Caucasian Genes in American Negroes

Measurement of non-African ancestry is difficult, but it is worthwhile for several genetic reasons.

T. Edward Reed

It is very difficult to describe the genetic history of a large, defined human population in a meaningful way. As a result there have been few opportunities, at the population level, to study the consequences of known genetic events in the recent past of modern populations. The Negro population of the United States, however, is one of the exceptions to these generalizations. The American individual to whom the term *Negro* is applied is almost always a biracial hybrid. Usually between 2 and 50 percent of his genes are derived from Caucasian ancestors, and these genes were very probably received after 1700. While it is obviously of social and cultural importance to understand Negro hybridity, it is less obvious that there are several pertinent genetic reasons for wishing to know about the magnitude and nature of Caucasian ancestry in Negroes. Recent data, both genetic and historical, now make possible a better understanding of American Negro genetic history than has been possible heretofore. Here I review and criticize the published data on this subject, present new data, and interpret the genetic significance of the evidence.

In order to put the genetic data in proper context, I must first give a little of the history of American slavery. The first slaves were brought to what is now the United States in 1619. Importation of slaves before 1700 was negligible, however, but after that date it proceeded at a high rate for most of the 18th century. Importation became illegal after 1808 but in fact continued at a low rate for several more decades (*1*, *2*). The total number of slaves brought into the United States was probably somewhat less than 400,000 (*3*). Charleston, South Carolina, was the most important port of entry, receiving 30 to 40 percent of the total number (*4*). More than 98 percent of the slaves came from a very extensive area of West Africa and west-central Africa—from Senegal to Angola—and, in these areas, from both coastal and inland regions. Shipping lists of ships that brought slaves to the United States—and to the West Indies, often to be sent later to the United States—provide a fairly detailed picture of the geographic origins of the slaves and a less complete picture of their ethnic origins. Table 1 gives the approximate proportions of American slaves brought from the eight major slaving areas of Africa. The contribution from East Africa is seen to be negligible, whereas the area from Senegal to western Nigeria contributed about half the total and the region from eastern Nigeria to Angola contributed the other half. An earlier tabulation for entry at Charleston alone (*5*) is quite similar, except that the contribution from the Bight of Biafra is much less (0.021 as compared to 0.233) and that from "Angola" is appreciably greater (0.396 as compared to 0.245).

At some early point in American slavery, matings between slaves and Caucasians began to occur. Quantitative data are lacking, and we can say only that most of these matings occurred after 1700. Our concern here is the genetic consequences of the matings—the introduction of Caucasian genes into the genome (or total complement of genetic material) of the American Negro. We could, in theory, estimate the Caucasian contribution to American Negro ancestry in a very simple way *if* certain strict criteria were met. In practice it is not possible to show that all these criteria are met, but this fact has not stopped geneticists, including myself, from making estimates.

The usual estimation procedure is simple and direct. Consider some gene —say the allele *A* of the ABO blood group locus, whose frequency was q_a in the African ancestors of American Negroes and q_c in the Caucasian ancestors, while in modern American Negroes the frequency is q_n. If M is the present proportion of genes at this genetic locus (and, ideally, at every other locus too) which are derived from Caucasians, and if race mixture is the only process affecting q_n, then, by definition,

$$q_n = Mq_c + (1 - M)q_a \qquad (1)$$

and therefore

$$M = (q_n - q_a)/(q_c - q_a) \qquad (2)$$

This formula for M, or an algebraic equivalent, was used for all estimates of M given in Table 2 except one. [This one differed only in that three alleles were used simultaneously at one locus to obtain a maximum likelihood estimate for M; for each allele an equation of the type of Eq. 1 could be written, and used to estimate M (*6*)]. We see that if we know q_a, q_c, and q_n (for a defined area) without error and if there were no factors affecting q_n other than race crossing, estimation of M would be simple. Unfortunately, such is not the case.

Criteria for Critical Estimation of M

Critical evaluation of estimates of M requires complete specification of the needed criteria and judgment on the degree to which these criteria are met. These criteria are simple and obvious, but the demands they make have not always been appreciated. They are as follows.

1) The exact ethnic compositions of the two ancestral populations, African Negro and Caucasian, are known.

2) No change in gene frequency (for the gene in question) between ancestral and modern populations either of African Negroes or of American Caucasians has occurred.

3) Interbreeding of the two ancestral populations is the only factor affecting gene frequency in U.S. Negroes —that is, there has been no selection, mutation, or genetic drift.

4) Adequate samples (that is, samples that are unbiased, from correct

Dr. Reed is professor in the departments of zoology and anthropology, and associate professor in the department of paediatrics, University of Toronto, Toronto, Canada.

populations, with small standard error) of the modern descendants of the ancestral African Negroes and U.S. Caucasians, and of modern U.S. Negroes, are available.

It should be said immediately that none of these criteria has been shown to be fully met in any study. In particular, point 1 is not met, because the detailed ethnic origins of slaves from the various slaving areas are unknown (*4*). Point 2 can never be met because ancestral gene frequencies are unknown and point 3, at best, can only be inferred from indirect evidence. Point 4 cannot be fully met for African Negroes, since the proportions of various ethnic contributions are only roughly known. The problem is simpler for U.S. Negroes and Caucasians, although marked heterogeneity in values of M between different Negro populations is now known to complicate the matter.

Somewhat more affirmative views on these criteria can also be given, however. *If* it can be shown that gene frequencies in neighboring modern tribes and in populations of adjacent former slaving areas do not differ appreciably, point 1 becomes less important. For example, this appears to be the situation for the ABO blood groups, the best-known genetic system throughout the slaving area. With regard to point 2, since the populations concerned usually were, and are, large, it is probable that this criterion is quite well satisfied. If point 1 is satisfied in the way suggested, point 4 may be met by using large, carefully collected samples. Unfortunately, it is less easy to overcome the problem posed by point 3. This is discussed below.

Table 1. African origins of slaves imported into the North American mainland [data of Curtin (*37*)]. Distribution by areas is approximate and is an average of data for Virginia (1710–1769), for South Carolina (1773–1807), and for the British slave trade (1690–1807).

