

Reconstructing Ancestral Character States Under Wagner Parsimony

DAVID L. SWOFFORD

*Section of Faunistic Surveys and Insect Identification,
Illinois Natural History Survey, Champaign, Illinois 61820*

AND

WAYNE P. MADDISON

*Museum of Comparative Zoology, Harvard University,
Cambridge, Massachusetts 01238*

Received 14 November 1986; revised 11 August 1987

ABSTRACT

The problem of assigning optimal character states to the hypothetical ancestors of an evolutionary tree under the Wagner parsimony criterion is examined. A proof is provided for the correctness of Farris's well-known, but previously unproven, algorithm for solving this problem. However, the solution is not, in general, unique, and Farris's method obtains only a subset (generally only one) of the possible solutions. Algorithms that discover other solutions and that resolve ambiguities through the imposition of ancillary criteria are developed and discussed. A method for determining the optimal length of a given tree without actually assigning character states to hypothetical ancestors is described.

1. INTRODUCTION

Several numerical methods have been developed for estimating phylogenetic trees under the principle of maximum parsimony [14]. These methods share the goal of finding minimum-length trees: those that minimize the total amount of evolutionary change needed to explain the variation in a given set of data. In the "Wagner method" for inferring phylogenies, character states are measured on an interval scale and no a priori restrictions are imposed either on the reversibility of character changes or on the number of times in which particular character-state transitions may occur [7, 21]. Algorithms that strive to optimize Wagner parsimony have achieved

Please address correspondence to David L. Swofford, Illinois Natural History Survey, 607 E. Peabody Drive, Champaign, Illinois 61820.

widespread popularity in phylogenetic analysis, largely because of their presumed freedom from assumptions about the nature of the evolutionary process. We will not address the validity of this presumption here; see Felsenstein [14] and Sober [27] for an interesting discussion on the use of parsimony in phylogenetic analysis.

Farris [7] described an algorithm for assigning optimal character states to each of the hypothetical ancestors (interior nodes) on a tree so as to minimize the tree length under the Wagner parsimony criterion. Since this procedure minimizes the *total* amount of change, it also minimizes the amount of extra change or homoplasy: character-state transitions that occur independently in different regions of the tree (parallelisms) and transitions that reflect a reversal in the evolutionary tendency of a character (reversals). When all characters are fully consistent on a tree (i.e., they can evolve on the tree with no homoplasy), Farris's procedure yields a unique solution: only one possible set of character-state assignments will be optimal for the specified tree topology. However, in the presence of homoplasy, there are often many different sets of character-state assignments that minimize the tree length [1, 12, 20, 22, 23], only some of which will be found by Farris's [7] method. If we are interested only in estimating the branching pattern of the tree, this ambiguity poses no particular problem, for any of the optimal solutions minimizes the length required for a given topology, and alternative topologies may be evaluated by comparing their optimized lengths. Frequently, however, we are interested not only in the branching pattern but also in the evolutionary hypothesis [9]: a phylogeny coupled with the reconstructed states of the characters in the hypothetical ancestors. This evolutionary hypothesis permits us to interpret branch lengths as minimal amounts of evolutionary change and to make inferences about the evolution of the characters themselves (e.g., [29]).

When multiple, equally parsimonious character-state reconstructions exist, we must be careful in interpreting any one solution. In this paper, we first describe a method for finding all equally parsimonious reconstructions and then introduce several procedures for reducing the arbitrariness in choosing among them. The implications of these results for actual biological studies will be expanded upon in a subsequent paper.

2. BASIC CONCEPTS

We are given a set of n operational taxonomic units (OTUs) whose phylogenetic relationships are summarized by an evolutionary tree or cladogram. The tree can be described mathematically (e.g., [16, 17]) as a connected acyclic graph comprising an ordered pair of sets (V, E) , where V is a nonempty, finite set of *nodes* (vertices) and E is a set of unordered pairs $\{v_i, v_j\}$, $v_i, v_j \in V$, representing branches (edges). (Because some biologists

TABLE 1

Notational Conventions	
Notation	Meaning
$a \in A$	Set membership: a is an element of set A .
$A \setminus B$	Set difference: $A \setminus B$ is a set containing all elements of A that are not also elements of B .
$A \cup B$	Union: $A \cup B$ is a set containing all items that are either elements of set A , of set B , or of both A and B .
$A \cap B$	Intersection: $A \cap B$ is a set containing all items that are elements of both A and B .
$A \oplus B$	Symmetric difference: $A \oplus B$ is a set containing all items that are either in A or B but not both A and B .
$\text{card}(A)$	Cardinality: The cardinality of set A is the number of elements it contains.
\mathbb{R}	The set of real numbers.
\mathbb{N}^+	The set of positive integers.

may be unfamiliar with the set theoretic notation used throughout this paper, we have briefly summarized this notation in Table 1.) Primarily we will consider *completely bifurcating* trees in which V is partitioned into a set V_O containing n *terminal nodes* of degree 1 and a set V_H containing $n - 2$ *interior nodes* of degree 3. The terminal nodes, labeled $1, 2, \dots, n$, correspond to the original OTUs and the interior nodes, (arbitrarily) labeled $n + 1, n + 2, \dots, 2n - 2$, represent *hypothetical taxonomic units* (HTUs). A *path* between points nodes j and k in a tree $T = (V, E)$ is a sequence of alternating nodes and branches $j = v_0, \{v_0, v_1\}, v_1, \dots, v_{m-1}, \{v_{m-1}, v_m\}, v_m = k$, where $\{v_i, v_{i+1}\} \in E$ for $0 \leq i < m$.

It will be useful to impose an arbitrary direction on the tree by designating one OTU, denoted ρ , as the *root* of the entire tree. In practice, this OTU often represents a synthesized "hypothetical ancestor" possessing putative ancestral states for each character [7], although this need not be the case. The topology of a tree may then be described by an *ancestor function* $f: V \setminus \{\rho\} \rightarrow V_H \cup \{\rho\}$ such that $f(k)$ represents the node adjacent to node k on the path between k and ρ . Let \leq be a binary relation on V such that

$v_i \leq v_j$ if and only if either v_i and v_j are coincident or v_j lies on the path from v_i to ρ . Since \leq is reflexive, antisymmetric, and transitive, it represents a partial order on V . The ancestor function f can easily be transformed into the *descendant functions* g and h . For any interior node k , $g(k)$ and $h(k)$ represent the two immediate descendants, respectively, of k [i.e., $f(g(k)) = f(h(k)) = k$]. The decision as to which descendant represents $g(k)$ can be made arbitrarily; all results below are invariant to this selection.

We define a *subtree* T_k as the subgraph (V_k, E_k) of T , where $V_k = \{v_i \in V | v_i \leq k\}$ and $E_k = \{(v_i, f(v_i)) \in E | v_i < k\}$. (Biologists refer to a rooted subtree as a monophyletic section of a tree.) A subtree consisting of a single terminal node k , $k \neq \rho$, is called *trivial*, in which case $E_k = \emptyset$. The full tree T may then be described as the triple $(T_\delta, \rho, e_{\delta\rho})$, where δ represents the interior node adjacent to ρ (coinciding with the initial bifurcation of the tree) and $e_{\delta\rho}$ is the *basal branch* connecting nodes δ and ρ . The data consist of a pair (T, \mathbf{X}) , where T is a tree and \mathbf{X} is a rectangular matrix that assigns a character state x_{ij} to each OTU i for each character j . Originally, \mathbf{X} is of dimension n by c , where c is the total number of characters. Our task is to assign states for all characters to each of the $n-2$ interior nodes (HTUs), augmenting \mathbf{X} to $(2n-2) \times c$. A *reconstruction* on tree T for character j is then given by the pair $R_{(j)} = (T, x_{(j)})$, where $x_{(j)}$ is a vector consisting of the original states $x_{1,j}, x_{2,j}, \dots, x_{n,j}$ followed by the states $x_{n+1,j}, x_{n+2,j}, \dots, x_{2n-2,j}$ chosen for the interior nodes. We will denote a reconstruction on the subtree T_k for character j as $R_{(j)k}$. The length associated with each branch of the tree in this reconstruction is then given by the *Manhattan distance* D between its incident nodes, where in general

$$D(k, l) = \sum_{j=1}^c |x_{k,j} - x_{l,j}| \quad (1)$$

for any pair of nodes k and l . The *total length* $L(R)$ of the reconstruction is simply the sum of the lengths of the branches:

$$\begin{aligned} L(R) &= \sum_{i \in V \setminus \rho} D[i, f(i)] \\ &= \sum_{i \in V \setminus \rho} \sum_{j=1}^c |x_{i,j} - x_{f(i),j}| \end{aligned} \quad (2a)$$

$$= \sum_{j=1}^c \sum_{i \in V \setminus \rho} |x_{i,j} - x_{f(i),j}|. \quad (2b)$$

Note that the equality of statements (2a) and (2b) allows us to treat each character independently; indeed, this independence among characters is a

crucial assumption of the Wagner method. Consequently, to simplify the presentation, we will usually consider only a single character, dropping unnecessary subscripts wherever possible.

