Admixture Studies and the Detection of Selection

Data analyses of admixture studies give little evidence of natural selection operating in U.S. black populations.

Julian Adams and Richard H. Ward

The estimation of genetic admixture in human populations has been used for a variety of purposes ranging from the confirmation of historical events, such as population fusion and migration, to estimation of the effects of other biological parameters. While the emphasis in early studies was on the analysis of continuous traits (*1*), investigators in later studies have utilized single locus markers to obtain a result easily interpretable in genetic terms. This has led to the use of admixture estimates to detect the action of natural selection on the many polymorphisms in human populations. Hitherto such techniques have been regarded as powerful methods for the detection of selection effects, especially when an appreciable number of generations have elapsed since the original fusion or migration. In this article, by analyzing the data for five U.S. black populations, we demonstrate that the methods proposed for the detection of natural selection are extremely susceptible to violation of certain restrictive assumptions (changing gene frequencies in parental populations, and incomplete mixture in the hybrid populations), to sampling errors, and to errors in gene frequency estimation. We conclude that to date, admixture studies have given little or no evidence for the action of natural selection.

The significance of these findings goes beyond an evaluation of our ability to detect natural selection, since studies designed to measure the genetic component of complex variables in hybrid groups require a sound knowledge of the distribution of admixture estimates at identifiable loci. Thus studies that propose to correlate factors such as intelligence quotient (IQ) with the degree of Caucasian admixture in black populations will need accurate estimates of admixture as a base datum. Because in this article we show that the available data on black populations give highly variable estimates of admixture, the results of any such study seem likely to be uninterpretable. Without the accumulation of good estimates of genetic admixture, the probability of our being able to define the genetic component of racial differences in characters such as IQ seems discouragingly small.

Description of the Model

The amount of genetic admixture may be simply defined as the proportionate contribution of one population to the genetic constitution of the hybrid population. Thus for the U.S. black population, if q_{af} and q_w represent the gene frequencies of the two parental populations (African and Caucasian, respectively) and a proportion m of the genes of the African population were replaced by genes from the Caucasian population, then the gene frequency of the hybrid population (q_n) is given by

$$q_n = (1 - m)q_{af} + mq_w \quad (1)$$

and

$$m = \frac{q_n - q_{af}}{q_w - q_{af}} \quad (2)$$

The variance of m is the variance of a ratio (*2*) and reduces to

$$\sigma^2_m = \frac{1}{(q_{af} - q_w)^2}[\sigma^2_{q_n} + m^2\sigma^2_{q_w} + (1 - m)^2\sigma^2_{q_{af}}] \quad (3)$$

These formulae are only appropriate in the absence of any systematic factors affecting gene frequencies in any of the three populations. If selection or any other force is acting at the locus in question the estimate of admixture will be biased. For the purposes of this article we describe this bias in the estimate of admixture at the i^{th} locus, $\hat{m}^i$, as

$$m^i = \hat{m} + f(s^i, q^i) + f(K^i, q^i) + e^i \quad (4)$$

where m is the amount of admixture that occurred. Thus the model partitions the bias in admixture estimates into three categories, the effect of selection [$f(s^i, q^i)$], the effect of biased estimation of gene frequencies [$f(K^i, q^i)$], and the effect of random drift, founder effects, and sampling errors in the hybrid population (e^i). Other factors, such as mutation, will be expected to have a negligible effect.

Comparison of Methods to Detect Selection

We now examine whether three methods that have been proposed for searching for selection are likely to differentiate effectively the effect of selection, $f(s^i, q^i)$, from the other sources of bias. Without such differentiation there would be little justification in attempting any further biological validation of the selective forces. Later we shall consider the results obtained by applying these methods.

1) The earliest method, that of Workman *et al.* (*3*), was designed as a screening device to identify specific loci at which selection might be operating. Workman was able to distinguish two groups of admixture estimates in 14 polymorphic traits for a population in Claxton, Georgia. He concluded that one group of traits showed no apparent evidence for selection and the other group, which had elevated estimates of admixture, provided evidence for the action of natural selection. In general there are three difficulties associated with this approach which suggest that it cannot reliably identify loci at which selection is operating. (i) Unless an investigator has a prior knowledge of which group gives unbiased estimates and which group is subject to selection, it is impossible for him to distinguish the group of loci at which selection is operating. In Workman's analysis the group of elevated estimates included G6PD and Hb S (*4*) and it

Dr. Adams is assistant professor of the department of botany and Dr. Ward is research associate in the department of human genetics at the University of Michigan, Ann Arbor 48104.

was reasonable to assume that these loci had been subjected to selection. However, it would usually be impossible to identify such groups. (ii) In practice it is difficult to assign the admixture estimates to one of two distinct groups because of the underlying continuous distribution of selection coefficients. (iii) Because this method assumes that $f(K^i, q^i)$ and e^i are close to zero for all loci, it is sensitive to poor estimates of gene frequency, the effect of genetic drift, and other factors causing changes in gene frequency (*5*).

