Stratocladistics: Integrating Temporal Data and Character Data in Phylogenetic Inference

Daniel C. Fisher

Museum of Paleontology and Department of Geological Sciences, University of Michigan, Ann Arbor, Michigan 48109; email: dcfisher@umich.edu

Key Words

cladograms, ancestor-descendant hypotheses, parsimony, phylogenetic trees, stratigraphic order

Abstract

Debate has long simmered over whether data on the order of appearance of taxa in the stratigraphic record should play any role in analyses of phylogenetic relationships among those taxa. Critics argue that temporal data are in principle inapplicable to questions of cladistic relationship, but specific versions of this claim all seem flawed. Stratocladistics offers a methodological context (patterned after that of cladistics itself) within which temporal data participate along with conventional character data in selecting most-parsimonious hypotheses. Stratocladistics outperforms cladistics in tests based on simulated histories, and additional testing will be facilitated by new software automating stratocladistic searches. As with any body of data, we may decide to include or exclude temporal data for specific reasons, but the explanatory power of hypotheses that use both temporal and conventional character data exceeds that of hypotheses based on character data alone.

INTRODUCTION

Early attempts to infer phylogenetic relationships accorded great significance to data from the fossil record (Simpson 1961, p. 83). The importance of fossils was twofold. They have always provided our only access to some of life's diversity, presenting new taxa, distinguished by new character combinations, with potential to affect our interpretations of relationships (Donoghue et al. 1989). In addition, fossils document the observed order of occurrence of taxa within stratigraphic sequences, which are correlated on various spatial scales to compile the temporal sequence of the rock record. Data on the temporal order of fossils were viewed as important because they influenced the relative plausibility of competing hypotheses of relationship. Gingerich (1979) formalized this approach to phylogenetic analysis as stratophenetics, according to which temporal data are crucial because they record the location of samples along a principal dimension (in addition to the multiple dimensions of "character-space") along which proximity, or similarity, of taxa is assessed. If phylogenetic inference consists of evaluating hypotheses about trajectories of morphologic change within lineages and how lineages branch through time, how can we ignore the temporal order of samples?

Against this backdrop, cladistics emerged (Edwards & Cavalli-Sforza 1963, Hennig 1966, Kluge & Farris 1969) as a method of inferring relationships based on shared patterns of distribution of character states, in conjunction with what initially seemed like minimal assumptions about higher-level relationships (recognition of outgroups). The purported genius of cladistics (and for proponents, this is no overstatement) was that it used data with no explicitly temporal content to generate inferences with temporal implications, with respect to relative recency of common ancestry. Thus, a cladistic history became accessible even for groups of extant organisms with no fossil record (Brundin 1968). The temporal component of the fossil record therefore seemed inessential, even when it was available.

Cladists' assertion of independence from explicitly temporal data soon hardened into stronger claims that temporal data were a poor source of information, or were even inapplicable to the problem of resolving cladistic relationships. Sources for such ideas include Schaeffer et al. (1972) and Eldredge (1979). Although their work preceded articulation of stratocladistic methodology (Fisher 1980, 1982, 1988, 1991), it represents a clear challenge to stratocladistics if not to other approaches to using temporal data in assessing relationships.

The birth of stratocladistics traces simply to my own struggle to understand why, if indeed it was true, temporal order did not count as evidence for phylogeny. The more I probed, the more I remained unconvinced, and I began to suspect that the fundamental workings of cladistic methods were not as incompatible with incorporation of temporal data as had been portrayed. I was intellectually intrigued by the prospect of a differently configured version of cladistics that could be conversant with a broader range of data types. As I searched for a defensible form for an expanded version of cladistics, I was encouraged by early signs of favorable performance on trial data sets. I also came to view this approach as enhancing the explanatory power of cladistics, making it less dependent on prior knowledge of outgroup relationships, and improving its ability to resolve phylogenetic history (Fisher 1994, 1998).

I attempt here to review and respond to the main arguments raised against stratocladistics, especially those that involve methodological issues independent of the quality of the temporal record. The quality of the record is important, but if methodological criticisms are cogent, even an excellent record would be irrelevant to inferring phylogenies. Stratocladistics is not alone in use of temporal data, but space constraints preclude treatment of other methods (e.g., Alroy 1995; Gingerich 1979; Huelsenbeck & Rannala 1997; Wagner 1995, 1998).

HOW STRATOCLADISTICS WORKS

Stratocladistics handles character data exactly like conventional cladistics. It is parsimony based; it accommodates specification of outgroups if they are available; and it seeks, as one level of output, the most parsimonious cladogram(s) for the taxa under analysis.

Stratocladistics differs from cladistics in having a modified criterion of parsimony and consequently an ability to distinguish one or more most-parsimonious phylogenetic trees compatible with the most-parsimonious (by its criterion) cladogram(s). These trees may specify ancestor-descendant (as well as collateral, sister-taxon) relationships, if data warrant. In cladistics, the parsimony criterion focuses on minimizing treelength, the number (or for weighted characters, the value) of transitions in character state needed to explain the distribution of observed states among taxa. This is equivalent to minimizing the number (or value) of homoplasies implied for the problem, which equals the number of ad hoc hypotheses of convergence, parallelism, or reversal needed to explain away logically independent cases where the observed distribution of character states fails to match the pattern expected, given the hypothesis of relationships under consideration (Farris 1983). In stratocladistics, the criterion of parsimony additionally records the number of logically independent cases where the order of occurrence of taxa does not match that expected, given the hypothesis of relationships and a working assumption of roughly equivalent preservation probabilities for each lineage present in a given time interval (preservation probabilities may vary between time intervals). This is equivalent to the minimum number of cases where a lineage is inferred to have passed completely through an interval of time without representation, when members of other lineages in the analysis were represented by fossils. It is also equivalent to the number of independent cases where the phylogenetic hypothesis and the observed stratigraphic order of taxa cannot be reconciled without assuming substantially unequal preservation probabilities, in a direction, and of a magnitude, sufficient to render the observed pattern an expected outcome.

On a terminological level, stratocladistics refers to the number of homoplasies required by a hypothesis as its morphologic parsimony debt, a quantity that reflects the number of ad hoc hypotheses that a phylogenetic hypothesis must accommodate because of its imperfect fit to morphologic data. If enough incongruence exists among the data, all competing hypotheses share a finite debt level attributable to that incongruence. Treelength, by comparison, is a compound of debt plus a finite quantity (the minimum conceivable number of steps for each character, summed over all characters) reflecting the number of characters (and states) used in the analysis. This quantity, included in the treelength of all hypotheses, is a function of the data set, not the fit of hypotheses to data. Stratocladistics focuses on debt and the retention index (Farris 1989), which normalizes debt values to the maximum debt value incurred by the worst of competing hypotheses; it ignores the consistency index and related measures because they include units of treelength not attributable to fit between hypothesis and data.

With respect to temporal data, stratocladistics assesses stratigraphic parsimony debt, the number of missing lineage segments or ad hoc hypotheses of differential preservation probability required by a phylogenetic hypothesis. This measures the fit between that hypothesis and data on the order of taxon occurrence. Total parsimony debt (morphologic+stratigraphic) is stratocladistics' criterion of hypothesis choice. Zero debt implies perfect fit to congruent data; positive debt implies some mismatch between hypothesis and data; and the best hypothesis(es) is one that explains data in terms of descent with modification, while relying least on ad hoc hypotheses of homoplasy and differential preservation probability. For additional detail, see Fisher (1992, 1994), Clyde & Fisher (1997), Fox et al. (1999), Fisher et al. (1999), Bodenbender & Fisher (2001), Bloch et al. (2001), Fisher et al. (2002), and Fisher & Bodenbender (2003).

WHAT'S WRONG WITH TEMPORAL ORDER?