Coastal region of origin	Approximate present area	Peoples	Approximate proportion from region
Senegambia	Senegal and Gambia	Mainly Bambara and Malinke (from interior)	0.133
Sierra Leone	Sierra Leone	Sierra Leone, Guinea, Portugese Guinea peoples, plus Bambara and Malinke	.055
Windward Coast	Ivory Coast, Liberia	Various peoples of area	.114
Gold Coast	Ghana	About ¾ Akan people from southern part, the rest from northern part	.159
Bight of Benin	Togo, Dahomey, Nigeria west of Benin river	Peoples of Togo, southern Dahomey, and western Nigeria	.043
Bight of Biafra	Nigeria (east of Benin river) to 1°S (Gabon)	About ¾ Ibo, the rest Ibibio and people from Cameroon	.233
"Angola"	1°S to southwest Africa (Gabon, Congo, Angola)	Many peoples of the area, from the coast to far inland	.245
Mozambique and Madagascar			.016
Region unknown			.002

Review of Published Estimates of M

Table 2 is a tabulation of published estimates of M for American Negroes, beginning with the well-known estimate of 0.31 for Baltimore Negroes given by Glass and Li in 1953 (*7*). The estimates are grouped according to the authors' statements as to their validity or lack of validity (due to selection) as estimates of M. They are further classified as "southern" (estimates for Georgia, South Carolina, and Tennessee) or "non-southern." As has been noted elsewhere (*6*, *8*, *9*), among the presumed valid estimates, all "non-southern" estimates are greater than "southern" estimates. Also, the estimates presumed to indicate selection are usually appreciably higher than the estimates presumed to be valid. Among the "valid" estimates of M, that of Glass and Li (*7*) is by far the best known, and is often quoted as "the" estimate for the amount of Caucasian ancestry in "the" American Negro (see, for example, *10–14*). A revision of this estimate from 0.31 to 0.216 (*15*) appears to have escaped general notice.

The estimates of Table 2 must be considered in the light of the four criteria given above. As already noted, criterion 1 cannot be strictly satisfied for any estimate because the detailed ethnic origins of the slaves are unknown. The estimates for M in Table 2, however, do not even roughly meet criterion 1, since none of them is based on quantitative information on distribution of origins, such as is given in Table 1. Typically, data from only one or two regions of West Africa are taken to represent the whole slaving area. Ironically, for the best-known estimate, that of Glass and Li (*7*), Rh blood group data from East and South Africa, as well as from Ghana, were used to represent ancestral Rh blood group frequencies because better data were not then available. Glass, for his revised estimate (*15*), used only Rh data from Nigeria and Ghana. Of the 540 individuals from Nigeria studied (*15*), 105 were Ibos, who may be representative of ancestral inhabitants of the Bight of Biafra region, the area of origin of about 23 percent of American slaves (Table 1); the remaining 435 individuals from Nigeria may be representative of the slaves (4 percent) who came from the Bight of Benin. The 274 individuals from Ghana studied (*15*) may be representative of the slaves (16 percent) from that region. The slaves (57 percent) from areas other than Nigeria and Ghana are unrepresented in Glass's revised estimate. These same Rh blood group data were used by later investigators in arriving at their own estimates (*8*, *9*, *16*, *17*). These critical comments on the best-known estimate are made to illustrate the nature of the problem; similar comments could be made about each of the other estimates of Table 2.

With regard to criterion 4 (adequacy of samples), one can distinguish between (i) adequate representation (by the mean gene frequency used) of the entire slaving area and (ii) adequate sample size (as shown by a small standard error for M). If the gene used has a uniform frequency over the entire slaving area, any large sample from one part of the area could adequately represent the whole. The problem, of course, is to demonstrate uniformity. If, as one would expect, gene frequencies vary from region to region of the slaving area, appropriate samples over the whole area are needed if one is to obtain a properly weighted mean frequency. Neither of these approaches has been used in making any of the estimates. [I made an attempt to con-

firm the belief that the frequency of certain Gm alleles is near zero in African populations (*6*) but found that not enough surveys had been made.]

To make the problem more concrete, let us consider Glass's estimate of M (*15*) in the light of more recent Rh data. For the R^0 allele of the Rh locus, he used 0.5512 for the frequency in West Africa (on the basis of the data from Nigeria and Ghana). The frequencies in present-day U.S. Negroes and Caucasians were found to be 0.4381 and 0.0279, respectively, so that, from Eq. 2, we estimate M to be $(0.5512-0.4381)/(0.5512-0.0279)$, or 0.216. However, the frequency of R^0 in Liberia is 0.60 (*18*), and in Bantu of the Congo (Leopoldville) it is also about 0.60 (*19*). If the true overall value for the slaving area were 0.60, the estimate for M would be 0.283.

With regard to the purely statistical accuracy of the estimates of M, as shown by standard errors, calculation of the standard errors for several pertinent estimates indicates that they may be much larger than the authors may have suspected (*20*). The standard error for Glass's estimate (*15*), for example, is 0.042, giving a 95-percent confidence interval of 0.133 to 0.299. The estimate in Table 2, of 0.13 for M for gene AK^2 (the lowest estimate for the non-southern region) has a standard error of 0.053, producing a 95-percent confidence interval of 0.025–0.234, overlapping Glass's estimate. This large error seems particularly surprising at first, in view of the large sample sizes, but it is explained by the very low AK^2 gene frequencies (<5 percent). The standard errors of the other estimates appear to be of comparable size or larger (due to smaller sample sizes).

