

Estimating Divergence Times in the Presence of an Overdispersed Molecular Clock

David J. Cutler¹

Center for Population Biology, University of California at Davis

Molecular loci that fail relative-rate tests are said to be “overdispersed.” Traditional molecular-clock approaches to estimating divergence times do not take this into account. In this study, a method was developed to estimate divergence times using loci that may be overdispersed. The approach was to replace the traditional Poisson process assumption with a more general stationary process assumption. A probability model was developed, and an accompanying computer program was written to find maximum-likelihood estimates of divergence times under both the Poisson process and the stationary process assumptions. In simulation, it was shown that confidence intervals under the traditional Poisson assumptions often vastly underestimate the true confidence limits for overdispersed loci. Both models were applied to two data sets: one from land plants, the other from the higher metazoans. In both cases, the traditional Poisson process model could be rejected with high confidence. Maximum-likelihood analysis of the metazoan data set under the more general stationary process suggested that their radiation occurred well over a billion years ago, but confidence intervals were extremely wide. It was also shown that a model consistent with a Cambrian (or nearly Cambrian) origination of the animal phyla, although significantly less likely than a much older divergence, fitted the data well. It is argued that without an a priori understanding of the variance in the time between substitutions, molecular data sets may be incapable of ever establishing the age of the metazoan radiation.

Introduction

In 1965, Zuckerkandl and Pauling suggested the use of protein sequences to infer the time since divergence of two or more related species. Their suggestion was based largely on the observation that proteins did, in fact, accumulate differences in an amount proportional to the time since their common ancestor. One could, therefore, use this empirical pattern to estimate divergence times. A precise population genetics model to account for this observation was not presented.

Kimura and Ohta (1971) offered the neutral theory as a model to explain this pattern. Under the neutral theory, it can be shown that the number of substitutions in a lineage in T generations will be Poisson distributed with mean uT , where u is the mutation rate per sequence per generation (Kimura 1983). Thus, the neutral theory provides a precise statistical model that can be tested with molecular sequence data.

The first test of the constant-rate Poisson model of protein substitutions occurred early in the history of molecular clocks (Ohta and Kimura 1971). Ohta and Kimura examined three proteins in several pairwise comparisons in mammals. They showed that for two of the proteins in a few of the pairs, a Poisson process with a constant rate could be rejected. This result was somewhat hard to interpret, as no explicit phylogenetic hypothesis was made, and the effect of phylogeny on, among other things, the independence of comparisons went unconsidered. The first attempts to use phylogeny in an explicit manner occurred in the following three

years (Langley and Fitch 1973, 1974). Langley and Fitch examined four proteins in 18 species. The species were assumed to have a known phylogeny. The numbers of amino acid substitutions along all branches in the phylogeny were inferred. Next, all divergence times and substitution rates were found by maximum likelihood. Finally, a χ^2 (Langley and Fitch 1973) and a likelihood ratio test (Langley and Fitch 1974) were performed to ask whether the observed numbers of mutations on the branches were statistically different from those expected under a constant-rate Poisson model. This model was rejected with high confidence. Discussions in both papers attribute the rejection of the Poisson model to rate variation.

Since these pioneering works were performed, testing the constant-rate Poisson model has become common practice in molecular evolution (Li and Graur 1991; Li 1997), and new tests continue to be proposed (Muse and Weir 1992). Tests of the constant-rate Poisson model are generally called “relative-rate tests,” but this name may be misleading. Relative-rate tests examine agreement between the constant-rate Poisson model and the observed sequence divergence. When a test achieves a significant result, the model is rejected. The reason for rejection of the model may be that the constant-rate assumption failed. On the other hand, the model may have been rejected because the Poisson assumption failed (Gillespie and Langley 1979).

Standard Poisson processes have the property that the mean number of events in a given amount of time is equal to the variance in the number of events in that time. All relative-rate tests use this fact. Thus, when a relative-rate test returns a significant result, it may be because the mean numbers of events in the lineages are unequal. Conversely, it may be that the mean numbers are identical, but the variance in the numbers is larger than the mean (Gillespie and Langley 1979). A choice between these two alternative interpretations will be necessary to model an overdispersed molecular clock.

¹ Present address: Institute for Genetic Medicine, Johns Hopkins Medicine.

Key words: metazoan radiation, Cambrian explosion, overdispersion, molecular clock.

Address for correspondence and reprints: David J. Cutler, McKusick-Nathans Institute for Genetic Medicine, Jefferson Street Building 2-109, Johns Hopkins Medicine, 600 N. Wolfe Street, Baltimore, Maryland 21287.

Mol. Biol. Evol. 17(11):1647–1660. 2000

© 2000 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

Nevertheless, this dichotomy of interpretations is not always clearly distinguished.

Suppose one believes that the mean numbers of substitutions in two lineages are identical, but the variances are higher than the mean. One might naturally try to model this situation with a negative binomial distribution. The negative binomial is a discrete distribution with a variance larger than the mean, and as the variance approaches the mean, the negative binomial reduces to a Poisson (Stuart and Ord 1987, p. 178). It seems a logical choice to model a high variance process.

Alternatively, one might believe that the number of substitutions in a lineage is Poisson distributed, but the mean number varies. One might attempt to model this situation by assuming that the mean numbers of substitutions in lineages are independent and gamma distributed, and given the mean, the actual number of substitutions is Poisson distributed. The resulting substitution process (a Poisson process with a gamma-distributed mean) has a negative binomial distribution (Stuart and Ord 1987, p. 182). In other words, both interpretations (Poisson with variable rate versus high-variance constant rate) lead to identically distributed substitution processes. Distinguishing these alternatives with a single set of observations is impossible. In principle, if the evolution process were repeated over and over again, the alternatives could be distinguished, but given that the tree of life happened only once, it may be fundamentally unknowable which interpretation is correct.

A locus that fails a relative-rate test is often called overdispersed. Knowing why the locus is overdispersed may not be possible, but some assumption about the cause will be necessary in order to formulate a practical method to estimate divergence times using loci that might be overdispersed. In the past, four basic approaches have been taken to estimating divergence times with such loci: (1) don't ask, don't tell; (2) molecule shopping; (3) taxa shopping; or (4) making an explicit model for the overdispersion. Each option will be considered in turn.

Sometimes researchers interested in estimating divergence times do not perform relative-rate tests, or if they do perform the tests, the results are not reported. This approach has little to offer other than expediency. A second approach is to search for a molecule that passes, or is more likely to pass, a relative-rate test. The procedure can be either formal (acquire sequences; perform the tests; if you reject the constant-rate Poisson model, repeat; Kumar and Hedges 1998) or informal (introns are thought to be more "neutral" than exons, so do the analysis on introns because they are more likely to pass a relative-rate test, even if the test is never performed). Molecule shopping is a sensible procedure if one believes that there do exist molecules for which the constant-rate Poisson model is correct. If such molecules do exist, the researcher merely needs to find them, and shopping may be the appropriate way.

A third approach is taxa shopping. The basic idea is to exclude from the analysis taxa that are likely to cause significant relative-rate tests. Again, this procedure can be formal (perform a test; if the test is signif-

icant, remove a taxon and repeat; Takezaki, Rzhetsky, and Nei 1995) or informal (taxon X, say, nematodes, is thought to be "unusual," and therefore X is excluded from the analysis). This approach is sensible if one believes that most species evolve according to a constant-rate Poisson model, but occasionally there are a few species evolving at some different rate and therefore need to be excluded.

Of course, any sort of data shopping can potentially introduce bias into the analysis. If one believes that overdispersion is intrinsic to the process of evolution (Gillespie 1991) and that loci are, on average, overdispersed (Ohta 1995), then restricting one's analysis to loci and taxa which happen to pass relative-rate tests is inappropriate. Moreover, it is easy to imagine that the taxa and loci that happen to pass relative-rate tests might be a biased subset of all taxa and loci. In particular, the power of any relative-rate test increases with the amount of divergence in the sample. By shopping for molecules or taxa, one may be restricting one's analysis to loci and/or taxa that evolve unusually slowly. Finally, even though a locus is overdispersed, it still almost certainly contains some information about divergence times. Ideally, one would like to be able to use that information.

Therefore, the fourth approach, to build an explicit model of overdispersion, is favored here. Several previous authors have already taken this approach, but each analysis assumed that the substitution process was fundamentally Poisson and that rates varied between lineages. Hasegawa, Kishino, and Yano (1989) assumed three different rates in a tree of four primates. Lynch and Jarrell (1993) demonstrated a generalized linear model that can allow for arbitrary rate variation. Uyenoyama (1995) gave different rates to different functional classes of sequences. Sanderson (1997) assumed that rates autocorrelated through the history of a lineage, and he maximized this covariance. Thorne, Kishino, and Painter (1998) proposed a different model of rate autocovariance. The approach taken here will be from the opposite end of the dichotomy. In this paper, a model will be developed that assumes the number of substitutions in a lineage is stationary, but, unlike a Poisson process, the variance in the number of substitutions will not necessarily be equal to the mean. To do this, the Poisson process assumption will be replaced by a stationary process assumption, and a central limit theorem approximation will be required. The performance of the model will be examined in simulation, and the model will be used to estimate divergence times among land plants and among the higher metazoans.

Materials and Methods

Langley and Fitch with a Stationary Process

Under the constant-rate Poisson model, the times between substitutions are independent of each other and exponentially distributed with constant mean μ . The exponential distribution imposes the property that the variance in the time between substitutions must equal μ^2 . By relaxing the independence assumption, not requiring an exponential distribution, and utilizing a central limit

theorem approximation, a more general stationary process model can be used.