Schaeffer et al. (1972) reminded readers that a sequence of samples recovered from any moderately complex phylogeny might not represent a series of successive ancestors and descendants. Lest this point be lost on readers, they offered a diagram showing a hypothetical phylogeny of 34 species bearing just five x's marking samples that actually got preserved; admittedly, interpreting the x's as a series of ancestors and descendants would have missed most of the complexity of the phylogeny. However, contrary to their claim that "no two [of the x's] have an ancestral-descendant relationship" (Schaeffer et al. 1972, p. 35), the next-to-last x was in fact ancestral to the last one (though separated from it by one unpreserved species; perhaps they were counting only direct ancestors; Foote 1996). Still, their diagram hints at a systematic pattern. Each x falls on a species-level lineage, and any two x's can be traced back to a most recent common ancestor; for each pair except the last two (the ancestor-descendant pair noted above), the lineage with the greater longevity also has the greatest waiting time (from the common ancestor to its x) before a member of its lineage is preserved, and greater waiting times imply lower probability of preservation or recovery for analysis. However, it seems unlikely that long-lived lineages, if they differ in any systematic way from shorter-lived relatives, would have consistently longer waiting times till preservation (lower preservation probabilities). If their longevity relates to population size or biogeographic range, the reverse should be true. Schaeffer et al. made no claim that the x's were distributed randomly, and they were trying to make a certain point, but a more nearly random distribution of x's, as expected for more nearly homogeneous preservation probabilities, would not have made the point as well. Finally, ancestor-descendant relationships aside, it is not self-evident that the distribution of x's is unrelated to the illustrated branching order.

To be fair to Schaeffer et al., much of their intent was apparently to discourage a rather crude application of data on temporal order: inferring polarity of morphoclines. Still, their case was made mostly by citing potential for error (ignoring potential informativeness) and by assertions of the form, "the fossil record for most groups . . . is too incomplete to allow the assumption that relative stratigraphic position is necessarily indicative of morphocline polarity" (Schaeffer et al. 1972, p. 37). The inclusion of "necessarily" makes use of temporal data appear extreme enough that their caution seems well advised, but it is worrisome that a quantitative threshold is implied ("too incomplete") but not defended in quantitative terms. How incomplete is too incomplete, and how do we know whether this threshold is exceeded in most groups?

In contrast to Schaeffer et al. (1972), I suspected that a probabilistic argument could be made in which temporal order has some bearing on morphocline polarity. However, I was dissatisfied with this approach because a finite set of stratigraphic observations (taxon A occurs before taxon B) might polarize any number of characters in those taxa, compromising the independence of synapomorphies deriving their polarity from the same set of stratigraphic occurrences. I therefore consider determining morphocline polarity an inappropriate way of using temporal information.

Another argument central to the origin of stratocladistics was made by Eldredge (1979), who elaborated on the distinction (attributed to Gareth Nelson) between cladograms, trees, and scenarios. Consistent usage of these terms is complicated in biological systematics by frequent use of tree in a more general sense that covers cladogram as well, but Eldredge's and Nelson's distinctions are still widely recognized. Cladograms are diagrams depicting synapomorphy distributions for a group of taxa. All taxa are in terminal positions, and the diagram as a whole lacks a temporal dimension per se. Nevertheless, when interpreted in an evolutionary context, cladograms imply patterns of relative recency of common ancestry because they specify an order of branching events within (not between) clades. Their nodes and branches are conventions for grouping taxa, not symbols for common ancestors and phyletic lines. Trees, in contrast, are maps of lineage genealogy with

an explicitly temporal dimension (even if uncalibrated). Taxa on trees occupy terminal or internal positions, according to a wide variety of conventions, and the nodes and branches of the tree are intended to represent schematically the pattern of lineage history, including cladogenetic events and ancestor-descendant relationships, if these are hypothesized. Trees vary in the extent to which they express character state mapping, but because taxa (with observed character states) may occupy nonterminal positions, and characters may be traced by optimization procedures, it is appropriate to consider trees as involving information on distribution of character states along branches. Finally, scenarios place the phyletic and cladistic history portrayed by trees into a geographic, environmental, or ecological context.

Eldredge (1979) pointed out that cladograms are more general statements than trees because they only portray relative proximity of relationship, as in (A(BC)), without going on, as a tree would do, to specify the nature of relationships as either collateral or ancestor-descendant. Eldredge then argued that trees also require additional assumptions beyond those entailed by cladograms, such as assumptions that certain taxa really were linked in an ancestor-descendant relationship and that reversals of character state had not occurred along the lineage leading from ancestor to descendant (this was supposedly required to make trees testable; 1979, p. 189–190). Finally, he argued that for both of these reasons, phylogenetic analysis should begin at the cladogram level, and only then should we evaluate which trees were most plausible among those associated with the preferred cladogram. Because explicitly temporal data are only expressed in trees, this means that temporal data cannot be entertained until the most parsimonious cladogram(s) has been identified and, therefore, temporal data cannot participate in selecting most-parsimonious cladograms. Eldredge's approach set the pattern for almost all cladists (e.g., Smith 1994, Wiley 1981) who were inclined to consider temporal data at all.

A HIERARCHY OF HYPOTHESES

The generality of cladograms relative to trees is now routinely portrayed as part of an even more inclusive scheme (Fisher 1994, Scotland 1992) in which unrooted networks (all taxa terminal, linked by unpolarized branches via internal nodes) form the most general level of a hierarchical system of hypotheses for describing relationships. Taken together, all unrooted networks comprise a set of mutually exclusive and collectively exhaustive (MECE) hypotheses about how a given set of taxa might be connected via intermediate nodes. Each unrooted network is then associated with a set of cladograms that represents different potential rootings of that network, and together, all cladograms (associated with all unrooted networks) comprise a set of MECE hypotheses about relative proximity of relationship. Continuing in this vein, each cladogram is associated with a set of trees that represents different specific hypotheses of relationship compatible with that cladogram, and together, all trees (associated with all cladograms) comprise a set of MECE hypotheses about the specific nature of relationships among all taxa under consideration. Only when we get to the level of scenarios does it become impossible to generate MECE hypotheses to explain the system, because at this level, the relevance of innumerable possible interactions means we have stepped outside the bounds of a combinatorial problem for which a closed hypothesis set can be proposed.

Given the hierarchical, one-to-many relationship between cladograms and trees, the generality of cladograms relative to trees seems incontrovertible (Fisher 1981, 1994). However, I view this as a structural relation with no implication regarding logical priority. I also differ from Eldredge with regard to the assumptions he thought trees made. A tree that shows B as ancestral to C does not assume this; it merely states it as its interpretation of relationships. Selecting that tree as most parsimonious is only legitimate if we have data supporting that interpretation, and this means we select the tree not because we choose to make an assumption of ancestry but because the data in

hand make competing trees less attractive. The same would be true for a tree showing B and C as collateral descendants of a common ancestor. If the data do not distinguish between these two trees, but do imply that all competing trees are less parsimonious, we retain these two as equally parsimonious, and we cannot then be said to assume both models of genealogy for the same taxa! Likewise, if an ancestor-descendant hypothesis implies that a reversal has taken place in one or more characters, this is not counted in its favor; it is evidence against the hypothesis. Contrary to the arguments of Eldredge, we do not need to assume "that no character reversal has occurred" (1979, p. 190) to make trees testable; Eldredge may have felt driven to such a position because he focused exclusively on character data, and this may be the only way, within a standard cladistic framework, to reject trees based on character data alone. However, if we include temporal data, the problem is transformed. Debates on these issues have admittedly been complicated by controversy over the nature of hypothesis choice in cladistics: Is it a matter of testing and Popperian falsification, or is it a matter of corroboration and refutation, with hypothesis choice being directed by an "optimality criterion" (Geiger et al. 2001)? Despite early popularity of the former view, current practice and algorithmic approaches to phylogenetic inference surely document the importance of the latter view. With Farris (1983), I take cladistics to be a search for the best-available hypotheses based on a performance criterion measuring explanatory power. This perspective does no violence to standard cladistic practice, and it works for stratocladistics as well. We simply need to evaluate trees on their merits.

If trees do not entail assumptions that cladograms avoid—if instead, trees simply make more explicit (less general) statements and are judged accordingly—then there is no necessity to treat cladistics as operating first at cladogram level and only then at tree level. We could equally well consider it a problem to be resolved at tree level, after which we have the option of expressing results at cladogram level. In short, given a hierarchical hypothesis set, implications should propagate equally well up or down the levels of the hierarchy, from wherever the data are applicable to wherever the consequences are relevant. This argument first took form in a different context (the role of functional analysis in phylogenetic inference), but it immediately suggested that data on the stratigraphic pattern of occurrence of taxa might be applied in comparative evaluation of trees and, depending on other data supporting those trees, have effects that extend to cladogram level (Fisher 1980, 1981).

