

7

**USING STRATIGRAPHIC
INFORMATION IN PHYLOGENETICS**

Perhaps the greatest single advance in systematics over the past 50 years has been the introduction of explicit methods for reconstructing phylogeny. Before about 1950, the construction of phylogenetic trees was entrusted only to those individuals with an expert knowledge of specific taxonomic groups. These experts were often paleontologists who constructed evolutionary trees such as the one shown in figure 7.1. Early phylogenies were constructed mostly on the basis of experts' intuitions about relationships among groups. Although today few systematists would argue that phylogenies should be reconstructed solely on intuition, it is interesting to note that early systematists frequently incorporated additional sources of information into their phylogenetic descriptions that modern systematists now rarely consider; the specific positions of fossil species in the stratigraphic record, for example, were usually taken into account. Early phylogenetic trees, such as the phylogeny of the elephants shown in figure 7.1, were generally consistent with the temporal distributions of fossil species.

The temporal order of fossil species is rarely considered in contemporary phylogenetic analyses that are based on DNA sequences or morphology of extant species. In fact, only a handful of modern researchers have either directly incorporated stratigraphic data into a phylogenetic analysis or investigated the agreement of phylogenetic trees with the stratigraphic record (Harper 1976; Gingerich 1979; Fisher 1982, 1988, 1991, 1992, 1994; Gauthier et al. 1988; Marshall 1990; Norell and Novacek 1992a, 1992b; Huelsenbeck 1994; Benton 1995; Wagner 1995; Benton and Storrs 1996; Brochu 1997; Huelsenbeck and Rannala 1997). Moreover, explicit methods for incorporating stratigraphic information into a phylogenetic analysis have been developed only recently, and

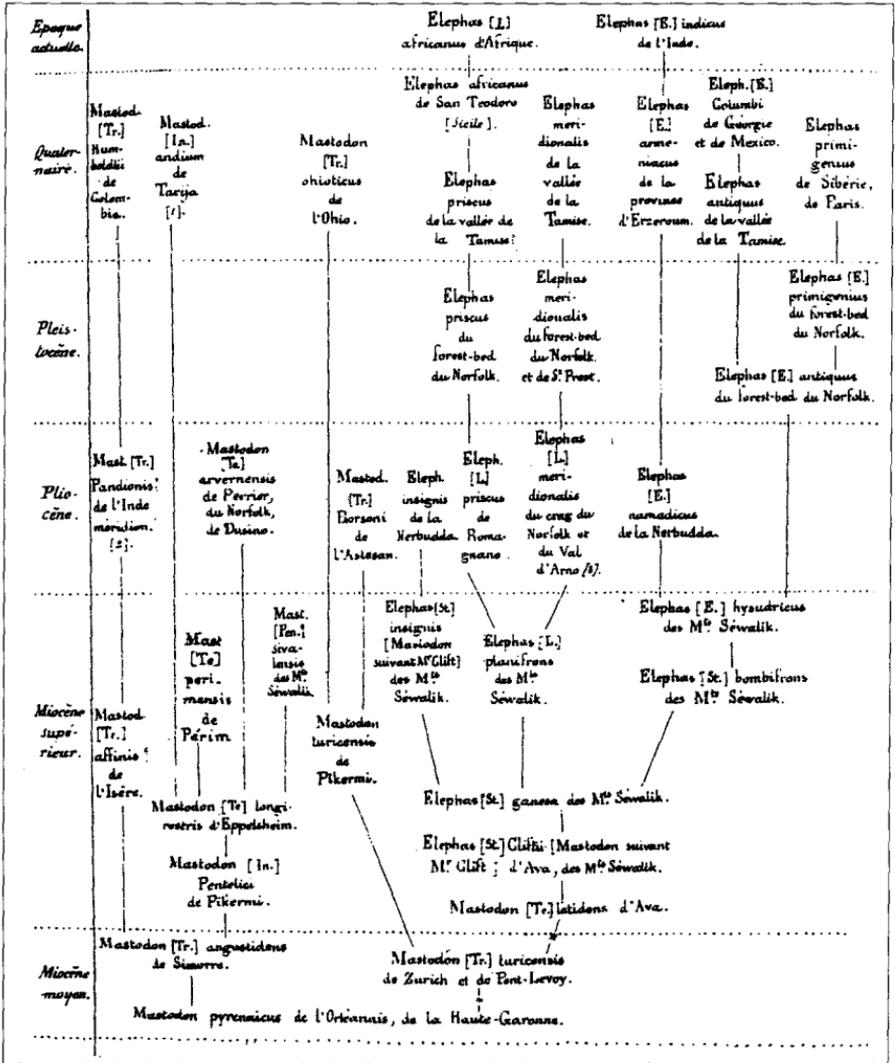


Figure 7.1. A phylogeny of elephants proposed by Gaudry (1862) based, in large part, on collections of Miocene mammals from Pikermi, Greece. Gaudry's phylogenies of mammals are among the first published after the *Origin of Species* (Darwin 1859). Note that the phylogenetic tree of elephants appears consistent with the occurrence of different ancestral elephant species in the stratigraphic record.

little is known concerning the relative performance of the different methods. In this chapter, we review several recently developed methods designed to incorporate stratigraphic data into phylogenetic analysis. Although in this chapter we discuss stratigraphic data, except in very rare cases (e.g., Krings et al. 1997) fossils can be used only with morphological data.

CHANGING PERCEPTIONS OF THE IMPORTANCE OF FOSSILS

For about a century after the publication of *On the Origin of Species* (Darwin 1859), paleontologists were the main practitioners of phylogenetics. Many, perhaps most, biologists believed that fossil evidence was necessary to construct accurate evolutionary trees (e.g., Simpson 1961). However, with the introduction of explicit phylogenetic methods and of new sources of data, such as amino acid or DNA sequences, attitudes changed about the relevance of fossils in a phylogenetic analysis. Because fossils contain only a small fraction of the characters present in extant species, they came to be thought of as less useful for resolving relationships (Jefferies 1979; Hennig 1981; Patterson 1981; Rosen et al. 1981; Forey 1986).

In the late 1980s several groups rigorously examined the effect fossils have on phylogenetic analysis (see review by Donoghue et al. 1989). For example, Doyle and Donoghue (1987) examined the importance of fossils for seed plant phylogeny. Their general conclusion was that although the fossil data did not overturn a phylogeny constructed from data on living taxa, the interpretation of character evolution did change. Gauthier et al. (1988) examined the importance of fossils by studying amniotes. The results of their analysis suggested that the inclusion of fossils was critical. The phylogenetic tree constructed in the absence of fossils placed birds and mammals as sister taxa, whereas the inclusion of fossils resulted in a tree with Reptilia and mammals as sister taxa. Importantly, the phylogenetic tree proposed by Gauthier et al. (1988) was more consistent with the order of occurrence of fossil groups in the rock record. Moreover, the tree proposed by Gauthier et al. (1988) has been confirmed by collection of additional data (e.g., the majority of the genes in Hedges 1994). The critical fossils in the Gauthier et al. (1988) analysis turned out to be synapsid fossils. However, the critical fossils were neither the oldest nor the youngest included in the analysis.

Huelsenbeck (1991) performed simulation studies to tease apart the relationship between the temporal position of a fossil and its completeness. All else being equal, including fossils that were more ancestral was more useful to phylogenetic analysis than adding living taxa. Also, all else being equal, fossils that were more complete were better than fossils that preserved only a small fraction of characters. Interestingly, there were many situations in which the inclusion of an ancestral fossil that preserved only a small fraction of the characters was better than the addition of a living taxon to the analysis. These findings were consistent with the conclusions of Gauthier et al. (1988) and of Donoghue et al. (1989)—that the importance of fossils in a phylogenetic

analysis is related to the combination of primitive and derived characters preserved in the fossils.

Fossils are important in phylogenetic analysis for what seems like the most obvious of reasons: fossils are remnants of the past, often the very distant past. Because of this, fossils often preserve unique combinations of characters that help resolve relationships. The purpose of this chapter is to discuss how the temporal uniqueness of fossils can be used in another manner: The temporal position of fossils can be used to test the fit of an estimated fossil to the rock record. Moreover, the stratigraphic position of fossils contains limited information on relationship. Although many readers may work with groups that have a poor fossil record or may work with molecular characters, for many problems it is important to consider the completeness of the stratigraphic record. For example, in several recent studies, the divergence times of major groups were examined using molecular data (for eubacteria/eukaryotes [Doolittle et al. 1996], metazoa [Wray et al. 1996], birds [Cooper and Penny 1997], mammalian orders and major lineages of vertebrates [Kumar and Hedges 1998]). In all of these studies, a molecular clock was assumed and at least one speciation time on the tree was considered known (to calibrate the clock). These researchers concluded that, for some major groups, the origin greatly predates the first fossils observed for the group. However, in none of these studies was uncertainty in the calibration times considered. To assess the uncertainty of the calibration times, models of fossil preservation are needed.

