Am J Hum Genet 26:421–443, 1974

Genic Variation Within and Between the Three Major Races of Man, Caucasoids, Negroids, and Mongoloids

MASATOSHI NEI[1] AND A. K. ROYCHOUDHURY[1]

INTRODUCTION

It is of considerable importance to know the number of gene differences among different races of man. Many geneticists and anthropologists have studied racial differences in gene frequency at various polymorphic loci, but little attention has been given to the overall gene differences between races. The main reason for this seems to be the fact that racial differences in gene frequency are mostly quantitative, so that it is difficult to determine whether a locus is the same between races. This difficulty, however, can be overcome if we consider not loci but genes (alleles) as units of comparison. Recently Nei [1–3] developed a statistical method by which the number of codon or mutational differences per locus as well as the proportion of different genes between populations can be estimated from gene frequency data. This method has already been applied to data on protein polymorphism in man and other organisms [1, 2, 4]. It has been shown by Nei and Roychoudhury [4] that the net codon differences between Caucasian, Negro, and Japanese populations are rather small compared to the codon differences between two randomly chosen genomes from the same population.

The genic variation within a population is usually measured by the average heterozygosity per locus. For a comparison of gene variations within and between populations, a better measure is the number of codon differences per locus between two randomly chosen genomes from the same populations [3, 4]. Ordinarily, however, there is not much difference between these two quantities. The average heterozygosity in man has been estimated to be about 7%–10% for protein loci [4–6], while the estimate for blood group loci is 16% [7].

In this paper we shall present the results of further analysis of genic variation within and between the three major races of man, Caucasoids, Negroids, and Mongoloids. We have used gene frequency data for protein and blood group loci. These two kinds of data were analyzed separately because the detectability of gene differences may be different.

Received May 30, 1973; revised February 11, 1974.

Supported by U.S. Public Health Service research grant GM 20293.

[1] Center for Demographic and Population Genetics, University of Texas, Houston, Texas 77025.

MATERIALS AND METHODS

In our previous investigation [4] we used 44 protein (including enzyme) loci for both Caucasoids and Negroids. Surveying the literature further, we have collected gene frequency data of 74 protein loci for Caucasoids and 62 loci for Negroids, the number of common loci being 62. The gene frequencies for these protein loci were all studied by electrophoresis. The majority of data were taken from the American Caucasoids (52 loci) and the American Negroids (42 loci). When there were several different data for the same race, we used the most recent and extensive one. However, the differences in gene frequency between different sources of data were generally very small within the same race. When Mongoloids were included, the number of common loci for which gene frequency data were available was 35. The majority (29 loci) of the Mongoloid data were taken from the Japanese population, data for two loci (third component of complement and parotid basic protein) coming from the Chinese population, data for another three loci (galactose-1-phosphate uridyl transferase, α-galactosidase, and 2,3-diphosphoglycerate mutase) from unspecified Oriental populations, and those for one locus (ceruloplasmin) from the Korean population. When gene frequency data were available for any two or all of the Japanese, Chinese, and Korean populations, we used the Japanese data. In general, however, there was not much difference in gene frequency among these three populations when large samples were surveyed. The names of proteins used in our investigation and the references (arranged alphabetically) from which gene frequency data were taken are given in Appendix A.

Although we do not know what proportion of the human genome is concerned with blood cell antigens, it is of interest to study the heterozygosity and genetic distance at these loci, since there is a large amount of data published. So far more than 100 blood cell antigens have been discovered [8], but for many of these blood groups, gene frequency data are not available. It is also often difficult to define a locus for blood groups, particularly when there are associated antigens. We followed Race and Sanger [8] in defining "loci" and collected gene frequency data for 57, 34, and 22 loci for Caucasoids, Negroids, and Mongoloids, respectively. We neglected all data about associated antigens. In our data analysis, we assumed that Rh and MNSs systems are controlled by three and two loci, respectively. (Actually, we analyzed the data also by assuming that each of these systems is controlled by a single locus, but it was found that either assumption gives virtually the same average heterozygosity and genetic distance.) The number of loci common to Caucasoids and Negroids was 34, of which 29 and 28 loci were taken from the American Caucasoids and the American Negroids, respectively. When Mongoloids were included, 21 loci common to the three races were obtained. Of the 21 loci, the gene frequency data of 19 loci were taken from the Japanese population, and the rest (Berrens and Radin) from the Chinese and other Mongoloid populations. The names of blood groups and their references are given in Appendix B.

In order to estimate the average heterozygosity and genetic distance per locus, it is important to choose gene loci at random from the genome. In the present case the gene frequency data were collected from the literature, so that this is not assured. However, there are some reasons to believe that the loci used here do not deviate grossly from a random sample of the genome, except for the Negroid and Mongoloid blood group loci. The weighted mean of average heterozygosities for enzymic loci (excluding nonenzymic protein loci) in the three populations is about 8.4%. This value is close to the value (7.0%) obtained by Harris [5] for 20 arbitrarily chosen enzymic loci in man. In a study of the average heterozygosity for blood group loci in the English population, Lewontin [7] believed that the 33 loci he used were close to a random sample of the genome because the cumulative average heterozygosity had reached an apparent asymptotic value when plotted against the year of discovery of each blood group (to 1962). We studied the same cumulative average heterozygosity to 1968 for the three populations separately. The results obtained are given in figure 1. This figure suggests that the cumulative average

