# Strongest Evidence: Maximum Apparent Phylogenetic Signal as a New Cladistic Optimality Criterion

## Benjamin A. Salisbury

*Department of Biology, University of Michigan, Ann Arbor, Michigan 48109-1048*

A new method for phylogenetic inference, Strongest Evidence (SE), is described. In this method, a character's support for a phylogenetic hypothesis, its apparent phylogenetic signal, is greatest when the amount of implied homoplasy is most remarkably small given background knowledge alone. Because evolutionary rates are not assumed to be slow, background expectations for character length can be derived through modeling complete dissociation between branching pattern and character state assignments. As in unweighted parsimony, SE holds that fewer required evolutionary steps in a character indicates stronger support for a tree. However, in SE, the relationship between steps and support differs by unlabeled tree topology and character state distribution. Strongest evidence is contrasted in detail with both unweighted parsimony and Goloboff's method of implied weights. An iterative process is suggested for incrementally resolving a phylogenetic hypothesis while conducting cladistic analyses at increasingly local levels. © 1999 The Willi Hennig Society

## INTRODUCTION

Many phylogeneticists suppose that all characters should be assigned the same weight, or "no" weight, in a phylogenetic analysis (e.g., Kluge, 1997). Farris' statement above does not directly conflict with advocating unweighted parsimony. While characters clearly are not equally indicative of the historically true branching patterns, there is no way to know before or even after an analysis which characters are truthful and thus more deserving of our trust. Although the method described herein does not resolve this problem, it does allow us to measure the inferential strength of a character's evidence regarding the cladistic relationships of each phylogenetic hypothesis.

In Farris' (1983) defense of parsimony as a basis for phylogenetic analysis, he placed great emphasis on explanatory power: ". . . we wish naturally to identify the genealogy that explains as much available observation as possible. . . . Any feature shared by organisms is so either by reason of common descent or because it is a homoplasy. The explanatory power of a genealogy is consequently measured by the degree to which

No one supposes, however, that characters in general all deserve the same weight—that they all yield equally strong evidence. Drawing conclusions despite conflicting evidence requires that some evidence be dismissed as homoplasy. It is surely preferable to dismiss weaker evidence in deference to stronger. (Farris, 1983:11)

it can avoid postulating homoplasies" (Farris, 1983:18). How to measure explanatory power, or the "degree" to which ad hoc hypotheses of homoplasy can be avoided, deserves further exploration.

Lack of explanatory power can be approximated as the tally of ad hoc hypotheses of homoplasy required to reconcile all observations with a tree, i.e., extra steps in unweighted parsimony's tree length. This measure is sometimes altered through differential character and step weighting, practices that are arguably consistent with parsimony (Farris 1969, 1983; Goloboff 1993, 1995; but see Kluge, 1997). A tangential complication is the assumption that all inferred homoplasies are independent and may therefore be treated as equivalently damning; this assumption is invalid as will be shown in the next section of this paper. In light of these issues, I present below an enhanced measure of phylogenetic explanatory power, apparent phylogenetic signal, based on detailed consideration of the background assumptions of parsimony. Finding the tree with maximum apparent phylogenetic signal, in what I call a "strongest evidence" analysis, is functionally similar to both parsimony and likelihood. In light of the unclear methodological affinity of strongest evidence, it is perhaps best to view it not as a variant of a traditional method but as a fundamentally new approach to evaluating alternative phylogenetic hypotheses.

## HOMOPLASY INDEPENDENCE

To appreciate the logical basis of strongest evidence, one must understand the characteristic of traditional parsimony to which it is a reaction. In unweighted parsimony, every extra evolutionary step required to reconcile a character with a tree is treated as an equally influential, independent, ad hoc hypothesis of homoplasy. The justification for this was explained well by Kluge (1997) in his dismissal of Goloboff's (1993) implied weights method. Kluge argued that denying the independence of homoplasy instances within a character "assumes the historical dependence among lineages, a proposition which is contradicted by the historical independence of exclusive clades" (Kluge, 1997:91). In other words, homoplastic evolutionary changes occur in different historical individuals and are therefore incomparable and causally separate.

(Note: I use the term "homoplasy" strictly to refer to historically true rather than presently inferred occurrence of parallelisms and reversals. A homoplasious character may or may not require extra steps, even on the correct tree.) Although Kluge's concept of homoplasy independence is reasonable, it is not sufficient justification for unweighted parsimony. Just because true evolutionary changes occur independently in a character does not mean that the *inference* of evolutionary change is likewise free of dependencies; it is not.

An unweighted parsimony analysis does not involve counting actual evolutionary changes, independent or otherwise. What parsimony seeks instead is the phylogenetic hypothesis requiring the fewest hypothesized evolutionary changes, while the actual number of historical changes in unknowable. Tree length is not even an estimate of how many character changes occurred, but only the lower bound for this number; homoplasy might greatly exceed the minimum. This property of length follows of course from the definition of parsimony and is not in itself problematic (in the sense of Siddall and Kluge, 1997). What is problematic is that a character's minimum length on a tree is not the sum of independent items as required by parsimony. If a DNA site changed independently from A to T in two terminal lineages and all the rest have state A, the length of this character will be either one or two depending how the taxa are arranged on the tree being evaluated, even though two events did occur and involved two taxa that may or may not be true sisters. The changes implied by parsimony are not observed changes; they are inferred possibilities. These inferences are influenced not only by the true history of the characters, but also by the characters' state structures and the topologies of the trees.

