

CLADISTIC ANALYSIS OF RESTRICTION ENDONUCLEASE CLEAVAGE MAPS WITHIN A MAXIMUM-LIKELIHOOD FRAMEWORK

RONALD W. DEBRY¹ AND NORMAN A. SLADE

Museum of Natural History, University of Kansas, Lawrence, Kansas 66045

Abstract.—Cladistic analysis of restriction endonuclease cleavage map data, particularly from animal mitochondrial DNA, is considered within a maximum-likelihood framework. The required probability model of evolutionary change is developed from current empirical evidence, and is specific to mtDNA. A large asymmetry is shown to exist between the probabilities of gaining a new site and losing an existing site, with the loss of a site being the much more common event. Since Wagner parsimony assumes that the probabilities of forward and reverse changes are equal, it is a very inefficient estimator of relationships for these data. Dollo parsimony, on the other hand, assumes that the probability of a loss is much greater than that of a gain, an assumption of Dollo parsimony which fits restriction map data well. Dollo parsimony is shown to be a consistent and efficient estimator of phylogeny for mtDNA restriction map data. [Mitochondrial DNA; restriction endonucleases; phylogenetic inference; parsimony; statistical inference; maximum-likelihood estimation.]

In the past several years, an increasing number of theoretical studies have appeared on the phylogenetic analysis of restriction enzyme cleavage maps of mitochondrial (mt) DNA. Most of these have focused on the flux of restriction sites in the mitochondrial genome as a whole (Gotoh et al., 1979; Kaplan and Langley, 1979; Nei and Li, 1979; Upholt and Dawid, 1979; Aoki et al., 1981; Kaplan and Risko, 1981). These studies have generally proposed measures of genetic distance between two or more species, which can be used in much the same way as electrophoretic distances to deduce phylogenies.

In a recent paper, Ferris et al. (1981b) took a different approach and examined mtDNA cleavage sites as individual characters. They used the parsimony method of Dayhoff and Eck (1968) and Farris (1970) to construct a phylogenetic hypothesis for anthropoid primates based on their mtDNA cleavage maps. Avise et al. (1983) have also used Wagner parsimony to analyze restriction-fragment data from *Peromyscus*, and Templeton (1983a) has recently developed a method for analyzing

restriction map data by a combination of Wagner parsimony and character compatibility-like methods.

In this paper, we examine the properties of maximum-parsimony analysis as it applies to mtDNA cleavage maps, using a maximum-likelihood framework similar to that of Felsenstein (1978). While we believe that a direct maximum-likelihood method of phylogenetic analysis would be desirable for cleavage map data, we admit that such an approach would involve immense computational difficulties (see also Felsenstein, 1973). In place of a direct method, we will examine existing parsimony-based tree-building criteria to see under what conditions they possess the same desirable statistical properties (i.e., consistency and efficiency) that maximum-likelihood methods often possess. To do this, we will examine in detail the behavior of a single restriction site through the evolution of a species group. We are particularly interested in the probability that the site will evolve in a pattern which would result in the correct phylogenetic hypothesis according to the tree-building algorithm under consideration. This requires a probabilistic evolutionary model, which we develop specifically for mtDNA.

¹ Present address: Department of Zoology, Michigan State University, East Lansing, Michigan 48824.

THE NATURE OF RESTRICTION SITES

Restriction endonucleases, when used to digest a purified sample of DNA, recognize and cleave a specific sequence of base pairs. For example, the enzyme *EcoRI* will recognize and cleave the sequence GAATTC every time it occurs in a sample of DNA. Although restriction enzymes are available which cleave sequences of four, five, or six base pairs, we are only considering in this study those cleaving six-base sequences (although the model could be adapted to the other forms). We will also limit this discussion to only those enzymes with nonambiguous recognition sequences.

If we define a sequence of six base pairs as a "word," then any given word either is or is not a recognition word. Those sites that are recognition words we designate "+," those that are not we designate "0." There is a subclass of 0, which contains those words one substitution away from being a recognition sequence, and which we designate 0' (these are the "one-off sites" of Templeton, 1983a, b). Since there are six bases, with three possible alternate states each, there are 18 different 0' sequences for each enzyme. If the DNA is a random collection of nucleotides, then for every + present in a species there will on average be 18 0' words for the same recognition sequence (Templeton, 1983a).

We know the mechanics of change in cleavage-site characters (although we may not know the cause of the change). At a site which is currently a recognition word, any single substitution will cause the site to be lost, changing it from a + to a 0'. At any site currently in a particular 0' state, only one of the 18 possible changes will cause the site to become a +. Fifteen of the 17 other possible changes will take the word to a 0 state which is at least two steps away from a +. The remaining two changes would take the site to a different 0' state. For example, one 0' state for *EcoRI* is AAATTC, lacking only a G in the first position. Any substitution in the last five bases would take the site to a state at least two changes from a +. However, the first

base could change to either a T or a C and still be a 0' sequence.

AN EVOLUTIONARY MODEL FOR RESTRICTION SITES

It is possible to use currently available information about the evolution of mtDNA to construct a plausible parametric model of restriction site evolution. Such a model is necessary for the use of maximum-likelihood methods in phylogenetic analysis. Previous models for other types of data have been greatly simplified, and have often assumed that changes are irreversible (e.g., Felsenstein, 1973), or have purposely attempted to make "the evolutionary assumption as simple and weak as possible so as to obtain an inference procedure applicable to a variety of types of comparative data" (Farris, 1973:252). The strength of the model presented here lies in the fact that it is specifically derived for mtDNA restriction-site maps. Thus, generality has been sacrificed in the interest of increased power and efficiency in analyses of this type of data.

Assumptions.—While a number of assumptions are necessary to make this model tractable, we have attempted to stay as consistent with actual observations on the nature of mtDNA as possible. We have limited this model to animal mtDNA for two reasons. First, mtDNA is much more commonly used for systematic studies than is nuclear DNA, primarily due to the convenient size of mtDNA (about 16 kilobases). Second, specific empirical observations indicate that there may be differences between the evolutionary behavior of mtDNA and nuclear DNA. This should be taken into account if an attempt is made to use the information presented here in the analysis of nuclear or chloroplast DNA.

The first assumption we have made is that the rate of substitution is constant across lineages. This is almost certainly not true; however, the extent to which rates vary—both across taxa and across regions of the mtDNA within a species—is not known. Certain genes, notably the tRNAs, appear to be quite highly conserved in the higher primates relative to the rest of the

mtDNA (Brown et al., 1982). Specifically, when we make numerical calculations, we assume that the average substitution rate is equal to about 8.5×10^{-8} substitutions per six nucleotide words per year. This is based on estimates obtained by Gotoh et al. (1979) and Brown et al. (1979). This is only an approximate average and should not be taken as exact. However, the substitution rate of purine-purine and pyrimidine-pyrimidine transitions is approximately nine times that of purine-pyrimidine transversions (Brown et al., 1982). Brown et al. (1982:fig. 5) found that the position of a nucleotide pair with respect to its codon affects its probability of substitution. In particular, approximately one-half of the substitutions observed were in the third position, while about one-third were at the first and one-sixth were at the second codon position. Both of these observations are incorporated into the model where appropriate.