2) The analysis of the correlations between the rankings of admixture estimates for various loci in more than one population (*6, 7*) is an interesting method that does not have some of the disadvantages of the first approach. If two parental populations give rise to more than one distinct hybrid population then the hybrid populations can be treated as "replicates." If the estimates of q_n are poor, or if the frequencies have been changed appreciably by random pressures, there will be little or no correlation between locus-specific admixture estimates. On the other hand, the action of natural selection at certain loci will produce estimates of admixture for those loci altered in the same pattern in each replicate population. A correlation between the population estimates will then exist. However, the values of q_{af} and q_w which are used in these calculations are treated as constants whereas the values of q_n are not. Thus any biases in the parental gene frequency estimates will be reflected in the admixture estimates and these biases will be of the same order for all replicate populations. Consequently, correlations between the rankings of the admixture estimates for the same group of loci from different hybrid populations can be a manifestation of biases in the parental gene frequency estimates and not a result of the action of natural selection. Even in the absence of any appreciable biases in the estimates of parental gene frequencies, a significant correlation will only indicate that at least one of the loci is being subjected to natural selection. It will not identify the locus or loci in question. Furthermore, the method suffers from the restriction that it requires data from more than one population. Hence it is necessary for selection to be operating in the same fashion on an identical set of loci in each population for the method to work.

3) The third method requires the calculation of the standard error of the admixture estimates, with the assumption of a normal approximation of the binomial distribution of gene frequencies; the estimates are tested for homogeneity by means of a chi-square test (*8*). In this method, most of the bias due to $f(K^i, q^i)$ is accounted for by large values of σ^2_{af} and σ^2_w. However, if the hybrid population is subject to drift, the estimate of admixture may vary markedly from locus to locus, without substantially altering the calculated variance associated with the gene frequency estimates. Because this also results in an increased chi-square value, a significant total chi-square may indicate either drift or selection. Furthermore, this method has the disadvantage that the chi-square test is notoriously weak and probably unable to detect the effect of selection unless it is particularly extreme.

Gene Frequency Estimates

The above discussion indicates that identification of natural selection from admixture estimates depends on the existence of good unbiased estimates of gene frequencies for the two parental populations and the hybrid population. For the hybrid population, the requirements are that a suitable random sample be taken and the sample size be large enough to keep the standard error small. Estimates of frequency from autosomal recessive genes have an additional source of bias if the population is incompletely mixed (*9*). However, the quality of the admixture

Table 1. Average gene frequencies (*4*) by geographical region [see Table 2 and Curtin (*17*)] and weighted mean gene frequencies for slave imports.

Locus	Regions								q_{af} estimates	
	I	II	III	IV	V	VI	VII	VIII	S.E. states	Whole country
R_o	.632	.611	.614	.558	.602	.563	.704		.624	.617
R_1	.055	.064	.082	.120	.058	.046	.050		.067	.066
R_2	.037	.086	.061	.027	.083	.114	.037		.057	.061
$r + R_o''$	.273	.249	.240	.296	.202	.258	.194		.240	.245
A	.159	.155	.137	.148	.119	.157	.175	.141	.156	.156
B	.127	.170	.140	.138	.121	.129	.140	.127	.136	.136
O	.714	.677	.723	.715	.760	.715	.685	.732	.708	.709
M	.339	.489	.461	.489	.470	.395	.618		.487	.474
N	.661	.512	.540	.511	.530	.605	.383		.513	.526
Fy^a			.000	.000	.000				.000	.000
P	.641	.769	.650		← .783 ——→*				.713	.723
Jk^a			.479		← .782 ——→				.666	.693
K			.002		← .004 ——→				.003	.004
Lu^a				←—— .047 ——→†					.047	.047
Js^a							.117		.117	.117
Di^a			.000						.000	.000
G6PD	.086			.178	.194	.180	.214	.197	.178	.176
Hb S	.058	.055	.052	.112	.101	.095	.116	.042	.091	.090
Hp^1	.614	.544	.727		.873	.655	.728		.696	.684
Tf^{D1}	.013	.025	.041			.074	.037		.041	.044
T					.631				.631	.631
A_1	.080		.100	.186	.085	.099			.115	.115
A_2	.082		.042	.018	.038	.043			.044	.044
Gc^1							.905		.905	.905
S					.117	.112	.241		.185	.172

* Mean gene frequency for regions V and VI. † Mean gene frequency for regions IV, V, and VI.

estimates is chiefly limited by the quality of the parental gene frequency estimates which have to be made from contemporary populations. The initial assumption is that gene frequencies have not changed over time.

