J. Mol. Evol. 11, 75–85 (1978)

Journal of Molecular Evolution

On Investigating the Statistical Properties of the Populous Path Algorithm by Computer Simulation

Counterconclusions to Those of Tateno and Nei

John Czelusniak[1], Morris Goodman[1], and G. William Moore[2]

[1]Department of Anatomy, Wayne State University, School of Medicine, 540 E. Canfield Avenue, Detroit, Michigan 48201, USA
[2]Department of Pathology, The Johns Hopkins Hospital, Baltimore, Maryland 21205, USA

Summary. Goodman et al.'s (1974) populous path algorithm for estimating hidden mutational change in protein evolution is designed to be used as an adjunct to the maximum parsimony method. When the algorithm is so used, the augmented maximum parsimony distances, far from being overestimates, are underestimates of the actual number of nucleotide substitutions which occur in Tateno and Nei's (1978) computer simulation by the Poisson process model, even when the simulation is carried out at two and a half times the sequence density. Although underestimates, our evidence shows that they are nevertheless more accurate than estimates obtained by a Poisson correction. In the maximum parsimony reconstruction, there is a bias towards overrepresenting the number of shared nucleotide identities between adjacent ancestral and descendant nodal sequences with the bias being stronger in those portions of the evolutionary tree sparser in sequence data. Because of this particular property of maximum parsimony reconstructed sequences, the conclusions of Tateno and Nei concerning the statistical properties of the populous path algorithm are invalid. We conclude that estimates of protein evolutionary rates by the maximum parsimony – populous path approach will become more accurate rather than less as larger numbers of closely related species are included in the analysis.

Key words: Maximum parsimony – Populous path algorithm – Computer simulation

Several authors (Nei and Chakraborty, 1976; Ohta, 1976) have stated that the populous path algorithm for detecting hidden nucleotide substitutions (Goodman et al., 1974; Moore et al., 1976) may produce too great an augmentation. Investigating this matter by computer simulation, Tateno and Nei (1978) now claim to have found the predicted systematic error of the populous path algorithm overcompensating for hidden substitutions. Their claim, however, is unsupported by valid evidence. The populous path algorithm is designed to be used as an adjunct to the maximum parsimony method (Moore et al., 1973). Yet Tateno and Nei did not use the maximum

0022-2844/78/0011/0075/$02.20

parsimony method in their study. If they had done so, they would have observed, as we have, that the augmented parsimony distances did not overestimate but, instead, underestimated the number of hidden nucleotide substitutions. Moreover underestimation still occurs even when computer simulation by the Poisson model is carried out on a considerably larger scale than in their study.

Tateno and Nei employ a computer simulation with a hypothetical cistron composed of 100 codons or 300 nucleotide residues, the first half of the cistron subject to random mutations and the second half maintained as invariable. The portion of variable nucleotide residues are then assigned a Poisson determined number of mutations which are distributed randomly. While this is a conceivable model for molecular evolution at the nucleotide level, it is only one of many possible models, and not necessarily the one which has taken place in nature. In fact, the computer simulation model of Tateno and Nei bears little relationship to either the actual course of globin evolution, or to the maximum parsimony reconstruction of this course other than for a superficial resemblance in the branching pattern in the tree used in the computer simulations to those found in maximum parsimony reconstructions (Goodman et al., 1975; Beard and Goodman, 1976). With regard to the actual course of globin evolution almost all residue positions have been found to be variable. Indeed even during the more restricted period of teleost-tetrapod phylogeny within either the alpha globin or the beta globin family, although any two contemporary sequences rarely vary by any more than 50% of their codons, our current data show that throughout either the alpha or beta family about 85% of codon positions have varied. Actually, in practice on using the Poisson correction for estimating hidden mutational change, Nei and Chakraborty (1976) found it necessary to assume that all positions are potentially variable. More importantly, the computer simulation model of Tateno and Nei fails to account for those aspects of the covarian phenomenon described by Fitch and Markowitz (1970) in which different sets of residue positions are first variable and then become invariable in different lines of descent in a pattern suggestive of the operation of natural selection rather than random change. A striking example of this is provided by the residue positions responsible for cooperativity in tetrameric hemoglobin of amniote vertebrates. The maximum parsimony reconstruction indicates that in pre- and early vertebrates these positions were highly variable, but then became virtually invariable in the descent of birds and mammals from their common ancestor of 300 or so million years ago (Goodman et al., 1975). Other contradictions to the random model also exist, e.g. the higher incidence of guanine to adenine and adenine to guanine transitions over other nucleotide substitutions in the mRNA codons revealed by the maximum parsimony method (e.g. Fitch, 1972; Goodman et al., 1974; Goodman and Moore, 1977) and by inspection of human hemoglobin variants (Vogel and Kopun, 1977) as well as the higher incidence of observable substitutions in the maximum parsimony reconstructions at the first nucleotide position of the codons rather than at the second, although in the random model just the opposite would be expected.