A second assumption is that the ancestral populations were not polymorphic at the time of speciation. Since intraspecies variation has been observed (Avisé et al., 1979; Ferris et al., 1981a), this must occur to some degree. Since eventual fixation of these variants should be random with respect to the descendant species (in the absence of selection), this represents a source of noise. Gillespie and Langley (1979) and Tajima (1983) have considered the effects of retained polymorphism in non-recombining DNA (such as mtDNA) on phylogenetic reconstruction. It is possible that two mtDNAs share a common ancestor which predates the common ancestor of the two species. If only one mtDNA sample is available per species, the best that can be said is that one is inferring the gene phylogeny of the mtDNA, which may not completely agree with the species phylogeny. Felsenstein (1979) discussed the general effects of high frequencies of retained polymorphism on maximum-parsimony estimates of phylogeny.

We are only considering those potential sites at which a + state is actually observed in one or more species; therefore, we will also assume that the character state

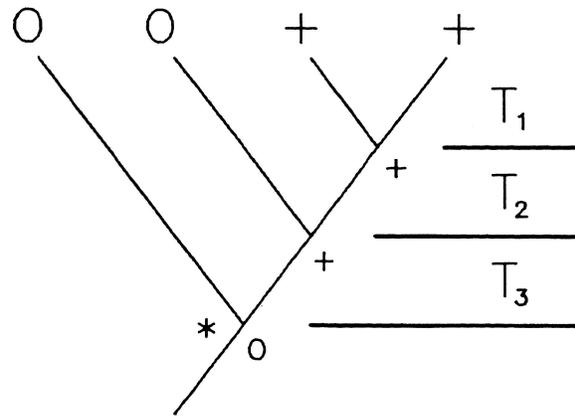


FIG. 1. A possible evolutionary tree for four species with the character states at the interior nodes specified. The speciation events divide the tree into the three time intervals T_1 , T_2 , and T_3 . Node 1 is denoted by an asterisk (*). Species from left to right are D (outgroup), A, B, and C.

at the base of any tree is either 0' or +. This is justified by the fact that the probability of becoming a + from a 0 state can be no greater than approximately the probability of two substitutions times 18^{-2} .

We also assume that sites mapped to the same location in different species are indeed at the same location. The accuracy of the maps will vary substantially, being best when side by side comparisons are made of double digests, and being by far the worst when data published from different sources are being compared. Finally, we have assumed that the base composition of the DNA is 1:1:1:1 (A:T:C:G).

The model.—The model, which is fully derived in Appendix A, provides the probability of each possible type of transformation in a segment of length T years. We may use these probabilities to calculate the overall probability (given the particular evolutionary model) of a tree such as the one in Figure 1. The interior nodes of the tree, which in this case represent speciation events, divide the tree into three time intervals (T_1 , T_2 , and T_3). In our model, p represents the substitution rate per word per year. Since p is presumed to be constant over time, each of these intervals can be assigned a λ value, where $\lambda_i = T_i p$. Using this notation, the λ for the total time interval represented by species D would be given as $\lambda_1 + \lambda_2 + \lambda_3$.

The transformations we are interested in are: $0' \rightarrow 0'$, $0' \rightarrow +$, $0' \rightarrow 0$, $+ \rightarrow 0'$, $+ \rightarrow +$, and $+ \rightarrow 0$. In this case, the probability of $0' \rightarrow 0$ includes both the $0'$ state remaining the same and changing to any state other than a $+$. In the evolutionary scenarios described below, this represents a case where a segment begins with a $0'$ and no subsequent (descendant) segment requires the gain of a $+$ state. The same holds for the probability of a $+ \rightarrow 0$ transformation. The probabilities for the six transformations of interest are:

$$P_{0' \rightarrow 0'}(\lambda) = e^{-\lambda} + \lambda e^{-\lambda}(\frac{2}{18}) + \frac{1}{2}\lambda^2 e^{-\lambda}(0.251), \quad (1)$$

$$P_{0' \rightarrow +}(\lambda) = \lambda e^{-\lambda}(\frac{1}{18}) + \frac{1}{2}\lambda^2 e^{-\lambda}(0.0074), \quad (2)$$

$$P_{0' \rightarrow 0}(\lambda) = 1 - P_{0' \rightarrow +}(\lambda), \quad (3)$$

$$P_{+ \rightarrow 0'}(\lambda) = \lambda e^{-\lambda} + \frac{1}{2}\lambda^2 e^{-\lambda}(0.037), \quad (4)$$

$$P_{+ \rightarrow +}(\lambda) = e^{-\lambda} + \frac{1}{2}\lambda^2 e^{-\lambda}(0.163), \quad (5)$$

$$P_{+ \rightarrow 0}(\lambda) = 1 - P_{+ \rightarrow +}(\lambda). \quad (6)$$

The remaining possible transformations ($0 \rightarrow +$, $0 \rightarrow 0'$, and $0 \rightarrow 0$) could be calculated in the same fashion; however, the probability of a $0 \rightarrow +$ event is approximately six orders of magnitude lower than that of a $0' \rightarrow +$ event. Therefore, any transformations beginning with a 0 state may be safely ignored.

Obtaining the probability of a scenario.—There are six segments in the tree in Figure 1, each bounded either by two interior nodes or by one interior node and one endpoint (taxon). An evolutionary scenario for such a tree consists of the observed character states at the terminal nodes as well as hypothetical character states at the interior nodes. For any particular scenario, each of the six segments has the state of the restriction site unambiguously defined at both endpoints. Starting with either a $+$ or a $0'$ at the basal node, we can derive a probability expression for each scenario. This probability is derived as by Felsenstein (1973). If the tree is taken to be a set of six independent segments, each with a transformation, then all the transforma-

tions must be true for the scenario to be true. There must actually be a $0'$ state present at the base of the tree ($P = \frac{18}{4,096}$, where 18 is the number of possible $0'$ sites and 4,096 is the total number of possible six-base words). In the segment between nodes 1 and 2, there must be a $0' \rightarrow +$ transformation. Between 2 and 3, and twice between 3 and endpoints B and C, there must be a $+ \rightarrow +$. From node 1 to endpoint D requires a $0' \rightarrow 0$ (i.e., anything except $0' \rightarrow +$), and from node 2 to endpoint A requires a $+ \rightarrow 0$ transformation. Multiplying these independent probabilities gives

$$P_{G_7} = P(0') \cdot P_{0' \rightarrow +}(\lambda_3) \cdot P_{+ \rightarrow +}(\lambda_2) \cdot [P_{+ \rightarrow +}(\lambda_1)]^2 \cdot P_{0' \rightarrow 0}(\lambda_1 + \lambda_2 + \lambda_3) \cdot P_{+ \rightarrow 0}(\lambda_1 + \lambda_2). \quad (7)$$

In this way the probability expression may be obtained for each possible scenario.