In analogy to $L(R)$, the length $L(R_k)$ of the reconstruction R_k is defined as the sum of the lengths of the $2n_k - 2$ branches included in the subtree rooted at k , where n_k is the number of terminal nodes in T_k . $L^*(T)$ and $L^*(T_k)$ represent the minimum possible lengths of reconstructions for the tree T and the subtree T_k , respectively. If T_k is trivial, then x_k must be assigned the state observed in the data, and both $L(R_k)$ and $L^*(T_k)$ equal zero. A *most parsimonious reconstruction* (MPR) for tree T is a reconstruction R for which $L(R) = L^*(T)$. As noted by Farris [7], when character states take only integer values, the minimum length of a reconstruction can be thought of as the number of steps required by the tree.

In addition to the Manhattan distance function D , we will also refer to the distance d between closed intervals in \mathbb{R} , the set of real numbers. Let $S_i = \{y | a_i \leq y \leq b_i\} = [a_i, b_i]$ and $S_j = \{y | a_j \leq y \leq b_j\} = [a_j, b_j]$. The distance between S_i and S_j is the smallest possible value of $|x_i - x_j|$ where $x_i \in S_i$ and $x_j \in S_j$. This distance can also be written

$$d(S_i, S_j) = \max(a_j - b_i, a_i - b_j, 0).$$

Similarly, the distance between a real number z and an interval S_i is defined as the smallest possible value of $|z - x_i|$, where $x_i \in S_i$:

$$d(z, S_i) = \max(z - b_i, a_i - z, 0).$$

The element of S_i closest to z is the median of z , a_i , and b_i , where a_i and b_i are greatest lower and least upper bounds, respectively, of S_i . Thus, $d(z, S_i)$ can also be computed as

$$d(z, S_i) = |z - \text{median}(z, a_i, b_i)|.$$

3. STATE SETS AND FARRIS OPTIMIZATION

Farris's [7] method for assigning HTU character states so as to obtain the minimum tree length required for a given topology consists of an initial pass during which *state sets* are computed for all interior nodes on the tree and a final pass in which nonsingleton state sets are replaced by singletons. This procedure has been referred to as "Farris optimization" [25] to distinguish it from "Fitch optimization" [15, 18] in which character states are unordered and any state may transform directly into any other state.

The state set S_i is a closed interval in \mathbb{R} ; the smallest and largest elements in this interval place lower and upper bounds, respectively, on the state x_i

that will eventually be assigned to an interior node i by Farris's method. He provides two rules for computing set S_i from the state sets of the immediately descendant nodes, $S_{g(i)}$ and $S_{h(i)}$. The first rule (R-1) dictates that $S_i = S_{g(i)} \cap S_{h(i)}$ if this intersection is not empty. If $S_{g(i)}$ and $S_{h(i)}$ are disjoint, Farris's second rule (R-2) states that S_i is the smallest closed interval of the form $[a, b]$ or $[b, a]$, where $a \in S_{g(i)}$ and $b \in S_{h(i)}$.

It will be convenient to define the operator \circ for the "state set operation": the definition of a state set from two pre-existing state sets according to R-1 and R-2. Hereafter, a_i and b_i will represent the smallest and largest elements, respectively, of the state set S_i ; that is, $S_i = \{x | a_i \leq x \leq b_i\} = [a_i, b_i]$. Farris's two rules can be expressed jointly as

$$S_i \circ S_j = U \setminus (S_i \oplus S_j),$$

where $U = \{x | \min(a_i, a_j) \leq x \leq \max(b_i, b_j)\}$. A convenient computational formula is

$$S_i \circ S_j = [\min(y, z), \max(y, z)], \quad (3)$$

where $y = \text{median}(a_i, b_i, a_j)$ and $z = \text{median}(b_i, a_j, b_j)$. State sets for all nodes on the tree are computed according to the following two steps:

- 1) For each terminal node $i \in V_0$, let $S_i = \{x_i\}$.
- 2) Visit an interior node k for which S_k has not been defined but for which the state sets of the two immediate descendants, $S_{g(k)}$ and $S_{h(k)}$, have been defined. Compute $S_k = S_{g(k)} \circ S_{h(k)}$ using Equation (3). Repeat step 2 until state sets have been assigned to each of the $n - 2$ interior nodes.

Step 2 is best performed as a postorder traversal of T_δ (proceeding from the tips of the tree toward the initial bifurcation), ensuring that $S_{g(k)}$ and $S_{h(k)}$ will have been defined prior to consideration of node k . Hereafter, state sets computed in the above manner will be referred to as *Farris intervals*. Note that the Farris interval calculated for an interior node depends on which OTU was designated ρ ; subsequent redirection of a tree requires a corresponding redefinition of its Farris intervals.

After having computed Farris intervals (state sets) for all nodes on the tree steps 1 and 2 above, a final pass over the tree proceeds as follows:

- 3) For all interior nodes k whose Farris intervals are singleton, let

$$x_k = a_k (= b_k).$$

- 4) Visit an interior node k for which S_k is nonsingleton but for which the state set of its ancestor, $S_{f(k)}$, is singleton. Let x_k equal $\text{median}(x_{f(k)}, a_k, b_k)$. Replace S_k with $\{x_k\}$. Repeat step 4 until all state sets are singleton.

Step 4 is best performed as a preorder traversal of T_k (proceeding from the initial bifurcation toward the tips), as this ensures that $S_{f(k)}$ will be singleton prior to consideration of node k .¹

Farris's [7] method was presented without proof. To provide his procedure with a more rigorous foundation and to aid in the development and proof of our new algorithms, we now state several useful properties and implications of Farris intervals.

THE BASIC LEMMA

Let T_k be a subtree of T rooted at an interior node k . If k 's immediate ancestor, $f(k)$, is assigned a state $x_{f(k)} = y$, then

$$\min_{x_k \in \mathbf{R}} \{ L(R_k) + |x_k - y| \} = L^*(T_k) + d(y, S_k), \tag{4}$$

where S_k is the Farris interval for node k .

Proof. We must show that

(a) for any state assignment $x_k \in \mathbf{R}$ to node k ,

$$L(R_k) + |x_k - y| \geq L^*(T_k) + d(y, S_k); \tag{5}$$

(b) there exists at least one reconstruction for T_k such that

$$L(R_k) + |x_k - y| = L^*(T_k) + d(y, S_k). \tag{6}$$

Let \mathcal{T} be the set of rooted subtrees contained in T_k such that $T_{v_i} \in \mathcal{T}$ iff $v_i \leq k$. Define the poset $\{\mathcal{T}, \leq\}$ where for all $T_{v_i}, T_{v_j} \in \mathcal{T}$, $T_{v_i} \leq T_{v_j}$ iff $v_i \leq v_j$. The proof proceeds by induction on the partial order of \mathcal{T} . As the basis, consider the situation when T_k is trivial, in which case we must assign to node k the corresponding state x_k from the data matrix. By definition, $L(R_k) = L^*(T_k) = 0$ and $S_k = \{x_k\}$. Since S_k is singleton, $d(y, S_k) = |x_k - y|$, so that Equation (6) is always satisfied.