Examples of these contextual dependencies can be demonstrated easily. Imagine a binary character with a state distribution completely dissociated from phylogenetic branching pattern because of mistaken homologies or rapid evolution. Even though this character has no inherent value in recovering cladistic relationships, it may influence the analysis in a biased fashion. Trees of topology X and Y in Fig. 1 are not equally likely to be refuted if this character (equiprobably) assigns two taxa to one state and six to the other. When interpreted parsimoniously on trees like those in Fig. 1, such a character would be twice as likely to appear nonhomoplasious on an X tree than on a Y tree because X has
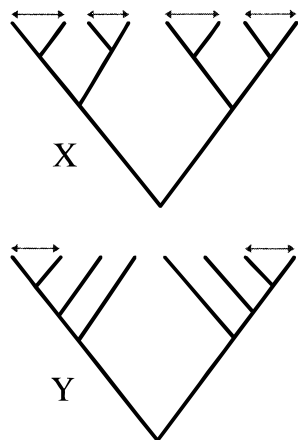
FIG. 1.  Two topologies for an eight-terminal tree. If a character state is shared by only the two terminal taxa indicated by one of the arrows, the origin of that state would appear nonhomoplasious, even if it in fact arose separately in the two taxa. A character that equiprobably assigns one state to two taxa and another state to six is twice as likely to appear nonhomoplasious on a tree of topology X than on a tree of topology Y.

twice as many pairs of terminal taxa. When data are not completely free of homoplasy, character structure and tree topology may influence a traditional parsimony analysis. Moreover, simulations strongly indicate that the topology of the true phylogeny affects the amount and inference of homoplasy, which may lead to further biases in phylogenetic inference using parsimony (Salisbury, 1999).

The Strongest Evidence (SE) approach to phylogenetic inference was designed to avoid biases like the one just described. As in unweighted parsimony, lower step counts indicate stronger support (weaker refutation) under SE, but this relationship is nonlinear and dependent upon the state structure of the character as well as the topology of the tree being considered. The relationship between character length and evidential strength is determined by reference to a background null model of nonrelationship between cladistic branching pattern and the taxon-to-state association. After finding the length of a character on a tree, SE calculates the probability of a similar character with states shuffled across the taxa being just as short or shorter on that same tree. From the SE perspective, the smaller that probability, the more convincing the observed character is as evidence supporting the phylogenetic hypothesis. Justification for use of the reference model and an example of SE analysis are given

further below after a description of the complete phylogenetic procedure.

## THE STRONGEST EVIDENCE PROCEDURE

For every phylogenetic hypothesis considered, each character, as an independent test of the tree, must be compared separately against the null model. First, the number of steps required to reconcile the character with the tree (i.e., its length) is determined. Second, an expected distribution of step counts is computed by counting the steps required for each distinct permutation (i.e., combination) of the character's state assignments across taxa. Third, of those permutations, the fraction that needs as few or fewer steps than the observed character is calculated; this fraction is the probability under the null model of the tree being at least as parsimonious an explanation of the character as observed. The smaller the probability, the greater the phylogenetic signal appears to be.

Because the characters in a data set are assumed to be independent, the probabilities individually associated with them are too; how homoplasious one randomized character appears on a tree is unrelated to how homoplasious another will seem. The probabilities of independent events can be multiplied to find the probability of both events occurring. Thus, the probability under the null model of all the characters simultaneously requiring no more steps in each than were actually "observed" on the tree in question is the product of their independent probabilities on that tree.

When discussing small probabilities, it is convenient to describe them logarithmically. I define a character's Apparent Phylogenetic Signal (APS) implied by a particular tree as the inverse $\log_{10}$ of its probability of being at least as parsimonious under the null model. Just as independent probabilities can be multiplied, their logarithms can be added. A tree's total APS, the apparent phylogenetic signal of the entire data set on the tree, is simply the sum of APS values for all the characters. Figure 2 contains a small example. The tree with the greatest total APS (tree A) is the phylogenetic hypothesis on which the data are least likely to be as parsimonious by chance and for which the evidential support is thus most remarkably "strong."
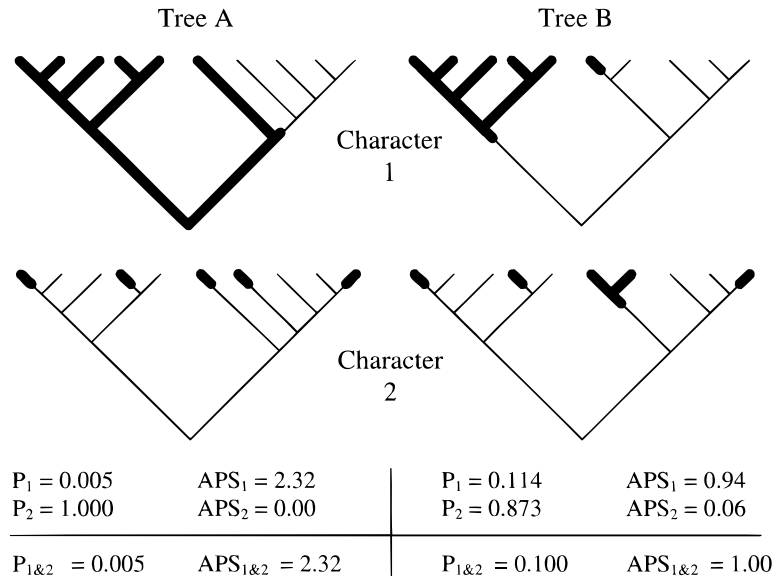
FIG. 2. A strongest evidence comparison of two trees of equal step count. Each character has two states, drawn with arbitrary optimization. $P$ is the probability of having such a parsimonious explanation of the character(s) by the tree if the tree's branching pattern and the terminal character states are unrelated. APS, the apparent phylogenetic signal of the character(s) on the tree, is $-\log_{10}(P)$. Character 1 gives much stronger support to Tree A than to Tree B, while Character 2 gives only slightly more support to B than A. Overall, the parsimony of the characters is 20 times less likely to have occurred by chance for topology A than B.

