#7

2

The Use of Genetic Markers of the Blood in the Study of the Evolution of Human Populations

MICHAEL H. CRAWFORD

Department of Anthropology
University of Kansas

Since Landsteiner's discovery in 1900 of the ABO blood group system in man, there has been a rapid proliferation of new genetic markers[1] for use in population studies. At present, nearly 50 polymorphic gene systems with more than 140 common alleles, or chromosomal segments—each with a populational frequency greater than, or equal to, 0.01—have been discovered by the examination of human blood. These marker genes and their dates of discovery are summarized in Table 2.

The discussion in this chapter will be restricted to genetic markers of the blood that are available for population studies, and will describe their use in interpretations of evolutionary theory and the analyses of phylogenetic relations between populations. The question of how these genetic polymorphisms may be maintained will not be dealt with here. (See Chapter 3 for a discussion of several explanatory models.) The purpose of this chapter is (1) to review the technical and methodological innovations that have enabled the use of an increased number of genetic

TABLE 2
POLYMORPHIC LOCI OF HUMAN BLOOD

Genetic Marker System	Year of Discovery	Common Alleles
A. Blood groups		
1. ABO	1900	I^A, I^O, I^B
A_1A_2 subdivision	1911	I^{A1}, I^{A2}
2. MN	1927	L^M, L^N
MNS subdivision	1947	*Ms*, *MS*, *Ns*, *NS*
Uu	1953	*U*, *u*
3. P	1927	P_1, P_2
4. Rh	1940	*CDE*, *cde*, *CDe*, *cDE*, C^wDe, *Cde*, *cDe*
5. Lutheran	1945	Lu^a, Lu^b
6. Lewis	1946	Le^a, Le^b
7. Kell	1946	*K*, *k*
Kp subdivision	1956	Kp^a, Kp^b
Sutter subdivision	1959	Js^a, Js^b
8. Duffy	1950	Fy^a, Fy^b, *Fy*
9. Kidd	1951	Jk^a, Jk^b
10. Vel	1952	Ve^a, Ve^b
11. Diego	1955	Di^a, Di^b
12. Cartwright	1956	Yt^a, Yt^b
13. Auberger	1961	*Au*, Au^a
14. Xg	1962	*Xg*, Xg^a
15. Dombrock	1965	Do^a, *Do*
16. Stoltzfus	1969	Sf^a, *Sf*
B. Red cell proteins		
1. Hemoglobin	1947	Hb^A, Hb^S, Hb^O, Hb^D, Hb^E
Thalassemia		*T*, t^i, t^o
2. Glucose-6-phosphate dehydrogenase (G6PD)		
Deficiency	1958	*Gd+*, *Gd–*
Structural	1962	Gd^A, Gd^B
3. Acetyl transferase	1958	Ac^S, *Ac*
4. Acid phosphatase	1963	P^a, P^b, P^c
5. 6-phosphogluconate dehydrogenase (6PGD)	1963	PGD^A, PGD^C
6. Phosphoglucomutase (PGM)	1964	$PGM_1{}^1$, $PGM_1{}^2$
7. Adenylate kinase	1966	AK^1, AK^2
8. Adenosine diaminase	1968	ADA^1, ADA^2
9. Phosphoglycerate kinase (PGK)	1970	PgK^1, PgK^2
10. Peptidase C	1970	Pep C^1, C^2, C^0
11. Glutamic—pyruvic transaminase	1971	Gpt^1, Gpt^2

TABLE 2, cont'd.

Genetic Marker System	Year of Discovery	Common Alleles
C. Serum proteins		
1. Haptoglobins	1955	Hp^1, Hp^2
Subdivision of Hp^1	1962	Hp^{1s}, Hp^{1f}
2. Gm groups	1956	1, 5, 13, 14; 1, 5, 6, 14; 1, 5, 6; 1, 13; 1; 1, 5, 14; 1, 5; 3, 5, 13, 14; 1, 3, 5, 13, 14; 1, 2
3. Transferrins	1957	Tf^B, Tf^O, Tf^D
4. Pseudocholinesterase (E_1)	1957	$E_1{}^u$, $E_1{}^a$, $E_1{}^A$, $E_1{}^F$
(E_2)	1963	E_2+, E_2-
5. Albumin	1959	AL^A, AL^B
6. Group specific (G_c)	1959	Gc^1, Gc^2
7. Inv groups	1961	Inv^1, $Inv^{1,2}$, Inv^3
8. Beta lipoprotein allotypes	1961	Ag^a, Ag^x
Lp system	1963	Lp^a, *Lp*
9. Placental alkaline phosphatase	1961	PL^F, PL^S
10. Xm group	1966	Xm^a, *Xm*
D. White cells and platelets		
1. Ko platelet groups	1962	Ko^a, Ko^b
2. Zw platelet groups	1963	Zw^a, Zw^b
3. HLA		
LA series	1964	*HL-A1, A2, A3, A9* 3002Q
4 series	1961	*HL-A12, 4C, A7, A8*

markers for population studies, (2) to indicate the extent of known genetic variability in man, (3) to examine the uses and limitations of data on gene frequency distributions, and (4) to suggest possible means by which marker genes may be used for future populational studies in genetics and physical anthropology.