I have said enough to show the deficiencies of most of the estimates of Table 2 with regard to both African gene frequency and statistical accuracy. I should also comment on the classification of M estimates as "valid" (not affected by selection) or as indicating the effects of selection. Classification of an estimate in this way requires a "standard" M that is thought to be free from the effects of selection. Such a "standard" can then be used to determine whether an M estimated for some other gene demonstrates selection. The M estimates from Rh genes R^0 and R^1 have been assigned this role of "standard" by various investigators [Parker *et al.* (*21*) chose R^0 alone; Workman and his associates (*8*, *9*) chose R^0 and R^1 in combination]. In addition, M estimates from frequencies of the Fy^a allele of the Duffy blood group locus (*8*) and the Gm^1 and Gm^5 alleles of the Gm serum group locus (*21*) have been considered as possible standards. Yet, as discussed above, it is not possible to prove directly that selection has not affected a particular gene frequency in American Negroes, and no evidence in support of the belief that it has not has been offered. We can only draw inferences of varying degrees of rigor as suitable data become available. I attempt in the remainder of this article to draw and apply such inferences.

Table 2. Published estimates of the proportion (M), in American Negroes, of genes that are of Caucasian origin. All estimates except those based on genes Fy^a, Gm^1, $Gm^{1,2}$, or Gm^5 (and perhaps AK^2) require an estimate of African gene frequency appreciably different from zero. Within regions, localities are listed in chronological order of the estimates. Standard errors for M were not given (except for reference *6*).

Region* and locality	Gene(s)†	Sample size		M	Reference
		Negro	Caucasian		
	Estimates for M *presumed by their authors to be valid*				
Non-southern					
Baltimore	R^0	907	7,317	0.306	(*7*)
Baltimore	R^0	907	7,317	.216	(*15*)
Five areas	R^0, R^1, Jk^b, *T*, *S*	96 to 3,156	189 to 7,317	~ .20	(*16*)
Cleveland and Baltimore	Gm^1, Gm^5	623	249	.310	(*11*)
Various	R^0, R^1, R^2, *r*, *M*, *S*, Jk^b, *k*, Fy^b			.232–.261	(*17*)
Chicago	AK^2‡	1,063	1,315	.13	(*14*)
Washington, D.C., Baltimore, New York City	R^0, R^1, Fy^a			.20–.24	(*8*)
Oakland, Calif.	Gm^1, $Gm^{1,2}$, Gm^5	260	478	.273±.037	(*6*)
Southern					
Evans and Bullock counties, Ga.	R^0, R^1	340	331	.104	(*9*)
Evans and Bullock counties, Ga.	Gm^1, $Gm^{1,5}$	189	295	.073	(*12*)
	Gc^1	231	292	~ .10	
Charleston, S.C.	R^0, R^1, Fy^a	515		.04–.08	(*8*)
James Island, S.C. and Evans and Bullock counties, Ga.	R^0, R^1, Fy^a	394		.09–.12	(*8*)
	Estimates of M *presumed by their authors to indicate selection*				
Non-southern§					
Four areas, mainly Seattle	Hp^1	936	865(?)	~.40	(*21*)
Seattle	Hp^1	1,657	?	.478	(*8*)
Seattle	Gd^{A-}	658 ♂ ♂		.490	(*8*)
Southern					
Evans and Bullock counties, Ga.	*T*	285	314	.466	(*9*)
	Hp^1‖	167	145	.42–.70	
	Gd^{A-}	76 ♂ ♂		.34–.44	
	Hb^S	247		.46–.69	
	Tf^{D1}‖	133	107	.495	
Memphis	Gd^{A-}	97 ♂ ♂		.175	(*8*)

* An estimate of 0.34, from Hb^S data on 10,858 Negroes, is based on 11 sources in both the North and the South (*38*). It is therefore not placed in a regional category. † Locus and alleles used are as follows. Blood groups: Rh (R^0, R^1, R^2, *r*), Kidd (Jk^b), M-N-S-s (*M*, *S*), Kell (*k*), Duffy (Fy^a, Fy^b); serum protein genes: Gm (Gm^1, $Gm^{1,2}$, Gm^5), haptoglobin (Hp^1), Gc (Gc^1), transferrin (Tf^{D1}); hemoglobin: HbS (Hb^S); red cell enzymes: adenylate kinase (AK^2), glucose-6-phosphate dehydrogenase deficiency (Gd^{A-}); phenylthiocarbamide tasting (*T*). ‡ Newly investigated gene. The African frequency of AK^2 is poorly known, but it is assumed to be zero. The 95-percent confidence interval for M is 0.03–0.23, according to my calculation. § Seven non-southern estimates ranging from 0.270 to 0.685, obtained by Workman (*8*) (using Hp^1 or Gd^{A-}) on small samples (79 to 238 Negroes) are omitted here. ‖ "Possibly" reflecting selection.

An Approach to a More Critical Estimate of M

To constitute a critical estimate in the light of the four criteria listed above, an estimate of M should substantially meet three of them—1, 3, and 4 (2 is, of course, untestable). This means that we should (i) have good survey data on gene frequency from most or all of the seven West African and west-central African slaving areas of Table 1; (ii) be able to calculate a mean African gene frequency properly weighted according to the origins shown in Table 1; (iii) have adequate data on Caucasians and U.S. Negroes; (iv) have samples large enough to give an acceptably small standard error for M; and, very importantly, (v) have some evidence that in U.S. Negroes the gene in question is not subject to strong selection. With regard to points (i) and (ii), an ideal situation is to have a gene which can be shown to be absent or rare in all parts of the slaving area but common in Caucasians. The problem of finding "the" African-ancestor gene frequency is then eliminated, and M is simply q_n/q_c. The Caucasian gene contribution is then directly determinable. It has been claimed that Gm alleles Gm^1, $Gm^{1,5}$, and Gm^5 are of this type (*22*); it is quite likely that they are, but not enough of the slaving area has been surveyed for Gm alleles for us to be sure (*6*).

