Int. J. Plant Sci. 165(4 Suppl.):S7–S21. 2004.

1058-5893/2004/1650S4-0002$15.00

DATING LINEAGES: MOLECULAR AND PALEONTOLOGICAL APPROACHES TO THE TEMPORAL FRAMEWORK OF CLADES

Susana A. Magallón[1]

Instituto de Biología, Universidad Nacional Autónoma de México, Circuito Exterior, Anexo al Jardín Botánico, A.P. 70-233, México D.F. 04510, México

The recent proliferation of methodological advances in molecular phylogenetic and paleobiological research has resulted in powerful approaches to investigate the temporal framework of lineages. This article is a review of molecular and paleontological methods to estimate ages of clades. Inferring ages of clades is complicated by the nature of the process of molecular substitution and the uncertainties of the paleontological record. Some of the greatest problems associated with molecular methods include the stochastic nature of molecular substitution, the assumption of rate constancy among lineages when such constancy is absent, and the inextricable link between substitution rate and elapsed time on the branches of phylogeny. Molecular methods to estimate ages are ultimately based on the fact that as time elapses, molecular differences accumulate among sequences. Under rate constancy, methods to estimate ages include linear regression of molecular distance on elapsed time and maximum likelihood optimization of node ages under a single rate. Recently developed methods that allow rate heterogeneity are powerful approaches to estimate rates and divergence times under more realistic assumptions. Among-lineage rate variation is introduced as a compound Poisson process or more frequently is guided by the principle of temporal rate autocorrelation. These methods are based on numerical, semiparametric, and Bayesian parametric approaches, and some allow incorporation of constraints on the ages of nodes (derived, e.g., from fossils), conferring additional realism to age estimates. The paleontological record provides times of first appearances of morphological traits but not of lineage divergences; nevertheless, it represents one of the few sources of absolute information to decouple rates and times in a phylogeny. Analytical methods applied to paleontological data provide an alternative source of information about lineage duration. Stratigraphic confidence intervals that contain the time of origin of a lineage under a known probability are based on the frequency and abundance of fossil finds through the lineage's fossil record. Tests of postulated lineage durations, derived, for example, from a molecular age estimate, are available under probabilistic or likelihood frameworks. A powerful approach toward achieving more robust inferences about evolutionary rates and timing of lineage divergence lies in the complementary use of molecular- and paleontological-based approaches. While incorporating fossil information as age constraints confers further realism to molecular-estimated rates and ages, such estimates may be evaluated against expectations derived from paleontological information.

Keywords: ages, Bayesian inference, branch lengths, calibration, constraints, confidence intervals, fossil record, lineage duration, maximum likelihood, minimum ages, molecular clock, nonparametric rate smoothing, penalized likelihood, preservation rate, Poisson process, substitution rate.

Introduction

The study of geographical disjunctions is inextricably linked with lineage divergence. A natural starting point to investigate the causes and mechanisms underlying disjunct distributions of closely related taxa is to obtain accurate information about the time when an ancestral lineage split to give rise to a pair of descendant lineages. The literature on the topic, from both empirical and methodological standpoints, is substantial and currently expanding (Hillis et al. 1996; Ayala et al. 1998; Sanderson 1998, 2002; Huelsenbeck et al. 2000; Heckman et al. 2001; Rodríguez-Trelles et al. 2001; Aris-Brosou and Yang 2002; Thorne and Kishino 2002; Benton and Ayala 2003; Yang and Yoder 2003). The main goal of this article is to summarize and review some of the currently available approaches to evaluate the temporal framework of lineages on the basis of molecular data and paleontological information. Because these approaches are based not only on different data but also on contrasting conceptual assumptions and empirical frameworks, together the approaches offer a system in which estimates derived from one set of methods can be tested through inferences derived from alternative methodologies.

Molecular estimates of time of lineage divergence are ultimately based on the observation that as time elapses, genetic differences between a pair of sequences accumulate. The hypothesis of the molecular clock (Zuckerkandl and Pauling 1962) postulates that the amount of molecular difference between a pair of protein sequences is approximately

[1] E-mail s.magallon@ibiologia.unam.mx.

Manuscript received July 2003; revised manuscript received January 2004.

proportional to the time elapsed since the divergence from their common ancestor. Although the concept of the molecular clock provides an explicit framework for estimating divergence times from molecular distances, many factors and methodological biases are known to disrupt the expected correlation between molecular distance and elapsed time. These factors and biases have been described and discussed extensively (Hillis et al. 1996; Ayala 1997; Sanderson 1998; Sanderson and Doyle 2001; many others). While the effect of some of these factors is evident, several others are cryptic, and many are pervasive.

The effects of several methodological biases in the estimation of times and rates have been empirically documented. The significant influence of taxonomic sampling on phylogeny estimation is well known (Felsenstein 1978; Graybeal 1998; Poe 1998), and it also has been shown to have a determinant influence in estimating rates and ages of clades (Sanderson and Doyle 2001). While the effect of different types of data in phylogeny estimation is significant (Lewis et al. 1997; Simmons and Freudenstein 2002), its effect in estimating rates and divergence times seldom has been evaluated. Empirical estimates show that rates and ages derived from different genes and gene partitions are substantially different (Magallón and Sanderson 2003; Yang and Yoder 2003; S. Magallón and M. J. Sanderson, unpublished manuscript). Other pervasive sources of error include an incorrect phylogeny hypothesis and an incorrect temporal calibration of the phylogeny.

Implementing a model of molecular substitution that appropriately describes the available data is vital to obtain accurate estimates of rates and divergence times. The optimal model should account for relevant characteristics of the data but should not include superfluous parameters that reduce its predictive power and yield excessive variance to estimated parameters (Zharkikh 1994; Lewis 1998; Sanderson 1998). The fit of a model to the data can be evaluated through hierarchical likelihood ratio tests (Page and Holmes 1998) applied to sequentially more complex nested models (e.g., Modeltest; Posada and Crandall 1998). The simultaneous use of multiple molecular markers in phylogeny reconstruction and estimation of rates and divergence times is now ubiquitous. Exploratory analyses indicate that different genes and gene partitions are explained significantly better by independent models and parameters than by a single model optimized on a combined data set (Sanderson and Doyle 2001; Pupko et al. 2002). While a possible solution may lie in performing independent analyses for separate partitions under optimal model selection and parameterization, an alternative view holds that accounting for the differences among the data in a combined data set yields more accurate estimates of rates and divergence times (Yang and Yoder 2003).

An additional level of complication arises from the linkage of substitution rate and temporal duration in the branches of a phylogram. Whereas estimating branch lengths as the product of rate and time is not problematic in phylogeny reconstruction, it represents a significant conflict for estimating rates and times. Decoupling rate and duration in a branch may only be achieved by incorporating independent information about the absolute magnitude of one or both parameters.

The nature of molecular substitution introduces unpredictable biases in estimating lineage divergence times by significantly disrupting the expected correlation between molecular distance and elapsed time. The stochastic periodicity of molecular substitution (Zuckerkandl and Pauling 1965) introduces an inherent error (Hillis et al. 1996), which increases as rates of substitution are higher but diminishes as the amount of data (i.e., number and length of sequences) increases (Sanderson 1998 and references therein). The correlation between molecular distance and elapsed time critically depends on an approximately constant rate of substitution among lineages. However, numerous studies (Langley and Fitch 1974; Goodman et al. 1975; Wu and Li 1985; Gillespie 1991; Li 1997; Muse 2000; many others) strongly suggest that rate heterogeneity is widespread rather than exceptional. Furthermore, lineages cannot be easily characterized as "fast"- or "slow"-evolving because each gene seems to manifest its own tendencies regarding the pace of molecular substitution (Rodríguez-Trelles et al. 2001). Some of the most serious problems in age estimation arise from incorrectly assuming a constant rate of substitution among lineages when constancy is absent (Sanderson 1998). Testing for rate constancy is therefore crucial for selecting an appropriate method for estimating rates and ages.

This article first discusses the incorporation of independent (nonmolecular) chronological information in a phylogenetic tree. Issues pertaining to the nature of the paleontological record and the information provided by fossils are discussed. The next section discusses molecular-based methods to estimate rates and time, which build on an explicit phylogeny. These methods differ in the degree to which they allow among-lineage rate heterogeneity and in the way in which rate heterogeneity is introduced. The final section describes analytical methods based on paleontological data to estimate temporal duration of lineages.

Introducing Chronological Landmarks into a Phylogenetic Tree

Chronological information derived from nonmolecular data provides an absolute framework to decouple substitution rate and temporal duration in the branches of a phylogenetic tree. Temporal landmarks can be incorporated as a calibration point, which establishes a tree-wide temporal reference against which relative divergence times of nodes can be scaled, or as constraints to the ages of nodes, which aid the estimation of divergence times.

Calibration consists of assigning an absolute age to a node in a phylogenetic tree in order to convert relative divergence times into absolute time units (e.g., millions of years). Because the timescale derived from calibration has pervasive effects on absolute age estimates for all nodes, an adequate selection of a calibration node is critical. An appropriate calibration node should depict a well-supported (presumably correct) phylogenetic relationship. Equally important, the age assigned to a calibration node should be such as to have a very high probability of being very close to the true time of lineage splitting (see below).

Nodes may be constrained with a minimum or maximum age. Constraints facilitate estimation of times by imposing upper and/or lower bounds on the possible ages for nodes. Constraints also help to decouple rates and times, especially

in absence of rate constancy, and may provide a rough assessment of rate heterogeneity in the phylogeny. A candidate constraint node should depict a correct phylogenetic relationship. Most frequently, constraints are introduced as minimum ages, indicating that the phylogenetic split had occurred at least by a given time. Nodes may be constrained with a maximum age (Sanderson and Doyle 2001) when the constraint age has a high probability of being very close to the true age of the node.