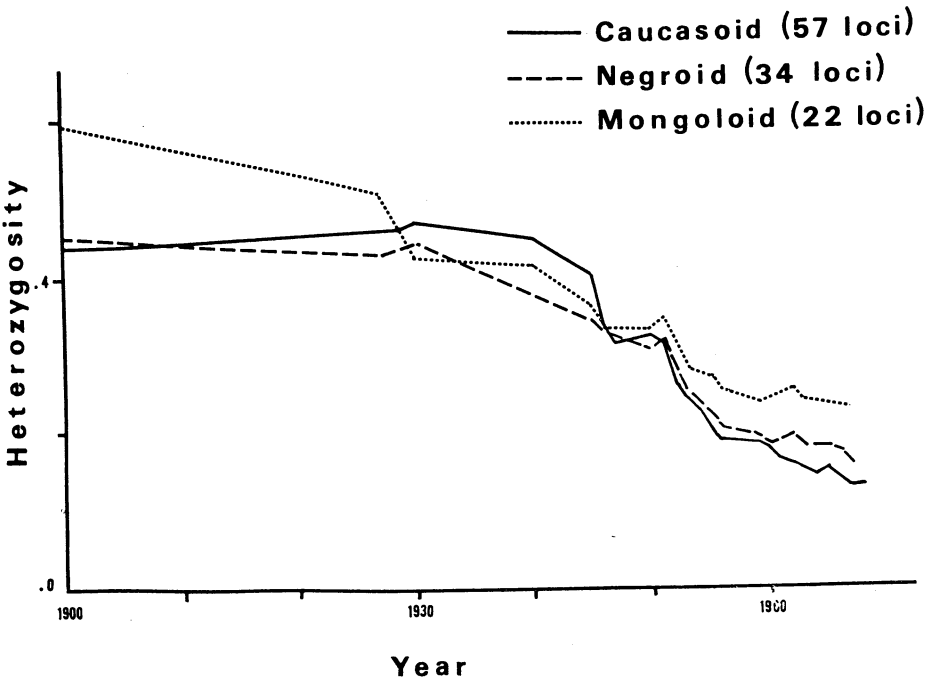

FIG. 1.—Asymptotic behavior of cumulative average heterozygosity for blood group loci in the three major races of man.

heterozygosity has not really reached the asymptotic values perhaps in all populations. This is particularly so in Negroids and Mongoloids, where the number of loci examined is much smaller than that in Caucasoids. As indicated by Lewontin, polymorphic loci are easier to detect than monomorphic loci, so there is a tendency to overestimate the average heterozygosity if the number of loci is small. Figure 1 clearly indicates this tendency, and the data for Negroids and Mongoloids probably deviate from a random sample of the genome. Our data for Caucasoids do not support Lewontin's belief that the 33 loci he used are close to a random sample of the genome for the English population. The asymptotic value of the cumulative average heterozygosity, which he estimated to be about 16%, now appears to be about 13% or less. Nevertheless, figure 1 shows an asymptotic behavior of the cumulative average heterozygosity as noted by Lewontin, and we tentatively assume that our data for Caucasoids to 1968 are close to a random sample of the genome.

The number of individuals sampled for determining gene frequencies varied considerably with each locus in both protein and blood group data. In a majority of cases (55%–97%), the number was larger than 100 and often more than 1,000. In one extreme case the number for a polymorphic locus was as small as 29. For studying the average heterozygosity and genetic distance between populations, however, a rather small sample per locus (about 20 individuals) is known to be sufficient [9]. To get reliable estimates of these quantities, a large number of loci rather than a large number of individuals per locus should be studied.

RESULTS

Heterozygosity

The proportions of polymorphic loci and average heterozygosities for protein and blood group loci are presented in table 1. In this paper a locus is defined as poly-

TABLE 1

PROPORTION OF POLYMORPHIC LOCI AND AVERAGE HETEROZYGOSITY FOR PROTEIN AND BLOOD GROUP LOCI IN THE THREE MAJOR RACES OF MAN

Race and No. Loci Used	Proportion Polymorphic Loci	Average Heterozygosity
	Protein Loci	
Caucasoid:		
74*	.31	.099 ± .021
62†	.32	.104 ± .023
35‡	.40	.142 ± .034
Negroid:		
62†	.40	.092 ± .019
35‡	.51	.122 ± .028
Mongoloid:		
35‡	.40	.098 ± .027
	Blood Group Loci	
Caucasoid:		
57*	.37	.130 ± .027
34†	.56	.197 ± .038
21‡	.71	.264 ± .049
Negroid:		
34†	.44	.162 ± .035
21‡	.62	.218 ± .046
Mongoloid:		
22§	.59	.231 ± .049
21‡	.62	.242 ± .050

* All loci for Caucasoids.
† Common loci for Caucasoids and Negroids.
‡ Common loci for Caucasoids, Negroids, and Mongoloids.
§ All loci for Mongoloids.

morphic if the frequency of the most common allele is less than or equal to .99. This definition is clearly arbitrary and therefore is not as good a measure of the heterogeneity of a population as the average heterozygosity.

It is seen that the proportion of polymorphic loci and heterozygosity for Caucasoid populations are .31 and .099 when all 74 loci are used. These values are both higher than those (.28 and .07, respectively) obtained by Harris and Hopkinson [6]. The difference is probably due to the fact that we have included 12 nonenzymic protein loci which are generally more polymorphic than enzymic loci, while Harris and Hopkinson have excluded them. If we exclude the nonenzymic loci, the average heterozygosity becomes .087, which is close to the value obtained by Harris and Hopkinson [6].

Comparison of the average heterozygosities of Caucasoids and Negroids by using 62 common protein loci indicates that Caucasoids are more heterogeneous than Negroids, although the difference in average heterozygosity is not statistically significant. On the other hand, the proportion of polymorphic loci suggests that Cauca-

soids are less polymorphic than Negroids. This is somewhat inconsistent with the result of average heterozygosity, but the difference is again not statistically significant. Table 1 also shows that in terms of average heterozygosity, Mongoloid populations are least heterogeneous, but little importance can be attached to the racial differences since none of them is statistically significant. From table 1 we may conclude that the average heterozygosity for protein loci in human populations is about 10%, as was the case in our previous work.