## USE OF THE NULL MODEL

I wish to emphasize, before presenting any further discussion, that SE *is not* a maximum likelihood procedure, at least not in the standard phylogenetic sense of Felsenstein (1973). SE produces results that are not conditional on the truth of evolutionary assumptions beyond those of standard parsimony, whereas maximum likelihood statistics depend on explicit assumptions about how characters evolve. The remaining discussion defends only the reasonableness of SE's model of noncorrespondence between branching pattern and character and the use of this model in evaluating alternative phylogenetic hypotheses.

The model used in SE is derived from Le Quesne's (1969, 1972) model of random agreement between characters. He posited that if a character were independent of phylogeny, its compatibility with other characters would be no better than if the character states were randomly shuffled among the taxa. Several other authors have also used Le Quesne's random shuffling model in a compatibility context (Sneath *et al.*, 1975; Meacham, 1981, 1994; Day *et al.*, 1998), but it has also

been extended to parsimony for tests of data nonrandomness (Archie, 1989; Faith and Cranston, 1991). While this null model has been widely used, the merits and demerits of its application in strongest evidence must be considered.

Unweighted parsimony (UP) and strongest evidence use phylogenetic data in very different ways. In a UP analysis, evidence for (or rather against) a phylogenetic hypothesis comes in the form of unitary inferences of evolutionary change. These inferred changes (i.e., steps) are considered equal and are added as if they were independent, regardless of whether they belong to the same or different characters. In contrast, SE treats the characters rather than the "steps" as the independent pieces of evidence. The character, with its observed distribution of states, is used to test the alternative phylogenetic hypotheses.

Goloboff (1993), in his method of Implied Weights (IW), made effectively the same distinction about characters as units of evidence as I have. IW uses a measure called "fit" to evaluate phylogenetic hypotheses. The "fit" of a character to a tree equals $k/(k + ES)$, where $k$ is a user-decided constant of concavity for the function and ES is the number of steps required beyond

the minimum possible. The individual character "fits" are totaled to find the "fit" for the whole data set on the tree. From the IW perspective, intracharacter homoplasies are biologically nonindependent; as implied by the tree, characters have different inherent evidential value. Specifically, IW approximates a biological, evolutionary concept of reliability, "the simple idea that characters which have failed repeatedly to adjust to the expectation of hierarchic correlation are more likely to fail again in the future" (Goloboff, 1993:84).

The use of reliability weighting, such as Goloboff's, has been criticized strongly by Kluge (1997) on philosophical grounds. Calling upon Karl Popper's notions of the evidential corroboration of hypotheses, Kluge (1997:92) argued that employing a concept of biological reliability adds to the assumed "background knowledge" and therefore makes phylogenetic hypotheses less testable and "decreases degree of corroboration." Kluge further noted that the biological processes responsible for a character's inherent reliability are undefined and consequently untestable. Although analogous in a some ways to IW, strongest evidence is not subject to these criticisms.

SE's character null model adds nothing to the background knowledge assumed in an unweighted parsimony analysis: descent with modification and the existence of a nonreticulate, fully bifurcating, true phylogeny of the taxa (Siddall and Kluge, 1997). SE's model stems specifically from "descent with modification" combined with not assuming that evolutionary rates are low. It follows from this absence of rate assumption that a character documenting merely the endpoints of phylogenetic descent is not assumed to retain any indication of the historical branching pattern. The permutation model of SE simply describes in detail this consequence of the minimum background knowledge. Moreover, if the phylogenetic hypothesis under investigation is wrong (as all but one are in the absence of reticulate evolution) then there is even less reason for the state distribution to reflect the hypothesized branching pattern.

According to Kluge's (1997:91) interpretation of Popperian philosophy, "a cladistic hypothesis receives corroboration from synapomorphies only to the degree that the evidence is improbable given the background knowledge alone." Calculating this degree of improbability for each tree's supporting evidence is precisely the aim of strongest evidence. The hypothetical case described in Fig. 2 shows how this logic can be applied. The complete lack of implied homoplasy for Character 1 on Tree A is far less likely given the background knowledge than the slight homoplasy implied by Tree B. Character 2, on the other hand, does little to distinguish the two tree hypotheses because its level of support for each tree is unsurprising given chance alone; consequently, Character 1 provides the more discriminating test of these two phylogenetic hypotheses. Based on just these two characters, Tree A is more highly corroborated, despite both trees requiring the same total count of inferred evolutionary steps.

The key point of the above discussion is that strongest evidence analysis does not rely on any added assumptions about biological reality. SE's primary concern is the inferential nonindependence of character state changes implied by the parsimonious interpretation of character data on a tree. Philosophical considerations aside, the following empirical comparisons may be useful for the researcher ultimately obligated to chose from the available methods of phylogenetic inference.

## AN EMPIRICAL EXAMPLE

To explore the behavior of the strongest evidence method, I conducted both SE and UP analyses of a small published data set: 20 binary characters for 14 grass species in the genus *Chloris* (Varadarajan and Gilmartin, 1983). Three of the characters were autapomorphic and (being irrelevant in both methods) were removed from the analysis. Although the character-to-taxon ratio is low, this data set is still valuable for illustrating differences between SE and UP.

The SE analysis required a new program, DNASEP, which I created by modifying DNAPARS (Felsenstein, 1993) from the PHYLIP software package. (DNASEP is available free of charge from my website, http://www.biology.lsa.umich.edu/~salisbur/.) Over 60 random taxon addition orders were used to seed the SE heuristic tree search. The UP analysis was performed using both Hennig86 and DNASEP. Decay values, which will be explained later in this paper, were estimated using DNASEP's successor program, SEPAL, which is also available from my website; 31,834 weaker

(suboptimal APS) trees and 15,816 shorter trees were examined.