USING STRATIGRAPHIC DATA IN A PHYLOGENETIC ANALYSIS

Indices Measuring the Fit of Stratigraphy to Phylogeny

Several methods currently exist that are intended either to quantify the agreement of a phylogenetic tree with the stratigraphic record or to estimate phylogeny on the basis of information from the stratigraphic record. In this section, we focus on methods that measure the agreement between a phylogenetic tree and the stratigraphic record.

Spearman rank correlation. Gauthier et al. (1988) describe a method that investigates the degree of congruence between a phylogenetic tree and the stratigraphic record by comparing the “clade rank” with the “age rank” for each taxonomic group. The clade rank of a taxonomic group is defined as the number of nodes that must be traversed on a direct path from the root of the tree to the

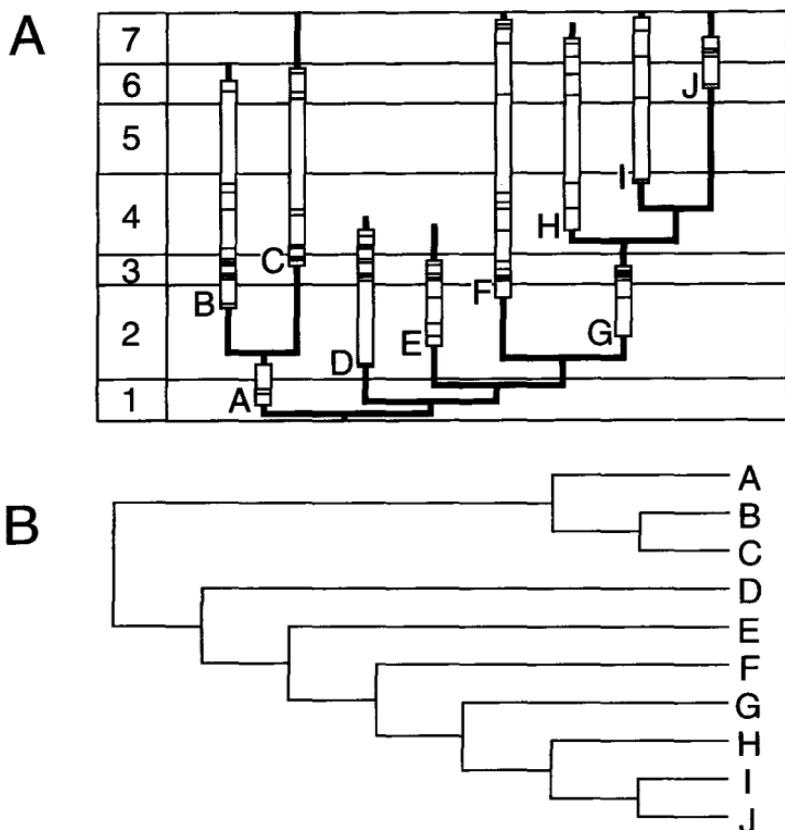


Figure 7.2. Sample data set. (A) The true evolutionary history of a group and its stratigraphic record. The true relationships and age ranges of the lineages are denoted by the thick solid lines. The observed age ranges are denoted by the rectangular boxes; within these boxes, the stratigraphic horizons at which a lineage is observed are denoted by dashes. (B) The cladogram that would be obtained if the relationships were correctly estimated.

branch-tip containing the group. The age rank of a taxonomic group is defined as the relative order of appearance of the group in the rock record (labeled from oldest to youngest). Gauthier et al. (1988) used Spearman's rank correlation (SRC; Sokal and Rohlf 1981) to evaluate the correlation of age and clade ranks for amniotes. The SRC coefficient, r_s , ranges in value from -1 to 1 and Gauthier et al. (1988) used it as a measure of the fit of the stratigraphic record to a phylogenetic tree. When $r_s = 1$, the order of appearance of groups agrees completely with their order of origination on the phylogeny, and when $r_s = -1$, these orderings are completely reversed. Figures 7.2 and 7.3 provide an example application of the method. Figure 7.2A depicts the actual evolutionary history of

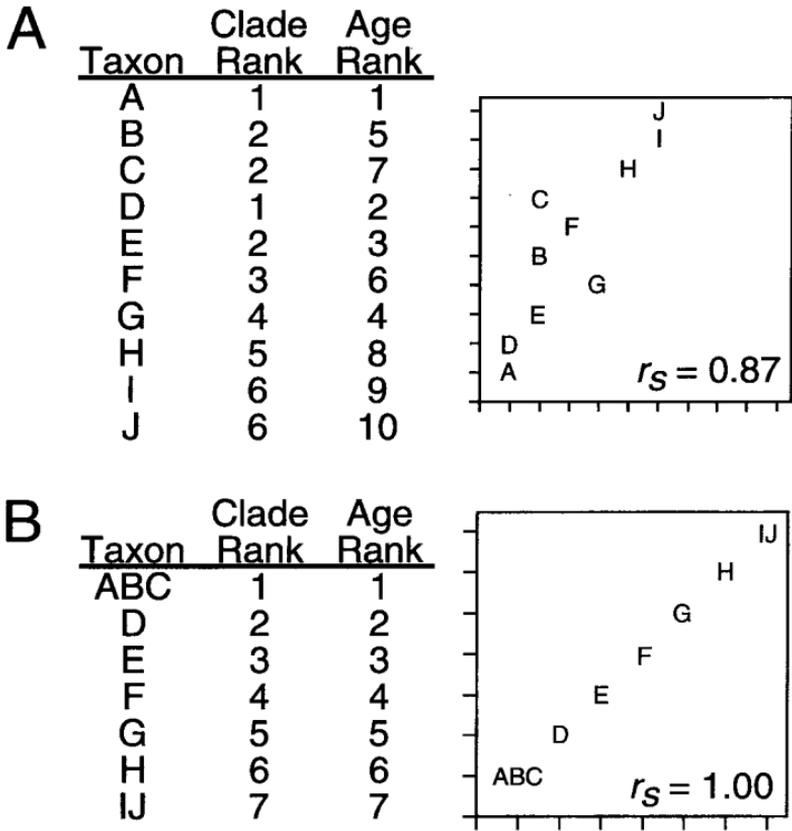


Figure 7.3. An example of Spearman rank correlation of clade and age ranks for the data from figure 7.2. (A) The relationship between clade and age rank according to the method of Gauthier et al. (1988). (B) The relationship between clade and age rank according to the method of Norell and Novacek (1992a).

a group of organisms, including the true age ranges and relationships of the species, as well as the observed age ranges. Figure 7.2B shows the cladogram that would result if the phylogeny of the 10 observed species were correctly estimated. Figure 7.3A shows the clade and age ranks of the 10 species, as well as the relationship between clade rank and age rank. For this example, $r_s = 0.87$.

The fact that the SRC coefficient is less than 1 for the tree in figure 7.2 indicates imperfect agreement between the phylogeny and the stratigraphic record. However, the imperfect fit results from the multiple clades of the same rank implied by the tree. Multiple clades of equal rank result when a tree is not maximally asymmetric (or pectinate) in shape. Figure 7.4 shows examples of maximally asymmetric and maximally symmetric trees. Norell and Novacek

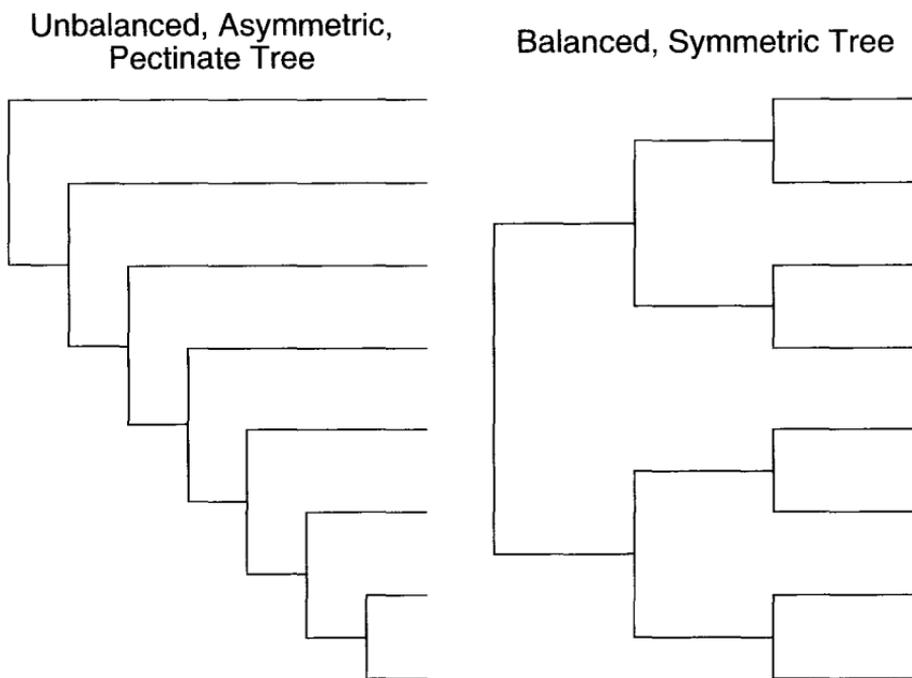


Figure 7.4. Examples of maximally balanced and unbalanced trees.