Let X_i be the time between substitution i and substitution $i + 1$. Under the constant-rate Poisson model, X_i 's are independent of one another and exponentially distributed. To relax these assumptions, first assume that the substitution process is stationary, such that the joint density of $X_{i_1}, X_{i_2}, \dots, X_{i_k}$ depends only on $i_2 - i_1, \dots, i_k - i_1$ (Cox and Isham 1980, p. 24). In other words, the substitution process is unaffected by translations of time. In place of assuming that all X_i 's are independent of one another, instead assume that the dependence of the X_i 's on one another decays over time. Thus, assume that

$$\sigma^2 = \lim_{k \rightarrow \infty} \left(\frac{1}{k} \text{var}\{X_1 + X_2 + \dots + X_k\} \right) \\ = \lim_{k \rightarrow \infty} \left(\text{var}\{X_1\} + 2 \sum_{i=1}^{k-1} \left[\left(1 - \frac{i}{k}\right) \text{cov}\{X_1, X_{i+1}\} \right] \right) \quad (1)$$

exists and that $\sigma^2 < \infty$. If $\mu = E\{X_i\}$ is the mean time between substitutions, for a Poisson process the times between substitutions are all independent of one another, such that $\text{Cov}\{X_1, X_{i+1}\} = 0$. As a result, $\sigma^2 = \mu^2$, and σ^2 is simply the variance in the time between substitutions. In fact, $\text{Cov}\{X_1, X_{i+1}\} = 0$ for any renewal process (such as Takahata's [1987] fluctuating neutral space), so that σ^2 in equation (1) is the variance $\text{Var}\{X_i\}$ of any single time between substitutions. For other stationary models, σ^2 is the variance in time between substitutions plus twice the sum of all the covariances in times between substitutions. For the sake of simplicity, σ^2 will be called the "cumulative variance per term" in time between substitutions. A finite σ^2 appeared to exist for every model of evolution studied by Gillespie (1993, 1994a, 1994b), although models without this property are certainly possible. Bickel and West (1998) model the substitution process as a doubly stochastic Poisson process with the assumption that σ^2 is infinite, and they obtain a good fit between observed mammalian substitutions and their model. Nevertheless, for this paper, σ^2 will be assumed to be finite.

Since the time between substitutions is not observable from molecular data, but the total number of substitutions may be inferred, the distribution for the total number of substitutions is needed. Finding this distribution for any given model of evolution is usually a formidable problem. As a result, this analysis will fall back on a central limit theorem approximation. If the total number of substitutions, S , in a given period of time, t , is large, S will be approximately normally distributed with mean t/μ and variance $\sigma^2 t/\mu^3$ (Cox and Isham 1980, p. 36).

The basic approach to finding divergence times is that of Langley and Fitch (1973). Amino acid or nucleotide sequences will be collected from L loci in K taxa. The K taxa will be assumed to have a known, accurate, and given phylogeny. The phylogeny will be assumed to be identical for each locus examined. The number of substitutions in each locus along each branch in the phy-

logeny will be assumed to be known without error. Of course, in practice, neither the phylogeny nor the number of substitutions along each branch is ever known without error, but must be inferred. Errors in this inference may lead to subsequent errors in analysis and inference (Gillespie and Langley 1979; Hudson 1983b). In particular, ancestral polymorphism can lead to separate loci having slightly different divergence times (Hudson 1983b). More rarely, ancestral polymorphism can lead to loci having different phylogenies (Hudson 1983b).

Let $S_{b,l}$ be the number of substitutions along branch b for locus l . By assumption, $S_{b,l}$ is normally distributed with mean $m_{b,l} = (T_i - T_j)/\mu_l$ and variance $s_{b,l}^2 = \sigma_l^2(T_i - T_j)/\mu_l^3$, where T_i is the time of divergence of the node above b , T_j is the time of divergence of the node below b , μ_l is the mean time between substitutions for locus l , and σ_l^2 is the cumulative variance per term. Thus, $S_{b,l}$ has probability density function

$$\phi_{b,l}(x) = \begin{cases} \frac{1}{\sqrt{2\pi s_{b,l}^2}} \exp\left[-\frac{(x - m_{b,l})^2}{2s_{b,l}^2}\right], & \text{if } x > 0 \\ \int_{-\infty}^0 \frac{1}{\sqrt{2\pi s_{b,l}^2}} \exp\left[-\frac{(y - m_{b,l})^2}{2s_{b,l}^2}\right] dy, & \text{if } x = 0. \end{cases} \quad (2)$$

The second condition assumes that a negative number of substitutions is impossible. Thus, $\phi_{b,l}(x)$ is a normal density on the positive real line, with an atom of probability mass at zero. Notice that the number of substitutions has been assumed to be real-valued (i.e., not an integer). This can be viewed as consistent with algorithms that infer real values for the number of substitutions on each branch, for example, the Fitch and Margoliash (1967) and neighbor-joining (Saitou and Nei 1980) algorithms.

Assuming that all loci and all branches are independent of one another, the overall likelihood of the observed numbers of substitutions is given by

$$\mathcal{L} = \prod_{b,l} \phi_{b,l}(S_{b,l}). \quad (3)$$

The parameter values (T_i , μ_l , and σ_l^2) that maximize \mathcal{L} are found numerically. To do this, $-\log(\mathcal{L})$ is minimized using the "Powell" algorithm found in Press et al. (1992). In principle, the mean time between substitutions (μ_l) and the divergence times (T_i) are confounded with one another. In order to obtain absolute estimates of these quantities, calibration using the fossil record is required. The treatment used here is inspired by Sanderson (1997) and is described in the next section.

Confidence intervals for all parameters can be estimated from the likelihood surface (Edwards 1992). If \mathcal{L}^* is the maximum likelihood, all points within two log likelihood units of \mathcal{L}^* are considered to be within the 95% confidence interval of the maximum (Edwards 1992). For each parameter A , the largest and smallest values of A are sought, such that there exists a point with the given value of A that is within two log likelihood units of \mathcal{L}^* . During this search, all other parame-

ters are allowed to take on any value. The search for these points is conducted using the Powell algorithm (Press et al. 1992).

Confidence interval estimation in this fashion suffers from three potential problems. First, if (A^{\min}, A^{\max}) is the estimated confidence interval for some parameter A , nothing guarantees that all values of A such that $A^{\min} < A < A^{\max}$ are also within two log likelihood units of the maximum. This will never be a problem for sufficiently convex likelihood surfaces, but should be kept in mind when small data sets are examined. Second, inadequate or incomplete searches of the likelihood surface will always lead to an underestimate of the “true” confidence interval. Third, the choice of two log likelihood units as defining the 95% confidence limits is based on a central limit theorem result for large data sets. It appears from simulation that small data sets tend to lead to an underestimate of the true interval (see *Results*).

Two distinct models are compared. The first is the stationary process (SP) model described above, and the second is the constant-rate Poisson process (PP) model, i.e., the model Langley and Fitch (1974) used. The index of dispersion, R , is the ratio of the variance in the number of substitutions to the mean number of substitutions. Under the PP model, $R = 1$ for all loci. Under the SP model, $R_l = \sigma_l^2/\mu_l^2$ for locus l .

Following Langley and Fitch (1973, 1974), model evaluation occurs in two ways. First, a χ^2 test is performed to assess the goodness of fit of the SP model and the PP model. For the SP model, χ^2 is calculated by

$$\chi^2 = \sum_{b,l} \frac{(S_{b,l} - m_{b,l})^2}{s_{b,l}^2}, \quad (4)$$

where $m_{b,l}$ is the expected number of substitutions on branch b for locus l , and $s_{b,l}^2$ is the variance in that number. For the PP model, χ^2 is given by

$$\chi^2 = \sum_{b,l} \frac{(S_{b,l} - m_{b,l}^*)^2}{m_{b,l}^*}, \quad (5)$$

where $m_{b,l}^*$ is the expected number of substitutions for branch b and locus l under the PP model. For both χ^2 tests, all branches with fewer than three substitutions were binned together.

Second, a series of likelihood ratio tests is performed to directly compare the fits of the two models. In order to create proper nesting between the SP and the PP models, for the likelihood ratio test only, the Poisson model is approximated by a normal with a mean and variance equal to the mean under the Poisson. This approximation is expected to be very poor when many branches have few substitutions. As a result, this test may be highly unreliable for data sets with very short branches (see *Results: Performance Evaluation*). The PP model has $K - 1 + L$ parameters ($K - 1$ divergence times and L substitution rates). The SP model has $K - 1 + 2L$ parameters ($K - 1$ divergence times, L mean times between substitutions, and L variances in time be-

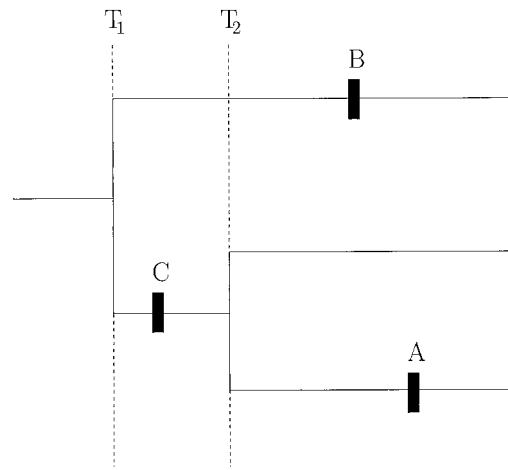


FIG. 1.—Fossil A provides a lower bound for times T_1 and T_2 . Fossil B provides a lower bound for time T_1 . Fossil C provides an upper bound for time T_2 and a lower bound for time T_1 .

tween substitutions). Thus, there are L degrees of freedom when comparing SP with PP.