WAGNER PARSIMONY

Under the Wagner parsimony criterion (Farris, 1970), in a study of three species and one outgroup, there are only six character distributions which contain phylogenetically useful information (Table 1). In three of these, the presumed primitive state for the group is presence of a cleavage site ($+$). Two members of the ingroup share the derived state—absence of the site (0). The other three are the reverse, where the ancestral condition is presumed to be 0 and two members of the ingroup share a newly acquired $+$. Using the parsimony method as described by Farris (1970), the tree chosen is the one that requires the fewest combined reversals and independent origins. Note that this amounts to an implicit assumption that reversals and independent origins are equally likely.

Under a strict Wagner parsimony criterion, each of the character distributions in Table 1 is matched to only one phylogeny. It is possible, of course, since evolution is not strictly parsimonious, for any or all of these character distributions to appear in a single lineage and, conversely, for any particular character distribution to be the result of any of the three possible phylogenies (Fig. 2). In this study we choose to

TABLE 1. Character distributions for three species which are "phylogenetically informative" according to the Wagner parsimony criterion. Species D is the outgroup.

| D | A | B | C |
|---|---|---|---|
| 0 | 0 | + | + |
| 0 | + | 0 | + |
| 0 | + | + | 0 |
| + | + | 0 | 0 |
| + | 0 | + | 0 |
| + | 0 | 0 | + |

hold the phylogeny constant and allow the distribution of restriction sites among species to vary. For the rest of this discussion we will assume that Figure 2a represents the true phylogenetic history of the group. Our goal is to discover, under various conditions, the approximate probability that a single restriction-site character will agree with the true phylogeny.

All possible scenarios.—On a tree with seven locations (three interior nodes and four endpoints, such as Fig. 1), a total of 128 different placements of +s and 0s are possible. For the moment we are concerned only with character distributions considered informative under Wagner parsimony and, thus, will ignore those scenarios in which A, B, and C are all + or all 0. This leaves 96 scenarios (2^4 possibilities for D, 1, 2, 3 times six possibilities for A, B, and C, without +++ or 000). In half of these, however, the state of the outgroup, D, is the same as the state shared by two members of the ingroup. These are symplesiomorphies and, thus, not considered under Wagner parsimony. This leaves 48 scenarios with shared-derived characters (synapomorphies), 24 of which show a derived loss of the restriction site and 24 which show derived gain of the site (see Appendix B). However, distributions of the form

| D | A | B | C |
|---|---|---|---|
| 0 | + | 0 | + |
| 0 | + | + | 0 |

are represented by simple rotation of species B and C, and so will have identical

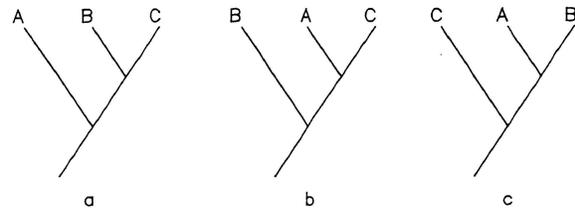


FIG. 2. The possible evolutionary trees for three extant species.

probabilities. Thus, a total of 32 probability expressions need actually be calculated, corresponding to 16 shared-gain and 16 shared-loss scenarios.

Consistency. We are not necessarily interested in determining exactly which of the 24 scenarios has the highest probability for any given restriction site. Rather, we wish to know whether the class of scenarios which indicate the correct phylogeny has a higher probability than either of the two classes which indicate incorrect phylogenies. In other words, we want to know if the phylogenetic methods employed to analyze this type of data will converge on the true phylogeny as more data are collected and, therefore, be statistically consistent (Felsenstein, 1978).

If we can show that an algorithm is consistent for all tree lengths that might be encountered in practice, then variation in substitution rate between characters will not pose a problem, because the time and substitution rate always enter the model in the form "pt". If one character has a low rate of substitution, it will be the equivalent of a shorter time and will still be consistent. The converse also holds; if one character has a high rate of substitution, it will be the equivalent of a longer time, but the character will still be consistent. This does not, however, allow violations of the between taxa constant rate assumption, or variation in rate between characters when the algorithm is consistent for some, but not all, tree lengths.

Efficiency.—It is also important that a phylogenetic algorithm be efficient. Efficiency of an algorithm is related to the probabilities associated with each of the possible outcomes (presuming that the

TABLE 2. Probabilities calculated for selected scenarios shown in Appendix B. Each interior stem is of the length indicated. Each probability has been multiplied by 10^5 to obtain the values listed. See text for an explanation of P_c .

| Scenario | Stem lengths (in millions of years) | | | | | |
|----------|-------------------------------------|------|------|------|------|------|
| | 1 | 2 | 4 | 6 | 8 | 10 |
| G3 | 1.57 | 2.18 | 2.15 | 1.63 | 1.12 | 0.73 |
| G7 | 0.26 | 0.66 | 1.07 | 1.02 | 0.79 | 0.56 |
| G8 | 0.13 | 0.31 | 0.46 | 0.40 | 0.29 | 0.12 |
| G15 | 0.68 | 1.48 | 1.85 | 1.41 | 0.90 | 0.54 |
| G16 | 0.33 | 0.68 | 0.79 | 0.56 | 0.33 | 0.18 |
| P_c | 0.73 | 0.68 | 0.67 | 0.68 | 0.69 | 0.71 |
| L1 | 0.10 | 0.20 | 0.19 | 0.12 | 0.06 | 0.03 |
| L2 | 0.21 | 0.42 | 0.45 | 0.29 | 0.16 | 0.08 |
| L3 | 1.24 | 1.36 | 0.88 | 0.46 | 0.23 | 0.11 |
| P_c | 0.76 | 0.65 | 0.54 | 0.50 | 0.47 | 0.46 |

method used is consistent and, therefore, the most likely outcome is the correct phylogeny). As an example, consider a situation where, with one algorithm, the probability of obtaining a single restriction site indicating the correct phylogeny is 0.4, while the incorrect phylogenies will each be indicated with a frequency of 0.3. This algorithm will be consistent; as more characters are collected, the data will indeed converge on the true phylogeny. However, with 10 restriction site characters the probability that five or more "correct" characters will be observed is only approximately 0.37. On the other hand, consider a different algorithm which will obtain the correct phylogeny with a probability of 0.6 and each of the incorrect phylogenies with probability 0.2. In this case, with 10 characters, the probability of observing five or more "correct" characters is approximately 0.83. This means that at least half of the observed sites would indicate the correct phylogeny about 83% of the time. The second method is clearly more efficient and far fewer characters need be examined in order to have a 95% chance of obtaining the correct topology.