(1992a) modified the SRC approach to account for multiple clades of equal rank. Their approach is to pass through the tree, collapsing taxonomic groups with equal clade rank into a single taxonomic group and assigning that clade one rank. The oldest observed taxon from the collapsed group represents the age rank. Although this approach makes a correlation coefficient of 1 possible, it also discards potentially useful information because much of the stratigraphic information present in the collapsed clades is lost. Figure 7.3B illustrates clade and age ranks calculated by Norell and Novacek's (1992a) method. The SRC coefficient is designed to detect correlations between ordered pairs of observations. Although the ordered appearances of species in the fossil record may be viewed as observations, the clade ranks should not be treated as observations, as these are parameters inferred under an evolutionary model. As a consequence, the results for this index, at least as currently construed, do not have a rigorous statistical basis.

Stratigraphic consistency index. Huelsenbeck (1994) took a different approach to measuring the fit of the stratigraphic record to a phylogenetic tree. For every internal node of the phylogenetic tree (excluding the root node), the old-

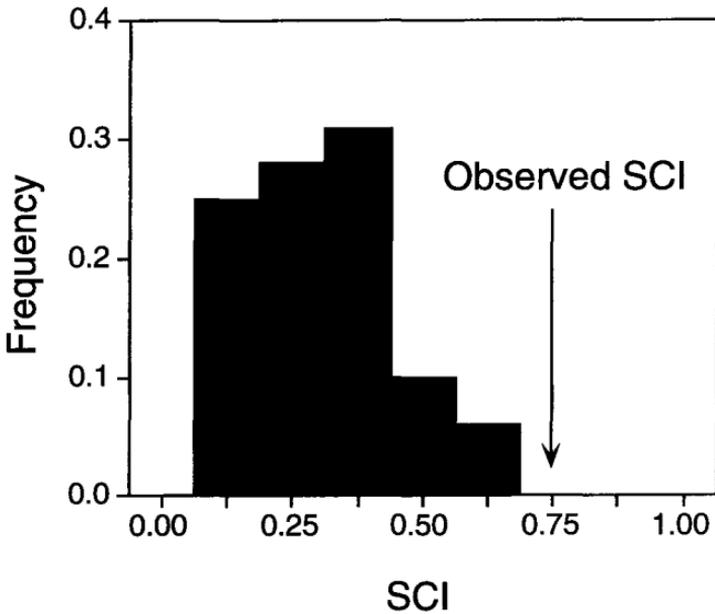


Figure 7.5. The distribution of stratigraphic consistency indices (SCI) expected under the null hypothesis of no association between phylogeny and the stratigraphic record for the data from figure 7.2.

est first occurrence in the fossil record of the species descended from the node is compared with the oldest first occurrence in the fossil record of its sister species. If the oldest first occurrence above the node is the same age as, or younger than, the oldest first occurrence of the sister taxon to the node, then the node is considered stratigraphically consistent. The stratigraphic consistency index (SCI) is intended to measure the proportion of stratigraphically consistent nodes.

The SCI varies between 0 and 1, and a low value may reflect disagreement between the fossil record and a phylogenetic tree or incomplete information on stratigraphic ranges. However, a high SCI value does not necessarily imply a good match between a phylogenetic tree and the stratigraphic record. For example, if all the fossil observations first occur at the same time instant, then $SCI = 1$. Huelsenbeck (1994) suggested that the observed SCI be compared to that expected under a null hypothesis of no association with the stratigraphic record for trees of the same shape; this can be done by permuting the ages among the tips of the tree many times. For each permutation, the SCI value is calculated. The proportion of the time that the observed SCI is exceeded in the permutations represents the probability of observing the SCI under the null hypothesis

of random agreement between the stratigraphic record and phylogeny. Figure 7.5 shows the distribution of SCI values for the example from figure 7.2. The observed SCI value of 0.75 would not be expected to occur under the null hypothesis of no association between the stratigraphic record and phylogeny ($P < 0.01$).

Measures of gaps. Several methods aim to evaluate the amount of missing time (or gaps) implied by a phylogeny. Benton (1987) proposed the simple completeness metric (SCM), which is the ratio of the number of time intervals that a specific taxon was observed to the minimum number of time intervals during which the taxon must have existed. For example, if a taxon was observed in time intervals 1, 2, 3, 6, 9, and 10, then $SCM = 0.6$ (i.e., the taxon must have existed during at least 10 time intervals but was observed in only 6). For the five fossil species shown in figure 7.6A, the SCM values are 0.53, 1.00, 0.75, 0.55, and 0.60 for species A, B, C, D, and E, respectively. Norell and Novacek (1992a, 1992b) proposed the minimum implied gaps (MIG) method, intended to evaluate the minimum amount of missing time on a tree. MIG is calculated as the sum of the missing time for each lineage and its sister taxon on the tree. Benton (1994) modified the MIG measure. His relative completeness index (RCI) is equal to one minus the ratio of the MIG to the simple range lengths (SRL), where the SRL is a sum of the observed duration of each fossil taxon included in the analysis. A problem with such gap measures is that they are purely descriptive. In the absence of a null model, it is not obvious what $SCM = 0.67$ implies about the processes of evolution and fossil preservation.

Estimating Phylogeny from the Stratigraphic Record

Stratocladistics. The methods described above aimed to measure the congruence between the stratigraphic record and a phylogenetic tree. Fisher's (1982, 1988, 1991, 1992, 1994) stratocladistic method is intended to provide an estimate of phylogeny based on stratigraphic and morphological data. The method is similar in spirit to the gap indices discussed above. However, the method uses the amount of missing time (i.e., non-preservations of a species in the fossil record) implied by a tree to choose among alternative phylogenies. Missing time is measured as the number of unobserved "stratum crossings"—or instances that a taxon should have been present in a time interval (as implied by the phylogeny) but was not observed—so that each phylogenetic tree has an integer number associated with it that represents the minimum number of unobserved stratum crossings implied by that tree. The amount of missing time implied by a tree is its "stratigraphic debt" (Fisher 1994) and the tree requiring