The Fy^a gene may be almost an ideal "Caucasian gene" for estimating M. Available survey data for regions from Liberia to the Congo (Leopoldville), presented in Table 3, show that in this region (from which about 56 percent of the ancestral slaves came) the mean frequency of Fy^a is probably not over about 0.02. The mean frequency for all Africans of the slave area is probably less than 0.03. The frequency for U.S. Caucasians is about 0.43 (Table 4). Moreover, recent extensive studies in a population of California Negroes revealed no evidence for natural selection (evidence pertaining to fetal and infant growth and viability and to adult growth and fertility) associated with Duffy blood group phenotypes (*23*). Strong selection due to this locus seems excluded, so there is some protection against bias in the estimation of M. Table 4 presents available Fy^a frequency data for U.S. Negroes and for some U.S. Caucasians, and the resulting M estimates. The M estimates for the three non-southern regions studied do not differ significantly, so the estimate 0.2195 ± 0.0093 for California Negroes—the largest of the three samples—may tentatively be used as the best estimate of M for a non-southern area. The very small standard error of this estimate reflects both the discrimination power of this "Caucasian gene" and the large sample sizes for the Negro and Caucasian populations. The two estimates from the "Deep South" do differ significantly and should be kept separate. The smaller one, 0.0366 ± 0.0091 from Charleston, appears to justify the statement that these Gullah Negroes have an unusually small amount of Caucasian ancestry (*5*). It is clear that the data of Table 4 are especially useful in comparing M for different U.S. Negro populations, because the same gene, Fy^a, is used as the basis for all estimates. Any bias due to selection should operate quite similarly in the different Negro populations. The difference between "southern" and "non-southern" M values evident in Table 2 is also marked in Table 4 and must be regarded as real.

Table 3. Frequencies of Duffy blood group Fy (a+) in West African and Congo (Leopoldville) populations.

Region	Sample size (N)	Proportion of Fy (a+)*	Reference
Liberia (many tribes)	661	0.00	(*18*)
Ivory Coast	163	.043†	(*18*)
Upper Volta	75	.00	(*18*)
Dahomey	20	.00	(*18*)
Ghana (Accra) and Nigeria (Lagos)	37	.00	(*39*)
Congo (Bantu)	501	.078‡	(*40*)

* Reacting positively with anti-Fy^a, indicating a genotype of Fy^aFy (most likely), or Fy^aFy^b, or Fy^aFy^a (rare) (*39*). † The true proportion is probably zero because the Ivory Coast positive reactions with anti-Fy^a are believed to be incorrect. ‡ The gene frequency for Fy^a is 0.040.

Thus Fy^a, for the reasons given, may be the best gene presently available for estimating M. When more African survey data are available, the "Caucasian" alleles Gm^1, $Gm^{1,5}$, and Gm^5 of the Gm locus, used jointly, may be as good. The AK^2 gene (Table 2) may be of some use if further African data establish a general zero frequency, but the low frequency, 0.047, of the AK^2 gene in Caucasians considerably reduces its discrimination power. The K gene of the Kell blood group system is sometimes thought of as a "Caucasian gene," but this is not strictly the case. This gene was present in 8 of 1202 Africans from the Liberia-Dahomey (*18*) and western Nigeria (*24*) region, at a mean frequency of 0.0033. The California Negroes of Table 4 ($N = 3146$) have a K gene frequency of about 0.0083, and the California Caucasians, a K gene frequency of about 0.046 (*25*). If we consider q_a to be zero, we obtain an estimate of 0.181 ± 0.026 for M for this population—clearly a maximum estimate and not reliable. This maximum does not differ significantly from the Fy^a estimate for this same population. The relatively large standard error here again reflects the low Caucasian gene frequency.

Although a zero q_a is generally preferable, there is one situation where a q_a value appreciably different from zero might yield a useful estimate of M. This could occur when there are sufficiently extensive and detailed data on African gene frequency to make it possible to calculate a mean African gene frequency, with weighting of regional gene frequencies according to the proportions of Table 1. At present, the ABO blood groups provide the only

Table 4. Estimates of M derived from Fy^a gene frequencies for American Negroes from various areas. The frequency of this gene in the African ancestors of American Negroes is assumed here to be zero; if it is not zero, these are *maximum* estimates. N = number in sample, $q = Fy^a$ gene frequency, S.E. = standard error of q (all estimates by T. E. Reed).

Region and locality	Negroes		Caucasians		$M \pm$ S.E.*	Reference
	N	$q \pm$ S.E.	N	$q \pm$ S.E.		
Non-southern						
New York City	179	0.0809 ± 0.0147			0.189 ± 0.034	(*39*)†
Detroit	404	$.1114 \pm .0114$			$.260 \pm .027$	(*41*)
Oakland, Calif.	3146	$.0941 \pm .0038$	5046	0.4286 ± 0.0058	$.2195 \pm .0093$‡	(*25*)
Southern						
Charleston, S.C.	515	$.0157 \pm .0039$			$.0366 \pm .0091$	(*5*)
Evans and Bullock counties, Ga.	304	$.0454 \pm .0086$	322	$.422 \pm .0224$	$.106 \pm .020$	(*9*)

* The q for Oakland Caucasians (who are of West European ancestry) was used in all estimates. $M = q_n/q_c$. † Two other New York City studies (*42*) are omitted because they involved selection for dark skin color. The data used here were grouped with both anti-Fy^a and anti-Fy^b. The observed distribution of four Duffy phenotypes differs from the Hardy-Weinberg expectation at the 0.025 level of significance. ‡ If the frequency of Fy^a in the African ancestors were 0.02, this estimate would be 0.181.

Table 5. Frequencies of genes *A* and *B* of the ABO blood-group system in surveys in the major slaving areas of Africa (see Table 1); p = frequency of *A* gene, q = frequency of *B* gene.

Region	Peoples or population	Sample size (N)	$p \pm$ S.E.*	$q \pm$ S.E.*	Reference
Senegambia	Bambara, Malinke	2,120	0.159 ± .006	0.174 ± .006	(*43*)
Sierra Leone	Gbah-Mende	1,015	.159 ± .009	.151 ± .008	(*44*)
Liberia	>18 tribes†	2,337	.143 ± .005	.148 ± .006	(*18*)
Gold Coast	Unspecified, from Accra	1,540	.130 ± .006	.122 ± .006	(*45*)
Bight of Benin	Yoruba of Lagos, Ibadan	1,003	.130 ± .008	.141 ± .008	(*46*)
Bight of Biafra	Ibo ("Eastern")	572	.161 ± .011	.089 ± .009	(*47*)
"Angola"	"Bantu"—8000 (mainly Bakongo) near Leopoldville and 8000 from Angola	16,000	.152 ± .002	.138 ± .002	(*48*)
Mean frequencies‡ over the entire slaving area			.150	.131	

* Maximum-likelihood estimate (*49*). † Exclusive of Americo-Liberians. ‡ Calculated from values for p and q given in the body of the table, weighted by the proportions of Table 1 (after the removal of values for Mozambique, Madagascar, and "region unknown").

such usable genetic marker [the gene for hemoglobin S is known to be affected by selection, and much less information is available for other loci (*26*); for selection data on hemoglobin S, see (*27*)]. Table 5 gives the gene frequencies for genes *A* and *B* of the ABO system from relevant surveys in the seven major slaving areas of Table 1.

