



# Population studies on 17 STR loci routinely used in forensic analyses

Bruce Budowle\*

*Laboratory Division, FBI, 935 Pennsylvania Ave., NW, Washington, DC 20535, USA*

---

## Abstract

STR population data are available for 17 loci: CSF1PO, FGA, TH01, TPOX, vWA, D2S1338, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D19S433, D21S11, Penta D, and Penta E. Thus, estimates of a multiple locus profile frequency can be made for most forensic scenarios. For those situations where a specific population is not available or for correcting for population substructure, the inbreeding coefficient  $\theta$  is recommended. The extant population data support that a  $\theta$  value of 0.01 is conservative for most population groups and 0.03 is conservative for genetically isolated populations.

© 2003 Elsevier Science B.V. All rights reserved.

*Keywords:* STR loci; Population data; Inbreeding;  $\theta$

---

## 1. Introduction

The analysis of short tandem repeat loci has become a routine procedure in forensic laboratories. In addition, because of the establishment of national DNA databases in many countries and the production of commercial kits, substantial population data exist to assess the rarity of a DNA profile. In fact, there are 17 STR loci that are well defined and many of these loci have been typed in populations on six continents. These loci are CSF1PO, FGA, TH01, TPOX, vWA, D2S1338, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D19S433, D21S11, Penta D, and Penta E. These loci have been found to be highly polymorphic in all major population groups, and population data are available.

---

\* Tel.: +1-202-324-9512; fax: +1-202-324-1462.

*E-mail address:* bbudowle@fbi.gov (B. Budowle).

## 2. Discussion

For calculating the rarity of a DNA profile, the National Research Council (NRC) II Report [1] recommended the use of Wright's  $F_{ST}$  statistic (or  $\theta$ ) [2,3]. This parameter estimates the degree of relatedness in a population and can be used to correct for departures from Hardy–Weinberg expectations that are due to population substructure. The NRC II Report recommended that a conservative value for  $\theta$  is 0.01, but 0.03 could be applied when making profile frequency estimates for more genetically isolated populations, such as Native Americans.

With extant population data, one can determine if the recommended values for  $\theta$  are indeed conservative. The approach described by Weir and Cockerham [4] was used to estimate  $\theta$ . Budowle et al. [5] determined  $\theta$  estimates over 13 of the STR loci to be 0.0006 for African Americans (11 population samples),  $-0.0005$  for US Caucasians (9 population samples), 0.0021 for Hispanics (8 population samples), and 0.0039 for Asians (6 population samples). Eleven European populations (i.e., Czech, Finn, Greek Cypriot, Italian (two samples), Slovene, Spanish (two samples), Swiss, Turk, US Caucasian) were analyzed, and the estimate over 13 STR loci was 0.0028 [6]. For Omani and Dubai Arabs typed for nine STR loci, the estimated value for  $\theta$  was 0.0010 (manuscript in preparation). Even across such a large geographic area as Latin American, the value for  $\theta$  is 0.0053 for 15 STR loci in five population groups (Argentina, Brazil, Chile, Costa Rica, and Uruguay) (manuscript in preparation). These values are much less than the advocated conservative estimate of 0.01.

For more isolated groups, Budowle et al. [7] reported a  $\theta$  estimate of 0.0090 for nine STR loci in Chamorros and Filipinos. For Native Americans (seven sample populations), a larger  $\theta$  estimate of 0.0282 was observed [5]. These larger values for  $\theta$  are consistent with expectations for more isolated groups. Recently, three Native Alaskan population samples (Athabaskan, Inupiat, and Yupik) were analyzed for 13 STR loci [8]. All sample populations