For our own part, we have sought methods for reconstructing evolutionary trees which make the fullest use of the data themselves in the reconstruction process, as opposed to *a priori* assumptions about statistical properties of the data. We use a mathematically rigorous procedure (Moore et al., 1973) for constructing maximum parsimony ancestral mRNA sequences on a given tree. Our procedure is grounded in

concepts described by Farris (1970), in particular those pertaining to unrooted trees, and generalizes the method of Fitch (1971) for nucleotides to codons. It meets the complications posed by the degeneracy of the genetic code and is guaranteed to give the most parsimonious solutions for each given tree.

The reconstructed ancestral nucleotide sequences can then be used to correlate changes in the primary structure of a given polypeptide with evolutionary changes in function, as was done with the globin molecule (Goodman et al., 1975). Computer methods which do not make such explicit predictions with respect to ancestral sequences are less subject to control by comparison with independent data about evolutionary change.

We have employed the same principles in developing the populous path augmentation procedure: primary orientation toward the data themselves rather than *a priori* statistical assumptions, methods supported by rigorous mathematical proof, and use of maximum parsimony ancestral sequences to find the minimal number of nucleotide changes in the descent of the contemporary sequences from their common ancestor. Our augmentation approach is based on the particular properties of genealogical trees constructed by the maximum parsimony method. We have observed in these trees that as the density of nodal points between two species increases with the addition of new sequence data from related species so does the unaugmented parsimony distance over the path increase. In conformity with this observation, the populous path algorithm fills out the amount of mutational change in portions of the evolutionary tree with few intervening links by making them comparable to portions with many intervening links. The procedure employed for propagating the mutational information in the densest portions of the tree throughout the tree as a whole is designed to minimize the number of added new mutations. The formal proof that the algorithm does what it purports to do (Moore, 1977) shows that in a maximally dense tree where each unaugmented link distance is zero or unity there is no augmentation, i.e. AF is always zero. This limiting parameter of the populous path algorithm contradicts the statement of Tateno and Nei that "if a tree has many branches or links, the chance of overaugmentation is high." On the contrary, as tree topology becomes sufficiently dense, the AD value approaches the TD value and increases linearly with it.

The concern of Tateno and Nei with the sole possibility that some links in the tree will be overaugmented is illogical in our opinion. It should be obvious that any augmentation method (Poisson correction, populous path etc.) which does not consistently underaugment is bound to overaugment some link distances. The only possible approach, so far as we know, which could attempt to find for the tree's links of the same DD different amounts of actual evolutionary change is the REH method developed by Holmquist and his collaborators (Holmquist et al., 1972; Jukes and Holmquist, 1972; Holmquist, 1976). The REH method would give different augmentations depending on the ratio of minimal 2-plus minimal 3 base type amino acid replacements to minimal 1 base type amino acid replacements on comparing two aligned amino acid sequences. Of course, this method too would be bound to make mistakes, overaugmenting on some links and underaugmenting on others.