METHODOLOGICAL INNOVATIONS

A number of methodological breakthroughs have contributed to the abundance of genetic markers presently available for use in the study of microevolution in human populations. The following are among the most significant recent developments.

(1) Zone electrophoresis permits the separation of proteins on the basis of molecular size, configuration, and charge. This teechnique has been applied primarily to the detection of genetic variants of red cell or serum proteins in human populations. Although electrophoresis has serious limitations (for example, in failing to distinguish the three types of hemoglobin that possess the same electrophoretic mobility as hemoglobin *D*), it is a valuable screening device for certain protein variants and has led to the recognition of a large number of polymorphic loci. It was this method that led to the recognition of sickle-cell anemia as a hemoglobinopathy resulting from a structural alteration.

(2) Some of the limitations of zone electrophoresis have been overcome by the development of chromatographic techniques and amino acid analysis. Used in conjunction with electrophoresis, these techniques permit the recognition of structural differences between proteins possessing similar or identical electrophoretic patterns. For example, several hemoglobins, on the basis of their electrophoretic patterns, were initially classified as hemoglobin *D*. However, amino acid analyses revealed that these hemoglobins differed structurally and resulted in their reclassification as $D\alpha$, $D\beta$, and $D\gamma$. These techniques are also sufficiently refined to establish that some proteins, formerly assumed to be different because of the distance of the populations in which they were found, are in fact identical. For example, two transferrin variants with similar electrophoretic patterns, *Tf*, D_1, one discovered in an Australian aborigine and the second found in an American Black, were regarded as different because of the geographic (and presumed genetic) distance between the two individuals. However, a comparison by enzyme digestion, followed by paper electrophoresis, chromatography, and amino acid analysis, suggested that the two variants were identical in structure as well as in mobility (Wang, Sutton, and Scott 1967). These findings have raised the question of the origin of the allele. Did the two variants result from two independent mutations, or are they the result of the common ancestry of Australian and African populations?

(3) Improvements in the methods of transport and storage of human blood have permitted the genetic description of geographically remote and isolated populations. Most striking among these is the use of liquid nitrogen for the storage of red blood cells for a number of years, with little loss of enzyme activity. It is now possible to store blood samples from several populations in liquid nitrogen tanks and to retrieve and analyze

the specimens whenever a new genetic polymorphism or variant is discovered. Thus, the investigator has access to a number of populations, from different geographical areas, in the laboratory. Although the preparation of the red blood cells for storage in liquid nitrogen is much more time-consuming and costly, this procedure is vastly superior to the citrated glycerol method of storing red blood cells. Storage in citrated glycerol results in considerable loss of enzyme activity, and a high percentage of erythrocytes are lost during the saline washing phase of cellular retrieval.

(4) Within the last decade, several rapid biochemical screening procedures have been developed for use in the field when dealing with blood enzymes that are unstable and subject to a rapid loss of activity or denaturation during prolonged periods in transit. The low cost and technical simplicity of these screening procedures are important considerations when compared with the logistics of shipping blood specimens in tanks of liquid nitrogen.

One of the most commonly used screening tests was developed by Motulsky and Campbell-Kraut (1960) for determining the presence of G6PD deficiency in red blood cells. This method is based on the decolorization of a brilliant cresyl blue dye in a prescribed period of time. It is sufficiently sensitive to reveal females who are heterozygous for the enzyme deficiency. This method does not require elaborate equipment and therefore can be used in the field for rapid identification of individuals who are G6PD deficient. Another factor in favor of such screening procedures is the small quantity of blood required, thus eliminating the need for venipuncture.

(5) Developments in computer technology have facilitated estimations of gene and chromosomal segment frequencies in loci with multiple alleles or segments. Maximum likelihood (ML) estimation computer programs, initially written by Balakrishnan and Sanghvi (1965), Kurczynski and Steinberg (1967), MacCluer et al. (1967), and Reed and Schull (1968), provide more accurate estimation of the *Rh* and *MNS* chromosomal segment frequencies in population studies. The computer program of Reed and Schull makes no prior assumptions concerning the presence or absence of alleles, or chromosomal segments, in contrast to the commonly employed method of Mourant (1954). Mourant's method assumes, in the absence of an appropriate indicator phenotype, that a particular gene is not present in the population. Gershowitz et al. (1970) compare the *Rh* gene frequencies for three Makiritare villages using five

different methods of estimating allelic frequencies, and only the eight-gene computer program of Reed and Schull assigns a frequency to the *r(cDe)* gene. However, the other methods of calculation of the *Rh* gene frequencies (Mourant 1954; Schull's four-gene maximum likelihood program; Layrisse, Layrisse, and Wilbert 1963; and a gene counting method using pedigree analysis) are in much closer agreement with each other than with the method of Schull and Reed. These findings present a dilemma: "Which genes and which frequencies shall be accepted as a proper description of the population?" (Gershowitz et al. 1970). Basically, this question contrasts the goodness-of-fit of the observed phenotypic frequencies with those predicted by the derived gene frequencies. Although both approaches are adequate, the Reed and Schull method may be better, because it makes no assumptions about the presence or absence of any of the eight alleles theoretically possible and detectable with the five major *Rh* antisera commercially available. However, this theoretical question is probably of little relevance to field studies, because the differences between the various estimates are probably much less than the sampling error resulting from nonrandom selection of a sample population.

