from HWE using the exact test (Tables 1–4). The departure from HWE was not highly significant ($p = 0.031$). One observed departure from HWE is within expectations. An inter-class correlation test analysis failed to detect any correlations between the alleles at any of the pair-wise comparisons of the 4 loci (Table 5).

Significant differences (by G-statistic) in allele frequencies between the African American, U.S. Caucasian, and Hispanic databases were observed for each of the 4 loci (data not shown) and might be expected. While the degree of similarity for D1S80 between subgroups within a major population group has been described previously (for example, see 24–27) such a comparison for the STR loci HUMTHO1, TPOX, and CSF1PO has not. There were only two significant differences in allele frequencies for the STR loci out of 17 comparisons between our U.S. population data and published within-group data (Table 6). At the HUMTHO1 locus the allele frequencies for U.S. Caucasians and Swedes were significantly different ($p = 0.025$), and at the TPOX locus allele frequencies U.S. Caucasians and Portuguese were different ($p = 0.007$). The two differences between the sample populations may be due to genetic differences or to small sample sizes. Regardless, the allelic frequencies are not substantially different. For example, at the HUMTHO1 locus, the allele frequencies in the Swedish and U.S. Caucasian samples, respectively, are: 0.007 and 0.000 (allele 5); 0.169 and 0.221 (allele 6); 0.199 and 0.175 (allele 7); 0.085 and 0.127 (allele 8); 0.174 and 0.168 (allele 9); 0.359 and 0.305 (allele 9.3); and 0.007 and 0.003 (allele 10). Also, the frequencies of the TPOX alleles in the Portuguese and U.S. Caucasians, respectively, are: 0.500 and 0.546 (allele 8); 0.070 and 0.124 (allele 9); 0.070 and 0.036 (allele 10); 0.311 and 0.256 (allele 11); and 0.040 and 0.038 (allele 12). Because of these similarities in allele

TABLE 4—*D1S80 allele frequencies in United States sample populations.*

Allele	African American* (N = 193)§	Caucasian† (N = 192)	Hispanic‡ (N = 203)
16	0.000	0.003	0.007
17	0.031	0.000	0.005
18	0.080	0.258	0.259
19	0.000	0.003	0.005
20	0.034	0.036	0.015
21	0.135	0.013	0.022
22	0.052	0.034	0.010
23	0.018	0.008	0.017
24	0.199	0.362	0.273
25	0.057	0.047	0.076
26	0.010	0.016	0.012
27	0.016	0.003	0.010
28	0.148	0.042	0.062
29	0.067	0.076	0.020
30	0.010	0.005	0.076
31	0.044	0.070	0.106
32	0.010	0.000	0.000
33	0.003	0.008	0.002
34	0.075	0.003	0.005
35	0.000	0.000	0.002
36	0.000	0.003	0.005
37	0.000	0.005	0.002
38	0.000	0.005	0.000
39	0.000	0.000	0.000
40	0.000	0.000	0.000
41	0.000	0.000	0.000
>41//	0.010	0.003	0.007

*Observed homozygosity = 0.124; expected homozygosity (unbiased) = 0.105; HWE – homozygosity test (p = 0.382), likelihood ratio test (p = 0.642), and exact test (p = 0.605).

†Observed homozygosity = 0.198; expected homozygosity (unbiased) = 0.213; HWE – homozygosity test (p = 0.606), likelihood ratio test (p = 0.827), and exact test (p = 0.885).

‡Observed homozygosity = 0.168; expected homozygosity (unbiased) = 0.168; HWE – homozygosity test (p = 0.978). likelihood ratio test (p = 0.570), and exact test (p = 0.479).

§N = number of individuals.

//All alleles migrating slower than the largest allele in the ladder (i.e., allele #41) are placed in the >41 allele class.

frequencies between the compared populations, there would be no anticipated substantial differences in DNA profile frequency estimates if another subgroup sample population were used as a reference database for the loci CSF1PO, TPOX, and HUMTHO1.

Since D1S80 population data are well characterized (12,24–27),

TABLE 5—*Two locus interclass correlation test (p-value)* for United States sample populations.*

Loci	African American (N = 198)‡	Caucasian (N = 197)	Hispanic (N = 206)
HUMTHO1/TPOX	0.853	0.796	0.401
HUMTHO1/CSF1PO	0.090	0.370	0.163
TPOX/CSF1PO	0.919	0.457	0.150
HUMTHO1/D1S8O	0.450	0.929†	0.738
TPOX/D1S8O	0.825	0.183	0.206
CSF1PO/D1S8O	0.228	0.810	0.934

*Two-sided probability values; based on 2000 shuffling experiments.

