

## Fixed Bin Frequency Distributions for the VNTR Locus D5S110 in General United States Reference Databases

**REFERENCE:** Budowle, B. and Giusti, A. M., "Fixed Bin Frequency Distributions for the VNTR Locus D5S110 in General United States Reference Databases," *Journal of Forensic Sciences*, JFSCA, Vol. 40, No. 2, March 1995, pp. 236–238.

**ABSTRACT:** Fixed bin frequencies for the D5S110 locus were determined in African Americans, Caucasians, Southeastern Hispanics, and Southwestern Hispanics. The data were generated by RFLP analysis of HAE III-digested genomic DNA. The D5S110 locus met Hardy-Weinberg expectations in the four sample populations, and there is no evidence for association of alleles between the D5S110 and other routinely used VNTR loci. The frequency data can be used in forensic analyses and paternity tests to estimate the frequency of a DNA profile in various general United States populations.

**KEYWORDS:** forensic science, VNTR, allele frequency, population databases, D5S110, equilibrium, RFLP

Restriction fragment length polymorphism (RFLP) typing of variable number of tandem repeat (VNTR) loci [1–4] provides a high degree of discrimination and can be accomplished, at times, with less than 50 ng of high molecular weight genomic DNA. The molecular biology aspects of the RFLP technique are well-defined and the method is a robust forensic tool for characterizing biological evidence [5,6]. By using a panel of VNTR probes, sufficient data can be obtained to produce a composite profile that is unique to an individual (excluding monozygotic twins). More importantly, typing VNTR loci currently provides the forensic scientist the best avenue to exclude a suspect who has been falsely associated with an evidentiary sample.

One factor that can augment the effectiveness of RFLP analysis is the availability of additional well-characterized VNTR loci. RFLP analysis can be made more effective by identifying other VNTR loci, compatible with the restriction enzyme utilized for RFLP analysis (for the FBI laboratory that would be HAE III [7]), whose alleles generally fall in a size range that is greater than 1000 base pairs and less than 10,000 base pairs, are highly polymorphic, and have a high degree of sensitivity of detection. The VNTR locus D5S110 [8] meets these criteria. This short communication

Received for publication 15 April 1994; accepted for publication 11 July 1994.

<sup>1</sup>Program Manager for DNA Research, and Chemist, respectively, Forensic Science Research and Training Center, FBI Academy, Quantico, VA.

This is publication number 94-09 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.

provides African American, Caucasian, Southeastern Hispanic and Southwestern Hispanic population data on the D5S110 locus.

### Materials and Methods

The source of the DNA samples was described previously [9]. The DNA was extracted and typed by RFLP analysis according to the method of Budowle and Baechtel [5]. The probe LH1 (for D5S110) was kindly provided by L. Klevan (Life Technologies, GibcoBRL, Gaithersburg, MD). Size measurements of RFLP bands were made using an interactive image analysis system [10]. The base pair size data were binned according to the method of Budowle et al. [11].

Possible divergence from Hardy-Weinberg expectations (HWE) was determined by the likelihood ratio test [12–14], and the exact test [15]. An interclass correlation criterion [16] was used for detecting disequilibrium between the D5S110 locus and other previously typed VNTR loci—D1S7, D2S44, D4S139, D10S28, and D17S79 [9]. Independence across the six VNTR loci also was determined by examining whether or not the observed variance of the number of heterozygous loci in the population sample is outside its confidence interval under the assumption of independence [17,18].

### Results and Discussion

The 31 fixed bin frequency distributions for the D5S110 locus for African Americans, Caucasians, Southeastern Hispanics and Southwestern Hispanics are shown in Table 1. The D5S110 locus is highly polymorphic in all sample populations studied, with the African American population sample demonstrating the highest degree of variation. In all the sample populations, most alleles are greater than 1000 base pairs and less than 10,000 base pairs in length. There is no evidence for departure from HWE for the D5S110 locus in the four population groups based on the likelihood ratio test [12–14] and the exact test [15] (Table 1). Furthermore, the probe for D5S110 provides a sensitivity of detection similar to pH30/D4S139 (that is, 10 ng of human genomic DNA may be typeable) (Fig. 1).