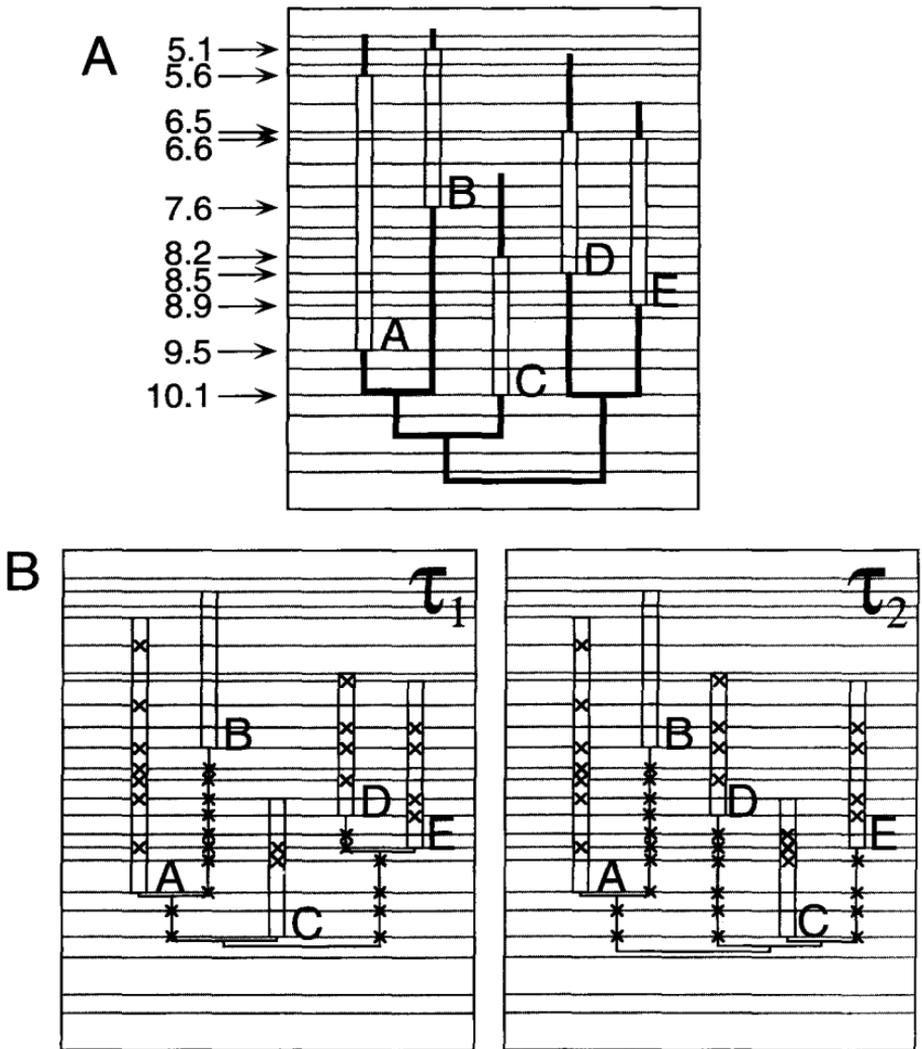


Figure 7.6. Consideration of stratigraphic debt. (A) The true genealogical relationships (thick solid lines) and observed stratigraphic ranges (rectangles and dashes therein) of five species. Numbers to the left represent millions of years. (B) Two alternative topologies, which have different levels of stratigraphic debt. Unobserved stratum crossings are denoted by "x."

the minimum amount of missing time is chosen as the best estimate of phylogeny in the absence of morphological information. An example calculation of stratigraphic debt is shown in figure 7.6B. Tree τ_1 implies a total of 33 unobserved stratum crossings, whereas tree τ_2 implies a total of 37 unobserved stratum crossings. On the basis of the stratigraphic record only, τ_1 is favored over

τ_2 . If instead of measuring debt in terms of unobserved stratum crossings one uses absolute time, τ_1 implies 4.1 million years of missing time, whereas τ_2 implies 5.3 million years of missing time.

Gap methods using implied first and last fossil occurrences. Wagner (1995) proposed a method for choosing among phylogenetic trees that uses information from the first and last observed occurrence of a fossil species, as well as the total number of occurrences for each species. Wagner's method uses the 95% confidence limits on the first and last occurrences to find the tree that implies the fewest gaps (using the unobserved confidence intervals included in the range of a species). This tree is chosen as the best estimate of phylogeny. For a single taxon, the confidence intervals for the true first and last occurrence may be calculated by assuming that fossil preservation can be described as a Poisson process with rate λ (Strauss and Sadler 1989).

Maximum Likelihood Estimation of Phylogeny with Stratigraphic Data

Maximum likelihood is a standard statistical method for estimating the parameters of a model by using observed data generated under the model. An example of a maximum likelihood estimate is the familiar sample mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

which provides an estimate of the true mean, the parameter μ , on the basis of n independent observations, $\mathbf{x} = \{x_1, \dots, x_n\}$, from a normal distribution, $N(\mu, \sigma^2)$. In an earlier study, we showed that a maximum likelihood approach can be used to estimate phylogenetic trees from stratigraphic data (Huelsenbeck and Rannala 1997). Fossil preservation is assumed to follow a Poisson process with preservation rate λ , as in Strauss and Sadler's (1989) method for establishing confidence intervals on the stratigraphic range of a single species. The maximum likelihood method uses information on the first and last occurrences of fossil taxa, as well as the total number of fossil occurrences for each taxon. The maximum likelihood approach is not restricted to a Poisson preservation model, however, and can be used with any stochastic model of fossil preservation. Likelihood ratio tests can be used to evaluate the relative fit of different models of fossil preservation to the observed data.

Models of fossil preservation. Several models of fossil preservation have been proposed. Here we review some of these models and discuss how they

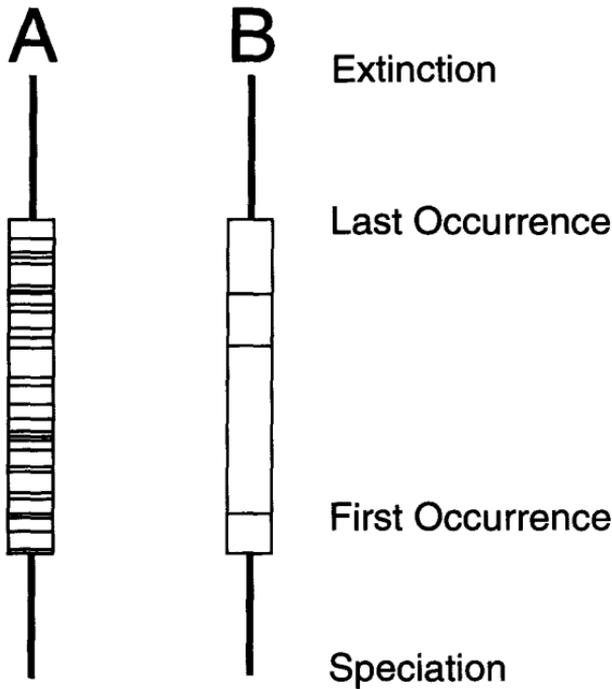


Figure 7.7. An example of two species with the same true range (the time between speciation and extinction) and the same observed range (the time between the first and last occurrences). However, species A was observed 30 times (including the first and last occurrence), whereas species B was observed only 5 times. Under a Poisson process of preservation, in which the preservation probability in any small interval of time, dt , is λdt , the probability of observing the data for species B is higher than that for species A if the preservation rate is the same for both species.

might be applied in a phylogenetic analysis using maximum likelihood methods. A Poisson process model of preservation assumes that in any instant of time, dt , the probability of a fossilization event occurring is λdt . The intuition underlying this model of fossil preservation is shown in figure 7.7. Here, two species have exactly the same speciation and extinction times. The observed first and last occurrence of each species is also the same. Species A has six times the number of observed preservation events as does species B, however. Hence, the probability of the actual taxon range is higher for species B than for species A; we would not expect such large gaps with no preservations for species A

(as is the case at the ends of its range) because the larger number of observed preservations implies a high preservation rate for A.

The model of preservation specifies the probability of observing the fossil occurrences of species, given the actual speciation and extinction times and the preservation rate. This is denoted as $f(o_f, o_l, n | t_f, t_l, \lambda)$, where o_f is the observed first occurrence, o_l is the observed last occurrence, n is the number of occurrences, t_f is the time at which the species first arose, t_l is the time at which the species went extinct, and λ is the preservation rate. Conditional on t_f and t_l , the distribution of fossil preservations is uniform on the interval $[t_f, t_l]$. The joint probability distribution of the first and last fossil observations is the joint distribution of the smallest and largest order statistics of a uniform distribution:

$$f(o_f, o_l | n, t_f, t_l) = \left(\frac{1}{(t_l - t_f)} \right)^n n(n-1)(o_l - o_f)^{n-2}.$$

The probability distribution $f(n | t_f, t_l, \lambda)$ under the model is Poisson with parameter $\lambda(t_l - t_f)$ and the joint distribution $f(o_f, o_l, n | t_f, t_l, \lambda)$ is then

$$f(o_f, o_l, n | t_f, t_l, \lambda) = f(o_f, o_l | n, t_f, t_l) f(n | t_f, t_l, \lambda) = \frac{(o_l - o_f)^{n-2} \lambda^n e^{-\lambda(t_l - t_f)}}{(n-2)!}.$$

This equation gives the probability of a particular set of first and last occurrences, o_f and o_l , and a total of n occurrences, given the actual speciation and extinction times, t_f and t_l , and the preservation rate, λ . This equation assumes that $n \geq 2$. Huelsenbeck and Rannala (1997) provide formulae for the other two possible cases ($n = 1$ and $n = 0$).