Results.—The variables in the probability expressions are the substitution rate and time. We are assuming that the substitution rates are constant, leaving the time

between speciation events as the remaining variable. It is, of course, not possible to know the times of speciation in a real group. We will instead investigate the properties of cleavage-map sites over a wide range of speciation times, and attempt to establish the approximate boundaries of statistical consistency and/or useful efficiency (Felsenstein, 1979) for this type of data.

We used a computer program that manipulates the speciation times and calculates the relative probability for each of the 32 scenarios. Sample results from the calculations are shown in Table 2. Only those scenarios that at some point contribute at least 5% of the total probability are listed. The cutoff means that less than 5% of the "phylogenetically informative" characters would be expected to evolve in conformance with any particular scenario which is not listed.

Comparison between Table 2 and the scenarios diagrammed in Appendix B reveals an important pattern. Among those site-gain synapomorphy scenarios with high probabilities, none contain either independent gains or a loss followed by regaining the site. Furthermore, all of the scenarios that do not contribute over 5% to the probability *do* require at least one independent gain or loss-regain event. The same is true for site loss synapomorphies, where scenarios with multiple losses are much more likely than scenarios with multiple gains. All shared-loss scenarios except L1, L2, and L3 have at least one independent gain term, and only the first three scenarios contribute significantly to the total probability. This is consistent with the findings of Templeton (1983a, b) and inconsistent with the assumption of Wagner parsimony—that forward and reverse changes are equally likely.

While asymmetry between probabilities of gains and losses will have a significant impact on the analysis of mtDNA restriction map data, a Wagner analysis of one character in three taxa will always identify only a single gain or a single loss for that

character. Thus, we may still consider whether or not such an analysis is appropriate for the three species case.

Those scenarios that contribute more than 5% of the total probability can generally be grouped into pairs. For site-gain synapomorphies these pairs are scenarios *G7* and *G8*, and *G15* and *G16*. Within each pair the states at the interior nodes are the same (e.g., 0++ in *G7* and *G8*), with the odd-numbered scenario representing the true phylogeny and the even-numbered scenario representing the two "incorrect" phylogenies in each case. The counterpart to *G3* (i.e., *G4*) does not, however, contribute significantly to the total probability, since *G4* requires two gains and one loss.

If we can show that the probability of the correct phylogeny is greater than one-third for each pair, then Wagner parsimony analysis of site-gain synapomorphies will be a consistent estimator of the phylogeny, regardless of the relative frequencies between pairs. This can be accomplished algebraically, and we will take scenarios *G7* and *G8* as an example. The two scenarios are identical except that in *G7* the site is lost in species A and is maintained in species B, while in *G8* the site is maintained in A but lost in B. If the probability of losing the site in species A is represented as L_A and the probability of losing the site in species B as L_B , then the criterion for consistency is

$$L_A(1 - L_B) > (1 - L_A)L_B, \quad (8a)$$

$$L_A - L_AL_B > L_B - L_AL_B, \quad (8b)$$

$$L_A > L_B, \quad (8c)$$

which will be true for any nontrichotomous tree where the assumption of constant rates between taxa is met. The argument is identical for the comparison between scenarios *G15* and *G16*. Since scenario *G3* is always at least an order of magnitude more likely than is scenario *G4*, Wagner parsimony analysis of site-gain synapomorphies must always be a consistent estimator of the phylogeny in the three species case, regardless of the length

of time or substitution rate at any one character, so long as the rate is the same for all taxa at that character. These arguments are supported by the results of the computer calculations listed in Table 2.

For shared-site losses, the situation is somewhat different. As mentioned above, multiple losses are far more likely than multiple origins. Examination of Appendix B shows that scenarios *L1*, *L2*, and *L3* are the only shared-loss scenarios without multiple site origins. In this case the criterion for consistency is that $P_{L1} + P_{L3} > P_{L2}$, which is nontrivial, since an argument analogous to that for scenarios *G7* and *G8* shows that the probability of *L2* will always be greater than that of *L1*. This is because both *L1* and *L2* have two independent losses; in *L1* they are both in the BC group, while in *L2* one loss is in the BC group and the other is in species A. Again, the loss is more likely to occur in the longer time interval (species A) for any case except a trichotomy. This means that the incorrect interpretation is more likely than the correct one. Thus, any time that $P_{L2} - P_{L1} > P_{L3}$ Wagner parsimony analysis of shared-site losses will be inconsistent, even with the assumption of constant rates being met.

Even when it is consistent, a Wagner analysis of shared-site losses may suffer from a substantial loss of efficiency relative to site-gain characters (Table 2). A measure which is related to efficiency, the probability that one restriction-site character will give the correct topology when analyzed by Wagner parsimony (P_c), is shown in Table 2. P_c is calculated by summing the probabilities for all the scenarios with the correct topology and dividing this by the sum of the probabilities of all scenarios with a "phylogenetically informative" character distribution. The efficiency of site-gain synapomorphies is relatively high under all conditions, while the efficiency of site-loss synapomorphies decreases as the tree gets longer. The efficiency of both types of characters is lower when the common ancestor of B and C is shortened relative to the other segments,

