

## Brief Communication

### ***Gm 3;5,13,14* and Stated-Admixture: Independent Estimates of Admixture in American Indians**

ROBERT C. WILLIAMS,<sup>1,2</sup> ARTHUR G. STEINBERG,<sup>3</sup> WILLIAM C. KNOWLER,<sup>4</sup>  
AND DAVID J. PETTITT<sup>4</sup>

#### SUMMARY

Bernstein's formula for the estimation of the amount of admixture ( $m$ ) in a hybrid population has been used frequently since its publication in 1931. While mathematically correct, it has not been shown to be correct in practice, because an independent estimate from a large sample has not been available. We have compared the estimate of  $m$  for Caucasian admixture derived by using Bernstein's formula with that estimated from stated-admixture ( $s_a$ ) within a sample of 5,759 Native Americans. There was a linear relationship between the two variables ( $m = -.000275 + .714s_a$ ;  $r = .976$  for the grouped data,  $P = .0001$ ).

#### INTRODUCTION

Admixture in Native Americans has multiple sources. It derives from, among others, European (white), Mexican-American, and black populations. Bernstein developed a method in 1931 for measuring admixture from a specific population that utilized a marker allele [1-3]. However, there was no independent admixture measurement available for comparison. It is shown below in a sample of 5,759 Native Americans that stated-admixture, an estimate from personal testimony and pedigrees, is an independent index that is highly cor-

---

Received February 11, 1986; revised April 21, 1986.

<sup>1</sup> Department of Anthropology, Arizona State University, Tempe, AZ 85281.

<sup>2</sup> Histocompatibility Laboratory, Blood Systems Incorporated, 6220 East Oak Street, Scottsdale, AZ 85257.

<sup>3</sup> Department of Biology, Case Western Reserve University, Cleveland, OH 44106.

<sup>4</sup> National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases, 1550 East Indian School Road, Phoenix, AZ 85014.

© 1986 by the American Society of Human Genetics. All rights reserved. 0002-9297/86/3903-0014\$02.00

related with Bernstein's estimate of Caucasian admixture using the haplotype *Gm 3;5,13,14* as a marker.

Traditional analyses of admixture have dealt with three populations, two from which the genes were derived, A and B, and the third hybrid population, H, which was solely derived from them [1].

Let  $P(H)$ ,  $P(A)$ , and  $P(B)$  be the allele frequencies of a given genetic marker in the three populations, and let  $m$  be the fraction of alleles contributed by population A and  $(1 - m)$  that contributed by B with  $m$  having the range  $0.0 \leq m \leq 1.0$ . Then:  $P(H) = mP(A) + (1 - m)P(B)$ . When the allele frequencies are known or estimated,  $m$  becomes

$$m = [P(H) - P(B)] / [P(A) - P(B)] . \quad (1)$$

The estimate of  $m$  will improve as the value of  $[P(A) - P(B)]$  increases and is best when the marker is always present in one population and absent in the other.

Let population A be Americans of European descent and population B be precontact Native Americans. The hybrid population is contemporary American Indians whose gene pool is a combination of native and European alleles. As stated above, an ideal genetic marker for measuring admixture between European Caucasians and Native Americans would be one that has a very high frequency in Europeans but which was absent from precontact Native Americans. It would be expected that, as the hybrid population became "less Indian," the frequency of the European allele would become higher. Correspondingly, "full-blooded Indians" should have none of the genetic marker for admixture.

Other than its application to a single pedigree of South American Indians [4], no systematic test of the accuracy of Bernstein's equation when applied to populations has, to our knowledge, been published. This is due to the fact that no independent estimate of admixture has been available. In theory, such an estimate and that derived from equation (1) should be identical. However, errors in the estimates of the fraction of Indian ancestry and in the estimates of the frequency of the "European alleles" in the hybrid population will occur. The former set of errors will occur because some people will report their ancestry incorrectly; the latter will occur because of sampling errors of the estimate. These errors, which are independent, will tend to reduce the value of the correlation coefficient. In brief, as the fraction of non-Indian ancestry increases, the fraction of the European-derived allele (haplotype) should increase and the correlation between the two estimates of  $m$  over the range of admixture should be close to one.