Treating cladistics as operating at tree level sounds like the polar opposite to claims that cladograms represent the only legitimate phylogenetic hypotheses (Engelmann & Wiley 1977). Yet other threads within cladistic practice, Eldredge notwithstanding, suggest that it has been operating at tree level for a long time. One indication of this is the standard currency for comparing cladograms: their treelength. This is not just a misappropriated term. To compare the performance, or explanatory power of competing cladograms, we must know how they each succeed or fail at interpreting traits of organisms as shared novelties originating within a branching system of lineages in a process of "descent with modification." This can only be evaluated by treating nodes and branches as common ancestors and connecting lineages, within which particular combinations of character states are presumed to exist. Thus, we must consider a tree in Eldredge's or Nelson's sense of the word. It is sufficient within conventional cladistics to restrict consideration to the tree that is isomorphic to (i.e., shows the same topology as) each cladogram, because a general property of trees evaluated only by character data is that trees isomorphic to cladograms are never less parsimonious than competing trees associated with that same cladogram that propose one or more ancestor-descendant relationships. In a similar vein, treelength can be construed as applying even to choosing the most parsimonious unrooted network, because the computation of treelength does not depend on polarity. PAUP (Phylogenetic Analysis Using Parsimony) thus searches on unrooted trees by default, because this obviates bookkeeping on the

large number of cladograms that represent only different rootings of the same unrooted network (Swofford 2001).

Another justification for treating conventional cladistics as a choice among competing trees, not cladograms, derives from considering the nature of synapomorphy. If we treat synapomorphies as indivisible, unitary hypotheses, they appear to operate at cladogram level, because a given hypothesis of synapomorphy (which specifies both the taxa in which a trait is observed and the postulate that the trait is synapomorphous in those occurrences) either is or is not compatible with a given cladogram and the traits of record for other taxa. The notion that synapomorphies falsify certain cladograms relies on the potential for such incompatibility. However, a hypothesis of synapomorphy is a compound hypothesis. It requires that a trait observed in two or more taxa be both homologous in those taxa and derived relative to occurrences of the corresponding primitive state. Neither of these components of the hypothesis operates at cladogram level, because no cladogram is necessarily inconsistent with a given hypothesis of homology or polarity, considered in isolation. On the other hand, trees specifying the topology of genealogy and assigning character states to nodes, and along branches, are sufficient for evaluating consistency with hypotheses of homology and polarity. For example, in a typical three-taxon problem, considering (for simplicity) only one character, multiple trees, some associated with each competing cladogram, are consistent with a given hypothesis of homology. Likewise, multiple trees, some associated with each cladogram, are consistent with a given hypothesis of polarity. However, only one cladogram, the one said to be supported by the corresponding synapomorphy, is associated with trees that are simultaneously consistent with both hypotheses of homology and polarity. We can thus use these hypotheses (either separately, or combined, in the form of a synapomorphy) at tree level and report the results, if we like, at cladogram level, and the effect will be the same.

TREES WITH MORE . . . AND LESS . . . EXPLANATORY POWER

Optimizing the fit of trees to character data is widely understood as finding trees whose topology matches the nested patterns of distribution of synapomorphies. In a case where no incongruent character data are observed, and where the number and pattern of distribution of synapomorphies permit complete resolution of cladistic relationships, there may be multiple trees that are equally parsimonious, but they are all associated with the same cladogram (Fisher 1994). These trees differ in proposing different combinations of collateral and ancestor-descendant relationships, but saying they are equally parsimonious admits that the data do not warrant any preference for one over another. If we now assume that all ingroup taxa have autapomorphies, the tree isomorphic to the most parsimonious cladogram will be selected as best overall, because placing any taxon as ancestral to another would imply homoplasy to reverse an autapomorphy in the putative ancestor and explain why its putative descendant shows a condition indistinguishable from the outgroup. Note here that the problem posed by reversal is not that the condition of the descendant has changed from that of its ancestor, but that the insertion of the putative ancestor into the direct line of descent connecting the putative descendant to the outgroup makes it impossible to interpret the similarity of descendant and outgroup as a matter of shared retention of the same condition from a common ancestor. Put differently, if we want to retain the ancestor-descendant hypothesis in the running, we cannot claim as one of its strengths that the topology of its genealogy explains the distribution of character states. This is because, though the genealogical topology would not necessarily require ancestor and descendant to be identical, it would not lead us to expect the descendant to resemble the outgroup. How could we justify continuing interest in this hypothesis of ancestor-descendant relationship when an alternative genealogical topology, taking the putative ancestor out of the direct line of descent between the outgroup and the putative descendant, does

not raise this problem, or fits the data better? The hypothesis of ancestor-decendant relationship might still be saved if we could claim that along the lineage leading to the putative descendant, the probability of a reverse-transition in the character in question was, for some unknown reason, elevated, or that the scoring of this character for the putative descendant taxon had been handled in an incautious, superficial manner, rendering the data untrustworthy. However, both of these are clearly examples of ad hoc hypotheses (Farris 1983). Although they might seem at first to prop up the ancestor-descendant hypothesis in the face of inconvenient data, their very arbitrariness raises questions about other data: Why should transition probabilities be elevated here and not elsewhere, or why should these data and not others be misleading? In short, any such arbitrary appeal reduces the overall explanatory power of hypotheses. For this reason, we avoid ad hoc hypotheses wherever possible and evaluate hypotheses in terms of the extent to which they explain data in hand without indulging in such appeals.

In the case considered above, tree selection was the only issue; whether one or many remained as equally parsimonious, they were all associated with the same cladogram. In a case with incongruent synapomorphies, there may be no tree, and no cladogram, that does not appeal to some ad hoc hypothesis to explain away (or minimize the impact of) data that are not explained by its genealogical or cladistic hypothesis. However, we still apply the same criterion for hypothesis choice (in the absence of temporal data): minimize instances of homoplasy in character data, minimize appeals to ad hoc hypotheses, minimize treelength (homoplasies add steps beyond the minimum number conceivable), and minimize morphologic parsimony debt. Yet in evaluating treelength or parsimony debt, how do we know that all synapomorphies are independent and equal in evidential value, or that all homoplasies are equally egregious? We do not, and they no doubt are not . . . exactly . . . but we (generally) have no grounds for quantifying such inequalities objectively and thus cannot justify less even-handed treatment. We therefore treat one ad hoc hypothesis as equivalent to another and quantify the weight of character evidence against hypotheses in a simple tally of morphologic parsimony debt.

Just as genealogical or cladistic hypotheses may explain, or fail to explain, character data, they may also explain or fail to explain temporal data. Temporal data recording the order of occurrence of samples within a stratigraphic sequence, or indeed within any independently documented temporal sequence (e.g., based on radiometric dating), are objective properties of samples considered within the context of their preservation and recovery. We may interpret aspects of this context on geological grounds, but we are not free to consider samples as having any relative age whatsoever.

Consider the stratigraphic sequence A<[lower than]B<C. Within a stratocladistic context this means that specimens have been collected at three successive levels and that specimens at the lowest level were referable only to A, at the next level, only to B, and at the highest level, only to C. Now consider the genealogical tree postulating the direct ancestor-descendant sequence A→[gives rise to]B→C (**Figure 1**). Without any equivocation, this genealogy can be said to predict the observed stratigraphic order and thus certainly explains it.

An alternative genealogy that is isomorphic to the cladogram (A(BC)) cannot make such clear predictions. However, given an occurrence of A in the lowest level, it does predict (given a minimal assumption of comparable preservation probability for members of different lineages within the same time interval) that members of the lineage leading to BC should have been found there as well, and none were. Likewise, given that B occurs in the middle level, members of the lineage leading to C should have occurred there as well, and none did. Thus, this genealogy does not explain the observed stratigraphic sequence as well as the hypothesis of direct ancestor-descendant relations (even though both are associated with the same cladogram). How bad is it? By making two ad hoc assumptions of lower preservation probability for each of the unrepresented lineage segments (*asterisks* in **Figure 1c**), we could reconcile this hypothesis with the record.
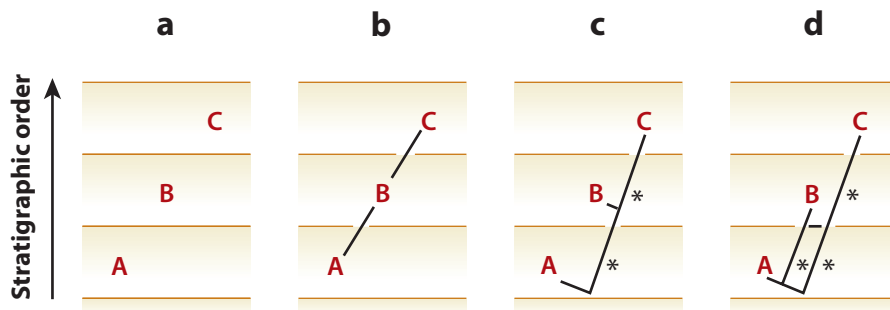
Stratigraphic order

a   b   c   d

Figure 1

Different genealogical hypotheses incur different stratigraphic parsimony debts. (*a*) Stratigraphic order of taxa A, B, and C. (*b* and *c*) Two genealogies compatible with cladogram (A(BC)) are shown. (*d*) A third genealogy compatible with cladogram ((AB)C). In each case, units of stratigraphic debt (*), if any, are incurred by lineage segments that are hypothesized but not recovered.