GENETIC VARIATION IN HUMAN POPULATIONS

On the basis of the number of the so-called private alleles, Lewontin (1967) has estimated that the proportion of polymorphic loci in man is 30 percent. This is probably an overestimate, since in the future, it may be shown that a great proportion of loci are familial (that is, less than 2 percent of the members of a population will be heterozygous for a given gene, a necessary prerequisite for classification as a polymorphism). However, this approximation was further supported by Harris (1966) who "randomly" selected 10 red cell enzyme loci and found that three of these loci maintained variants at polymorphic levels. Assuming that man has at least 120,000 loci encoding for protein structure (a conservative estimate made by Neel and Schull [1968] assumed the existence of 100,000), and further assuming that one locus controls the synthesis of one polypeptide chain and that an "average" protein possesses four polypeptide chains, man may possess up to 30,000 proteins, of which 10,000 would be

structurally polymorphic. Those few polymorphic loci presently known provide only a small indication of the spectacular human diversity that must have been etched on the genome of man through time.

The proliferation of genetic markers in the last two decades has revealed great phenotypic variability both within and between populations. Giblett (1969) applied Race and Sanger's (1962) method of paternity exclusion to 17 genetic markers and showed that in western Europe or in the White portion of the United States population, less than one in 350,000 people would be expected to have the same combination of phenotypes. By considering not just 17 genetic markers, but 10,000, it is possible to account for all of the molecular genetic individuality of the hominids in existence for the last million years.

In a recent review article, Neel and Schull (1968:566) state that genetic polymorphisms "have now reached such numbers and new ones are being recognized at such a rate, that the understanding of the significance is unquestionably a focal problem in modern biomedicine." With this realization comes the necessity to explain how and why this genetic variation is maintained in human populations. As a result, it is not surprising to find that much of the research in physical anthropology and human genetics is presently directed to the elucidation of these questions. In the last decade, there has been much activity in the study of genetic markers and their frequencies in human populations, because the markers permit the description and quantification of the degree of genetic variation in a population. For analytical purposes, this variation can be viewed either on the intrapopulational or the interpopulational level, although these levels are interrelated.

INTRAPOPULATIONAL VARIATION

Populations are often divided into various subunits, such as bands, hordes, or villages. These may be either endogamous or exogamous, the latter being part of larger social units such as tribes, communities, or nations, which are usually the units of genetic analysis, or Mendelian populations, and are inclusive in terms of mate selection.

In several studies, new genetic markers have been used to study intrapopulation variability in man. Salzano (1968) compared the intratribal genetic variability of the Caingang and the Xavante Indians of Brazil

with the intertribal variation among the South American Indians. Sixteen genetic markers were used for a comparison of seven Caingang settlements and three Xavante bands. Salzano found that the range of gene frequencies observed among the Caingang was larger than the range observed among the Xavante groupings (see Figure 6 and Table 3). Caution, however, must be exercised in interpreting these data, because of the disparity of sample numbers and the varying sizes of the subpopulations. The observed ranges of gene frequencies for six markers did not show overlapping ranges, suggesting that the two tribes can be distinguished genetically.

Moreover, there is great difficulty in interpreting results of genetic studies in terms of the demographic or social structure of a population. Workman and Niswander (1970) applied Wright's F statistics to the gene frequencies of the 10 subdivisions (districts) of the Papago Indian tribe of the southwestern United States. The degree of heterogeneity among the districts was expressed by means of one of the F statistics, F_{ST}, which is based on the ratio of the actual variance in gene frequencies among the groups to the product of the weighted mean gene frequencies. Thus, $F_{ST} = \sigma^2 p/(\bar{p}\bar{q})$, where $(\bar{p}\bar{q})$ are the weighted means and $\sigma^2 p$ is the variance of the gene frequencies.

A comparison of the degree of heterogeneity between the Papago and the Yanomama (of Venezuela) revealed that Yanomama villages were more heterogeneous genetically than the Papago districts. The comparison is based on the F ratio, made up of the sum of the χ^2 ($\Sigma\chi^2$) of the F_{ST} values for each of the bands. Further comparisons of the Papago with the Xavante and the Caingang Indians revealed that the Papago districts have significantly greater heterogeneity than either the Caingang or the Xavante. The F ratios permit comparisons of the degrees of differentiation of the subpopulations within a Mendelian population, as Workman and Niswander (1970:47) conclude: "These studies also show that demonstration of significant intra-population heterogeneity indicates nothing about the real structure of a population, past or present."