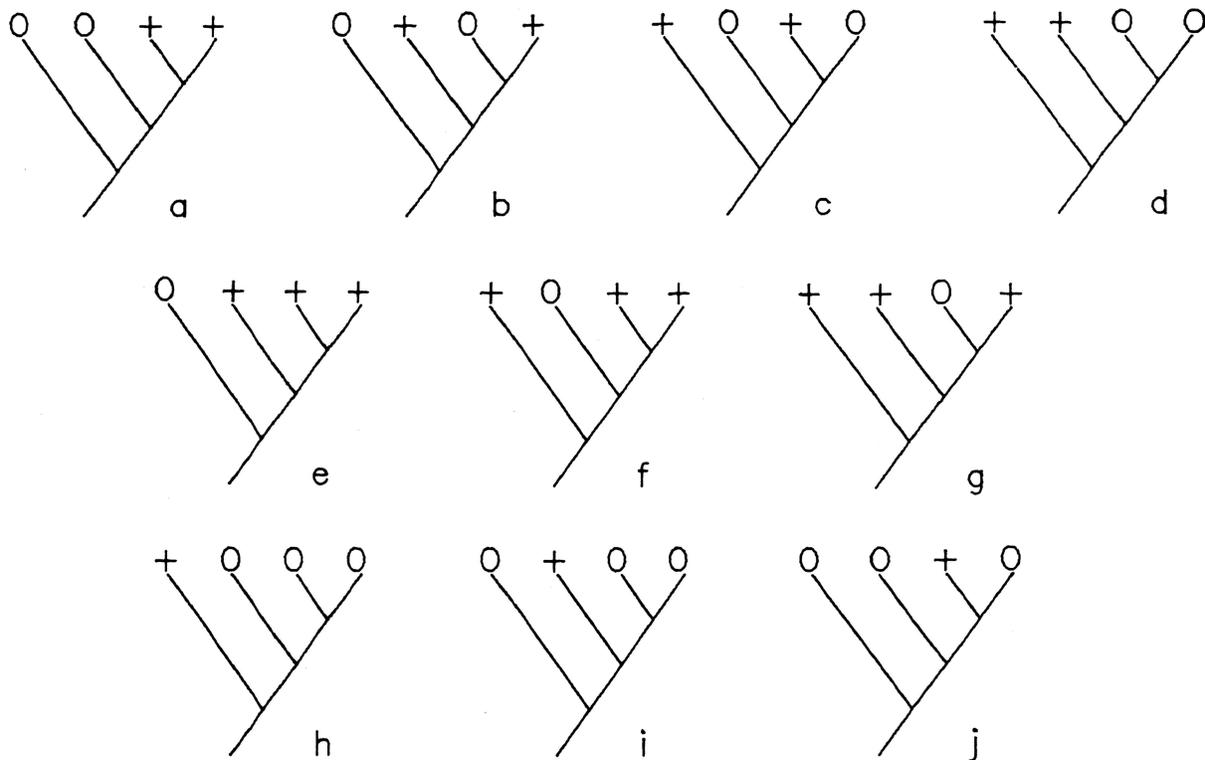


FIG. 3. Possible asymmetrical evolutionary relationships between four species. (a-d) Trees with the restriction site present in two species (+) and absent in the other two (0). (e-g) Trees with the three + states and one 0 state. (h-j) Trees with one + state and three 0 states.

reflecting the decrease in the time available for a gain (or loss) in the B-C ancestor.

Cladistics without outgroups.—The purpose of an outgroup is to provide an estimate of the character state at the base of a tree. An interesting property of restriction-site characters is that grouping two +s together as sister taxa when the third species has a 0 is statistically consistent, whether the state at the base is 0' or +. If the primitive state is +, then the situation is identical to scenarios G7/G8 and G15/G16, which we have already shown to be consistent for all stem lengths. The greater length of A compared to B or C will make a loss in A the more likely event, leaving B and C with +s. However, if 0' is primitive, it is equivalent to the pairs G3/G4 and G1/G2. The single gain in G3 will always be much more likely than the two gains and one loss in G4. Although G2 will sometimes be more likely than G1, the total probability of G1 + G2 (which each re-

quire two 0' → + events) will be insignificant compared to the one gain-multiple loss scenarios.

DOLLO PARSIMONY

Although we have demonstrated that Wagner parsimony will always be a consistent estimator of the phylogeny in the three taxa case when only site-gain synapomorphies are used, we will show this not to be true when the analysis is extended to four or more species. Rather, the asymmetry between the probabilities of gaining a site and losing a site suggests that Dollo parsimony (LeQuesne, 1974; Farris, 1977a, b) is the best candidate for a statistically well-behaved parsimony method (Felsenstein, 1983). The Dollo method assumes that each character arose only once, and selects as the estimate of the phylogeny that tree which minimizes the number of subsequent losses. This will

not always be the same as a maximum-likelihood estimate, however, since the model predicts that a certain large number of independent losses are as equally unlikely as two independent gains.

We will examine Dollo parsimony quantitatively for restriction-site maps by considering the case with four species. The methods for generating all the possible scenarios are the same as for three species, except that there are more "informative" character distributions and, therefore, more possible scenarios. The asymmetrical tree topologies associated with each distribution are shown in Figure 3.

There are three types of topologies to consider, those with two +s and two 0s; those with one + and three 0s; and those with three +s and one 0. The first case is identical to our treatment of Wagner parsimony above, except that the fourth species is not constrained to be the outgroup. The four topologies are shown in Figure 3a-d. Table 3 gives the probabilities (summed over all the interior node assignments) for each of these topologies for several stem lengths. Although trees 3a and 3d would be equally favored under Wagner parsimony (the choice would depend on the state found in the outgroup), it is clear that tree 3a is actually the most probable outcome. This relationship holds for all of the conditions for which we performed calculations. As in the case of three species, this conclusion is independent of the character state at the base of the tree and, therefore, does not change if a fifth species is added as an outgroup, even if the outgroup has a 0.

A similar situation is found for character distributions with three +s and one 0. Dollo parsimony scores the + as a synapomorphy of the three species sharing the character (Fig. 3e), although topologies 3f and 3g are also possible outcomes. The calculations summarized in Table 3 show that topology 3e is more probable than either 3f or 3g for any length of tree and, therefore, is both a consistent and efficient estimate of the phylogeny. This is true whether the state at the base is a + or a 0.

TABLE 3. Probabilities calculated for the tree topologies in Figure 3 (summed over all possible interior node assignments). Each interior stem is of the length indicated. Each probability has been multiplied by 10^5 to give the values shown. See text for an explanation of P_c .

| Topology | Stem lengths (in millions of years) | | | | | |
|----------|-------------------------------------|-------|-------|-------|------|------|
| | 1 | 2 | 4 | 6 | 8 | 10 |
| A | 2.53 | 4.34 | 5.12 | 4.11 | 2.86 | 1.87 |
| B | 0.47 | 1.03 | 1.32 | 1.03 | 0.67 | 0.41 |
| C | 0.23 | 0.48 | 0.55 | 0.39 | 0.23 | 0.12 |
| D | 1.39 | 1.69 | 1.26 | 0.73 | 0.39 | 0.19 |
| P_c | 0.47 | 0.48 | 0.50 | 0.54 | 0.57 | 0.60 |
| E | 4.66 | 4.87 | 2.84 | 1.35 | 0.60 | 0.27 |
| F | 2.18 | 2.13 | 1.10 | 0.47 | 0.19 | 0.07 |
| G | 1.04 | 0.96 | 0.45 | 0.17 | 0.06 | 0.02 |
| P_c | 0.52 | 0.54 | 0.59 | 0.62 | 0.66 | 0.69 |
| H | 6.98 | 10.67 | 12.57 | 11.18 | 8.86 | 6.59 |
| I | 4.02 | 6.82 | 9.20 | 8.92 | 7.55 | 5.94 |
| J | 1.98 | 3.53 | 5.39 | 5.78 | 5.26 | 4.37 |
| P_c | 0.47 | 0.43 | 0.39 | 0.35 | 0.33 | 0.31 |

When the four species show one + and three 0s (Fig. 3h-j), Dollo parsimony normally simply scores an autapomorphy for the species with the + (Fig. 3h). The results in Table 3, however, show that this may not always be the most informative interpretation for restriction map data. In fact, three 0s can be used as a synapomorphy, but only when the tree is so short that a single loss in the common ancestor of the three is much more likely than the combined probabilities of the double loss or the gain-loss required for any other topology. Also, in a short enough tree, if the + is truly an autapomorphy then it is most likely to have arisen in the longest branch. This knowledge will only be useful if the group is known to fall within the limits of consistency indicated by the results in Table 3. Otherwise, it is best to simply treat such characters as autapomorphies.