The frequency distributions of heterozygosity of protein loci for the three major races are given in figures 2, 3, and 4. All the distributions are inverted J-shaped, although there is a small peak in the tail for Caucasoids and Mongoloids, and about 70%–80% of the loci have a heterozygosity lower than .10. Since the number of loci used is still small, the racial difference in the distribution is not clear.

Although the racial differences in average heterozygosity are not statistically significant, there are some loci at which a considerable difference in single-locus heterozygosity exists between races (see Appendix A). For example, Caucasoid populations have a higher heterozygosity in placental alkaline phosphatase, third component of complement (C'3 system), and peptidase A than Negroids, while the latter have a higher heterozygosity in parotid basic protein, G6PD, and peptidase C than Caucasoids. The heterozygosities at individual loci for Mongoloid populations are similar to those for Caucasoids, but there are some loci at which

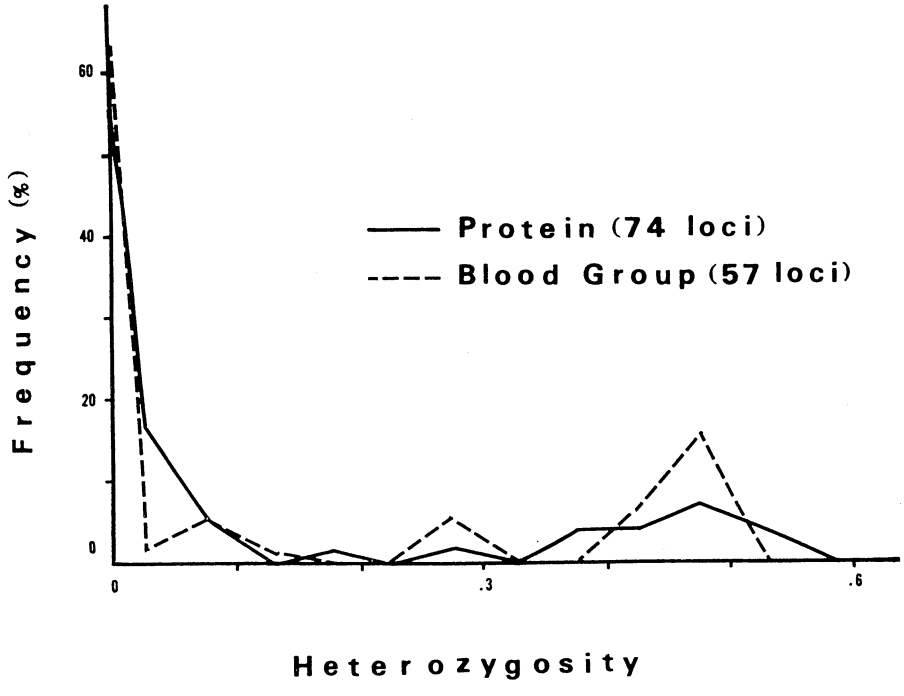

FIG. 2.—Frequency distributions of heterozygosity for protein and blood group loci in Caucasoids.

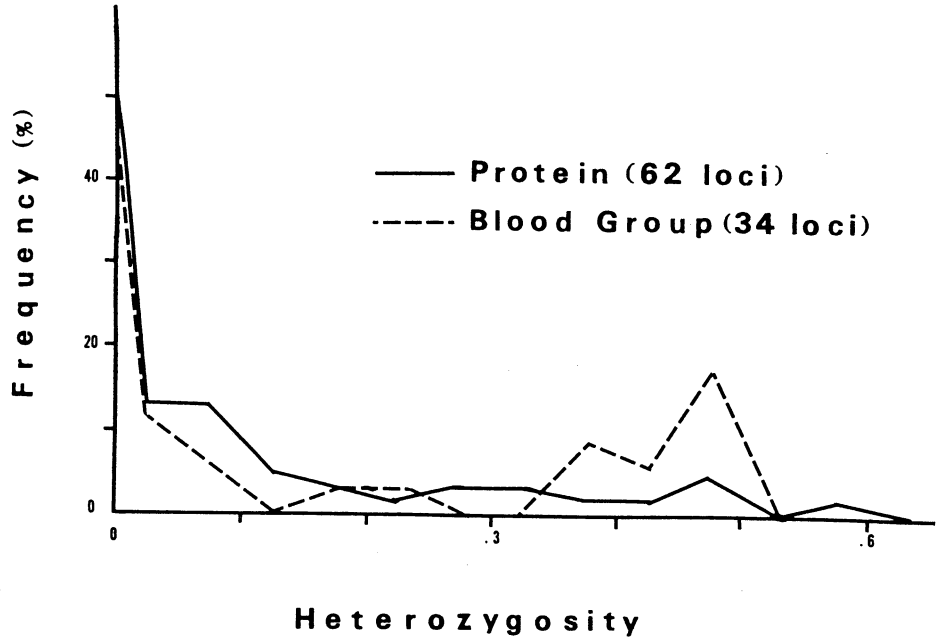

FIG. 3.—Frequency distributions of heterozygosity for protein and blood group loci in Negroids