Sources of Independent Chronological Information

The fossil record. By far the most common source of nonmolecular information of ages of clades is the fossil record. To use a fossil to date a node in a phylogeny, its placement in the phylogeny first should be unequivocally identified, and then the fossil itself should be dated accurately, either through reliable stratigraphic correlations or radiometric dating. When assigning a date to a calibration node, an additional condition is that the age of the fossil is close to the true age of the clade. Some types of fossils have a higher probability of fulfilling these three requirements; for example, organs or body parts that are structurally resistant and abundantly produced and that bear unambiguous indication of their membership to a particular clade. The first two attributes provide a higher probability of a small time lapse between the origination of the morphological traits and their appearance in the fossil record (but see below). The third attribute allows reliable recognition of clade membership.

The fossil record documents the first appearance of morphologies in stratigraphic sequences (Foote et al. 1999). For a morphological trait that characterizes a clade (i.e., a synapomorphy) to appear in the fossil record, it was probably abundantly produced. Thus, fossil first appearances most likely document the time when a structure became abundant rather that its time of origin. This observation raises the questions of how close in time is the first fossil occurrence of a distinctive morphological trait (i.e., a synapomorphy) to its origin, and how close is the origin of that morphological trait to the time of lineage divergence (fig. 1). Even with an excellent fossil record, the time lapse between lineage divergence and first fossil occurrence is unknown. The fossil record and molecular dating methods inherently measure different events. The first documents first appearances of morphologies. The second (possibly) dates splits of molecular lineages. The extent to which the two events are temporally coupled is unknown. Fossils therefore can provide only minimum ages for molecular lineages. The best fossils (e.g., decay-resistant structures with unambiguous synapomorphies documenting clade membership) provide a high probability of a small time lapse between becoming abundant and first fossil occurrence. However, the question of the time elapsed between lineage divergence and the origin and abundance of the synapomorphy still remains.

Nevertheless, if critically interpreted, the fossil record still represents the best available source of independent chronological information for ages of lineages. As decisively stated by Thorne and Kishino (2002), fossil information is crucial because not even an infinite amount of molecular data can surmount uncertainties introduced by fossils to obtain exact

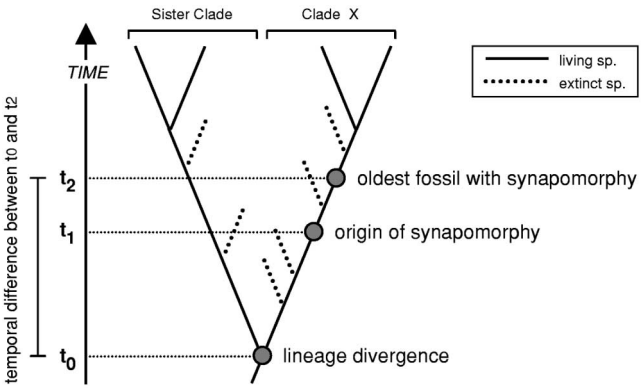

Fig. 1 Timing of lineage divergence, origin of a synapomorphy, and occurrence of the oldest fossil with the synapomorphy. A temporal gap of unknown magnitude exists between the divergence of a lineage and its sister from their common ancestor (t_0), the origin of a synapomorphy (t_1), and the occurrence of the oldest fossil bearing the synapomorphy. Molecular dating methods provide evidence of the first event (t_0). The fossil record provides evidence of the third (t_2).

estimation of divergence times. The fossil record provides information about minimum ages of clades, which allows us to decouple rate and time in the branches of a phylogeny, that would otherwise be unavailable.

To introduce ages derived from the fossil record into a tree, the correct phylogenetic placement of fossils should be identified. Given the current extensive availability of plant molecular data sets and the scarceness of comparable morphological data sets needed to include fossils directly in phylogenetic analyses, it frequently becomes necessary to identify the phylogenetic placement of fossils on the basis of their morphological characters and their distribution on an already available tree. Any fossil taxon is either an extinct member of a crown group or a member of the stem group (fig. 2*A*). While the conceptual distinction is unambiguous, determining in practice the exact relationship between a fossil and its living relatives is not straightforward. Useful criteria rely on the combination of morphological characters of the fossil and the distribution of synapomorphies in the phylogeny (fig. 2*B*–2*E*).

Because fossils display only morphological characters, phylogenetic analyses including fossils should be based at least partially on this type of data and be conducted under the parsimony criterion. The compilation of descriptive, critically evaluated, and widely documented morphological data sets for living and fossil taxa is therefore crucial. There is an urgent need for development and explicit testing of alternative methodologies to scoring missing fossil molecular characters as a question mark. Parsimony is prone to be misguided under particular circumstances, for example, reconstructing relationships among clades separated by long evolutionary distances (Felsenstein 1978). Under these circumstances it might be preferable to conduct phylogenetic analysis using parametric optimization criteria (i.e., maximum likelihood or Bayesian inference), which usually rely on molecular data. It is paradoxical that the questions that would most greatly benefit from the inclusion of fossils represent the greatest

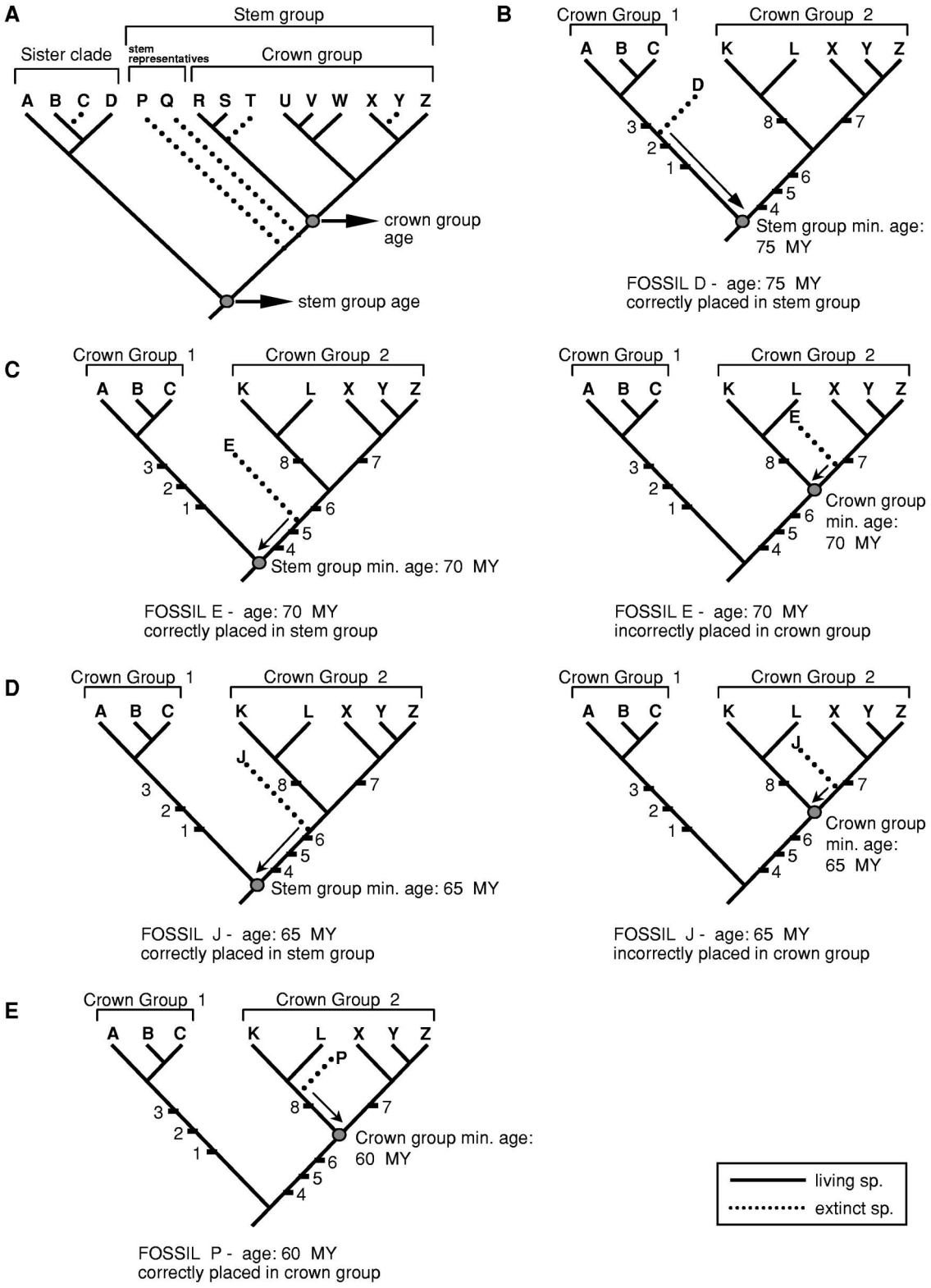

Fig. 2 Recognizing a fossil's crown or stem group membership. *A*, The crown group is the least inclusive monophyletic group that contains all the living members of a clade plus all extinct species that diverged after the oldest split that gave rise to two living species (e.g., *T* and *Y*). The stem group contains the crown group plus all the extinct species that diverged from the lineage leading to the crown group after it diverged from its extant sister group (e.g., *P* and *Q*). These definitions differ from those of Jefferies (1979) and de Queiroz and Gauthier (1990). *B*, A fossil that displays some but not all the synapomorphies of the crown group can be unambiguously identified as a stem representative. Fossil *D* lacks a crown

methodological challenges to their inclusion. It becomes increasingly necessary to design and test methodologies to reliably identify the phylogenetic placement of fossils. These may include, for example, the use of constrained topologies derived from phylogenetic parametric analyses of living taxa on parsimony analyses of living and fossil taxa based (partially or completely) on morphological characters.

