Proceedings of the National Academy of Sciences
Vol. 65, No. 2, pp. 323–330, February 1970

Village and Tribal Genetic Distances among American Indians, and the Possible Implications for Human Evolution*

James V. Neel and Richard H. Ward†

DEPARTMENT OF HUMAN GENETICS, UNIVERSITY OF MICHIGAN MEDICAL SCHOOL, ANN ARBOR

Communicated November 14, 1969

Abstract. Pair-wise genetic distances based on six genetic systems (Rh, MNSs, Kidd, Duffy, Diego, and haptoglobins) are presented for seven villages of Makiritare Indians, seven villages of Yanomama Indians, and 12 Indian tribes of Central and South America. It is shown that the mean genetic distance between Indian villages is 85–90 per cent of the distance between tribes. Since in the past, the initial event in the formation of a new tribe was probably the breaking away of one or several related villages from an established tribe, it is clear that this initial event could have profound consequences for subsequent tribal gene frequencies. By the criterion of changes in gene frequency in polymorphic genetic systems, the maximal rate of evolution in the American Indian has been approximately 100 times more rapid than the mean rate suggested by calculations based on amino acid substitutions in certain polypeptides of a variety of organisms. The above-mentioned findings could account in part for this 100-fold difference. Some other factors which might diminish the apparent difference between the results of these two types of calculations are mentioned.

Introduction. The tempo and driving forces of biological evolution are central issues in human population genetics. Recently we have described the genetic distances between 12 tribes of Central and South American Indians, and on the basis of these distances between tribes and the probable time of arrival of the Indian in Central and South America, constructed a phylogeny of the tribes and estimated the *maximal* rate of gene substitution in these groups. This estimate was 130,000 years/gene substitution/locus in any one line of descent since the arrival of the Indian in Central and South America, where gene substitution was defined in terms of the additive components of the vector space utilized to find the genetic distance between populations.[1] The estimate is based on cumulative change at only six genetic loci (the MNSs, Rh, Kidd, Duffy, Diego, and haptoglobin loci), which may or may not be representative of the genome as a whole. Among the several assumptions in this approach, two which are paramount are that changes in diverse directions at various loci can be equated to directional change at one locus, and that the coefficient of selection is approximately the same at all gene frequencies—i.e., that frequency-dependent selection is relatively unimportant. The tribes were chosen for consideration from a tabulation of all

the gene frequency data on the American Indian[2] solely on the basis of meeting the following critiera: (1) sample size greater than 200, (2) non-Indian admixture estimated at less than 5 per cent, and (3) data on the above six genetic loci plus the ABO locus. Since all the unmixed Indians of Central and South America appear to be type O, this latter locus is of little value in distance measurements but very important in opinons concerning admixture with non-Indians.

We have also recently described the gene frequencies encountered in ten villages of the Yanomama Indians of Southern Venezuela and Northern Brazil (included as one of the 12 tribes mentioned above),[3] and will shortly describe the gene frequencies of seven more villages of an adjacent tribe, the Makiritare Indians.[4-6] In choosing these two tribes for that type of study, we were strongly influenced by their relative lack of acculturation, i.e., retention of the aboriginal tribal population structure. In both of these tribes, the degree of genetic differentiation from one village to the next was felt to be noteworthy. These genetic differences were primarily attributed to the fact that a new Indian village usually results from the fissioning of an established village, this fissioning structured by factors of kinship, so that the division of the village (and tribal) gene pool at the time of a split is highly nonrandom.[7] Thus we see these differences as usually nonadaptive, although in theory a set of gene frequencies closer to a new adaptive peak than those of the mother village might occasionally result.

Finally, we have also suggested that in the past, the initial event in the formation of a new tribe may often have consisted of one of the products of a village fission, or several closely related villages, wandering so far from the other villages of the mother tribe that it or they became the nucleus for a new tribe.[8] There was thus seen the opportunity for a large stochastic element in the establishment of the gene frequencies of a newly formed tribe, with the present gene frequency (i.e., the "now" plane) determined by the subsequent interplay of selection, migration, genetic drift, and mutation.

The present note will atempt to begin to explore the extent to which the above-mentioned aspects of the social organization of the American Indian may have influenced the nature and tempo of his evolution, as measured by changes in gene frequency. We recognize that not all students of evolution will accept the events measured as evolution (rather than random fluctuations in gene frequency), but we believe an exploration at this time to be warranted. It will be shown that stochastic events related to social structure *may* play a very important role in rates of gene frequency change. By inference, the same could be true of differences between the tribal populations of other major land masses, and to some extent of differences between larger ethnic groupings.