To complete the proof, we must show that Equation (4) is true when k is an interior node, assuming that the lemma holds for $T_{g(k)}$ and $T_{h(k)}$. Without loss of generality, we can denote nodes $g(k)$ and $h(k)$ as u and v such that $a_u \leq a_v$. For example, if $a_{g(k)} \leq a_{h(k)}$, the situation can be depicted graphically as shown in Fig. 1. The length of the subtree rooted at

¹Actually, Farris's [7] algorithm specifies that for all nodes k having nonsingleton Farris intervals, x_k is obtained as $S_k \cap S_{f(k)}$ (rule "R-3"). The substitution of our step 4 results in state assignments identical to those obtained using Farris's rule when $S_k \cap S_{f(k)}$ is not empty, but also handles the case that sometimes arises where $k = \delta$ and S_k and $S_{f(k)}$ are disjoint.

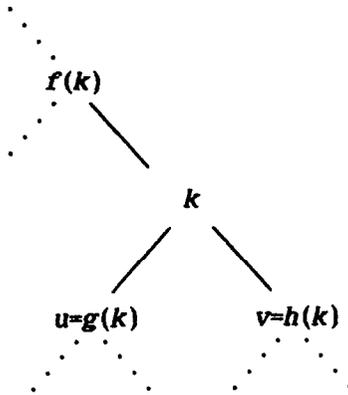


FIG. 1. A region of the tree for which the Basic Lemma is to be proved (see text).

k can then be written

$$L(R_k) = L(R_u) + |x_u - x_k| + L(R_v) + |x_v - x_k|. \quad (7)$$

Since we assume that the lemma is true for the subtrees rooted at k 's immediate descendants, we have from Equation (4) that for any given value x_k ,

$$\min\{L(R_u) + |x_u - x_k|\} = L^*(T_u) + d(x_k, S_u) \quad (8)$$

and

$$\min\{L(R_v) + |x_v - x_k|\} = L^*(T_v) + d(x_k, S_v). \quad (9)$$

It follows from Equations (7-9) that the minimum possible length of a reconstruction for T_k , given the state x_k assigned to k , is

$$L^*(T_k|x_k) = L^*(T_u) + d(x_k, S_u) + L^*(T_v) + d(x_k, S_v). \quad (10)$$

Since all terms on the right hand side of Equation (10) are nonnegative, the sum of the minimum possible lengths of reconstructions for the subtrees T_u and T_v is a lower bound on the length of an MPR for the subtree rooted at k :

$$L^*(T_k) \geq L^*(T_u) + L^*(T_v). \quad (11)$$

All possible situations with respect to S_u, S_v , and $y = x_{f(k)}$ are considered, in turn, below.

Case 1. $S_u \cap S_v \neq \emptyset$, $y \in S_k$.

Let $x_k = y$. Observe that $x_k \in S_k$, $S_k \subseteq S_u$ and $S_k \subseteq S_v$; hence, $d(x_k, S_u) = d(x_k, S_v) = 0$. Then from Equation (10), $L^*(T_k|x_k = y) = L^*(T_u) + L^*(T_v)$, which is the (unconditional) minimum possible length of a reconstruction for T_k [$= L^*(T_k)$]. Since $|x_k - y| = d(y, S_k) = 0$, Equation (6) is satisfied. If we let $x_k \neq y$, the minimum value of $L(R_k) + |x_k - y|$ exceeds $L^*(T_k) + d(y, S_k)$, for even if $L(R_k) = L^*(T_k)$, $|x_k - y| > 0 = d(y, S_k)$.

Case 2. $S_u \cap S_v \neq \emptyset$, $y \notin S_k$ ($y < a_k$).

Let $x_k = a_k$. Note that $a_k \in S_k$, $S_k \subseteq S_u$ and $S_k \subseteq S_v$; hence $d(x_k, S_u) = d(x_k, S_v) = 0$. Then from Equation (10), $L^*(T_k|x_k = a_k) = L^*(T_u) + L^*(T_v)$, which is the (unconditional) minimum possible length of a reconstruction for T_k [Inequality (11)]. Observe that $d(y, S_k) = a_k - y = |x_k - y|$, so that Equation (6) is satisfied. To establish Inequality (5) as well, we must show that $L(R_k) + |x_k - y|$ can be minimized by letting $x_k = a_k$. Suppose instead that we let $x_k > a_k$. Even if $L(R_k) = L^*(T_k)$, $x_k - y$ exceeds $a_k - y$, so that $L(R_k) + |x_k - y| > L^*(T_k) + |a_k - y|$. Alternatively, suppose that we let $x_k < a_k$. Note that since $a_k \in [a_u, b_u] \cap [a_v, b_v]$ and $a_u \leq a_v$, $a_k = a_v$. Thus, $d(x_k, S_v) = a_v - x_k = a_k - x_k = |a_k - x_k|$. Substituting into Equation (10) and recalling Inequality (11),

$$\begin{aligned} L^*(T_k|x_k < a_k) &= L^*(T_u) + d(x_k, S_u) + L^*(T_v) + |a_k - x_k| \\ &\geq L^*(T_k) + |a_k - x_k|. \end{aligned}$$

Thus, when $x_k < a_k$,

$$L(R_k) + |x_k - y| \geq L^*(T_k) + |a_k - x_k| + |x_k - y|.$$

Applying the triangle inequality,

$$L(R_k) + |x_k - y| \geq L^*(T_k) + |a_k - y| = L^*(T_k) + d(y, S_k).$$

Case 3. $S_u \cap S_v \neq \emptyset$, $y \notin S_k$ ($y > b_k$).

The proof for this case is analogous to that for Case 2 and is omitted.

Case 4. $S_u \cap S_v = \emptyset$, $y \in S_k$.

Since S_u and S_v are disjoint, $a_u \leq b_u < a_v \leq b_v$, and $S_k = [b_u, a_v]$. If $x_k \in S_k$, then $d(x_k, S_u) = x_k - b_u$ and $d(x_k, S_v) = a_v - x_k$. Then from Equation (10),

$$\begin{aligned} L^*(T_k|x_k \in S_k) &= L^*(T_u) + L^*(T_v) + (x_k - b_u) + (a_v - x_k) \\ &= L^*(T_u) + L^*(T_v) + (a_v - b_u). \end{aligned} \tag{12}$$

Equation (12) represents the minimum value of $L(R_k) + |x_k - y|$, subject to

$x_k \in S_k$, as we can reduce both terms of this expression to their minimum possible (conditional) values by letting $x_k = y$. Since $d(y, S_k)$ is then zero, Inequality (5) and Equation (6) are satisfied iff $L^*(T_k | x_k \notin S_k) \geq L^*(T_k | x_k \in S_k)$. Suppose that we let $x_k < b_u$. Substituting into Equation (10),

$$\begin{aligned} L^*(T_k | x_k < b_u) &= L^*(T_u) + L^*(T_v) + d(x_k, S_u) + (a_v - x_k) \\ &\geq L^*(T_u) + L^*(T_v) + (a_v - x_k) \\ &= L^*(T_u) + L^*(T_v) + (a_v - b_u) + (b_u - x_k) \\ &\geq L^*(T_u) + L^*(T_v) + (a_v - b_u) \\ &= L^*(T_k | x_k \in S_k). \end{aligned}$$

Likewise, if we let $x_k > a_v$, then $L(R_k)$ exceeds the value of Equation (12) by at least $(x_k - a_v)$. Thus, not only have we established the lemma for this case, but also we have proven that

$$L^*(T_k) = L^*(T_u) + L^*(T_v) + (a_v - b_u), \quad (13)$$

when $S_u \cap S_v = \emptyset$ and $a_v > b_u$.