INTERPOPULATIONAL VARIATION

The evidence provided by marker genes of the high degree of interpopulational variability in man has raised more questions than it has answered. Salzano's (1968) compilation of the gene frequency distribu-

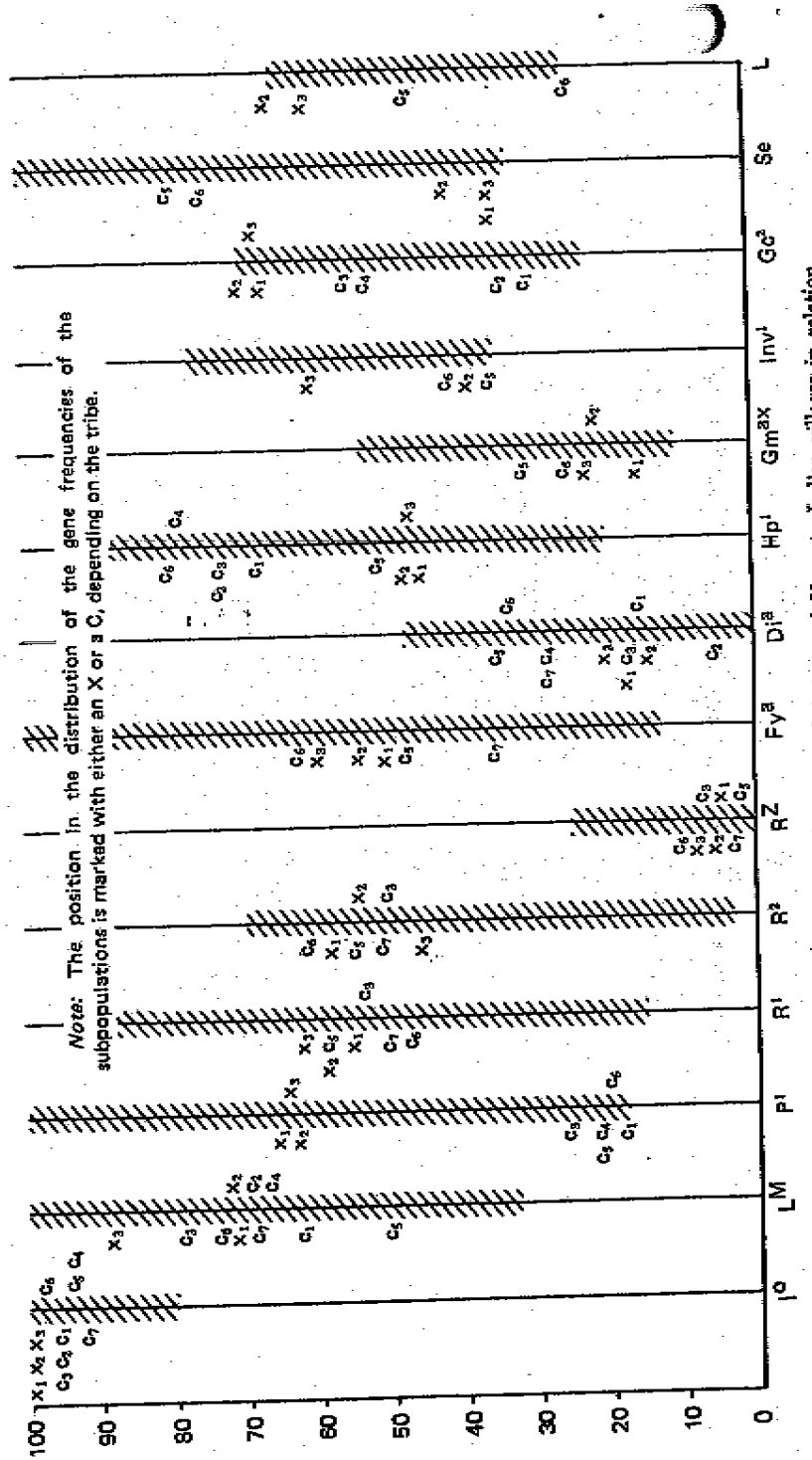

FIGURE 6. Range of gene frequencies of the Caingang and Xavante Indian villages in relation to 14 genetic markers in South American Indians (the histograms). Source: F. M. Salzano, "Intra- and Inter-Tribal Genetic Variability in South American Indians," *American Journal of Physical*

TABLE 3
A COMPARISON OF THE GENETIC VARIABILITY FOUND IN FOUR AMERICAN INDIAN POPULATIONS WITH THE AVAILABLE GENE DISTRIBUTION OF THE SOUTH AMERICAN INDIANS