Consider now another genealogical hypothesis isomorphic to the competing cladogram ((AB)C). Given that A occurs in the lowest level, a member of the lineage leading to B should have occurred as well, but did not, and given that C occurs in the highest level, it must have been in existence during, and should have occurred within, each of the lower levels, but it did not. Thus, it takes three ad hoc hypotheses to reconcile the genealogy with the observed temporal order. The lineage leading to B has to have a lower preservation probability than the lineage including A; the lineage leading to C must have a lower preservation probability, during the lowest interval of time, than the lineage including A; and the lineage leading to C must have a lower preservation probability, during the middle interval of time, than the lineage leading to B. If these inequalities are allowed as auxiliary hypotheses, the genealogy+auxiliaries could be said to explain the temporal data, but only by accepting three ad hoc hypotheses (*asterisks* in **Figure 1***d*).

Continuing in this manner, any genealogical hypothesis is associated with expectations about patterns of occurrence of taxa that may or may not match observation. What makes these failed expectations a liability is that other genealogical hypotheses explain the observed pattern without recourse to arbitrary assumptions of preservation probabilities unequal enough to explain observed patterns. The number of ad hoc hypotheses regarding preservation probabilities is tallied as the stratigraphic parsimony debt and is a measure of the mismatch between genealogical hypotheses and temporal data. Sound familiar?

Can temporal data rule out genealogical hypotheses? This depends on the way units of analysis are viewed. As a stratocladist, I prefer considering the units of analysis to be temporally bounded samples of species-level lineages, or lineage segments (Fisher 1992). They may still be referred to as taxa, because that is common parlance (though operational phylogenetic units, or OPUs, would be more precise), but in this usage, there is no presumption that these taxa have a range that extends beyond the single interval in which they are recorded. We might choose to refer, however, to taxa as units of descent, as well as units of analysis (Fisher et al. 2002). If we mean by this to study a taxon recognized as persisting through two or more coded intervals, stratocladistics reminds us that this assumes a relationship that might better emerge as a result of the analysis. Still, acknowledging that there may be circumstances where a species-level lineage is accepted as persisting through multiple time intervals, such taxa are scored from a temporal standpoint in terms of the entire range attributed to them. The difference between treating lineage segments and units of descent is that with lineage segments, finding A>[higher than]B rules out a genealogical hypothesis according to which A→B; a lineage segment higher in a stratigraphic sequence cannot

be ancestral to one lower in the sequence. However, if units of analysis are considered units of descent, whether or not they are observed to occur above or below the time interval in which they are recorded, A>B is not necessarily incompatible with A→B. However, the genealogical hypothesis A→B cannot be said to explain a temporal pattern in which A>B, and in the face of such data, an ad hoc hypothesis regarding preservation probability would be required to explain A's failure to be represented in the interval in which B occurred.

Finally, we must consider how trees will be evaluated when they show mismatch to both character and temporal data. Stratocladistics takes the position that an ad hoc hypothesis regarding preservation probabilities represents just as much loss in explanatory power, just as much a mismatch between a genealogical hypothesis and the data being used to evaluate it, as an ad hoc hypothesis regarding homoplasy. Morphologic parsimony debt frequently involves diverse types of character data, and yet we treat it as a simple tally because the exact nature of an ad hoc hypothesis of homoplasy or the identity of the character to which it applies is not viewed as critical. Rather, what matters is that data that are predictable under a competing hypothesis of relationships, and that can therefore be interpreted as support for that hypothesis, fail to match expectations conditioned on this hypothesis of relationships. If we had the option simply to accept a genealogical hypothesis showing no mismatch to data, we would do so and avoid any appeal to ad hoc hypotheses, but when incongruence forces our hand, we must acknowledge each logically independent case of mismatch as calling for a separate ad hoc hypothesis. Ad hoc hypotheses regarding preservation probability are indeed different in content from ad hoc hypotheses of homoplasy, but the same considerations apply. Independent ad hoc hypotheses regarding preservation probability are thus summed in a tally of stratigraphic parsimony debt, and the total number of ad hoc hypotheses attributed to a genealogical hypothesis represents its total parsimony debt. This is stratocladistics' optimality criterion. This approach, unlike an explicitly probabilistic treatment, makes no assumption about how improbable mismatches between expected and observed patterns might be, nor even that they all be (strictly) equi-probable. Because only the number of mismatches is tallied, all that matters is that the posterior probability of hypotheses entailing more mismatches be lower than that of a hypothesis entailing fewer, and no combination of probabilities compatible with this outcome will be misleading.

In stratocladistics, although temporal data influence hypothesis choice, their impact is limited by the number of temporal intervals coded in an analysis, which is commonly greatly exceeded by the number of available characters. Moreover, in successive rounds of analysis, any result may be tested by addition of new characters or of newly resolved time intervals (between previously coded intervals) with newly recovered specimens. The only case in which temporal data might threaten to swamp character data is when there are few identifiable characters and a dense, well-resolved temporal record.

PERFORMANCE OF STRATOCLADISTICS

Clyde & Fisher (1997) published the first study of stratocladistics' performance using real data. Case by case and overall, results were similar to, but variably different from, conventional cladistic results (comparing at cladogram level). Stratocladistics' most-parsimonious trees had morphologic parsimony debts slightly higher than the morphologically most-parsimonious trees selected by cladistics, but they more than made up for this with lower stratigraphic parsimony debts. Incorporation of temporal data thus modified interpretations of phylogeny somewhat, but provided greater explanatory power overall. Stratigraphic retention indices were lower than morphologic retention indices when measured on cladistic results but not significantly different when measured on stratocladistic trees.

Beyond simply comparing the results of different analyses, Wagner (1998) showed that strato-cladistics outperformed conventional cladistics in a simulation modeled on real data for hyaenids. Likewise, a simulation-based study by Fox et al. (1999) explored a range of different probabilistic evolutionary models and levels of preservation probability, comparing results for stratocladistics and cladistics in a blind, head-to-head competition. Under all models of character change and most levels of preservation probability, stratocladistics performed significantly better than cladistics. At the lowest levels of preservation probability, the difference in performance dropped below significance, probably because too few taxa remained under this preservational model for the test to have sufficient power. There was no indication of a level of preservation probability below which stratocladistics was consistently less effective than cladistics.

Additional studies explored two extreme cases. With a dense and relatively continuous fossil record and relatively congruent character data, cladistics and stratocladistics yielded the same (single) cladogram, and the two stratocladistic trees closely matched a stratophenetic genealogy (Bloch et al. 2001). In a much larger and more complex problem for both character and temporal data, cladistics and stratocladistics both struggled with large sets of equally parsimonious alternatives, though stratocladistics seemed to achieve greater resolution of cladistic relationships (Bodenbender & Fisher 2001). Sumrall & Brochu's (2003) comment on this pointed out that the reported stratocladistic result was not the most parsimonious interpretation, even strato-cladistically, which highlights the difficulty of a nonautomated stratocladistic search [implemented using PAUP* (Swofford 2001) and MacClade (Maddison & Maddison 1992)]. Searching manually, Bodenbender & Fisher (2001) clearly fell short of the stratocladistic optimality criterion, but we stand by other aspects of our analysis.

RESPONSE TO STRATOCLADISTICS

From the cladistic camp, response to stratocladistics has been predictable and severe. It mattered little that stratocladistics was designed to emulate cladistic practice, because inclusion of temporal data violates basic precepts of cladistics. Stratocladistics was represented (Fisher 1994) in a 1992 symposium, "Systematics and Process," at the Field Museum, Chicago, but was criticized in the summary chapter of the published proceedings (Rieppel & Grande 1994). Another symposium featuring proponents (Fisher 1997; Huelsenbeck 1997, Marshall 1997b, Wagner 1997) and opponents (Norell & Novacek 1997, Rieppel 1997, Siddall 1997) of temporal data in phylogenetic inference was organized for the 1997 meeting of the Society of Vertebrate Paleontology in Chicago (reviewed by Lockwood 1998). A different venue was an online debate hosted by *Nature* (Smith 1998b). The report by Fox et al. (1999) of a performance advantage for stratocladistics relative to cladistics was criticized by Heyning & Thacker (1999) and Smith (2000). Other critical views were presented by Geiger et al. (2001) and Sumrall & Brochu (2003). More favorable impressions (and discussions of other views) were aired by Alroy (2002), Lockwood (1998), Novacek (2001), and Wagner (2002).