Case 5. $S_u \cap S_v = \emptyset$, $y \notin S_k$ ($y < a_k$).

As for Case 4, $S_k = [b_u, a_v]$. Let $x_k = a_k (= b_u)$. Then $d(x_k, S_u) = 0$ and $d(x_k, S_v) = a_v - x_k = a_v - b_u$. Substituting into Equation (10) and recalling Equation (13),

$$\begin{aligned} L^*(T_k | x_k = a_k) &= L^*(T_u) + L^*(T_v) + (a_v - b_u) \\ &= L^*(T_k). \end{aligned}$$

Observe that $d(y, S_k) = a_k - y = x_k - y = |x_k - y|$, hence Equation (6) is satisfied. To establish Inequality (5) as well, we must show that $L(R_k) + |x_k - y|$ can be minimized by letting $x_k = a_k$. If instead we let $x_k > a_k$, then even if $L(R_k) = L^*(T_k)$, $|x_k - y|$ exceeds $d(y, S_k) = a_k - y$.

Alternatively, suppose that we let $x_k < a_k$. Then $d(x_k, S_v) = a_v - x_k$. Substituting into Equation (10),

$$\begin{aligned} L^*(T_k | x_k < a_k) &= L^*(T_u) + L^*(T_v) + d(x_k, S_u) + (a_v - x_k) \\ &\geq L^*(T_u) + L^*(T_v) + (a_v - x_k) \\ &= L^*(T_u) + L^*(T_v) + (a_v - b_u) + (b_u - x_k) \\ &= L^*(T_k) + (b_u - x_k) = L^*(T_k) + |a_k - x_k|. \end{aligned}$$

Thus, when $x_k < a_k$,

$$\begin{aligned} L(R_k) + |x_k - y| &\geq L^*(T_k) + |a_k - x_k| + |x_k - y| \\ &\geq L^*(T_k) + |a_k - y| = L^*(T_k) + d(y, S_k). \end{aligned}$$

Case 6. $S_u \cap S_v = \emptyset$, $y \notin S_k$ ($y > b_k$).

The proof for this case is analogous to that for Case 5 and is omitted. ■

The following result establishes the sufficiency of Farris intervals for the purpose of calculating the minimum length required by a reconstruction; the actual assignment of character states to HTUs and evaluation of equations like Equation (2) is not necessary if our only goal is to determine $L^*(T)$ or $L^*(T_k)$.

THEOREM 1

Let T_k be the subtree of T rooted at node k . Then,

$$L^*(T_k) = L^*(T_{g(k)}) + L^*(T_{h(k)}) + d(S_{g(k)}, S_{h(k)}). \tag{14}$$

Proof. Observe that

$$L(R_k) = [L(R_{g(k)}) + |x_{g(k)} - x_k|] + [L(R_{h(k)}) + |x_{h(k)} - x_k|]. \tag{15}$$

We know from the Basic Lemma that for a given value of x_k , the minimal values of the two bracketed terms on the right-hand side of Equation (15) are $L^*(T_{g(k)}) + d(x_k, S_{g(k)})$ and $L^*(T_{h(k)}) + d(x_k, S_{h(k)})$, respectively.

Consider first the case where $S_{g(k)} \cap S_{h(k)} \neq \emptyset$. If we let $x_k \in S_k$, then $x_k \in S_{g(k)}$ and $x_k \in S_{h(k)}$ as well, so that $d(x_k, S_{g(k)}) = d(x_k, S_{h(k)}) = 0$. Hence $L(R_k) = L^*(T_{g(k)}) + L^*(T_{h(k)})$. Since this is the minimum possible length of a reconstruction for T_k and $d(S_{g(k)}, S_{h(k)}) = 0$, the theorem is satisfied. Note that if $x_k \notin S_k$, either $d(x_k, S_{g(k)}) > 0$ or $d(x_k, S_{h(k)}) > 0$ (or both); hence, $L(R_k)$ exceeds $L^*(T_k)$ unless $x_k \in S_k$.

Now consider the case where $S_{g(k)}$ and $S_{h(k)}$ are disjoint. If we let $x_k \in S_k$, then $d(x_k, S_{g(k)}) + d(x_k, S_{h(k)}) = (b_k - x_k) + (x_k - a_k) = b_k - a_k = d(S_{g(k)}, S_{h(k)})$. Logic identical to that used in developing Equation (13) in the proof of the Basic Lemma then demonstrates that $L^*(T_{g(k)}) + L^*(T_{h(k)}) + d(S_{g(k)}, S_{h(k)})$ is indeed the minimum possible length, and that $L(R_k) > L^*(T_k)$ if $x_k \notin S_k$. ■

Thus, to calculate the minimum length of a reconstruction on a given tree, we can perform the initial pass of the Farris optimization procedure in the usual manner, accumulating the length required for each subtree by successive applications of Equation (14) as we proceed from the tips of the tree to

the initial bifurcation. When we arrive at node δ (the initial bifurcation), we can obtain the minimum length of a reconstruction for the full tree as

$$L^*(T) = L^*(T_\delta) + d(x_\rho, S_\delta). \quad (16)$$

The Basic Lemma guarantees that Equation (16) obtains the minimum value of $L(R) = L(R_\delta) + |x_\delta - x_\rho|$. Elimination of the final pass saves time when comparing the lengths of alternative tree topologies.

Explicit in the proof of Theorem 1 is the following corollary, which is a generalization of the “root proposition” of Maddison et al. [24]:

COROLLARY 1.1. AN MPR FOR T_k IS POSSIBLE IF AND ONLY IF $x_k \in S_k$.

This result establishes that to obtain an MPR for any subtree, it is both necessary and sufficient that the state assignment for the root node of the subtree be selected from the Farris interval computed for that node.

We are now in a position to prove that Farris optimization does indeed obtain an MPR. From the corollary above, we know that an MPR for the subtree T_δ is possible if x_δ is chosen from the interval S_δ as defined during step 2 of the initial pass. Therefore we can obtain an MPR for the full tree by choosing $x_\delta \in S_\delta$ so that the length of the basal branch, $|x_\delta - x_\rho|$, is minimized. This length is minimal when $x_\delta = \text{median}(x_\rho, a_\delta, b_\delta)$ according to step 4 of the algorithm. [The Basic Lemma guarantees that $L(R_\delta) + |x_\delta - x_\rho|$ cannot be reduced further by letting $x_\delta \notin S_\delta$.] Having established that there exists an MPR with x_δ chosen according to Farris’s final pass, we now can show by induction that further application of Farris’s method indeed does obtain state assignments that yield an MPR. That is, for any node k , if an MPR exists with $f(k)$ assigned a state $x_{f(k)}$ according to Farris’s final pass, then an MPR also exists with x_k assigned according to the algorithm. This induction step follows directly from the Basic Lemma and the fact that choosing $x_k = \text{median}(a_k, b_k, x_{f(k)})$ minimizes $d(x_{f(k)}, S_k)$.

