Population Genetics and Molecular Evolution (1985) (Ohta, T. and Aoki, K., ed.), pp. 41–64,
Japan Sci. Soc. Press, Tokyo/Springer-Verlag, Berlin

Human Evolution at the Molecular Level

MASATOSHI NEI

Center for Demographic and Population Genetics,
University of Texas at Houston, Houston, Texas 77225, U.S.A.

For the past 15 years, my colleagues and I have been studying human evolution at the molecular level by using statistical methods we developed (*1–5*). Using electrophoretic data, we first showed that the net gene differences between the three major races of man, Caucasoid, Negroid, and Mongoloid, are much smaller than the differences between individuals of the same races, but this small amount of gene differences corresponds to a divergence time of 50,000 to 100,000 years. Later, we extended our analysis to various human populations to study their evolutionary relationships in relation to geographical distribution. Recently, we have been studying the interracial variation of mitochondrial DNA (mtDNA) in man and the genetic relationships of man and apes. In these studies, we are using data on both restriction-site polymorphism and sequence variation of mtDNA. In this paper, I shall present the results of our recent studies which have been conducted in collaboration with Arun Roychoudhury, Clay Stephens, and Naruya Saitou. Specifically, I shall discuss three problems: i) evolutionary divergence of the three major races of man, ii) genetic relationships of various

human populations, and iii) phylogenetic relationship of man and apes.

EVOLUTIONARY DIVERGENCE OF THREE MAJOR RACES OF MAN

In the process of human racial evolution, gene migration seems to have occurred frequently among neighboring populations. Indeed, the genetic distances between neighboring populations are generally very small, as shown by Nei and Roychoudhury (*3*). However, European Caucasoids, Central African Negroids, and Far-Eastern Mongoloids seem to have been isolated for a long time. Coon (*6*) argued that this isolation was caused mainly by two barriers, *i.e.*, the Sahara Desert in Africa and the Movius line in Eurasia (high mountains in the west and south of Tibet). It is, therefore, interesting to know how long these three major races have been separated. This problem can be studied by using Nei's (*7*) genetic distance based on protein loci since this distance is expected to be proportional to evolutionary time. The evolutionary time can also be estimated from data on restriction site polymorphism in mtDNA (*8–11*).

1. *Electrophoretic Data*

The genetic distances (the number of codon substitutions per locus that are detectable by the biochemical technique used) between Caucasoid, Negroid, and Mongoloid for protein and blood group loci are given in Table I. Here, Caucasoid, Negroid, and Mongoloid are represented by northern Europeans (mainly English), central Africans, and far-eastern Asians (Japanese, Chinese, Koreans), respectively. The protein data in Table I indicate that Caucasoid and Mongoloid are more closely related to each other than to Negroid, so that the evolutionary relationship among the three major races becomes as given in Fig. 1A. The genetic distances computed from blood group data do not give the same genetic relationship (Fig. 1B), but the relationship is similar to that given by Cavalli-Sforza and Bodmer (*12*). If this relationship is correct, it suggests that there was a considerable amount of gene migration between Caucasoid and Negroid in the past. However, the relationship between blood group phenotype and nucleotide sequence in the gene is not clear, so that protein data seem to be more reliable. If we accept the genetic relationship obtained from protein data, we can estimate the times of divergence of these races using Nei's (*13*) method. (Actually, we estimate "effective divergence times," since our genetic distances might have been affected by migration (*2*)). The results obtained suggest that Negroid and the Caucasoid-Mongoloid group diverged about 110,000 years ago, whereas Caucasoid and Mongoloid diverged about 40,000 years ago (*3*). These estimates are in agreement with our earlier results obtained from a smaller number of loci (*2*).

TABLE I
Genetic Distances and Effective Divergence Times between the Three Major Races of Man (*3*)

Comparison	Proteins (62 loci)	Blood groups (23 loci)	Total (85 loci)	Effective divergence time (years)
Caucasoid/Negroid	0.030	0.038	0.032	113,000±34,000
Caucasoid/Mongoloid	0.011	0.043	0.019	41,000±15,000
Negroid/Mongoloid	0.031	0.096	0.047	116,000±34,000

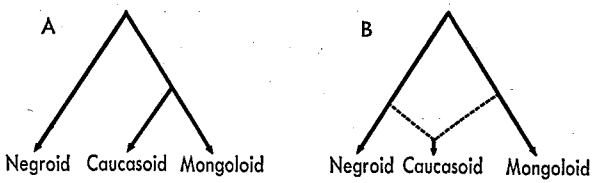

Fig. 1. Evolutionary schemes of Caucasoid, Negroid, and Mongoloid as suggested by genetic distance estimates for protein (A) and blood group (B) loci.

Around 1974, when we first published our estimates of divergence time, most anthropologists believed that modern man (*Homo sapiens*) evolved only about 25,000 years ago, after the disappearance of Neanderthals (*14*). They did not pay much attention to our estimates. In the last decade, however, a number of authors have reported fossils of modern men which are as old as 120,000 years (*15*, *16*). Therefore, our estimates are no longer incompatible with the fossil records, even if Neanderthals are not genuine *Homo sapiens*.

2. *Mitochondrial DNA*

Brown (*17*) studied the restriction-site patterns of mtDNAs of 21

individuals from the three major races. Nei (*5*) estimated the nucleotide differences per site for all pairs of these mtDNAs using the methods of Nei and Li (*8*) and Nei and Tajima (*18*). He then computed the number of net nucleotide differences between two races (d) using the following equation,

$$d = d_{XY} - (d_X + d_Y)/2, \quad (1)$$

where d_{XY} is the average number of nucleotide differences between genes of populations (or races) X and Y, and d_X and d_Y are the average number of nucleotide differences between two randomly chosen genes in populations X and Y, respectively (*8*). The expectation of d is known to be equal to $2\lambda t$, where λ is the rate of nucleotide substitution per year and t is the number of years since divergence of the two races (*8*). Nei's estimates of d are presented in Table II.