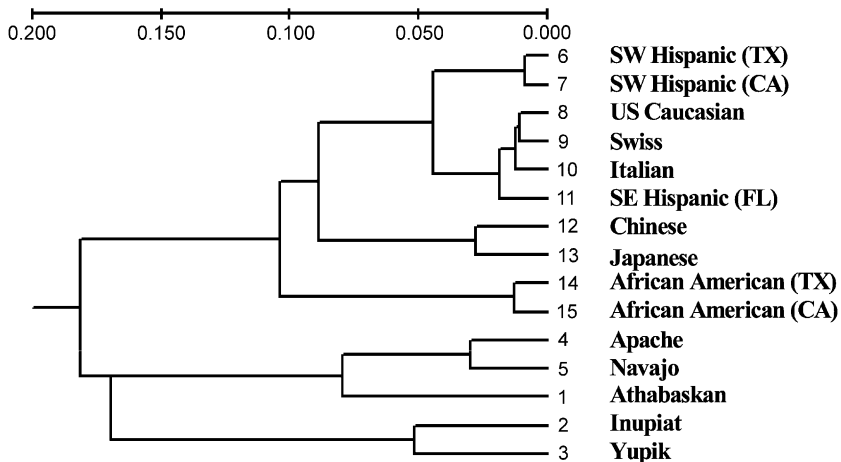


Fig. 1. UPGMA tree for subpopulation affinities based on 13 CODIS STR loci.

were highly polymorphic. The estimate for  $\theta$  over 13 STR loci was 0.0309 for the three Native Alaskan groups. Thus, a  $\theta$  value of 0.03 likely is a reasonable value to use when estimating a DNA profile frequency for Native Americans in Alaska.

However, distance measures (represented as a UPGMA tree using Nei's [9,10] unbiased estimate and 1000 permutations) can provide clarification on whether or not a  $\theta$  estimate of 0.03 is conservative for such isolated groups as Native Alaskans (Fig. 1). The data from 15 population groups, including African American, Asians, Europeans, Hispanics, and Native Americans, are consistent with the ethnohistory of the populations. Moreover, Yupiks and Inupiat are more closely related to each other than to the Athabaskans. In fact, the Athabaskans are closer to the Apaches and Navajos. This observation is not surprising since Athabaskans, Apaches and Navajos belong to the same linguistic group [11]. An estimate for the populations that cluster together is more indicative of whether or not a value of 0.03 is conservative. For Inupiat and Yupiks,  $\theta$  over 13 STR loci was 0.0167, and for the Athabaskans, Apaches and Navajos the value was 0.0180. These values support that the recommended  $\theta$  of 0.03 is conservative.

### 3. Conclusion

In conclusion, these results support that the effects of inbreeding on the estimates of profiles frequencies are negligible for most population groups. If  $\theta$  is used, a value of 0.01 is very conservative. When more isolated groups are considered a value for  $\theta$  of 0.03 is more than sufficient.

### Acknowledgements

This is publication number 01-23 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. Special thanks to the members of the Alaska State Scientific Crime Detection Laboratory of the Department of Public Safety, Jose Lorente and GITAD, and Abdul Alkhatay and Farida Shamali for their contributions of population data.

### References

- [1] National Research Council II Report, The Evaluation of Forensic Evidence, National Academy Press, Washington, DC, 1996.
- [2] S. Wright, Coefficients of inbreeding and relationship, *Am. Nat.* 56 (1922) 330–338.
- [3] S. Wright, The interpretation of population structure by  $F$ -statistics with special regard to systems of mating, *Evolution* 19 (1965) 395–420.
- [4] B.S. Weir, C.C. Cockerham, Estimating  $F$ -statistics for the analysis of population structure, *Evolution* 38 (1984) 1358–1370.
- [5] B. Budowle, B. Shea, S. Niezgodna, C. Chakraborty, CODIS STR loci data from 41 sample populations, *J. Forensic Sci.* 46 (2001) 453–489.
- [6] B. Budowle, R. Chakraborty, Population variation at the CODIS core short tandem repeat loci in Europeans, *Leg. Med.* 3 (2001) 29–33.

- [7] B. Budowle, D.A. Defenbaugh, K.M. Keys, Genetic variation at nine short tandem repeat loci in Chamorros and Filipinos from Guam, *Leg. Med.* 2 (1) (2000) 26–30.
- [8] B. Budowle, Population genetic studies on forensically-employed short tandem repeat loci, Presented at the American Academy of Forensic Sciences Meeting, Seattle, Washington, 2001.
- [9] M. Nei, Genetic distance between populations, *Am. Nat.* 106 (1972) 283–292.
- [10] M. Nei, Estimation of average heterozygosity and genetic distance from a small number of individuals, *Genetics* 89 (1978) 583–590.
- [11] J.G. Lorenz, D.G. Smith, Distribution of four founding mtDNA haplogroups among native North Americans, *Am. J. Phys. Anthropol.* 101 (1996) 307–323.