Criticisms Claiming to Vitiate Use of Temporal Data

The most strident charges against stratocladistics claim that the logical foundations of cladistics absolutely preclude incorporation of temporal data. These views address matters of principle that no quibbling over the accuracy or precision of temporal data should be able to finesse.

Generality and logical priority of cladograms. If temporal data are thought of as native only to trees, and trees as always secondary to cladograms (Smith 1994), temporal data could never

affect choice of cladistic hypotheses. Stratocladists grant cladograms' generality and use them as efficient summaries of the cladistic aspect of sets of trees, but we reject the logical priority of cladograms over trees. We acknowledge that phylogenetic analysis can be viewed as occurring at cladogram level, but we consider it also to work at tree level (as most algorithms do), searching for most-parsimonious genealogies and patterns of character state change.

Temporal data depend on negative evidence. Temporal data have been described as flawed because they depend on negative evidence (Heyning & Thacker 1999). You know the chant: "Evidence of absence is absence of evidence, and absence of evidence is no evidence at all." Yet absence of evidence in stratocladistics is when you find no fossils, and the method ignores these intervals in coding temporal data. Absences that are charged against phylogenetic hypotheses are when some lineages are represented but other lineages that should have been observed (treating preserved taxa as taphonomic controls) are not. It is not just absence of data that is meaningful (Wagner 2000), but presence of certain data conjoined with absence of other data. Any data involving distributions depend to some degree on where we do not find things that should have been found under a given sampling regime, were they present. Biogeography has not fallen prey to charges of negative evidence, though it clearly operates in the same way.

Extrinsic attributes are not subject to descent with modification. Most phylogenetic analysis is based on distributions of intrinsic character states, either morphological or molecular. These carry information about genealogy because they participate directly in a process of descent with modification, which leaves clues about lineage history in the form of trait distributions. Extrinsic attributes such as geographic location or geologic age, from which we derive relative temporal position, are not recorded genetically (Geiger et al. 2001). Yet age does descend with modification, in the sense that each descendant has an incremented variant of its ancestor's position in time; indeed, the process of modification—the succession of generations and time's unidirectionality— is more regular than evolution. Genes and heritability may be necessary to get a morphological or molecular character from one generation to the next, but the continuity of genealogy itself does the job for extrinsic attributes. More degrees of freedom are available for modification of geographic location in a community of descent, but the process still shows broadscale continuity. "The apple falls not far from the tree."

Extrinsic attributes have no associated concept of homology. I agree with this statement by Norell & Novacek (1997), and it follows that sharing the same age is not sufficient evidence that taxa are closely related (Forey 1995). However, temporal data simply do not work in the same way as character data. This point may have been clouded by the decision to refer to stratigraphic order as a character in MacClade 3.0 (Maddison & Maddison 1992; I had suggested calling it simply an attribute). Yet users of MacClade 3.0 are cautioned that the stratigraphic character is a unique character type and that calculation of stratigraphic parsimony debt required its own algorithm (Maddison & Maddison 1992). The bottom line is that homology is not the only carrier of genealogical information in the system of taxa and traits. Temporal order carries information as well, but in a way that can only be understood (as explained above) at tree level.

Time and temporal order are linear, not hierarchical, traits. This point (Smith 1998a) is related to the previous one in that homologies, interpreted as synapomorphies, are what display the hierarchical distributions central to cladistics. However, are singular synapomorphies really hierarchical? Some taxa have them; others do not. There is no hierarchy in this, unless we adopt the arbitrary position that "presence is always a subset of absence" (Patterson in Fisher 1994).

Hierarchy first appears when we add evidence that roots a formerly unrooted network depicting trait distribution (Fisher 1994), and it only becomes manifest as nested sets of taxa and synapomorphies when we consider comparable data for multiple characters. The single trait of temporal order (some taxa occur earlier, others later) is no more effective (without further input) at generating hierarchy; but for stratocladistics, this is irrelevant because its tree-based tests of genealogical hypotheses still provide grounds for hypothesis choice (Fisher 1998). Wagner (2000) makes similar comments on hierarchical versus linear information.

Extrinsic attributes are not traits of specimens themselves. Geiger et al. (2001, p. 177) argue that "Age is not a property of specimens; age cannot be directly observed from the specimen alone...." The second clause in this quote may be true (ignoring cases where direct dating is feasible), but age or stratigraphic position, like locality information or other contextual data, is very much a property of specimens as they occur in nature. Stewardship of that data is our obligation to the record; few natural scientists would be this dismissive of contextual data!

Ancestors are not recognizable in the fossil record. This much-debated proposition is one of the more protean claims in this controversy. One of its early forms is that ancestors cannot be identified with certainty, but this is an ironic charge for a research program that fashioned itself so explicitly after Popper (Platnick 1979). In other cases the claim boils down to recognition that hypotheses of ancestry never emerge as more parsimonious than hypotheses of collateral descent—based on the evidence of character state distribution (Engelmann & Wiley 1977). This is true, as far as it goes, but is strictly a limitation of standard cladistic practice. Generalizing beyond this domain implies, if it can't be done with cladistics, it can't be done at all. Hypotheses of ancestry may also be considered vulnerable if they seem like just another appeal to negative evidence, i.e., absence of autapomorphies, but in stratocladistics, absence of autapomorphies is neither necessary nor sufficient for hypotheses of ancestry. In other cases the ancestor problem asserts that stratigraphic data cannot falsify hypotheses of ancestry. Treating a potential ancestral taxon as a unit of descent (as well as a unit of analysis), we know that finding the putative ancestor higher in a stratigraphic sequence than its presumed descendant does not preclude that lineage— that unit of descent—from being ancestral. However, if units of analysis are lineage segments, a later segment cannot give rise to an earlier segment. With either type of taxa, stratocladistics evaluates hypotheses of ancestry using a combination of character and temporal data, and either rejects them or finds them equal or preferable to hypotheses of collateral descent, according to weight of evidence. As argued by Foote (1996), ancestors of known taxa are likely present in fossil assemblages; if our methods cannot distinguish them, we should develop methods that will. Stratocladistics is one of those methods.

Appeals to temporal data reflect an obsession with verification. This idea (Rieppel 1997) again stems from the Popperian predilections of many proponents of cladistics. Yet the asymmetry between verification and falsification that matters so much for open hypothesis sets (where no set of MECE hypotheses can be proposed) is fundamentally blurred for a problem like phylogeny, structured as a closed hypothesis set (with MECE alternatives). In the clearest case, falsification of all but one alternative is logically equivalent to verification of the remaining candidate. Reflecting this, stratocladistics uses temporal data as much to show shortcomings of some hypotheses as to show strengths of others. Still, the conditional nature of prediction and explanation in phylogenetics, and correspondingly, our methods of hypothesis evaluation, are not well captured by the concepts of verification or falsification (see also Wagner 2000). Rather, we evaluate support using

a performance criterion to express the relative merits of competing hypotheses. Stratocladistics does not make truth claims outside the virtual world of simulation.

Criticisms of Stratocladistic Practice

Unlike the points discussed above, other comments explicitly or implicitly acknowledge the evidential value of temporal data, but criticize stratocladistics for how it handles such data. Where no reference is given, the stated criticism has been raised in verbal debates.

Temporal data can only select among equally parsimonious cladograms. Although many cladists reject the idea that temporal data might overturn a morphologically most-parsimonious cladogram, some of these same workers are content to allow temporal data to select one among a group of cladograms that are equally parsimonious morphologically (Smith 1998a,b). However, if temporal data can be influential in this context, why not in the original analysis (Lockwood 1998)? Letting temporal data sort only among morphologically most-parsimonious cladograms implies that they never provide enough phylogenetic signal to counter even one homoplasy, but this presumes greater homogeneity and lower error rates among character data than is suggested by typical morphologic retention indices (Clyde & Fisher 1997).

Temporal data should be held apart as an independent test of phylogenies. This suggestion (Forey in Smith 1998b; Sumrall & Brochu 2003) would be easier to take seriously if it were implemented more frequently. In actual practice, evaluations of the fit, or agreement, between cladistic hypotheses and temporal data are routinely treated as measuring the quality of the fossil record, assuming phylogeny has been inferred without error (Hitchin & Benton 1997).