†The HUMTHO1 8.3 allele (observed in one individual) was merged with the 9 allele for this analysis.

‡N = number of individuals.

TABLE 6—*G-statistic test (p-values ± SD) for homogeneity on HUMTHO1, TPOX, and CSF1PO, allele distributions between Caucasian sample populations, between African American sample populations, and between Hispanic sample populations.*

Populations	HUMTHO1
U.S. Caucasian/Spanish Basque*	0.078 ± 0.009
U.S. Caucasian/Swedish†	0.025 ± 0.005
U.S. Caucasian/U.S. Caucasian‡	0.342 ± 0.015
U.S. Caucasian/Hungarian§	0.663 ± 0.015
U.S. Caucasian/Spanish//	0.454 ± 0.016
U.S. Caucasian/Swiss//	0.396 ± 0.016
African American/African American‡	0.275 ± 0.014
Hispanic/Hispanic‡	0.463 ± 0.016
	TPOX
U.S. Caucasian/Portuguese**	0.007 ± 0.003
U.S. Caucasian/Galicians**	0.670 ± 0.015
U.S. Caucasian/Spanish//	0.348 ± 0.015
U.S. Caucasian/Swiss//	0.084 ± 0.009
	CSF1PO
U.S. Caucasian/Spanish//	0.253 ± 0.014
U.S. Caucasian/Swiss//	0.760 ± 0.014
U.S. Caucasian/U.S. Caucasian††	0.836 ± 0.012
African American/African American††	0.291 ± 0.014
Hispanic/Hispanic††	0.219 ± 0.013

*Data from Alonso et al.
†Data from Holgersson et al.
‡Data from Puers et al.
§Data from Furedi et al.
//Data from Martin et al.
¶Data from Hochmeister et al.
**Data from Gusmao et al.
††Data from Hammond et al.

the U.S. allele frequency data presented in this study were compared with U.S. data on the D1S80 locus described previously by Budowle et al. (12). There were no statistical differences for the within group comparisons.

In conclusion, United States African American, Caucasian, and Hispanic databases have been established for HUMTHO1, TPOX, CSF1PO, and D1S80. The application of the product rule is appropriate for estimating the rarity of a multiple loci profile for these four loci. Generally, there are few differences between subgroups within a major population. Moreover, the STR multiplex methodology provides sufficient resolution to separate the HUMTHO1 9.3 and 10 alleles, is relatively simple, and can be implemented into most application-oriented laboratories at minimal cost.

This is publication number 96-10 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.

References

1. Edwards A, Civitello A, Hammond HA, Caskey CT. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 1991;49:746–56.
2. 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.
3. Polymeropoulos MH, Xiao H, Rath DS, Merril CR. Tetranucleotide repeat polymorphism at the human coagulation factor XIII A subunit gene (F13A1). Nucleic Acids Res 1991;19:4036.
4. Sullivan KM, Pope S, Gill P, Robertson JM. Automated DNA profiling by fluorescent labeling of PCR products. PCR Meth Appl 1992;2:34–40.
5. Polymeropoulos MH, Xiao H, Rath DS, Merril CR. Tetranucleotide repeat polymorphism at the human tyrosine hydrolase gene (TH). Nucleic Acids Res 1991;19:3753.