The population samples were typed previously for other VNTR loci [9]. Therefore, analyses were performed to determine whether there were any detectable deviations from independence between D5S110 and D1S7, D2S44, D4S139, D10S28, or D17S79 loci. An interclass correlation test [16] analysis demonstrated that there is no evidence for correlation between the alleles at any of the pairs of loci (Table 2).

An alternate method that assesses all six VNTR loci at one time was used for testing for detectable deviation from expectation

TABLE 1—D5S110 bin frequencies.

BIN (bps)	African American <sup>a</sup> (N = 353)	Caucasian <sup>b</sup> (N = 511)	Southeastern Hispanic <sup>c</sup> (N = 286)	Southwestern Hispanic <sup>d</sup> (N = 165)
0-639	.000	.000	.000	.000
640-772	.000	.000	.002	.000
773-871	.001	.001	.004	.000
872-963	.000	.001	.002	.003
964-1077	.006	.001	.003	.000
1078-1196	.006	.005	.007	.012
1197-1352	.010	.006	.007	.006
1353-1507	.013	.015	.009	.018
1508-1637	.026	.027	.021	.021
1638-1788	.024	.027	.056	.024
1789-1924	.035	.031	.038	.018
1925-2088	.017	.031	.030	.039
2089-2351	.040	.112	.086	.067
2352-2522	.035	.037	.058	.039
2523-2692	.047	.077	.077	.045
2693-2862	.042	.060	.075	.045
2863-3033	.028	.091	.052	.076
3034-3329	.076	.089	.093	.073
3330-3674	.061	.117	.072	.109
3675-3979	.055	.066	.066	.064
3980-4323	.084	.069	.072	.085
4324-4821	.044	.039	.038	.048
4822-5219	.058	.018	.026	.052
5220-5685	.052	.027	.033	.015
5686-6368	.076	.025	.037	.045
6369-7241	.040	.005	.016	.024
7242-8452	.057	.018	.012	.045
8453-10093	.037	.004	.009	.021
10094-11368	.011	.000	.000	.003
11369-12829	.007	.000	.000	.000
12830-	.011	.000	.000	.000

<sup>a</sup>There were four individuals carrying three-banded patterns that are not included in this table. The number of individuals carrying a single band pattern = 6. HWE—Likelihood Ratio Test ( $P = 0.756$ ) and Exact Test ( $P = 0.850$ ).

<sup>b</sup>There was one individual carrying a three-banded pattern that is not included in this table. The number of individuals carrying a single band pattern = 26. HWE—Likelihood Ratio Test ( $P = 0.449$ ) and Exact Test ( $P = 0.554$ ).

<sup>c</sup>There was one individual carrying a three-banded pattern that is not included in this table. The number of individuals carrying a single band pattern = 20. HWE—Likelihood Ratio Test ( $P = 0.604$ ) and Exact Test ( $P = 0.475$ ).

<sup>d</sup>There were no individuals observed carrying a three-banded pattern. The number of individuals carrying a single band pattern = 10. HWE—Likelihood Ratio Test ( $P = 0.239$ ) and Exact Test ( $P = 0.163$ ).

when applying the product rule to derive a multiple loci frequency estimate. The test examines whether or not the observed variance ( $s_k^2$ ) of the number of heterozygous loci in a population sample is outside its confidence interval under the assumption of independence using the procedure described by Brown et al. [17]. While this test criterion results in some loss of information in summarizing multiple locus genotype data, it is more powerful than the classical goodness-of-fit  $\chi^2$  test [18]. There was no evidence of association for the six loci described in our African American ( $N = 241$ ), Caucasian ( $N = 348$ ), Southeastern Hispanic ( $N = 172$ ), and Southwestern Hispanic ( $N = 110$ ) sample populations using the  $s_k^2$  criterion ( $s_k^2_{\text{African American}} = 0.524$ , 95% confidence interval of variance is 0.457–0.689;  $s_k^2_{\text{Caucasian}} = 0.622$ , 95% confidence interval of variance is 0.511–0.713;  $s_k^2_{\text{Southeastern Hispanic}} = 0.556$ , 95% confidence interval of variance is 0.395–0.650,  $s_k^2_{\text{Southwestern Hispanic}}$  of variance is 0.363–0.686).