Alleles	*Range of Gene Freq., 5 Caingang Communities**	*Max. Deviation*	*Range of Gene Freq., 3 Xavante Communities**	*Max. Deviation*	*Range of Gene Freq., 7 Makiritare Villages†*	*Max. Deviation*	*Range of Gene Freq., 10 Papago Districts‡*	*Max. Deviation*	*Range of Gene Freq., South American Indians**	*Max. Deviation*
I^o	0.93–1.00	0.07	1.00	0.00	1.00	0.00	0.86–1.00	0.14	0.80–1.00	0.20
L^M	0.50–0.78	0.28	0.71–0.87	0.16	—	—	—	—	0.32–1.00	0.68
P^1	0.16–0.27	0.11	0.63–0.66	0.03	0.30–0.52	0.22	0.37–0.59	0.22	0.16–1.00	0.84
R^1	0.47–0.57	0.10	0.54–0.62	0.08	0.27–0.44	0.17	0.54–0.71	0.17	0.12–0.96	0.84
R^2	0.28–0.40	0.12	0.24–0.39	0.15	0.47–0.62	0.15	0.27–0.40	0.13	0.02–0.73	0.71
R^o	0.00–0.10	0.10	0.03–0.08	0.05	0.00–0.10	0.10	0.00–0.04	0.04	0.00–0.25	0.25
Fy^a	0.36–0.63	0.27	0.50–0.59	0.09	0.64–0.80	0.16	0.76–1.00	0.24	0.13–1.00	0.87
Di^a	0.07–0.36	0.29	0.15–0.20	0.05	0.02–0.52	0.50	0.01–0.12	0.11	0.00–0.48	0.48
Hp^1	0.52–0.79	0.27	0.46–0.48	0.02	0.24–0.65	0.41	0.34–0.58	0.24	0.20–0.89	0.69
Gm^{ax}	0.26–0.32	0.06	0.16–0.23	0.07	—	—	—	—	0.10–0.54	0.44
Inv^1	0.33–0.40	0.07	0.37–0.63	0.26	—	—	—	—	0.33–0.76	0.43
Gc^2	0.28–0.56	0.28	0.67–0.70	0.03	0.11–0.36	0.25	0.09–0.27	0.17	0.22–0.70	0.48
Se	0.73–0.81	0.08	0.34–0.42	0.08	1.00	0.00	—	—	0.34–1.00	0.66
L	0.25–0.47	0.22	0.61–0.67	0.06	—	—	—	—	0.25–0.67	0.42

* Salzano (1968).
† Gershowitz et al. (1970) and Arends et al. (1970).
‡ Workman and Niswander (1970).

tions among aboriginal populations of South America demonstrates the existence of a high level of genetic variation between historically related human populations. The range of gene frequencies of South American Indians varies markedly from one locus to another. For example, the Duffy allele, Fy^a, has a range of 0.87 (gene frequencies of from 0.13 to 1.00 have been reported in various populations), while the I^o allele exhibits a range of 0.20, with a distribution of 0.80 to 1.00. Why do the frequencies of certain alleles vary only within narrow ranges, although the populations are widely dispersed, while those of other genetic loci, even in subdivisions of the same population, vary within broader ranges? There is no simple answer to this question. One must take into account the history of the population, many social and ecological factors, and the demographic structure, any or all of which may provide clues as to the relative actions of the forces of evolution. It is no longer acceptable to explain gene frequency differences, within or between populations, strictly on the basis of a single force of evolution, such as genetic drift; the exception is if the action of natural selection or gene flow can be shown to be minimal or unlikely through historical evidence or some peculiarity of the social structure or population size. All too often, investigators attribute genetic differentiation of populations to specific causes without adequate data. This is, to some extent, an artifact of the genetic models that were formulated with many assumptions and unfounded simplifications, and that only rarely considered the simultaneous action of more than one possible source of evolutionary change. However, the observable gene frequencies in human populations are the products of all the forces of evolution, acting not one at a time but simultaneously.

ANTHROPOLOGICAL USE OF GENE MARKERS

One of the major theoretical preoccupations of physical anthropologists in the 1930s and 1940s was the definition and classification of the races of man. Hooton (1946:448) defined *race* as "a great division of mankind, the members of which, though individually varying, are characterized as a group by a certain combination of morphological and metric features, principally non-adaptive,[2] which have been derived from their common descent." Racial classifications were based on various mor-

phological criteria that were considered nonadaptive, and, therefore, similarities in these traits between populations denoted biological relationships. Thus, racial history, according to the early physical anthropologists, could be traced on a worldwide basis. The early students of race assumed they knew what characteristics of man were nonadaptive.

In the late 1940s and early 1950s, the so-called nonadaptive traits were being examined with suspicion. For example, Boyd (1952:18–19) stated, "it is doubtful if any hereditary characters are completely non-adaptive, and . . . probably we can deal only with different degrees of adaptive value." Hooton (1946), however, altered his opinion about the existence of nonadaptive traits in the second edition of his book: "the use of non-adaptive characters in human taxonomy now seems to me impractical and erroneous."

In the United States, with the demise of racial classification based on morphological, nonadaptive traits, came the use of blood group data for the same purposes. Boyd (1952:27), while discrediting morphologically based racial taxonomy, equivocally substituted blood groups in its place: "Among the racial characters which we would be tempted to pick out at the present time as non-adaptive, there are certain serological features of the blood, such as the genes O, A, B, M, N, etc."

Anthropologists in the 1940s, 1950s, and even into the late 1960s continued to classify races of the world, using allelic frequency distributions instead of the cephalic index or the bizygomatic breadth. Greater emphasis was placed on the process of racial differentiation than on the classification and taxonomy of man. Garn (1961) distinguished between three different kinds of races—geographical, local, and microraces—and thus "modernized" the age-old concept. However, even with these new frills and the acknowledgment of a genetic approach, Garn still attempted to define nine specific geographical races in his book.