The two methods arrived at dramatically different conclusions. The UP analyses found three equally parsimonious trees (Fig. 3A; unweighted length = 36; APS = 20.98 for all three trees). The SE analysis found one best tree (Fig. 3B; unweighted length = 38; APS = 21.49). The SE and UP trees agree on only four taxon bipartitions (1, 7, 8, 9). DNASEP also found four other local APS optima (separate islands under subtree pruning and regrafting) better than the UP trees.
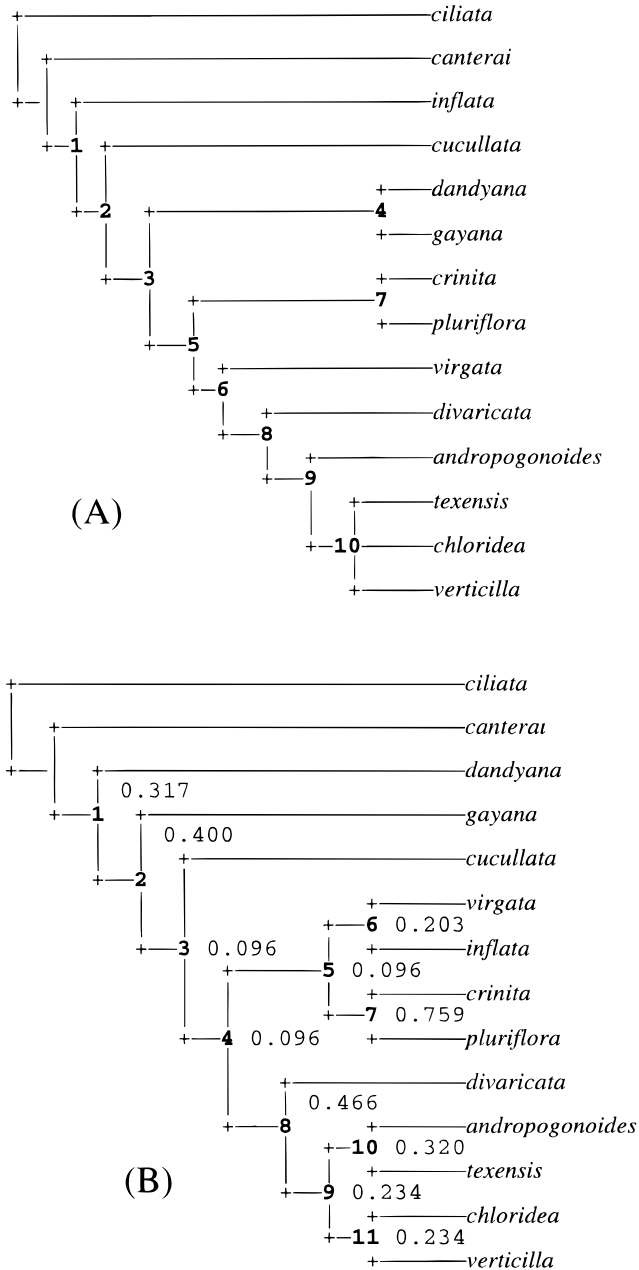
Table 1 shows the number of characters requiring



FIG. 3. Optimal trees for 14 *Chloris* species, arbitrarily drawn with *C. ciliata* in the outgroup position. (A) Strict consensus of the three shortest UP trees (unweighted length = 36; APS = 20.98). All ten bipartitions have a decay value (Bremer support) of one step. (B) Strongest SE tree (unweighted length = 38; APS = 21.49). The boldface bipartition labels are followed by support values measured as APS decay.

Number of Characters Requiring 1 through 5 Steps on the Optimal Trees Identified by UP and SE

| Method | Number of Steps | | | | |
|--------|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| UP | 3 | 10 | 3 | 1 | 0 |
| SE | 6 | 4 | 5 | 1 | 1 |

*Note.* See Fig. 3.

one, two, three, four, or five unweighted steps on the best trees. The best SE tree implies no homoplasy in 6 of the 17 characters, twice as many as on the two UP trees. Correspondingly, the SE tree requires a greater number of steps than the UP trees for some other characters. These empirical results will be discussed further below.

## COMPARING SE, IW, AND UP

Although strongest evidence was not conceived of as a weighting scheme for parsimony, it could be translated into one. Similarly, unweighted parsimony could be translated from a refutation-based to a support-based framework. I employ such transformations below to enable simpler comparison of their properties.

Seen as parsimony procedures, SE, IW, and UP can be compared as different frameworks for step weighting. In unweighted parsimony, by definition, each step in each character has the same value. The implied weights method, like SE, is support-based, but step weights can be derived easily from it. Because a constant minimum number of steps is required for a given character regardless of the tree, IW only starts giving weights to the steps in excess of the minimum. If desired, a constant weight could be assigned to all the universally required steps, but the outcome of the analysis would not be affected. The additional steps in IW are weighted separately for each character. Each step added beyond the minimum receives a weight less that the one before it. The weight of the $N$th step above the minimum equals $(k/(k + N - 1)) - (k/(k + N))$.

A similar translation for strongest evidence is conceivable but far more complex. The first layer of added complexity is that the support function, APS, is different for each unlabeled tree topology and for each character structure (i.e., number, membership size, and ordering, if ordered, of the states), unlike iterative weights with its uniform "fit" function. Moreover, the topological influences revealed by considering SE's null model demand that the "weights" of the *universally* required steps are not the same for every topology; as discussed above, a character with no inherent signal = (i.e., any hierarchical agreement it has with the tree and other characters derives from chance and does not stem from the true branching pattern *per se.*) may appear nonhomoplasious more frequently on some tree topologies than on others. With these attributes of SE taken into account, step weights could be defined, although the implementation would be needlessly complex.