The Poisson process model of fossil preservation, as formulated above, assumes that fossils can be dated with a high degree of accuracy and that the fossil preservation rate is constant through time and does not vary among lineages. Although the constraints of a constant preservation rate through time and of no rate variation among lineages can be relaxed, most paleontological data are not collected with accurate enough time resolution to implement the Poisson process model as described. Often, the paleontologist will have only partial information on the time intervals between fossil observations. We consider three cases in which the relative amount of information about the temporal distribution of fossils differs: (1) the duration of the time interval between observed first and last occurrences, o_f and o_l , and the number of observations in between the first and last occurrences are known, (2) the duration of each time interval (such as the geological period or epoch) and the presence or absence of a fossil lineage in each time interval are known ($n = 0$ or $n \geq 1$), and (3) the durations of the time intervals are roughly equal in length and the presence or absence of a fos-

sil lineage in each time interval is known. The duration of the i th time interval is denoted d_i . In the first case, the information on the preservation of a lineage may be represented as a vector, $\mathbf{n} = \{n_i\}$, where n_i represents the number of observations of the lineage in the i th time interval (there are I time intervals in total). As an example, if a lineage was thought to exist during 11 time intervals ($I = 11$), the data might be coded $\mathbf{n} = \{0, 0, 2, 0, 0, 5, 0, 10, 112, 9, 0\}$. In this case, there are no observations of the lineage during intervals 1 and 2, two observations during interval 3, and so on. For the second and third cases, only the presence or absence of a species in each time interval is known. Hence, the stratigraphic data is coded as $\mathbf{n} = \{n_i\}$, where n_i is now a binary (0, 1) random variable. An example of preservation data coded in this form is $\mathbf{n} = \{0, 0, 1, 0, 0, 1, 0, 1, 1, 1, 0\}$, where 1 denotes the presence of the lineage in the i th time interval and 0 denotes its absence.

If the duration of the i th interval is known (d_i), then the probability under a Poisson model of preservation of observing n_i fossil horizons is

$$f(n_i | \lambda, d_i) = \frac{e^{-\lambda d_i} (\lambda d_i)^{n_i}}{n_i!},$$

where λ is the preservation rate. If only the presence or absence of a fossil lineage in a time interval is known (i.e., $n_i = 0$ or $n_i = 1$), then the probability of n_i under a Poisson model of preservation is

$$f(n_i | \lambda, d_i) = n_i + (-1)^{n_i} e^{-\lambda d_i}.$$

The probability of observing the distribution of fossil preservations among the I intervals is the product of the probabilities of observing the data in each interval, because the intervals are independent under a Poisson process:

$$f(\mathbf{n} | \lambda, \mathbf{d}) = \prod_{i=1}^I f(n_i | \lambda, d_i).$$

Foote and Raup (1996) considered a model of fossil preservation in which the probability of preservation during each time interval is p , regardless of the duration of the time interval. The probability of observing a given distribution of fossil presences ($n_i = 1$) or absences ($n_i = 0$) during I time intervals is then

$$\prod_{i=1}^I p^{n_i} (1-p)^{1-n_i}.$$

Foote and Raup (1996) derive the probability of observing a taxon range of t time intervals given that the true range was I intervals:

$$f(t|I) = \begin{cases} (1-p)^I, & t=0 \\ T(1-p)^{(I-1)}p, & t=1. \\ (I-t+1)(1-p)^{(I-t)}p^2, & t>1 \end{cases}$$

Maximum likelihood estimation. Assuming independent preservation among lineages, the likelihood of a specific phylogenetic tree is the product of the probabilities of observing the data for each of the s branches of a tree. If preservations among lineages are independent and follow a Poisson process, the likelihood is

$$L(\tau, \tilde{t}_f, \tilde{t}_p, \lambda | \tilde{q}_f, \tilde{q}_p, \tilde{n}) = \prod_{i=1}^s f(o_f^i, o_l^i, n^i | t_f^i, t_l^i, \lambda),$$

where the product is over the s lineages of a phylogenetic tree (τ), and

$$\begin{aligned} \tilde{q}_f &= \tilde{q}_f^\lambda, \dots, o_f^s, \\ \tilde{q}_l &= o_l^\lambda, \dots, o_l^s, \\ \tilde{t}_f &= \tilde{t}_f^\lambda, \dots, \tilde{t}_f^s, \\ \tilde{t}_l &= \tilde{t}_l^\lambda, \dots, \tilde{t}_l^s, \text{ and} \\ \tilde{n}_o &= n_o^\lambda, \dots, n_o^s. \end{aligned}$$

The parameters of the model (branch lengths, topology, and preservation rate) are estimated by maximizing the likelihood function. The maximum likelihood tree maximizes the probability of observing the fossil data given the model of fossil preservation. The likelihood for the other models of preservation described above can be calculated similarly (i.e., by taking the product of the probabilities of observing the data over the s lineages).

As an example calculation of the likelihood of possible phylogenetic trees for stratigraphic data, consider the two trees of figure 7.6B. Table 7.1 shows the

Table 7.1

The observations (o_f , o_l , and n) and maximum likelihood estimates of speciation and extinction times (t_f and t_p , respectively) for tree τ_1 of figure 7.6B

Lineage	o_f	o_l	n	t_f	t_l
A	9.5	5.6	8	9.5	5.6
B	7.6	5.1	9	9.5	5.1
C	10.1	8.2	6	10.1	8.2
D	8.5	6.5	5	8.9	6.5
E	8.9	6.6	6	8.9	6.6
G ₁	—	—	0	10.1	9.5
G ₂	—	—	0	10.1	10.1
G ₃	—	—	0	10.1	8.9

Tree τ_1 implies three ghost taxa, denoted G₁, G₂, and G₃.

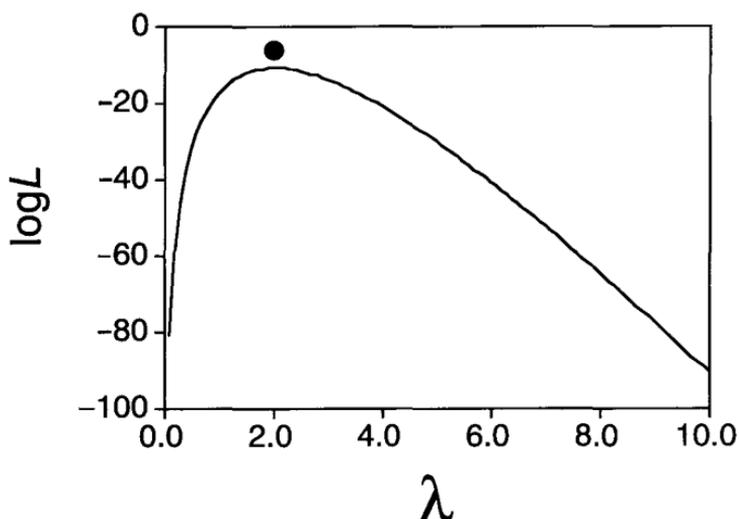


Figure 7.8. The log likelihood as a function of preservation rate (λ) for tree τ_1 from figure 7.6B. The maximum likelihood estimate of the fossil preservation rate is 2.02 (denoted by the large dot).

maximum likelihood estimates of speciation and extinction times for τ_1 , as well as the information on the times of first and last occurrence and the total number of occurrences for each species. Note that tree τ_1 implies that there are three unobserved lineages (“ghost taxa” in Norell’s [1992] terminology). The log likelihood, $\log L$, of τ_1 is -10.72 , whereas for τ_2 it is -12.88 . Like the stratocladistic method, the maximum likelihood criterion favors τ_1 . Although the maximum likelihood method provides an estimate of the phylogenetic tree and speciation and extinction times on that tree, it also provides an estimate of the preservation rate, λ . The maximum likelihood estimate of λ changes slightly depending on which tree is assumed. The estimate of λ is 2.02 for τ_1 and 1.90 for τ_2 (λ is in units of expected number of preservation events per million years). Figure 7.8 shows how the log likelihood changes as a function of λ for τ_1 .

COMPARISON OF STRATIGRAPHIC METHODS

Most methods considered here for measuring the fit of the stratigraphic record to a phylogenetic tree are simple indices. These can be useful, but it is often not clear how a specific index value should be interpreted. For example, the SCI, by itself, is not very useful; that is, if someone were to state that the SCI for a particular data set was 0.5, should this be interpreted as a good or poor fit of the stratigraphic record to a tree? The same problem exists for other measures of stratigraphic fit (e.g., SRC, MIG, RCI). A potential solution is to compare the

observed fit (using SCI, SRC, MIG, RCI, or some other method) to that expected if there were no correspondence between phylogeny and stratigraphy. One way to do this is to permute the observed stratigraphic ages among the tips of a phylogenetic tree (Huelsenbeck 1994). For each permutation, the index is recalculated and the proportion of the time that the observed value is exceeded in the permutations estimates the probability that the observed value would be generated given a random association of phylogeny with stratigraphy. Unfortunately, this approach has rarely been used to evaluate data.

Another problem with the various indices we have described is that their values may be strongly influenced by the shape of a tree (Siddall 1996). For example, for the SCI, maximally asymmetric trees can be made to more easily fit the stratigraphic record than can symmetrical trees. Caution must therefore be exercised when comparing indices among trees that have different shapes.