4. MULTIPLE SOLUTIONS

The proof of the validity of Farris’s [7] optimization algorithm establishes that no reconstruction exists which yields a shorter length than the one obtained using his method. The possibility that other equally parsimonious reconstructions exist is not excluded, however. The simplest possible case where equally most parsimonious reconstructions exist is illustrated in Fig. 2. Fig. 2A shows a tree topology in which the terminal nodes take the states indicated for a particular character. Fig. 2B shows the state sets calculated according to the initial pass of the Farris algorithm. In Fig. 2C, these state sets have been replaced by the states assigned according to the final pass. Obviously, the reconstruction shown in Fig. 2D has the same length (2 steps)

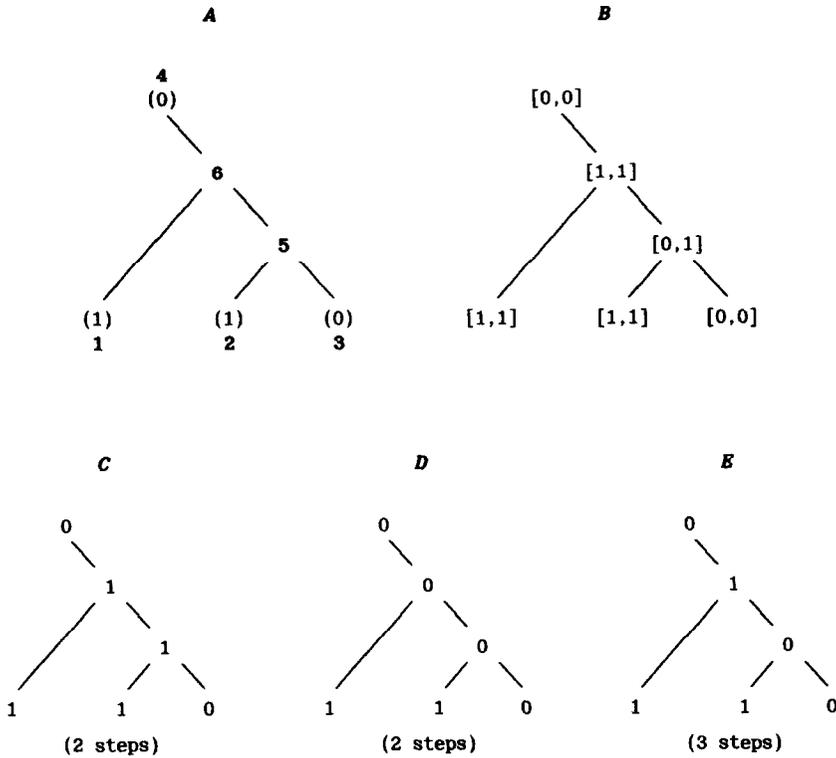


FIG. 2. An example illustrating the existence of alternative equally most parsimonious reconstructions. **A:** States taken by OTUs 1–4 are shown in parentheses, node labels in boldface. **B:** State sets assigned according to Farris's rules. **C:** Reconstruction obtained using Farris optimization. **D:** Another reconstruction requiring the same length as C. **E:** A less parsimonious reconstruction.

as that of Fig. 2C. This example makes clear the reason for the failure of Farris's [7] algorithm to find all MPRs. Although $L(R_6)$ is minimal only if $x_6 = 1$ (Corollary 1.1), we can permit a less parsimonious reconstruction for T_6 if the difference can be made up elsewhere, in this case the branch connecting HTU 6 to OTU 4. If the direction of the tree in this example is not arbitrary (i.e., if we have assumed that state 0 is plesiomorphic), the two reconstructions lead to very different interpretations of the mode of evolution of the character: Fig. 2C implies a reversal from state 1 to state 0, whereas Fig. 2D implies parallel transformations from state 0 to state 1. Of course, this is not a novel point; numerous authors have recognized the existence of alternative equally most parsimonious reconstructions (e.g., [1, 2, 6, 12, 20, 22]).

Unfortunately, the discovery of equally parsimonious reconstructions is seldom as easy as the example above might indicate, particularly when characters are not binary (0,1). One approach to recognizing the existence of alternative MPRs is to determine, for each interior node k on the tree, the set of states that might be assigned to k in at least one MPR. We will denote this set \mathcal{S}_k and will refer to it informally as k 's *MPR-set*. If \mathcal{S}_i is singleton for all interior nodes i on the tree, the reconstruction is unique. Otherwise, alternative MPRs exist. Faith [6] described an algorithm for discovering "equivocal" character state assignments for binary data (actually a special case of Fitch's [15] method for unordered multistate characters), but this method does not apply to the general case where characters are measured on an interval scale. Nonetheless, any character with linearly ordered states can be decomposed into a set of additive binary characters that, for our purposes, are jointly equivalent to the original single character [13]. Thus we could, at least in principle, use Fitch's method on the binary-coded data and obtain the MPR-sets for the original character by backtransforming the resulting single-character MPR-sets. However, this approach is both tedious and unnecessary; the methods we describe below apply equally well to binary and (linearly ordered) multistate character data. Furthermore, operating directly on the ordered characters removes a potentially confusing transformation and adds a simpler geometric interpretation, aiding comprehension of the algorithms.

Suppose that we are interested in determining the MPR-set for some interior node k (Fig. 3A). To simplify the notation to follow, let $u = g(k)$, $v = h(k)$, and $w = f(k)$. We can reroot the tree so that node k is the ancestor of the entire tree, producing a basal trichotomy in the otherwise bifurcating tree (Fig. 3B). Now calculate the Farris intervals S_u , S_v , and S_w by performing the initial pass of Farris optimization on the three subtrees descending from node k . The following result then provides a simple method for calculating \mathcal{S}_k .

THEOREM 2

The set \mathcal{S}_k of states x_k that may be assigned to node k in an MPR is

$$\mathcal{S}_k = [(S_u \circ S_v) \circ S_w] \cap [(S_u \circ S_w) \circ S_v] \cap [(S_v \circ S_w) \circ S_u]. \quad (17)$$

Proof. Let us create the new nodes p , q , and r as shown in Fig. 3C. Consider the problem of finding all possible states that might be assigned to node p in an MPR. We can do this by rooting the tree at p (Fig. 3D) and calculating the resulting Farris interval $S_p = (S_v \circ S_w) \circ S_u$. By Corollary 1.1, S_p contains all states that may be assigned to p in at least one MPR (regardless of the tree direction), so that $S_p = \mathcal{S}_p$. Likewise, we can reroot the

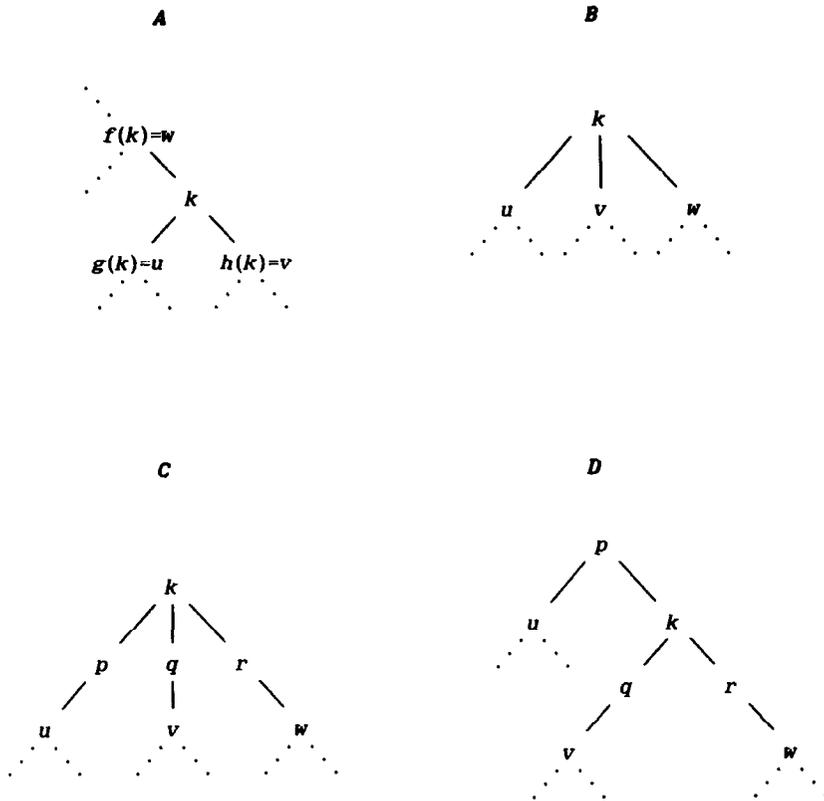


FIG. 3. Computing the MPR-set for node k (see text). **A:** Original tree. **B:** Tree after redirection with k as the root. **C:** Addition of nodes p , q , and r . **D:** Tree C after rerooting at node p .