Geological information. Ages of geological events may also be used to assign ages to nodes. If each member of a sister pair is distributed in disjunct fragments of a formerly continuous land mass and land mass separation was the immediate cause of lineage splitting, then the age of land separation can be assigned to the node from which the sister lineages diverge (fig. 3*A*). Minimal conditions to be met include uncontroversial evidence that lineage splitting occurred immediately after the geological event took place and that an unequivocal date can be assigned to the geological event itself. In all cases, the necessary evidence may be difficult to obtain. The split of an original lineage into a pair of sister groups as a direct and immediate consequence of a geological event constitutes a hypothesis to be tested, not merely assumed. Lineage splitting may have taken place before or after the geological event (fig. 3*B*, 3*C*). The uncertainty between the age of a geological event and the time of lineage splitting may be in two directions; the former may represent either an underestimate (fig. 3*B*) or an overestimate (fig. 3*C*) of the true age of the latter. Comparatively, the fossil record offers the advantage that, as long as clade membership is correctly identified, associated errors can only represent underestimates of the true age of a clade.

Molecular-based age estimates. Ages derived from independent molecular dating inferences have sometimes been used to date calibration nodes (Martin et al. 1989, 1993; Heckman et al. 2001). Ages estimated through inferences based on molecular data are hypotheses of the true age of nodes. How close these estimates are to true ages depends at least on the adequacy of the available molecular data, on the implemented model of molecular evolution, and on the method used to infer the ages. As extensively discussed elsewhere (Hillis et al. 1996; Sanderson 1998; Sanderson and Doyle 2001), numerous factors (e.g., stochasticity of molecular substitution, rate heterogeneity, model adequacy, estimated phylogeny, estimated molecular distances, accurate calibration, and many others) can severely bias age estimation. Thus, using an independent molecular-based age to date a calibration node is strongly discouraged and should be implemented only when absolutely no other options are available.

Estimating Divergence Times and Rates from Molecular Data

The process of molecular substitution implies that as time elapses, the molecular difference between a pair of sequences that diverged from a common ancestor will increase. While this principle provides a general framework to estimate elapsed time since lineage splitting, exploratory studies have identified important sources of error in estimating time from molecular distance (Ayala et al. 1998; Yoder and Yang 2000; Rodríguez-Trelles et al. 2001; Sanderson and Doyle 2001). The linkage of evolutionary rate and temporal duration in a phylogeny introduces an additional level of complexity (Sanderson 1998; Thorne and Kishino 2002). The following section is a summary of methods to estimate time of lineage divergence and rates of molecular evolution from molecular data. While detailed description of the functions and algorithms involved in each of these methods is beyond the scope of this summary, some recently formulated approaches that relax the restrictive assumption of rate constancy are emphasized.

Estimating Rate and Divergence Times under Rate Constancy

The hypothesis of the molecular clock proposes that rate of molecular change of genes and proteins is sufficiently constant as to reflect elapsed time in a linear fashion (Zuckerkandl and Pauling 1962). Under the molecular clock, phylogenetic trees are ultrametric (i.e., all the tips are equidistant from the root), and the genetic distance between sequences, corrected for saturated sites, is proportional to the elapsed time since their divergence. Molecular substitution was proposed to behave approximately as a stochastic Poisson process (Zuckerkandl and Pauling 1965), and while the Poisson behavior of molecular substitution has been contested (Langley and Fitch 1973, 1974), it implies that rate of molecular substitution does not occur with exact precision but rather, with probabilistic regularity.

Linear regression. Under rate constancy, the molecular distance between each member of a sister pair and their most recent common ancestor is one-half of the distance between the two sequences (D). If divergence time between a pair of sequences (species) is known (e.g., from the fossil record), this absolute age provides a time frame for a molecular distance of magnitude $1/2D$. If divergence times among several pairs of sequences are available, the rate of substitution can be estimated by regressing molecular distances on divergence

group synapomorphy, and its age is correctly used to date the stem group. *C*, Fossils (especially plant fossils) usually represent a body part of an extinct organism. Although the fossilized organ may display all the crown group synapomorphies of a given clade, it is possible that body parts that did not enter the fossil record lacked crown group synapomorphies, and thus the extinct organism is a stem representative. Fossil E represents a body part of an organism that displays the crown group synapomorphies corresponding to that body part. Nevertheless, the extinct organism lacked some crown group synapomorphies. Fossil E provides an age for the stem group. Erroneously placing fossil E in the crown group will lead to an overestimation of the crown group age. *D*, Fossil J represents an extinct organism with all the synapomorphies of crown group 2 but lacks synapomorphies of any of the subclades within crown group 2. Fossil J is the extinct sister of the crown group, and its age correctly indicates the stem group's age. Erroneously placing fossil J within the crown group will lead to an overestimation of the crown group age. *E*, A reliable date for a crown group can be obtained from the oldest unequivocal fossil member of a subclade within the crown group. Fossil P displays the synapomorphies of crown group 2 and also a synapomorphy of one subclade within crown group 2. Fossil P can be reliably used to date crown group 2.

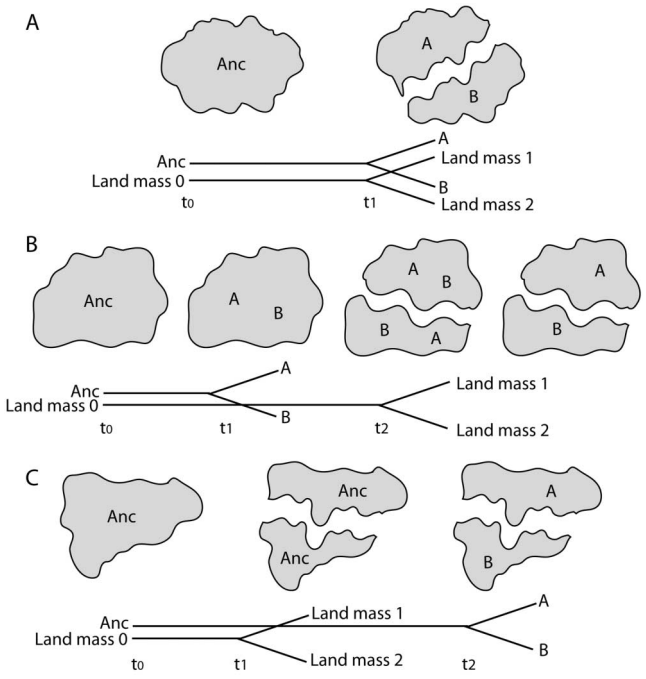

Fig. 3 Timing of land mass and lineage splitting. *A*, Land mass separation and lineage divergence occur simultaneously. The time of land mass separation correctly depicts the time of lineage divergence (t_1). *B*, Lineage divergence precedes land mass separation. Lineage divergence occurs at t_1, whereas land mass separation occurs at t_2. Each of the daughter lineages becomes extinct in one of the land masses, resulting in a disjunct distribution of sister species. Assigning the time of land mass separation (t_2) to lineage splitting (t_1) is an underestimation of the age of lineage splitting. *C*, Land mass separation precedes lineage divergence. Land mass separation occurs at t_1. The ancestral species survives in each of the new land masses for some time, ultimately evolving into a new species in each land mass at t_2. Assigning the time of land mass separation (t_1) to lineage splitting (t_2) is an overestimation of the age of lineage splitting.

time. At the time of divergence (t_0), the molecular distance between two sequences is 0, and therefore the regression line should be forced through the origin. Because the underlying Poisson process introduces greater variance in molecular distance as time increases, a weighted linear regression is more appropriate. The scatter of data points around the regression line provides a confidence interval around estimated ages.

Conditions to be fulfilled include a constant rate of substitution across all sites and lineages, availability of a correct phylogenetic hypothesis, and absence of error in estimated branch lengths, in age of calibration points, and in the regression of molecular change on time (Hillis et al. 1996). Even supposing these conditions were fulfilled, uncertainties introduced by the stochastic substitution process may result in broad confidence intervals, especially at high rates of molecular substitution and at high levels of molecular change. While including a representative number of data points in estimating the regression is desirable, nonindependence among points may be introduced by phylogenetic structure (Hillis et al. 1996).

Mean path lengths. The method of mean path lengths estimates rate and divergence times based on the mean branch length between a node and each of its terminals (Bremer and Gustafsson 1997; Britton et al. 2002). The required data are a tree with known branch lengths, expressed as number of substitutions between nodes corrected for multiple hits and a calibration node. The number of substitutions between the calibration node and each of its terminals is obtained, and the mean number of substitutions, that is, the mean path length (MPL) between a node and its terminals, is calculated. By dividing the MPL by the known age of the node, the substitution rate is obtained. Because rate constancy is assumed, this rate is extrapolated to the entire tree. To estimate the age of a node, its MPL is divided by the estimated rate. Rate constancy may be tested by comparing the MPLs of each of the two subclades derived from a given node; if significantly different, rate constancy is rejected, and different rates may be estimated on different parts of the tree (Britton et al. 2002).

Maximum likelihood clock optimizations. Divergence times can be estimated via maximum likelihood by optimizing the single constant rate of substitution that best fits the entire phylogeny. Divergence times, together with all remaining parameters, are then estimated according to the constant rate of substitution. A maximum likelihood method for estimating divergence times under rate constancy was proposed by Langley and Fitch (1974). Starting with a phylogeny for which branch lengths are available, a constant rate of substitution is optimized through maximum likelihood. Using this constant-rate, branch lengths and ages of nodes are estimated. Rate constancy is tested by comparing branch lengths estimated under the single rate with observed branch lengths through a χ^2 test. If differences are nonsignificant, ages of nodes estimated under the single rate can be accepted.

Accommodating the Molecular Clock When Rate Heterogeneity Is Recognized

While the existence of a universal molecular clock is debated, several authors have proposed locally constant rates that characterize functionally related genes or closely related lineages. Under these conditions, methods that estimate rates and divergence times using the molecular clock while correcting for observed global rate heterogeneity have been implemented.