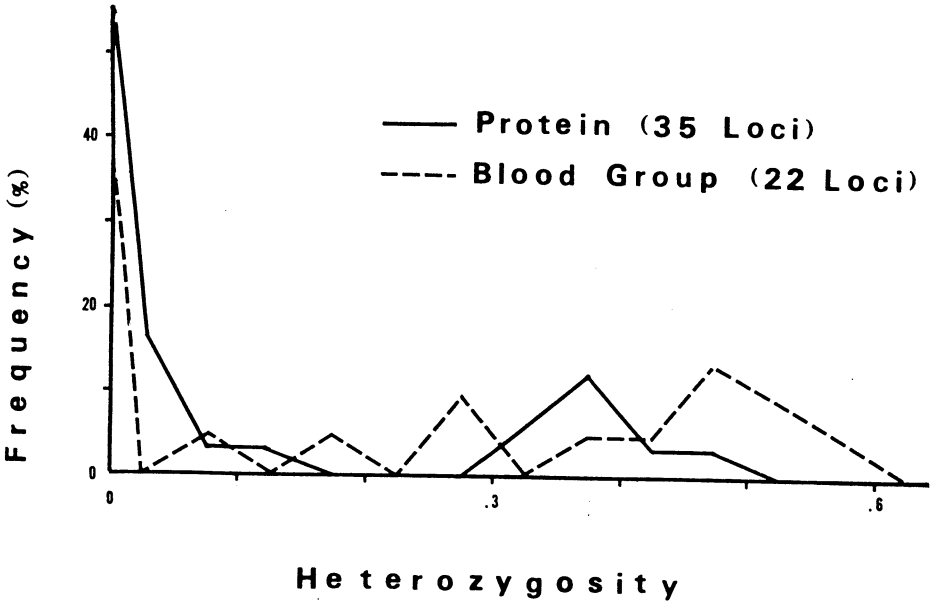

FIG. 4.—Frequency distributions of heterozygosity for protein and blood group loci in Mongoloids.

Mongoloids are uniquely polymorphic (soluble glutamic oxaloacetic transaminase) or uniquely monomorphic (adenylate kinase and pepsinogen). Nevertheless, the correlations of heterozygosity between the three major races are very high (table 2).

TABLE 2

CORRELATION COEFFICIENTS OF HETEROZYGOSITIES BETWEEN THE THREE MAJOR RACES OF MAN

Comparison	Proteins		Blood Groups	
	No. Loci	r	No. Loci	r
Caucasoid/Negroid	35	.816	21	.862
Caucasoid/Mongoloid	35	.860	21	.886
Negroid/Mongoloid	35	.720	21	.873

Appendix A includes the standard errors of heterozygosity at each locus. It is noted that this standard error is generally small, whereas the standard error of the average heterozygosity is rather large (table 1). This is because the latter includes the interlocus variation of heterozygosity. The standard errors of the heterozygosity for each locus and the average heterozygosity were obtained by the method given by Nei and Roychoudhury [9].

Table 1 shows that the proportion of polymorphic loci and the average heterozygosity for 57 blood group loci for Caucasoids are .37 and .130, respectively. The value of average heterozygosity is lower than that obtained by Lewontin by using 33 loci for the English population. This is because the blood group loci discovered after 1962 are largely monomorphic. At face value, it is higher than that for protein loci. However, we do not know which group of loci are more variable at the codon level, since the difference in detectability of gene differences by electrophoresis and immunological reaction is not well known.

As noted earlier, the blood group loci for Negroids and Mongoloids probably deviate from a random sample of the genome. However, the average heterozygosities of Caucasoids and Negroids can be compared by using 34 common loci: they are .197 and .162, respectively. Thus, Caucasoids appear to be genetically more heterogeneous than Negroids, though the difference is not statistically significant. A higher heterozygosity for Caucasoids is also obtained by using 21 loci which are common to all three major races. The average heterozygosity for Mongoloids is intermediate between the values for Caucasoids and Negroids, but again there is no statistically significant difference between any pair of the three major races. These results suggest that Negroid and Mongoloid populations are no more heterogeneous than Caucasoids with respect to blood group loci, and if as many loci as that for Caucasoids were studied, the average heterozygosity for these two major races would be reduced to the same order of magnitude as that for Caucasoids.

Figures 2, 3, and 4 indicate that the frequency distributions of heterozygosity for

blood group loci are basically similar to those for protein loci and there is a small peak between .3 and .5 in all populations. Loci such as ABO, Kidd, Auberger, Lewis, MN, P, Secretor, and Xg belong to the high heterozygosity group (Appendix B). There are some racial differences in heterozygosity at individual loci. For example, Caucasoids have a higher heterozygosity at the loci for Duffy, Rh (*Cc*), Kidd, and P than Negroids, while the latter have a higher heterozygosity at the loci for the Lewis and Xg blood groups. On the other hand, Mongoloids have a high heterozygosity at the ABO and Rh (*Ee*) loci. The correlations of heterozygosity between the three major races are, however, very high, as in the case of protein loci (table 2).

Interracial Gene Differences

The genetic distance or gene differences between two populations may be measured in terms of the number of net codon differences per locus. Nei [3] devised three measures of genetic distance: the minimum, standard, and maximum distances. The units of these distances are the accumulated number of gene substitutions or number of net codon differences per locus. (Strictly speaking, the term "codon differences" is not appropriate, since there is a small probability that the mutation could be due to deletion, insertion, frameshift, etc. However, for lack of a better alternative, we use this terminology.) In practice, of course, we consider only those codon differences that are detectable by the technique used (electrophoresis or immunological reaction in the present paper).

Let x_i and y_i be the frequencies of the ith allele at a locus in populations X and Y, respectively. We can then compute the probabilities of identity of two randomly chosen genes within and between populations; that is, $j_X = \Sigma x_i^2$, $j_Y = \Sigma y_i^2$, and $j_{XY} = \Sigma x_i y_i$. We denote by J_X, J_Y, and J_{XY} the arithmetic means of these quantities over all loci, including monomorphic ones. Similarly, we denote the geometric means by J'_X, J'_Y, and J'_{XY}. The minimum, standard, and maximum genetic distances are then given by $D_m = D_{XY} - (D_X + D_Y)/2$, $D = -\ln I$, and $D' = -\ln I'$, respectively, where $D_X = 1 - J_X$, $D_Y = 1 - J_Y$, $D_{XY} = 1 - J_{XY}$, $I = J_{XY}/\sqrt{J_X J_Y}$, and $I' = J'_{XY}/\sqrt{J'_X J'_Y}$. The D_X and D_Y may be interpreted as the minimum estimate of codon differences per locus between two randomly chosen genomes from populations X and Y, respectively [3]. They are also equal to the average heterozygosity. In general, $D_m < D < D'$, but this relation may not hold if the absolute values are very small. The standard errors of D_m, D, and D' may be obtained by the method given by Nei and Roychoudhury [9]. For other properties of these measures of genetic distance, see Nei [1–3].