The Data. On the basis of the above-mentioned data, three matrices have now been calculated, one for all the pair-wise genetic distances between 12 Indian tribes (the tribes having been selected for consideration without reference to their genetic distance from one another), the other two similar matrices for 7 Yanomama and 7 Makiritare villages. Although the argument to be developed could have been based solely on the distance between tribes and between Yanomama villages, it has been delayed until the recent confirmation of the Yanomama findings by those on the Makiritare. Distance is based on the cumulation of differences in the frequencies of each of the corresponding

alleles of each of the six genetic systems employed. In this instance we have elected to represent the pair-wise distance between populations in Euclidean hyperspace, with the contributions from each locus summed in the hyperplane by the theorem of Pythagoras, after the method of Cavalli-Sforza and Edwards.[9] Hence each set of populations can be conceived as being embedded in Euclidean hyperspace with a set of Cartesian coordinates for each population derived as a function of the calculated distance. Such a procedure eliminates the problem of representing the set of distances in curvilinear space, enabling the comparison of distance functions for different sets of populations to be made with greater validity, despite the fact that the distance deviates from the actual number of transformed gene substitutions and thus our measure of genetic distance is not directly defined in terms of gene substitutions. For a single locus with m alleles, two populations are given the distance $d = (2\sqrt{2}/\pi)\sqrt{1 - \cos\theta}$, where $\cos\theta = \Sigma_{i=1}^{m}\sqrt{p_i q_i}$ and p_i, q_i are the frequencies in the respective populations of the ith allele. The total distance, D, is obtained from the distances resulting for each of the six loci by squaring each and extracting the square root of the sum, i.e., $D = \sqrt{\Sigma^{6}_{j=1} d_j^2}$.

The tribal and the two village matrices are given in Tables 1 and 2. Since only 7 Makiritare villages have been investigated, for the symmetry and stability of the comparison we have reduced the 10 Yanomama villages studied to the same number by excluding one "mixed" village and the two smallest. On a simple percentage scale, the

TABLE 1. *Matrix of pair-wise genetic distances for 12 South American tribes. (Description of construction in text.)*

Tribe	Cakchiquel	Cayapa	Cuna	Guayami	Jivaro	Pemon	Quechua	Shipibo	Xavante	Yanomama	Yupa
Aymara	.260	.301	.355	.485	.370	.381	.288	.393	.374	.514	.450
Cakchiquel		.297	.224	.364	.342	.302	.278	.363	.250	.439	.326
Cayapa			.283	.446	.289	.346	.224	.486	.343	.473	.328
Cuna				.327	.381	.283	.331	.466	.227	.479	.239
Guayami					.444	.469	.398	.645	.410	.437	.433
Jivaro						.402	.270	.521	.375	.536	.433
Pemon							.319	.460	.371	.510	.354
Quechua								.433	.336	.479	.392
Shipibo									.335	.660	.479
Xavante										.549	.249
Yanomama											.453

TABLE 2. *Matrices of genetic distance between paired villages of (a) Makiritare and (b) Yanomama Indians. (Explanation in text.)*

(*a*) Makiritare: 7 Villages (6 loci)

	Distance Matrices					
Village	BD	C	E	F	G	HI
A	.362	.558	.353	.345	.268	.336
BD		.250	.221	.432	.314	.296
C			.393	.588	.485	.444
E				.379	.249	.273
F					.394	.383
G						.158

(*b*) Yanomama: 7 Villages (6 loci)

	Distance Matrices					
Village	BD	C	D	E	H	I
A	.227	.228	.385	.157	.416	.243
B		.367	.506	.144	.537	.360
C			.298	.295	.346	.297
D				.464	.154	.364
E					.486	.296
H						.350

mean Yanomama and Makiritare pair-wise village distances are 85.5 and 92.4%, respectively, of mean tribal distances (Table 3). However, following our argument that a village may break away to become the nucleus of a new tribe, the distance which is more germane is the mean distance of each village from a tribal mean based on all the others (Table 4). This figure for the Yanomama is 0.260 and for the Makiritare, 0.263, amounting to 67.5 and 68.1%, respectively, of the mean tribal distance. Although measures of dispersion and tests of the significance of differences for data such as this are highly desirable, in our opinion suitable statistics are not yet available. However, the general argument which follows would not in principle be dependent upon the outcome of such tests.