In conclusion, this report provides D5S110 point estimate fre-

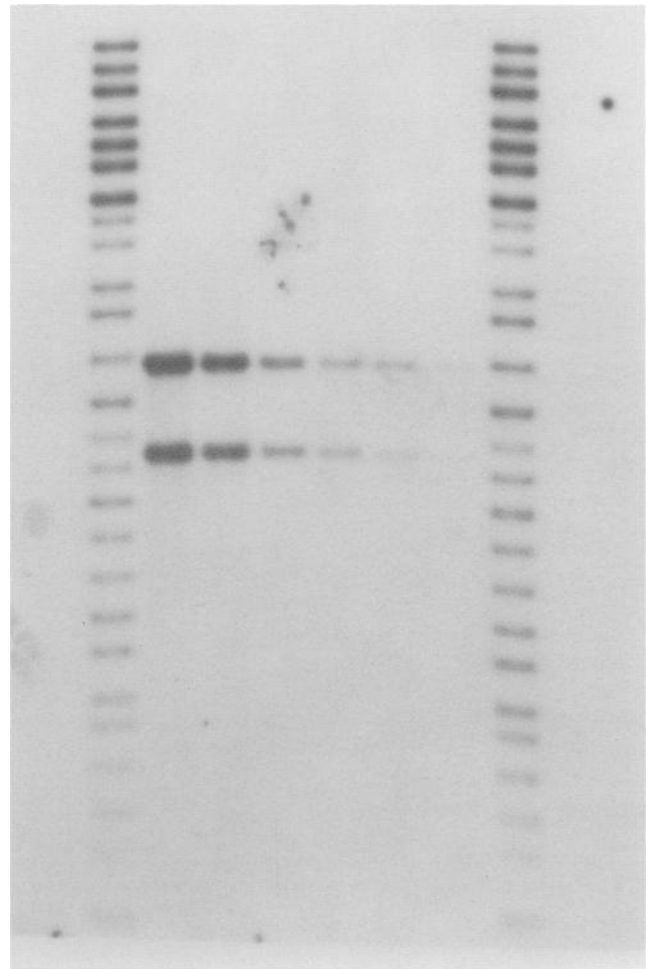


FIG. 1—D5S110 RFLP dilution series with K562 cell line DNA. The quantities of DNA from left-to-right are: 400 ng, 250 ng, 100 ng, 50 ng, 25 ng, and 10 ng.

TABLE 2—Two locus inter-class correlation test (two-sided probability values) for pairs of VNTR loci. No evidence for correlations.

Loci Pair	African American	Caucasian	Southeast Hispanic	Southwest Hispanic
D5S110/D1S7	0.402	0.665	0.421	0.986
D5S110/D2S44	0.929	0.930	0.360	0.985
D5S110/D4S139	0.743	0.912	0.576	0.460
D5S110/D10S28	0.067	0.357	0.591	0.793
D5S110/D17S79	0.091	0.494	0.477	0.782

quency data in a fixed bin format for African Americans, Caucasians, Southeastern Hispanics, and Southwestern Hispanics. The D5S110 locus is highly polymorphic in these four general population groups. The data demonstrate that it is appropriate to apply the product rule for estimating D5S110 single locus frequencies, as well as multiple locus frequencies.