The assumption of temporal stability of the gene frequencies is probably unjustified for the African population. Cultural changes and disease prevention measures have contributed substantially to changing both the West African environment and the population composition over the last 200 years (*10*). In addition, all evidence indicates a considerable degree of variability among present-day populations. Most of the African gene frequency estimates used in the literature are derived from Glass and Li (*11*) and Glass (*12*), referred to here as the "original" estimates. At that time there was very little information available on the genetic structure of African populations, and gene frequencies from one part of the continent to the other were thought to be relatively homogeneous (*13, 14*). Recent studies (*15*) have suggested that African gene frequencies are highly heterogeneous both within and between geographical regions (see Table 1), in accord with studies on other nonurban populations (*16*). We now develop a set of "improved" estimates which take into account the variation within and between African populations.

Any reliable estimate of the African parental gene frequencies must consider the proportional contribution of the various geographical regions to the original immigrant slave population. Curtin (*17*) has listed the geographical origins and approximate frequencies of slaves entering North America; the great majority entered during the relatively short period between 1710 and 1810. Two estimates of the frequency of slaves have been calculated, one for the Southeastern states, and one for the whole country (see Table 2).

Gene frequency estimates for the various regions are not so readily available. Except for the common systems, ABO, Rh, and MN, the distribution of gene frequencies is little known. In Table 1 we summarize the available gene frequencies for the various regions (*18*). The improved q_{af} estimates in Table 1 are derived from the regional means weighted by the proportional contribution of the geographical region, as shown in Table 2. Even these improved estimates may not be totally unbiased. First, several of the estimates are by necessity based on a small amount of data. It is unfortunate that more reliable estimates for T and Js^a are not available because in both loci the differences between q_w and q_{af} are large and therefore could provide reliable estimates of admixture. Second, the detection of allelic states is unsatisfactory for some loci [for example, Hp^1 (*19, 20*)]. Third, the averages for each region were unweighted averages of the gene frequency estimates for tribes of this region. No information was available for certain tribes that were known to contribute to the slave trade (*21*).

Table 2. Geographical origin of the slaves imported into the North American mainland (*17*). The data for Southeastern states represent means of figures for South Carolina and British slave trade imports; data for the whole country represent means of figures for Virginia, South Carolina, and British slave trade.

Geographical region	Southeastern states	Whole country
I. Senegal, Gambia	.12	.13
II. Sierra Leone, Portuguese Guinea, French Guinea	.05	.06
III. Liberia, Ivory Coast	.14	.11
IV. Ghana	.16	.16
V. Togo, Dahomey, Western Nigeria	.07	.04
VI. Eastern Nigeria, Cameroons	.16	.23
VII. Angola, Congo	.29	.25
VIII. Mozambique, Madagascar	.01	.02

The assumption of constant gene frequencies is probably reasonable for the white ancestral population. The environment of the Caucasians has not changed substantially over the time period involved and the effective population size is sufficiently large to make the effect of genetic drift negligible. Gene frequencies in the white population appear to be invariant over a wide geographical range. Estimates of gene frequencies for Oakland whites (*22*) lie within two standard deviations of the estimates for Claxton whites (*23*), with the exception of the B allele (Table 3).

Populations

Data from five U.S. black populations are suitable for inclusion in this study. (i) Claxton, Georgia (*23*); the samples ranged in size from 167 to 304 (blacks), and 91 to 333 (whites). (ii) Sapelo Island, Georgia (*6*); all but 20 of the Island's 211 inhabitants were sampled. (iii) Charleston, South Carolina (*24*); sample size varied between 119 and 515. (iv) James Island, South Carolina (*25*); sample sizes ranged between 54 and 151. For both Sapelo Island and James Island, demographic data indicate that the effective size of the population is quite low and random drift could have had a considerable effect. (v) Oakland, California (*22*); the sample size was large, and the individuals were apparently unrelated. However, this population has the disadvantage of being more heterogeneous than the first four populations, and occupying a different environment; other populations were within 150 miles (240 kilometers) of each other. Gene frequencies for the five hybrid populations as well as the two white populations are summarized in Table 3.

Bimodal Distribution of Admixture Estimates

Table 4 shows estimates of migration rates calculated for selected loci from the original and the improved African gene frequency estimates. In this and the following section, we only calculated migration rate estimates for those loci where $|q_w - q_{af}|$ is greater than 0.09 (*3, 4*). The criterion for this exclusion was based on the improved q_{af} estimates. Where data were available, the loci were divided into the same two groups, as in the paper of Workman (*5*). Although the estimates of admixture based on the original q_{af} estimates tend to be elevated in group II, this distinction disappears when the improved q_{af} estimates are used (*26*). Furthermore, there is little consistency from one hybrid population to another. The extreme variability of admixture estimates across loci, many of which are biologically unreasonable, indicates the virtual impossibility of separating out a clearly demarcated group of loci with recognizably higher (or lower) admixture estimates. While it is tempting to interpret the elevated value for Hb S, as a function of the joint action of selection and migration, this interpretation is based on the assumption that selective pressures operating at the Hb S locus are substantially different in the United States compared with West Africa. This assumption may well be invalid because malaria was endemic in the Southeast until the 1930's (*24*), in which case balanced polymorphism

Table 3. Gene frequency estimates for U.S. white (q_w) and U.S. black (q_n) populations. Maximum likelihood estimates of gene frequency were calculated from the phenotype arrays (where available). This accounts for the slight difference between our estimates and previously published estimates.