TABLE 3. *A comparison of certain aspects of the pair-wise genetic distance for various populations.*

Group	No. of populations	Mean pair-wise genetic distance	Smallest pair-wise genetic distance	Largest pair-wise genetic distance	Dispersion index*
Makiritare	7	.356	.158	.588	.197
Yanomama	7	.330	.144	.537	.164
South American Indians	12	.385	.224	.660	.238

* The dispersion index is based on the ratio of minimum network length to the number of populations studied, where minimum network length was obtained employing Edwards' procedure of embedding the component populations in a Euclidean space using the "tightest" topological configuration possible.

TABLE 4. *Genetic distance between each of the villages sampled and the sum of all the other villages sampled from the same tribe, as approximated by the weighted average of the other six villages.*

	Village							
	1	2	3	4	5	6	7	Mean
Tribal remainder:								
Makiritare	.327	.222	.377	.167	.359	.181	.206	.263
Yanomama	.178	.313	.187	.324	.245	.351	.225	.260

In any comparison involving Yanomama village distances, it is well to recall that one of the six systems on which the distance function is based (Diego) is invariant in this group of Yanomama villages, but quite variable in both the Makiritare villages and the other tribes. Furthermore, the sample of Yanomama villages is drawn from a limited area in the central portion of the Yanomama distribution; it seems possible the more extensive sampling now in progress will result in a revision upwards of mean village genetic distance. The Makiritare villages sampled are scattered through a proportionately greater portion of the tribal distribution and probably reflect more accurately mean village distances.

Discussion. Although the implications of this demonstration of the magnitude of the mean genetic distance of any one village from the remaining villages of a tribe are not yet entirely clear, several alternatives present themselves for consideration. If the final gene frequencies of a tribe stemming from one of these break-away villages are determined primarily by the further action of what for these alleles appear to be nondirectional (i.e., stochastic) forces, then the nature of this initial step in the formation of a new tribe may have greater implications for the ultimate gene frequencies of the tribe than all subsequent events. If, on the other hand, one maintains a strictly deterministic position, that the ultimate genetic differences between tribes of American Indians result from the

response to localized selective forces, then selection must undo the result of a major stochastic event and/or conserve those villages whose gene frequencies are most adaptive. In point of fact, of course, this is not an either-or situation—both drift and selection may have significant roles to play, with the contribution of these factors varying from one locus to the next, and the challenge is to apportion to each its contribution. In the absence of comparable data on other relatively equally *undisturbed* mammalian populations, it is impossible to state whether the subpopulations of man are more variable than other mammalian subpopulations, but this may well be the case. In any event, these data on village distances, together with the findings as to how the distances arise, make it clear that the equilibrium population so necessary to many genetic formulations is a statistical abstraction, rarely if ever realized, the misleading character of which is only justified by its usefulness to the mathematician.

A direct approach to the evaluation of the role of selection in determining Indian gene frequencies will be difficult, because of the labor involved in amassing, for the sparse and inaccessible populations still suitable for such studies, the large numbers necessary to indicate (or exclude) reasonable selection coefficients,[10] especially in view of the imminent acculturation of so many of the remaining primitive groups. More promising are indirect approaches, in which one searches for patterns of allele frequencies, in relation to other alleles and ecological situations, patterns most reasonably attributed to selection. Another type of indirect approach will be simulation, in which, given knowledge of tribal breeding structure, one studies the extent to which drift, opposed by reasonable amounts of in-migration, could account for the genetic distances between villages and tribes. A review of present knowledge of gene frequencies in American Indian tribes fails to reveal convincing regularities in relation to ecological features such as would suggest the operation of systematic events.

Cavalli-Sforza, Barrai, and Edwards[11] have presented similar data on genetic distances on some 15 world populations, including representatives of most of the major ethnic groupings. The genetic systems which they employed (A_1A_2BO, MNS, Rh, Duffy, and Diego) differ somewhat from those utilized in the present paper, so that a precise comparison with our findings is precluded. However, as an approximation one can say that just as the mean genetic distance between Indian villages is well over half the mean distance between Indian tribes, so the mean distance between Indian tribes appears approximately half the mean distance between these 15 populations. Thus on a world-wide basis we must consider the possibility that in the peopling of the world, when and if a tribe destined to found a major ethnic grouping moved into a new area, then, subject to the interpretations to be discussed below, half of the genetic differentiation may already have been accomplished.