## References

- [1] Wyman, A. R. and White, R., "A Highly Polymorphic Locus in Human DNA," *Proceedings of the National Academy of Sciences USA*, Vol. 77, 1980, pp. 6754-6758.
- [2] Jeffreys, A. J., Wilson, V., and Thein, S. L., "Individual-Specific Fingerprints of Human DNA," *Nature*, Vol. 316, 1985, pp. 76-79.
- [3] Jeffreys, A. J., Royle, N. J., Wilson, V., and Wong, Z., "Spontaneous Mutation Rates to New Length Alleles at Tandem-Repetitive Hypervariable Loci in Human DNA," *Nature*, Vol. 323, 1988, pp. 278-281.
- [4] Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., and White, R., "Variable Number of Tandem Repeat Markers for Human Gene Mapping," *Science*, Vol. 235, 1987, pp. 1616-1622.
- [5] Budowle, B. and Baechtel, F. S., "Modifications to Improve the Effectiveness of Restriction Fragment Length Polymorphism Typing," *Applied and Theoretical Electrophoresis*, Vol. 1, 1990, pp. 181-187.
- [6] Adams, D. E., Presley, L. A., Baumstark, A. L., Hensley, K. W., Hill, A. L., Anoe, K. S., Campbell, P. A., McLaughlin, C. M., Budowle, B., Giusti, A. M., Smerick, J. B., and Baechtel, F. S., "DNA Analysis by Restriction Fragment Length Polymorphisms of Blood and Other Body Fluid Stains Subjected to Contamination and Environmental Insults," *Journal of Forensic Sciences*, Vol. 36, 1991, pp. 1284-1298.
- [7] Budowle, B., Waye, J. S., Shutler, G. G., and Baechtel, F. S., "Hae III—A Suitable Restriction Endonuclease for Restriction Fragment Length Polymorphism Analysis of Biological Evidence Samples," *Journal of Forensic Sciences*, Vol. 35, 1990, pp. 530-536.
- [8] Amour, J. A., Povey, S., Jeremiah, S., and Jeffreys, A. J., "Systematic Cloning of Human Minisatellites from Ordered Charomid Libraries," *Genomics*, Vol. 8, 1990, pp. 501-512.
- [9] Budowle, B., Monson, K. L., Anoe, K., Baechtel, F. S., Bergman, D., et al., "A Preliminary Report on Binned General Population Data on Six VNTR Loci in Caucasians, Blacks, and Hispanics from the United States," *Crime Laboratory Digest*, Vol. 18, No. 1, 1991, pp. 9-26.
- [10] Monson, K. L. and Budowle, B., "A System for Semi-Automated Analysis of DNA Autoradiograms," In, *Proceedings of the International Symposium on the Forensic Aspects of DNA Analysis*, Government Printing Office, Washington, D.C., 1989, pp. 127-132.
- [11] Budowle, B., Giusti, A. M., Waye, J. S., Baechtel, F. S., Fourney, R. M., Adams, D. E., Presley, L. A., Deadman, H. A., and Monson, K. L., "Fixed-Bin Analysis for Statistical Evaluation of Continuous Distributions of Allelic Data from VNTR Loci, for Use in Forensic Comparisons," *American Journal of Human Genetics*, Vol. 48, 1991, pp. 841-855.
- [12] Edwards, A., Hammond, H., Jin, L., Caskey, C. T., and Chakraborty, R., "Genetic Variation at Five Trimeric and Tetrameric Repeat Loci in Four Human Population Groups," *Genomics*, Vol. 12, 1992, pp. 241-253.
- [13] Chakraborty, R., Fornage, M., Guegue, R., and Boerwinkle, E., "Population Genetics of Hypervariable Loci: Analysis of PCR Based VNTR Polymorphism Within a Population," In, *DNA Fingerprinting: Approaches and Applications*, T. Burke, G. Dolf, A. J. Jeffreys, and R. Wolff, Eds., Birkhauser Verlag, Berlin, 1991, pp. 127-143.
- [14] Weir, B. S., "Independence of VNTR Alleles Defined by Fixed Bins," *Genetics*, Vol. 130, 1992, pp. 873-887.
- [15] Guo, S. W. and Thompson, E. A., "Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles," *Biometrics*, Vol. 48, 1992, pp. 361-372.
- [16] Karlin, S., Cameron, E. C., and Williams, P. T., "Sibling and Parent-Offspring Correlation Estimation with Variable Family Size," *Proceedings of the National Academy of Sciences U.S.A.*, Vol. 78, 1981, pp. 2664-2668.
- [17] Brown, A. H. D., Feldman, M. W., and Nevo, E., "Multilocus Structure of Natural Populations of *Hordeum Spontaneum*," *Genetics*, Vol. 96, 1980, pp. 523-536.
- [18] Chakraborty, R., "Detection of Nonrandom Association of Alleles from the Distribution of the Number of Heterozygous Loci in a Sample," *Genetics*, Vol. 108, 1984, pp. 719-731.

Address requests for reprints or additional information to  
 Bruce Budowle, Ph.D.  
 FSRTC, FBI Academy  
 Quantico, VA 22135