The three estimates of genetic distances or net codon differences between Caucasoids and Negroids are given in table 3. The estimates from 62 common protein loci are virtually the same as our previous estimates obtained from 44 protein loci. The difference between the minimum and maximum distances is very small. The estimates of genetic distances obtained from 34 blood group loci are slightly higher than those for protein loci, but still in the range of .01–.02 per locus. As we noted in our previous study, the interracial net codon differences are small compared to

TABLE 3

MINIMUM, STANDARD, AND MAXIMUM ESTIMATES OF NUMBER OF NET CODON DIFFERENCES PER LOCUS BETWEEN CAUCASOID AND NEGROID POPULATIONS

	Proteins (62 Loci)	Blood Groups (34 Loci)
Minimum	.010 ± .003	.013 ± .005
Standard	.011 ± .004	.016 ± .007
Maximum	.014 ± .006	.020 ± .008

the intraracial codon differences between two randomly chosen genomes within the same race. The ratio of the interracial net codon differences to the intraracial codon differences $[R = 2D_m/(D_X + D_Y)]$ is 7%–10% between Caucasoids and Negroids.

The genetic distances between Caucasoids, Negroids, and Mongoloids can be compared by using 35 common protein and 21 common blood group loci (table 4).

TABLE 4

MINIMUM, STANDARD, AND MAXIMUM ESTIMATES OF NUMBER OF NET CODON DIFFERENCES BETWEEN CAUCASOID, NEGROID, AND MONGOLOID POPULATIONS

	Caucasoid/Negroid		Caucasoid/Mongoloid		Negroid/Mongoloid	
	Proteins (35 Loci)	Blood Groups (21 Loci)	Proteins (35 Loci)	Blood Groups (21 Loci)	Proteins (35 Loci)	Blood Groups (21 Loci)
Minimum	.014 ± .006	.021 ± .008	.010 ± .004	.025 ± .009	.017 ± .008	.070 ± .034
Standard	.017 ± .007	.027 ± .012	.011 ± .005	.034 ± .014	0.19 ± .009	.095 ± .049
Maximum	.021 ± .010	.031 ± .012	.012 ± .005	.043 ± .016	.026 ± .013	.144 ± .075

In the case of protein loci, the genetic distance between Negroids and Mongoloids is slightly higher than that between Caucasoids and Negroids. On the other hand, the distance between Caucasoids and Mongoloids is the least among the three pairs of groups. This suggests that Caucasoids and Mongoloids are more closely related to each other than to Negroids. However, the results from blood group loci are somewhat inconsistent with this conclusion; that is, Caucasoids and Negroids are more closely related than Caucasoids and Mongoloids, while Negroids and Mongoloids are least related. We will discuss this problem in more detail later.

The absolute values of the estimates of genetic distance between Caucasoids and Negroids are considerably larger when 21 common blood group loci are used than when 34 loci are used. This apparently reflects nonrandom sampling of blood group loci when the number of loci is small, as discussed earlier. The genetic distances between Negroids and Mongoloids are about three times larger than the values between the other pairs of races. The large value is mainly due to the Duffy locus,

in which the single-locus distance $[d = (j_X + j_Y)/2 - j_{XY}]$ is .63. The gene frequency of Fy^a is .05 in Negroids and .84 in Mongoloids. If this locus is excluded from the data, the minimum and maximum distances become .041 and .077, respectively. The interracial net codon differences relative to the intraracial codon differences (R) are again small (.08–.30).

The frequency distributions of single-locus distance between Caucasoids and Negroids for protein and blood group loci are given in figure 5. It is again inverted J-shaped. In 60 of 62 protein loci, the genetic distance is less than .08, while the

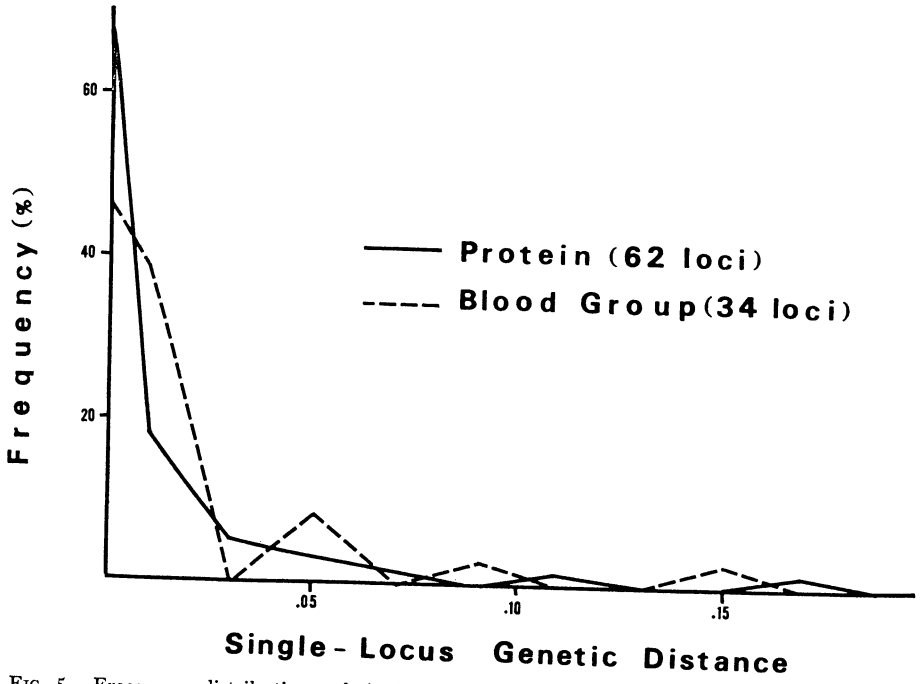

FIG. 5.—Frequency distributions of single-locus genetic distance between Caucasoids and Negroids for protein and blood group loci.

other two loci (phosphoglucomutase locus 3 and soluble glutamic pyruvic transaminase) show a distance of more than .10. The frequency distribution of genetic distance for blood group loci is approximately the same as that for protein loci.