Most anthropologists today prefer a genetic definition of *races* as "populations which differ in the frequency of some genes" (Livingstone 1964b); classification has, for the most part, ceased. Livingstone commented that the concept of race seemed to him to be of no use in describing or explaining human genetic variability, which is the central problem of physical anthropology today. It is for this reason that reference was made earlier only to intrapopulational and interpopulational variability, without invoking race or racial categories.

During the late 1950s and early 1960s, several scholars (Oschinsky

1959, Bielicki 1962, Wiercinski 1962) became highly critical of the use of the so-called genetical methods for the formulation of racial classification and for tracing phylogenetic relationships between populations. What initially appeared to be the panacea for racial classification in a mathematically precise manner later became bogged down in the contradictory and inconsistent schemes based on gene frequencies. Bielicki and Wiercinski advocated the return to conventional morphological and anthropometric criteria for racial classifications. Boyd (1963), in defense of the genetic method, reviewed what he considered its four most important contributions to anthropological theory. These contributions concerned theories of the ethnogenesis of the Gypsies, Lapps, American Blacks, and Pygmies of Equatorial Africa, and using these four groups, Boyd showed that some ethnologically derived theories may be tested biologically. Although most investigators shy away from tracing the ethnogenesis of various populations and races across the maps, cautious tests of possible relationships between populations with some suspected historical connections are possible through the use of marker gene distribution. The presence or absence of certain alleles may permit the testing of various ethnographic, linguistic, historical, and evolutionary hypotheses. For example, because of the linguistic similarities between the language of the Gypsies and Sanskrit, and because of the oral tradition of the Gypsies (which traces their origins to India), the ethnogenesis of these itinerants has been periodically questioned by various anthropologists. The blood group frequencies support the Gypsies' own claim of an Indian origin, at least for the Hungarian Gypsies (Boyd).

However, not all of the Gypsy populations are related to the Indian populations. The Irish itinerant populations (Tinkers)—sometimes referred to as Irish Gypsies—are genetically unrelated to the Romany Gypsies of the European mainland. The Irish itinerants are culturally distinct from the Irish people, speaking a language (Gammon) unintelligible to the Irish and moving about the countryside in caravans consisting of their nuclear or extended families.

A recent study (Crawford and Leyshon 1971) has shown that on the basis of 17 genetic loci, the Tinkers are more similar genetically to the Irish controls than they are to any of the Gypsy or Indian populations (Table 4). There are some differences between control and Tinker gene frequencies in the Lewis and Duffy systems, but these differences may be due to inbreeding or to sampling error. Even so, the frequencies peculiar

TABLE 4
A COMPARISON OF SOME BLOOD GROUP FREQUENCIES BETWEEN AN IRISH CONTROL POPULATION, AN INDIAN POPULATION, AND THE IRISH TINKERS

Blood Group	*Alleles*	*Irish Controls* (N = 97)	*Indian Population** (N = 1168)	*Irish Tinkers* (N = 119)
ABO	*A*	0.1878	0.2244	0.2010
	B	0.1208	0.2183	0.1208
	O	0.6811	0.5573	0.6859
P	*P*	0.7216	0.32–0.50	0.7983
	p	0.2783	0.49–0.68	0.2016
MN	*M*	0.5979	0.6917	0.5466
	N	0.4020	0.3083	0.4533
Duffy	Fy^a+	0.4258	0.2886	0.3519
	Fy^a-	0.5742	0.7114	0.6481

* Mourant (1954).

to the Tinkers are much closer to those of the Irish than to the Indian populations.

Allelic frequencies can be utilized to estimate the degree of admixture and gene flow between populations. The study of Glass and Li (1953) of the American Blacks applied Bernstein's (1931) method to the frequencies of seven marker genes (R_o, R_1, T, r, B, A, and R_2) to estimate the amount of gene flow from the Caucasion population into the West African slave gene pool. This crude method of estimation revealed that approximately 30 percent of the American Black gene pool was of Caucasian origin. In the last decade, various estimates of the genetic makeup of the American Black populations have been made using more refined techniques. Both the estimates and the techniques are discussed by P. L. Workman in Chapter 6 of this book.

Most recently, Crawford, McClean, and Workman (1971) have attempted to reconstruct the amount of genetic admixture in Tlaxcala, Mexico, using the inhabitants of a Mestizo town (Tlaxcala) as the hybridized population and those of an Indian town (San Pablo), in the valley, as representative of one of the parental populations. Spanish gene frequencies were obtained from the literature for 23 different loci and are assumed to represent the second parental population. One of the investigators performed multiple regression analyses using three alternative

models of the base populations for the Tlaxcala Mestizo. The first computer runs were based on San Pablo, representing the Indian parental population, and the European component was based on data from the largest available sample from Spain, or from other western European populations if the data from Spain could not be obtained. This method gives the composition of the Mestizo gene pool as 31.4 percent Spanish in origin with a standard deviation of 3.4 percent and a standard error of the slope of the regression line of 29.7.

Recognizing that the present population of San Pablo may imperfectly represent the Indian gene pool, a second run using estimated mean Nahua Indian gene frequencies for 16 loci and data from eight populations, including the San Pablo sample, was made. When the mean Nahua values were run in multiple regression, the estimate of the Mestizo composition was unchanged (31.6 percent of the gene pool was of Spanish origin), but the standard deviation more than doubled (to 7.0), and the standard error of the regression line increased to 53.5.