Further comparison of the methods is facilitated by viewing each from the perspective of a character's differential support for a phylogenetic hypothesis. The plots in Fig. 4 show the support given to a tree by seven character types for each possible parsimonious step count. The support values are all scaled relative to zero, the value assigned when a character requires the most extra steps to fit a tree; improvements over the worst case are given positive values. Under all three methods, the sum of the support values is used as the cladistic optimality criterion to be maximized. Although it might seem odd to describe unweighted parsimony as a process of maximizing "steps saved," minimizing a function is the same as maximizing its complement; a glass contains the same amounts of water and air whether it is described as half full or half empty.

The three methods in question have much in common, as Fig. 4 shows. In all three, support for a tree hypothesis decreases as the length of a given character increases. In all three, some characters can make greater distinctions among phylogenetic hypotheses than others can. Yet, there are important differences. Under the conditions in Figs. 4A–4C, the ratios between the maximum support values (the $y$ intercepts) for a binary character distinguishing 7 taxa from 7 and a character distinguishing 2 from 12 are 6, 1.71, and 1.95, respectively, for UP, IW, and SE; from this sample, character structure appears to affect discriminating power far more in unweighted parsimony than in the other two methods.

**(A)**



**(B)**



**(C)**



**(D)**
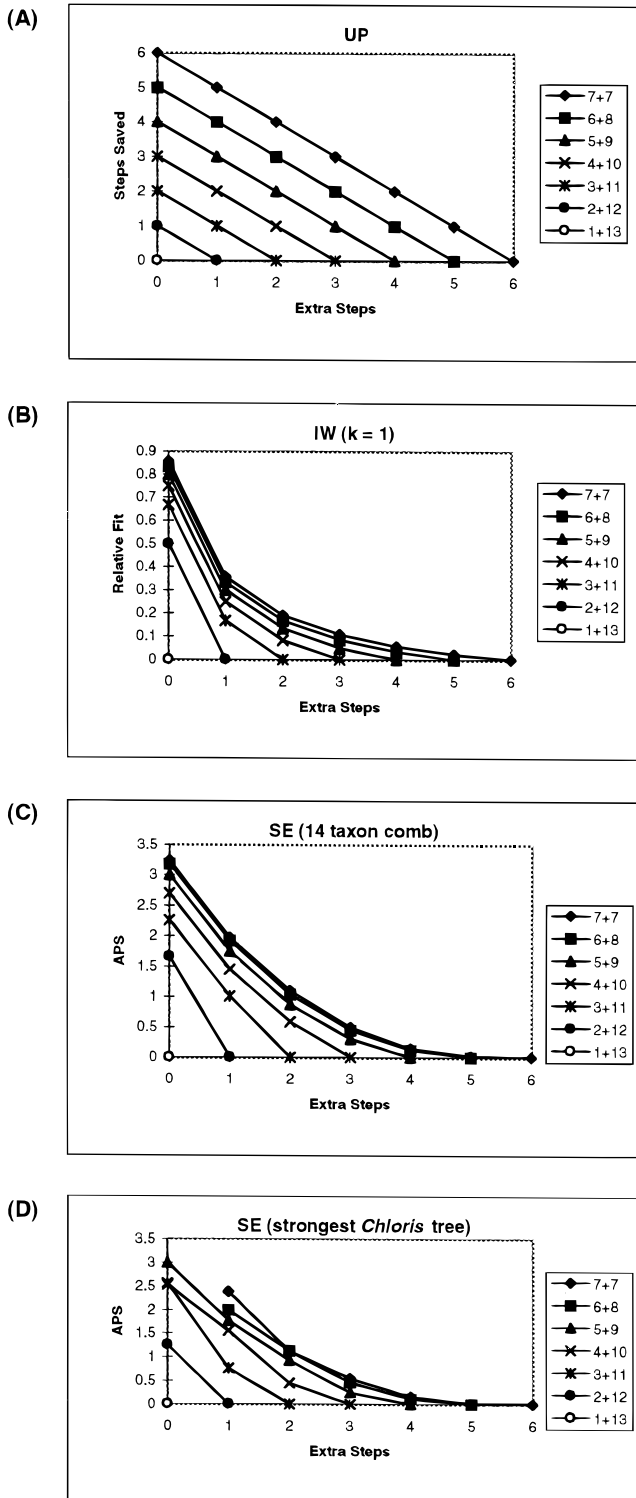


Perhaps the most remarkable observation to make regarding Fig. 4 is the similarity between the curves for IW and SE. The curves for IW would resemble the curves shown in Fig. 4C for SE even more closely if the constant of concavity were slightly higher, somewhat straightening the IW curves. This similarity is especially remarkable given that the two methods are related only by their foundation in parsimony and their recognition of a difference between intra- and inter-character homoplasy. As previously mentioned, IW is based on the assumption that characters are inherently predisposed to be more or less homoplasious in their evolution or in our interpretation of the organismal traits. In contrast, strongest evidence is not concerned with biological reliability but with inferential reliability arising from the dependencies involved in measuring a character's length on a tree. Yet, from either perspective, equal weighting of inferred character steps is contraindicated.

The resemblance of the support curves for IW and SE stems largely from their concavity. Both methods, though for different reasons, generally imply lower costs for successive step additions to a character. However, the curves are quite different when examined in finer detail. The apparent similarity seen in Figs. 4B and 4C between the two methods is not nearly so evident in a comparison of Figs. 4B and 4D. The SE support curves adjust to differences in tree topology and character structure while IW's curves are unchanging.