tree, in turn, at nodes q and r , obtaining $\mathcal{S}_q = (S_u \circ S_w) \circ S_v$ and $\mathcal{S}_r = (S_u \circ S_v) \circ S_w$. First, we must show that any state in $\mathcal{S}_p \cap \mathcal{S}_q \cap \mathcal{S}_r$ is also in \mathcal{S}_k . Root the tree once more at node p (Fig. 3D). Let us assign any state $x_q \in \mathcal{S}_p \cap \mathcal{S}_q \cap \mathcal{S}_r$ to node q . After this assignment, we can simply treat node q as a terminal node (pruning off the portion of the tree descending from q) and minimize the length of the remainder of the tree. Then the (revised) Farris interval for k is equal to $\{x_q\} \circ S_w$. If $x_q \in S_w$, then $S_k = \{x_q\}$, and thus there exists an MPR with $x_k = x_q$ (as would be obtained in the final pass of Farris optimization). If instead $x_q \notin S_w$, then x_q must be either the least or greatest element of S_k according to the rules for the calculation of Farris intervals. Since $x_q \in \mathcal{S}_p \cap \mathcal{S}_q \cap \mathcal{S}_r$, it is also contained in \mathcal{S}_p . Consequently, x_p may equal x_q in an MPR, and hence we

may also assign x_q to node k (again according to the final pass of Farris optimization). Thus any $x_q \in \mathcal{S}_p \cap \mathcal{S}_q \cap \mathcal{S}_r$ can be assigned to k in an MPR.

Next, we must show that any state in \mathcal{S}_k is also in $\mathcal{S}_p \cap \mathcal{S}_q \cap \mathcal{S}_r$. Consider Fig. 3C. Suppose x_k is any element of \mathcal{S}_k and we have an MPR with $x_k, x_p, x_q,$ and x_r assigned to nodes $k, p, q,$ and $r,$ respectively. The state x_k must be the median of $x_p, x_q,$ and x_r [7]. Let us take the case where $x_p \leq x_q = x_k \leq x_r$; other cases can be argued analogously. There exists another MPR having the same state assignments as this one for all nodes other than p and $r,$ but with p and r assigned x_k as well (any increase in $|x_p - x_u|$ is exactly countered by a decrease in $|x_p - x_k|$ and likewise for $|x_r - x_k|$ and $|x_r - x_w|$). Consequently, there exists an MPR with x_k assigned to nodes p, q, r ; hence, $x_k \in \mathcal{S}_p, x_k \in \mathcal{S}_q,$ and $x_k \in \mathcal{S}_r$. ■

By examining all possible cases with respect to the intersection of $\mathcal{S}_p, \mathcal{S}_q,$ and $\mathcal{S}_r,$ we can show that \mathcal{S}_k may also be calculated as follows. Choose a pair of intervals from $S_u, S_v,$ and S_w that are maximally distant (breaking ties arbitrarily). If, for example, $d(S_u, S_v)$ is maximal, then $\mathcal{S}_k = (S_u \circ S_v) \circ S_w$; calculation of the other two terms in (17) is unnecessary. While recognizing the most distant pair of intervals and applying the appropriate sequence of state set operations is easier for hand computations, application of Theorem 2 is more efficient in computer implementations, since it avoids the need to evaluate $\max[d(S_u, S_v), d(S_u, S_w), d(S_v, S_w)]$ and then branch to separate instructions for computing \mathcal{S}_k .

A COMPUTATIONAL ALGORITHM

To obtain the MPR-sets for all interior nodes, we could visit each interior node in turn and apply Theorem 2 to each. However, this approach would require considerable duplication of effort (in calculating the Farris intervals); the following two-pass algorithm is much more efficient. Note in particular that only the original direction of the tree need be considered; successive rerooting and recomputation of Farris intervals for each rooting is unnecessary.

The first pass is identical to the initial pass of Farris optimization, visiting the interior nodes in a postorder traversal. After completing this pass we will have computed the Farris intervals S_k for each of the $n - 2$ interior nodes k . Then a second pass is performed as follows:

- 1) Let $S'_g = \{x_p\}$. Let $k = \delta$.
- 2) Let $S'_{g(k)} = S_{h(k)} \circ S'_k$ and $S'_{h(k)} = S_{g(k)} \circ S'_k$. Note that S'_k , which will have been defined in either step 1 or a previous performance of step 2, is equivalent to the Farris interval that would be calculated for node $f(k)$ if the tree had instead been rooted at k .
- 3) \mathcal{S}_k is then obtained as

$$\mathcal{S}_k = (S_{g(k)} \circ S'_{g(k)}) \cap (S_{h(k)} \circ S'_{h(k)}) \cap (S_k \circ S'_k).$$

4) If \mathcal{S}_i has been computed for all interior nodes i , stop. Otherwise let k equal the next interior node in a preorder traversal of the tree and go to 2.

The reader can easily verify that the above procedure is equivalent to successively rerooting the tree at each interior node k and defining \mathcal{S}_k according to Theorem 2.

5. OBTAINING A RECONSTRUCTION

Although definition of the MPR-sets is useful, these definitions do not lead immediately to a reconstruction unless all of the sets \mathcal{S}_k are singleton. When ambiguity exists, we cannot freely assign a state $x_k \in \mathcal{S}_k$ for each node k , as the state assignment made to one node usually places constraints on the states eligible for assignment to other nodes. That these constraints occur is clarified by the example of Fig. 2, where $\mathcal{S}_5 = \mathcal{S}_6 = [0,1]$. For any combination of states $y, z \in [0,1]$, there exists at least one MPR where $x_5 = y$ and at least one MPR in which $x_6 = z$. Yet in this example, x_5 must equal x_6 ; for instance, we cannot simultaneously let $x_5 = 0$ and $x_6 = 1$ and still obtain an MPR (Fig. 2E).

Although enumeration of all possible MPRs occasionally may be of interest, it generally will not be feasible, for if the MPR-set for even one interior node is nonsingleton, an infinite number of state assignments could be made. However, if we add the restriction that states assigned to the interior nodes must come from some predefined finite set (e.g., the character states actually observed in the data), we can enumerate all possibilities via a recursive algorithm that recalculates the set of permissible states for each interior node given the state already assigned to its ancestor. By following the implications of every possible assignment at each stage further from the root, we guarantee the discovery of all MPRs.

AN EXAMPLE

At this point an example will be useful to clarify the points made thus far and to introduce the final segment of the paper. Consider the tree shown in Fig. 4A, where the character states observed in OTUs 1 through 7 take values from the interval $[0,6]$ as indicated. Suppose that we are interested only in determining \mathcal{S}_{10} , the set of states that may be assigned to HTU 10 in at least one MPR. First, we reroot the tree at HTU 10 and compute the Farris intervals S_8, S_9 , and S_{12} as $[2,4], [5,6]$, and $[1,1]$, respectively, as shown in Fig. 4B. Using Theorem 2, we calculate the MPR-set for node 10 as

$$\begin{aligned} \mathcal{S}_{10} &= [(S_8 \circ S_9) \circ S_{12}] \cap [(S_8 \circ S_{12}) \circ S_9] \cap [(S_9 \circ S_{12}) \circ S_8] \\ &= ([4,5] \circ [1,1]) \cap ([1,2] \circ [5,6]) \cap ([1,5] \circ [2,4]) \\ &= [1,4] \cap [2,5] \cap [2,4] \\ &= [2,4]. \end{aligned}$$

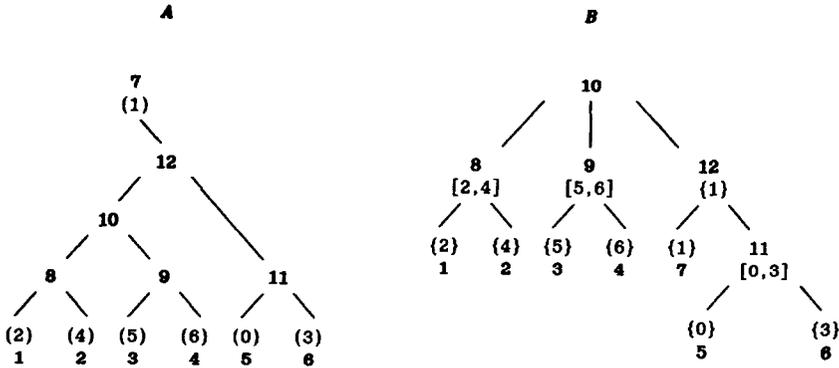


FIG. 4. **A:** Tree topology, node labelings (boldface), and character-state data (in parentheses) for examples discussed in text. **B:** Farris intervals after rerooting at HTU 10.