Kimura[12] has calculated on the basis of species differences in the precise amino acid composition of a number of well studied proteins that evolution proceeds over a wide range of the animal kingdom at the rate of approximately one amino acid substitution in 2.8×10^7 years for a polypeptide chain consisting of 100 amino acids (a relatively small polypeptide). On the basis of more extensive data, King and Jukes[13] have revised this estimate, for a similarly sized polypep-

tide, to 1.6×10^7. It may be presumed that the allele differences with which our treatment is concerned are also ultimately based on amino acid substitutions (or their equivalent). If this is granted, then our *maximal* rate of amino acid substitution, of 1.3×10^5 years, would thus be more than 100 times greater than the rate determined by King and Jukes,[13] a difference especially striking in view of the greater length of the human life cycle than that of most of the other organisms involved in Kimura's calculation.

The initial event in animal speciation is often probably the geographic isolation of a small group of organisms, which, lacking man's unique social organization, may usually be more representative of the total gene pool from which they are drawn than is an Indian village (even though they too depart widely from the population mean). If this assumption is correct, and if the gene frequencies established at the initial event (village break-away) need not always be "undone" by selection (see below) but can be the point of departure on which drift or selection operates, the rate of gene (amino acid) substitution in these Indian populations subsequent to the initial step towards a new tribe would be substantially less than (perhaps half) the rate of 1.3×10^5 years which we have derived. But even when allowance is made for this initial event in tribalization, the subsequent *maximal* tempo of human gene substitution still appears to be about 50 times greater than the mean calculated for a variety of other forms. Four further factors suggest that there may be no real inconsistency between our maximal estimate and the mean rate derived by others: (1) Most polypeptides are longer than 100 amino acids. (2) In view of the marked genetic microdifferentiation noted above, the sampling of the tribes yielding this maximal estimate may have been inadequate. (3) By basing the estimate on polymorphisms, there may have been selection for the more rapidly evolving polypeptides, i.e., no correction has been made for the fact that in Drosophila, mouse, and man, only about 30 per cent of the loci are polymorphic at any one time.[14-16] (4) The maximum exceeds the mean, by an amount which cannot be estimated at present. Further, as noted above, the fluctuations in gene frequency on which our estimate is based do not have the directionality of the genetic change entailed when one amino acid is replaced by another in all the members of a species. While the downward adjustments in evolutionary rate suggested by these factors would tend to bring our estimate more in line with Kimura's estimate, it is not now clear whether these adjustments would entirely erase the apparent difference. Although the propriety of a direct comparison of the results of these two approaches, as well as the "adjustments" in both approaches, will no doubt be under discussion for some time, the facts as they stand can be interpreted as indicating that in the American Indian evolution as regards "structural" genes was somewhat more rapid than the average in other forms, in part because of the initial event in the genesis of a new tribe but also we suspect in part the result of subsequent forces.

Kimura[12] (see also King and Jukes[13]) has argued that the majority of these amino acid substitutions are selectively equivalent to their predecessors, i.e., selectively neutral, much of this evolution thus being non-Darwinian. In this argument he depends heavily on the apparent constant rate of amino acid sub-

stitution with time in all lines of descent, and on a formulation of the "cost" of natural selection, i.e., the genetic deaths resulting from the substitution of an allele with more favorable phenotypic effects for its predecessor. With respect to the latter point, Sved[17] and Maynard Smith[18] have questioned this formulation and developed an alternate approach which greatly lowers the "cost" of the substitution of a favorable gene. O'Donald[19] has pointed out factors which reduce the force of Maynard Smith's argument (without a full return to Kimura's). Furthermore, Franklin and Lewontin[20] demonstrate how linkage disequilibrium undermines calculations concerning the cost of gene substitution based on the assumption of independent gene segregation. With respect to the former point, the apparent relatively constant rate of substitution of amino acids in various lines of descent, we suggest that this may be related to factors of breeding structure (for example, small effective population size correlated with longer generation time), so that the apparent time independence is a misleading coincidence. Incidentally, some of the Indian villages included in this survey are lacking genes present in other villages in frequencies as high as 0.24 (Arends *et al.*[3]), pointing to the possibility of gene fixation at the time of fission rather than through subsequent selection and drift. Similarly, one must consider the possibility that sometimes the amino acid substitutions which distinguish species arise very early in the course of speciation.