In computer simulations modeled after a Poisson process we expect in any dense portions of the tree that the unaugmented link distances, using the ancestral sequences

produced by the pseudorandom number generator, will be comparable to those found from the maximum parsimony reconstruction starting with only the contemporary sequences. However, in sparser portions of the tree, we expect that the unaugmented link distances using sequences produced by the pseudorandom number generator will become progressively larger than those from the maximum parsimony reconstruction. This follows because in the sparse region many nucleotide positions in the ancestral sequence will undergo multiple or superimposed nucleotide substitutions. Thus frequently at such a variable nucleotide position the ancestral node may have a nucleotide, say G, whereas its two offspring nodes might share convergently a different nucleotide, say A. In this situation, the maximum parsimony method will reconstruct that shared nucleotide, A, as the common ancestral nucleotide of the two descendant sequences and thus produce a smaller direct distance between the ancestor and each descendant than the DD would be between the real sequences.

In contrast, for a pair of densely connected nodal points, i.e. for two points with many intervening links between them in the dense region of the tree, the DD obtained from using the maximum parsimony reconstructed sequences should be about the same as the DD obtained from using the actual sequences produced in the simulation. We can, therefore, predict that the populous path algorithm applied to maximum parsimony reconstructions will tend to increasingly underaugment as the links to be augmented become larger. Our best chance, then, to reduce such gross underaugmentation is indeed to choose the most populous paths in the maximum parsimony reconstructed tree as the source of the augmentation factors.

Computer Simulations

We followed the essentials of the procedure described by Tateno and Nei (1978) using a cistron containing only the 50 codons subject to random mutation and omitting the 50 invariant codons as irrelevant to the experiment. The 150 nucleotides were assigned mutations distributed randomly; chain terminators (nonsense codons) were excluded. The number of mutations assigned to each link or evolutionary period was set equal to the corresponding "True Distance" ("TD") in either replication 1 or replication 2 of the computer simulation of Tateno and Nei. These "TD" values were graciously provided to us by Dr. Nei. We carried out the simulation several times. Each time the direct distances (DDs) from the sequences produced in the simulation were employed to find augmented distances (ADs) by the populous path algorithm. In addition for each run of the simulation the 21 contemporary or terminal nucleotide sequences were translated into amino acid sequences. These amino acid sequences were then used to construct A-solution mRNA ancestral and descendant sequences by the maximum parsimony method; these maximum parsimony DDs were employed to find the ADs by the populous path alogrithm. If there are n contemporary sequences in the dendrogram, and m_i represents the sum of mutations from the root of a dendrogram to the contemporary sequence i, the A-solution reconstructs mRNA sequences so that

$$\sum_{i=1}^{n} m_i$$

Table 1. Direct distances (DD), augmented distances (AD), and "true distances" ("TD") for the links of the evolutionary tree in Figure 1 of Tateno and Nei (1977). When all ancestral and descendant sequences generated in the computer simulation are employed for the augmentation procedure, DD and AD are designated as sDD and sAD. When the maximum parsimony method is used with the terminal amino acid sequences produced in the computer simulation to reconstruct the ancestral and descendant mRNA sequences and these then employed for the augmentation procedure, DD and AD are designated as mpDD and mpAD; also "TD" is multiplied by the factor 0.75 becuase most synonymous substitutions are not detected in the reconstruction. The "TDs" of replication 1 of Tateno and Nei are used in our computer simulation designated Rep 1 and the "TDs" of replication 2 of Tateno and Nei are used in our computer simulation designated as Rep 2