6. Anker R, Steinbreuck T, Donnis-Keller H. Tetranucleotide repeat polymorphism at the human thyroid peroxidase (hTPO) locus. Hum Mol Genet 1992;1:137.
7. Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R. Evaluation of 13 short tandem repeat loci for use in personal identification applications. Am J Hum Genet 1994;55:175–89.
8. Comey CT, Koons BW, Presley KW, Smerick JB, Sobieralski CA, Stanley DM, et al. DNA extraction strategies for amplified fragment length polymorphism analysis. J Forensic Sci 1994;39:1254–69.
9. Waye JS, Presley L, Budowle B, Shutler GG, Fourney RM. A simple method for quantifying human genomic DNA in forensic specimen extracts. BioTechniques 1989;7:852–5.
10. Budowle B, Baechtel FS, Comey CT, Giusti AM, Klevan L. Simple protocols for typing forensic biological evidence: chemiluminescent detection for human DNA quantitation and RFLP analyses and manual typing of PCR amplified polymorphisms. Electrophoresis 1995;16:1559–67.
11. Budowle B, Chakraborty R, Giusti AM, Eisenberg AJ, Allen RC. Analysis of the variable number of tandem repeats locus D1S80 by the polymerase chain reaction followed by high resolution polyacrylamide gel electrophoresis. Am J Hum Genet 1991;48:137–44.
12. 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;40:38–44.
13. Chakraborty R, Smouse PE, Neel JV. Population amalgamation and genetic variation: observations on artificially agglomerated tribal populations of Central and South America. Am J Hum Genet 1988;43:709–25.
14. 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, and Wolff R, editors, DNA fingerprinting: approaches and applications, Birkhauser Verlag, Berlin, 1991;127–43.
15. Nei M, Roychoudhury AK. Sampling variances of heterozygosity and genetic distance. Genetics 1974;76:379–90.
16. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 1978;89:583–90.
17. Weir BS. Independence of VNTR alleles defined by fixed bins. Genetics 1992;130:873–87.
18. Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 1992;48:361–72.
19. Karlin S, Cameron EC, Williams PT. Sibling and parent-offspring correlation estimation with variable family size. Proc Nat Acad Sci USA 1981;78:2664–8.
20. 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.
21. Lewontin RC, Felsenstein J. The robustness of homogeneity tests in 2 × N tables. Biometrics 1965;21:19–33.
22. Puers C, Hammond HA, Jin L, Caskey CT, Schumm JW. Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUMTHO1 $[AATG]_n$ and reassignment of alleles in population analysis by using a locus-specific allelic ladder. Am J Hum Genet 1993;53:953–8.
23. Hochmeister MN, Budowle B, Schumm JW, Sprecher CJ, Borer UV, Dirnhofer R. Swiss population data and forensic efficiency values on 3 tetrameric short tandem repeat loci—HUMTHO1, TPOX, and CSF1PO—derived using a STR multiplex system. Int J Leg Med 1995;107:246–9.
24. Kloosterman AD, Budowle B, Daselaar P. PCR-amplification and detection of the human D1S80 VNTR locus: amplification conditions, population genetics, and application in forensic analysis. Int J Leg Med 1993;105:257–64.
25. Huang NE, Chakraborty R, Budowle B. D1S80 allele frequencies in a Chinese population. Int J Leg Med 1994;107:118–20.
26. Scholl S, Budowle B, Radecki K, Salvo M. Navajo, Pueblo, and Sioux population data on the loci HLA-DQA1, LDLR, GYPA, HBGG, D7S8, Gc, and D1S80. J Forensic Sci 1996;41:47–51.
27. Woo KM, Budowle B. Korean population data on the PCR-based loci LDLR, GYPA, HBGG, D7S8, Gc, HLA-DQA1, and D1S80. J Forensic Sci 1995;40:645–8.
28. Alonso S, Castro A, Fernandez I, Gomez de Cedron M, Garcia-Orad A, Meyer E, et al. Population study of 3 STR loci in the Basque Country (northern Spain). Int J Leg Med 1995;107:239–45.
29. Holgersson S, Karlsson JA, Kihlgren A, Rosen B, Savolainen P, Gyllensten U. Fluorescent-based typing of the two short tandem repeat loci HUMTHO1 and HUMACTBP2: reproducibility of size measurements and genetic variation in the Swedish population. Electrophoresis 1994;15:890–5.
30. Furedi S, Woller J, Padar Z. Hungarian population data for the STR systems THO1 and VWA. Int J Leg Med 1995;108:48–9.
31. Martin P, Alonso A, Budowle B, Albarran C, Garcia O, Sancho M. Spanish population data on 7 tetrameric short tandem repeat loci. Int J Leg Med 1995;108:145–9.
32. Gusmao L, Prata MJ, Amorim A. The STR system hTPO: population and segregation data. Int J Leg Med 1995:108:167–9.

Additional information and reprint requests:
Bruce Budowle, Ph.D.
Forensic Science RTC
F.B.I. Academy
Quantico VA 22135

Erratum

The paper Budowle B, Smerick JB, Keys KM, Moretti TR. United States population data on the multiplex short tandem repeat loci—HUMTHO1, TPOX, and CSF1PO—and the variable number tandem repeat locus D1S80. J Forensic Sci 1997 Sep;42(5):846–49 contained an error in Table 3 on page 847 in the column headed "African American." The correct information is shown below.

Future citations of the above-referenced paper should read: Budowle B, Smerick JB, Keys KM, Moretti TR. United States population data on the multiplex short tandem repeat loci—HUMTHO1, TPOX and CSF1PO—and the variable number tandem repeat locus D1S80 [published erratum appears in J Forensic Sci 1998 Jan;43(1)] J Forensic Sci 1997 Sep;42(5):846–49.

Allele	African American[a] (N = 197)
6	0.000
7	0.043
8	0.086
9	0.030
10	0.279
11	0.198
12	0.302
13	0.053
14	0.008
15	0.000

[a]Observed Homozygosity = 0.203; Expected Homozygosity (unbiased) = 0.220; HWE—Homozygosity Test (p = 0.577), Likelihood Ratio Test (p = 0.927), and Exact Test (p = 0.993).