It is seen that the pattern of racial divergence as revealed by the net nucleotide differences is in agreement with that obtained from protein loci rather than with that obtained from blood group loci. Brown *et al.* (*19*) estimated the rate of nucleotide substitution (λ) in mtDNA to be 10^{-8} per site per year from their data on restriction-site maps for primates. However, a more reliable estimate is obtained from Brown *et al.*'s (*20*) nucleotide sequence data, as will be discussed later. It becomes $\lambda = 7.15 \times 10^{-9}$ per nucleotide site per year. If we use this

TABLE II
DNA Divergences (d) and Estimates of Effective Divergence Time (t) between the Three Major Races of Man

	$d \times 100$	t (years)
All 21 individuals used		
Caucasoid/Negroid	0.050±0.096	35,000
Caucasoid/Mongoloid	0.019±0.110	13,000
Negroid/Mongoloid	0.045±0.124	31,000
20 individuals used		
Caucasoid/Negroid	0.107±0.105	75,000
Caucasoid/Mongoloid	0.019±0.110	13,000
Negroid/Mongoloid	0.087±0.135	61,000

d represents the number of net nucleotide substitutions per site. t was computed under the assumption that the substitution rate (λ) is 7.15×10^{-9} per year. Eighteen restriction enzymes were used. (The data used are those of Brown (*17*))

rate, the divergence time between Negroid and Caucasoid or between Negroid and Mongoloid is estimated to be about 35,000 years, whereas the divergence time between Caucasoid and Mongoloid is about 13,000 years. Brown's (*17*) study includes one American black who was suspected to have a white female in his or her maternal lineage. Even if we exclude this individual from our analysis, the estimate of the divergence time between Negroid and Caucasoid is 75,000 years. Note, however, that the standard errors of these estimates are so large that these estimates are not really reliable.

In the hope of obtaining more reliable estimates of the divergence times, we recently analyzed Cann's (*21*) new restriction-site data. These data were obtained by comparing all restriction sites with Anderson *et al.*'s (*22*) DNA sequence so that they are more reliable than Brown's. Furthermore, since Cann used mainly four-base enzymes ($r=4$ in Nei and Tajima's (*11*) classification), her data are more informative. She studied 121 individuals from various human populations, but in our study we used 10 randomly chosen individuals from each of Caucasoid (English-origin Caucasians or northern Europeans), Negroid (Nigerians and American Blacks), and Mongoloid (Japanese, Koreans, and Chinese). Using data for 11 four-base enzymes (including five-base enzymes with $r=4$), we first estimated the number of nucleotide substitutions for all pairs of individuals. (We did not use the data for the two six-base enzymes because they were not very informative.) We then constructed a phylogenetic tree for the 30 individuals. The tree obtained is given in Fig. 2. Figure 2 shows that the individuals from different races are intermingled, though there is some tendency for the individuals from the same race to cluster. A similar intermingling of individuals from different races was observed by Cann (*21*) and Cann *et al.* (*23*). Cann (*21*) interpreted this pattern as being a result of gene migration. However, the intermingling of individuals belonging to different races is expected to occur even without migration if the ancestral population was polymorphic and the time since divergence between the populations is relatively short (*24*, *25*). This is because many of the polymorphic genes in the current populations are expected to have diverged before population splitting (see Fig. 3). It should also be noted that the time of gene splitting is usually much earlier than the time of population

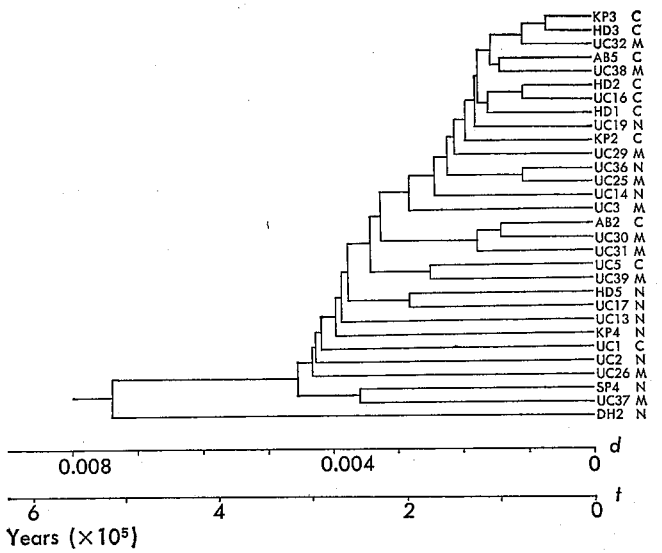

Fig. 2. Phylogenetic tree of mtDNAs for 30 individuals sampled from the Caucasian (C), Negroid (N), and Mongoloid (M) populations. (Data from Cann (*21*) were used)

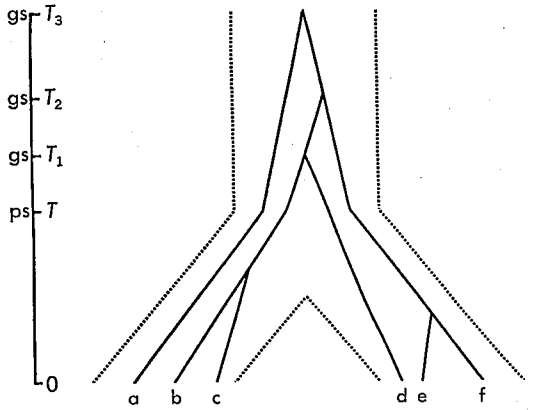

Fig. 3. Diagram showing that the time of gene splitting (T_1, T_2, or T_3) is usually earlier than the time of population splitting (T) when polymorphism exists.

splitting, and thus the former cannot be used for estimating the latter (*25*). Figure 2 shows that the oldest Mongoloid or Caucasoid gene diverged from one old Negroid gene about 300,000 years ago, but this

TABLE III

Estimates of Interpopulational (d_{XY}), Intrapopulational (d_X or d_Y), and Net (d) Nucleotide Differences among the Three Major Races of Man

	Caucasoid	Negroid	Mongoloid
Caucasoid	0.255±0.039	0.008±0.045	0.0059±0.030
Negroid	0.379±0.049	0.487±0.067	−0.0024±0.044
Mongoloid	0.308±0.029	0.416±0.044	0.350±0.032

All values are multiplied by 100. The figures on the diagonal refer to d_X (or d_Y), and those below the diagonal d_{XY}. The figures above the diagonal represent the values of $d=d_{XY}-(d_X+d_Y)/2$. (The data used are those of Cann (*21*))

time of gene divergence is almost certainly earlier than the time of racial splitting, since many other Caucasoid and Mongoloid genes diverged from Negroid genes at later times (Fig. 2).

Table III shows the estimates of d_{XY}, d_X, d_Y, and d for each pair of races that were obtained from Cann's data. It is noted that the number of intrapopulational nucleotide differences (d_X) for Negroid is twice that for Caucasoid. This is caused by the fact that a number of Negroid individuals, particularly DH2, have diverged extensively from other individuals. It is also noted that d is very small compared with d_X and has a large standard error for all pairs of races. Thus, d is again unreliable for estimating evolutionary time.