Calibration

In order to assess absolute divergence times, it is necessary to calibrate the clock with fossil data. The traditional approach (Wray, Levinton, and Shapiro 1996) is to use fossils to “assign” dates to some of the nodes in the phylogeny. Next, given these fixed dates, the substitution rate is estimated under the Poisson model for taxa descended from these nodes. Finally, using the inferred substitution rate, divergence dates are estimated for the remaining nodes in the phylogeny. Unfortunately, using fossils to assign dates to nodes in a phylogeny is problematic.

Fossils never occur at nodes in a phylogenetic tree. One is never so lucky as to find a fossil that dates to the precise moment at which two taxa diverged from one another. Fossils do not provide estimates of divergence dates, but they do provide bounds for divergence dates (Sanderson 1997). The best possible situation would be to find a fossil that occurs on a branch above a node and find another fossil on a branch below a node (see fig. 1). Any fossil that occurs on a branch above a node provides an upper bound on the divergence time of that node. A fossil that occurs on a branch below a node provides a lower bound for that divergence time. In general, all fossils are assigned to branches. Fossils that occur on exterior branches provide only lower bounds for nodes. Fossils on interior branches provide both upper bounds and lower bounds. In order for divergence times to be bounded above, either at least one interior branch fossil must be identified (see appendix), or one must artificially assign an absolute upper bound for the entire tree. Presumably any assigned upper bound reflects some natural boundary (e.g., the age of the universe) or is a hypothesis being directly tested (e.g., the hypothesis that the higher metazoans radiated no more than 600 MYA, which is examined below).

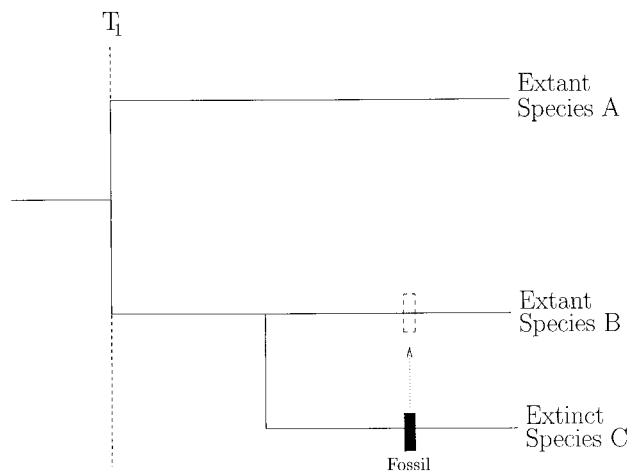


FIG. 2.—In a phylogeny of extant taxa, a fossil from an extinct sister taxon can be “assigned” to its closest extant relative. The fossil provides a lower bound for time T_1 . In order to do this, the researcher must be convinced that B and C are more closely related to each other than either is to any other taxon in the phylogeny.

Doyle and Donoghue (1993) point out that fossils can seldom even be assigned to branches in a phylogeny of extant taxa. Fossils are often not ancestors of extant taxa, but instead may be extinct sister groups to modern taxa or clades. This presents little difficulty when the extinct fossil is the sister of an extant terminal taxon, since the fossil can simply be “assigned” to the extant lineage without changing the analysis (see fig. 2). The fossil provides a lower bound for the node above the extant lineage, whether it occurred on the lineage leading to the extant taxa or on the lineage leading to a now extinct sister. Unfortunately, this does mean that fossils will seldom be unequivocally assignable to interior branches in the phylogeny. Since interior branch fossils are likely to be rare, and since fossils on interior branches provide upper bounds on divergence times, confidence intervals for divergence times will often be highly asymmetric with lower bounds tightly fixed, but with upper bounds often many times as large as the point estimates.

Thus, the approach to finding divergence times is to allow fossils to provide constraints on divergence estimates. The parameter values (T_i , μ_i , and σ_i^2) that maximize equation (3) are sought simultaneously, but if any divergence date (T_i) violates a fossil bound, the likelihood is set to 0. Fossils create the domain from which divergence dates may be drawn.

Clock calibration as described above, and as done by Sanderson (1997) in a different manner, has several advantages over the traditional approach. First, this method accurately represents the proper scope of one’s knowledge. One does not ever know for certain when two lineages diverged, and building an analysis by assuming that such a quantity is known just seems logically flawed. Second, under this approach, adding additional fossils can only improve estimates. Adding additional fossils will either create better bounds or have no effect (if there are already fossils above and below on the branch). In the traditional approach, by assuming

that fossils occur at nodes when they do not, each additional fossil adds error to the analysis. If this error has mean 0, adding additional fossils will probably improve the analysis. On the other hand, if the error exhibits bias (e.g., divergence times are generally underestimates of the true times; see Doyle and Donoghue 1993; Springer 1995), adding more fossils can have the effect of increasing one’s confidence in an erroneous estimate.

A third advantage to treatment of fossils in this manner is that it does not require the Poisson model to be correct. Traditionally, one estimates rates of substitution under a Poisson model, and then those substitution rates are used to estimate divergence times. Unfortunately, if the Poisson model is rejected, it is not at all clear how calibration should proceed. The approach taken here avoids such problems.

Finally, it should be noted that the approach to fossil calibration taken here can result in nonunique estimates of divergence times. A very simplistic example of this can be found in the appendix. In general, it is difficult to know a priori whether or not an inferred divergence time represents a unique maximum or, instead, represents one possible maximum from some interval of possibilities. For this reason, it is important not only to focus on estimated divergence times, but also to consider inferred confidence intervals. Assuming the likelihood surface is sufficiently convex, the inferred confidence intervals should include all possible maximum-likelihood estimators. For this reason, confidence intervals should be carefully considered along with point estimates.

C language source code to estimate divergence times and confidence intervals under both the PP model and the SP model is available from the author on request. The basic input files are PHYLIP tree files. The program relies on the Powell algorithm (Press et al. 1992) to find parameter values. This algorithm assumes parameters are unconstrained (i.e., can take on any value from $-\infty$ to ∞). As a result, the algorithm will often try negative μ ’s, negative σ^2 ’s, and divergence times which violate fossil constraints. Whenever such an event occurs, an extremely low likelihood ($e^{-10^{50}}$) is assigned to the offending set of parameters. This creates discontinuities at the boundaries of legal parameter space, and as a result, the initial parameter guesses must come from legal parameter space (i.e., initial μ ’s and σ^2 ’s must be positive, and initial divergence times must not violate fossil constraints).

Performance Evaluation

Several computer simulations were performed to examine the effectiveness of the SP model to estimate divergence times when the Poisson model was correct, as well as the effectiveness of the PP model to estimate divergence times when the stationary model was correct. The basic simulation approach follows Sanderson (1997). Several variables were examined in simulation: number of substitutions per branch, number of taxa, rate of fossilization, and cumulative variance per term (for the SP model simulations).

In all simulations, the number of terminal taxa examined was K , with $K = 10$ or $K = 25$. The phylogeny of these taxa is generated by a pure-birth process (Yule) with the time between taxon births being exponentially distributed with mean B/n , where n is the number of taxa currently “alive.” The simulations used the time reversibility property of exponential distributions and formally modeled a pure-death process. As a result, the implemented code bore a strong similarity to a neutral coalescent (Hudson 1983a), but one where the time between “common ancestor” events had mean B/n , rather than $2B/n(n-1)$. Thus, in these simulations, all branches are on average length $B/2$. The mean time between substitutions, μ , was fixed in all simulations with $\mu = 1$, but average branch length was allowed to vary, with B ranging from 5 to 50. Therefore, the average number of substitutions per branch ranged from 2.5 to 25. The majority of the simulations focused on $B = 5$ or $B = 50$, and only these values are reported.

The number of loci examined per simulation, L , was either 1, 5, or 10. For a branch of length T , under the Poisson model, the number of substitutions at a given locus along this branch was drawn from a Poisson distribution with mean T/μ . Under the SP model, the number of substitutions was drawn from a normal distribution with mean T/μ and variance $T\sigma^2/\mu^3$. All negative numbers of substitutions were replaced by 0. Simulated branch lengths were not rounded when estimates were made under the SP model. When estimated under the PP model, branch lengths were rounded to the nearest integer. The cumulative variance per term, σ^2 , ranged from 5 to 50 for the SP model. Within any one simulation, all loci were assumed to have the same σ^2 .

By assumption, the number of fossils on a branch of length T that are recovered and can be assigned to this branch is Poisson distributed with mean T/f . Here, f varies with B so that the average number of fossils per branch ranges from 0.005 to 1. Given that a fossil occurred on a branch with endpoint t_1 and t_2 , the date of the fossil is assumed to be uniformly distributed on (t_1, t_2) .

In all simulations, the phylogeny and the number of substitutions per branch were assumed to be known without error. Any simulation with no substitutions on any branch for a locus was discarded. Similarly, any simulation without any fossils on an internal branch was discarded (without internal branch fossils, divergence times cannot be bounded from above). For any given set of parameters, 100 simulations were performed.