The remaining approaches that we have discussed—stratocladistics, Wagner's method, and the maximum likelihood approach—are all intended to estimate phylogenetic trees. Simulation analysis could be used to evaluate which of the methods performs best in estimating phylogeny. Such studies have not yet been performed, although we have investigated the statistical properties of the maximum likelihood method by means of simulations (Huelsenbeck and Rannala 1997). The maximum likelihood and stratocladistic methods should have similar performance because there is a strong relationship between the two methods; under a Poisson process of fossil preservation, the likelihood for a given phylogenetic tree is maximized when the amount of missing time implied by the tree is minimized (Huelsenbeck and Rannala 1997), and stratocladistics also seeks to minimize missing time. Apart from the accurate estimation of phylogeny, other criteria are also important in evaluating methods. One of the most important criteria is the ability to extend a method to address additional questions that may depend on phylogeny. Almost all of the methods discussed here can be modified to address additional questions concerning such processes as fossil preservation, support for different tree rootings, and the degree of phylogenetic agreement among diverse forms of data. However, maximum likelihood is unique among the methods so far developed in providing a unified set of procedures for testing hypotheses through the use of likelihood ratio tests.

COMBINING DATA

How can one best combine different sources of information, such as morphological data, stratigraphic data, or DNA sequences, to obtain an estimate of phylogeny? Fisher (1992) argued that the best overall estimate of phylogeny is the tree with the minimum overall length, as determined by adding the minimum

number of character transformations implied by a tree (under a maximum parsimony criterion) to the minimum number of stratum crossings implied by that same tree overlaid on the fossil record. Fisher's method then allows various sources of data (morphological, molecular, and stratigraphic) to be combined in a phylogenetic analysis. MacClade 3.0 (Maddison and Maddison 1992) can be used to calculate stratigraphic debt for user-defined trees. Gingerich (1979) suggested another possible means of combining morphological and stratigraphic data—stratophenetics. The approach relies on the fine-scale resolution of the stratigraphic record and morphological comparisons among fossil taxa. Working from the base of the stratigraphic record upward, one views phenetically similar samples in different strata as being ancestors and descendants. Finally, the maximum likelihood method (Huelsenbeck and Rannala 1997) provides an alternative means of combining stratigraphic, morphological, and molecular data. If these data are independent, then their likelihoods, for a particular phylogeny, can be multiplied. As mentioned previously, the calculation of the likelihood for stratigraphic data under some simple models of fossil preservation is computationally simple. Similarly, calculating the likelihood of a tree for molecular data, such as DNA sequences, is easily done with existing methods (Felsenstein 1981a). Calculating the likelihood of morphological data is more difficult because few reasonable models exist for morphological evolution. A likelihood analysis can be performed for continuously varying characters by assuming a Brownian motion model of evolution (e.g., using CONTML; Felsenstein 1981b, 1995). This model has been applied to discrete morphological data with the traits treated as frequencies (Wiens 1998) and appears to behave similarly to frequency parsimony. Moreover, simulations suggest that CONTML performs well under a variety of conditions (Wiens and Servedio 1998).

The advantage of maximum likelihood methods over stratigraphic parsimony is that the weights given to the various sources of data are (1) proportional to their probabilities under the evolutionary models and (2) automatic, requiring no hidden assumptions on the part of the user (apart from the explicit assumptions specifying the model under which the analysis is performed). By contrast, the weight assigned to stratigraphic data by the stratocladistic method can be arbitrarily increased or decreased by considering more or less finely resolved stratigraphic units and has no obvious relationship to the probability of observing the data.

LIKELIHOOD RATIO TESTS WITH STRATIGRAPHIC DATA

There are a number of questions in paleontology that likelihood ratio tests can address. The ratio of the likelihoods calculated under two hypotheses,

$$\Lambda = \frac{\max[L(\text{Null Model} \mid \text{Data})]}{\max[L(\text{Alternative Model} \mid \text{Data})]},$$

is a measure of the tenability of one hypothesis over the other. Typically, likelihood ratio tests are constructed in such a way that the hypotheses considered are nested (i.e., the null hypothesis is a special case of the alternative hypothesis). In the case of nested hypotheses, $-2\log L$ is χ^2 distributed with q degrees of freedom, where q is the number of additional parameters under the alternative hypothesis. For many problems in phylogenetics, even though the hypotheses considered may appear to be nested, the χ^2 distribution does not closely approximate the null distribution. Monte Carlo simulation (or parametric bootstrapping) must then be used to approximate the null distribution of $-2\log L$ (Cox 1961, 1962; Felsenstein 1988; Goldman 1993). The parametric bootstrap procedure uses maximum likelihood estimates of model parameters, obtained under the null hypothesis, to simulate many data sets of the same size as the original, under the assumption that the null hypothesis (and estimated parameters) are true. For each simulated data set, the test statistic $-2\log L$ is calculated. The proportion of the time that the original value of $-2\log L$ is exceeded in the simulated data estimates the significance level of the test.

Congruence of Trees Estimated from Diverse Data

One possible application of likelihood ratio tests is to evaluate the congruence of phylogenetic trees estimated from diverse forms of data (Huelsenbeck and Bull 1996). For example, two possible types of data considered in paleontology are morphological characters for particular taxa and their stratigraphic occurrences. Alternatively, one might compare molecular data, such as DNA sequences, with stratigraphic occurrences. There are several reasons why one would want to test the congruence of different data partitions (see Huelsenbeck and Bull 1996). One reason is that one may want to evaluate the congruence of the stratigraphic data with an estimate of phylogeny based on morphological or molecular data. The null hypothesis assumes that the same phylogeny underlies both data sets. The alternative hypothesis relaxes the constraint that the same phylogeny underlies both data partitions. (This does not literally mean that different trees underly each data partition; rather, it is used to ask whether the different data sets are making significantly different statements about phylogeny.) The significance of the likelihood ratio test statistic, $-2\log L$, is determined through parametric bootstrapping (Huelsenbeck and Bull 1996). Rejection of the null hypothesis suggests that the data partitions support different phylogenetic trees and that this support is greater than would be expected from sampling error alone.

Often it will be difficult to perform the test of congruence as described above. For example, the likelihood of the morphological data may often be difficult to calculate due to the lack of an appropriate model describing the evolution of morphological characters. An alternative test might consider the fit of stratigraphic data to a tree constructed on the basis of morphology. This test assumes that the tree based on morphology is known without error. The null hypothesis assumes that this tree is correct and calculates the likelihood of the stratigraphic data given this tree. The likelihood of the stratigraphic data under the alternative hypothesis is calculated by allowing topology to be a free parameter; that is, the likelihood under the alternative hypothesis is the tree that maximizes the likelihood for the stratigraphic data. The null distribution of the likelihood ratio test statistic can be determined by parametric bootstrapping.

Other tests examining the concordance of a phylogenetic tree with the stratigraphic record that do not use likelihood ratios have been suggested (using Spearman's rank correlation—Gauthier et al. 1988, Norell and Novacek 1992a; using SCI—Huelsenbeck 1994; using SRC—Benton 1994). Very little can be said about the performance of these tests because no studies (e.g., simulations) have been done to evaluate the relative performance of these tests. However, it is known that likelihood ratio tests often perform well even when the hypotheses considered are composite (Rice 1995).

Testing Models of Fossil Preservation

Models of fossil preservation are important for several reasons. First, these models provide a basis for constructing confidence intervals of fossil species ranges over time. In fact, a statistical test for a mass extinction event has been proposed which relies on a Poisson process model of fossil preservations (Solow 1996). A model of fossil preservation also underlies the recent method of Foote (1997) for estimating extinction and preservation rates. Second, models of fossil preservation provide important information about the rate at which fossils are preserved. The field of taphonomy is concerned with the process of fossil preservation and a model-based approach might be used in this area to ask more detailed questions concerning the nature of the process of fossil preservation.

In as much as models of fossil preservation are important, it is also important to be able to improve preservation models. To test the fit of stratigraphic data to one model of fossil preservation versus another, one can use a likelihood ratio test. More complex models of fossil preservation than the Poisson process can be studied by introducing additional parameters into a model. For example, one could designate the Poisson process model typically used by paleontologists as the null model (Strauss and Sadler 1989). This model assumes that preservation rates are constant through time and identical across lineages. An

alternative hypothesis might relax the constraint that preservation rate is constant through time or that rates are identical across lineages. Additional preservation parameters could then be estimated for different time intervals or for different lineages. The likelihood ratio test statistic calculated for the null and alternative models, $-2\log L$, is then compared to a χ^2 distribution with the appropriate degrees of freedom. The importance of the likelihood ratio test for models of fossil preservation is that it provides a framework for the gradual improvement of our current models.