Actually, the unreliability of d for estimating evolutionary time for this case is expected from our recent theoretical study. Using the infinite-site model of neutral mutations (*26*, *27*), Takahata and Nei (*25*) studied the theoretical variance of d for various values of effective population size (N), mutation rate per nucleotide site (μ), and divergence time (g). Some of their results are presented in Table IV. It is clear that when the time since divergence between two populations is relatively small, the standard error (s_d) of d is expected to be larger than the expectation, even if a large number of genes are sampled. In the case of divergence of mtDNA between Negroid and Caucasoid, we may assume $g=5{,}000$ generations, $N=2{,}500$, and $\mu=10^{-7}$ per generation to get a rough idea of s_d. In this case, the expectation of d is $E(d)=0.001$, and s_d becomes 0.00116 for sample size $m=10$. Therefore, the standard error is expected to be larger than $E(d)$. It is noted that if we assume a larger value of N, s_d becomes even larger. This large value of

TABLE IV
Theoretical Standard Errors (s_d) of $d=d_{XY}-(d_X+d_Y)$

N	Generations since divergence	$E(d)\times 100$	Standard error ($s_d\times 100$) $m=n$		
			2	10	100
2,500	5,000	0.1	0.149	0.116	0.111
2,500	50,000	1.0	0.339	0.327	0.326
25,000	5,000	0.1	0.728	0.211	0.122
25,000	50,000	1.0	1.08	0.786	0.748
25,000	500,000	10.0	1.58	1.45	1.44
250,000	5,000	0.1	6.54	1.18	0.215
250,000	50,000	1.0	7.08	2.05	1.18
250,000	500,000	10.0	10.30	7.38	7.00

It is assumed that a gene consists of 1,000 nucleotide pairs and the rate of nucleotide substitution (mutation rate) is 10^{-7} per nucleotide site per generation. $E(d)$ is the expected value of d. (Adapted from Takahata and Nei (*25*))
N, effective population size; m, n, sample sizes from populations X and Y.

s_d is mainly due to the stochastic errors of nucleotide substitution, and thus it is not reduced appreciably by increasing sample size.

It is unfortunate that mtDNA is not very useful for estimating the time of divergence of human races despite the fact that it can easily be studied experimentally. For DNA data to be useful for our purpose, we must use many independent genes (*25*). It is, therefore, hoped that in the future many different genes from nuclear DNA will be studied. Of course, if one is interested in the evolution of more distantly related organisms, such as those of man and apes, even a single genome of mtDNA is quite useful, as will be discussed later.

It should be noted that, although it is difficult to obtain a reliable estimate of divergence time from mtDNA in the present case, some idea about the pattern of racial differentiation can be obtained from the evolutionary tree given in Fig. 2. This tree shows that one Negroid mtDNA is quite different from the other mtDNAs and all others diverged from this mtDNA about 500,000 years ago. It is also noted that many mtDNAs from Caucasoid and Mongoloid are derived from Negroid mtDNAs. Although we cannot regard this tree as the true tree, this observation suggests that Negroid diverged from the Caucasoid-Mongoloid group earlier than Caucasoid and Mongoloid diverged. This interpretation agrees with the pattern of racial differentiation inferred

from protein data (Fig. 1A). It is also in agreement with the pattern observed by Nei (*5*) in his phylogenetic analysis of Brown's (*17*) mtDNA data and that observed by Johnson *et al.* (*28*) for their own mtDNA data. Cann *et al.* (*23*) presented a phylogenetic tree of 110 mtDNAs from various human populations, showing that the oldest mtDNA exists in Australian Aborigines. However, Cann's (*21*) more extensive and careful study has shown that the oldest mtDNA actually exists in the Negroid population rather than in the Australian Aborigines. Her later study of DNA sequences (R.N. Cann, personal communication) has confirmed this pattern of mtDNA differentiation. In this connection, it is interesting to note that the oldest fossils of *Homo sapiens* were discovered in Africa (*15*, *16*).

The pattern of racial differentiation can also be inferred by using d_{XY}, which has a smaller coefficient of variation than d (see Table III). As defined earlier, d_{XY} is the average number of nucleotide differences between genes of populations X and Y, and is composed of two components, *i.e.*, i) the average number of nucleotide differences between two randomly chosen genes at the time of population splitting (d_0) and ii) the number of nucleotide substitutions after population splitting (d). If we assume that d_0 is the same for all of the three major races, d_{XY}'s give a rough idea of the pattern of population splitting. (In Eq. (1), d_0 is estimated by $(d_X+d_Y)/2$.) In Table III, the d_{XY} value between Caucasoid and Mongoloid is considerably smaller than the values between Caucasoid and Negroid and between Negroid and Mongoloid. Therefore, the pattern of population splitting is roughly similar to that of Fig. 1A.

3. *Skin-color Differentiation between Caucasoid and Negroid*

One classical study which is relevant to the times of divergence between the three major races of man is that of skin-color difference between Caucasoid and Negroid. Studying the distribution of skin pigment intensity in Caucasoid/Negroid admixed populations, Stern (*29*, *30*) estimated that the skin-color differences between Caucasoid and Negroid are controlled by 4–6 loci at which different alleles are fixed in the two populations. If Stern's estimate is correct, how many years would have been necessary for the skin-color difference to be

established after the two populations were separated? A crude answer to this question can be obtained if we assume that dark skin color is more advantageous than light skin color in the tropical region, whereas in the northern temperate region light skin color is more advantageous.

A simple way of estimating the time of skin-color divergence is to use a deterministic model of gene frequency change. For simplicity, let us assume that there are five loci involved in the skin-color difference, the genotypes for Negroid and Caucasoid being $A_{1N}A_{1N}A_{2N}A_{2N}$ . . . $A_{5N}A_{5N}$ and $A_{1C}A_{1C}A_{2C}A_{2C}$ . . . $A_{5C}A_{5C}$, respectively. We also assume that that the original ancestral population had black skin, living in tropical Africa, and later a group of individuals moved to a northern temperate region and accumulated light skin-color alleles. (At the present time, we do not know which population was ancestral in terms of skin pigmentation, Negroid or Caucasoid (or even Mongoloid), but it does not matter for our computation.) We further assume that the fitness difference between whites and blacks in the northern temperate region is 0.1 and that the allelic effect (selection coefficient) at a locus is $s=0.1/(2\times 5)=0.01$. That is, the fitnesses of $A_{iN}A_{iN}$, $A_{iN}A_{iC}$, and $A_{iC}A_{iC}$ are 1, $1+s$, $1+2s$, respectively. One might argue that s is larger than 0.01, but the fitness difference between whites and blacks does not seem to be much larger than 0.1 either in northern temperate regions or in tropical Africa.