Land Plants

The land plant data set was provided by Sanderson (1997). This data set consisted of a single locus (the chloroplast *rbcl* gene) of nucleotide sequence. The phylogeny and the number of substitutions per branch were both given by Sanderson. The number of substitutions per branch ranged from 3 to 121, with a mean of 36.3 substitutions and median of 26.5 (the median would correspond to $B = 52$ in the simulations). Two internal fossil calibration points were also given by Sanderson

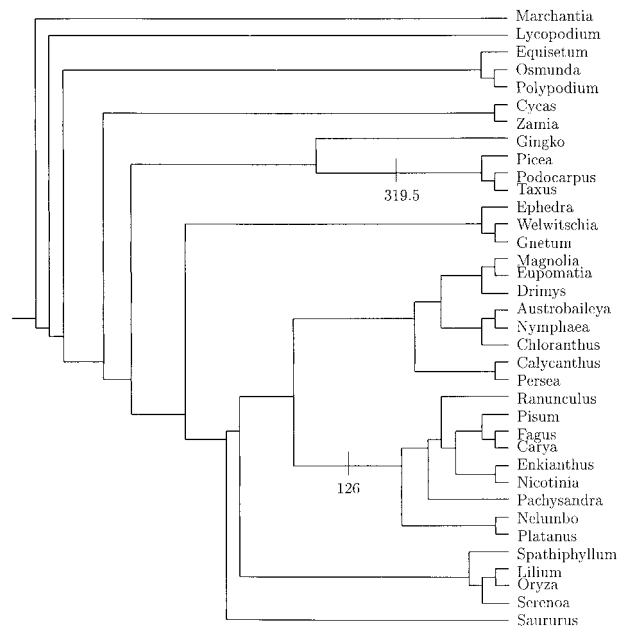


FIG. 3.—Land plant tree topology. Branch lengths are not to scale. Fossil calibration points are indicated.

(fig. 3). Two hypotheses were considered. The first hypothesis imposed the condition that the root of the phylogeny was no older than 450 Myr (Sanderson 1997). The second hypothesis made no assumption about the root.

Metazoa

Five mitochondrial (*ATPase subunit 6*, *cytochrome B*, *cytochrome C oxidase subunit 1*, *cytochrome C oxidase subunit 2*, *NADH subunit 1*) and two nuclear (*actin*, *histone H3*) amino acid sequences were acquired from 14 higher metazoan species (plus one outgroup species). Sequences were obtained primarily from Greg Wray's website (Wray, Levinton, and Shapiro 1996) with additional data from GenBank. The amino acid sequences were aligned with CLUSTAL W using the default parameters. All pairwise per-site divergences were calculated with the PHYLIP program *protdist*, with default parameters using a Dayhoff PAM substitution model to correct for multiple hits.

Since the phylogeny of the higher metazoans is far from established (McHugh and Halanych 1998), seven distinct phylogenetic hypotheses were considered (figs. 4–10). For each of the seven unrooted (but containing a bacterial—*Escherichia coli* or *Rhodobacter sphaeroides*—outgroup to estimate the number of substitutions on the most interior branches) trees, for each locus, substitutions per branch were inferred by least-squares minimization using the PHYLIP program *fit* with default parameters. The average number of substitutions per branch varied depending on the tree being considered, but the mean numbers of substitutions per branch were 65, 37, 42, 52, 64, 2.5, 0.5 for *ATPase subunit 6*, *cytochrome B*, *cytochrome C oxidase subunit 1*, *cytochrome C oxidase subunit 2*, *NADH subunit 1*, *actin*,

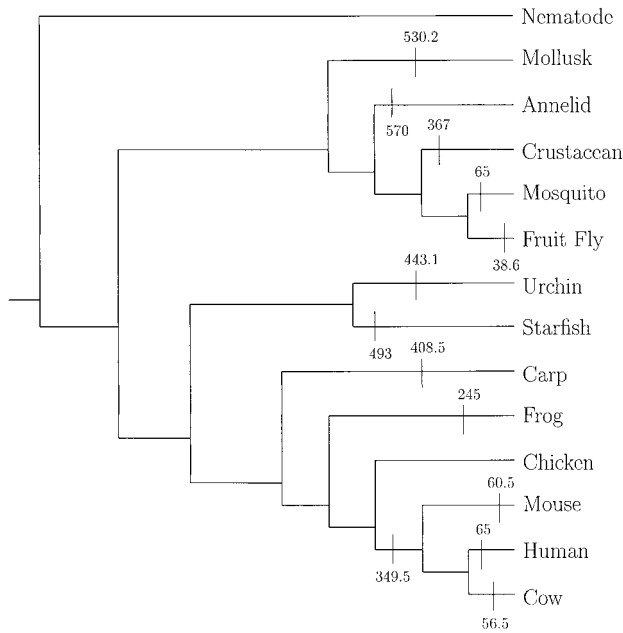


FIG. 4.—Metazoan tree 1 topology. Branch lengths are not to scale. Fossil calibration points are indicated.

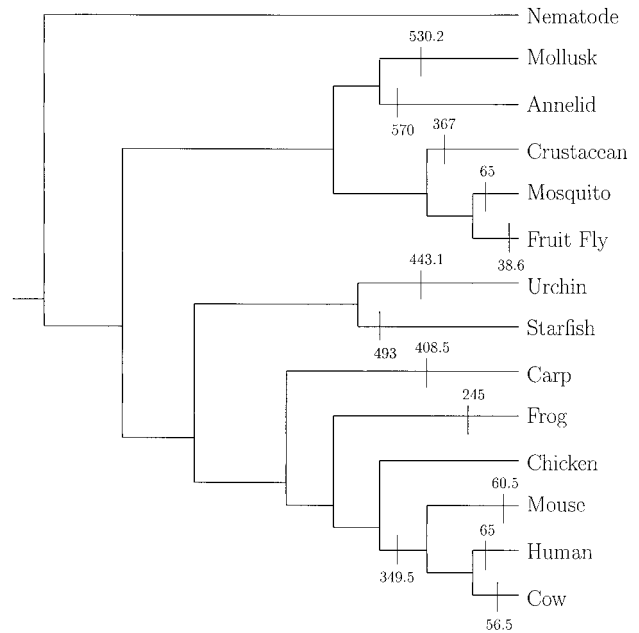


FIG. 6.—Metazoan tree 3 topology. Branch lengths are not to scale. Fossil calibration points are indicated.

and *histone H3*, respectively. Rooted trees were then obtained by removal of the outgroup.

Following Wray, Levinton, and Shapiro (1996), fossil calibration was taken from Benton (1993) and is indicated in figures 4–10. For each phylogeny, two hypotheses were considered. The first hypothesis might be called the “Cambrian explosion hypothesis” and imposes the condition that the root of the phylogeny is no older than 600 Myr (30 Myr older than the oldest identified annelid fossil). The second hypothesis makes no assumption about the age of the root. All sequences, alignments, fossil calibrations, and phylogenies, along

with branch lengths, are available from the author on request.

Results

Performance Evaluation

Both PP and SP simulations were performed. Analysis of the simulations focused on three areas: estimation of the age of the entire clade, confidence intervals associated with that estimate, and model selection. Each of these areas will be discussed in turn.

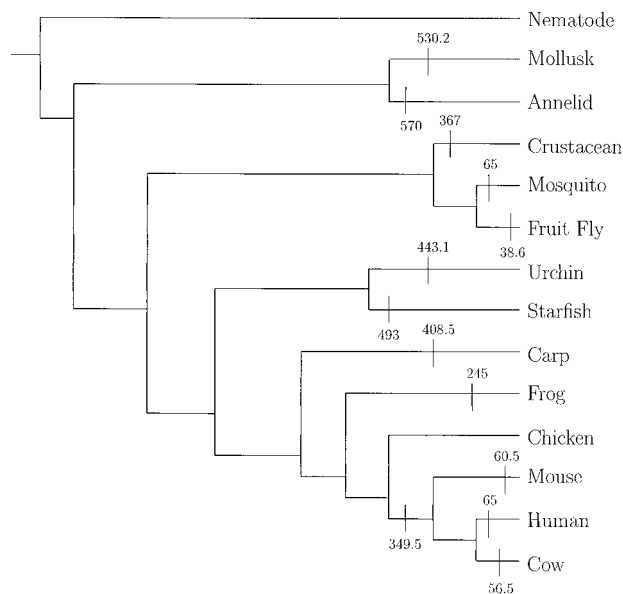


FIG. 5.—Metazoan tree 2 topology. Branch lengths are not to scale. Fossil calibration points are indicated.

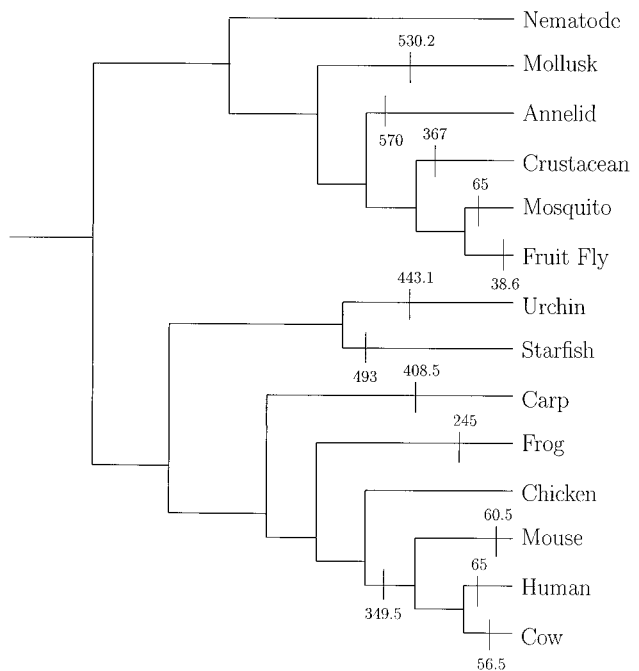


FIG. 7.—Metazoan tree 4 topology. Branch lengths are not to scale. Fossil calibration points are indicated.