As mentioned earlier, about two-thirds of the Negroid data were taken from American Negroids. It is known that about 20% of American Negroid genes are of Caucasoid origin, while virtually no Negroid genes have entered into the Caucasoid gene pool [10]. It is, therefore, of interest to know the genetic distances between the three major races before the gene migration occurred. The correction for this gene migration can be made by the method given by Nei [2]. If p is the proportion of Caucasoid genes in Negroids, then the average probabilities of identity of genes within and between populations before the migration are given by $J_N^{(b)} = (J_N + p^2J_C - 2pJ_{CN})/(1 - p)^2$, $J_{CN}^{(b)} = (J_{CN} - pJ_C)/(1 - p)$, and $J_{NM}^{(b)} = (J_{NM} -$

$pJ_{CM})/(1 - p)$, where subscripts C, N, and M refer to Caucasoids, Negroids, and Mongoloids, respectively. Obviously, $J_C^{(b)} = J_C$ and $J_M^{(b)} = J_M$.

Using the above formula, we recomputed the estimates of standard genetic distances from the 35 common protein and 21 common blood group loci. In this computation we used $p = .2 \times .743 = .149$ for protein loci and $p = .2$ for blood group loci, since 26 (74.3%) of the 35 protein loci and all 21 blood group loci (including one doubtful locus) were taken from American Negroids. The results obtained show that the genetic distance between Caucasoids and Negroids in increased by 35%–56% by this correction, but the absolute distance still remains small (table 5).

TABLE 5

STANDARD ESTIMATES OF NET CODON DIFFERENCES AFTER CORRECTION FOR MIGRATION OF CAUCASOID GENES INTO NEGROID GENE POOL AND EFFECTIVE DIVERGENCE TIMES BETWEEN THE THREE MAJOR RACES OF MAN

Comparison	Proteins (35 Loci)	Blood Groups (21 Loci)	Effective Divergence Time* (Yr)
Caucasoid/Negroid	.023	.042	115,000
Caucasoid/Mongoloid	.011	.034	55,000
Negroid/Mongoloid	.024	.118	120,000

* Based on protein data.

As expected, the amount of increase of the genetic distance between Negroids and Mongoloids is smaller than that between Caucasoids and Negroids. One interesting result is that the genetic distance for blood group loci between Caucasoids and Negroids is now larger than that between Caucasoids and Mongoloids, which agrees with the results from protein loci. Of course, the difference between the two genetic distances is not statistically significant.

DISCUSSION

In the section on heterozygosity, we noted that the distribution of heterozygosity at individual loci is almost the same for the three major races and that there is a small peak in the tail of the distribution located near .5 in all cases. This type of distribution could be explained by the hypothesis that a majority of the loci are selectively neutral or subject to directional selection but that a few loci exist at which two or three alleles are maintained in high frequency by means of balancing selection. This hypothesis is attractive if we note that many of the loci contributing to the peak in the tail for blood group loci are those which appear to have been polymorphic for a long time in human evolution (e.g., ABO, MN, Lewis, and Rh): these blood groups are polymorphic even in some apes and monkeys [11]. However, F. Stewart (personal communication) showed that such a peak in the tail may arise even if all loci are completely neutral. His demonstration was in terms of the model of three possible alleles at a locus with the effects of mutation and random genetic drift being balanced, but it is possible that such a peak arises even with a model of

a large number of possible alleles per locus. In fact, this type of peak is observed in the result of computer simulations by Latter [12] with a model of an infinite number of alleles per locus, although he did not pay much attention to it. Therefore, it is difficult to distinguish between the neutral and selective gene hypotheses in terms of the distribution of heterozygosity.

There is another way to test the neutral mutation hypothesis. F. Stewart (personal communication) showed that the steady-state variance of population heterozygosity at a locus (h) under the assumption of neutral mutations is given by $V(h) = 2M/(1 + M)^2(2 + M)(3 + M)$, while the mean of population heterozygosity is known to be $M/(1 + M)$, where $M = 4Nu$; N is the effective population size and u is the mutation rate per locus per generation. Even though human populations have increased rapidly in the last few centuries, the evolutionary change of heterozygosity is so slow that the above formula may be used.

Equating the average heterozygosity to the mean of the population heterozygosity, the value of M can be estimated. For example, the average heterozygosity for 74 protein loci in Caucasoids is .099, from which we estimate $M = .110$. Using this value, we get the theoretical variance $V(h) = .0272$. On the other hand, the variance of actual heterozygosity among different loci is .0318. Subtracting from this value the average sampling variance of heterozygosity (the mean of the square of the standard error of each heterozygosity value given in Appendix A), the observed variance of heterozygosity is .0317. Similar computations were made for the other heterozygosity data. The results obtained are given in table 6. The theoretical

TABLE 6

THEORETICAL AND OBSERVED VALUES OF INTERLOCUS VARIANCE OF HETEROZYGOSITY

	PROTEINS			BLOOD GROUPS		
RACE	No. Loci	Theoretical	Observed	No. Loci	Theoretical	Observed
Caucasoid	74	.0272	.0317	57	.0334	.0400
Negroid	62	.0257	.0231	34	.0388	.0425
Mongoloid	35	.0269	.0259	22	.0468	.0518

and observed variances of heterozygosity agree quite well, although in blood group loci the observed values are consistently higher than the theoretical to a slight extent, suggesting that the polymorphisms at some loci are selective or the mutation rate per locus varies considerably with each locus. These results are consistent with the view that a majority of the polymorphic alleles at protein and blood group loci are selectively neutral. It should, however, be noted that this analysis is very crude and a certain combination of different modes of selection at different loci will probably produce a similar relationship between the mean and variance of heterozygosity. Clearly, further study should be made of this problem.