Suppose that the original population had the A_C allele at a locus with a frequency of $p_0=0.001$ and that if the allele frequency reaches $p_t=0.999$, the allele is regarded to have been fixed. Then the time required for the frequency of allele A_C to change from p_0 to p_t is

$$t=\frac{1}{s}\log_e\frac{p_t(1-p_0)}{p_0(1-p_t)}=1{,}381 \text{ generations.}$$

In primitive human populations, one generation probably corresponds to 25 years. If this is the case, the required time will be 34,525 years.

This time, however, would be minimal because in natural populations, which are always finite, even advantageous mutations would not be fixed in the population with probability 1 (*31–33*). When the effective population size is of the order of 5,000–10,000, the probability of fixation of rare alleles is quite small. Once the A_C allele is lost from the

TABLE V

Expected Evolutionary Times (T) Required for the Skin-color Differentiation between Whites and Blacks under Various Assumptions of Effective Population Size (N), Selection Coefficient (s), Mutation Rate (v), and Initial Gene Frequency (p_0)

Case	N	s	v	T (years)		
				$p_0=0$	$p_0=0.001$	$p_0=0.01$
1	5,000	0.005	5×10^{-7}	525,000	504,000	234,000
2	2,500	0.01	10^{-7}	2,402,500	2,288,000	943,500
3	2,500	0.01	10^{-6}	262,500	252,000	117,000
4	1,250	0.02	5×10^{-6}	131,250	126,000	58,500

See text for the details.

population, the population can no longer develop light skin-color unless new mutations are introduced. New mutations are also required when the original population happens to have no A_C alleles at the time of population splitting.

Thus, a more realistic model would be a stochastic one in which gene frequency is subject to mutation, selection, and genetic drift. We consider a population of effective size N and assume that the initial frequency (p_0) of A_C at the time of population splitting is low (0, 0.001 or 0.01), but because of recurrent mutation from A_N to A_C, the population eventually develops light skin-color. Let v be the rate of mutation from A_N to A_C (advantageous allele) and s be the selective advantage of A_C over A_N as before. The expected time to fixation of the A_C allele in the population can then be studied by using Li and Nei's (*34*) theory.

This expected time to fixation depends on the values of N, s, v, and p_0 in a complicated way. However, as is clear from Table V, it is quite long unless population size is extremely small. Here, we have assumed $v\leq 5\times10^{-6}$ per locus per generation, because we are dealing only with those mutations that would affect the intensity of skin-color pigmentation. Even if $v=5\times10^{-6}$, $s=0.02$, and $p_0=0.001$, the expected time is 126,000 years. We note that the effective population size for mitochondrial genes is unlikely to be smaller than 1,250, though it is only about a quarter of that for nuclear genes (*8*). This is because the average number of nucleotide differences between two randomly chosen sequences within populations (d_x) is quite large (Table III). Note that the expectation of d_x for *neutral* mutations is $4N\mu$, where μ is the

mutation rate per nucleotide site per generation. This becomes 9×10^{-4} if N=1,250 and $\mu=25\times7.15\times10^{-9}=1.8\times10^{-7}$. The present computation therefore suggests that the divergence time between Caucasoid and Negroid is likely to be of the order of at least 100,000 years rather than the order of 25,000 years.

GENETIC RELATIONSHIPS OF VARIOUS HUMAN POPULATIONS

There seems to be no agreement concerning the classification of human races among anthropologists. Some anthropologists (*e.g.*, Boyd (*35*)) prefer to classify human races into five major groups, adding Amerindian and Australoid (or Oceanian) to the three major races we have considered. We have therefore studied the genetic relationships of various human populations within each of these five groups by using Nei's genetic distance (*7*). In this study, we have used gene frequency data for protein and blood group loci jointly because the data for protein loci were limited in some populations. The total number of loci used varied from 10 to 25, depending on the population group examined. In this paper, I shall not present all of the results because of space limitation. Rather, I present the genetic relationship of 18 representative populations of the world and then discuss the factors that caused genetic differentiation of populations.

1. Eighteen Representative Populations

The 18 populations given in Table VI have been chosen mainly because they are of anthropological interest and the gene frequency data for them are available. All five racial groups are represented by them. The genetic distances in Table VI were obtained by using 14 protein loci and nine blood group loci. Figure 4 shows the dendrogram obtained from these distances by using the unweighted pair group method (UPGMA). It is clear that the Caucasoid and Mongoloid populations are again more closely related to each other than to the Negroid populations. However, Amerindians and Australoids, who are supposed to be closely related to Asian Mongoloids, make separate clusters. This is apparently due to the effect of inbreeding that has occurred in these small tribal populations. The relatively large distance

TABLE VI
Average Heterozygosities and Genetic Distances ($D\times10^3$) Based on 23 Genetic Loci for 18 Representative Populations from the World (*3*)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
(1) Lapp	27.1																	
(2) English	1.7	26.4																
(3) Italian	1.4	0.02	26.6															
(4) Iranian	1.7	0.7	0.7	27.4														
(5) Indian	1.5	0.8	0.7	0.3	28.0													
(6) Chinese	3.4	3.7	3.9	3.2	2.4	23.3												
(7) Japanese	2.9	3.4	3.5	2.9	2.2	0.3	23.2											
(8) Malay	2.6	3.0	3.2	2.1	1.5	0.5	0.5	24.1										
(9) Polynesian	3.9	4.1	4.2	3.8	3.3	0.7	0.9	1.2	24.6									
(10) Micronesian	4.6	4.6	5.0	4.5	3.9	2.0	1.4	1.7	1.7	21.0								
(11) Braz. Indian	5.0	3.6	3.8	4.5	4.0	3.4	3.7	4.0	2.7	4.3	24.0							
(12) Alaskan Indian	5.2	5.9	6.1	7.1	6.2	4.3	4.1	4.4	4.3	5.4	4.0	17.5						
(13) Eskimo	3.4	4.7	4.6	5.4	4.7	3.8	3.2	3.8	4.1	5.3	5.1	1.4	19.3					
(14) Aust. Aborig.	5.3	5.9	5.5	6.8	5.3	3.5	3.1	4.3	5.4	4.8	6.5	5.7	4.6	13.8				
(15) Papuan	5.1	4.6	4.7	6.1	5.6	4.7	4.4	5.2	4.8	4.0	6.3	7.9	6.2	3.9	21.1			
(16) Nigerian	8.0	4.5	4.3	6.0	6.5	11.6	11.0	10.1	10.6	9.9	8.6	13.8	12.6	13.2	10.4	22.4		
(17) Bantu	7.9	4.4	4.3	5.6	6.1	11.5	10.6	9.8	10.9	10.0	9.1	13.8	11.9	13.1	11.3	0.6	20.4	
(18) Bushman	7.5	4.4	4.3	5.5	5.7	10.3	9.1	8.4	10.4	8.2	8.8	10.7	9.5	10.5	11.0	2.2	1.3	20.5

The figures on the diagonal are the average heterozygosity per locus in percent.