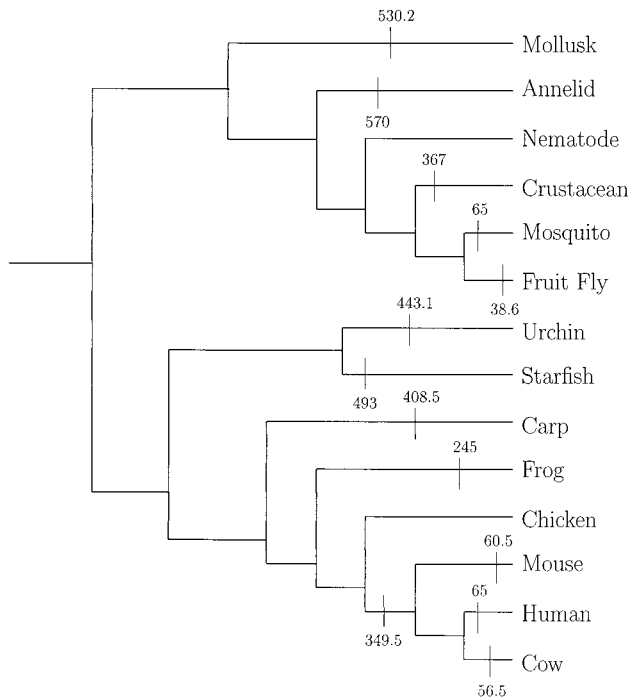


FIG. 8.—Metazoan tree 5 topology. Branch lengths are not to scale. Fossil calibration points are indicated.

Special attention was given to estimation of the age of the entire clade for three reasons. First, because of the stochastic nature of the treatment of fossils, interior nodes were often tightly bounded in some simulations, while being nearly unconstrained in others. The root of the tree was the only node never bounded from above by fossils; hence, only the root had consistently large confidence intervals. Second, the age of the root of the tree is often a question of great interest (e.g., in the metazoan data set). Third, because maximum-likelihood estimates of divergence times need not be unique (see appendix), confidence intervals may be more descriptive than single-point estimates.

When simulations were performed under a Poisson model, point estimates of the age of the root of the tree were almost always nearly identical under either a PP or an SP estimation procedure. The correlation coefficient between these age estimates under PP and SP was greater than 99%. Thus, when the true model is Poisson, assuming that the model is actually a stationary process does very little to change estimates of divergence times. The average error of these estimates (Sanderson 1997) ranged from approximately 5% (from simulations with long branches, many loci, and many fossils) to approximately 45% (from simulations with short branches, few loci, and few fossils) (see tables 1 and 3).

Similarly, when the true model was stationary, both the PP and the SP models tended to make similar estimates of the age of the clade, with the correlation coefficient between their estimates being greater than 95%. The mean error ranged from approximately 13% to over 66% and was again inversely proportional to branch length, number of loci, and number of fossils. Additionally, the mean error was also an increasing function of

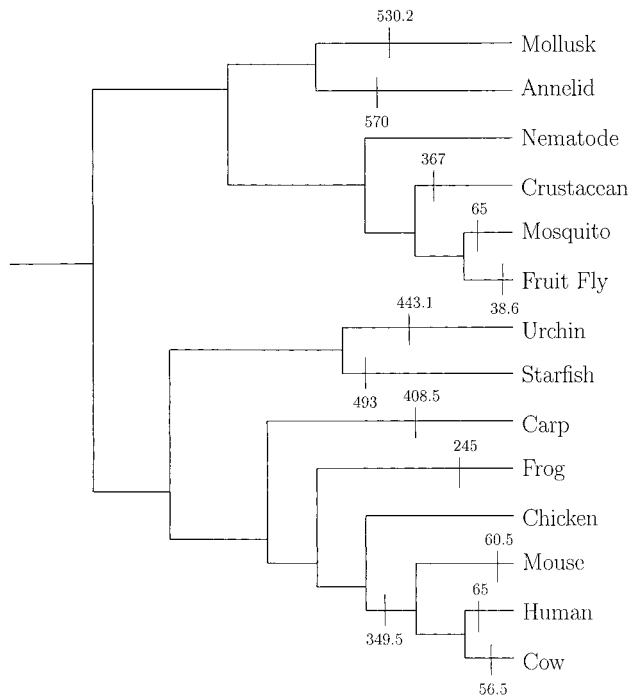


FIG. 9.—Metazoan tree 6 topology. Branch lengths are not to scale. Fossil calibration points are indicated.

the index of dispersion. Higher variance loci produced greater errors in estimates (see tables 2 and 3).

In an infinitely large data set, with a comprehensive search of the likelihood surface, inferred confidence intervals should include the true parameter value 95% of the time. Under a Poisson substitution model and the PP

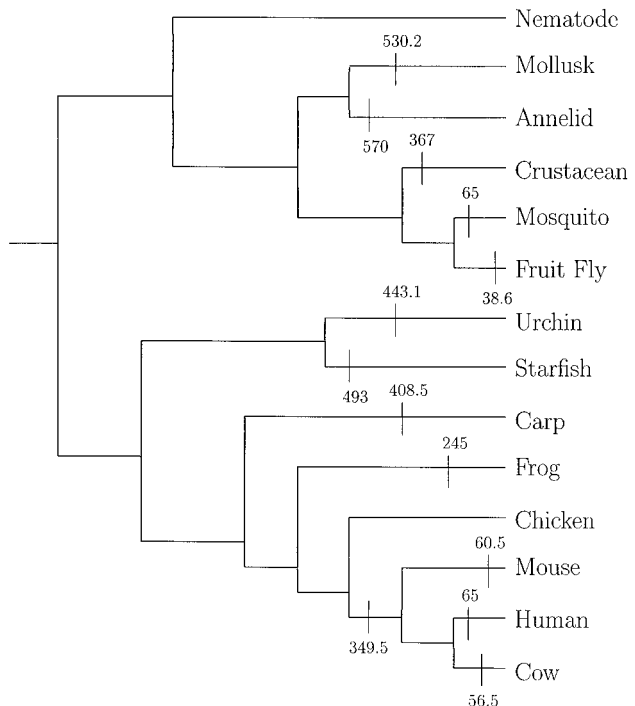


FIG. 10.—Metazoan tree 7 topology. Branch lengths are not to scale. Fossil calibration points are indicated.

Table 1
Poisson Process Simulations for the Age of the Root

True Model Poisson		<i>L</i> = 1	<i>L</i> = 5	<i>L</i> = 10
% error of estimated age	PP	18.6	16.4	15.4
	SP	18.6	16.2	15.2
Frequency of confidence interval including true time	PP	84.5	84.5	87.5
	SP	79.3	82	84.3
% rejected by χ^2 test	PP	2.5	2.67	2.83
	SP	1.8	1.17	1.0
Frequency of PP model selected by likelihood ratio test	2.5 substitutions per branch	73	57.3	32
	25 substitutions per branch	52	80.3	82.3
Frequency of SP model selected by likelihood ratio test	2.5 substitutions per branch	27	42.7	68
	25 substitutions per branch	48	19.7	17.7

NOTE.—*L* is the number of loci. PP indicates estimation under a Poisson process assumption. SP indicates estimation under a stationary process assumption. All values are averaged over several fossilization levels.

estimation procedure, confidence intervals included the true age of the clade roughly 83% of the time with 1 locus, 84% of the time with 5 loci, and 87% of the time with 10 loci. When the correct substitution model was Poisson but estimation was performed with the SP model, confidence intervals included the true value 79%, 82%, and 84% of the time for 1, 5, and 10 loci, respectively. Thus, when the true model is Poisson, inferred 95% confidence intervals might be thought of as the 80% confidence intervals.

When the true model was stationary, the PP and SP confidence intervals behave fundamentally differently. PP confidence intervals are often too small. With only one locus, PP confidence intervals include the true time as infrequently as 38% of the time, and averaged over all one locus stationary process simulations, no more than 64% of the time. As expected, the larger the index of dispersion, the worse the confidence intervals. On the other hand, for single-locus simulations, the SP model confidence intervals include the true time 75%–80% of the time, regardless of the index of dispersion. When 5 or 10 loci are examined, the PP model confidence interval includes the true time approximately 80% of the time, and the SP model includes the true time 85% and 87% of the time, respectively.

Thus, when the true model is stationary, the confidence interval from the SP model can be loosely thought of as the 80% confidence interval. The confi-

dence interval for the PP model should not be considered so. In fact, from one-locus simulations, the PP confidence intervals appear worthless (often less than a 50% chance of including the true value). This result should be cause for deep concern, since many divergence estimates reported in the literature are based on a single locus and a Poisson model.

Model selection is performed in two ways. First, a χ^2 test is performed to assess goodness of fit of the PP and the SP models. When the true model is Poisson, the PP model is rejected on average 2.67% of the time at a 5% significance level. When the Poisson model is correct, the SP model is rejected 1.3% of the time, also at a 5% significance level. Thus, the χ^2 test seems somewhat conservative when the Poisson model is correct. When the stationary process model is correct, the SP model is rejected 1.4% of the time, averaged over all simulations. The PP model is rejected from 25% (1 locus, *R* = 5) to over 99% of the time (10 loci, *R* = 50). Thus, there is moderate to great power to reject PP as long as the index of dispersion is large enough and enough loci are considered.