Linearized trees. This method begins by implementing tests to identify branches in a phylogeny that depart significantly from rate constancy, which are then excluded. The remaining constant-rate branches constitute a "linearized tree" because time and genetic distance are linearly related. Rate and divergence times are estimated on the linearized tree using methods that assume rate constancy. To obtain a truly clocklike tree, several passes of rate constancy tests and elimination of branches should be implemented until significant differences in rates are no longer detected. A major objection to this method is that it relies on eliminating data that do not fit the expected constant-rate behavior. If rate heterogeneity were widespread, this approach would lead to a massive elimination of data (e.g., Aris-Brosou and Yang 2002).

Li and Tanimura (1987) compared rates of synonymous and nonsynonymous substitution in globin genes among

primates and other mammals. By using relative rates tests, they identified slower rates in the human lineage, which was then excluded from the tree. Divergence times for the remaining species and genes were then calculated. Takezaki et al. (1995) implemented a statistical method to test rate constancy, to identify and eliminate sequences that do not fulfill constancy, and to construct a tree with the remaining sequences—a linearized tree. Two tests of rate constancy were proposed: the two-cluster test evaluates constancy of substitution rates for two sister clades with respect to one or several outgroups, and the branch-length test examines deviation of a particular root-to-tip distance versus the average root-to-tip distance for the tree (except for the outgroup). After rate-divergent sequences have been excluded, ages are estimated under rate constancy on the linearized tree.

Model selection methods. Model selection relies on the assumption that closely related lineages or functionally similar genes are characterized by similar rates of substitution. Regions of the tree with different rates are identified through explicit tests (Yoder and Yang 2000), or different rates are assigned on the basis of available data (Hasegawa et al. 1989) or independent information, such as generation time or gene function (Uyenoyama 1995). By applying models that best fit each region of the tree, rates and other parameters are optimized. Divergence times are estimated assuming rate constancy on each part of the tree characterized by a common model. This method can be viewed in effect as applying two or more molecular clocks on a phylogeny. One substantial difficulty is to identify correctly branches or regions of a tree in which rates of substitution change significantly. The use of biological (e.g., similar life-form, generation time, metabolic rate) or functional (e.g., gene function) information may aid their recognition. Additional obstacles arise from the increasing complexity of tests for adequacy of models and implementation of different rate parameters on several parts of a tree. Nevertheless, beginning with Hasegawa et al. (1989), several studies have used model selection to estimate substitution rates and divergence times.

Hasegawa et al. (1989) used transition and transversion rates to detect rate heterogeneity in a primate phylogeny. After identifying a clade characterized by a slower rate, rate parameters and divergence times were optimized via maximum likelihood on two regions of the tree. Subsequently, an additional slowdown was detected within the "slow" clade. Three rates were assigned to different parts of the tree and used to estimate divergence times via maximum likelihood. Using a similar approach, Kishino and Hasegawa (1990) estimated divergence dates for hominids by applying models with different rates of transitions and transversions to different lineages.

In a thorough application of model selection, Uyenoyama (1995) investigated the timing of change of function of alleles to produce sporophytic self-incompatibility in *Brassica* and relatives. The known differential function of alleles guided the identification of sites of rate change in the phylogeny. Estimates and standard errors of rates and divergence times were obtained through a generalized least squares (GLS) procedure. Rate constancy within each of the assigned function-specific classes was also tested with a multivariate relative rates test that involved comparing individual with mean root-to-tip distances, removing significantly different sequences and repeating the test with the remaining sequences.

Yoder and Yang (2000) implemented local constant-rate models via maximum likelihood to assess divergence times for rodents and primates. Yang and Yoder (2003) expanded their previous model to incorporate multiple calibration points under a global molecular clock and to use multiple genes under local clocks. Mouse lemur species divergence times were estimated using two concatenated mitochondrial genes for a taxonomic set that included 26 additional mammal species. Codon positions were evaluated separately and jointly. Local clocks were implemented by assigning three different rates in the tree on the basis of branch lengths estimated via maximum likelihood in absence of rate constancy. Ages resulting from the simultaneous use of the three codon positions (with their differences accounted for) under three local clocks were very similar to ages estimated with the same data using a Bayesian method (Thorne and Kishino 2002, discussed below) and to fossil ages, thus suggesting a general methodological robustness (Yang and Yoder 2003). The use of two local clocks estimated a much older age for the clade, probably because its long branches were interpreted as long durations, rather than accelerated rates.

A variation of model selection that incorporates likelihood ratio tests was proposed by Rambaut and Bromham (1998). Quartets, consisting of two pairs of sister groups, are delimited in a phylogeny. Rate constancy is tested within each quartet through a likelihood ratio test in which a two-rate model, which assigns a rate to each of the sister pairs, is tested against a multiple-rate model, which allows a different rate on each branch of the quartet. If the multiple-rate model is significantly better, the quartet is excluded from the tree. Unknown divergence times are estimated through maximum likelihood by applying a constant rate to each half of the remaining quartets.

Incorporating Rate Heterogeneity in Estimating Rates and Divergence Times

Studies in which rates among lineages have been assessed provide mounting evidence of the widespread occurrence of rate heterogeneity (Langley and Fitch 1974; Goodman et al. 1975; Wu and Li 1985; Gillespie 1991; Gaut et al. 1992; Rodríguez-Trelles et al. 2001; Sanderson and Doyle 2001; many others). Recently proposed methods that relax rate constancy are promising approaches to estimate rates and divergence times under more realistic approximations of the process of molecular evolution.

Methods that relax rate constancy must necessarily be guided by specifications about how rates are expected to change among lineages. Although rate heterogeneity among lineages has been extensively documented, significantly fewer ideas about how rates change from lineage to lineage have been proposed. Available methods introduce rate heterogeneity on the basis of two very different approaches. One is a compound Poisson process of rate change in which a Poisson process that introduces changes in the rate of substitution in different places in a phylogeny is imposed on the primary Poisson process of molecular substitution (Huelsenbeck et al. 2000).

The other is based on the important concept of temporal autocorrelation in rates (Gillespie 1991). According to temporal autocorrelation, rates of substitution are likely to be similar among closely related lineages but increasingly different as lineages are more distantly related. Reasons why temporal autocorrelation might exist include descendant lineages inheriting the rates of their ancestor, with subsequent independent evolution of new rates (Takahata 1987; Gillespie 1991; Sanderson 1997), or descendant lineages inheriting from their ancestor traits that regulate rates of evolution, such as habit, life-form, metabolic rate, generation time, and/or descendant lineages being subject to similar environmental conditions (Thorne et al. 1998). Temporal autocorrelation is an explicit a priori criterion to guide inference of among-lineage rate change and is implemented in several methods to estimate rates and divergence times.

Parametric methods with a Bayesian implementation. In a Bayesian framework, a priori knowledge about parameter values is introduced by assigning probability distributions known as priors. The observed data, the priors, and a likelihood function are used to obtain posterior probability distributions, which represent uncertainty about the parameters after observing the data (Thorne et al. 1998; Lewis 2001; Aris-Brosou and Yang 2002; Huelsenbeck et al. 2002). Posterior distributions are the basis of Bayesian inference. Methods based on a Bayesian framework to estimate rates and ages differ mainly in the principle guiding rate change among branches.

A compound Poisson process of rate change. A method that relaxes rate constancy and simultaneously allows estimation of the ages of nodes on the basis of the previously described compound Poisson process of rate change was presented by Huelsenbeck et al. (2000). On a fixed tree topology, the number of rate-change events and speciation ages are variables. Branch lengths are defined as the expected number of substitutions per site and are scaled so that tips of the tree are at time t_0 and the root is at t_1. Branch lengths can be scaled to million years to yield time estimates as absolute ages. Rates on the tree are determined by the number of rate-change events, the point in the tree where they occur, and the magnitude of change at each event (Huelsenbeck et al. 2000). The number, position, and magnitude of rate-change events are determined from the posterior probabilities of the expected number of substitutions per site on a branch (m), branch lengths (ν_k), the Poisson parameter that regulates rate change (λ), and the shape parameter of the gamma distribution (α_p) from which a random multiplier for m is drawn. An exponential prior for λ is assigned, to decrease probability of having numerous but small rate-change events. Prior distributions for m, α_p, number of rate-change events, and speciation times are drawn from independent uniform distributions with speciation times conditional on tree topology (Huelsenbeck et al. 2000). Posterior distributions are determined mainly by the likelihood function.

Posterior distributions of parameters are estimated via a Markov Chain Monte Carlo (MCMC), implemented as the Metropolis-Hastings-Green algorithm (Green 1995; in Huelsenbeck et al. 2000). From the starting point of the chain, a new state is nominated. The acceptance probability (R) of the new state is calculated, and a uniformly distributed random variable is generated. If the random variable is smaller than the acceptance probability of the new state, the new state is accepted. In this method (Huelsenbeck et al. 2000), the chain is updated by implementing 10 different types of moves that include, for example, adding or deleting a rate-change event from a tree or changing the time of internal nodes. Posterior distribution of parameters, including divergence times, are obtained directly from the chain after it has performed a large number of generations (excluding burn-in values). The proportion of time the chain remains at a given interval is an approximation to its posterior distribution (Huelsenbeck et al. 2000).

Bayesian implementation of rate autocorrelation. Thorne, Kishino, and collaborators formulated models of the evolution of the rate of substitution that allow estimating divergence times without assuming rate constancy. These methods use temporal autocorrelation to introduce rate change among branches (Thorne et al. 1998; Kishino et al. 2001; Thorne and Kishino 2002). By modeling autocorrelation, rates are not allowed to change too quickly from ancestral to descendant branches.

The original hierarchical model of rate evolution (Thorne et al. 1998) was modified to achieve better parameterization (Kishino et al. 2001) and subsequently extended to include multiple gene data sets (Thorne and Kishino 2002). However, the Bayesian technique to obtain posterior distributions of rates and divergence times through an MCMC remained basically unchanged. Model parameters are rates on branches (R), the rate of one of the branches derived from the root node (R_0), unknown divergence times of nodes (T), and an autocorrelation parameter (ν). As the magnitude of ν increases, rates are more heterogeneous, and as ν decreases, rates become more clocklike. Parameters are estimated on a fixed, rooted, and bifurcating phylogeny, based on the aligned sequence data (X) and branch lengths (B), measured as expected amount of sequence change along a branch and determined by rate and time.