The curves for the best SE *Chloris* tree (Fig. 4D) may be more typically diverse than those for a perfect comb (Fig. 4C) and show many striking features not seen in the curves for a comb or for IW. First, two of the character types require a minimum of one "extra" step on this topology, no matter how the taxa names are arranged. On this tree, there are no internodes dividing

---

FIG. 4. Support curves for seven binary character types for a data set with 14 taxa using (A) unweighted parsimony, (B) implied weights with $k = 1$, (C) strongest evidence on a perfect comb, and (D) on the topology of the strongest *Chloris* tree (Fig. 3B). The units of support (scaled to a lower bound of zero) given by a character to a tree are indicated on the $y$ axes. Note that the uninformative character type, $1 + 13$, never requires extra steps and provides no tree support in any method.

seven taxa from seven or dividing six from eight. However, just because one extra step is the shortest some characters can be on a tree with this topology does not mean such a character receives a "perfect score," as revealed by contrasting Figs. 4C and 4D. Second, in two instances, curves for two characters cross each other ($4 + 10$ and $3 + 11$ at 0 extra steps and $6 + 8$ and $7 + 7$ at 2 extra steps). Third, the value of every step is different. The first extra step for a $3 + 11$ character is nearly twice as costly as the first extra step for a $4 + 10$ character; under IW the costs would be the same. Fourth, the $4 + 10$ curve is convex over the first two steps; the second extra step is actually more costly than the first, in direct conflict with Goloboff's premise. IW's fixed function does not respond to these interactions between tree topology and character structure.

The difference in homoplasy/support measurement that distinguishes SE from UP leads to some interesting contrasts, as exemplified in the results of the *Chloris* analysis. Due to the mostly decreasing costs for successive step additions, SE (like IW) will tend to cluster homoplasy in some characters while reducing the number of steps in others. UP, because it equates the cost of all steps, will spread out homoplasy more evenly among characters; it is no more reluctant to add a step to a perfect character than to a character already requiring many extra steps. This behavioral difference is arguably an attractive feature of SE. In the *Chloris* example, the UP tree indicates homoplasy in 14 of the 17 characters, whereas the SE tree indicates homoplasy in only 11 (Table 1). SE thus supports the systematists' homology determinations (Varadarajan and Gilmartin, 1983) for twice as many characters in this system, while indicating that the remaining, already homoplasious, characters are even more homoplasious, on average, than indicated by the UP trees.

Another interesting feature of SE is that typically fewer trees are expected to be equal under its optimality criterion because APS, the logarithmic SE support measure, marks smaller differences between competing tree hypotheses than does unweighted tree length. When the data conflict, there may be many shortest trees but only one or a few strongest trees. Although this precise selection might seem attractive, it may be misleading when an "unambiguous" resolution rests on only a slight improvement in apparent phylogenetic signal. UP might provide less resolution than SE, but the clades retained in a strict UP consensus

would tend to be those with greater support. For example, UP cannot resolve the relationships among the *Chloris* species *C. texensis, C. chloridea*, and *C. verticillata*; the lack of agreement in the data is shown clearly by the consensus tree (Fig. 3A) of the three shortest trees. SE identifies a single strongest tree for this same data, but one of its clades (5 in Fig. 3B) is not supported by any character state transformations, not even an ambiguous one; in the alternative sister arrangements for clades 6, 7, and 8, the unweighted lengths of all the characters are the same as in the preferred arrangement. Clades 6 and 7 are preferred as sister taxa by SE because the unweighted character lengths are (on average) slightly less likely on this tree topology under the background model.

The counterintuitive phenomenon above may be partially dealt with by adapting Bremer's support values (Bremer, 1988) (a.k.a. the decay index) as has been done similarly for IW (see example in Rognes, 1997). For each clade in a UP tree, a Bremer support value can be assigned that measures the clade's robustness as the difference in steps between the overall shortest tree and the shortest tree in which that clade does not appear. For a clade derived from an SE analysis, the decay value would be the difference in APS between the strongest tree and the strongest tree in which the clade does not appear. In the case of clade 5 in the *Chloris* SE tree (Fig. 3B), the decay value is only 0.096. Recall that APS is logarithmic; an APS decay value of 1 would indicate an order of magnitude decrease in probabilistic distinction from the null model for the whole data set. The small decay value for clade 5 indicates that the data's implication of monophyly for this clade is extremely weak, as should be expected given the lack of any supporting character transformation. (Note that none of the clades in Fig. 3 appears strongly supported under UP or SE, which is unsurprising given the low character-to-taxon ratio.) Alternatively, one might use bootstrap analysis (Felsenstein, 1985) to estimate clade support, although I find simple decay values more persuasive. No matter what method is used to infer phylogeny, some estimate of clade support must be used if one is to avoid being misled by spurious resolution. From decades of cladistic phylogenetics, it should be clear that no matter what phylogenetic inference method is used, the resulting trees are very likely to be wrong in some of the details.

There are other properties of SE that deserve comment. Because SE requires more calculations that UP, a full analysis takes longer. SE must consider all distinct permutations of the characters, so the analysis quickly becomes impossible as more taxa are added. This difficulty may be overcome by a combination of techniques including referencing a data base of precalculated step distributions for the null model and calculating (rather than enumeratively determining) the distributions as demonstrated by Maddison and Slatkin (1991). More disturbing is that SE's use of state frequencies makes it unusually sensitive to taxon sampling. Even the addition of a taxon identical to one already included can alter an SE analysis. Another aspect of SE (and IW) that might be criticized is that homoplasy in one part of a tree will deflate a character's value even if it is highly informative in a different part of the phylogeny. Fortunately, these demerits of SE may be mitigated by iterative fixation of taxon bipartitions, as is described next.