Stratocladistics is impractical because it relies on multiple optimality criteria. This claim (Norell & Novacek 1997) is a reference to tallying both morphologic and stratigraphic parsimony debts, but our sole optimality criterion is total debt, the sum of those two components. Counting homoplasies from multiple sets of characters does not imply multiple criteria. The sum of state changes, or homoplasies, is all that matters. Likewise, the sum of morphologic and stratigraphic parsimony debts is stratocladistics' single criterion for hypothesis choice.

Combining debts from different sources confounds the effects of each. As a matter of practice, stratocladistic analyses have always featured a prior cladistic analysis to evaluate the signal in character data alone. This is not essential, but it is helpful for exploring the impact of temporal data. Although separate morphologic and stratigraphic debts are not used as criteria of hypothesis choice, I report them, along with retention indices for each, as a measure of the fit of the best hypothesis(es) to each source of data. Effects are thus not confounded (contra Smith 2000).

Character data and temporal data are too different to be combined. This claim is that the fundamental natures of character (intrinsic, biological) data and temporal (extrinsic, abiological; Adrain & Westrop 2001) data are so qualitatively incomparable that it makes no sense to measure morphologic and stratigraphic parsimony debts in common units or combine them additively. Adding these debts was certainly not the only option. Both could have been portrayed as in a price-performance chart, one of the ways we handle situations where there really are multiple criteria. Stratocladistics does not take that approach because I regard it as unnecessary and because I wanted to conform as closely as possible to standard cladistic practice. Cladistics already combines data that are radically different when it treats characters of different types and anatomical sources. The

common currency in which these contributions are integrated is the tally of evolutionary steps, units of treelength, or morphologic parsimony debt. Character steps are combined additively because each step corresponds to an independent ad hoc hypothesis reflecting mismatch between the data and the phylogenetic hypothesis in question. At this level of analysis, all steps, all ad hoc hypotheses, are considered equivalent. Phylogenetic trees are a framework on which both character and temporal data can be mapped and on which mismatches between hypotheses and data of both types can be tallied. However different in detail, each case of mismatch is motivated by the need to reconcile one of our interpretations with some inconvenient fact. At this level, they are all the same.

Stratocladistics requires more assumptions than cladistics. The question of what assumptions cladistics makes has been much debated (e.g., Farris 1983, Felsenstein 1979, Sober 1988) but not fully resolved. I propose a perspective that can be outlined here, but will need further discussion elsewhere. Use of each data type in phylogenetics brings with it certain assumptions without which observations cannot be effectively combined to constrain interpretation. "No model, no inference" (Sober 1988, p. 199). What matters is not the number of kinds of assumptions, but the number of times, or cases, in which assumptions are invoked. For character data, a degree of mutual independence is assumed, but this is not covered by invoking one all-purpose declaration of independence. Rather, for each character added, we implicitly accept the proposition that change in this part of the system proceeds (effectively) independently of changes in previously analyzed parts of the system. Likewise, forward- and back-transition probabilities for each character are not necessarily required to be small (though we suspect they often are) or equal (exactly), but if they are too large or too unequal, a simple count of steps will not give a fair account of weight-of-evidence. It would not be realistic, nor would it count as just one assumption, to claim that cladistics restricts transition probabilities to certain values. Instead, for each character, we implicitly assume that transition probabilities will not be large enough to render the account of parsimony misleading, and the more congruent the data, the less restrictive such assumptions need be. Much inequality is incorporated in this discrete system of accounting, but as long as these assumptions hold, we will still judge evidence fairly. Use of temporal data does require new kinds of assumption (e.g., Fisher 1994, Smith 2000). Preservation of members of different lineages must qualify as independent events, but we would have no reason to expect otherwise. Somewhat more risky, though still a good bet in most cases, is the assumption of comparability of preservation and recovery probabilities for members of different lineages occurring within the same time interval. These probabilities do not have to be equal (exactly), but if they are too unequal, we may misrepresent the fit of hypotheses to data. In summary, all new data require new assumptions, but temporal data do not require more support in this way than would new character data. Where temporal data do provide an advantage is that assumptions regarding preservation probability are testable through taphonomic studies, outside our understanding of the evolutionary process (Fisher 1994).

Use of temporal data leads to circularity in evolutionary studies. Smith (2000) and Adrain & Westrop (2001) argued that we should not include temporal data in phylogenetic analyses because inclusion would subject studies of evolutionary rates within lineages or clades to charges of circularity. However, Fisher et al. (2002) pointed out that the temporal data used for phylogenetic inference by stratocladistics involves order only, whereas the temporal data used for quantifying evolutionary rates would be absolute ages. Although order can be derived from age data, the reverse is not true, so rate studies require new data. Moreover, at no point in this sequence are we presuming to extract the same information that was input in the beginning.

Impact of temporal data is subject to arbitrary manipulation. The potential for temporal data to affect the outcome of a stratocladistic analysis is in part determined by the total range of variation in stratigraphic parsimony debt values (from the best to the worst hypotheses), compared to the total range of character debt values, and the former range is a function of the number of taxa in the analysis, their distribution in time, and the number of temporal intervals recognized. Smith (2000) and Adrain & Westrop (2001) asserted that the number of intervals was essentially under control of the analyst and could be varied arbitrarily to give stratigraphic data any level of impact desired. However, as argued by Fisher et al. (2002) and Alroy (2002), stringent and quite conservative criteria have been proposed for judging how refined a coding of the temporal scale is warranted. At essentially all scales meaningful for analysis, the stratigraphic record is fundamentally discrete. These data cannot be multiplied at will.

Parsimony searches conducted manually are ineffective. Following release of MacClade 3.0, which implemented calculation of stratigraphic parsimony debt via its stratigraphic character, we conducted stratocladistic searches by (*a*) starting with a cladistic search using only character data, (*b*) adding temporal data to determine total debt of the morphologically most-parsimonious trees, and (*c*) exploring neighboring regions of treespace, mostly manually, attempting to find trees with still lower total debts. For problems with relatively few taxa and relatively congruent data, this is a workable, if time-consuming, solution. However, as noted by Sumrall & Brochu (2003), we agree this is not satisfactory. Fortunately, a stratocladistic search algorithm is now available (Marcot & Fox 2008). Nonetheless, this is a matter of implementation, separate from considering the merits of stratocladistics' optimality criterion.

Stratocladistics is really more about time than stratigraphy. Alroy (2002) made cogent remarks to this effect, and I understand his reservations about the "strato" root. The data in question are significant because they address temporal order. However, the source of these data is still (usually) interpretation of the stratigraphic record that forms the context of fossil discoveries. Whether the method is named according to the nature of its distinctive data or their contextual source is a matter of choice. Originally, "stratocladistics" was a deliberate homage to "stratophenetics," and that seemed sufficient to warrant its use. However, spatial distribution, another extrinsic attribute, is also a source of data relevant to phylogeny. Analyzing it will require definition of a further component of parsimony debt, together with assumptions and conventions for coding and scoring patterns of geographic occurrence. Although beyond the scope of this discussion, the prospect of including spatial data along with temporal data is concisely conveyed by citing their common source in stratigraphy. Looking to the future, I thus prefer "stratocladistics."

Criticisms Citing the Imperfect Nature of the Fossil Record

If use of temporal data cannot be ruled out in principle, some critics have turned to issues of practice, arguing that temporal data are too noisy or error-prone to include. It is not clear that such a perspective could ever justify considering the record to be more often than not misleading ($\neq$ noisy), but it is an empirical problem to determine just how bad the record is.

Probability of preservation is so low that temporal data cannot be meaningful. The probability that any given individual organism will be preserved in the fossil record is widely acknowledged to be low. However, this is not the probability on which stratocladistics depends. We are interested instead in the probability that any one or more individuals out of the millions that might have lived will be preserved within a given time interval, and this value is often not especially low

(Foote & Sepkoski 1999). Preservation probability depends on the taxon and time interval, but in stratocladistics, the preservation probabilities being compared are between closely related taxa, within the same time interval. Without evidence to the contrary, it does not seem unreasonable that these values, however large or small, would be closely comparable.