In the initial model, the logarithm of the rate on a descendant branch was normally distributed around a mean equal to the logarithm of the rate of the ancestral branch, and its variance depended on the duration of the ancestral branch (Thorne et al. 1998). In subsequent implementations (Kishino et al. 2001; Thorne and Kishino 2002), rates were decoupled from duration of an ancestral branch. For any given branch (except for one of the branches derived from the root node), the logarithm of the rate at its ending node had a normal distribution around the logarithm of the rate at its starting node. Also, a tendency in the initial model for rates to be higher toward the tips of the tree was corrected (Kishino et al. 2001).

The autocorrelation parameter ν regulates the prior distribution of rates on different branches, given internal node times. Because ν has great influence on prior rates, flexibility was introduced by drawing its magnitude from an exponential distribution (Thorne et al. 1998). Priors for divergence times (T) were determined by the birth rate on a pure-birth (Yule) branching process. The prior rate for one of the branches diverging from the root node (R_0) was obtained from an exponential distribution. The logarithm of the rate of its sister branch was sampled from a normal distribution

with a mean equal to $\log(R_0)$. Prior distributions for rates of all other branches were estimated through the model (Thorne et al. 1998).

Sources of priors for some parameters were modified in the more recent implementation of the model (Kishino et al. 2001; Thorne and Kishino 2002). The prior for v was drawn from a gamma distribution. The prior for T depends on the time difference between the root and the tips of the tree, derived from a gamma distribution. Internal node times are assigned by using paths and by jointly sampling divergence times of branches traversing each path from a Dirichlet distribution (Kishino et al. 2001). The prior for R_0 is drawn from a gamma distribution. The rate of the node ending one of the branches diverging from the root node is forced to be equal to the rate at the starting node. This constraint improves the approximation of posterior distributions by the MCMC by allowing a better mixing rate (Kishino et al. 2001). Rates for all other branches are determined by the model of rate evolution. Empirical results show that relatively large changes in the shape of the gamma distribution from which the prior for R_0 is drawn and on the age of the root node result in small differences in estimated posterior distributions (Yang and Yoder 2003).

Posterior probability distributions of T, R, and v depend on the aligned (nucleotide or amino acid) sequences (X) on a bifurcating and rooted tree (Thorne et al. 1998; Kishino et al. 2001; Thorne and Kishino 2002). An MCMC similar to the one described for the previous method (Huelsenbeck et al. 2000) is used to obtain posterior distributions for $(T, R, v|X)$. After the MCMC is run for a large number of generations, the achieved stationary distributions correspond to the posterior probabilities for rate and divergence times. Because estimating the likelihood of the data given branch lengths may become a limiting factor, this likelihood is approximated through a multivariate normal approach (Thorne et al. 1998).

Kishino et al. (2001) included fossil data to estimate posterior probabilities of divergence times. This was achieved by modifying the posterior probability equation into $p(R, T, v|X, C)$, with C corresponding to a set of fossil constrained ages. Constraining the age of nodes in a tree resulted in reasonably narrow posterior distributions of divergence times. Thorne and Kishino (2002) emphasize that fossil constraints represent the single source of information available for the MCMC to decouple rates and times. However, in this method, introducing the constraints is empirically problematic (Kishino et al. 2001).

Thorne and Kishino (2002) extended the model of Kishino et al. (2001) to include multigene data sets and provided a technique to detect correlated rate changes between two genes. The implementation is based on the assumption that, while different genes may exhibit uncorrelated patterns of rate change, divergence times are shared. The rate of each gene at the root node is obtained from independent draws from a gamma distribution. Given the rate of each gene at the root and a common set of divergence times, the rate of each gene was modeled independently on the common phylogenetic tree. Rate autocorrelation can vary among genes by independently drawing the magnitude of v for each gene from a gamma distribution. Alternatively, a v of constant magnitude, obtained from a specified gamma distribution, may be assigned to all genes. The Metropolis-Hastings algorithm was extended straightforwardly to multigene data sets (Thorne and Kishino 2002).

Nonparametric and semiparametric methods. Two methods to estimate divergence times and variable rates of evolution on a phylogenetic tree that do not rely on modeling rate change among branches are also available. Both methods use temporal autocorrelation as the principle explaining rate heterogeneity in estimation of rates and times (Gillespie 1991; Sanderson 1997, 2002). Instead of being modeled, autocorrelation is introduced as a numerical penalty against drastic rate change between an ancestral branch and its descendants. The two methods differ in the elements they use to estimate rates and times; whereas the nonparametric method relies solely on the penalty against drastic rate changes, the semiparametric method combines this penalty with a fully parametric model that optimizes rates on branches.

Nonparametric rate smoothing. Nonparametric rate smoothing (NPRS; Sanderson 1997) relies on minimizing ancestor-descendant rate changes to estimate rates and times. Starting with a fixed phylogeny for which branch lengths, expressed as expected number of substitutions, are available, the rate of substitution for each branch (r_k) is estimated through $r_k = b_k/t_k$, which follows the definition of branch length (b_k) as the product of rate and time (t_k). Penalization against rapid rate change is introduced by minimizing the difference in rates between a branch and its immediate descendants. By summing all local minimizations across the phylogeny, an overall objective function (W), which depends on unknown divergence times, is obtained. Estimates of unknown divergence times are obtained by finding the set of internal node times that minimize the objective function. This minimization is a nonlinear optimization function that can be solved by numerical techniques (Sanderson 2003).

The optimization technique allows the introduction of minimum and/or maximum age constraints on ages of nodes obtained, for example, from the fossil record. Age constraints are introduced by specifying that the age of a given internal node cannot be younger than the age of the fossil evidence (in the case of a minimum age constraint). After a set of age constraints have been specified, a feasible space of internal node ages, comprising only those that satisfy the conditions imposed by the constraints, is determined. The optimization is performed only over those internal node times permitted in the feasible space.

Penalized likelihood. A semiparametric method to estimate variable rates and divergence times across a phylogeny is based on penalized likelihood (PL; Sanderson 2002). It relies on a parameter-rich model to estimate rates on branches and a numerical penalty to avoid rapid rate variation among nearby branches (as in NPRS). The relative contribution of the parametric model and the penalty into estimating rates and divergence times is regulated by a smoothing parameter.

A saturated parametric model in which each branch on the tree has its own rate of substitution can be optimized on a fixed phylogeny of known branch lengths. Estimating rates and times using a saturated model may be problematic, first, because the available data may be insufficient to explain the necessary parameters, and second, because of the reduced

predictive power resulting from overfitted data. An option to impose constraints on rate variation is to introduce rate autocorrelation as a roughness penalty (Φ), which increases as rates among branches change more rapidly, forcing smooth changes in rates across the tree. The penalized likelihood for this constrained model is the likelihood of the saturated model given the number of substitutions on a branch (x_k) minus the product of the roughness penalty and a smoothing parameter (λ) for the rate of each of the branches. By maximizing the penalized likelihood, estimates of rates and times are obtained.

The roughness penalty (Φ) is the same penalization against drastic rate change implemented in NPRS (Sanderson 1997). The smoothing parameter (λ) regulates the relative contribution of the saturated model and the penalization in estimating rates and times. When it is small, the saturated model predominates, and rate variation is extensive. As it increases, the penalty against rate change predominates, and rates approach constancy.

The smoothing parameter is a determinant factor in estimating rates and divergence times; therefore, selecting a value that accurately fits the data is crucial. Its optimal magnitude can be approximated through a cross-validation procedure implemented on the available data. The cross validation consists of pruning terminal branches from the phylogeny, estimating model parameters (including λ) on the remaining branches through penalized likelihood, and using the estimated parameters to predict the length (x_k^*) of the removed branches. The difference between predicted (x_k^*) and observed (x_k) branch lengths represents the prediction error (CV), which is used as an index of predictive power. A tree-wide prediction error for a set of parameters is obtained by averaging prediction errors obtained from pruning each of the terminal branches. The λ associated with the set of parameters resulting in the lowest tree-wide prediction error should be chosen (Sanderson 2002). As in NPRS, the PL method straightforwardly allows including constraints to node ages from the fossil record. These constraints can be implemented in estimating the optimal magnitude of the smoothing parameter as well as rates and divergence times.

Comparative performance. Development of methods to estimate rates and divergence times allowing rate heterogeneity is an area of current expansion, and relatively few exploratory comparative analyses of their accuracy, tendencies, and possible biases have been conducted. The effect of different models of rate change and sensitivity to rate heterogeneity in Bayesian estimation of rates and divergence times was evaluated by Aris-Brosou and Yang (2002). Their study also compared estimates derived from Bayesian and from model selection methods (local clocks). Their method is directly based on the Bayes theorem, from which posterior distributions of rates and divergence times are obtained. Priors for divergence times were obtained from a generalized birth and death process (Kendall 1948) with constant speciation and extinction rates. Priors for rates of branches were obtained from a distribution with a mean equal to the rate of the ancestral branch and variance proportional to branch duration. The hyperparameter of rate change (σ^2) specifies the increase in rate variance as a function of time. If σ^2 is small, the model approaches rate constancy; if large, more heterogeneity is allowed. Five distributions, corresponding to different models of rate change, were implemented: lognormal (LND), equivalent to the hierarchical model of Thorne et al. (1998); stationary lognormal (SLD), equivalent to the model of Kishino et al. (2001); gamma (GD); exponential (ED), in which rate variance is a direct function of the mean; and an Ornstein-Uhlenbeck process (OUP) based on a model of the speed of a particle in Brownian motion as a function of time. In this model, an additional parameter (β) influences the mean of the distribution (Aris-Brosou and Yang 2002). The five models approximate the molecular clock if σ^2 is close to 0. Posterior distributions were approximated through an MCMC, as described for other Bayesian methods. The optimal value for the hyperparameter of rate change under LND, SLD, GD, and OUP, and the β parameter for OUP, were obtained via an empirical Bayesian approach. Performance of different methods was assessed by considering the marginal probability of a model, denoted $p(X|M_k)$, which contains information about its performance. The posterior Bayes factor, estimated from the MCMC outputs, was used to compare models.