Locus	q_n					q_w	
	Claxton	Sapelo Island	James Island	Charleston	Oakland	Claxton	Oakland
R_o	.533	.696	.536	.496*	.486	.022	.023
R_1	.109	.084	.129	.081*	.161	.429	.413
R_2	.109	.109	.089	.063*	.071	.137	.140
r	.230	.057	.226	.313*	.253	.374	.405
A	.145	.141	.158	.138	.175	.241	.247
B	.113	.058	.114	.153	.125	.038	.068
M	.484	.470	.464	.509	.486	.507	.550
S	.157	.180	.110		.161†	.279	.315
Fy^a	.045	.029	.038	.016	.094	.422	.429
P	.757	.562			.737	.525	.491
Jk^a	.743	.681				.536	
Js^a	.123	.198				.002	
T	.670	.674				.527	
Hp^1	.518	.630				.413	
G6PD	.118	.127				.000	
Hb S	.043	.051		.080		.000	

* Distribution of phenotypes departs significantly from Hardy-Weinberg expectations at $\alpha < .05$.
† Distribution of phenotypes departs significantly from Hardy-Weinberg expectations at $\alpha < .001$.

would have been maintained as in West Africa. It thus seems unwise to attempt to explain the deviant admixture estimates in terms of selection. Random drift, or biased African gene frequency estimates are more likely explanations.

Correlations between Admixture Estimates

The method of correlating estimates of admixture can potentially distinguish between selection and genetic drift in the hybrid population; only selection will give positive correlations between admixture estimates. The consistency of rankings of admixture estimates in all five populations may be tested for those alleles common to all five populations (R_0, R_1, R_2, r, A, B, Fy^a) (*27*). The results may be expressed in terms of the mean pairwise rank correlation coefficient between admixture estimates as .035, and .211 for the original and improved q_{af} estimates, respectively. Neither of these figures is significantly different from zero, providing no evidence for the action of natural selection.

Data from all possible alleles may be used only to calculate individual pairwise correlation coefficients. Table 5 shows the correlation matrices derived from the original and improved q_{af} estimates, all alleles common to each pair of populations being used. The correlations for the most part center around zero; only 3 out of 20 correlations differ significantly from zero. In addition, the correlations between the Oakland population and the other four populations do not form a distinct set in comparison with the other coefficients in the matrix. Thus there is no evidence that the Oakland population is subject to a different set of selective pressures.

Although the significant pairwise correlations may suggest that further studies should be undertaken in these populations, the overall result is that there is little or no correlation between the rankings of the admixture estimates. Thus with the available data, this method is unable to show any statistically significant effects that may be ascribed to natural selection.

Chi-Square Tests for Homogeneity

Variances in the admixture estimates, σ^2_m, were calculated from the data in Tables 1, 3, and 4 (by means of Eq. 3). Variances in gene frequencies for U.S. whites and U.S. blacks were derived from the maximum likelihood estimates where phenotype distributions were available, otherwise the lower limit was used (*28*). Variances of the African gene frequencies were taken to be the variance of the mean gene frequencies for the eight different regions (see Table 1) weighted by the frequencies of geographical origin shown in Table 2. Two sets of variances were calculated, one for the Southeast U.S. populations, and one for the Oakland population. For those loci and regions where gene frequency estimates were unavailable, the appropriate weighted mean gene frequency was used, a method which lowers the variance. Furthermore, these variances of the African gene frequencies are almost certainly underestimates of the

Table 4. Admixture estimates. Results in roman type are estimates obtained from the original q_{af} estimates; results in italics are estimates obtained from the improved q_{af} estimates.

Locus	Claxton		Sapelo Island		James Island		Charleston		Oakland	
					Group I					
R_o	.107	*.151*	− .178	*− .119*	.101	*.146*	.172	*.214*	.189	*.220*
R_1	.110	*.116*	.041	*.047*	.167	*.172*	.033	*.039*	.268	*.274*
R_2	.446	*.647*	.450	*.650*	.059	*.402*	− .460	*.072*	− .284	*.121*
r	.117	*− .077*	− .948	*− 1.378*	.092	*− .108*	.628	*.546*	.216	*.049*
A	− .037	*− .128*	− .076	*− .170*	.108	*.033*	− .113	*− .216*	.268	*.204*
B	.333	*.234*	.825	*.798*	.330	*.244*	− .019	*− .170*	.315	*.161*
S	.157	*− .299*	.317	*− .053*	− .163	*− .805*			.148	*− .086*
Fy^a	.108	*.108*	.069	*.069*	.090	*.090*	.037	*.037*	.220	*.220*
P	.092	*− .236*	.856	*.804*					.148	*− .064*
Jk^a	.164	*− .586*	.413	*− .114*						
Js^a		*− .052*		*− .701*						
					Group II					
T	.466	*− .375*	.452	*− .413*						
Hp^1	.619	*.627*	.28	*.231*						
G6PD	.395	*.338*	.33	*.288*						
Hb S	.614	*.532*	.49	*.438*			.273	*.119*		