## ITERATIVE FIXATION

An unbiased phylogenetic analysis makes no prior assumption about how the study taxa are in fact related. Yet, once an analysis has been completed, there will often be strong indication that the data have revealed some of the phylogenetic relationships. An appealing notion is that a preliminary set of results might be used in some way to improve our analysis. One approach based on this notion, successive approximations, was suggested by Farris (1969). In Farris' method, after each parsimony analysis, characters are reweighted according to the number of extra steps they require on the most parsimonious trees derived from the previous set of weights. This procedure is repeated until the weights are unchanging.

Another method of successive analytic refinement is what I call "iterative fixation." In this new procedure, the most robustly supported taxon bipartition in each analysis is fixed and then the two groups of taxa remaining on either side of the bipartition are analyzed separately. Each excluded group is included as a hypothetical terminal taxon with optimized states. One additional bipartition is established in each round of analysis until no further resolution is warranted by the

data. While successive approximations reassesses the phylogenetic value of the characters over the whole tree (much like IW), iterative fixation allows a character to express itself differently at each scale of taxonomic consideration.

Iterative fixation might be criticized on the grounds that it increases the assumptional load of a study and therefore detracts from the corroborative potential of the analysis. However, this procedure is really just an automation of what many systematists have traditionally done by hand: restrict the scope of a study by lumping together taxa whose monophyly is undoubted such as when replacing the diversity of a genus with a single hypothetical taxon (or a representative species) or when analyzing only the members of one family plus token outgroups. Of course, these traditional practices may themselves be challenged on various grounds. Yet, there is another way to look at the process of phylogenetic analysis that nearly demands iterative fixation.

A cladogram is not a phylogenetic hypothesis; it is many. Each tree contains many hypotheses of relationship, which though not contradictory are also not fully interdependent. Because the historical truth of one clade (or bipartition more generally) in a cladogram is not dependent upon the truth of the others, the clades need not be accepted collectively. Perhaps each analysis should seek only the least-refuted clade, which might be judged by its Bremer support in a UP analysis. Subsequently, this process could continue until all significantly supported clades have been added. In unweighted parsimony, there is no logical reason to disregard global parsimony, so the resulting collection of accepted clades would be the original most parsimonious tree (or a strict consensus of the shortest trees).

In SE, on the other hand, iteration requires that the background knowledge be remodeled as bipartitions are fixed. Accepting a bipartition credits the data with phylogenetic signal regarding the newly accepted split; however, nothing need be assumed about the signal strength on either side. After the first split, the bipartition hypotheses remaining to be tested belong to two sets, those involving taxa on one side of the split and those involving the others. No hypothesis on either side can contradict a hypothesis on the other side, so the two sets can be evaluated separately. The best SE tree on each side may draw support from a different subset of the characters, different also from the subset

responsible for the first split. The ability to evaluate signal strength separately for these different scales and phylogenetic regions seems to hold great potential for improving analyses based on strongest evidence.

After each round of iterative fixation, the taxa on one side of a split collectively take on the role of a terminal taxon in phylogenetic analysis of the remaining taxa. Imagine a study of species A–F. It may be found by an initial SE analysis that the data most strongly support a close sister relationship for species A and B. This bipartition (splitting A and B from the rest), judged to be the least-refuted hypothesis, is accepted; subsequently, there is no need to test any hypothesis that contradicts this split of the taxa. The A plus B clade, "(AB)," is one of the taxa whose relationships are under continued investigation and it can replace the separate entries for A and B in the data matrix. The background model now relates only to data useful for resolving relationships among (AB), C, D, E, and F. The states of (AB), optimized as unambiguously as possible, can be permuted along with those of C–F to find both the strongest SE cladogram for this reduced set of taxa and the least-refuted partition within that overall branching pattern.

As promised above, iterative strongest evidence (ISE) is largely immune to the unfavorable properties of non-iterative SE. Consider first the compelling issue of taxon sensitivity. This problem is transformed when recognizing a cladogram as a composite hypothesis. In the example just described for ISE, nearly identical sister taxa were quickly identified as belonging together and were thereafter treated as a unified, individual taxon. A similar analysis might work out differently, however. The hypothesis that two nearly identical species in fact form a monophyletic clade might not be as well supported as some other aspect of the phylogeny, such as a division between two genera. If the relationship of the putative sister species is as clear as was believed at the outset, then they should still coalesce before exerting their numerical influence over the weaker-supported hypotheses regarding the remaining unresolved regions of the tree, else their inclusion and influence may have been justified after all.

Noniterative SE ignores the possibility that a character might have strong signal within one phylogenetic region and weak signal in another because it makes no distinction about where homoplasy is implied to have occurred in a cladogram. Iterative fixation allows APS support values to be calculated separately for each subproblem considered along the course of complete cladistic resolution. Where evolutionary rates have varied among lineages, ISE should consequently make better use of the data. Given these arguments, an advocate of Goloboff's implied weights method might consider a similar application of iterative fixation to IW. What could be lost, however, is the ability to recognize when a character is more globally predisposed to homoplasy; focused in on a small number of taxa, an analysis will not be able to detect the broader trend. In fact, when only four terminal taxa are under investigation, both IW and SE reduce to UP because only one extra step is possible for any character and only one unrooted tree topology exists.