Overall, the Dollo parsimony method appears to be a consistent estimator of phylogeny for four taxa for all stem lengths. The P_c values listed in Table 3 can also be interpreted as an indication of the efficiency of the algorithm. In both char-

acter distributions normally considered informative under Dollo parsimony (two +s and two 0s; three +s and one 0) the efficiency of the algorithm increases with an increase in the length of the tree.

TEMPLETON'S ALGORITHM

Templeton (1983a) has advocated analyzing restriction map data through a combination of Wagner parsimony and character compatibility methods. In his method, all the sites cut by one particular endonuclease are pooled and a Wagner parsimony algorithm is used separately for each enzyme. Sites are treated within enzyme groups, because Adams and Rothman (1982) have shown that cleavage-site locations are nonrandomly distributed within enzymes, and that there is considerable heterogeneity among different endonucleases. Character compatibility methods are then used between the enzyme-specific trees. Under conditions where some characters are evolving at a high rate while others are evolving much slower, it has been suggested that compatibility methods are more likely to be consistent than are maximum-parsimony methods (Felsenstein, 1981). Apparently, this choice of method relies on the assumption that the nonrandom distribution of restriction sites reflects differences in evolutionary rate between enzyme recognition sequences. In effect, those characters that are most compatible are assumed to have the lowest rate of change, while those showing homoplasy relative to the consensus tree are assumed to do so because of a higher rate of evolution.

We see two objections to the Templeton algorithm. First, as shown above, the high probability of convergent loss means that Dollo parsimony is a much better choice for the analysis of restriction map data than is Wagner parsimony. The second concerns both the use of compatibility methods and the need to treat all sites recognized by one enzyme as a single character. Despite the fact that these sites are nonrandomly distributed, there is no reason to believe that all *EcoRI* (GAATTC) sequences are evolving at the same rate,

while all *HindIII* (AAGCTT) sequences are evolving at another, lower, rate. On average, the GAATTC sequences may well show a higher rate of change, but there will be variation in rate between the different sites. Character compatibility would in this case assign all *EcoRI* sites a lower weight than all *HindIII* sites, despite the fact that some individual *EcoRI* sites will actually have a lower rate of change and, therefore, a lower probability of homoplasy than some individual *HindIII* sites. We believe instead that all restriction sites are best treated as individual characters throughout the analysis.

CONCLUSIONS

The major assumption of Dollo parsimony, that any particular character is rarely gained but often lost, is fit rather closely by mtDNA restriction site characters. On the other hand, these data do not satisfy the assumptions of a method which counts forward and reverse changes equally, such as Wagner parsimony. Our analysis confirms that restriction map data are more effectively analyzed by Dollo parsimony.

Dollo parsimony does not require outgroup rooting; hence, certain results are predicted to be statistically consistent regardless of the state at the base of the tree. For example, when four species show the states 00++, it should be consistent to group as the most apomorphic sister species the two with the character present, whether the base of the tree had a 0 or a +. Our results confirm that this is true for restriction-site characters, at least so long as the assumption of constant evolutionary rate between taxa at one character is met. With that assumption satisfied, Dollo parsimony is a consistent estimator of the phylogeny despite any between-character variation in evolutionary rate. As a result, one need not worry about distinguishing evolutionary "hot spots" from highly conserved regions of the DNA. However, it is reasonable to expect that, as the substitution rate increases (or, conversely, the time of divergence increases), a Dollo parsimony analysis will become less and less robust to rate variation between taxa, as

suggested for Wagner parsimony by Felsenstein (1978).

If, in a group with the character states 00++, one adds an outgroup and attempts to use Wagner parsimony, that method will be consistent only when the outgroup happens to have a 0. If the outgroup has a +, then Wagner parsimony will be statistically inconsistent. In other words, as more and more such characters are collected, the data will converge on the wrong phylogeny. Thus, use of Wagner parsimony will result in a large number of incorrectly interpreted characters and, therefore, is a very inefficient method of analysis. Dollo parsimony is clearly preferred over Wagner parsimony for the analysis of restriction map data. Although the Dollo tree will not always be exactly the same as the results of a maximum-likelihood analysis, the substantial computational advantages of Dollo algorithms make this method very desirable for restriction map data.

These conclusions were reached under a particular parametric evolutionary model. The assumptions underlying this model may or may not be valid, although we believe that they are justified by current empirical evidence. However, the preference for Dollo parsimony is not the result of a particular choice for the probability of, say, an A to G substitution event at a certain nucleotide position, although those substitution probabilities are the only direct products of the evolutionary model. Rather, the asymmetry expected between the probabilities of losing a site and gaining a site is a result of the purely mechanical properties of restriction word characters. Therefore, we expect that Dollo parsimony would be robust to the selection of other (non-Poisson) sampling models.

Even if the empirical observations about mtDNA prove not to be generally applicable to nuclear or plant DNA, the differences would likely be minor compared with the effects caused by using restriction words as characters. Thus, it is also expected that Dollo parsimony will be the

method of choice when analyzing restriction map data for other types of DNA.

The conclusions reached both in this study and by Templeton (1983a, b) were made possible by the construction of an explicit probability model for the evolution of mtDNA. The assumptions of this model are themselves explicitly stated and open to testing by further empirical studies. DNA data, either from restriction site maps or from direct nucleotide sequences, may be the only type of systematic data currently available for which we can have confidence in the details of such a model. This is not to say that molecular data are any "better" than other forms of biochemical or morphological data for reconstructing phylogenies. However, since a probability model of evolution is required in order to use maximum-likelihood methods, DNA data (of the various types of data currently used for phylogeny reconstruction) is the best suited for developing and using explicitly statistical phylogenetic reconstruction methods.

ACKNOWLEDGMENTS

We thank the following people for reading, commenting on, and improving this work: R. S. Hoffmann, E. O. Wiley, D. O. Straney, and three anonymous reviewers. We particularly thank J. Felsenstein for a large number of comments which led to numerous improvements in this paper. We would also like to thank J. E. Feder, S. D. Ferris, W. M. Brown, and D. L. Swofford for discussions about the evolution of mtDNA.