Model selection is also performed by way of a likelihood ratio test. For a given simulation, when the SP model fits significantly better (at the 5% level) than the PP model, we say the SP model is selected; otherwise, the PP model is selected. In order for this testing procedure to behave properly, there must be enough data

Table 2
Stationary Process Simulations for the Age of the Root

TRUE MODEL STATIONARY		<i>L</i> = 1			<i>L</i> = 5			<i>L</i> = 10		
		$\sigma^2 = 5$	$\sigma^2 = 10$	$\sigma^2 = 50$	$\sigma^2 = 5$	$\sigma^2 = 10$	$\sigma^2 = 50$	$\sigma^2 = 5$	$\sigma^2 = 10$	$\sigma^2 = 50$
% error	PP	28.2	31.4	35.3	24.7	26.4	26.6	25.1	24.5	25.6
	SP	29.4	32.3	37	24.6	26.7	30.5	25.1	24.4	27.6
% within confidence interval	PP	76.2	66.4	50	88.6	83.2	68.8	87.4	81.4	73.2
	SP	75.8	75.6	79.3	88.4	85.6	79.3	89	88.4	84.5
% χ^2 rejected	PP	24.4	45.5	79.5	52.2	79.2	96	64	83	99.3
	SP	3	3.6	5.3	0.2	0	0	0.2	0	0.3
% PP model selected	<i>B</i> = 5	26	11.5	0.5	0	0	0	0	0	0
	<i>B</i> = 50	20.7	1.7	0	0	0	0	0	0	0
% SP model selected	<i>B</i> = 5	74	88.5	99.5	100	100	100	100	100	100
	<i>B</i> = 50	79.3	98.3	100	100	100	100	100	100	100

NOTE.—*L* is the number of loci. σ^2 is the cumulative variance per term. Note that μ (the expected time between substitutions) = 1. PP indicates estimation under a Poisson process assumption. SP indicates estimation under a stationary process assumption. *B* = 5 implies that branches, on average, had 2.5 substitutions. *B* = 50 implies that branches, on average, had 25 substitutions. All values are averaged over several fossilization levels.

Table 3
Effect of Fossilization Level on the Age of the Root

% ERROR OF ESTIMATED AGE	EXPECTED NO. OF FOSSILS PER BRANCH ^a			
	0.005	0.25	0.5	1
Estimated under Poisson process ^b . . .	0.41	0.19	0.16	0.08
Estimated under stationary process ^b . .	0.43	0.19	0.17	0.08

^a Listed values are the expected numbers of fossils per branch in a given simulation. Any simulation with no interior branch fossils was discarded; therefore, the actual number of fossils in the simulations analyzed was biased to be larger than this value.

^b Averages taken over all other parameter values.

so that likelihood surfaces are nearly Gaussian. Small data sets can lead to incorrect inferences, and this is seen in simulations. Moreover, the SP model makes the assumption that a large number of substitutions occur on each branch. Violation of this assumption also leads to incorrect inference.

With only one locus and few substitutions per branch (2.5 per branch), when the Poisson model is correct, the PP model is selected 73% of the time, and the SP model is selected 27% of the time. As the number of loci increase, with short branches, the inference procedure gets worse. With five loci, the two models (PP and SP) are selected 57% and 43% of the time, respectively. With 10 loci, the models are selected 32% and 68% of the time, respectively. Thus, when the true model is Poisson and the number of substitutions per branch is very small, the Poisson model is selected too rarely, and this bias becomes more pronounced as the number of loci increases. The reason for this, almost certainly, can be traced to the normal approximation. In order to have proper model nesting when selecting models with a likelihood ratio test, the Poisson model is approximated by a normal distribution. The approximation is very poor for short branch lengths, and as a result, the likelihood ratio test with short branches is highly unreliable.

With large numbers of substitutions (25) per branch and under Poisson simulation, a bias in selection still exists, but adding loci appears to diminish it. With one locus, the model selection splits are 52% and 48%. With five loci, one sees 80% and 20%, and with 10 loci, the splits are 82% and 18%. Thus, there is still a bias (ideally, one should select the PP model 95% of the time), and the size of this bias is similar to that observed for confidence interval estimation.

When the true model is stationary, the likelihood ratio test is more reliable. With one locus, the PP model is selected less than 25% of the time, and the SP model is selected more than 75% of the time (over all simulation parameters). With five or more loci, the SP model is selected 100% of the time (over all simulation parameters). Thus, when the true model is stationary, the likelihood ratio test selects the correct model the vast majority of the time, as long as there are at least five loci. Overall, though, the likelihood ratio test should not be seriously relied upon unless there are several loci and many substitutions per branch.

Land Plants

Sanderson (1997), in addition to implementing a rate autocovariance model of the substitution process, also implemented the PP model to estimate divergence times. His implementation of PP and the implementation of PP done here, when applied to his land plant data set under the assumption that the root of the tree is no older than 450 Myr, produce the same divergence time estimates to three significant digits.

Sanderson (1997) highlights three nodes in the tree where the PP procedure is likely to have inferred an incorrect divergence estimate. Under the PP model, the cycad crown group is estimated to have diverged 112 MYA, with a confidence interval of (85 MYA, 144 MYA). The fossil record suggests that this divergence probably took place in the Carboniferous (Sanderson 1997), roughly 300 MYA. Thus, the estimated date is, in all likelihood, more than 150 Myr too late. Similarly, the fossil record suggests a Mesozoic divergence of *Fagus* and *Carya*, as well as *Nelumbo* and *Platanus*. The PP estimation procedure places both of these divergences in the Cenozoic: 39 MYA, with a confidence interval of (28 MYA, 50 MYA), for *Fagus/Carya* and 54 MYA, with a confidence interval of (37 MYA, 74 MYA), for *Nelumbo/Platanus*.

Interestingly, the SP model produces even worse point estimates for these problem taxa, but produces confidence intervals that should lead one to the proper inference. The cycad crown group is estimated to have diverged 57 MYA, but the confidence interval runs from 16 MYA to 261 MYA. Similarly, the *Fagus/Carya* split is placed at 28 MYA, with a confidence interval of (8 MYA, 78 MYA), and the *Nelumbo/Platanus* split is placed at 21 MYA, with a confidence interval of (4 MYA, 124 MYA). Thus, the *Nelumbo/Platanus* split has a confidence interval that includes times that are both a factor of five larger and a factor of five smaller than the point estimate. The confidence interval for the cycad crown group does not quite reach the supposed divergence time for these groups (estimated to be roughly 300 MYA), but the confidence intervals for *Nelumbo/Platanus* and *Fagus/Carya* are likely to include the true divergence times. Moreover, in all three cases, the confidence intervals are so wide as to leave one highly suspicious that the maximum-likelihood estimates are not unique, and there may be a very large region of equally likely divergence estimates.

In this data set, the PP model can be rejected by a simple goodness-of-fit test ($\chi^2 = 259.5$, which yields $P < 10^{-9}$); the SP model cannot be ($\chi^2 = 69.7$, which yields $P \approx 0.32$). Using a likelihood ratio test, the PP model is rejected in favor of the SP model ($P < 10^{-5}$). Consistent with both of these results, the inferred index of dispersion of this locus is 19.7, i.e., nearly 20 times as large as that expected under a constant-rate Poisson model.

Overall, a PP model is rejected. The SP model cannot be rejected by a χ^2 test. Moving from a Poisson model to an SP model enlarges the inferred confidence intervals on nearly all divergence estimates and shows

Table 4
Metazoan Data (unconstrained time to root)

	$-\log(\mathcal{L})$	Stationary process root	Divergence (MYA)	Poisson process root	Divergence (MYA)
Tree 1	615.3	1,407.5	(1,129.6, 1,752.35)	1,033.8	(988, 1,177)
Tree 2	620.6	1,532.6	(1,223, 1,931.1)	1,096.1	(1,041.9, 1,154.3)
Tree 3	629.4	1,502.9	(1,193.4, 1,912.5)	1,088.9	(1,037.2, 1,144.4)
Tree 4	638.5	1,022.1	(839.8, 1,295.1)	863.5	(833.9, 965.6)
Tree 5	641	680.8	(647.2, 1,349)	730.2	(713.8, 836)
Tree 6	644.6	1,056.1	(848.9, 1,305)	899.3	(866, 934.7)
Tree 7	650	1,038.1	(881.8, 1,286.7)	915.4	(880.6, 932.5)

that all three of the problematic taxa have enormous confidence intervals. Thus, in the three cases in which we suspected for a priori reasons that our point estimates were inaccurate, the confidence intervals told us that the point estimates were highly unreliable.

Sanderson (1997) constrains the root of the land plants to be no older than 450 Myr. If this constraint is removed, several interesting results occur. First, the PP model is still rejected with extremely high confidence ($P < 10^{-9}$) and the SP model is not ($P = 0.38$). The SP model without the constraint also fits significantly better than the SP model with it ($P < 0.01$, $df = 1$) by a likelihood ratio test. Neither the point estimates nor the confidence intervals for the three problem nodes change greatly (less than 10%). The root of the tree changes considerably. When the root is constrained to be <450 MYA, it is estimated to be 450 MYA, with a confidence interval of (430 MYA, 450 MYA). Without this constraint, the estimate becomes 612 MYA, with a confidence interval of (492 MYA, 840 MYA). Thus, the data become significantly more likely if one assumes that there has been 162 Myr of evolution in the land plants that has not been observed in the fossil record. A date this early, at first blush, appears preposterous, for it suggests a common ancestor of land plants that predates land on our planet. Of course, there is no necessary reason to suspect that the common ancestor of modern land plants was itself terrestrial. Nevertheless, either this date is incorrect, or very little is known about the early history of land plants. Finally, by removing the constraint, the index of dispersion for this locus is reduced to 16.8, as opposed to 19.7.

Metazoa

Seven separate phylogenies were considered, and two types of hypotheses were considered. One hypothesis, the Cambrian explosion hypothesis, constrained the root of the tree to be no older than 600 Myr. The other hypothesis made no constraint on the root. We begin by discussing the unconstrained hypothesis.

All seven trees have several features in common. In all cases, the PP model was rejected by a χ^2 test with extraordinarily high confidence ($P < 10^{-9}$). In all cases, the SP model could not be rejected by a χ^2 test. In all cases, the SP model fit significantly better than the PP model by a likelihood ratio test ($P < 10^{-5}$). Thus, for all seven trees, the SP model was selected as the best fitting model.