These extensive surveys reveal an overall uniformity in gene frequency, with the one exception of a somewhat low *B* frequency for the Bight of Biafra (Ibos). From these mean values for African frequencies of genes *A* and *B* and from extensive data on ABO-system distribution in California Negroes and Caucasians (*25*), a maximum likelihood estimate for M of 0.200 ± 0.044 was obtained (*28*). This estimate is not greatly affected by the accuracy of the proportions given in Table 1 or by the exactness of the values for individual regional gene frequencies (*29*). A good fit of the observed number of individuals in each of the eight race and blood-group classes with the corresponding number expected from the estimated parameters (gene frequencies and M values) tested by the chi-square method, indicates both that the estimation is reasonable and that there are no large selective differences between genes *A* and *B* in U.S. Negroes (*28*). This procedure therefore seems justified for the case of ABO blood groups. Practically, however, the large standard error for M indicates that, in spite of large samples, the estimate for this locus is too imprecise to be very useful.

Since there are now three different estimates of M, and since extensive data on other aspects of the problem, including selection, are available for this one large California population of Negroes, these estimates are presented in a single table, Table 6. We note that they do not differ significantly from each other; this is due at least in part to the relatively large standard errors for the Gm and ABO estimates. The marked superiority, for estimating M, of Fy^a over *A* and *B* for samples of equal size is evident (*30*), whereas, if the sample sizes were the same for Fy^a and the three Gm alleles, it would be found that these are equally efficient for estimating M. An extensive search for evidence of natural selection due to the presence of ABO blood groups in these Negroes, similar to the search reported above for the Duffy blood group, also failed to reveal any consistent selective effect (*23*). This finding, plus the good chi-square fit in the estimation of M, which implies that the *A* and *B* genes are not very different with respect to their selective values in U.S. Negroes, gives some assurance that the ABO estimate is not greatly disturbed by selection (*28*). No selection studies for Gm were made on these California Negroes, but extensive studies on a Brazilian population which was about 30 percent Negro, 11 percent Indian, and 59 percent Caucasian (*13*) revealed no evidence of selective effect (*31*). Further evidence is provided by the good chi-square fit in the multi-allelic estimation obtained with the three Gm alleles (*6*). It seems reasonable to conclude that strong selective effects on these three estimates of M may be excluded. The existence of weaker effects, however, still sufficient to bias these estimates appreciably, cannot be ruled out. As more independent estimates on these and other genes become available, each having regard to the criteria listed above and including some safeguard against a strong bias due to selection and having a relatively small standard error (say, less than 0.02), it should become possible to obtain a "consensus" on the true value of M (for specified Negroes). Estimates biased upward or downward by selection will be separated from those little affected by selection, and so, in time, the former can be identified and rejected.

Table 6. Estimates of M calculated from data on Gm serum groups, Duffy blood group, and ABO blood group from Negroes and Caucasians of the Oakland, California, area. [Data of the Child Health and Development Studies (*6*, *25*).]

Locus	Alleles used	Sample size (N)		M
		Negroes	Caucasians	
Gm	Gm^1, $Gm^{1,5}$, Gm^5	260	478	0.273 ± 0.037*
Duffy	Fy^a	3146	5046	.220 ± .009†
ABO	*A*, *B*	3146	5046	.200 ± .044†

* See (*6*). † See text.

Use of M To Detect Selection

Several investigators (*8*, *9*, *21*, *32*) have argued that selection for or against a gene may be clearly inferred from the M value that the gene produces. From the foregoing section it is clear that if (i) the true (unbiased) value of M (say, M_0) is known, (ii) the estimate in question (M_e) is calculated with regard to the criteria given above, and (iii) M_e differs significantly from M_0, then we may reasonably suspect that selection has caused the observed deviation. These conditions have not been met. In particular, we have no M_0. The M estimates obtained with R^0 (*8*, *9*, *21*), R^1 (*8*, *9*), and Fy^a (*8*) were considered to be valid estimates unbiased by selection, but no objective evidence was offered to support these views. With one or more of these M estimates used as reference standards, it has been claimed that the deviant M estimates of the following genes demonstrate selection on these genes in U.S. Negroes: Hp^1, *T*, Gd^{A-}, Hb^S, and Tf^{D1} (see Table 2). These results can, at present, be considered only suggestive, but it must be admitted that the usually high M estimates obtained with Hp^1 and Gd^{A-} argue for an effect of selection (*27*).

A different approach was used to

show that M estimates obtained with r, R^0, and R^1 alleles of the Rh locus ranked in this (decreasing) order of size for a Georgia population and also for two Brazilian populations (*32*). Accepted at face value, this is evidence of differences between M values from different Rh alleles. The investigators attribute these differences to selection. This same approach in these populations also indicates that M for the B allele is greater than $1.5M$ for the A allele (*32*). African Rh and ABO gene frequencies, weighted by slaving-area origins, were not used, however, although the African areas of origin of Brazilian Negroes are known (*2*). Again, these findings are interesting and suggestive but far from conclusive.

Workman (*8*), from inspection of A_1, A_2, and B allele frequencies in various West African, U.S. Negro, and U.S. Caucasian populations, concludes that there has been strong selection in U.S. Negroes against A_1 and for A_2. He identifies the various African data only as "West Africa," and does not use significance tests. Since Workman and also Hertzog and Johnson claim to find selection in the ABO system, it is pertinent here to recall that the M estimate obtained from ABO-system distributions that is discussed earlier in this article (an estimate based on *large* populations and good estimates for African gene frequency) did not suggest selective differences between the A and B alleles.