true parameters because no allowance is made for the variance in gene frequency within a geographical region. For some loci these variances can be considerably larger than the gene frequency variance between regions (*29*). Consequently, the estimates of the variance of the amount of admixture are undoubtedly conservative. Table 6 shows these variances together with the chi-square value for homogeneity of admixture estimates calculated from

$$\chi^2_{(K-1)} = \sum_{i=1}^{K} \frac{(m^i - \bar{m})^2}{\sigma^2_{m^i}}$$

where $\bar{m}$ is the weighted mean of the admixture estimates for all loci (*30*).

Two populations, Claxton and Sapelo Island, give significantly large chi-square values, indicating heterogeneity of the admixture estimates. Although it is possible to identify the loci which give large chi-square values, no conclusions can be drawn concerning the action of natural selection on these loci. A large chi-square value merely indicates a large deviation from the weighted mean. In these data the weighted mean is determined principally by the value for the Fy^a locus which has the smallest variance for the admixture estimate in all populations. If we could rule out the action of selection at the Fy^a locus, then the admixture estimate for Fy^a could be taken as a base value (*31*) and conclusions could be drawn from the magnitude of individual values. However, recent evidence suggests that the admixture estimate for the Fy^a locus does not reflect the action of migration alone (*32*). Therefore, we must consider the total chi-square values only.

The significant chi-square values must be interpreted with caution because, as outlined previously, large total chi-square values can result as much from the action of random drift as from the action of selection. The Sapelo Island population is small and its history suggests that gene frequencies could have been changed substantially by random processes, leading to the large total chi-square value. Furthermore, the population sampled in the Oakland study, which was by far the largest of the five populations considered here, is associated with a small chi-square value. Overall then, the apparent inverse relationship between total chi-square value and population size appears more indicative of random fluctuation operating in the smaller hybrid populations than of the action of selection.

Conclusions

We have described three techniques that have been used to detect the action of natural selection in hybrid populations. When the data from five U.S. black populations are compared, the results give little indication of selection. Although each approach viewed apart appears to give some indication that natural selection operates in one or more populations, there is little corroboration from one method to another. Four of the five populations examined are within 150 miles of each other and are presumed to have a similar parental composition (Table 2) and hence to be subject to the same selective pressures. If selection were operating at a set of loci in these hybrid populations, all three methods should give concordant results. The data in Tables 4 to 6 indicate that they do not. Even the apparent concordance between the Claxton and Sapelo Island populations (*6*) breaks down when the heterogeneity of the admixture estimates is considered (Table 6). Different loci in each population contribute disproportionately (that is, values > 3.84) to the total chi-square value (Table 6), whereas a similar pattern of values would be expected for both populations. Hence, it is unreasonable to build any case for selection from these data.

None of the three methods discussed herein is entirely satisfactory for identifying the action of selection in admixture. It seems that the first method should be abandoned in favor of the third method, which depends upon the same principle of evaluating the variability of admixture estimates. The greatest drawback to the third method appears to be its inability to correct for genetic drift occurring in small hybrid populations (*33*). In addition, when a single locus can contribute disproportionately to the mean (in this instance, the Fy^a locus), no interpretation can be given to the significance of locus-specific chi-square values.

Table 5. Rank correlation coefficients between the admixture estimates. The lower triangular matrix (data in roman type) represents correlation coefficients calculated from original q_{af} estimates, and the upper triangular matrix (data in italics) represents correlation coefficients calculated from improved q_{af} estimates.

Populations	Populations				
	Claxton	Sapelo Island	James Island	Charleston	Oakland
Claxton		*.493*	*.952**	*.214*	*.516*
Sapelo Island	.420		*.738†*	*− .310*	*− .050*
James Island	− .262	− .071		*.262*	*.404*
Charleston	.095	− .381	− .357		*− .143*
Oakland	− .167	− .200	.857*	− .286	

* Significant at $\alpha < .01$ for a two-tailed test. † Significant at $\alpha < .05$ for a two-tailed test.

Table 6. Standard deviation of admixture estimates for each locus, overall weighted means, and chi-square values for homogeneity.