Because of the presence of the Rh_o factor (*cDe* chromosome segment present at a frequency of 0.09263 ± 0.0177) in the Mestizo population, suggestive of African admixture, a West African parental component was added for a triracial estimate of the Mestizo gene pool. In this model, the estimated contribution of West African genes is small (6.7 percent), with a standard deviation of 1.0, and appears to replace part of the Spanish component without affecting the estimated proportion of Indian genes. The estimate of Spanish contribution to the Mestizo gene pool decreases from 31.4 to 22.9 in the triracial fitting of the curve. The degree of fit is no better with a triracial estimate than that calculated for Spanish-Indian admixture alone. The diminution of the Spanish component as a result of the West African gene frequencies, together with demographic and historical documentation, is suggestive of an introduction of the African genes into the Mestizo gene pool through the Moorish elements in the Spanish army. It may be that this genetic survey has picked up African genes segregating in the Mestizo gene pool that were introduced into Europe through the Moorish invasion many centuries ago.

Although early writers such as Haddon (1925) and Deniker (1900) classified the Lapps with various Mongolian and proto-Mongolian groups, analysis of the blood group frequencies of the Lapps (Allison et al. 1956) revealed low gene frequencies of the *B* gene, a high frequency of the A_2 allele, and a low frequency of the *M* gene. These results contrast markedly

with the high frequencies of blood group B, absence of A_2, and high frequencies of *M*, which would be expected from Mongoloid or related populations. From this evidence, it appears that the Lapps are distinctive European populations with no special affinity or relationship to the Mongoloids, despite their linguistic similarities. More recent investigations of the Lapp populations, based on haptoglobin types, Gc, and transferrins, also support the ethnographic conclusion that the Lapps are similar to their European neighbors, although heterogeneous genetically. Table 5 summarizes the gene frequencies at the haptoglobin locus among the Lapps and their neighbors.

TABLE 5
HAPTOGLOBINS IN LAPP AND NON-LAPP SCANDINAVIAN POPULATIONS

Population	Sample Size	Gene Frequency Hp[1]	Hp[2]	Reference
Norwegian Lapps	301	0.311	0.689	Fleischer and Monn (1970)
Non-Lapp Norwegians	5,811	0.376	0.624	Fleischer and Mohr (1962)
Swedish Lapps	329	0.317	0.683	Beckman and Mellbin (1959)
Non-Lapp Swedes	1,272	0.393	0.607	Beckman, Heiken, and Hirschfeld (1961)
Finnish Skolt Lapps	300 (approx.)	0.434	0.566	Eriksson (1968)
Non-Lapp Finns	891	0.362	0.638	Makela, Eriksson, and Lehtovaara (1959)

The gene frequencies of the Gc^1 allele vary between 0.719 and 0.896 among the various Lapp populations, reaching the highest levels among the Swedish Lapps from the south (Reinskow and Kornstad 1965, Melartin 1965). There also appears to be a high frequency of the CD_1 variant of transferrin, but Melartin and Kaarsalo (1965) explain these in terms of possible consanguinity.

One of the most fascinating problems in physical anthropology is the origin and genetic affinities of the pygmy populations of the world. Populations of short stature, dark pigmentation, and Negrito morphological traits have been reported in a number of tropical forest regions of the world: the Philippine Islands, Malaya, Highland New Guinea, Central

Africa, and the Andaman Islands. A number of anthropologists have proposed an African genesis for these Negrito populations. Howells (1959: 339) concludes, "it is as plain as the nose on your face that the Negritos are intimately related to fully developed Negroes—a specialized kind of man—in skin, hair form, nose shape, and so on. The Negritos are really all similar and must have a common origin. And Negritos and Negroes cannot have appeared on separate continents; they too must have had a common origin." However, this example demonstrates the possible danger of misinterpretation when morphological traits such as skin color and hair form are used as criteria for establishing populational relationships. More recently, Howells has reinterpreted the question of the ethnogenesis of the Pygmies. Using multivariate analyses of cranial evidence, he postulates separate origins of the various populations of Negritos (see Chapter 8).

A comparison of the blood group frequencies of the African and Oceanic Negritos shows significant differences in the presence of certain marker genes in three populations compared by Boyd (1963). Table 6 illustrates the extent of variation between the three populations. The most significant differences between the Oceanic and African Pygmies were in the presence of the R_o chromosomal segment and the relative frequency of the *M* and *N* antigens. The R_o chromosomal segment is usually referred to as an African marker gene, because of its high frequency in Africa and its low frequency, or absence, in other parts of the world that had not experienced gene flow from African populations. The Pygmies of the Congo exhibit a frequency of the R_o segment similar to frequencies found in other parts of Africa, in contrast to the low frequencies of this allele exhibited by Papuan Negritos and its total absence in the populations of the Andaman Islands. The highest frequency on earth of the *M* allele is found in New Guinea among the Papuan Pygmies. From these data, it apears that (1) the pygmy populations differ significantly among themselves, and (2) genetically, they resemble the surrounding populations more than these populations resemble each other. This is not surprising, considering the fact that in all cases, the Pygmies speak the language of their surrounding populations and apparently do not share a common language.