Rates and divergence times were estimated under different models of rate change on a small data set of tRNA-coding mitochondrial genes for six hominoid genera, and on an 18S rRNA data set for 39 metazoan genera. General results showed that similar magnitudes of the hyperparameter of rate change implied different levels of rate heterogeneity under different models. When the optimal hyperparameter for a given model was implemented, divergence times estimated with different models were similar. Nevertheless, OUP had a significantly better fit to the data, possibly because it is governed by an additional hyperparameter (Aris-Brosou and Yang 2002).

Additional results indicated that the Bayesian models exhibit different sensitivity to the level of introduced rate heterogeneity and that using the estimated optimal magnitude for σ^2 (according to each model) lead to estimated ages more in agreement with independent (fossil) evidence. Finally, when rate heterogeneity was allowed, Bayesian methods estimated much younger ages for nodes than model selection methods, although the latter were associated with large sampling errors. While concluding that Bayesian methods represent a promising alternative for estimating rates and divergence times in absence of rate constancy, Aris-Brosou and Yang (2002) caution about errors and biases introduced by erroneous tree topology and the use of single genes.

Comparative performance assessments of nonparametric and semiparametric methods that relax rate constancy provide explicit guidelines about conditions under which one method outperforms the other or the molecular clock (Sanderson 1997, 2002). NPRS and the constant-rate method of Langley and Fitch (1974; see above) were used to estimate internal divergence times on a set of 250 simulated trees generated from a pure birth (Yule) process. The performance of both methods was assessed under variable quantities of data (sequence length), different levels of autocorrelation, and different magnitudes of rate change. The general results indicate that both methods perform equally well when "infinite" data are available and rates are highly correlated. NPRS outperforms the constant-rate method when rates are unclocklike

but highly autocorrelated and also when the amount of data decreases moderately. However, NPRS did not perform well when the amount of data and autocorrelation were further reduced and the magnitude of rate change was increased (Sanderson 1997).

The relationship between prediction errors (CV) and level of smoothing (implicitly or explicitly) used in NPRS, PL, and the Langley-Fitch constant-rate method were analyzed on five significantly unclocklike molecular data sets for different groups of land plants plus a clocklike data set simulated through a pure birth stochastic process (Sanderson 2002). The magnitude of the smoothing parameter was varied on a log scale from 0.1 to 10,000, and the prediction error associated with each of the points was calculated. In all data sets, the smoothing values used by PL were associated with the lowest prediction errors, as expected, given the explicit choice of an optimal smoothing parameter derived from a data-driven cross-validation procedure. Smoothing values used by NPRS and the constant-rate method were usually associated with prediction errors distant from minimum, except when the constant-rate method was applied to the simulated clocklike data. For this data set, the cross-validation procedure selected an essentially clocklike smoothing parameter, which was implemented in PL calculations. In unclocklike data sets, PL outperformed NPRS and the constant-rate method as long as a cross-validation-determined optimal smoothing value was used. NPRS was found to overfit the data, resulting in a reduced predictive power, but nevertheless was associated with lower prediction errors than the constant-rate method (except when data were clocklike or nearly so). Estimating rates and divergence times was problematic when the smoothing parameter was too small, that is, when rate heterogeneity was high (Sanderson 2002). Overall, estimated divergence times depended on the implemented smoothing value, but nevertheless, branches reacted differently to changes in its magnitude.

Paleobiological Approaches to Estimating Lineage Duration

In spite of inherent biases (previously discussed), the fossil record represents one of the few, and arguably the best, sources of nonmolecular information about ages of clades. Fossils can provide minimum ages that can be compared with molecular-derived estimates or implemented as calibrations or constraints on phylogenetic trees in the context of molecular-based methods. In addition to this significantly relevant information, paleobiological research has yielded sophisticated quantitative approaches for investigating the duration of lineages through the integrative use of paleontological data. In contrast with molecular-based methods, the emphasis of paleobiological methods is to estimate probabilities of different parameters associated with lineage duration, which may then be compared with point estimates derived from alternative approaches (e.g., molecular-based methods). Paleontological-based methods allow us to estimate confidence intervals likely to contain the time of origination of a lineage or to evaluate the plausibility of a missing fossil history postulated by an extensive difference between a lineage's first fossil occurrence and its independently estimated age.

Confidence Intervals on Fossil Ages

As previously discussed, a temporal difference of unknown magnitude exists between lineage splitting and a clade's appearance in the fossil record (fig. 1). The stratigraphic range or time period likely to contain the origination of a clade can be obtained by estimating a confidence interval of known probability below the clade's earliest fossil record.

Absence of a fossil species at a given level below (or above) its observed stratigraphic range means either that the species had not originated (or was extinct) by the time of deposition or that the species was alive but was not preserved in the fossil record. A useful guideline to distinguish instances of nonpreservation from preorigination (or postextinction) of a species is the use of taphonomic controls. Taphonomic controls are a set of species expected to be present in all biotic and fossil assemblages where the species of interest is present. The co-occurrence of these species is due to shared geographical and environmental distributions and similar preservation potential. The occurrence of control groups below or above the observed stratigraphic range of the focal species aids in distinguishing between nonpreservation and preorigination (or postextinction) of the focal group. Bottjer and Jablonski (1988) developed an explicit methodology implementing this approach. While compelling, applicability of this method is nevertheless limited to organisms with relatively widespread and dense records and, especially, for which a set of reliably co-occurring taxa can be identified.

With what certainty can it be assumed that absence of a species in the record truly indicates its nonexistence? Methods that assess the uncertainty of times of origin and extinction are based on the idea that the difference between the origin of a clade and its first fossil appearance is inversely correlated with the quality of its fossil record. Confidence intervals that contain true ages of origin and extinction of lineages rely on observed stratigraphic ranges and quality of the fossil record, measured as the number of fossil horizons in which the focal species is present. A straightforward equation to estimate confidence intervals on local stratigraphic ranges was proposed by Strauss and Sadler (1989). This equation was extended by Marshall (1990) to estimate confidence intervals on temporal durations of entire species or clades using their distribution in composite stratigraphic sections. The size of the confidence interval at either end of the stratigraphic section (α) is expressed as a proportion of the total stratigraphic range, and its magnitude depends on the number of observed fossil horizons and desired confidence probability (Marshall 1990). While the method is conceptually straightforward, it relies on two strict assumptions: fossil horizons are randomly distributed across the stratigraphic section, and collecting intensity is uniform between first and last occurrences. Statistical tests for the random distribution of fossil horizons should be applied; however, their effectiveness is reduced as the record is poorer because departures from randomness are more difficult to detect.

A random distribution of fossil horizons can only result from stochastically constant sedimentation rates and

preservation potential during the life span of a lineage, a situation probably rarely met. Rather, the distribution of gaps between fossil horizons argues against random fossilization (Wise 1991, in Marshall 1994). A method that requires no assumptions about distribution of fossil horizons in stratigraphic sections is based on the probability that the size of gaps between all observed fossil horizons is smaller or larger than the median of the underlying distribution of gap sizes (Marshall 1994). Whereas this method only assumes absence of correlation between the size of gaps and their stratigraphic position, it introduces uncertainty around the size of the confidence interval itself and requires a rich fossil record if a high confidence probability on a particular interval is desired. The method allows the estimation of bounds on the size of an interval for a given number of gaps according to a selected confidence level and a chosen confidence probability (Marshall 1994). Equations allow the identification of the lower and upper bounds of a confidence interval under a given confidence probability. A value that satisfies one or both equations may not be found, and thus, assigning bounds to confidence intervals may not be possible. This situation may arise especially when estimating an inclusive confidence interval (e.g., 95%) with high confidence probability. A test to detect increasing or decreasing trends in gap size in relation to stratigraphic position is included (Marshall 1994). While this distribution-free method has the undesired property of introducing uncertainty to confidence intervals, its requirement for large amounts of data ensures its application only when available data are sufficiently abundant to provide high confidence probabilities on confidence intervals.

A method that allows introducing known or predicted collecting and preservation biases into calculation of confidence intervals was subsequently presented by Marshall (1997). This generalized method requires explicit quantification of collecting and preservation biases and assumes that the observed distribution of fossil horizons is consistent with such quantification. Confidence intervals at lower or upper endpoints of stratigraphic ranges are obtained from the weighted average of the size of gaps between fossil horizons. Weights are assigned on the basis of differential fossilization potential within the observed stratigraphic range as well as the number of observed gaps, desired confidence probability, and fossilization potential within the confidence interval (Marshall 1997). The original method for estimating confidence intervals at endpoints of stratigraphic ranges (Strauss and Sadler 1989; Marshall 1990) represents a special case of this generalized method. In addition to primary data of fossil occurrence through measured stratigraphic sections, the generalized method requires quantification of preservation and collection biases in the form of a fossil recovery potential function (FRPF). The FRPF can be obtained from independent information about preservation and recovery probabilities for a given taxon, for example, surface of bedding planes or water depth, shown to correlate with preservation potential for marine invertebrates (Holland 1995, in Marshall 1997). Importantly, estimation of FRPF should not rely on empirically observed distributions of fossil finds. Consistency between observed distribution of fossil finds and distribution expected from the independently derived FRPF is critical and should be explicitly evaluated (e.g., through a Kolmogorov-Smirnov goodness-of-fit test; Marshall 1997). If consistency is not found, the generalized method should not be applied, and instead, the distribution-free method (Marshall 1994) may be implemented. Efficacy of the generalized method depends on how accurately fossil recovery potential can be determined.