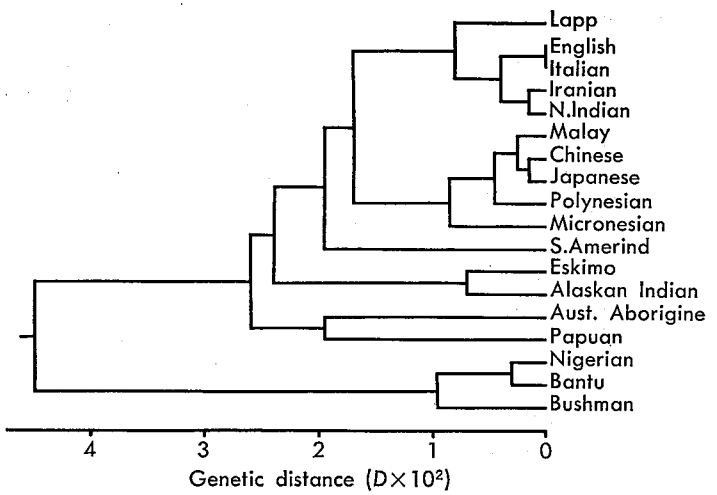

Fig. 4. Dendrogram for 18 representative races of man.

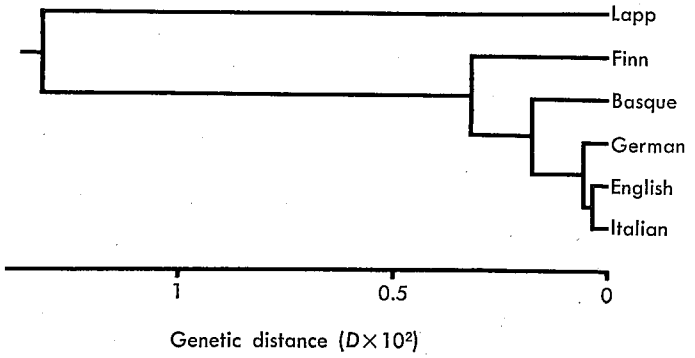

Fig. 5. Dendrogram for six European populations.

between Australian Aborigines and Papuans is also apparently caused by inbreeding. It is known that when population size is reduced drastically, genetic distance rapidly increases (*36*).

It is noted that the genetic distance between English and Italian is very small compared with the distances between other populations. It is less than 1/10 of the distance between Chinese and Japanese, and 1/200 of that between Australian Aborigines and Papuans. Actually, all western European populations are genetically closely related, as

seen in Fig. 5. Note that even the Basques who speak a non-Indo-European language are closely related to the other European populations. This is, of course, expected since there has been a substantial amount of gene migration in the recent history of Europe. In Europe, the only distantly related population is the Lapps, who are considered to have been isolated from other populations for a long time.

2. *Factors Affecting Gene Differentiation*

Our study of the genetic relationships of various human populations suggests that the most important factors affecting gene differentiation among populations are isolation and genetic drift. Figure 4 clearly shows that a pair of populations that have been isolated for a long time (*e.g.*, Bushmen and Japanese) generally show a large genetic distance. Thus, isolation is obviously the most important factor. The importance of genetic drift is indicated by the fact that a pair of tribal populations, such as Australian Aborigines and Papuans, generally have a large distance. This tendency was observed in many tribal populations in America, Africa, and Southeast Asia (*3*).

Gene migration has the opposite effect of isolation. In the process of human evolution, migration apparently occurred quite often among neighboring populations. A good example is the Beja in Sudan. These tribesmen are nomads who have lived for thousands of years in the semi-desert areas of the Red Sea Coast and the hilly country behind. They belong to Negroid, but because of geographical proximity they seem to have had gene admixture with eastern Mediterranean Caucasoids (*37*). This is reflected in the genetic distance matrix for the African and Mediterranean populations (*3*). Genetic distance data indicate that, although they are closely related to sub-Saharan Negroid populations, they are also genetically close to Italians and northern Caucasians. This clearly shows the importance of migration in making two populations genetically close.

Gene migration makes it difficult to reconstruct a phylogenetic tree that reflects the evolutionary pathways of the populations concerned. In the reconstruction of phylogenetic trees, bifurcation of populations is generally assumed. In the presence of migration, however, this assumption is no longer satisfied, and thus the phylogenetic tree

reconstructed does not necessarily reflect the true evolutionary scheme. It is, therefore, important to examine both the dendrogram and distance matrix when one wants to make any inference about evolution.

In this connection, it should be noted that the dendrogram reconstructed is subject to errors caused by the stochastic changes of gene frequencies even if there is no disturbance due to migration (*38*). These errors are quite serious when the number of loci used is small. The only way to reduce these errors is to increase the number of loci. Ideally, any dendrogram should be based on at least 30 loci.

In human populations, language can be a barrier to interracial hybridization. In practice, however, genetic distance is not clearly related to linguistic difference, except among very closely related populations (*3*). This is understandable because the language of a human population can rapidly change under certain circumstances. The relationship between genetic distance and morphological difference is also generally weak. Figure 6 shows the dendrograms for 10 human populations. The Negritos and Aboriginal Malays in Southwest Asia, Papuans in New Guinea, and Pygmies and Bushmen in Africa have a number of common morphological features such as short stature, dark skin, and frizzled hairs. Because of these similarities, some anthropologists believe that they have originated from the same common stock.

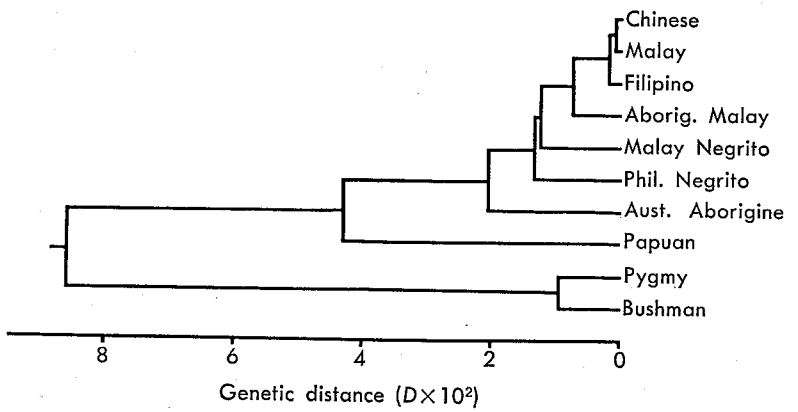

Fig. 6. Dendrogram for Negritos and their neighboring populations. Pygmies and Bushmen are included because they are phenotypically similar to Negritos.