More recent genetic investigations have shown that, although the Pygmies are closer genetically to the Bantu than to the Bushman or Hottentots, they are also distinct populations differing from all others. For

TABLE 6
BLOOD GROUP GENE FREQUENCIES IN PYGMY POPULATIONS

Population	Reference	Number Tested	Frequency of Gene								
			A	B	O	M	N	R_0	R_1	R_2	r
Pygmies Belgian Congo	Hubinot and Snoek (1949)	2,557	0.198	0.249	0.553	0.468	0.523	0.630	0.074	0.194	0.101
Papuan Pygmies	Graydon et al. (1958)	139	0.075	0.139	0.786	0.102	0.898	0.030	0.850	0.119	0.000
Andaman Negritos	Lehmann and Ikin (1954)	52	0.540	0.080	0.380	0.610	0.390	0.000	0.920	0.000	0.000

SOURCE: After Boyd (1963).

example, Benerecetti, Modiano, and Negri (1969) have demonstrated that a phosphoglucomutase (PGM) variant PGM_2^{Pw} is common only among the Babinga Pygmies of Africa, with a gene frequency of 7 percent. It has been suggested that this variant may be limited to the Pygmies and would therefore be a useful marker gene.

Another polymorphism that may be limited to the Pygmies was recently reported by Benerecetti and Negri (1970) at the red cell peptidase C locus. Although polymorphisms of other red cell peptidases have been reported (Lewis and Harris 1967, 1968, 1969), none have been detected for peptidase C until the report by Benerecetti and his colleagues. They describe a polymorphism controlled by three alleles at an autosomal locus, with Pep C^1 and Pep C^2 as co-dominant and Pep C^0 as a silent gene lacking a detectable electrophoretic component in the homozygous genotype.

Bodmer and Bodmer (1970) have described population differences in the HL-A system of leukocytic antigens in the Pygmies. The Pygmies differ from other populations on the basis of the total absence of the HL-A antigen and the very high incidence of the blank alleles for both the LA and 4 series. This suggests the existence of as yet unidentified antigens that have much higher frequencies in Pygmies than in Caucasians, and that may eventually serve as marker genes.

This concludes the review of the use of genetic markers for tracing phylogenetic relationships between human populations. Barnicot (1964), however, cautions against attempting to make historical inferences from gene distribution maps. He correctly argues that the presence of a common marker gene may not imply a phylogenetic relationship, but that each marker may have arisen independently in any population (Barnicot 1964:954): "To envisage human history as nothing but a series of migrations or intermixtures of peoples, each carrying its array of gene frequencies, like a convenient label to show its point of departure, may be an oversimplification."

Despite the contradictions and problems that have come to light through the use of genetic markers, a form of racial classification still persists. Although present statistical techniques are more sophisticated than those employed by the anthropologists of earlier decades, and although recent authors do not claim to be classifying races, the difference between earlier and present-day anthropologists is largely semantic—now they derive "trees of likely descent of populations" or "phylogenetic trees." Irrespective of the statistical trappings and the use of sophisticated

computer programs, these exercises are basically as much a waste of time and money today as they were 50 years ago. Instead of attempting to classify the populations of the world on the basis of gene frequencies, basic research is needed to explain how and why the genetic variability is maintained at such high levels in human populations.

Two studies, Edmondson (1965) and Cavalli-Sforza, Barrai, and Edwards (1964), have proposed schemes of racial affinities based on gene frequency distributions. Edmondson, using 24 genetic traits from 124 populations, estimated the genetic distance between populations by averaging the differences in gene frequencies of all the populations of the sample. He then reconstructed the probable patterns of divergence of the major races of man. The Cavalli-Sforza study selected 15 populations and five blood group systems for estimating the most likely phylogenetic tree for the history of human racial separation. Surprisingly, although the methods used to calculate the phylogenetic affinities were different, the results were grossly similar. However, Cavalli-Sforza and his colleagues do not define the criteria used for selecting the genetic markers for their scheme. Moreover, neither of these methods assumes directional selection, nor does either control for gene flow and hybridization.

In conclusion, genetic markers can be cautiously employed in testing specific hypotheses of genetic relationship or affinity, as long as the limitations of the method are understood. Gene frequencies of simple Mendelian traits are the new materials for both population genetics and anthropological analyses. However, gene frequencies describe only a portion of the gene pool and permit a "glimpse" of the genetic structure of the population. Therefore, gene marker systems and their frequencies in human populations, when prudently used, can be a means to an end, but they are not an end in themselves.

NOTES

1. The term *genetic marker* will be restricted here to discrete, segregating, genetic traits which can be used to characterize populations by virtue of their presence, absence, or high frequency in some populations and low frequency in others.

2. By *non-adaptive*, Hooton means those traits not subject to the action of natural selection. He listed (1931) the following bodily characters as mainly nonadaptive variations: hair color and distribution, shape of lips, form of the incisors, the length of the forearm relative to the arm, and so forth.