Preservation Probability Consistent with Postulated Missing History

An extensive temporal difference between a lineage's first fossil occurrence and an independently estimated age (e.g., a molecular-based age) implies a missing segment in the lineage's fossil record. Considering that absence in the fossil record of an already existent lineage requires explanation on the basis of a poor preservation potential, Foote et al. (1999) presented an approach to address "how low the rate of fossil preservation must be for all species of a group to escape detection over a specified interval of geologic time" (Foote et al. 1999, p. 1310). To test the plausibility of a clade's postulated missing record, the maximal preservation probability consistent with such a missing record can be estimated and compared with the clade's independently derived fossil preservation potential (Foote and Raup 1996; Foote 1997). If the former is much lower than the latter, the postulated missing history cannot be supported.

Foote et al. (1999) developed equations to estimate a lineage's maximal preservation rate (r) and maximal extinction rate (q), consistent with a postulated missing history, on the basis of the known diversity (number of lineages) of the clade at its first fossil occurrence (N) and duration of the postulated missing record (T). Preservation and extinction rates can be estimated according to four different models of species diversification. The first model is exponential diversification, the second considers the survival of the clade at least until its first appearance in the fossil record, the third considers that the clade's diversity is exactly equal to the observed at its first fossil occurrence, and the fourth considers that the clade's diversity is equal or greater than the observed at its first fossil occurrence.

The test was used to evaluate the plausibility of the postulated 60-m.yr. missing record for eutherian mammals implied by the difference between their first fossil occurrence in the early Tertiary (Foote et al. 1999 and references therein) and the much older age derived from a molecular clock estimate (Kumar and Hedges 1998). The sum of missing species durations (S) and magnitude of preservation rate (r) below which the probability of a nonexistent fossil record is >0.5 were estimated under the four diversification models. Estimated preservation probabilities consistent with the postulated missing history for eutherian mammals are much smaller than independently estimated preservation rates (Foote and Raup 1996; Foote 1997), the discrepancy being at least one order of magnitude. Thus, the postulated missing history for eutherian mammals cannot be supported (Foote et al. 1999).

Introducing Uncertainty around Parameters in Tests of Postulated Durations

Methods that evaluate the probability of a postulated lineage duration implement several relevant parameters as fixed

constants. In reality, these parameters are variables of unknown magnitude. In an integrative likelihood approach, Wagner (2000) postulates a number of tests that evaluate hypothesized durations while simultaneously introducing uncertainty around parameters of unknown magnitude, namely the probability of fossil preservation.

The likelihood of a hypothesized duration (D) given observed stratigraphic range (R) and number of fossil finds (n) and the likelihood of no finds over a postulated range extension, both for Poisson-distributed and binomial-distributed counts of fossil finds, are the basis for simultaneously evaluating the likelihood of postulated duration and of sampling intensity (θ, equivalent to preservation rate). A comparison of the likelihoods of postulated duration estimated including or excluding the likelihood of sampling intensity shows that accommodating uncertainty around sampling intensity results in a higher likelihood score for the postulated duration. Nevertheless, in the case of the considered example, the hypothesized duration is highly unlikely under either assessment (Wagner 2000). A test for the likelihood of a postulated duration over all sampling intensities is described. Relative likelihoods of alternative postulated durations are extremely similar to relative likelihoods found when solving for both duration and sampling intensity (Wagner 2000).

The likelihood of two or more sampling intensities on a lineage or over a phylogeny are tested through likelihood ratio tests against the simpler hypothesis of a single sampling intensity. In the first example, a two-sampling-intensities hypothesis along a lineage could not be rejected in favor of a three-sampling-intensities hypothesis. When assessing optimal number of sampling intensities over a phylogeny, a two-sampling-intensities hypothesis was favored over a single sampling intensity, but the two sampling intensities could not be rejected in favor of three or more sampling intensities across the phylogeny. One of the main conclusions of this study is that while tests that accommodate uncertainty about parameters of unknown magnitude provide more likely results; results from simpler tests, which treat unknown parameters as given models, cannot be rejected in favor of the former.

Concluding Remarks

Evolutionary thinking has been inextricably linked with questions about ages of clades. For a substantial part of the development of evolutionary thought, ages of clades were derived from paleontological information. Subsequently, molecular biology provided a framework to estimate lineage divergence times from molecular distances (Zuckerkandl and Pauling 1962). The nature of the paleontological record and of the process of molecular substitution complicate estimating the age of clades, which is further aggravated by the link between substitution rate and temporal duration in the branches of a phylogeny, by methodological biases, and the fact that unless the process of cladogenesis is actually observed (e.g., in viral lineages), realistic data against which to test dating methods are unavailable.

The past few decades have witnessed the proliferation of conceptual and methodological advances, both from molecular phylogenetic and paleobiological fields, resulting in promising approaches to investigate the temporal framework of lineages. These methods are based on alternative data and analytical procedures (i.e., numerical, probabilistic, or parametric), and on explicit assumptions. Of particular relevance is the formulation of molecular-based methods that relax rate constancy, which are remarkable approaches toward a more realistic estimation of substitution rates and divergence times. Since their original presentations, several methods have experienced modifications aimed toward greater realism and analytical power (Sanderson 1997, 2002; Thorne et al. 1998; Thorne and Kishino 2002). However, because of their recent formulation, their possible tendencies and biases have only partially been evaluated (Sanderson and Doyle 2001; Aris-Brosou and Yang 2002; Yang and Yoder 2003). While already extremely sophisticated and powerful, further improvements can be envisaged in two areas: allowing uncertainty around temporal calibrations and constraints, and modifying the criterion that determines how variation of rates among lineages is introduced.

The possibility of incorporating nonmolecular absolute temporal information (e.g., from fossils) as calibration or constraints permits a more accurate estimation of divergence times and rates by molecular rate-heterogeneous methods. Because fossil ages represent one of the few sources of absolute information to decouple rates and times in a phylogeny, a substantially greater effort needs to be devoted to obtaining information of first fossil occurrences and to develop and test explicit methodologies for integrating fossils into a phylogenetic context. Uncertainties around ages of fossils stem at least from two sources: the temporal gap of unknown magnitude between the origin of a lineage and its first fossil occurrence (depending on its preservation probability and/or recognition in the record), and the usual availability of the age of a fossil in terms of its membership in a stratigraphic unit (e.g., Aptian) but not in terms of a fixed point in geologic time. From a methodological standpoint, rate-heterogeneous methods could allow uncertainty on ages of calibration and constraint nodes based on the confidence interval expected to contain the time of origin of the lineage under a known probability and the geologic time interval associated with stratigraphic units.

Most of the currently available rate-heterogeneous methods introduce among-lineage rate change on the basis of temporal autocorrelation (Gillespie 1991). Rate autocorrelation represents a relevant and useful precept about the process of rate change among ancestor and descendant species over a microevolutionary timescale. However, it is unclear if the autocorrelation principle applies to molecular rate change among major branches of the tree of life over a macroevolutionary timescale. Little is known about the process of molecular change on a macroevolutionary scale, probably owing to the fact that inferences about macroevolutionary processes are usually derived from paleontological patterns, which provide no direct information about events at the molecular genomic level. Direct investigation of the process of molecular change can be carried out mostly on a short timescale. Perhaps it might be possible to postulate processes of molecular change associated with evolutionary radiations, or with the establishment of major branches of the tree of life, guided by the patterns of appearance and diversification

of major lineages through geologic time. Postulation of macroevolutionary molecular processes is a research area to be further developed and its predictions evaluated (Marshall et al. 1999; Jablonski 2000; Shubin and Marshall 2000).

Molecular-based methods for estimating divergence times and rates could model among-lineage rate change by superimposing a (newly postulated) process of macroevolutionary rate change on temporal autocorrelation. While autocorrelation would act as a constant process that introduces rate change among ancestral and descendant branches, the macroevolutionary process would act as a discrete temporal event that disrupts autocorrelation by introducing major rate changes at certain points in a phylogenetic tree.

A bigger picture about the temporal framework of lineages can result from the combined use of molecular and paleontological data and the complementary implementation of different methodological approaches. Incorporation of fossil information to constrain ages of nodes provides enhanced realism to divergence times and rates estimated via molecular-based, rate-heterogeneous methods. A greater power lies in evaluating molecular-derived age estimates against predictions derived from paleontological data. These include assessing whether a molecular-based age estimate falls within the confidence interval expected with a known probability to contain the true age of the lineage and explicitly evaluating the plausibility of a lineage's postulated missing fossil history.

Acknowledgments

I thank S. S. Renner and T. J. Givnish for organizing the symposium "Tropical Intercontinental Disjunctions" and for inviting me to participate in it, M. J. Sanderson for conversations and insights on molecular dating, and G. A. Salazar and two anonymous reviewers for useful comments on earlier drafts of the manuscript, particularly regarding observations on the nature of the paleontological record. I also thank M. E. Olson for his critical reading of one of the latest versions of the manuscript and for suggesting numerous improvements.

Literature Cited

Aris-Brosou S, Z Yang 2002 Effects of models of rate evolution on estimation of divergence dates with special reference to the metazoan 18S ribosomal RNA phylogeny. Syst Biol 51:703–714.

Ayala FJ 1997 Vagaries of the molecular clock. Proc Natl Acad Sci USA 94:7776–7783.

Ayala FJ, A Rzhetsky, J Ayala 1998 Origin of the metazoan phyla: molecular clocks confirm paleontological estimates. Proc Natl Acad Sci USA 95:606–611.

Benton MJ, FJ Ayala 2003 Dating the tree of life. Science 300: 1698–1700.

Bottjer DJ, D Jablonski 1988 Paleoenvironmental patterns in the evolution of post-Paleozoic benthic marine invertebrates. Palaios 3: 540–560.