Preservation probabilities are so variable that temporal order is not informative. Although Rieppel & Grande (1994) make this argument, Norell & Novacek (1992) examined the relation between clade rank and temporal order and found these to be well correlated. Likewise, Benton & Hitchin (1997, Hitchin & Benton 1997) demonstrated considerable structure in temporal data, though one of their measures used data on observed range rather than temporal order. Stratocladistics does not consider absolute time relevant for hypothesis choice precisely because preservation probabilities can vary between intervals (Adrain & Westrop 2001). When Clyde & Fisher (1997) used stratigraphic debt and retention index to evaluate fit to temporal data, it was relatively good even when cladistic results (with no input from temporal data) were the basis for comparison.

Temporal data are likely to change as new fossils are found. Although this is true (Rieppel & Grande 1994), the newsworthiness of significant changes in our perception of temporal order should be a clue that these are not all that common, compared to the number of lineages whose history we study. Moreover, stratocladistics, like cladistics, addresses implications of data in hand; it does not claim to anticipate new discoveries any more than any other method that focuses on pattern in available data. If new data become available, it is a virtue, not a weakness, for methods to recalculate their assessments of hypotheses.

Our ability to correlate strata is so poor that temporal data are meaningless. In the on-line *Nature* debate cited above, Macleod (in Smith 1998b) argued that the stratigraphic record was poorly constrained even in terms of relative age judgments. Yet even Smith (1998b), in introducing the debate, acknowledged that variable levels of uncertainty were associated with varying scales of temporal correlation; just as correlations at the finest scale become tenuous and indefensible, correlations at coarser scales are considered well supported by most of the geological community. In addition, if temporal data were really as poor as suggested by Macleod, the relatively high levels of fit between phylogenetic hypotheses and temporal data must be coincidence. It seems more parsimonious to consider the temporal data and the phylogenetic hypotheses in these circumstances to be reasonably well corroborated.

CONCLUSIONS

One way to exclude temporal data from phylogenetic analyses is to pronounce them formally inapplicable, to deny that they could, even in principle, constrain interpretations of the branching relationships of organisms. However, arguments offered for such claims by proponents of cladistic analysis are explored here and found inadequate. Temporal data naturally inhabit the space of phylogenetic hypotheses at tree level, and they interact there with hypotheses of homology in such a way as to affect the explanatory power of genealogical hypotheses. This in turn varies inversely as the level of ad hoc support required by such hypotheses.

Another approach is to argue that temporal data are the product of too many variable and poorly constrained factors to record an informative signal related to the topology of genealogy. If temporal data happen to be consistent with a given hypothesis, it may be a fortunate outcome of no great significance. However, if temporal data appear to disconfirm a hypothesis of relationships, we discover what is at stake. The cladistic response of rejecting temporal data is then seen as a

strategy to avoid dealing point-by-point with disconfirming evidence. It sounds so simple—just deny the sufficiency of this type of data to hold sway in the analysis. Farris (1983) describes this type of strategy as a "covering assumption." In his example, the covering assumption covers, or applies to, sets of homoplasies that work against one hypothesis but could be reinterpreted as support for a competing hypothesis.

> If contradictory observations could be dismissed as uninformative without regard to their abundance, the link between theory and observation would be tenuous at best. . . . Covering assumptions must be forbidden in scientific study, not only because they are ad hoc, but more particularly because they provide false license to dismiss any amount of evidence whatever (Farris 1983, p. 23).

Despite the fact that Farris is talking about conflicts between incongruent characters, the force of his argument applies equally to attempts to dismiss temporal data. All stratocladistics demands is that cladists count their ad hoc hypotheses of nonpreservation along with their ad hoc hypotheses of homoplasy. If we do this, we will be reconstructing history in a way that is more open to evidence and less reliant on assumptions about models of evolutionary change.

SUMMARY POINTS

1. Cladograms are more general than trees but not logically prior. Temporal data bear on the relative merits of competing trees, and selection among trees is a legitimate model for phylogenetic inference.

2. Although temporal data are extrinsic attributes, and not "heritable" as such, they are still entrained in a process of "descent with modification." Use of temporal data does not depend on a concept of homology but does involve the relative probability of preservation and recovery of members of different lineages in the same time interval.

3. Arguments that temporal data offer no consistent "signal" reflecting phylogenetic patterns are challenged by finding high levels of agreement between genealogical hypotheses and patterns of temporal order. Rejecting temporal data en masse goes against the principle of basing hypothesis choice on explanatory power.

4. Like cladistics, stratocladistics searches for the best fit between hypotheses and data. The difference is in the range of data types brought to the problem. Advantages of stratocladistics include the possibility of better performance (as on simulated histories), better resolution (fewer equally parsimonious cladograms and tree-level hypotheses in addition to cladograms), and reduced reliance on assumptions about the evolutionary process.

FUTURE ISSUES

1. Additional simulations should probe the limits of stratocladistics. Is there a threshold on preservation probability below which incorporation of temporal data is unhelpful? No such limit has been resolved (Fox et al. 1999), and it seems unlikely that even random occurrences would systematically mislead, but this should be demonstrated.

2. Spatial, or biogeographic, data should be accommodated in stratocladistic analysis. This will require expansion of the concept of parsimony debt, but it would represent a major advance in explanatory power.

3. Data on the density or number of occurrences of specimens within temporal intervals might be incorporated into a method of weighting units of stratigraphic parsimony debt. This is related to, but different from, the issue of confidence limits on ranges (Marshall 1997a), which are currently treated as part of the coding of temporal data.

4. If ancestor-descendant hypotheses come to be viewed more generally as legitimate, we may need to reopen the question of whether supraspecific taxa can be considered ancestors, along with related issues of classification. Treating cladograms as the only legitimate formulation of phylogenetic hypotheses supports expectations that all taxa should be monophyletic, but transferring our focus to trees may shift this balance.

DISCLOSURE STATEMENT

The author is not aware of any biases that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

LITERATURE CITED

Adrain JM, Westrop SR. 2001. Stratigraphy, phylogeny, and species sampling in time and space. In *Fossils, Phylogeny, and Form: An Analytical Approach*, ed. JM Adrain, GD Edgecombe, BS Lieberman, pp. 291–322. New York: Kluwer Acad.

Alroy J. 1995. Continuous track analysis: a new phylogenetic and biogeographic method. *Syst. Biol.* 44:152–78

Alroy J. 2002. Stratigraphy in phylogeny reconstruction-reply to Smith 2000. *J. Paleont.* 76:587–89

Benton MJ, Hitchin R. 1997. Congruence between phylogenetic and stratigraphic data on the history of life. *Proc. R. Soc. London Ser. B* 264:885–90

Bloch JI, Fisher DC, Rose KD, Gingerich PD. 2001. New genus of Late Tiffanian carpolestid and stratocladistic analysis of North American Carpolestidae (Mammalia, Plesiadapiformes). *J. Vert. Paleont.* 21:118–30

Bodenbender BE, Fisher DC. 2001. Stratocladistic analysis of blastoid phylogeny. *J. Paleont.* 75:351–69

Brundin L. 1968. Application of phylogenetic principles in systematics and evolutionary theory. In *Current Problems of Lower Vertebrate Phylogeny*, ed. T Ørvig, *Nobel Symp.* 4:473–95. New York: Wiley

Clyde WC, Fisher DC. 1997. Comparing the fit of stratigraphic and morphologic data in phylogenetic analysis. *Paleobiology* 23:1–19

Cracraft J, Eldredge N, eds. 1979. *Phylogenetic Analysis and Paleontology*. New York: Columbia Univ. Press

Donoghue MJ, Doyle JA, Gauthier J, Kluge AG, Rowe T. 1989. The importance of fossils in phylogeny reconstruction. *Annu. Rev. Ecol. Syst.* 20:431–60

Edwards AWF, Cavalli-Sforza LL. 1963. The reconstruction of evolution. *Ann. Hum. Genet.* 27:104–5

Eldredge N. 1979. Cladism and common sense. See Cracraft & Eldredge 1979, pp. 165–98

Engelmann GF, Wiley EO. 1977. The place of ancestor-descendant relationships in phylogeny reconstruction. *Syst. Zool.* 26:1–11

Farris JS. 1983. The logical basis of phylogenetic analysis. In *Advances in Cladistics*, Vol. 2, ed. NI Platnick, VA Funk, pp. 7–36. New York: Columbia Univ. Press

Farris JS. 1989. The retention index and the rescaled consistency index. *Cladistics* 5:417–19

Felsenstein J. 1979. Alternative methods of phylogenetic inference and their interrelationships. *Syst. Zool.* 28:49–62

Fisher DC. 1980. The role of stratigraphic data in phylogenetic inference. *Geol. Soc. Am.*, *Abstr. Prog.* 12(7):426 (Abstr.)