The present study supports our previous conclusion that the interracial gene

differences are rather small compared to the intraracial gene differences between individuals [4]. A similar conclusion was obtained by Lewontin [13] using a different method. The interracial codon differences per locus appear to be of the same order of magnitude as those in mice and *Drosophila* [1, 2]. This conclusion, of course, does not apply to those genes which control morphological characters such as pigmentation and facial structure. Note, however, that the number of loci concerned with the difference in pigmentation between Caucasoids and Negroids appears to be rather small [14]. We believe, as we stated previously, that the genes controlling these morphological characters were subjected to stronger natural selection than "average genes" in the process of racial differentiation.

The estimates of genetic distance between the three major races given in table 5 may be used to get a rough idea about human evolution. As noted earlier, the results from protein and blood group loci are somewhat contradictory. They may be expressed as shown in figure 6. Protein data suggest two alternative evolutionary

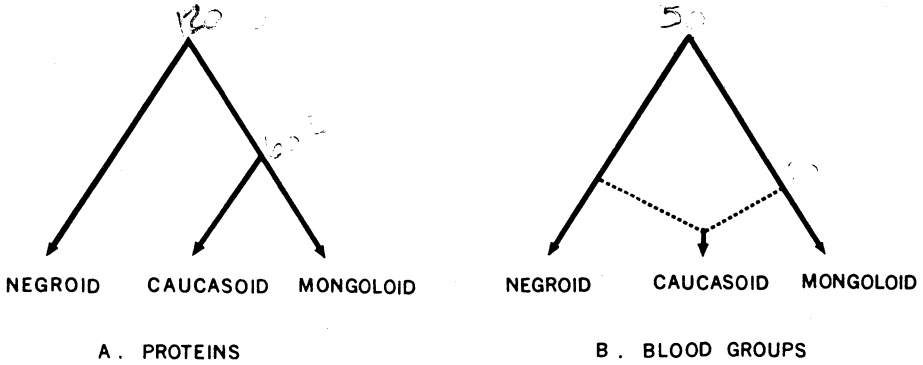

FIG. 6.—Evolutionary schemes of Caucasoids, Negroids, and Mongoloids as suggested by protein and blood group data. See text for details.

schemes. In one scheme the Negroid and the Caucasoid-Mongoloid group first diverged and then some time later the Caucasoid and Mongoloid groups evolved. In the other the divergence of the three groups occurred almost at the same time but there was more migration between the Caucasoid and Mongoloid groups than between any other pairs of groups after the divergence. On the other hand, blood group data suggest that there was some migration between the Caucasoid group and the other two groups but very little migration between the Negroid and Mongoloid groups. In this connection it should be noted that even a small amount of migration slightly higher than mutation rate prevents the gene differentiation between populations to a great extent [15]. At present, it is difficult to determine which alternative is correct, since the fossil record of modern man is very fragmentary [16]. However, since the blood group loci used here are probably not a random sample of the genome, the first two alternatives are more likely to be correct than the third. Analyzing gene frequency data for 16 blood group loci by an entirely different method, Cavalli-Sforza [17] reached the conclusion that the three major races have diverged probably in a scheme similar to our first alternative. However,

his estimates of genetic distance between the three groups fit our third alternative better than the first.

Nei [1, 18] has shown that if two populations are sexually isolated for t years and gene substitution occurs at a constant rate, the standard genetic distance may be expressed as $D = 2cn\lambda t$, where c, n, and λ are the detectability of codon differences by the technique used, the total number of codons per gene, and the rate of codon (gene) substitution per codon site per year. Therefore, if c, n, and λ are known, it is possible to estimate the divergence time. In the process of the divergence of the three major races of man, however, sexual isolation was probably incomplete. As mentioned earlier, migration between populations retards gene differentiation. Nevertheless, it is of interest to compute the divergence time between human races *on the assumption of no migration*. We call this *effective divergence time*. This effective divergence time is generally a minimum estimate of divergence time.

Virtually nothing is known about c, n, and λ for blood group loci, so that divergence time cannot be estimated from these loci. In the case of protein loci, however, it is possible to get rough estimates of these parameters. King and Jukes [19] have estimated λ to be 1.6×10^{-9} from data on amino acid substitution during evolution for seven different polypeptides. The value of c has also been estimated from amino acid substitution data in evolution [20]: it is .27. On the other hand, n can be estimated from the molecular weight of a protein subunit, although the molecular weight is known only for 24 single subunits of the proteins used here. The average molecular weight obtained for these subunits was 43,781. Since the average molecular weight of an amino acid is 110, the number of codons per locus is about 400. Therefore, the rate of electrophoretically detectable codon substitutions per locus per year ($\alpha = cn\lambda$) is estimated to be 1.7×10^{-7}. However, as argued by Kimura and Ohta [21], there is the possibility that the electrophoretically detectable changes are restricted to amino acids that are exposed to the surface of the protein molecule. Therefore, we take $\alpha = 10^{-7}$. This is identical with the estimate obtained by Kimura and Ohta under slightly different assumptions. The effective divergence time may then be estimated by $t = 5 \times 10^6 D$.