			REP 1			
Link	sDD	sAD	"TD"	mpDD	mpAD	3/4 "TD"
10–900	0	0	0	1	1	0
17–900	4	4	4	3	3	3.0
900–901	5	5	5	1	1	3.8
23–901	9	10	9	7	7	6.8
901–902	6	6	6	4	4	4.5
41–902	9	10	9	8	8	6.8
902–903	5	5	5	4	4	3.8
1–903	11	12	11	7	7	8.3
903–915	5	5	6	3	3	4.5
49–904	9	10	9	8	8	6.8
56–904	9	10	9	6	6	6.8
904–915	9	10	9	4	4	6.8
915–918	20	25	24	16	18	18.0
74–918	40	61	48	20	22	36.0
918–905	16	21	16	14	16	12.0
78–905	48	75	71	35	42	53.3
905–912	40	61	48	16	18	36.0
217–907	34	46	44	24	26	33.0
140–907	22	27	25	16	18	18.8
907–908	20	25	22	9	9	16.5
146–908	38	59	49	28	33	36.8
908–909	13	15	15	12	12	11.3
149–909	56	95	81	34	41	60.8
909–911	16	21	18	12	12	13.5
153–911	59	106	93	29	34	69.8
911–912	28	38	28	19	21	21.0
912–906	10	11	10	11	11	7.5
209–906	79	187	133	48	62	99.8
906–910	15	20	16	9	9	12.0
158–910	79	187	144	45	59	108.0
910–919	35	56	40	15	17	30.0
2–919	93	286	192	39	50	144.0
919–913	28	38	30	8	8	22.5
218–913	99	292	290	51	65	217.5
913–916	10	11	10	8	8	7.5
312–916	93	286	247	35	42	185.3
916–914	16	21	17	5	5	12.8
183–914	103	296	252	49	63	189.0
268–914	113	310	522	42	56	391.5

Table 1. (cont.)

	REP 2					
Link	sDD	sAD	"TD"	mpDD	mpAD	3/4 "TD"
10–900	5	5	5	5	5	3.8
17–900	3	3	5	3	3	3.8
900–901	6	6	6	5	5	4.5
23–901	8	8	8	5	5	6.0
901–902	5	5	5	3	3	3.8
41–902	15	17	16	11	11	12.0
902–903	6	6	6	10	10	4.5
1–903	15	17	15	10	10	11.3
903–915	4	4	4	3	3	3.0
49–904	14	16	14	9	9	10.5
56–904	17	19	21	15	16	15.8
904–915	5	5	5	4	4	3.8
915–918	22	24	23	9	9	17.3
74–918	32	39	39	26	34	29.3
918–905	16	18	16	12	12	12.0
78–905	45	63	63	24	29	47.3
905–912	31	38	43	21	26	32.3
217–907	31	38	38	26	34	28.5
140–907	32	39	35	16	17	26.3
907–908	14	16	17	12	12	12.8
146–908	43	53	54	24	29	40.5
908–909	27	33	28	8	8	21.0
149–909	53	78	67	31	41	50.3
909–911	15	17	17	6	6	12.8
153–911	62	110	99	41	51	74.3
911–912	16	18	18	12	12	13.5
912–906	6	6	6	9	9	4.5
209–906	83	178	137	44	61	102.8
906–910	17	19	18	13	13	13.5
158–910	85	180	152	31	41	114.0
910–919	27	33	35	13	13	26.3
2–919	76	142	171	44	61	128.3
919–913	40	50	49	8	8	36.8
218–913	98	239	233	46	67	174.8
913–916	15	17	16	10	10	12.0
312–916	100	248	269	48	69	201.8
916–914	17	19	17	7	7	12.8
183–914	108	270	279	54	78	209.3
268–914	108	270	558	42	59	418.5

is at a minimum. Thus a mutation which may appear in a lineage leading either to few or many contemporary species will choose the few-lineage in the A-solution because this decreases the number of times the mutation is counted between the root and each contemporary species.

Table 1 lists from two of our simulations the DD, AD, and "TD" values for the branches or links of the hypothetical evolutionary tree shown by Tateno and Nei in their Figure 1. The two simulations described are quite representative of other runs

of this same experiment. The DD and AD value obtained from using the simulated or stochastically produced nucleotide sequences are labelled sDD and sAD in our Table, whereas the DD and AD values obtained by the maximum parsimony reconstructions are labelled mpDD and mpAD. Inasmuch as 50 of the 576 mutations possible for the 64 codons involve chain terminators (nonsense codons) and inasmuch as 134 of the remaining 526 mutations are to synonymous codons, the "TDs" were multiplied by the factor 0.75 (from 392/526) in order to compare them to the mpDD and mpAD values. Although our maximum parsimony method works with up to 45 codons and in certain situations can detect synonymous substitutions, in practice the frequency of these situations is too low to affect the 3/4 factor used to compare the "TD" values to the maximum parsimony values.