By a direct comparison of likelihoods, the best fitting tree is tree 1, the second best fitting tree is tree 2, . . . , and the worst fitting tree is tree 7. Using a naive comparison of log likelihoods for these incompatible tree topologies, each tree fits significantly better than the tree that follows it. Point estimates for the root of the tree are given in table 4. The three best fitting trees all have the property that nematodes are an outgroup to all the other species in the tree. The four worst fitting trees do not have this property. The confidence interval for the root of all seven trees includes 1,223–1,286 MYA, suggesting that no matter which tree is correct, a metazoan divergence time of 1.2 billion years ago can never be ruled out. For all seven trees, confidence intervals are asymmetric, with larger ranges above the point estimates than below them. This asymmetry can be traced to the dearth of interior branch fossil calibration points (a single point is used), since only interior branch fossils help to provide upper bounds for divergence estimates. Finally, simulation studies suggest that these confidence intervals should be thought of as perhaps the 80% bounds and nothing better, even when the SP model is correct. If this model is not a correct description of the evolutionary process, no conclusions should be made.

For the best fitting tree, estimates of the index of dispersion were 27.4, 10.6, 18.6, 12.2, 16.3, 51.7, and 107.7 for *ATPase6*, *coxI*, *coxII*, *cytB*, *NADH1*, *actin*, and *histone*, respectively. The last two estimates come from slow-evolving nuclear loci, and these estimates of the index of dispersion are highly unreliable. The confidence interval for the estimates of the mean time between substitutions ranged over four orders of magnitude for both *actin* and *histone*. Estimates of the variance in time between substitutions had even larger confidence intervals. For trees 2–7, estimates of the index of dispersion for the first five loci were generally larger than for tree 1. No trend existed for *actin* and *histone* (as might be expected from the enormous confidence intervals).

To facilitate comparison with previous studies, table 4 also includes the root divergence estimates under the PP model (constant-rate Poisson model). These estimates should not be considered seriously, since the PP model has been clearly rejected. Nevertheless, the reader may note that these estimates are generally consistent with Wray, Levinton, and Shapiro (1996), but slightly smaller in absolute value. The favored SP model estimates are perhaps slightly larger than Wray, Levinton,

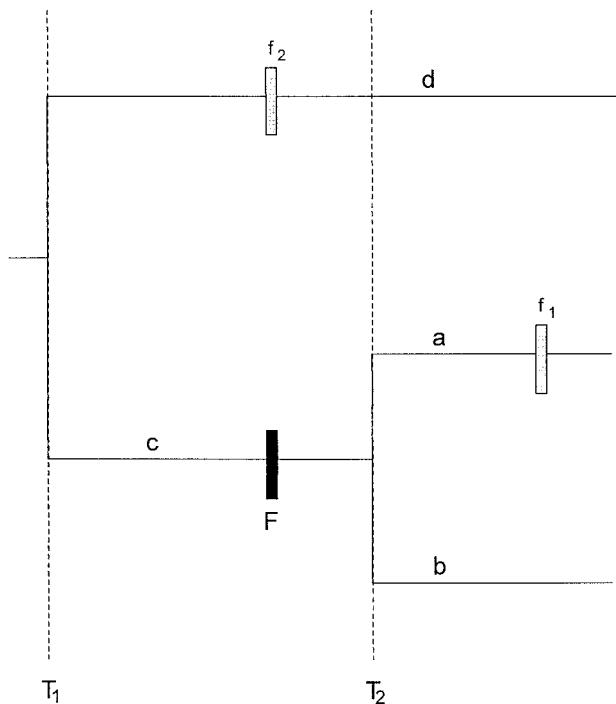


FIG. 11.—The fossil labeled *F* creates both upper and lower bounds on all divergence times. Fossils at either f_1 or f_2 create only lower bounds. The labels *a*, *b*, *c*, and *d* are the numbers of substitutions found on each of the four branches.

and Shapiro's (1996) for the three most favored trees, but the confidence intervals are very wide and certainly include Wray, Levinton, and Shapiro's date. For trees 1–3 under the SP model, the chordate/arthropod divergences are estimated to be 808 MYA (722 MYA, 958 MYA), 871 MYA (754 MYA, 1,045 MYA), and 840 MYA (740 MYA, 1,004 MYA). The chordate/arthropod split is the root for trees 4–7, given in table 4. These estimates are generally consistent with Gu (1998).

The Cambrian explosion hypothesis constrains the root of the phylogeny to be no older than 600 Myr. Under this constraint, the best fitting tree was tree 5, followed by trees 1, 2, 4, 6, 3, and 7. There was no significant difference between the fits of trees 2 and 4 and trees 3 and 7. All other pairs were significant. The best constrained tree (tree 5) did not fit significantly worse than the worst fitting unconstrained tree (tree 7). Otherwise, all unconstrained trees fit better than all constrained trees. Thus, a logical inference (but see discussion below) is rejection of the Cambrian explosion hypothesis, since unconstrained trees make the data significantly more likely.

For all seven constrained trees, the PP model could be rejected by a χ^2 test with extremely high confidence ($P < 10^{-9}$). Conversely, the SP model could never be rejected. In all cases, the SP model fit the data significantly better than the PP model by a likelihood ratio test. Thus, the SP model was always favored and never rejected for any of the seven constrained trees. The natural interpretation is that even under a Cambrian explosion hypothesis, the SP model fits the data well in absolute terms. The data become even more likely if the

Cambrian constraint is removed, but there is no way to reject the Cambrian explosion hypothesis solely on the basis of the goodness of fit of the SP model.

Discussion

In both data sets, land plants and metazoa, the constant-rate PP model was rejected in favor of an SP model. The improvement in the fit was highly statistically significant, and in no cases could the SP model be rejected by a χ^2 goodness-of-fit test. In every case, the PP model was rejected by the χ^2 test. Thus, it seems appropriate to estimate divergence times for these taxa and these loci only with the SP model. Under the SP model, the age of the higher metazoan radiation is estimated to be between 680 and 1,532 MYA, depending on the true phylogeny of these taxa. For the best fitting tree, this radiation is estimated to be 1,407 MYA, with a confidence interval ranging from 1,130 MYA to 1,752 MYA. Simulation results suggest that this confidence interval probably has little more than an 80% chance of including the true date under the assumption that the stationary model is correct.

In simulation and in these data sets, the primary distinction between the PP model and the SP model occurs in confidence interval estimation. Large cumulative variance per term creates wide confidence intervals. With few interior branch fossils and indices of dispersion larger than 10, confidence intervals in simulation and in both data sets are often many times as large as the point estimates. The only hope for improving on confidence intervals is to add additional loci and/or additional interior branch fossils. Moreover, this work strongly suggests that one should be highly skeptical of any confidence interval estimated using the constant-rate Poisson model without a simultaneous demonstration that this model fits the data well.

The approach to parameter estimation and model selection taken here is a typical likelihood approach (Edwards 1992). The model and parameters that imply the highest likelihood for the data are favored. Therefore, this study chooses tree 1 to be the favored metazoan tree and 1,407 MYA to be the age of the root, since this tree and parameter value imply a significantly higher likelihood for the data than any other combination. Nevertheless, this should not be construed as disproof of a Cambrian origin of the animal phyla.

For all seven trees, the Cambrian explosion hypothesis (a metazoan radiation no more than 600 MYA) led to a significantly lower likelihood for the data than did a much older divergence. Nevertheless, for all seven trees, the Cambrian hypothesis fit well. The hypothesis could never be rejected by a χ^2 test. The question is, why does an older divergence fit better?

From equation (4), it is clear that any model with sufficiently high variance will not be rejected by a χ^2 test. In effect, one can increase the variance until the model cannot be rejected by a goodness-of-fit test. When comparing the Cambrian explosion hypothesis with the unconstrained hypothesis, several features are immediately apparent: (1) the age of the root of the un-

constrained hypothesis is much older, (2) the likelihood of the unconstrained hypothesis is much higher, and (3) the cumulative variance per term of the unconstrained hypothesis is much lower. For instance, for the five mitochondrial loci, tree 1 produced indices of dispersion of 27, 11, 19, 12, and 16 for the unconstrained hypothesis and 92, 33, 28, 37, and 39 for the Cambrian explosion hypothesis. This pattern is reproduced for all five mitochondrial loci (the index of dispersion could not be reliably estimated from the slow-evolving nuclear loci) for all seven trees. Nevertheless, a priori there is no more reason to suspect that the index of dispersion for these loci ought to be around 15 than there is to suspect that it ought to be around 40. No Bayesian prior has even been suggested for the index of dispersion (other than the neutral prediction, which is rejected when the PP model is rejected).

This study has modeled the overdispersed clock by assuming that the substitution process is a stationary process. As argued in the introduction, given that the tree of life happened only once, it may be fundamentally unknowable whether the description of the world taken here, or some different model with, say, a Poisson-distributed number of substitutions along each branch, but with a mean that changed over time, is correct. Seen through the lenses of this world view, the much greater divergence times for the unconstrained trees and the increase in the index of dispersion associated with the Cambrian explosion hypothesis taken together imply that there must have been a tremendous increase in evolutionary rates at the time of the metazoan radiation. Interestingly, the land plant data set shows the same pattern. Either there is a 160-Myr evolutionary history missing from the fossil record, or there was an increase of evolutionary rates associated with the radiation of this group (as observed by the lower index of dispersion for the unconstrained hypothesis).

Thus, one is left with two hypotheses, neither one of which can be rejected with these data sets. The first hypothesis is that the higher metazoans are an ancient lineage which originated hundreds of millions of years before the Cambrian and left few fossils prior to the Cambrian. The second hypothesis postulates a Cambrian, or near Cambrian, radiation of this group, accompanied by a tremendous increase in evolutionary rates.