ROOTING PHYLOGENETIC TREES

A powerful use of stratigraphic data is for rooting phylogenetic trees. Figure 7.9 shows an example of an unrooted four-taxon tree and the five possible rootings of the tree. The tree can be rooted along any one of the five branches. Phylogenies are most often rooted according to the outgroup criterion; species that

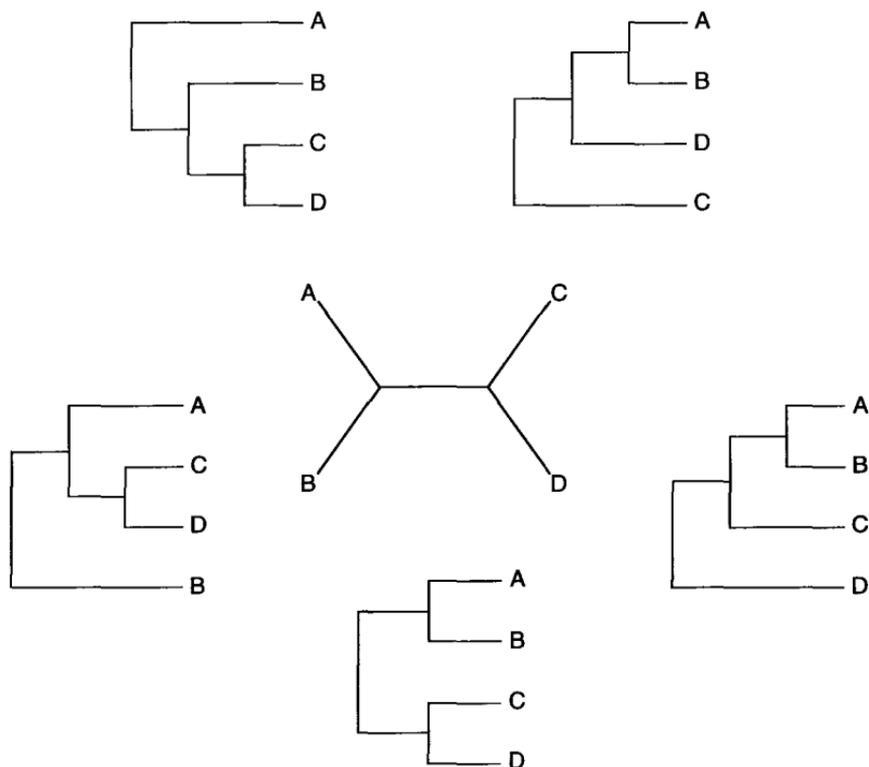
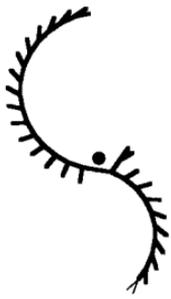
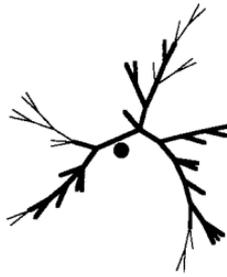


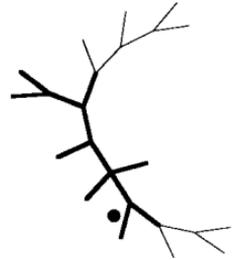
Figure 7.9. The unrooted and rooted trees for four species. The rooted trees are derived from the unrooted tree by considering each branch of the unrooted tree as a potential root.



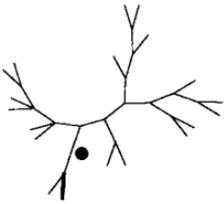
Amniota



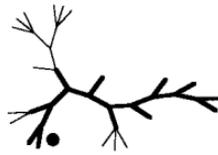
Ostreidae



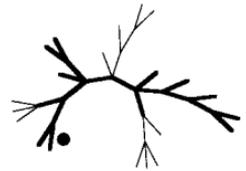
Diplograptidae



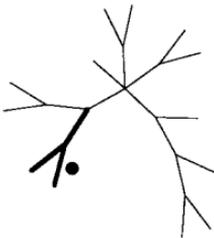
Monograptidae



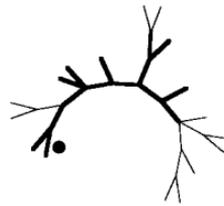
Orthograptidae



Taperioids



Eleutherozoa



Peccaries

Figure 7.10. The branches chosen by the outgroup criterion (thin lines, rooted at the dot) and by the stratigraphic criterion (thick lines) on eight trees. Morphological data and analyses based on the outgroup criterion are from the following sources: Amniota, Gauthier et al. (1988); Ostreidae, Huelsenbeck (1992); Diplograptidae, Monograptidae, and Orthograptidae, Mitchell (1987); taperioids, Schoch (1989); Eleutherozoa, Smith (1988); peccaries, Wright (1993). Reanalyses of these data sets according to the stratigraphic criterion are from Huelsenbeck (1994).

are considered to be outside of the group of interest are included in the analysis to provide a root. Methods that do not rely on the outgroup criterion can also be used for rooting trees. In particular, rooted trees can be obtained from molecular data when these data fit a molecular-clock model. Methods have been proposed for polarizing characters on the basis of ontogeny or other criteria (see de Queiroz [1985] or Bryant [1991] for reviews of character polarity criteria). The stratigraphic record can also be used to root a phylogenetic tree (Huelsenbeck 1994). The stratigraphic rooting criterion considers each possible rooting position for an unrooted phylogenetic tree. The root position that provides the best fit to the stratigraphic record (under some optimality criterion) can be chosen as the root of the tree.

How well does the stratigraphic record root phylogenetic trees? Figure 7.10 shows eight trees that Huelsenbeck (1994) evaluated by using Fisher's (1992) stratocladistic method to estimate the optimal root(s). The root(s) indicated by the stratigraphic data were very close (or identical) to the root chosen by applying the outgroup criterion. Although the probability that the outgroup rooting position falls along the same region of the tree is often very high because of the large number of branches indicated as possible rooting positions by the stratigraphic criterion, when all of the studies are taken together, the concordance is remarkable. The congruence of the rooting position chosen by two independent methods strengthens our confidence of the accuracy of both methods for rooting trees. The tree root may also be chosen by using the maximum likelihood criterion (see Huelsenbeck and Rannala 1997).

CONCLUSIONS

Although stratigraphic data have been incorporated into phylogenetic analyses since at least the 1860s, such information is only infrequently considered by modern systematists. Quite apart from phylogenetic information arising from unique combinations of character states (Gauthier et al. 1988), fossils provide an additional source of information in phylogenetics. Minimally, stratigraphic data can be used to (1) estimate phylogeny, (2) examine the congruence between the stratigraphic record and phylogeny, (3) examine models of fossil preservation, and (4) root phylogenetic trees. In this chapter, we have attempted to outline some of the approaches that have been suggested for incorporating stratigraphic data into a phylogenetic analysis. We show that all of the fossil applications outlined above can be performed by using a unified set of tools arising from the well-known statistical procedure of likelihood analysis. It is our hope that this discussion will motivate systematists to further explore the potential uses of stratigraphic data and likelihood methods in phylogenetic analysis.

ACKNOWLEDGMENTS

This research was supported by a Miller postdoctoral fellowship awarded to J.H. and a Natural Sciences and Engineering Research Council (NSERC) of Canada postdoctoral fellowship awarded to B.R.

REFERENCES

- Benton, M. J. 1987. Mass extinctions among families of non-marine tetrapods: The data. *Mémoires de la Société Géologique de France, Nouvelle Série* 150: 21–32.
- Benton, M. J. 1994. Paleontological data and identifying mass extinctions. *Trends in Ecology and Evolution* 9: 181–185.
- Benton, M. J. 1995. Testing the time axis of phylogenies. *Philosophical Transactions of the Royal Society of London B* 349: 5–10.
- Benton, M. J., and G. W. Storrs. 1996. Diversity in the past: Comparing cladistic phylogenies and stratigraphy. Pages 19–40 *in* *Aspects of the Genesis and Maintenance of Biological Diversity* (ed. M. E. Hochberg, J. Clobert, and R. Barbault). Oxford University Press, Oxford.
- Brochu, C. A. 1997. Morphology, fossils, divergence timing, and the phylogenetic relationships of *Gavialis*. *Systematic Biology* 46: 479–522.
- Bryant, H. 1991. The polarization of character transformations in phylogenetic systematics: Role of axiomatic and auxiliary assumptions. *Systematic Zoology* 40: 433–445.
- Cooper, A., and D. Penny. 1997. Mass survival of birds across the Cretaceous–Tertiary boundary: Molecular evidence. *Science* 275: 1109–1113.
- Cox, D. R. 1961. Tests of separate families of hypotheses. *Proceedings of the Fourth Berkeley Symposium*. University of California Press, Berkeley.
- Cox, D. R. 1962. Further results on tests of separate families of hypotheses. *Journal of the Royal Society of London B* 24: 406–424.
- Darwin, C. 1859. *On the Origin of Species by Means of Natural Selection*. J. Murray, London.
- de Queiroz, K. 1985. The ontogenetic method for determining character polarity and its relevance to phylogenetic systematics. *Systematic Zoology* 34: 280–299.
- Donoghue, M. J., J. A. Doyle, J. Gauthier, A. G. Kluge, and T. Rowe. 1989. The importance of fossils in phylogeny reconstruction. *Annual Review of Ecology and Systematics* 20: 431–460.
- Doolittle, F. F., D.-F. Feng, S. Tsar, G. Cho, and E. Little. 1996. Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science* 271: 470–477.
- Doyle, J. A., and M. J. Donoghue. 1987. The importance of fossils in elucidating seed plant phylogeny and macroevolution. *Review of Paleobotany and Palynology* 50: 63–95.
- Felsenstein, J. 1981a. Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution* 17: 368–376.