Now suppose we are interested in defining the complete family $\{\mathcal{S}_k\}_{8 \leq k \leq 12}$. We can use the two-pass algorithm for this purpose. The first pass consists of defining the Farris intervals S_k for each interior node k (Fig. 5A). We then let $S'_{12} = [1, 1]$ and begin the second pass, first considering node 12. Following step 2 of the algorithm, we let

$$\begin{aligned}
 S'_{g(12)} &= S_{h(12)} \circ S'_{12} \\
 S'_{10} &= S_{11} \circ S'_{12} \\
 &= [0, 3] \circ [1, 1] \\
 &= [1, 1]
 \end{aligned}$$

and

$$\begin{aligned}
 S'_{h(12)} &= S_{g(12)} \circ S'_{12} \\
 S'_{11} &= S_{10} \circ S'_{12} \\
 &= [4, 5] \circ [1, 1] \\
 &= [1, 4].
 \end{aligned}$$

We can now compute \mathcal{S}_{12} as

$$\begin{aligned}
 \mathcal{S}_{12} &= (S_{10} \circ S'_{10}) \cap (S_{11} \circ S'_{11}) \cap (S_{12} \circ S'_{12}) \\
 &= ([4, 5] \circ [1, 1]) \cap ([0, 3] \circ [1, 4]) \cap ([3, 4] \circ [1, 1]) \\
 &= [1, 4] \cap [1, 3] \cap [1, 3] \\
 &= [1, 3].
 \end{aligned}$$

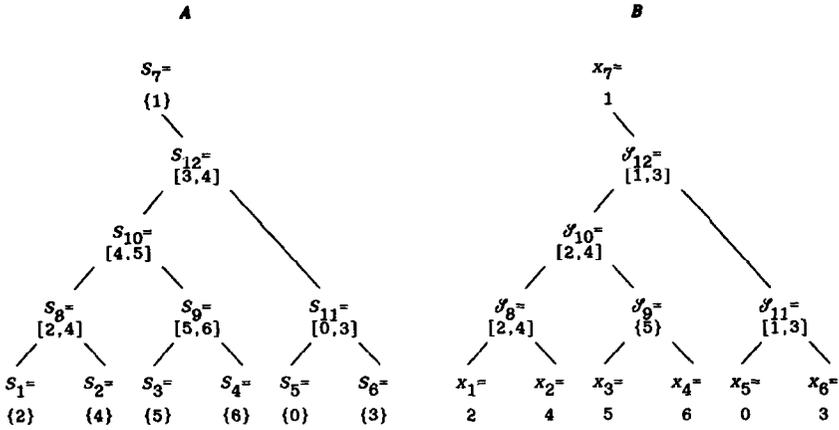


FIG. 5. A: Farris intervals resulting from the first pass of the two-pass algorithm. B: Sets of state (\mathcal{S}_k) that may be assigned to each interior node in at least one MPR.

We now move to node 10, letting

$$\begin{aligned} S'_8 &= S'_9 \circ S_{10} \\ &= [5, 6] \circ [1, 1] \\ &= [1, 5] \end{aligned}$$

and

$$\begin{aligned} S'_9 &= S'_8 \circ S_{10} \\ &= [2, 4] \circ [1, 1] \\ &= [1, 2]. \end{aligned}$$

Then,

$$\begin{aligned} \mathcal{S}_{10} &= (S_8 \circ S'_8) \cap (S_9 \circ S'_9) \cap (S_{10} \circ S'_{10}) \\ &= ([2, 4] \circ [1, 5]) \cap ([5, 6] \circ [1, 2]) \cap ([4, 5] \circ [1, 1]) \\ &= [2, 4] \cap [2, 5] \cap [1, 4] \\ &= [2, 4], \end{aligned}$$

in accord with our earlier result. We may now visit, in turn, nodes 8, 9, and 11, computing \mathcal{S}_8 , \mathcal{S}_9 , and \mathcal{S}_{11} by analogous calculations. The resulting sets \mathcal{S}_k are shown in Fig. 5B.

Now we can use the recursive algorithm briefly outlined above to generate all possible MPRs (subject to the restriction that each state assigned to an

interior node must be observed in at least one OTU). Initially, we observe that any of the states 1, 2, or 3 may be assigned to node 12. We first let $x_{12} = 1$ and then recalculate the states that may be assigned to nodes 10 and 11, given this assignment. For example, we can temporarily reroot the tree at node $k = 11$ and let $u = 5$, $v = 6$, $w = 12$, and $S_{12} = \{1\}$. Substituting into (17),

$$\begin{aligned} \mathcal{S}_{11}(x_{12} = 1) &= ([0,0] \circ [3,3]) \circ [1,1] \cap ([0,0] \circ [1,1]) \circ [3,3] \\ &\quad \cap ([3,3] \circ [1,1]) \circ [0,0] \\ &= [0,3] \circ [1,1] \cap [0,1] \circ [3,3] \cap [1,3] \circ [0,0] \\ &= [1,1] \circ [1,3] \circ [0,1] \\ &= [1,1]. \end{aligned}$$

Thus, having made the decision to assign $x_{12} = 1$, we must also assign $x_{11} = 1$. Now we must consider the possibilities for node 10 given the assignment $x_{12} = 1$. Again we temporarily reroot the tree, this time at node $k = 10$, letting $u = 8$, $v = 9$, $w = 12$, and (as before) $S_{12} = \{1\}$. Substituting into (17),

$$\begin{aligned} \mathcal{S}_{10}(x_{12} = 1) &= ([2,4] \circ [5,6]) \circ [1,1] \cap ([2,4] \circ [1,1]) \circ [5,6] \\ &\quad \cap ([5,6] \circ [1,1]) \circ [2,4] \\ &= [4,5] \circ [1,1] \cap [1,2] \circ [5,6] \cap [1,5] \circ [2,4] \\ &= [1,4] \circ [2,5] \circ [2,4] \\ &= [2,4]. \end{aligned}$$

In this case, no further restrictions were placed on the permissible states for node 10 given that $x_{12} = 1$. Therefore we must consider, in turn, the implications of assigning states 2, 3, and 4 to node 10. We first let $x_{10} = 2$. Letting $k = 8$, $u = 1$, $v = 2$, and $w = 10$, we employ calculations analogous to those above and determine that $\mathcal{S}_8(x_{12} = 1, x_{10} = 2) = \{2\}$. Since \mathcal{S}_9 is singleton, we have obtained one possible MPR, the one shown in Fig. 6A. Now we let $x_{10} = 3$. Again by calculations analogous to those above, we find that $\mathcal{S}_8(x_{12} = 1, x_{10} = 3) = \{3\}$, yielding the MPR of Fig. 6B. Similarly, $\mathcal{S}_8(x_{12} = 1, x_{10} = 4) = \{4\}$, implying the MPR shown in Fig. 6C.