This critical review of claims for selection would be incomplete if I did not mention that there *is* an important theoretical reason to look for selection in hybrid populations such as the American Negro. As has been previously recognized (*6, 8, 32*), selection in U.S. Negroes over several generations can produce a cumulative effect in present-day individuals appreciably greater than the effect of a single generation of selection—the type of data usually available. There is thus a possibility of detecting, in hybrids, selection due to common polymorphisms which is too small [usually less than 5 to 10 percent of the mean (*23*)] to be detectable by ordinary one-generation studies. This possibility, together with the probability that some of the genes are selective [because it is most unlikely that a new genotype (the hybrid) in a new environment would be exactly neutral in selective value], makes the search for selection here especially worthwhile. Some of these selective genes may already have been identified.

Other Uses of M Estimates

In addition to the definite likelihood of their yielding valuable information on the action of natural selection in human populations, good estimates of the amount of Caucasian ancestry in U.S. Negro populations have at least two other "uses."

1) They provide objective information about the genetic heterogeneity among various populations of U.S. Negroes. Evidence of marked differences between southern and non-southern Negroes with respect to the amount of Caucasian ancestry, as shown in Tables 2 and 4, is the first clear result from this use of M estimates. As more good estimates from defined U.S. Negro populations become available, we may expect further heterogeneity to be revealed.

2) They provide an understanding of the distribution in American Negroes of those genetic traits, including diseases, that are due primarily to genes of Caucasian origin. There are few examples of such genes at present, but, aside from common genetic polymorphisms, like blood groups, few genes have been sufficiently studied to permit possible identification of racial differences in gene frequency. One probable example of such a genetic trait is phenylketonuria—a condition resulting from homozygosity for a rare autosomal recessive gene, producing a deficiency of phenylalanine hydroxylase and resulting (if untreated) in severe mental defect. This occurs in about 1 in 10,000 births of persons of North European ancestry (*33*) but appears to be much rarer in U.S. Negroes (*34*). This rarity is understandable if the gene frequency in African Negroes is much lower than that in Caucasians (about 0.01). For example, if U.S. Negroes have, on the average, 20-percent Caucasian ancestry, the frequency of occurrence of phenylketonuria at birth in U.S. Negroes would be only 1/25th that in Caucasians, or roughly 1 in 250,000—rare indeed.

An example of a disease which is not simply inherited but which may show a similar racial distribution is cirrhosis of the liver. A study in Baltimore Negro cirrhotics revealed, relative to Negro controls, a significant increase in Fy (a+b+) Duffy blood group phenotype and a decrease in Fy (a−b−) phenotype, whereas Caucasian cirrhotics showed no such difference from Caucasian controls (*35*). The simplest interpretation is that the disease is more frequent in Caucasians, and that Negroes with some degree of Caucasian ancestry, as shown by their Duffy blood group, are more likely to develop the disease than those lacking such ancestry (*35*). Other examples of traits whose frequency of occurrence in U.S. Negroes is affected by the amount of their Caucasian ancestry will surely be reported (*36*). Accurate information on M will be clinically useful here.

Summary

Published estimates of the proportion, in American Negroes, of genes which are of Caucasian origin are critically reviewed. The criteria for estimating this proportion (M) are discussed, and it is argued that all estimates published to date have either deficiencies pertaining to the African-gene-frequency data used or statistical inaccuracies, or both. Other sources of error may also exist.

Evidence is presented that the Fy^a gene of the Duffy blood group system may be the best gene now available for estimating M. Estimates based on Fy^a frequencies have been obtained for Negroes in three non-southern and two southern areas. The value of M is found to be appreciably greater in non-southern areas, the best estimate being 0.2195 ± 0.0093 (Oakland, California). This estimate is still subject to some uncertainty. The value of M in the South is appreciably less.

Natural selection can introduce a bias in the estimate of M. Claims that selection acting on certain genes in American Negroes have been demonstrated are reviewed, and it is concluded that they are not yet proved. The approach discussed here may be valuable in the future as a sensitive method for detecting the action of natural selection. In addition, knowledge of the amount of Caucasian ancestry may be of medical value in explaining the frequencies of occurrence of certain hereditary diseases in Negroes.

References and Notes

1. J. H. Franklin and T. Marshall, in *World Book Encyclopedia* (Field Enterprises Educational Corporation, Chicago, 1966), vol. 14, p. 106.
2. P. D. Curtin, personal communication (1969).
3. J. Potter, in *Population in History: Essays in Historical Demography*, D. V. Glass and D. E. Eversley, Eds. (Univ. of Chicago Press, Chicago, 1965), p. 641.
4. P. D. Curtin, personal communication (1968).
5. W. S. Pollitzer, *Amer. J. Phys. Anthropol.* **16**, 241 (1958).
6. T. E. Reed, *Amer. J. Hum. Genet.* **21**, 71 (1969).
7. B. Glass and C. C. Li, *ibid.* **5**, 1 (1953).

8. P. L. Workman, *Hum. Biol.* **40**, 260 (1968).
9. ———, B. S. Blumberg, A. J. Cooper, *Amer. J. Hum. Genet.* **15**, 429 (1963).
10. C. Stern, *Principles of Human Genetics* (Freeman, San Francisco, ed. 2, 1960), p. 356.
11. A. G. Steinberg, R. Stauffer, S. H. Boyer, *Nature* **188**, 169 (1960).
12. B. S. Blumberg, P. L. Workman, J. Hirschfeld, *ibid.* **202**, 561 (1964).
13. H. Krieger, N. E. Morton, M. P. Mi, E. Azevêdo, A. Freire-Maia, N. Yasuda, *Ann. Hum. Genet.* **29**, 113 (1965).
14. J. E. Bowman, H. Frischer, F. Ajmar, P. E. Carson, M. K. Gower, *Nature* **214**, 1156 (1967).
15. B. Glass, *Amer. J. Hum. Genet.* **7**, 368 (1955) (non-D^u-tested data for Africans, Negroes, and Caucasians; use of D^u-tested data for Africans and non-D^u-tested data for others gives $M = 0.281$).
16. D. F. Roberts, *ibid.*, p. 361. The estimate is "provisional"; 66 separate estimates were made, ranging from 0.0404 to 0.3341.
17. ——— and R. W. Hiorns, *ibid.* **14**, 261 (1962). No sample sizes are specified.
18. F. B. Livingstone, H. Gershowitz, J. V. Neel, W. W. Zeulzer, M. D. Solomon, *Amer. J. Phys. Anthropol.* **18**, 161 (1960).
19. P. V. Tobias, in *The Biology of Human Adaptability*, P. T. Baker and J. S. Weiner, Eds. (Clarendon Press, Oxford, 1966), p. 161.
20. The following formula for the standard error (S.E.) of a ratio $R = y/x$ was used:

$$\text{S.E. } R = R\left[\frac{V_y}{y^2} + \frac{V_x}{x^2} - \frac{2C_{xy}}{xy}\right]^{1/2}$$

where the variance of y is V_y, that of x is V_x and the covariance between x and y is C_{xy} [see, for example, L. Kish, *Survey Sampling* (Wiley, New York, 1965), p. 207]. This formula is adequate for large or moderate-sized samples when it is unlikely that x is near zero. In terms of Eq. 2 for M,

$$\text{S.E. } M = M\left[\frac{V(q_a - q_n)}{(q_a - q_n)^2} + \frac{V(q_a - q_c)}{(q_a - q_c)^2} - \frac{2V(q_a)}{(q_a - q_n)(q_a - q_c)}\right]^{1/2}$$

where V represents the variance of the adjoining quantity in parentheses. The covariance between numerator and denominator of Eq. 2, due to the presence of q_a in both, is allowed for in this standard error.
21. W. C. Parker and A. G. Bearn, *Ann. Hum. Genet. London* **25**, 227 (1961). The total number of American, Canadian, and British individuals tested in the study reported is 865. A weighted estimate for the frequency of gene Hp^1 in U.S. Negroes, based on the data of Parker and Bearn, is 0.55; this value gives an M of 0.53.
22. A. G. Steinberg, in *Symposium on Immunogenetics*, T. J. Greenwalt, Ed. (Lippincott, Philadelphia, 1967), pp. 75–98.
23. T. E. Reed, *Amer. J. Hum. Genet.* **19**, 732 (1967); *ibid.* **20**, 119 (1968); *ibid.*, p. 129.
24. B. S. Blumberg, E. W. Ikin, A. E. Mourant, *Amer. J. Phys. Anthropol.* **19**, 195 (1961).
25. T. E. Reed, *Amer. J. Hum. Genet.* **20**, 142 (1968).
26. For gene distributions, see J. Hiernaux, *La Diversité Humaine en Afrique Subsaharienne* (Institut de Sociologie, Université Libre de Bruxelles, Brussels, 1968), figs. 2, 3, 7, 8, 12, 14.
27. There are good a priori reasons, entirely separate from M values, for expecting, in U.S. Negroes, a decrease in the frequency of the genes for sickle-cell hemoglobin, Hb^S, and for glucose-6-phosphate dehydrogenase deficiency, Gd^{A-}. (i) There is good evidence that in Africa the high frequency of the Hb^S gene is due to a selective advantage of heterozygotes for Hb^S in regions where malaria is endemic [see, for example, F. B. Livingstone, *Abnormal Hemoglobins in Human Populations* (Aldine, Chicago, 1967), pp. 105–107; A. C. Allison, in *Abnormal Haemoglobins in Africa*, J. H. P. Jonxis, Ed. (Davis, Philadelphia, 1965), pp. 369–371; D. L. Rucknagel and J. V. Neel, in *Progress in Medical Genetics*, A. G. Steinberg, Ed. (Grune & Stratton, New York, 1961), vol. 1, pp. 158–260]. There is strongly suggestive evidence that the Gd^{A-} gene in Africa is similarly kept at high frequencies due to selective advantage in malarious areas [see F. B. Livingstone, *Abnormal Hemoglobins in Human Populations* (Aldine, Chicago, 1967); A. G. Motulsky, in *Abnormal Haemoglobins in Africa*, J. H. P. Jonxis, Ed. (Davis, Philadelphia, 1965), pp. 181–185)]. (ii) Both genes are known to have selective disadvantages which can explain their rarity in nonmalarious areas. It is therefore to be expected that Negroes moved from their highly malarious homelands to the less malarious, and now nonmalarious, regions of North America would have lower frequencies of these two genes. This selective decrease would raise M estimates above the true value.
28. The computer program [see T. E. Reed and W. J. Schull, *Amer. J. Hum. Genet.* **20**, 579 (1968)] estimated M and Caucasian A and B gene frequencies, given the two African mean frequencies as constants and the two California populations determined by the gene frequencies to be estimated, subject to the constraints that, for *both* A and B, $q_n = Mq_c + (1 - M)q_a$. This equation is Eq. 1 applied to both alleles and is true when there is simple gene mixture without selection (see *6*). Comparison of the observed numbers of the eight race and blood-group classes (2 races × 4 groups) with the corresponding numbers expected on the basis of parameter estimates gives a chi-square value of 5.910 for 3 d.f., $P > .10$.
29. When the negligible contribution from Mozambique, Madagascar, and "Unknown" is excluded, the proportions of Table 1, column 4, become (in order): 0.135, 0.056, 0.116, 0.162, 0.044, 0.237, and 0.249. The corresponding proportions for South Carolina (1773–1807) are 0.197, 0.068, 0.164, 0.134, 0.016, 0.021, and 0.399 [data of Curtin (*4*)], yielding overall African mean values of 0.149 and 0.144 for p and q. These two series differ appreciably with respect to the final two values, yet when the South Carolina series is used, the estimate of M is 0.256 ± 0.042, a difference of just over one standard error. Also, q for the Bight of Biafra is the only markedly variant gene frequency among the frequencies for the seven regions, but replacing the p and q for this region by the p and q for the Bight of Benin or for "Angola" does not significantly change M (0.281 ± 0.040 or 0.251 ± 0.042, respectively, when corrected proportions of Table 1 are used).
30. The Fy^a estimate is based on $(0.044)^2/(0.0093)^2$, or 22 times as much statistical information as the ABO estimate.