The data recorded in Table 1 demonstrates that the populous path algorithm did not overaugment the maximum parsimony distances but, instead underaugmented them[1]. It may be noted that not only are mpAD values gross underestimates of the evolutionary change on the long links (in particular on 910–158, 919–2, 913–218, 916–312, 914–183, and 914–263) but also of the evolutionary change on certain shorter links, especially on the tree's stem or trunk from which many lineages descend (e.g. links 914–916, 913–919, and 919–910). The underestimation here is not surprising because mRNA sequences reconstructed by our maximum parsimony method were A-solution ones.

In Table 2 the ratios of sDD/"TD" for single links are compared to the corresponding mpDD/.75 "TD" ratios in three regions of the evolutionary tree: the region densest in sequence data (single link "TD"s from 0–21), the intermediate region which becomes increasingly sparser in sequence data (single link "TD"s from 26–192), and the sparsest region (single link "TD"s from 247–522). In the densest region, both sets of ratios approach unity, the direct distances approximating the true distances regardless of whether the DDs were determined from the actual stochastically produced sequences or from the maximum parsimony reconstructed sequences. In the intermediate region, both sets of ratios show that DD tends to become progressively smaller than "TD" as link length increases, but the decrease in ratio is steeper when determined from the maximum parsimony reconstructed sequences, averaging at 0.716 for sDD/"TD" and at 0.563 for mpDD/.75 "TD". This trend is especially evident in the sparsest region of the tree, the ratios averaging at 0.337 for the actual stochastically produced sequences and at only 0.206 for the maximum parsimony reconstructed sequences.

---

[1]Essentially the same findings were obtained with the actual 21 terminal mRNA sequences produced in replication 1 of the computer simulation study of Tateno and Nei. These 21 sequences were provided to Dr. Richard Holmquist by Yoshio Tateno and Professor Masatoshi Nei. Dr. Holmquist then translated these RNA sequences, using the genetic code table, into amino acid sequences. He sent us the 21 amino acid sequences and at his request we constructed the A-solution maximum parsimony ancestors and determined the augmented distance values by the populous path algorithm. We used the tree topology provided by Dr. Holmquist which was the same as Figure 1 in Tateno and Nei (1978). We also found a more parsimonious topology by our branch swapping algorithm; several relatively small changes in topology lowered the nucleotide replacement score from 686 to 679. The augmented distance values found for this topology were just as small as those found for the original topology. In both cases, the AD values grossly underestimated the hidden nucleotide substitutions.

Table 2.

	REP 1		REP 2	
	sDD	mpDD	sDD	mpDD
Link	"TD"	.75 "TD"	"TD"	.75 "TD"
Denser region ("TD"s: 0–21)				
10–900			1	1.316
17–900	1	1	.6	.789
900–901	1	.263	1	1.111
23–901	1	1.029	1	.833
901–902	1	.889	1	.789
41–902	1	1.176	.938	.917
902–903	1	1.053	1	2.222
1–903	1	.843	1	.885
903–915	.833	.667	1	1
49–904	1	1.176	1	.857
56–904	1	.822	.810	.949
904–915	1	.588	1	1.053
AVERAGE	.985	.870	.946	1.060
Intermediate region ("TD"s: 25–192)				
74–918	.833	.556	.821	.887
78–905	.676	.657	.714	.507
905–912	.833	.444	.721	.650
217–907	.733	.727	.816	.912
140–907	.880	.851	.914	.608
146–908	.776	.761	.796	.593
149–909	.691	.559	.791	.616
153–911	.634	.415	.626	.552
209–906	.594	.481	.606	.428
158–910	.549	.417	.559	.272
910–919	.875	.500	.771	.494
2–919	.484	.271	.444	.343
AVERAGE	.717	.553	.715	.572
Sparsest region ("TD"s: 233–558)				
218–913	.341	.234	.421	.263
312–916	.377	.189	.372	.238
183–914	.409	.259	.387	.258
268–914	.194	.107	.194	.100
AVERAGE	.330	.197	.344	.215