The only hope of distinguishing these hypotheses probably does not lie in further analysis of these taxa. Instead, one must develop a broader and more general understanding of the cumulative variance per term, or, viewed from the other side of the dichotomy, a more general understanding of the variation in evolutionary rates. If this understanding can be developed, it will only occur through some sort of replication. The tree of life may have evolved only once, but within the next few years, thousands of loci in many species will be sequenced. If these loci are, in any sense, independent replicates of a common evolutionary pattern, it may be possible to develop a proper Bayesian prior on the index of dispersion/variation in evolutionary rates. With such a prior, it may be clear that an index of dispersion of 40 is out of the question, or perhaps it will turn out to be

perfectly common. If these sequences cover many instances of radiation of multiple clades, it may become clear that a sharp increase in the index of dispersion/evolutionary rates always accompanies major clade radiations, or it may become clear that this is an unusual occurrence. Only with secure answers to these questions can molecular clocks provide more certain estimates of divergence times.

Acknowledgments

This work benefited greatly from suggestions by S. Sawyer, J. Gillespie, M. Sanderson, C. Langley, M. Turilli, and an anonymous reviewer.

APPENDIX

A single interior branch fossil can be both necessary and sufficient to bound from above all divergence times. Consider the simplest possible phylogeny under the PP model containing only a single interior fossil calibration point (fig. 11). Assume that a single locus is examined, and $a, b, c,$ and d are the numbers of substitutions along each of the four branches in the phylogeny. The PP model has three unknown parameters: $T_1, T_2,$ and μ . Let $F \neq 0$ be the observed age of a single fossil. Scale T_1 and T_2 by the age of the fossil so that $t_1 = T_1/F$ and $t_2 = T_2/F$. If L is the likelihood of the observations under the PP model,

$$\log(L) = -\mu F(2t_1 + t_2) + S \log(\mu) + (a + b)\log(t_2) + c \log(t_1 - t_2) + d \log(t_1) + C, \tag{6}$$

where $S = a + b + c + d$ and C is a constant. Differentiating equation (6) with respect to t_1 and μ , setting the result to 0, and solving for t_1 and t_2 yields

$$t_1 = \frac{\phi}{\mu F} \tag{7}$$

$$t_2 = \frac{S - 2\phi}{\mu F}, \tag{8}$$

where

$$\phi = \frac{2S + c + 3d \pm \sqrt{(2S - 3d)^2 + c(c + 4S + 6d)}}{12}. \tag{9}$$

The fossil F is an upper bound on time T_2 and a lower bound on T_1 . Therefore, $t_1 \geq 1$ and $t_2 \leq 1$. The second condition implies that $\hat{\mu}$ (the maximum-likelihood estimator of μ) $\geq (S - 2\phi)/F$. This implies that the maximum-likelihood estimator of the root divergence time is less than or equal to $\phi F/(S - 2\phi)$. Thus, a single interior branch fossil, F , creates an upper bound on the maximum-likelihood estimator of all divergence times.

Notice that if fossils are found at either f_1 or f_2 (fig. 11), no upper bounds are created. A fossil at f_1 implies that both t_1 and t_2 are greater than 1, and as a result, $\hat{\mu}$ is bounded only by 0; consequently, neither T_1 nor T_2 have any upper bound. Similarly, a fossil at f_2 implies

that $t_1 > 1$, but has no implications for t_2 , such that maximum-likelihood estimators of neither T_1 nor T_2 are bounded above.

Finally, it should be noted that in the simple situation shown in figure 11, the maximum-likelihood estimators of divergence times need not be unique. In fact, all positive real values of T_1 and T_2 such that

$$F \leq T_1 \leq \frac{\phi F}{S - 2\phi} \quad \frac{F(S - 2\phi)}{\phi} \leq T_2 \leq F$$

are maximum-likelihood estimators of T_1 and T_2 . For this reason, it is clearly possible for estimated divergence times not to be unique, but to instead represent one of an infinite number of equally likely divergence times within a bounded interval. Of course, increasing the number of fossil constraints can restrict the range of these divergence times and can potentially result in a single maximum. In general, for a complicated phylogeny with multiple fossils, determining whether or not a divergence time is unique appears to be a difficult problem.

LITERATURE CITED

- BENTON, M. J. 1993. *The fossil record 2*. Chapman and Hall, London.
- BICKEL, D. R., and B. J. WEST. 1998. Molecular evolution modeled as a fractal Poisson process in agreement with mammalian sequence comparisons. *Mol. Biol. Evol.* **15**:967–977.
- COX, D. R., and V. ISHAM. 1980. *Point processes*. Chapman and Hall, London.
- DOYLE, J. A., and M. J. DONOGHUE. 1993. Phylogenies and angiosperm diversification. *Paleobiology* **19**:141–167.
- EDWARDS, A. W. F. 1992. *Likelihood*. Expanded edition. Johns Hopkins University Press, Baltimore, Md.
- FITCH, W. M., and E. MARGOLISH. 1967. Construction of phylogenetic trees. *Science* **155**:279–284.
- GILLESPIE, J. H. 1991. *The causes of molecular evolution*. Oxford University Press, New York.
- . 1993. Substitution processes in molecular evolution. I. Uniform and clustered substitutions in a haploid model. *Genetics* **134**:971–981.
- . 1994a. Substitution processes in molecular evolution. II. Exchangeable models from population genetics. *Evolution* **48**:1101–1113.
- . 1994b. Substitution processes in molecular evolution. III. Deleterious alleles. *Genetics* **138**:943–952.
- GILLESPIE, J. H., and C. H. LANGLEY. 1979. Are evolutionary rates really variable? *J. Mol. Evol.* **13**:27–34.
- GU, X. 1998. Early metazoan divergence was about 830 million years ago. *J. Mol. Evol.* **47**:369–371.
- HASEGAWA, M., H. KISHINO, and T. YANO. 1989. Estimation of branching dates among primates by molecular clocks of nuclear DNA which slowed down in hominoidea. *J. Hum. Evol.* **18**:461–476.
- HUDSON, R. R. 1983a. Properties of a neutral allele model with intragenic recombination. *Theor. Popul. Biol.* **23**:183–201.
- . 1983b. Testing the constant rate neutral model with protein sequence data. *Evolution* **37**:711–719.
- KIMURA, M. 1983. *The neutral allele theory of molecular evolution*. Cambridge University Press, Cambridge, England.
- KIMURA, M., and T. OHTA. 1971. Protein polymorphism as a phase of molecular evolution. *Nature* **229**:467–469.
- KUMAR, S., and S. B. HEDGES. 1998. A molecular timescale for vertebrate evolution. *Nature* **392**:917–920.
- LANGLEY, C. H., and W. M. FITCH. 1973. The constancy of evolution: a statistical analysis of the α and β haemoglobins, cytochrome c, and fibrinopeptide A. Pp. 246–262 in N. E. MORTON, ed. *Genetic structure of populations*. University of Hawaii Press, Honolulu.
- . 1974. An estimation of the constancy of the rate of molecular evolution. *J. Mol. Evol.* **3**:161–177.
- LI, W.-H. 1997. *Molecular evolution*. Sinauer, Sunderland, Mass.
- LI, W.-H., and D. GRAUR. 1991. *Fundamentals of molecular evolution*. Sinauer, Sunderland, Mass.
- LYNCH, M., and P. E. JARRELL. 1993. A method for calibrating molecular clocks and its application to animal mitochondrial DNA. *Genetics* **135**:1197–1208.
- MCHUGH, D., and K. HALANYCH. 1998. Introduction to the symposium: evolutionary relationships of metazoan phyla: advances, problems and approaches. *Am. Zool.* **38**:813–817.
- MUSE, S. V., and B. S. WEIR. 1992. Testing for equality of evolutionary rates. *Genetics* **132**:269–276.
- OHTA, T. 1995. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *J. Mol. Evol.* **40**:56–63.
- OHTA, T., and M. KIMURA. 1971. On the constancy of the evolutionary rate of cistrons. *J. Mol. Evol.* **1**:18–25.
- PRESS, W. H., S. A. TEUKOLSKY, W. T. VETTERLING, and B. P. FLANNERY. 1992. *Numerical recipes in C: the art of scientific computing*. 2nd edition. Cambridge University Press, Cambridge, England.
- SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
- SANDERSON, M. J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* **12**:1218–1231.
- SPRINGER, M. S. 1995. Molecular clocks and the incompleteness of the fossil record. *J. Mol. Evol.* **41**:531–538.
- STUART, A., and J. K. ORD. 1987. *Kendall's advanced theory of statistics, Vol. 1. Distribution theory*. 5th edition. Oxford University Press, New York.
- TAKAHATA, N. 1987. On the overdispersed molecular clock. *Genetics* **116**:169–179.
- TAKEZAKI, N., A. RZHETSKY, and M. NEI. 1995. Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* **12**:823–833.
- THORNE, J. L., H. KISHINO, and I. S. PAINTER. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* **15**:1647–1657.
- UYENOYAMA, M. K. 1995. A generalized least-squares estimate for the origin of sporophytic self-incompatibility. *Genetics* **139**:975–992.
- WRAY, G. A., J. S. LEVINTON, and L. H. SHAPRIO. 1996. Molecular evidence for deep precambrian divergences among metazoan phyla. *Science* **274**:568–573.
- ZUCKERKANDL, E., and L. PAULING. 1965. Evolutionary divergence and convergence in proteins. Pp. 97–166 in V. BRYSON and H. J. VOGEL, eds. *Evolving genes and proteins*. Academic Press, New York.

STANLEY SAWYER, reviewing editor

Accepted July 17, 2000