Bremer K, MHG Gustafsson 1997 East Gondwana ancestry of the sunflower alliance of families. Proc Natl Acad Sci USA 94:9188–9190.

Britton T, B Oxelman, A Vinnersten, K Bremer 2002 Phylogenetic dating with confidence intervals using mean path lengths. Mol Phylogenet Evol 24:58–65.

de Queiroz K, J Gauthier 1990 Phylogeny as a central principle in taxonomy: phylogenetic definitions of taxon names. Syst Zool 39: 307–322.

Felsenstein J 1978 Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool 27:401–410.

Foote M 1997 Estimating taxonomic durations and preservation probability. Paleobiology 23:278–300.

Foote M, JP Hunter, CM Janis, JJ Sepkoski Jr 1999 Evolutionary and preservational constraints on origins of biologic groups: divergence times of eutherian mammals. Science 283:1310–1314.

Foote M, DM Raup 1996 Fossil preservation and the stratigraphic ranges of taxa. Paleobiology 22:121–140.

Gaut BS, SV Muse, WD Clark, MT Clegg 1992 Relative rates of nucleotide substitution at the *rbcL* locus of monocotyledonous plants. J Mol Evol 35:292–303.

Gillespie JH 1991 The causes of molecular evolution. Oxford University Press, Oxford. 336 pp.

Goodman M, GW Moore, G Matsuda 1975 Darwinian evolution in the genealogy of hemoglobin. Nature 253:603–608.

Graybeal A 1998 Is it better to add taxa or characters to a difficult phylogenetic problem? Syst Biol 47:9–17.

Green PJ 1995 Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82:711–732.

Hasegawa M, H Kishino, 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.

Heckman DS, DM Geiser, BR Eidell, RL Stauffer, NL Kardos, SB Hedges 2001 Molecular evidence for the early colonization of land by fungi and plants. Science 293:1129–1132.

Hillis DM, BK Mable, C Moritz 1996 Applications of molecular systematics: the state of the field and a look to the future. Pages 515–543 *in* DM Hillis, C Moritz, BK Mable, eds. Molecular systematics. Sinauer, Sunderland, Mass.

Holland SM 1995 The stratigraphic distribution of fossils. Paleobiology 21:92–109.

Huelsenbeck JP, B Larget, RE Miller, F Ronquist 2002 Potential applications and pitfalls of Bayesian inference of phylogeny. Syst Biol 51:673–688.

Huelsenbeck JP, B Larget, D Swofford 2000 A compound Poisson process for relaxing the molecular clock. Genetics 154: 1879–1892.

Jablonski D 2000 Micro- and macroevolution: scale in hierarchy in evolutionary biology and paleobiology. Pages 15–52 *in* DE Erwin, SL Wing, eds. Deep time: paleobiology's perspective. Paleontological Society, Lawrence, Kans.

Jefferies RSP 1979 The origin of chordates: a methodological essay. Pages 443–447 *in* MR House, ed. The origin of major invertebrate groups. Systematics Association Special Vol 12. Academic Press, London.

Kendall DG 1948 On the generalized birth-and-death process. Ann Math Stat 19:1–15.

Kishino H, M Hasegawa 1990 Converting distance to time: application to human evolution. Methods Enzymol 183:550–570.

Kishino H, JL Thorne, WJ Bruno 2001 Performance of a divergence time estimation method under a probabilistic model of rate evolution. Mol Biol Evol 18:352–361.

Kumar S, SB Hedges 1998 A molecular timescale for vertebrate evolution. Nature 392:917–920.

Langley CL, WM Fitch 1973 The constancy of evolution: a statistical analysis of a and b haemoglobins, cytochrome c, and fibrinopeptide A. Pages 246–262 *in* NE Morton, ed. Genetic structure of populations. University of Hawaii Press, Honolulu.

——— 1974 An examination of the constancy of the rate of molecular evolution. J Mol Evol 3:161–177.

Lewis LA, BD Mishler, R Vilgalys 1997 Phylogenetic relationships of the liverworts (Hepaticae), a basal embryophyte lineage, inferred from nucleotide sequence data of the chloroplast gene *rbcL*. Mol Phylogenet Evol 7:377–393.

Lewis PO 1998 Maximum likelihood as an alternative to parsimony for inferring phylogeny using nucleotide sequence data. Pages 132–163 *in* DE Soltis, PS Soltis, JJ Doyle, eds. Molecular systematics of plants: DNA sequencing. Kluwer Academic, Boston.

——— 2001 Phylogenetic systematics turns over a new leaf. Trends Ecol Evol 16:30–37.

Li W-H 1997 Molecular evolution. Sinauer, Sunderland, Mass. 487 pp.

Li W-H, M Tanimura 1987 The molecular clock runs more slowly in man than in apes and monkeys. Nature 326:93–96.

Magallón S, MJ Sanderson 2003 Estimating the age of angiosperms using constraints from the fossil record. Dating in the 21st century: theory and reality of finding a date for your clade. American Society of Plant Taxonomists/Botanical Society of America symposium. Program and abstracts.

Marshall CR 1990 Confidence intervals on stratigraphic ranges. Paleobiology 16:1–10.

——— 1994 Confidence intervals on stratigraphic ranges: partial relaxation of the assumption of randomly distributed fossil horizons. Paleobiology 20:459–469.

——— 1997 Confidence intervals on stratigraphic ranges with nonrandom distribution of fossil horizons. Paleobiology 23:165–173.

Marshall CR, HA Orr, NH Patel 1999 Morphological innovation and developmental genetics. Proc Nat Acad Sci USA 96:9995–9996.

Martin W, A Gierl, H Saedler 1989 Molecular evidence for pre-Cretaceous angiosperm origins. Nature 339:46–48.

Martin W, D Lydiate, H Brinkmann, G Forkmann, H Saedler, R Cerff 1993 Molecular phylogenies in angiosperm evolution. Mol Biol Evol 10:140–162.

Muse SV 2000 Examining rates and pattern of nucleotide substitution in plants. Plant Mol Biol 42:25–43.

Page DM, EC Holmes 1998 Molecular evolution: a phylogenetic approach. Blackwell, Oxford. 346 pp.

Poe S 1998 Sensitivity of phylogeny estimation to taxonomic sampling. Syst Biol 47:18–31.

Posada D, KA Crandall 1998 Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818.

Pupko T, D Huchon, Y Cao, N Okada, M Hasegawa 2002 Combining multiple data sets in a likelihood analysis: which models are the best? Mol Biol Evol 19:2294–2307.

Rambaut A, L Bromham 1998 Estimating divergence dates from molecular sequences. Mol Biol Evol 15:442–448.

Rodríguez-Trelles F, R Tarrío, FJ Ayala 2001 Erratic overdispersion of three molecular clocks: GPDH, SOD, and XDH. Proc Natl Acad Sci USA 98:11405–11410.

Sanderson MJ 1997 A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol Biol Evol 14: 1218–1232.

——— 1998 Estimating rate and time in molecular phylogenies: beyond the molecular clock? Pages 242–264 *in* DE Soltis, PS Soltis, JJ Doyle, eds. Molecular systematics of plants: DNA sequencing. Kluwer Academic, Boston.

——— 2002 Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol Biol Evol 19:101–109.

——— 2003 r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19:301–302.

Sanderson MJ, JA Doyle 2001 Sources of error and confidence intervals in estimating the age of the angiosperms from *rbcL* and 18S rDNA data. Am J Bot 88:1499–1516.

Shubin NH, CR Marshall 2000 Fossils, genes, and the origin of novelty. Pages 324–340 *in* DE Erwin, SL Wing, eds. Deep time: paleobiology's perspective. Paleontological Society, Lawrence, Kans.

Simmons MP, JV Freudenstein 2002 Artifacts of coding amino acids and other composite characters for phylogenetic analysis. Cladistics 18:354–365.

Strauss D, PM Sadler 1989 Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. Math Geol 21:411–427.

Takahata N 1987 On the overdispersed molecular clock. Genetics 116:169–179.

Takezaki N, A Rzhetsky, M Nei 1995 Phylogenetic test of the molecular clock and linearized trees. Mol Biol Evol 12: 823–833.

Thorne JL, H Kishino 2002 Divergence time and evolutionary rate estimation with multilocus data. Syst Biol 51:689–702.

Thorne JL, H Kishino, IS Painter 1998 Estimating the rate of evolution of the rate of molecular evolution. Mol Biol Evol 15: 1647–1657.

Uyenoyama M 1995 A generalized least-squares estimate for the origin of sporophytic self-incompatibility. Genetics 139: 975–992.

Wagner PJ 2000 Likelihood tests of hypothesized durations: determining and accommodating biasing factors. Paleobiology 26: 431–449.

Wise KP 1991 The use of fossil occurrence data to distinguish between instantaneous and stepwise extinction. *In* Abstract with Programs, Rocky Mountain Section 44th Annual Meeting. Geol Soc Am Bull 23:A184.

Wu C-I, W-H Li 1985 Evidence for higher rates of nucleotide substitution in rodents than in man. Proc Natl Acad Sci USA 82: 1741–1745.

Yang Z, AD Yoder 2003 Comparison of likelihood and Bayesian methods for estimating divergence times using multiple loci and calibration points, with application to a radiation of cute-looking mouse lemur species. Syst Biol 52:705–716.

Yoder AD, Z Yang 2000 Estimation of primate speciation dates using local molecular clocks. Mol Biol Evol 17:1081–1090.

Zharkikh A 1994 Estimation of evolutionary distances between nucleotide sequences. J Mol Evol 39:315–329.

Zuckerkandl E, L Pauling 1962 Molecular disease, evolution, and genetic heterogeneity. Pages 189–225 *in* M Kasha, B Pullman, eds. Horizons in biochemistry. Academic Press, New York.

——— 1965 Evolutionary divergence and convergence in proteins. Pages 97–166 *in* V Bryson, HJ Vogel, eds. Evolving genes and proteins. Academic Press, New York.