Whereas the average of the mpDD/.75 "TD" ratios is a little more than three-fourths of the sDD/"TD" average ratio in the intermediate region, it is less than two-thirds of it in the sparsest region.

These findings from the data in Tables 1 and 2 clearly show that the populous path algorithm, when used as an adjunct to the maximum parsimony method, has no systematic error towards overaugmentation. If anything, the algorithm has a systematic

error towards underaugmentation. Consequently, it would indeed seem that the best chance of reducing steep underaugmentation is to do what the algorithm does, augment on the basis of the most populous or densest paths in the maximum parsimony reconstructed tree.

Despite the underaugmentation observed, the computer simulation has nevertheless revealed a surprising result. The populous path algorithm generally gives better estimates of mutational distance between contemporary amino acid sequences than the Poisson correction[2], when measured with a "chi-squared" criterion, i.e.

$$\sum_{\text{all comparisons}} \frac{(\textit{True Distance-Predicted Distance})^2}{\text{True Distance,}}$$

and when at least one of the two amino acid sequences in each comparison was in the denser region of the genealogical tree, i.e. had various close or moderately distant as well as the more distant relatives. Our finding is surprising, because in this simulation all the conditions of the Poisson model are met. The following chi-squared values were found when all 20 comparisons to amino acid sequence 1 in replication 2 were carried out. For the Poisson correction the sum of the 20 chi-squared distances was 1328.7 and the populous path corrections 1022.5. Other series of pairwise comparisons examined yielded similar values. That the populous path algorithm gives better estimates can be explained because this method in determining the amount of change which has occurred from one sequence to another takes into account all the information provided by *all* sequences being compared. The Poisson estimate only works with the information in the two sequences being compared and thus much genealogical information is lost.

As the populous path algorithm is designed to be used as an adjunct to the maximum parsimony method, we consider it important to emphasize that the one bias of the maximum parsimony method is a tendency to overrepresent the nucleotide indentities between adjacent ancestral and descendant sequences. Tateno and Nei ignore or misunderstand this bias when they state: "However, if the number of species used is large, *and the relative values of DD's (obtained by the maximum parsimony method) to TD's are more or less the same as those in our simulation*, then overaugmentation would occur." In view of the particular nature of the parsimony bias, Tateno and Nei have stated an impossible proposition. We have shown that the ratio of DD (obtained from maximum parsimony reconstructed sequences) to TD must tend to become progressively smaller than the ratio of DD (obtained from the actual simulated sequences) to TD as link length increases. This relationship should hold even if the number of species used in the simulation are large, because the maximum parsimony method would still construct on any long links many more nucleotide identities between adjacent ancestral and descendant nodal sequences than would be expected for real sequences. In fact we have now carried out the computer simulation with two and a half times the density of sequences (52 contemporary sequences rather than 21) but

[2]Following Nei et al. (1976) the Poisson correction used is $-1.4 \ln (1 - P_{aad}) \times 50$. P_{aad} is the fraction of amino acid sites which differ in the two sequences being compared

have otherwise maintained the conditions in the model used by Tateno and Nei. Exactly the same trends as in the smaller simulation manifest themselves: progressive underaugmentation on the longer links and the steeper decrease of mpDD/0.75"TD" ratios than of the sDD/"TD" ratios as link length increases.