Our genetic distance study, however, indicates that the African and Southeast Asian populations are genetically quite different, and they are generally more closely related to their neighboring populations. This indicates that the evolutionary change of morphological characters are quite different from those of average genes. Apparently morphological characters are subject to stronger natural selection than average genes (*1*).

PHYLOGENETIC RELATIONSHIP OF MAN AND APES

In recent years, the phylogenetic relationship of man and apes has been studied intensively by using various molecular data. However, relatively little attention has been paid to the accuracy of the phylogenetic tree reconstructed. In view of this circumstance, we have developed a statistical method for computing the standard errors of branching points of a tree reconstructed by UPGMA and have examined the reliability of the reconstructed trees of man and apes from four different sets of data, *i.e.*, amino acid sequences, nucleotide sequences, restriction-site polymorphisms, and electrophoretic data (*39*).

Although amino acid sequencing was started more than 20 years ago, the sequence data for man and apes are still limited. The only data that could be used for our purpose were those for hemoglobins α and β, myoglobin, fibrinopeptides A and B, and two partial sequences (13 amino acids each) of the duplicate hemoglobin γ chains for the human, chimpanzee, gorilla, and orangutan. The total number of amino acids

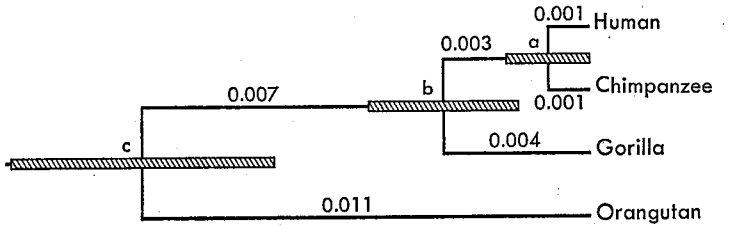

Fig. 7. Evolutionary tree for four hominoid species, which was reconstructed from amino acid sequence data. The number given for each branch represents the branch length or the number of amino acid substitutions per site. The hatched box represents 1 S.E. on each side of the mean branching point.

used was 496. We first estimated the number of amino acid substitutions per amino acid site using the Poisson correction method for all pairs of species. We then reconstructed a phylogenetic tree by using UPGMA. The tree obtained is presented in Fig. 7. The standard error of the branching points of this tree were obtained by Nei *et al.*'s (*39*) method.

Figure 7 suggests that the human and chimpanzee are more closely related to each other than to gorilla, but the standard errors of the two branching points a and b are so large that the difference between them is not statistically significant. That is, we cannot decide which organism diverged first among the human, chimpanzee, and gorilla. By contrast, the difference between branching points b and c in Fig. 7 is significantly different from 0. Therefore, the orangutan apparently diverged from the human line significantly earlier than the gorilla and chimpanzee did.

The second set of data we used was Brown *et al.*'s (*20*) nucleotide sequences of mtDNAs (a segment of 896 nucleotides) from the human, chimpanzee, gorilla, orangutan, and gibbon. The numbers of nucleotide substitutions per site were estimated by Jukes and Cantor's formula. The UPGMA tree obtained is given in Fig. 8. It is seen that the standard errors of branching points are much smaller than those of the tree in Fig. 7. Yet, the difference between the two branching points a and b is not statistically significant, whereas the difference between the branching points b and c is again significant.

The third set of data used is that of restriction-site differences for

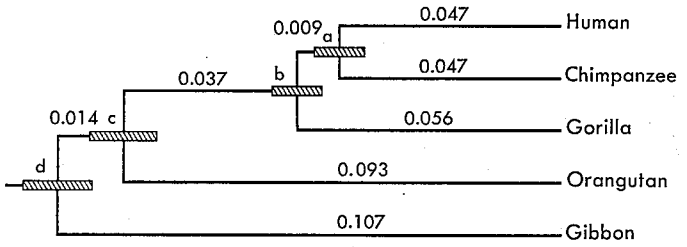

Fig. 8. Evolutionary tree for five hominoid species, which was reconstructed from nucleotide sequence data for a segment of mtDNA. The number given for each branch represents the branch length or the number of nucleotide substitutions per site. The hatched box represents 1 S.E. on each side of the mean branching point.

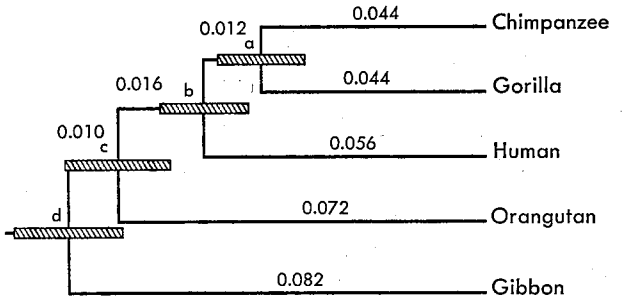

Fig. 9. Evolutionary tree for five hominoid species, which was reconstructed from restriction-site data for mtDNA. The number given for each branch represents the branch length (the number of nucleotide substitutions per site). The hatched box represents 1 S.E. on each side of the mean branching point.

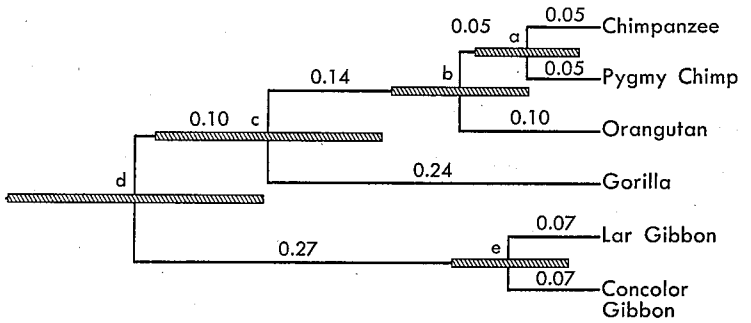

Fig. 10. Evolutionary tree for six hominoid species, which was reconstructed from electrophoretic data. The number given for each branch represents the branch length (genetic distance). The hatched box represents 1 S.E. on each side of the mean branching point.

mtDNA (*40*). In this case, the number of nucleotide differences for each pair of species was estimated by using Nei and Tajima's (*11*) maximum likelihood method. The topology of the phylogenetic tree obtained from this set of data (Fig. 9) is different from that of Figs. 7 and 8 in that the chimpanzee is closer to gorilla than to the human. However, the standard errors of the branching points of this tree are much larger than those of the tree from nucleotide sequence data. Therefore, the branching order of the human, chimpanzee, and gorilla cannot be determined from this set of data. In the present case, even the difference between branching points b and c is not statistically significant.