Locus	Claxton	Sapelo Island	James Island	Charleston	Oakland
	Standard deviation of admixtures, σm^i				
R_o	.094	.113	.094	.083*	.075
R_1	.076	.078	.079	.075	.071
R_2	.241*	.271*	.317	.372	.406
$r + R_o{}^u$	.367	.705*	.360	.218	.220
A	.277	.307	.316	.203	.138
B	.159	.139*	.207	.172	.142
Fy^a	.021	.021	.161	.010	.009
P	.363	.210*			.238
Jk^a	1.098	.785			
G6PD	.252	.181			
Hb S	.168*	.209		.288	
Hp^1	.149*	.203			
S	.652	.552	.913		.356
	Overall weighted means, $\bar{m}$				
	.126	.088	.153	.040	.219
	Chi-square values for homogeneity				
	26.92†	56.24‡	2.79	13.09	3.59

* Individual chi-square values > 3.84. † Significant at $\alpha < .05$. ‡ Significant at $\alpha < .001$.

The second method can be considered from two standpoints. First, we can perform multiple population comparisons on all common loci. Here a significant result would present a strong case for the action of natural selection, although for the data from the five populations discussed here the results were nonsignificant. Second, we can examine the individual pairwise correlation coefficients. In this case the existence of 3 significantly positive correlations out of 20 cannot be considered a reliable indicator of selection.

One interpretation of the lack of consistency and significance of these results is that selection does not operate on the loci in question. Although this may be true for some loci, it is unlikely that all of the loci are neutral (*34*). A more reasonable conclusion would be that for the data available, the methods we have outlined are not sensitive enough to detect selective coefficients below a certain size. Since the results do not give unequivocal evidence of the action of selection at the Hb S locus, it seems that even where the magnitude of selection is appreciable, none of the methods of analysis will be entirely satisfactory.

The studies cited herein give little indication of the action of natural selection, and it is doubtful whether any study of admixture will yield unequivocal conclusions. There are two basic requirements for successful admixture studies: good estimates of gene frequency and a knowledge of how the parental gene frequencies have changed over time. We have mentioned the problems involved in estimating gene frequencies in African populations, but the populations and sample sizes used for estimating gene frequencies in U.S. black and Caucasian populations should be adequate for the detection of small amounts of selection. The main problem thus centers around the usual assumption that gene frequencies did not change in the parental populations, but do change by natural selection in the hybrid population. Unless special circumstances prevail, one or other of the parental populations should also have experienced some change in gene frequencies in the time interval concerned. Although we can postulate situations that would result in constant gene frequencies for the parental populations, the postulates are untestable and must surely have been violated for certain loci. We must therefore question the usefulness of admixture studies in the search for selection. Problems in gene frequency estimation may be eliminated by judicious choice of the populations, but it is unlikely that it will ever be possible to distinguish between the effects of selection and the effects of bias due to changing parental gene frequencies.

Summary

Three methods can be used to search for evidence of natural selection from admixture studies. These include classification of the admixture estimates into two groups; calculation of the rank correlation between estimates from more than one population; and the testing of admixture estimates for homogeneity. The use of these methods is discussed with special reference to black-white admixture in the United States. Using revised estimates of African gene frequencies, derived from a consideration of the geographical origin of the slaves, we calculated admixture estimates and their variances for five U.S. black populations; Claxton (Georgia); Sapelo Island (Georgia); James Island (South Carolina); Charleston (South Carolina); Oakland (California). Two out of the five populations yielded heterogeneous admixture estimates but all other tests were nonsignificant. The data provide little evidence for the action of selection. The few, inconsistent significant results are more indicative of the action of random drift or biased gene frequency estimates than natural selection, and in general these effects cannot be differentiated. It seems doubtful that admixture studies can ever provide unequivocal evidence for the action of natural selection in human populations. In the search for natural selection, perhaps admixture studies should only be used as a preliminary screening device.