We agree with Tateno and Nei that in the real process of protein evolution as compared to a simple Poisson process there is much greater variance of the rate of nucleotide substitution. In fact the genealogical tree constructed for 55 eukaryotic globin sequences by the maximum parsimony method has helped establish this point (Goodman et al., 1975). It has also demonstrated that vertebrate globin sequences evolved at a much faster rate than had previously been thought from calculations based on the Poisson correction method. Inasmuch as the reconstructed globin tree was a relatively sparse one, it is very unlikely that the augmented parsimony distances on the tree were overaugmented distances. Indeed now that many more globin chains have been sequenced and much denser trees can be constructed, we are finding that the unaugmented parsimony distances themselves are approaching and even overtaking the previous augmented parsimony distances. To give just several examples, in the earlier globin tree, eight links separated rhesus monkey alpha sequence from dog alpha sequence, and the augmented parsimony distance between these two sequences was 44 nucleotide replacements. Now in a current genealogy twenty links separate the two sequences, and the unaugmented parsimony distance between them is itself 44 nucleotide replacements. In the earlier globin tree nine links separated rhesus and horse alpha sequences, and the augmented parsimony distance between these two sequences was 35 nucleotide replacements. Now eighteen links separate these two sequences and the unaugmented parsimony distance between them is 51 nucleotide replacements. In the earlier tree, four links separated human and rhesus monkey alpha sequences, and the parsimony distance (augmented same as unaugmented) was 5 nucleotide replacements. Now eleven links separate human and rhesus sequences, and the parsimony distance (augmented same as unaugmented) is 9 nucleotide replacements, almost twice that of the earlier sparser tree.

On the basis of present evidence, we conclude that the populous path procedure for augmenting maximum parsimony distances rests on a sound empirical and theoretical foundation. The phylogenetic reconstructions and corresponding estimates of evolutionary rates should become more accurate rather than less as larger numbers of closely related species are included in the analysis. However, it is true that when augmentations are chosen from a very restricted number of paths in the parsimony tree they become subject to random uncertainties. Thus we are presently testing algorithms which reduce these random uncertainties while retaining the basic concept of the populous path approach.

Acknowledgement. We thank Dr. Masatoshi Nei for providing us with detailed information concerning the computer simulation studies from his laboratory and for his challenging discussion. We also thank Dr. Richard Holmquist for helpful discussions. This work was supported by grant DE72–02429 from the National Science Foundation.

References

Beard, J.M., Goodman, M. (1976). In: Molecular Anthropology, (M. Goodman, R.E. Tashian, eds.), New York: Plenum
Farris, J.S. (1970). Syst. Zool. **19**, 83
Fitch, W.M. (1971). Syst. Zool. **20**, 406
Fitch, W.M. (1972). Humangenetic **16**, 67
Fitch, W.M., Markowitz, E. (1970). Biochem. Genet. **4**, 579
Goodman, M., Moore, G.W. (1977). J. Mol. Evol. **10**, 7
Goodman, M., Moore, G.W., Barnabas, J., Matsuda, G. (1974). J. Mol. Evol. **3**, 1
Goodman, M., Moore, G.W., Matsuda, G. (1975). Nature **253**, 603
Holmquist, R. (1976). In: Molecular Anthropology, (M. Goodman, R.E. Tashian, eds.), New York: Plenum
Holmquist, R., Cantor, C.R., Jukes, T.H. (1972). J. Mol. Biol. **64**, 145
Jukes, T.H., Holmquist, R. (1972). J. Mol. Biol. **64**, 163
Moore, G.W. (1977). J. Theor. Biol. **66**, 95
Moore, G.W., Barnabas, J., Goodman, M. (1973). J. Theor. Biol. **38**, 459
Moore, G.W., Goodman, M., Callahan, C., Holmquist, R., Moise, H. (1976). J. Mol. Biol. **105**, 15
Nei, M., Chakraborty, R. (1976). J. Mol. Evol. **7**, 313
Ohta, T. (1976). J. Mol. Evol. **8**, 1
Tateno, Y., Nei, M. (1978). J. Mol. Evol.
Vogel, F., Kopun, M. (1977). J. Mol. Evol. **9**, 159

Received October 13, 1977; Revised November 28, 1977