REFERENCES

- ADAMS, J., AND E. D. ROTHMAN. 1982. The estimation of phylogenetic relationships from DNA restriction patterns and selection of endonuclease cleavage sites. *Proc. Natl. Acad. Sci. USA*, 79:3560-3564.
- AOKI, K., Y. TATENO, AND N. TAKAHATA. 1981. Estimating evolutionary distance from restriction maps of mitochondrial DNA with arbitrary G+C content. *J. Mol. Evol.*, 18:1-8.
- AVISE, J. C., R. A. LANSMAN, AND R. O. SHADE. 1979. The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. I. Population structure and evolution in the genus *Peromyscus*. *Genetics*, 92:279-295.
- AVISE, J. C., J. F. SHAPIRA, S. W. DANIEL, C. F. AQUADRO, AND R. A. LANSMAN. 1983. Mitochondrial DNA differentiation during the speciation process in *Peromyscus*. *Mol. Biol. Evol.*, 1:38-56.

- BROWN, W. M., M. GEORGE, JR., AND A. C. WILSON. 1979. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA*, 76:1967-1971.
- BROWN, W. M., E. M. PRAEGER, A. WANG, AND A. C. WILSON. 1982. Mitochondrial DNA sequences of primates: Tempo and mode of evolution. *J. Mol. Evol.*, 18:225-233.
- DAYHOFF, M. O., AND R. V. ECK. 1968. Atlas of protein sequence and structure 1967-1968. National Biomedical Research Foundation, Silver Spring, Maryland.
- FARRIS, J. S. 1970. Methods for computing Wagner trees. *Syst. Zool.*, 19:83-92.
- FARRIS, J. S. 1973. A probability model for inferring evolutionary trees. *Syst. Zool.*, 22:250-256.
- FARRIS, J. S. 1977a. Phylogenetic analysis under Dollo's Law. *Syst. Zool.*, 26:77-88.
- FARRIS, J. S. 1977b. Some further comments on Le Quesne's methods. *Syst. Zool.*, 26:220-223.
- FELSENSTEIN, J. 1973. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Syst. Zool.*, 22:240-249.
- FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.*, 27:401-410.
- FELSENSTEIN, J. 1979. Alternative methods of phylogenetic inference and their interrelationship. *Syst. Zool.*, 28:49-62.
- FELSENSTEIN, J. 1981. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. *Biol. J. Linn. Soc.*, 16:183-196.
- FELSENSTEIN, J. 1983. Parsimony in systematics: Biological and statistical issues. *Annu. Rev. Ecol. Syst.*, 14:313-333.
- FERRIS, S. D., W. M. BROWN, W. S. DAVIDSON, AND A. C. WILSON. 1981a. Extensive polymorphism in the mitochondrial DNA of apes. *Proc. Natl. Acad. Sci. USA*, 78:6319-6323.
- FERRIS, S. D., A. C. WILSON, AND W. M. BROWN. 1981b. Evolutionary tree for apes and humans based on cleavage maps of mitochondrial DNA. *Proc. Natl. Acad. Sci. USA*, 78:2432-2436.
- GILLESPIE, J. H., AND C. H. LANGLEY. 1979. Are evolutionary rates really variable? *J. Mol. Evol.*, 13:27-34.
- GOTOH, O., J.-I. HAYASHI, H. YONEKAWA, AND Y. TAGASHIRA. 1979. An improved method for estimating sequence divergence between related DNA's from changes in restriction endonuclease cleavage sites. *J. Mol. Evol.*, 14:301-310.
- KAPLAN, N., AND C. H. LANGLEY. 1979. A new estimate of sequence divergence of mitochondrial DNA using restriction endonuclease mappings. *J. Mol. Evol.*, 13:295-304.
- KAPLAN, N., AND K. RISKÓ. 1981. An improved method for estimating sequence divergence of DNA using restriction endonuclease mappings. *J. Mol. Evol.*, 17:156-162.
- LEQUESNE, W. J. 1974. The uniquely evolved character concept and its cladistic application. *Syst. Zool.*, 18:201-205.
- NEI, M., AND W. LI. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA*, 76:5269-5273.
- TAJIMA, F. 1983. Evolutionary relationships of DNA sequences in finite populations. *Genetics*, 105:437-460.
- TEMPLETON, A. R. 1983a. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and apes. *Evolution*, 37:221-244.
- TEMPLETON, A. R. 1983b. Convergent evolution and non-parametric inferences from restriction fragment and DNA sequence data. Pages 151-179 in *Statistical analysis of DNA sequence data* (B. Weir, ed.). Marcel Dekker, New York.
- UPHOLT, W. B., AND I. B. DAWID. 1979. Mapping of mitochondrial DNA of individual sheep and goats: Rapid evolution in the D-loop region. *Cell*, 11:571-583.

Received 6 June 1984; accepted 10 October 1984.

APPENDIX A

A Probability Model for mtDNA Evolution

Consider the probability of change within a single lineage over an interval of time of length T years. Any site at a 0 state (two or more substitutions away from a +) is assumed to have a negligible chance of eventually becoming a +; thus, there are six combinations of states at the endpoints that we will be interested in:

| t_0 | $t_0 + T$ |
|-------|-----------|
| 0' | 0' |
| 0' | + |
| + | 0' |
| + | + |
| 0' | 0 |
| + | 0 |

0' → +.—We will define r_i as the probability of substitution per nucleotide pair per year at codon position i , where $i = 1$ to 6 and where $r_4 = r_1$, etc. Although it is not possible to know exactly at which codon position any given change occurs, any six-base recognition word will contain exactly two first-position sites, two second-position base pairs, and two third-position sites.

The probability of no change at any of the six base pairs in one year is

$$q = \prod_{i=1}^6 (1 - r_i). \quad (A1)$$

Thus, the probability of one change in one year is $p = 1 - q$ (since the r_i s are $\ll 1$, the probability of multiple changes in one year is negligible). Note that when there is variation between the different r s that $p > 6r_{ave}$, which means that the variation in codon-position substitution rates results in a somewhat higher overall substitution rate per word than that indicated by a simple average rate.

For any particular 0' sequence, only 1 of the 18

possible base changes will result in a +. Therefore, the probability of gaining a + from a particular 0' in one year is $\frac{1}{18}$. The probability of beginning a segment T with a 0' and finishing with a + can be approached in the following way.

Let $T = t_1 + t_2$, where t_1 is the time interval up to and including the year when the substitution occurs and t_2 is the remaining time in interval T. For any given year within the interval T, the probability of the first and only substitution occurring that year is the product of the following probabilities: no substitution until that year; the correct substitution occurring in that year; and no further substitution until the end of the interval. This gives

$$P(t_1) = (1 - p)^{t_1 - 1} p (1 - p)^{t_2} = p(1 - p)^{T - 1}. \quad (\text{A2})$$

Summing over all t_1 s gives

$$P(T) = Tp(1 - p)^{T - 1}. \quad (\text{A3})$$

Recalling that $p \ll 1$, the probability of a single substitution can be approximated by the appropriate term of the Poisson distribution as

$$P(T) = Tpe^{-Tp}. \quad (\text{A4})$$

If we define $\lambda = Tp$, the mean number of substitutions in the interval T, then

$$P(T) = \lambda e^{-\lambda}. \quad (\text{A5a})$$

This allows us to consider the probability of more than one change per time interval as well. Most of the intervals we will be considering here are less than 10 million years long. With the r values we are using, at 10 million years the first three terms of the Poisson distribution (0, 1, and 2 changes per interval) account for 93% of the total probability. Thus, we will not consider any events requiring three or more substitutions in one interval.