Fisher DC. 1981. The role of functional analysis in phylogenetic inference: examples from the history of the Xiphosura. *Am. Zool.* 21:62–77

Fisher DC. 1982. Phylogenetic and macroevolutionary patterns within the Xiphosurida. In *Proc. Third North American Paleontological Convention*, ed. B Mamet, MJ Copeland, 1:175–80. Montreal: Geol. Surv. Can.

Fisher DC. 1988. Stratocladistics: integrating stratigraphic and morphologic data in phylogenetic inference. *Geol. Soc. Am.*, *Abstr. Prog.* 20(7):A186 (Abstr.)

Fisher DC. 1991. Phylogenetic analysis and its application in evolutionary paleobiology. In *Analytical Paleobiology*, ed. NL Gilinsky, PW Signor, Short Courses Paleontol., No. 4, pp. 103–22. Knoxville, TN: Paleontol. Soc.

Fisher DC. 1992. Stratigraphic parsimony. See Maddison & Maddison 1992, pp. 124–29

Fisher DC. 1994. Stratocladistics: morphological and temporal patterns and their relation to phylogenetic process. In *Interpreting the Hierarchy of Nature*, ed. L Grande, O Rieppel, pp. 133–71. San Diego, CA: Academic

Fisher DC. 1997. Stratocladistics and hypothesis choice. *J. Vert. Paleont.* 17(Suppl. 3):46A (Abstr.)

Fisher DC. 1998. Stratigraphic data and phylogenetic hypotheses. *Nature* [online debate: "Is the fossil record adequate?"] (Dec. 3, 17, 24,1998). **www.nature.com/nature/debates/fossil/fossil_frameset.html**

Fisher DC, Bodenbender BE. 2003. Blastoid stratocladistics—Reply to Sumrall and Brochu. *J. Paleont.* 77:195–98

Fisher DC, Foote M, Fox DL, Leighton LR. 2002. Stratigraphy in phylogeny reconstruction—Comment on Smith 2000. *J. Paleont.* 76:585–86

Fisher DC, Fox DL, Leighton LR. 1999. Phylogenies, temporal data, and negative evidence. *Science* 285:1179, **http://www.sciencemag.org/cgi/content/full/285/5431/1179a**

Foote M. 1996. On the probability of ancestors in the fossil record. *Paleobiology* 22:141–51

Foote M, Sepkoski JJ Jr. 1999. Absolute measures of the completeness of the fossil record. *Nature* 398:415–17

Forey P. 1995. Review: Interpreting the hierarchy of nature. *J. Vert. Paleont.* 15:861–63

Fox DL, Fisher DC, Leighton LR. 1999. Reconstructing phylogeny with and without temporal data. *Science* 284:1816–19

Geiger DL, Fitzhugh K, Thacker CE. 2001. Timeless characters: a response to Vermeij 1999. *Paleobiology* 27:177–78

Gingerich PD. 1979. Stratophenetic approach to phylogeny reconstruction. See Cracraft & Eldredge 1979, pp. 41–77

Hennig W. 1966. *Phylogenetic Systematics*. Urbana: Univ. Ill. Press

Heyning JE, Thacker CE. 1999. Phylogenies, temporal data, and negative evidence. *Science* 285:1179

Hitchin R, Benton MJ. 1997. Congruence between parsimony and stratigraphy; comparisons of three indices. *Paleobiology* 23:20–32

Huelsenbeck JP. 1997. Maximum likelihood estimation of phylogeny using stratigraphic data. *J. Vert. Paleont.* 17(Suppl. 3):54A (Abstr.)

Huelsenbeck JP, Rannala B. 1997. Maximum likelihood estimation of topology and node times using stratigraphic data. *Paleobiology* 23:174–80

Kluge AG, Farris JS. 1969. Quantitative phyletics and the evolution of anurans. *Syst. Zool.* 18:1–32

Lockwood C. 1998. Stratigraphy and phylogeny. *Evol. Anthropol.* 6:153–54

Maddison WP, Maddison DR, eds. 1992. *MacClade: Analysis of Phylogeny and Character Evolution, Version 3*. Sunderland, MA: Sinauer

Marcot JD, Fox DL. 2008. StrataPhy: a new computer program for stratocladistics analysis. *Palaeontol. Elect.* 11(1, 5A):1–16

Marshall CR. 1997a. Confidence intervals on stratigraphic ranges with nonrandom distributions of fossil horizons. *Paleobiology* 23:165–73

Marshall CR. 1997b. Can Bayes' rule be used to combine stratigraphic and morphologic data in phylogeny reconstruction? *J. Vert. Paleont.* 17(Suppl. 3):62A (Abstr.)

Norell MA, Novacek MJ. 1992. The fossil record and evolution: comparing cladistic and paleontologic evidence for vertebrate history. *Science* 255:1690–93

Norell MA, Novacek MJ. 1997. The Ghost Dance: a cladistic critique of stratigraphic approaches to paleobiology and phylogeny. *J. Vert. Paleont.* 17(Suppl. 3):67A (Abstr.)

Novacek MJ. 2001. Character distributions are phylogenetic evidence. *Paleobiology* 27:172–73

Platnick NI. 1979. Philosophy and the transformation of cladistics. *Syst. Zool.* 28:537–46

Rieppel O. 1997. Falsificationist versus verificationist approaches to history. *J. Vert. Paleont.* 17(Suppl. 3):71A (Abstr.)

Rieppel O, Grande L. 1994. Summary and comments on systematic pattern and evolutionary process. In *Interpreting the Hierarchy of Nature*, ed. L Grande, O Rieppel, pp. 227–55. San Diego, CA: Academic

Schaeffer B, Hecht MK, Eldredge N. 1972. Phylogeny and paleontology. In *Evolutionary Biology*, ed. T Dobzhansky, MK Hecht, WC Steere, 6:31–46. New York: Appleton-Century-Crofts

Scotland RW. 1992. Character coding. In *Cladistics: A Practical Course in Systematics*, ed. PL Forey, CJ Humphries, IJ Kitching, RW Scotland, DJ Siebert, DM Williams, pp. 14–21. Syst. Assoc. Publ. No. 10. Oxford: Clarendon

Siddall ME. 1997. Stratigraphic fit, bias, and a proposed solution. *J. Vert. Paleont.* 17(Suppl. 3):76A (Abstr.)

Simpson GG. 1961. *Principles of Animal Taxonomy*. New York: Columbia Univ. Press

Smith AB. 1994. *Systematics and the Fossil Record: Documenting Evolutionary Patterns*. Oxford: Blackwell

Smith AB. 1998a. What does palaeontology contribute to systematics in a molecular world? *Mol. Phylogenet. Evol.* 9:437–47

Smith AB (moderator). 1998b. Is the fossil record adequate? *Nature* [online debate]. **www.nature.com/nature/debates/fossil/fossil_frameset.html**

Smith AB. 2000. Stratigraphy in phylogeny reconstruction. *J. Paleont.* 74:763–66

Sober E. 1988. *Reconstructing the Past: Parsimony, Evolution, and Inference*. Cambridge, MA: MIT Press

Sumrall CD, Brochu CA. 2003. Resolution, sampling, higher taxa, and assumptions in stratocladistic analysis. *J. Paleont.* 77:189–94

Swofford DL. 2001. PAUP*: Phylogenetic Analysis Using Parsimony (* and Other Methods), Version 4.0b7. Sunderland, MA: Sinauer

Wagner PJ. 1995. Stratigraphic tests of cladistic hypotheses. *Paleobiology* 21:153–78

Wagner PJ. 1997. Predicting amounts of stratigraphic debt and morphologic change. *J. Vert. Paleont.* 17(Suppl. 3):83A (Abstr.)

Wagner PJ. 1998. A likelihood approach for estimating phylogenetic relationships among fossil taxa. *Paleobiology* 24:430–49

Wagner PJ. 2000. Phylogenetic analyses and the fossil record: tests and inferences, hypotheses and models. In *Deep Time: Paleobiology's Perspective*, ed. DH Erwin, SL Wing. Lawrence, KS: The Paleontol. Soc. *Paleobiology* 26(Suppl. 4):341–71

Wagner PJ. 2002. Testing phylogenetic hypotheses with stratigraphy and morphology-a comment on Smith 2000. *J. Paleontol.* 76:590–93

Wiley EO. 1981. *Phylogenetics: The Theory and Practice of Phylogenetic Systematics*. New York: Wiley

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Ecology, Evolution, and Systematics*
articles may be found at http://ecolsys.annualreviews.org/errata.shtml