Iterative fixation can also solve SE's time problem if the least controversial splits are established by faster means than full SE evaluation [e.g., parsimony jack-knifing (Farris, *et al.*, 1996) or user-imposed constraints] until the unresolved tree portions are small enough for exact ISE analysis. Iteration also reduces the problem of clades with insubstantial support appearing in the final tree. As argued above, a clade supported by no synapomorphy is likely to have a small decay value. Such a clade would not then be considered for fixation until most of the remaining cladistic relationships had been decided. At worst, the clade would be one of the last fixed and would be identified as such by its fixation order number. At best, the clade would not be fixed at all because it would be part of a tetrachotomy, which SE treats just like UP does, and no resolution would be supported.

## GENERALIZED SE AND CONCLUSIONS

Unweighted length is not the only measure of a character's homoplasy that can be modeled for use in strongest evidence. A priori weighting schemes designed for parsimony can also be applied to SE. If one feels justified in devaluing transitions relative to transversions, the background length distribution can be calculated accordingly. If one feels justified in downweighting certain classes of data (e.g., third codon positions), APS values can be scaled by those weights. Perhaps, though, SE reduces the perceived need for

such a priori decisions. In SE, any individual character may demonstrate a strong phylogenetic signal regardless of the reliability of its peers.

A further consequence of SE's flexibility and its treatment of each character as an independent source of evidence is that all kinds of characters can be analyzed together without auxiliary assumptions. Different character-specific measures of length can be employed in parallel quite simply. For instance, when sequence data and morphological data are both used, transition versus transversion weights could be set without requiring a conversion ratio between those changes and morphological changes. SE's calculation of APS requires only that the same assumptions applied in measuring a character's length are also used when generating the expected distribution of lengths under the background model. Even more remarkably, quantitative characters can be added to the mix without having to lump the measurements into discrete states. Minimum change in the trait measurements over the tree serves as quantitative length and the observed measurements are permuted as for any other character to produce a distribution of expected values from which APS can be calculated. APS serves handily as a common currency when parsimony is measured in different units of evolutionary change.

Farris (1983:11) wrote, "It is surely preferable to dismiss weaker evidence in deference to stronger." Strongest evidence analysis accomplishes this goal by calculating evidential strength through explicit consideration of the background knowledge. A character that cannot be distinguished from the null model on any of the competing phylogenetic hypotheses is surely weak evidence. In a strongest evidence analysis, such weak evidence is effectively dismissed in deference to characters whose lengths are more remarkably parsimonious. When the whole process is iterated and the analytical focus is shifted to testing alternative taxon bipartitions, analysis is refined further and characters are most influential at the levels where their signals are strong.

## ACKNOWLEDGMENTS

## REFERENCES

Archie, J. W. (1989). A randomization test for phylogenetic information in systematic data. *Syst. Zool.* **38,** 239–252.

Bremer, K. (1988). The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* **42,** 795–803.

Day, W. H. E., Estabrook, G. F., and McMorris, F. R. (1998). Measuring the phylogenetic randomness of biological data sets. *Syst. Biol.* **47,** 604–616.

Faith, D. P., and Cranston, P. S. (1991). Could a cladogram this short have arisen by chance alone? On permutation tests for cladistic structure. *Cladistics* **7,** 1–28.

Farris, J. S. (1969). A successive approximations approach to character weighting. *Syst. Zool.* **18,** 374–385.

Farris, J. S. (1983). The logical basis of phylogenetic analysis. *In* "Advances in Cladistics" (N. Platnick, and V. A. Funk, Eds.), Vol. 2, pp. 7–36. Columbia Univ. Press, New York.

Farris, J. S., Albert, V. A., Källersjö, Lipscomb, D., and Kluge, A. G. (1996). Parsimony jackknifing outperforms neighbor-joining. *Cladistics* **12,** 99–124.

Felsenstein, J. (1973). Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Syst. Zool.* **22,** 240–249.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39,** 783–791.

Felsenstein, J. (1993). "DNAPARS." Version 3.5p. University of Washington.

Goloboff, P. A. (1993). Estimating character weights during tree search. *Cladistics* **9,** 83–91.

Goloboff, P. A. (1995). Parsimony and weighting: A reply to Turner and Zandee. *Cladistics* **11,** 91–104.

Kluge, A. G. (1997). Testability and the refutation and corroboration of cladistic hypotheses. *Cladistics* **13,** 81–96.

Le Quesne, W. J. (1969). A method of selection of characters in numerical taxonomy. *Syst. Zool.* **18,** 201–205.

Le Quesne, W. J. (1972). Further studies based on the uniquely derived character concept. *Syst. Zool.* **21,** 281–288.

Maddison, W. P., and Slatkin, M. (1991). Null models for the number

of evolutionary steps in a character on a phylogenetic tree. *Evolution* **45,** 1184–1197.

Meacham, C. A. (1981). A probability measure for character compatibility. *Math. Biosci.* **57,** 1–18.

Meacham, C. A. (1994). Phylogenetic relationships at the basal radiation of angiosperms: further study by probability of character compatibility. *Syst. Bot.* **19,** 506–522.

Rognes, K. (1997). The Calliphoridae (blowflies) (Diptera: Oestroidea) are not a monophyletic group. *Cladistics* **13,** 27–66.

Salisbury, B. S. (1999). Misinformative characters and phylogeny shape. *Syst. Biol.* **48,** 153–169.

Siddall, M. E., and Kluge, A. G. (1997). Probabilism and phylogenetic inference. *Cladistics* **13,** 313–336.

Sneath, P. H. A., Sackin, M. J., and Ambler, R. P. (1975). Detecting evolutionary incompatibilities from protein sequences. *Syst. Zool.* **24,** 311–332.

Varadarajan, G. S., and Gilmartin, A. J. (1983). Phenetic and cladistic analyses of North American *Chloris* (Poaceae). *Taxon.* **32,** 380–386.