#### METHODS

A unique set of data for demonstrating the utility of just such a marker has been collected by the Southwestern Field Studies Section of the National Institutes of Arthritis, Diabetes, and Digestive and Kidney Diseases (NIADDK) in Phoenix, Arizona, over the past 20 years. Thousands of Native Americans, primarily from the Pima and

Papago tribes, have been typed for Gm allotypes by standard methods [5, 6]. The Gm haplotype *Gm 3;5,13,14* has a high frequency in European-derived populations, .665, and a low, variable frequency in samples of Native Americans, .010–.050 [7].

In addition, NIADDK has developed its own admixture index that is determined without reference to the person's genetic phenotype. When NIADDK began this long-range study in the Gila River Indian Community in 1965 [8], each person who was seen in the clinic was asked about his or her pedigree and the amount of admixture from other tribes and non-Indian populations. This, as well as information from individuals who were familiar with the members of the tribes, established a basic admixture number for each member of the community who was interviewed. Since that time, these numbers have been used with pedigree data to establish the admixture index of subsequent generations.

The stated-admixture is determined in increments of one-eighth. For instance, a person with one parent who is a "full-blooded" Indian (0/8 admixture) and one who is a Caucasian (8/8 admixture) would be classified as 4/8. For the present report, all participants in the NIADDK study who have been typed for the Gm allotypes are included. These are primarily from the Pima and Papago tribes although members of other tribes are also present. The tribal affiliation is of no importance to this analysis because the *Gm 3;5,13,14* haplotype has a very low frequency in all Native American populations that have been typed [7]. Only non-Indian admixture will be dealt with here.

#### RESULTS

Table 1 presents the stated-admixture, the haplotype frequency of *Gm 3;5,13,14*, the sample size, and  $m$  as measured by this haplotype. European-derived, Caucasian populations have a pooled frequency of *Gm 3;5,13,14* of .665 [9–11]. The frequency of this haplotype in precontact Native Americans is taken as .015 (although it is probably zero as evidenced by other data [7]), the allele frequency in the 0/8 stated-admixture category. Therefore  $m$  becomes  $m = [P(H) - .015]/(.665 - .015) = [P(H) - .015]/.650$ , where  $P(H)$  is the frequency of the haplotype at a given level of stated-admixture. Note that among the 22 subjects in the group with stated-admixture 1.0, the frequency of *Gm 3;5,13,14* is .545, resulting in  $m = .815$ . These people state their heritage as

TABLE 1  
STATED ADMIXTURE AND ESTIMATED CAUCASIAN ADMIXTURE ( $m$ )  
IN 5,759 NATIVE AMERICANS USING THE *Gm 3;5,13,14* HAPLOTYPE  
FOR BERNSTEIN'S ESTIMATE OF  $m$

Stated-Admixture	Frequency <i>Gm3;5,13,14</i> *	No.	$m$ †
1.000 .....	.545	22	.815
.750 .....	.540	25	.808
.625 .....	.307	44	.449
.500 .....	.218	344	.312
.375 .....	.063	16	.074
.250 .....	.145	138	.200
.125 .....	.089	192	.114
0 .....	.015	4,978	.000

\* Frequency of the haplotype *Gm 3;5,13,14*.

†  $m$ , estimate of admixture based on Bernstein's formula using the haplotype.