31. N. E. Morton, H. Krieger, M. P. Mi, *Amer. J. Hum. Genet.* **18**, 153 (1966).
32. K. P. Hertzog and F. E. Johnston, *Hum. Biol.* **40**, 86 (1968).
33. V. A. McKusick, *Mendelian Inheritance in Man* (Johns Hopkins Press, Baltimore, ed. 2, 1968), p. 346.
34. For example, H. P. Katz and J. H. Menkes [*J. Pediat.* **65**, 71 (1964)] report the first definite case of phenylketonuria in a U.S. Negro. R. G. Graw and R. Koch [*Amer. J. Dis. Child.* **114**, 412 (1967)] report two "pale skinned" Negro brothers with phenylketonuria, bringing the total for U.S. Negroes at that time to five.
35. N. C. R. W. Reid, P. W. Brunt, W. B. Bias, W. C. Maddrey, B. A. Alonso, F. L. Iber, *Brit. Med. J.* **2**, 463 (1968).
36. Differences between Caucasians and Japanese with respect to gene frequency are instructive here. Phenylketonuria appears to be much rarer (perhaps a tenth as frequent) among Japanese than among Caucasians [K. Tanaka, E. Matsunaga, Y. Hanada, T. Murata, K. Takehara, *Jap. J. Hum. Genet.* **6**, 65 (1961)]. Another single-gene trait, Huntington's chorea, also appears to be about ten times as frequent in Caucasians as in Japanese, according to T. E. Reed and J. H. Chandler, *Amer. J. Hum. Genet.* **10**, 201 (1958). Examples of congenital abnormalities which are commoner in Caucasians than in Japanese, and vice versa, are given by J. V. Neel, *Amer. J. Hum. Genet.* **10**, 398 (1958).
37. P. D. Curtin, *The Atlantic Slave Trade: A Census* (Univ. of Wisconsin Press, Madison, in press).
38. J. V. Neel and W. J. Schull, *Human Heredity* (Univ. of Chicago Press, Chicago, 1954), p. 255; J. V. Neel, personal communication (1969). The estimate refers to "non-Negro" ancestry.
39. R. R. Race and R. Sanger, *Blood Groups in Man* (Blackwell, Oxford, ed. 5, 1968).
40. M. Shapiro and J. M. Vandepitte, *Int. Congr. Blood Transfusion, 5th, Paris* (1955), p. 243; M. Shapiro, personal communication (1969).
41. H. Gershowitz, unpublished data.
42. E. B. Miller, R. E. Rosenfield, P. Vogel, *Amer. J. Phys. Anthropol.* **9**, 115 (1951); R. Sanger, R. R. Race, J. Jack, *Brit. J. Haematol.* **1**, 370 (1955).
43. R. Koerber, *C. R. Seances Soc. Biol.* **141**, 1013 (1947); R. Koerber and J. Linhard, *Bull. Soc. Anthropol. Paris* **2**, 158 (1951). The frequencies for Bambara and Malinke do not differ significantly.
44. P. Julien, *Z. Rassenphysiol.* **9**, 146 (1937).
45. G. M. Edington, *West Afr. Med. J.* **5**, 71 (1956).
46. H. R. Muller, *Proc. Soc. Exp. Biol.* **24**, 437 (1927); J. P. Garlick, quoted in A. E. Mourant, A. C. Kopeć, K. Domaniewska-Sobczak, *The ABO Blood Groups* (Blackwell, Oxford, 1958), p. 173; J. P. Garlick and N. A. Barnicot, *Ann. Hum. Genet.* **21**, 420 (1957). The frequencies in these three surveys do not differ significantly from each other.
47. J. Hardy, *Roy. Anthropol. Inst.* **92**, 223 (1962). I have not used Hardy's data on "Onitsha Ibo" (sample size, 228) because I consider the subjects to be probably not of Ibo origin. I have not used data of J. N. M. Chalmers, E. W. Ikin, and A. E. Mourant [*Ann. Eugenics* **17**, 168 (1953)] on "southeastern" Nigeria (105 Ibo and 1 Tiv) because information on the type of Ibo was not given. Table 1 describes this region as "about ¾ Ibo," and I could find no suitable data for the remaining quarter.
48. G. van Ros and R. Jourdain, *Ann. Soc. Belge Med. Trop.* **36**, 307 (1956); L. Mayor, *Bull. Clin. Statistics* **7**, No. 3, suppl. 126 (1954). The first study (for the Congo) gave values of 0.1556 ± 0.0030 for p and 0.1244 ± 0.0027 for q. The second (for Angola gave values of 0.1486 ± 0.0029 and 0.1514 ± 0.0030, respectively. An unweighted average of these values was used. The values for Angola do not differ significantly from the mean p and q values for seven central-coastal named tribes in Angola (total population, 1285) tabulated by Hiernaux (see *26*).
49. T. E. Reed and W. J. Schull, *Amer. J. Hum. Genet.* **20**, 579 (1968).

50. Preparation of this article was begun while I was engaged in work for the Child Health and Development Studies (Division of Biostatistics, School of Public Health, University of California, Berkeley, and the Kaiser Foundation Research Institute, Oakland, California), on leave from the University of Toronto, and was supported there by U.S. Public Health Service research grants HD 00718 and HD 00720 from the National Institutes of Health. The analysis was supported in part by a grant from the Medical Research Council of Canada. I thank Professor Philip D. Curtin for making unpublished data available and for commenting on the manuscript, Dr. Arthur E. Mourant and Mrs. K. Domaniewska-Sobczak for recent references to African blood group distributions, and Professors Curt Stern, Donald Rucknagel, and Peter Carstens for their comments. Dr. H. Gershowitz and Dr. M. Shapiro made available unpublished data on Duffy blood groups in Negroes.