- Felsenstein, J. 1981b. Evolutionary trees from gene frequencies and quantitative characters: Finding maximum likelihood estimates. *Evolution* 35: 1229–1241.
- Felsenstein, J. 1988. Phylogenies from molecular sequences: Inference and reliability. *Annual Review of Genetics* 22: 521–565.
- Felsenstein, J. 1995. PHYLIP (Phylogeny Inference Package), version 3.57c. Distributed by the author. Department of Genetics, University of Washington, Seattle.
- Fisher, D. C. 1982. Phylogenetic and macroevolutionary patterns within the Xiphosurida. *North American Paleontological Convention III, Proceedings* 1: 175–180.
- Fisher, D. C. 1988. Stratocladistics: Integrating stratigraphic and morphologic data in phylogenetic inference. *Geological Society of America, Abstracts with Program* 20: A186.
- Fisher, D. C. 1991. Phylogenetic analysis and its application in evolutionary paleobiology. Pages 103–122 in *Analytical Paleobiology* (ed. N. L. Gilinsky and P. W. Signor). *Short Courses in Paleontology*, No. 4. Paleontological Society.
- Fisher, D. C. 1992. Stratigraphic parsimony. Pages 124–129 in *MacClade: Analysis of Phylogeny and Character Evolution* (ed. W. P. Maddison and D. R. Maddison). Sinauer Associates, Sunderland, Mass.
- Fisher, D. C. 1994. Stratocladistics: Morphological and temporal patterns and their relation to phylogenetic process. Pages 133–171 in *Interpreting the Hierarchy of Nature—from Systematic Patterns to Evolutionary Theories* (ed. L. Grande and O. Rieppel). Academic Press, Orlando, Fla.
- Foote, M. 1997. Estimating taxonomic durations and preservation probability. *Paleobiology* 23: 278–300.
- Foote, M., and D. M. Raup. 1996. Fossil preservation and the stratigraphic ranges of taxa. *Paleobiology* 22: 121–140.
- Forey, P. L. 1986. Relationships of lungfishes. *Journal of Morphology, Supplement* 1: 75–91.
- Gaudry, A. 1862. Animaux fossiles et géologie de l'attique d'après les recherches faites en 1855–56 et 1860 sous les auspices de l'Académie des Sciences, Paris. F. Savy, Paris.
- Gauthier, J., A. G. Kluge, and T. Rowe. 1988. Amniote phylogeny and the importance of fossils. *Cladistics* 4: 105–209.
- Gingerich, P. D. 1979. The stratophenetic approach to phylogeny reconstruction in vertebrate paleontology. Pages 41–77 in *Phylogenetic Analysis and Paleontology* (ed. J. Cracraft and N. Eldredge). Columbia University Press, New York.
- Goldman, N. 1993. Statistical tests of models of DNA substitution. *Journal of Molecular Evolution* 36: 182–198.
- Harper, C. 1976. Phylogenetic inference in paleontology. *Journal of Paleontology* 50: 180–193.
- Hedges, S. B. 1994. Molecular evidence for the origin of birds. *Proceedings of the National Academy of Sciences, U.S.A.* 91: 2621–2624.
- Hennig, W. 1981. *Insect Phylogeny*. Wiley, New York.
- Huelsenbeck, J. P. 1991. When are fossils better than extant taxa in phylogenetic analysis? *Systematic Zoology* 40: 458–469.

- Huelsenbeck, J. P. 1992. Oyster phylogeny: Fossils and confidence. Master's thesis, University of Texas, Austin.
- Huelsenbeck, J. P. 1994. Comparing the stratigraphic record to estimates of phylogeny. *Paleobiology* 20: 470–483.
- Huelsenbeck, J. P., and J. J. Bull. 1996. A likelihood ratio test to detect conflicting phylogenetic signal. *Systematic Biology* 45: 92–98.
- Huelsenbeck, J. P., and B. Rannala. 1997. Maximum likelihood estimation of phylogeny using stratigraphic data. *Paleobiology* 23: 174–180.
- Jefferies, R. P. S. 1979. The origin of chordates: A methodological essay. Pages 443–477 in *The Origin of Major Invertebrate Groups* (ed. M. R. House). Academic Press, London.
- Krings, M., A. Stone, R. W. Schmitz, H. Krainitzki, M. Stoneking, and S. Paabo. 1997. Neandertal DNA sequences and the origin of modern humans. *Cell* 90: 10–30.
- Kumar, S., and S. B. Hedges. 1998. A molecular timescale for vertebrate evolution. *Nature* 392: 917–920.
- Maddison, W. P., and D. R. Maddison. 1992. *MacClade: Analysis of Phylogeny and Character Evolution*, version 3.0. Sinauer Associates, Sunderland, Mass.
- Marshall, C. R. 1990. Confidence intervals on stratigraphic ranges. *Paleobiology* 16: 1–10.
- Mitchell, C. E. 1987. Evolution and phylogenetic classification of the Diplograptacea. *Paleontology* 30: 353–405.
- Norell, M. 1992. Taxic origin and temporal diversity: The effect of phylogeny. Pages 89–118 in *Extinction and Phylogeny* (ed. M. J. Novacek and Q. D. Wheeler). Columbia University Press, New York.
- Norell, M., and M. Novacek. 1992a. The fossil record and evolution: Comparing cladistic and paleontologic evidence for vertebrate history. *Science* 255: 1690–1693.
- Norell, M., and M. Novacek. 1992b. Congruence between superpositional and phylogenetic patterns with fossil records. *Cladistics* 8: 319–337.
- Patterson, C. 1981. Significance of fossils in determining evolutionary relationships. *Annual Review of Ecology and Systematics* 12: 195–223.
- Rice, J. A. 1995. *Mathematical Statistics and Data Analysis*. Duxbury Press, Belmont, Calif.
- Rosen, D. E., P. L. Forey, B. G. Gardiner, and C. Patterson. 1981. Lungfishes, tetrapods, paleontology, and plesiomorphy. *Bulletin of the American Museum of Natural History* 167: 159–267.
- Schoch, R. M. 1989. A review of the taperioids. Pages 298–320 in *The Evolution of Perisodactyls* (ed. D. R. Prothero and R. M. Schoch). Oxford University Press, New York.
- Siddall, M. 1996. Stratigraphic consistency and the shape of things. *Systematic Biology* 45: 111–115.
- Simpson, G. G. 1961. *Principles of Animal Taxonomy*. Columbia University Press, New York.
- Smith, A. B. 1988. Patterns of diversification and extinction in Early Paleozoic echinoderms. *Paleontology* 31: 799–828.

- Sokal, R. R., and F. J. Rohlf. 1981. *Biometry*, 2nd edition. W. H. Freeman, New York.
- Solow, A. R. 1996. Tests and confidence intervals for a common upper endpoint in fossil taxa. *Paleobiology* 22: 406–410.
- Strauss, D., and P. M. Sadler. 1989. Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. *Mathematical Geology* 21: 411–427.
- Wagner, P. J. 1995. Stratigraphic tests of cladistic hypotheses. *Paleobiology* 21: 153–178.
- Wiens, J. J. 1998. Testing phylogenetic methods with tree-congruence: Phylogenetic analysis of polymorphic morphological characters in phrynosomatid lizards. *Systematic Biology* 47: 411–428.
- Wiens, J. J., and M. R. Servedio. 1998. Phylogenetic analysis and intraspecific variation: Performance of parsimony, distance, and likelihood methods. *Systematic Biology* 47: 228–253.
- Wray, G., J. S. Levinton, and L. Shapiro. 1996. Molecular evidence for deep Precambrian divergences among metazoan phyla. *Science* 274: 568–573.
- Wright, D. 1993. Evolution of sexually dimorphic characters in peccaries (Mammalia, Tayassuidae). *Paleobiology* 19: 52–70.