It should be emphasized that the estimate obtained by the above formula is subject to a large standard error, since our estimate of α is very crude. Note also that this method gives an underestimate of t if D is large, say, more than 1 [20]. Furthermore, if the rate of gene substitution varies with the locus, as it surely does, t is again underestimated [18]. Nevertheless, Nevo et al. [22] have shown that the estimates of divergence time obtained by this method agree quite well with the fossil record in pocket gophers.

The estimates of effective divergence times obtained from the above formula are given in table 5. Divergence times of about 120,000 years between Caucasoids and Negroids and between Negroids and Mongoloids are not inconsistent with the fossil record of modern man [16]. Our estimate of divergence time between Negroids and Mongoloids is, however, somewhat larger than that (25,000–100,000 years) of Cavalli-Sforza [17], who analyzed the blood group gene frequency data by using

an entirely different method. The effective divergence time between Caucasoids and Mongoloids is about one-half of that between Caucasoids and Negroids. This may, of course, be due to a relatively high rate of migration which might have occurred between Caucasoids and Mongoloids during the evolutionary process.

SUMMARY

The genic variation within and between the three major races of man, Caucasoids, Negroids, and Mongoloids, was studied. Surveying the literature, we collected gene frequency data for 74, 62, and 35 protein loci and 57, 34, and 22 blood group loci in Caucasoids, Negroids, and Mongoloids, respectively. Protein loci appeared to be close to a random sample of the genome, while blood group loci for Negroids and Mongoloids apparently deviated from a random sample. The proportion of polymorphic loci was 31%–51% for protein loci and 37%–72% for blood group loci. The average heterozygosity for protein loci was about 10% in all three races, while it was 13% for blood group loci in Caucasoids and 16%–24% in Negroids and Mongoloids. However, the racial differences in average heterozygosity for blood group loci virtually disappeared when the loci common to all three races were used. The distribution of heterozygosity at individual loci was inverted J-shaped with a small peak in the tail in all cases. There were high correlations in heterozygosity at individual loci between the three races. The estimate of interracial net codon differences per locus between two randomly chosen genomes for protein loci was 1%–3%, while it was 2%–14% for blood groups. The interracial net codon differences relative to the intraracial codon differences was small, indicating that the genic variation between the three races is small compared to that within the same race. Using data on the rate of amino acid substitutions in some proteins, the minimum divergence time between the three races was estimated to be about 55,000 to 120,000 years.

REFERENCES

1. Nei M: Genetic distance between populations. *Am Natur* 106:283–292, 1972
2. Nei M: A new measure of genetic distance, in *Genetic Distance,* edited by Crow JF, New York, Plenum. In press, 1974
3. Nei M: The theory and estimation of genetic distance, in *Genetics of Population Structure,* edited by Morton NE, Honolulu, Univ. Hawaii Press, 1973, pp 45–54
4. Nei M, Roychoudhury AK: Gene differences between Caucasian, Negro, and Japanese populations. *Science* 177:434–436, 1972
5. Harris H: Enzyme and protein polymorphism in human populations. *Br Med Bull* 25:5–13, 1969
6. Harris H, Hopkinson DA: Average heterozygosity per locus in man: an estimate based on the incidence of enzyme polymorphisms. *Ann Hum Genet* 36:9–20, 1972
7. Lewontin RC: An estimate of average heterozygosity in man. *Am J Hum Genet* 19:681–685, 1967
8. Race RR, Sanger R: *Blood Groups in Man.* Oxford, Blackwell, 1968
9. Nei M, Roychoudhury AK: Sampling variances of heterozygosity and genetic distance. *Genetics* 76:379–390, 1974
10. Reed TE: Caucasian genes in American Negroes. *Science* 165:762–768, 1968
11. Wiener AS, Moor-Jankowski J: Blood groups of non-human primates and their

relationship to the blood groups of man, in *Comparative Genetics in Monkeys, Apes and Man,* edited by CHIARELLI AB, New York, Academic Press, 1971, pp 71–95
12. LATTER BHD: Selection in finite populations with multiple alleles. III. Genetic divergence with centripetal selection and mutation. *Genetics* 70:475–495, 1972
13. LEWONTIN RC: The apportionment of human diversity, in *Evolutionary Biology,* vol 6, edited by DOBZHANSKY TH, HECHT MK, STEERE WC, New York, Appleton-Century-Crofts, 1972, pp 381–398
14. STERN C: Model estimates of the number of gene pairs involved in pigmentation variability of the Negro-American. *Hum Hered* 20:165–168, 1970
15. NEI M, FELDMAN MW: Identity of genes by descent within and between populations under mutation and migration pressures. *Theor Pop Biol* 3:460–465, 1972
16. BIRDSELL JB: *Human Evolution.* Chicago, Rand McNally, 1972
17. CAVALLI-SFORZA LL: Human diversity, in *Proceedings 12th International Congress of Genetics,* Tokyo, 1968, vol 3, pp 405–416
18. NEI M: Interspecific gene differences and evolutionary time estimated from electrophoretic data on protein identity. *Am Natur* 105:385–398, 1971
19. KING JL, JUKES TH: Non-Darwinian evolution. *Science* 164:788–798, 1969
20. NEI M, CHAKRABORTY R: Genetic distance and electrophoretic identity of proteins between taxa. *J Mol Evol* 2:323–328, 1973
21. KIMURA M, OHTA T: Protein polymorphism as a phase of molecular evolution. *Nature* (Lond) 229:467–469, 1971
22. NEVO E, KIM YJ, SHAW CR, THAELER CS: Genetic variation, selection and speciation in *Thomomys talpoides* pocket gophers. *Evolution* 28:1–23, 1974

(Two appendixes and their references follow.)