The final set of data used was Bruce and Ayala's (*41*) isozyme gene frequencies. The genetic distances based on 20 loci were computed for

all pairs of species by using Nei's (*7*) method. In the present case, however, the human had to be excluded because gene frequency data were not available for this species. Instead, two species of the chimpanzee and two species of the gibbon were included. The branching pattern of the chimpanzee, gorilla, orangutan, and gibbon is quite different from that of the previous trees, the orangutan now being closer to the chimpanzee than to the gorilla (Fig. 10). However, the standard errors of the branching points of this tree are so large that this tree is not very reliable.

This low reliability of electrophoretic data is partly due to the small number of loci used. In a computer simulation, Nei *et al.* (*38*) have shown that when the number of loci used is less than 30, the topology of a reconstructed tree is subject to a large stochastic error. The accuracy of a reconstructed tree also depends on the detectability of protein differences by electrophoresis. The higher the detectability, the higher the reliability. It should be noted that in Bruce and Ayala's experiment, this detectability was not particularly high. Previously, King and Wilson (*42*) had studied the genetic distance between the human and chimpanzee and obtained $D=0.62$, which is nearly two times higher than the estimate (0.39) obtained by Bruce and Ayala (*41*).

Comparison of the four trees obtained from different sets of data suggests that the tree obtained from nucleotide sequence data is more reliable because this tree has the smallest standard errors of branching points. The topology of this tree is the same as that of the tree from amino acid sequence data, though the latter does not include the gibbon. It should also be noted that this topology is in agreement with that of the trees reconstructed from both chromosomal studies (*43*) and DNA hybridization (*44*). The topology of the tree inferred from restriction-site data for ribosomal DNA (*45*) also agrees with that of Fig. 8. Therefore, this topology seems to be the most reliable one.

In a statistical analysis of the parsimony tree reconstructed by Ferris *et al.* (*40*), Templeton (*46*) concluded that the topology in Fig. 6 is significantly better than that in Fig. 8. However, his conclusion is not justified, since the parsimony method he used introduces many statistical biases when it is applied to restriction-site data (*47*).

TABLE VII
Estimates of the Times of Divergence from the Human Lineage (Million Years)

Divergence node	Sarich and Wilson (1967)	Sibley and Ahlquist (1984)	This paper[a]
Chimpanzee	5	6.3	6.6
Gorilla	5	8.0	7.8
Orangutan	8	13.0	13.0
Gibbon	10	18.2	15.0
Old World monkey	30	27.0	

[a] The rate of nucleotide substitution used is $\lambda=7.15\times10^{-9}$ per site per year. (Brown *et al.*'s (*20*) data were used)

Divergence Times of Man and Apes

If the topology given in Fig. 8 is correct, we can estimate the times of divergence among the human and ape species. For this purpose, however, we must first know the rate of nucleotide substitution (λ). Brown *et al.* (*19*) estimated the rate to be about 10^{-8} per site per year, as mentioned earlier. However, their estimate is based on restriction-site data under the assumption that the estimates of the divergence times of man and apes based on the albumin clock are correct. In practice, the estimates obtained from the albumin clock are subject to a rather large stochastic error (*48*). Therefore, it is desirable to estimate λ from nucleotide sequence data and fossil records. According to Andrews (*49*) and Pilbeam (*50*), fossil apes *Ramapithecus* and *Sivapithecus* are considered to be ancestors of the present orangutan line. If this is the case, the orangutan line seems to have diverged from the human line about 13 million years ago (see *Note added in proof* in ref. *44*). We can then estimate λ from the estimate of the number of nucleotide substitutions in the orangutan line in Fig. 5. It becomes $\lambda=0.093/(13\times10^{6})=7.15\times10^{-9}$ per nucleotide site per year. We can now estimate the times of divergence of various organisms from the human line using this value of λ and the branch lengths of the tree in Fig. 8. The results obtained are presented in Table VII, together with those obtained by Sarich and Wilson (*51*) and Sibley and Ahlquist (*44*).

It is interesting to see that our estimates of divergence times are very close to Sibley and Ahlquist's obtained from DNA hybridization data, except for the divergence time for the gibbon. This indicates that the relative branch lengths of the Sibley-Ahlquist tree are similar to

those of ours. (Here, I have used Sibley and Ahlquist's estimates based on the assumption that the divergence time for the orangutan is 13 million years.) By contrast, Sarich and Wilson's estimates obtained from their albumin clock are somewhat different from ours. They used the old world monkeys for calibrating time. At the present time, however, it is premature to make any definite conclusion about the divergence times. The fossil records we have now are not sufficient for estimating a reliable evolutionary time. We also need more extensive molecular data.

SUMMARY

1) The genetic relationship of the three major races of man, Caucasoid, Negroid, and Mongoloid, was studied by using gene frequency data for 62 protein loci and 23 blood group loci. Genetic distance estimates obtained suggest that Caucasoid and Mongoloid are somewhat closer to each other than to Negroid and that Negroid and the Caucasoid-Mongoloid group diverged about 110,000±34,000 years ago, whereas Caucasoid and Mongoloid diverged about 41,000±15,000 years ago. This pattern of racial differentiation is supported by mtDNA data, but the latter data do not give reliable estimates of divergence time.

2) The genetic relationships of various populations in each group of Caucasoid, Negroid, and Mongoloid were also studied. All European populations are genetically close to one another except the Lapps, whereas many African, Oceanian, and Amerindian tribes show extensive genetic differentiation. The major factor for this differentiation seems to be the bottleneck effect. There are indications that migration played an important role in forming the current genetic relationships of human populations. The extent of genetic differentiation among human populations is not always correlated with the degree of morphological differentiation.

3) The evolutionary relationship of the human, chimpanzee, gorilla, orangutan, and gibbon was studied by using data on amino acid substitutions in proteins, nucleotide sequences and restriction-site maps of mtDNA, and electrophoretic allele frequencies. The phylogenetic tree obtained from DNA sequence data seems to be most reliable, and this tree indicates that the species that is closest to man is the chimpanzee, and the next closest species is the gorilla.

Acknowledgments

I thank Arun Roychoudhury, Clay Stephens, and Naruya Saitou for their help in data analysis, and Peter Smouse for his detailed comments on an earlier version of this manuscript. This study was supported by grants from the National Institutes of Health and the National Science Foundation.