For a known 0' site, of the 18 possible substitutions only one will result in a +. Thus,

$$P(T) = \lambda e^{-\lambda} (\frac{1}{18}). \quad (\text{A5b})$$

It is possible, however, for a 0' site to change to a + by two substitutions, although under only a restricted set of circumstances. The first change must be at the pair with the incorrect base, and must be to another 0' state, of which there are two available. The second change must be at the same base pair and must be to the nucleotide that gives a plus.

If each of the six r_i values were the same, then the probability of the second substitution occurring at the same position would be $\frac{1}{6}$. This is not the case, however, so we must first derive the probability of having two consecutive substitutions at the same site. This will simply be

$$p_d = \frac{\sum_{i=1}^6 r_i^2}{\sum_{i=1}^6 r_i} \quad (\text{A6})$$

which, if $r_1 = r_4 = 0.165 \times 6(r_{ave})$, $r_2 = r_5 = 0.085 \times 6(r_{ave})$, and $r_3 = r_6 = 0.25 \times 6(r_{ave})$, then $p_d \approx 0.2$. Thus,

the probability of acquiring a + from a 0' in T years is

$$P_{0' \rightarrow +}(\lambda) = \lambda e^{-\lambda} (\frac{1}{18}) + \frac{1}{2} \lambda^2 e^{-\lambda} (\frac{2}{18}) p_d (\frac{1}{6}) \quad (\text{A7a})$$

$$= \lambda e^{-\lambda} (\frac{1}{18}) + \frac{1}{2} \lambda^2 e^{-\lambda} (0.0074). \quad (\text{A7b})$$

0' → 0'.—We can also derive the probability of starting and finishing a time interval with a 0'. The simplest way for this to happen is to have no substitutions at all. As mentioned above, it is possible to have one substitution and still be at a 0', as long as the change is at the correct base and is to one of the two nucleotides which will not result in a +.

It is also possible, under an even more complicated set of restrictions, to have two substitutions in a time interval and both start and finish with a 0'. We will divide the possible combinations into two categories—one with both changes at the same base, the other with the two substitutions at different positions. The latter way is the simpler; the first change must take the word to a +, then a second change at any other position will result in a 0' state. These two substitutions could also occur in the reverse order. An example of this would be:



These two substitutions could also occur in the reverse order. Given that two substitutions do occur in an interval, the probability of this is $2(\frac{1}{6})(\frac{1}{6})(1 - p_d)$. If, on the other hand, both changes are at the same position, there are two more possibilities. The two substitutions may both be at the particular position which needs to be changed to acquire a +, in which case the first may take the word to a +, and the second back to a 0', or both substitutions may be to 0' sequences. The probability of this is $p_d [(\frac{1}{18}) + (\frac{2}{3})(\frac{1}{6})]$. The two changes may also both be at a position which already has the correct base, in which case the second change must be the exact reverse of the first, thus preserving the 0' sequence.

This last term, for a reversal, would be simple if the probabilities of substituting to and from each of the four bases were equal. Brown et al. (1982), however, have found that transitions are about nine times as likely as are transversions. This means that the probability of a reverse substitution is much greater than might otherwise be expected. In fact, it is probable that a particular base pair will toggle back and forth between either the two purine bases or the two pyrimidines before a transversion occurs. If the base present is A, the first change may be to G, with a probability of 0.9; or to either T or C, with probabilities of 0.05 each. Thus, the probability of an exact reversal is

$$p_r = 0.9^2 + 0.05^2 + 0.05^2 = 0.815.$$

So the probability that the two substitutions will both be at the same position, which already has the correct nucleotide, and that the second change is the reverse of the first, is $p_d (\frac{5}{6}) p_r$. Thus, overall, the probability of a 0' → 0' segment is

$$P_{0'-0}(\lambda) = e^{-\lambda} + \lambda e^{-\lambda}(\frac{1}{18}) + \frac{1}{2}\lambda^2 e^{-\lambda} \cdot \{2(\frac{1}{6})(\frac{1}{3})(1 - p_d) + p_d[\frac{1}{18}] + (\frac{1}{3})(\frac{1}{3})(\frac{1}{6})\} + p_d(\frac{1}{6})p_r \quad (\text{A8a})$$

$$= e^{-\lambda} + \lambda e^{-\lambda}(\frac{1}{18}) + \frac{1}{2}\lambda^2 e^{-\lambda}(0.251). \quad (\text{A8b})$$

$+\rightarrow+$.—A site which starts as a + may still be a + at the end of the interval under two circumstances. The more likely condition is that there are no substitutions in the interval. The only other possibility requires two substitutions, both at the same base pair and with the second being the exact reversal of the first. Thus the probability of a $+\rightarrow+$ segment is:

$$P_{+\rightarrow+}(\lambda) = e^{-\lambda} + \frac{1}{2}\lambda^2 e^{-\lambda} p_d p_r \quad (\text{A9a})$$

$$= e^{-\lambda} + \frac{1}{2}\lambda^2 e^{-\lambda}(0.163). \quad (\text{A9b})$$

Note that the second term of expression (A9a) is equivalent to the loss-gain convergence of Templeton (1983b: eq. [5]). This type of convergence would bias estimates of nucleotide substitution distance, but

it would only result in homoplasy when the convergence occurs between segments, not within a segment.

$+\rightarrow 0'$.—A + can change to a 0' either by having only one substitution or by having two—both of which must be at the same site and the second must not be the reverse of the first. The probability of this happening is

$$P_{+\rightarrow 0'}(\lambda) = \lambda e^{-\lambda} + \frac{1}{2}\lambda^2 e^{-\lambda} p_d (1 - p_r) \quad (\text{A10a})$$

$$= \lambda e^{-\lambda} + \frac{1}{2}\lambda^2 e^{-\lambda}(0.037). \quad (\text{A10b})$$

$0' \rightarrow 0$ and $+\rightarrow 0$.—These transformations are appropriate when there is no need to maintain a 0' state, because no further + states need be derived. The probabilities of these are simply

$$P_{0'-0}(\lambda) = 1 - P_{+\rightarrow+}(\lambda) \quad (\text{A11})$$

and

$$P_{+\rightarrow 0}(\lambda) = 1 - P_{+\rightarrow+}(\lambda). \quad (\text{A12})$$

APPENDIX B

All Possible Wagner Scenarios for Four Taxa