References and Notes

1. C. Davenport and M. Steggerda, *Race Crossing in Jamaica* (Carnegie Institute of Washington, D.C., 1929).
2. R. Geary, *J. Roy. Statist. Soc.* **93**, 442 (1930).
3. P. Workman, B. Blumberg, A. Cooper, *Amer. J. Hum. Genet.* **15**, 429 (1963).
4. Abbreviations: Loci and alleles described in this article are as follows. Blood groups: Rh (R_0, R_1, R_2, r, R_0^u), ABO (A_1, A_2, B, O), MNSs (M, N, S), Duffy (Fy^a), P (P), Kidd (Jk^a), Kell (K), Lutheran (Lu^a), Sutter (Js^a), Diego (Di^a). The serum protein genes are Haptoglobin (Hp^1), Gc (Gc^1); transferrin (Tf^{D1}). Hemoglobin (Hb^s). Erythrocyte enzyme, glucose 6-phosphate dehydrogenase deficiency (G6PD). Phenylthiocarbamide tasting (T).
5. P. Workman, *Hum. Biol.* **40**, 260 (1968).
6. B. Blumberg and J. Hesser, *Proc. Nat. Acad. Sci. U.S.A.* **68**, 2554 (1971).
7. K. Hertzog and F. Johnston, *Hum. Biol.* **40**, 86 (1968).
8. L. Cavalli-Sforza and W. Bodmer, *The Genetics of Human Populations* (Freeman, San Francisco, 1971).
9. H. Kalmus, *Ann. Hum. Genet.* **32**, 311 (1969).
10. F. Livingstone, *Evolution* **18**, 685 (1964); S. Wiesenfeld, *Science* **157**, 1134 (1967).
11. B. Glass and C. Li, *Amer. J. Hum. Genet.* **5**, 1 (1953).
12. B. Glass, *ibid.* **7**, 368 (1955).
13. A. Mourant, *The Distribution of Human Blood Groups* (Blackwell, Oxford, 1954).
14. A. Mourant, A. Kopěc, K. Domaniewska, *The ABO Blood Groups* (Blackwell, Oxford, 1958).
15. J. Hiernaux, *La diversité Humaine en Afrique Subsaharienne* (Université Libre de Bruxelles, Brussels, 1968).
16. J. Neel and R. Ward, *Proc. Nat. Acad. Sci. U.S.A.* **65**, 323 (1970).
17. P. Curtain, *The Atlantic Slave Trade* (Univ. of Wisconsin Press, Madison, 1969).
18. The main source of data was Hiernaux (*15*) and F. Livingstone, *Abnormal Hemoglobins in Human Populations* (Aldine, Chicago, 1967), the latter providing data for abnormal hemoglobin and G6PD frequencies. Other sources of data utilized were: H. Kalmus, *Human Biol.* **39**, 32 (196 → N. Barnicot and S. Lawler, *Amer. J. Phys. Anthropol.* **11**, 83 (1953); J. Chalmers, E. Ikin, A. Mourant, *Ann. Eugen.* **17**, 168 (1953); G. Fraser, E. Giblett, A. Motulsky, *Amer. J. Hum. Genet.* **18**, 546 (1966); E. Ikin and A. Mourant, *Man* **52**, 21 (1952); F. Livingstone, H. Gershowitz, J. Neel, W. Zuelzer, M. Solomon, *Amer. J. Phys. Anthropol.* **18**, 161 (1960); A. Mourant (*13*); A. Mourant *et al.* (*14*); D. Roberts, *Amer. J. Hum. Genet.* **7**, 361 (1955); and E. Giblett *et al.* (*19*).
19. E. Giblett, A. Motulsky, G. Fraser, *Amer. J. Hum. Genet.* **18**, 553 (1966).
20. E. Giblett and A. Steinberg, *ibid.* **12**, 160 (1960).
21. For example, no genetic data were available for the Akan tribe despite the fact that this tribe appears to have contributed disproportionately to the original slave population (G. Uzoigwe, personal communication).
22. T. Reed, *Amer. J. Hum. Genet.* **20**, 142 (1968).
23. A. Cooper, B. Blumberg, P. Workman, J. McDonough, *ibid.* **15**, 420 (1963).
24. W. Pollitzer, *Amer. J. Phys. Anthropol.* **16**, 241 (1958).
25. ———, R. Menegaz-Bock, R. Ceppellini, L. Dunn, *ibid.* **22**, 393 (1964).
26. The Mann-Whitney test was used to test the difference between the groups of admixture estimates. The two groups of estimates from the Claxton population were significantly different at $\alpha < .01$ (group II having higher estimates of admixture), when the original q_{af} values were used. However, the groups of admixture estimates were not significantly different when the improved q_{af} estimates were used. The two groups of admixture estimates from the Sapelo Island population (the only other population with more than one entry in the second group of estimates) were not significantly different whichever set of q_{af} estimates were used.
27. Kendall's coefficient of concordance, W, is a measure of the overall agreement in rankings of the variates when there are more than two populations. This test, which is related to the Friedman test statistic, can be expressed in terms of the mean value for all pairwise correlations (Spearman's rho).
28. When phenotypic frequencies were not available the variance of a multinomial distribution was used. This is the lower limit of the variance associated with a gene frequency estimate and besides assuming an appropriate sampling procedure it also assumes that all genotypic classes can be distinguished.
29. For example the average variance of the frequency of the Hb S allele within geographical regions (.0013) is larger than the gene frequency variance of the means for the geographical regions (.0007) [Compare R. Lewontin, *Evol. Biol.* **6**, 381 (1972)]. This situation also arises for other loci where there is a reasonable amount of information (ABO, MN, Rh).
30. The weighted mean of the admixture estimates for each population was calculated from

$$\bar{m} = \sum_{i=1}^{K} \left[\frac{m^i}{\sigma^2_{m^i}} \Big/ \left(\sum_{i=1}^{K} \frac{1}{\sigma^2_{m^i}} \right) \right]$$

This procedure gives most weight to alleles that apparently contain the most information about admixture (that is, the smallest variance associated with the admixture estimate). However, should there be perturbating factors influencing alleles associated with small variances, then the bias in the mean estimate of admixture may be considerable.