REFERENCES

1 Nei, M. and Roychoudhury, A.K. (1972) *Science* **177**, 434–436.
2 Nei, M. and Roychoudhury, A.K. (1974) *Am. J. Hum. Genet.* **26**, 421–443.
3 Nei, M. and Roychoudhury, A.K. (1982) *Evol. Biol.* **14**, 1–59.
4 Nei, M. (1978) *Japan. J. Hum. Genet.* **23**, 341–369.
5 Nei, M. (1982) In *Human Genetics, Part A: The Unfolding Genome* (Bonné-Tamir, B. *et al.*, eds.), pp. 167–181, Alan R. Liss, New York.
6 Coon, C.S. (1965) *The Living Races of Man*, Knopf, New York.
7 Nei, M. (1972) *Am. Nat.* **106**, 283–292.
8 Nei, M. and Li, W.-H. (1979) *Proc. Natl. Acad. Sci. U.S.* **76**, 5269–5273.
9 Gotoh, O., Hayashi, J.-I., Yonekawa, H., and Tagashira, Y. (1979) *J. Mol. Evol.* **14**, 301–310.
10 Kaplan, N. and Langley, C.H. (1979) *J. Mol. Evol.* **13**, 295–304.
11 Nei, M. and Tajima, F. (1983) *Genetics* **105**, 207–217.
12 Cavalli-Sforza, L.L. and Bodmer, W.F. (1971) *The Genetics of Human Populations*, Freeman, San Francisco.
13 Nei, M. (1975) *Molecular Population Genetics and Evolution*, North-Holland, Amsterdam.
14 Birdsell, J.B. (1972) *Human Evolution*, Rand McNally, Chicago.
15 Day, M.H., Leakey, M.D., and Magori, C. (1980) *Nature* **284**, 55–56.
16 Kennedy, G. (1980) *Nature* **284**, 11–12.
17 Brown, W.M. (1980) *Proc. Natl. Acad. Sci. U.S.* **77**, 3605–3609.
18 Nei, M. and Tajima, F. (1981) *Genetics* **97**, 145–163.
19 Brown, W.M., George, M., Jr., and Wilson, A.C. (1979) *Proc. Natl. Acad. Sci. U.S.* **76**, 1967–1971.
20 Brown, W.M., Prager, E.M., Wang, A., and Wilson, A.C. (1982) *J. Mol. Evol.* **18**, 225–239.
21 Cann, R.L. (1982) Ph.D. thesis, The University of California, Berkeley.
22 Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin,

J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schrier, P.H., Smith, A.J.H., Staden, R., and Young, I.G. (1981) *Nature* **290**, 457–465.
23 Cann, R.L., Brown, W.M., and Wilson, A.C. (1982) In *Human Genetics, Part A: The Unfolding Genome* (Bonné-Tamir, B. *et al.*, eds.), pp. 157–165, Alan R. Liss, New York.
24 Tajima, F. (1983) *Genetics* **105**, 437–460.
25 Takahata, N. and Nei, M. (1985) *Genetics* **110**, 325–344.
26 Kimura, M. (1971) *Theor. Pop. Biol.* **2**, 174–208.
27 Watterson, G.A. (1975) *Theor. Pop. Biol.* **7**, 256–276.
28 Johnson, M.J., Wallace, D.C., Ferris, S.D., Rattazzi, M.C., and Cavalli-Sforza, L.L. (1983) *J. Mol. Evol.* **19**, 255–271.
29 Stern, C. (1953) *Acta Genet.* **4**, 281–298.
30 Stern, C. (1970) *Hum. Hered.* **20**, 165–168.
31 Fisher, R.A. (1930) *The Genetical Theory of Natural Selection*, Dover Publ., Inc., New York.
32 Wright, S. (1931) *Genetics* **16**, 97–159.
33 Kimura, M. (1957) *Ann. Math. Statist.* **28**, 882–901.
34 Li, W.-H. and Nei, M. (1977) *Genetics* **86**, 901–914.
35 Boyd, W.C. (1963) *Science* **140**, 1057–1064.
36 Chakraborty, R. and Nei, M. (1977) *Evolution* **31**, 347–356.
37 El Hassan, A.M., Godber, M.G., Kopec, A.C., Mourant, A.E., Tills, D., and Lehmann, H. (1968) *Man* **3**, 272–283.
38 Nei, M., Tajima, F., and Tateno, Y. (1983) *J. Mol. Evol.* **19**, 153–170.
39 Nei, M., Stephens, J.C., and Saitou, N. (1985) *Mol. Biol. Evol.* **2**, 66–85.
40 Ferris, S.D., Wilson, A.C., and Brown, W.M. (1981) *Proc. Natl. Acad. Sci. U.S.* **78**, 2432–2436.
41 Bruce, E.J. and Ayala, F.J. (1979) *Evolution* **33**, 1040–1056.
42 King, M.-C. and Wilson, A.C. (1975) *Science* **188**, 107–116.
43 Yunis, J.J. and Prakash, O. (1982) *Science* **215**, 1525–1530.
44 Sibley, C.G. and Ahlquist, J.E. (1984) *J. Mol. Evol.* **20**, 2–15.
45 Wilson, G.N., Knoller, M., Szura, L.L., and Schmickel, R.D. (1984) *Mol. Biol. Evol.* **1**, 221–237.
46 Templeton, A.R. (1983) *Evolution* **37**, 221–244.
47 Nei, M. and Tajima, F. (1985) *Mol. Biol. Evol.* **2**, 189–205.
48 Nei, M. (1977) *J. Mol. Evol.* **9**, 203–211.
49 Andrews, P. (1982) *Nature* **295**, 185–186.
50 Pilbeam, D. (1984) *Sci. Am.* **250**(3), 84–96.
51 Sarich, V.M. and Wilson, A.C. (1967) *Science* **158**, 1200–1203.

Population Genetics and Molecular Evolution (1985) (Ohta, T. and Aoki, K., ed.), pp. 65–83, Japan Sci. Soc. Press, Tokyo/Springer-Verlag, Berlin

Two Elements of a Unified Theory of Population Genetics and Molecular Evolution

ROGER MILKMAN

Department of Biology, The University of Iowa, Iowa City, Iowa 52242, U.S.A.

The past 25 years, and more specifically 16 years of the neutral theory of molecular evolution (*1*), have brought new data and new approaches to the understanding of population genetics and molecular evolution.

This understanding is now beginning to take the form of a unified theory (*2*). Two important aspects of this unified theory are the following: first, the genetic structure of a species appears to be determined fundamentally by stabilizing selection acting on a composite phenotype influenced by many genes whose alleles have small, essentially additive effects. At phenotypic equilibrium, these alleles become neutral in that their frequencies are governed by random genetic drift. Second, it has been possible to find strong empirical evidence of the neutrality of a class of genetic differences, and these can now be used as a standard in examining species structure and evolutionary relationships. I should like to discuss each of these elements.