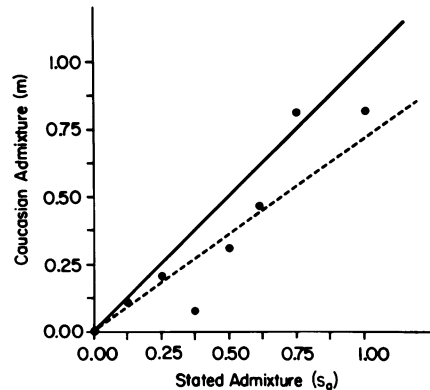


FIG. 1.—A regression of two estimates for admixture. The stated-admixture variable ( $s_a$ ) was taken from personal testimony and pedigree data and is measured in eighths. The variable  $m$  is an estimate of Caucasian admixture based on Bernstein's formula using the *Gm 3;5,13,14* haplotype. The fitted line (dashed), using a weighted procedure, is  $m = -.000275 + .714s_a$  ( $P < .0001$ ), while the correlation coefficient for the grouped data is .976. The solid line is the line of identity.

a mixture of "white," "Spanish," or "Mexican," and may have an unrecognized degree of Native American ancestry. Thus, they may not be entirely of Caucasian heritage.

A line was fitted to a graph of ( $x,y$ ) pairs where the variable  $x$  is the stated-admixture with a range from 0/8(0.000) to 8/8(1.000) and  $y$  is the value of  $m$  from 0.000 to .815 (table 1, fig. 1). There is a strong linear association between the two variables ( $P < .0001$ ). Inclusion of quadratic or higher-order polynomial terms did not significantly improve the fit of the regression equation. The regression—estimated by the least squares method using the sample sizes as weights—of European admixture ( $m$ ) on stated-admixture ( $s_a$ ), as measured by the haplotype *Gm 3;5,13,14*, is  $m = -.000275 + .714s_a$ . The correlation coefficient for the grouped data is .976.

#### DISCUSSION

There are at least two sources of error contributing to the regression: (1) the estimate of the frequency of *Gm 3;5,13,14* haplotype among Caucasians has a variance of estimate and in addition may not be precise for those who contributed the haplotype to the Native Americans; and (2) the estimate of admixture based on the pedigree data is not without error. For instance, the *Gm 3;5,13,14* haplotype occurs with a frequency of 1.5% among the individuals with no stated-admixture. Recent data suggest that this haplotype was absent in pre-contact Native Americans [7]. Finally, Bernstein's formula underestimates the total amount of admixture in this sample of American Indians. One explanation for this is that the *Gm 3;5,13,14* haplotype is a marker of European admixture while that represented in the stated-admixture variable has small components from other populations such as Mexicans and blacks. A second reason might be that the stated-admixture category, 1.000, is small: no. = 22. It would be better if there were a very large number of individuals who said that they are non-

Indian, but who were drawn from the same population as the other categories. Practically, this is not possible. Totally admixed individuals, non-Indians, do not live in large numbers in the Gila River Indian Community.

The estimate of admixture based on the haplotype frequency is congruent with that derived from personal testimony and pedigrees and confirms the utility of Bernstein's formula when used on population data.

#### REFERENCES

1. BERNSTEIN F: *Comitato Italiona per lo Studio dei Problemi della Popolazione*. Rome, Instituto Poligrafico dello Stato, 1931
2. CAVALLI-SFORZA LL, BODMER WF: *The Genetics of Human Populations*. San Francisco, W.H. Freeman, 1971
3. GLASS B, LI CC: *Am J Hum Genet* 5:1-20, 1953
4. LONG JC, SMOUSE PE: *Am J Phys Anthropol* 61:411-422, 1983
5. STEINBERG AG: *Prog Med Genet* 2:1-33, 1962
6. STEINBERG AG, COOK CE: *The Distribution of Human Immunoglobulin Allotypes*. London, Oxford Univ. Press, 1981
7. WILLIAMS RC, ET AL.: *Am J Phys Anthropol* 66:1-19, 1985
8. BENNETT PH, BURCH TA, MILLER M: *Lancet* 2:125-128, 1971
9. REED TE: *Am J Hum Genet* 21:71-83, 1969
10. STEINBERG AG, STAUFFER R, BLUMBERG BS, FUDENBERG H: *Am J Hum Genet* 13:205-213, 1961
11. BLUMBERG BS, WORKMAN PL, HIRSCHFIELD J: *Nature* 202:561-563, 1964