31. T. E. Reed, *Science* **165**, 762 (1969).
32. Recent evidence suggests the Duffy locus may be linked to a locus, such as skin color, at which assortative mating occurs in the black population [H. Gershowitz, *Amer. J. Hum. Genet.* **24**, 38a (1972)].
33. As an appropriate correction one could take into consideration the variance component due to genetic drift. Thus, with information about the size of the hybrid population, the number of generations involved, and the presumed parental composition, the distribution of the variance of gene frequencies through time could be obtained as a joint function of migration rate and genetic drift [compare M. Kimura, *Diffusion Models in Population Genetics* (Methuen, London, 1964), pp. 1–57].
34. For practically all the loci studied some kind of genotype-viability or genotype-fertility has been postulated [see Cavalli-Sforza and Bodmer (*8*), pp. 1–965].

35. Supported in part by PHS grant GM 18840 and Atomic Energy Commission contract AT(11-1)-1552. We have profited from discussions with H. Gershowitz, J. Neel, T. Reed, C. Sing, R. Spielman, and P. Workman.

Racial Aspects of Zero Population Growth

Ernest B. Attah

The possibility of zero growth in the population of the United States has stimulated some recent investigations (*1–3*) of the implications of alternative paths to that condition, based on projections of the population as a whole. Given the history (*4*) of marked differentials in fertility and mortality rates between the white and nonwhite segments of the population, this article is concerned with the consequences of different rates of approach to zero growth. Specifically, what would be the effects of different rates on the short- and long-term growth of the respective segments of the population? How long would it take the population to stabilize, and how much would the population have increased by then? What intermediate trends would appear in the proportion of nonwhites in the population, and what would be the relative sizes of the white and nonwhite segments in the long run?

These questions are explored on the basis of separate projections of the white and nonwhite segments of the population, incorporating different assumptions about the patterns through which zero growth is attained (intermixture of the two segments is not considered). The assumptions are based on alternative trends in the reproduction rate, considered in terms of the interaction between alternative fertility and mortality trends.

Modeling Approach

Since the discussion will mainly emphasize large-scale and long-term questions, it will be sufficient to work with only the main demographic factors of population growth for these projections. Like Frejka (*1*, p. 380), I shall "ignore factors which can play important roles in the short run, but which seem fairly stable in the long run (changes in proportion married, changes in average age at marriage, etc.)" and base the projections only on fertility rates, mortality rates, and reproduction rates. Model specification will turn primarily on the net reproduction rate (N) and will take the form of alternative assumptions about the speed with which replacement rates of fertility ($N = 1.0$) are attained, beginning from the initial date. Fertility rates will be assumed to remain indefinitely at replacement once that is attained. Zero population growth will thereby emerge with time, as stipulated by the theory of stable populations (*5*).

The initial year for all projections is 1965. Thus the initial demographic characteristics of both the white and nonwhite populations are those of 1965, shown in Tables 1 and 2. To simplify the arithmetic, projections are carried out for the female population only. For present purposes, it is sufficient simply to double the projected female population in order to obtain the corresponding total population. Projections are generally made for 5-year intervals and extend for about 200 years. Five separate projections, based on a fixed pattern of mortality decline combined with five different patterns of fertility decline, have been made for each of the white and nonwhite populations. Fertility rates predominate heavily over mortality rates, with respect to influencing possible zero growth. I therefore emphasize in these projections the effects of alternative trends in fertility.

Assumptions about mortality are based on the life expectancy at birth (e_0), which provides a convenient summary measure of the overall circumstances of sickness and death, over the entire span of life, in a given population (*6*). For the white population, life expectancy is assumed to increase linearly (by 6 months each period) over five projection periods after the initial period, until it attains a fixed level of 77.5 years [for convenience, life expectancy in the initial period is set at 75.0 years (Table 2)]. For the nonwhite population (life expectancy in the initial period set at 67.5 years), the assumed linear increase in life expectancy occurs in two stages: first with a 1-year increase each period up to a life expectancy of 74.5 years; then with a 6-month increase each period until the final, fixed level of 77.5 years is attained. The supporting argument for these assumptions is that, whereas mortality rates in the white population in 1965 are already quite low and future declines are therefore likely to be gradual, the corresponding rates in the nonwhite population are high enough that there is room for relatively rapid declines until rates similar to those for whites in 1965 are attained, after which declines may follow the same pattern as in the white population. The same final life ex-

The author is assistant professor of sociology at Brown University, Providence, Rhode Island 02912. This is a revised version of a paper presented at the meetings of the Population Association of America in Toronto, April 1972.