For the moment we can only accept the facts of amino acid substitution while the theory remains *sub judice*. There is likewise no basis at present for deciding whether the polymorphisms on which our treatment is based are neutral, balanced, or transient (or some mixture thereof), as these terms are conventionally used. If, however, after the initial break-away, evolution is essentially Darwinian rather than non-Darwinian, the possible faster tempo of human evolution compared to that of other organisms is best explained by the subsidiary hypothesis that those break-away villages whose gene pool had the highest adaptive content would be most apt to become the nucleus for a new tribe. Thus, whether subsequent evolution be regarded as Darwinian or non-Darwinian, the initial step has great importance.

In addition to the manner in which a village divides at the time of a fission, there are other aspects of tribal social organization, such as the prerogatives of headmanship, migration of nuclear families, and the fate of captured women, as a result of which the small villages of primitive populations may rapidly become genetically differentiated from one another.[21] Wright[22] has argued that evolution occurs most rapidly when the population structure of the species results in local, genetically differentiated subdivisions, the adaptive value of whose corporate genome can then be tested by the environment. Thus although the maximal rate of gene substitution which we have estimated is undoubtedly an overestimate of the mean rate, the possibility exists that man's social structure has created opportunities for a more rapid rate of evolution, both deterministic and non-deterministic, than exist for many other species. Herein may lie a partial explanation of the opinion of some students of evolution that in some respects man has evolved more rapidly than other large mammals.[23, 24] Obviously, the precise population structure may result in relatively faster or slower

rates of evolution. It will be of interest to compare the present findings with the results of a similar treatment of such populations as the Australian aborigines or the Polynesians, where the data concerning gene frequencies, social structure, and the time of dispersal throughout an area are becoming available.

* This investigation was supported in part by U.S. Atomic Energy Commission grant AT(11-1)-1552.

† Rackham Predoctoral Fellow of the University of Michigan.

[1] Fitch, W., and J. V. Neel, *Amer. J. Hum. Genet.*, **21**, 384 (1969).

[2] Post, R. H., J. V. Neel, and W. J. Schull, in *Biomedical Challenges Presented by the American Indian* (Washington: PAHO Scientific Publications, 1968), no. 165, p. 144.

[3] Arends, T., G. Brewer, N. Chagnon, M. Gallango, H. Gershowitz, M. Layrisse, J. Neel, D. Shreffler, R. Tashian, and L. Weitkamp, these PROCEEDINGS, **57**, 1252 (1967).

[4] Gershowitz, H., M. Layrisse, J. V. Neel, C. Brewer, N. Chagnon, and M. Ayres, *Amer. J. Hum. Genet.*, in press.

[5] Arends, T., L. R. Weitkamp, M. L. Gallango, J. V. Neel, and J. Schultz, *Amer. J. Hum. Genet.*, in press.

[6] Weitkamp, L., and J. V. Neel, *Amer. J. Hum. Genet.*, in press.

[7] Neel, J. V., and F. M. Salzano, *Amer. J. Hum. Genet.*, **19**, 554 (1967).

[8] Neel, J. V., *Jap. J. Hum. Genet.*, **12**, 1 (1967).

[9] Cavalli-Sforza, L. L., and A. W. F. Edwards, *Amer. J. Hum. Genet.*, **19**, 233 (1967).

[10] Neel, J. V., and W. J. Schull, *Persp. Biol. Med.*, **11**, 565 (1968).

[11] Cavalli-Sforza, L. L., I. Barrai, and A. W. F. Edwards, in *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 29 (1964), p. 9.

[12] Kimura, M., *Nature*, **217**, 624 (1968).

[13] King, J. L., and T. H. Jukes, *Science*, **164**, 788 (1969).

[14] Shaw, C. R., *Science*, **149**, 936 (1965).

[15] Harris, H., *Proc. Roy. Soc., London, Ser. B.*, **164**, 298 (1966).

[16] Lewontin, R. C., and J. L. Hubby, *Genetics*, **54**, 595 (1966).

[17] Sved, J. A., *Amer. Nat.*, **102**, 283 (1968).

[18] Smith, J. M., *Nature*, **219**, 1114 (1968).

[19] O'Donald, P., *Nature*, **221**, 815 (1969).

[20] Franklin, I., and R. C. Lewontin, *Genetics*, in press.

[21] Chagnon, N. A., J. V. Neel, L. Weitkamp, M. Layrisse, H. Gershowitz, and M. Ayres, *Amer. J. Phys. Anthrop.*, in press.

[22] Wright, S., *Genetics*, **16**, 97 (1931).

[23] Haldane, J. B. S., *Evolution*, **3**, 51 (1949).

[24] Mayr, E., *Animal Species and Evolution* (Cambridge: Belknap Press, 1963), pp. xiv and 797.