Now we have exhausted the possible state assignments for node 10 given that $x_{12} = 1$. We continue the recursion by returning to node 12, letting $x_{12} = 2$, and considering the implications of this assignment for nodes 10 and 11. The reader may verify that $\mathcal{S}_{11}(x_{12} = 2) = \{2\}$ and that $\mathcal{S}_{10}(x_{12} = 2) = [2,4]$. By assigning states 2, 3, and 4, in turn, to node 10, we find the additional reconstructions shown in Fig. 6D–F. If we then backtrack once again and let $x_{12} = 3$, we ultimately obtain reconstructions 6G and 6H,

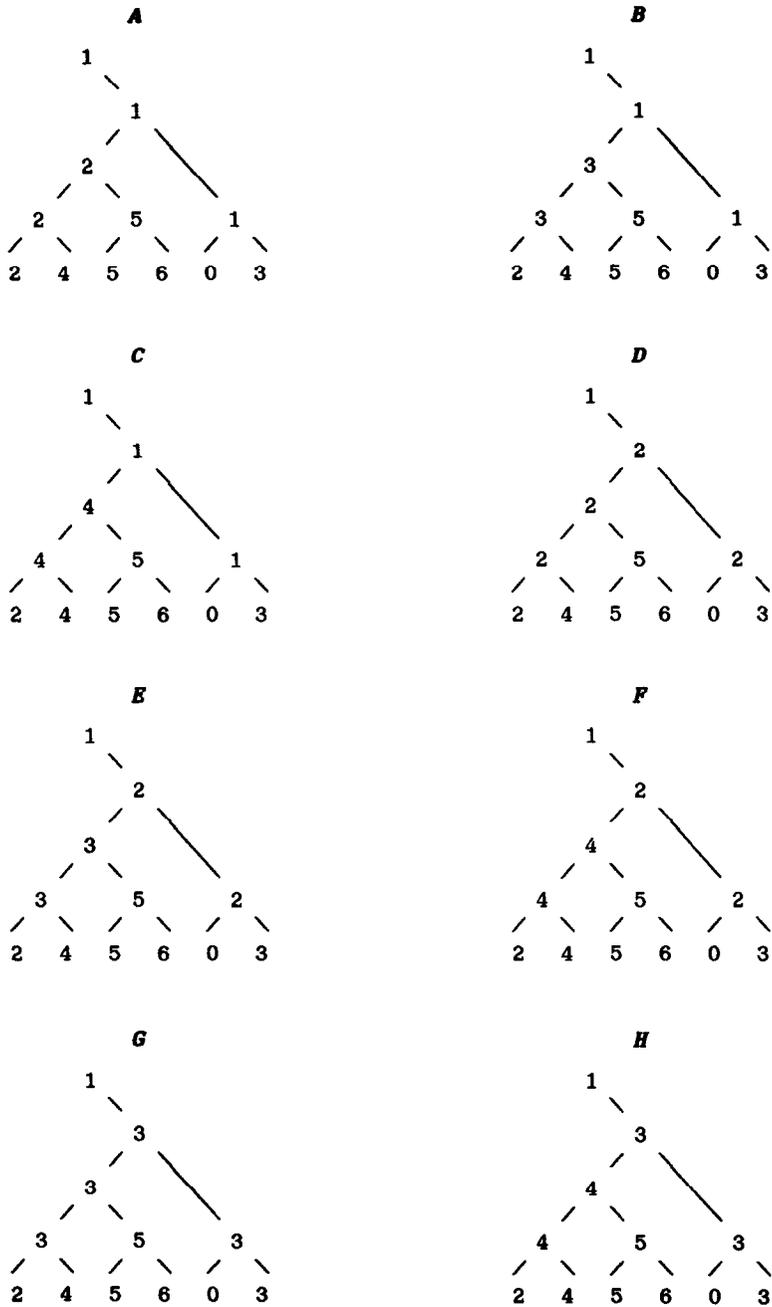


FIG. 6. MPRs for example discussed in text.

completing the search for all MPRs under the constraint that all character states assigned to interior nodes are observed in the input data matrix.

6. DEALING WITH AMBIGUITY

Although characters used in actual biological studies are seldom as homoplastic as the one in the example above, a potential dilemma is suggested nonetheless. The example considered a single character, but in real data sets, of course, there will be a number of characters, each of which may yield a nonunique set of optimal character-state assignments. How are we to choose from among the potentially vast array of suitable reconstructions? Several approaches to the resolution of the ambiguity created by the existence of alternative MPRs are described below. The specific method chosen in practical applications will depend ultimately on the investigator's willingness to make assumptions, the nature of the data, and the purpose of the study.

In many studies, certain assumptions regarding the nature of character evolution may be reasonable. For example, the transformation $0 \rightarrow 1$ may seem less probable than the transformation $1 \rightarrow 0$ for a binary character in which 0 means "absence" and 1 means "presence" (i.e., losses are more probable than gains). Thus, for a character such as the one shown in Fig. 2, reconstruction 2C (requiring one $0 \rightarrow 1$ transformation) would be preferred over that of reconstruction 2D (two $0 \rightarrow 1$ transformations), despite their being "equally parsimonious" in requiring 2 units of change. Evolutionary changes such as those shown in Fig. 2C usually are called *reversals*, while Fig. 2D exemplifies *parallelism*. Of course, if the researcher is able to develop a specific model that assigns probabilities to the various events, a maximum likelihood method tailored specifically to these assumptions would provide a better framework for reconstructing ancestral character-state distributions. For certain extreme cases, heuristic alternatives to full maximum likelihood estimation may provide satisfactory results. For example, if the probability of a gain-loss sequence is *much* greater than the probability of parallel gains, the "Dollo method" [10], in which multiple gains are expressly prohibited, may be employed [4]. Similarly, if parallelisms are *much* more likely than reversals, the method of Camin and Sokal [3], which does not allow reversals, is appropriate.

In most cases, however, alternatives to unrestricted parsimony are inappropriate or simply not available. The requirement for a meaningful probability model often excludes the use of maximum likelihood methods. Camin-Sokal and Dollo methods may be far too strict in their prohibitions of reversals and parallel gains, respectively. Nonetheless, it may be desirable to use the relative amounts of parallelism and convergence as an ancillary criterion in choosing among competing MPRs. A more complete presentation of this point may be found in Swofford (in preparation).

For example, although one might be unwilling to assume that reversals are impossible, one might prefer a reversal-free hypothesis to a hypothesis that requires one or more reversals if the two hypotheses are “equally parsimonious.” By post-poning character changes as long as possible as we move from the root of the tree toward the tips, we will obtain a reconstruction that maximizes the proportion of the total homoplasy accounted for by parallelism as opposed to reversal; the *delayed transformation optimization* (DELTRAN) procedure, described below, facilitates this approach.

DELAYED TRANSFORMATION ALGORITHM

- 1) Determine the sets \mathcal{S}_k for all interior nodes k using the two-pass algorithm described above.
- 2) Initially, let $k = \delta$.
- 3) Let x_k equal the state in \mathcal{S}_k that minimizes $|x_k - x_{f(k)}|$. Note that $x_{f(k)}$ will have been defined in a previous step.
- 4) If a state x_k has been assigned to each interior node k , stop. Otherwise, let k equal the next interior node in a preorder traversal of the tree and go to 3.

That the DELTRAN algorithm yields an MPR follows by induction using the following proposition: If there exists an MPR with node v assigned x_v , then there exists an MPR with v assigned x_v and v 's descendant node, w , assigned the state in \mathcal{S}_w closest to x_v . Suppose that node v is assigned a state $x_v \in \mathcal{S}_v$. Consider two possible assignments to w : c_w , the element in \mathcal{S}_w closest to x_v , and $d_w \in \mathcal{S}_w - \{c_w\}$. The assignment d_w could be more parsimonious than c_w only if the reduction it allows in the length of T_w is greater than the cost in additional length ($|d_w - c_w|$) incurred by not selecting $x_w = c_w$. However, if d_w reduces $L(R_w)$ by more than $|d_w - c_w|$, then it would always be best to let $x_w = d_w$ rather than c_w , since the increased length between nodes v and w would be at most $|d_w - c_w|$, regardless of what v were assigned. Thus, c_w could not be assigned to w in any MPR. But this contradicts $c_w \in \mathcal{S}_w$. Thus the choice of d_w over c_w is not more parsimonious, and c_w may be assigned to node w , given that x_v has been assigned to node v .

For the example data of Fig. 4A, the DELTRAN procedure first obtains the MPR-sets shown in Fig. 5B. Then the following state assignments are made in turn: $x_{12} = 1$, $x_{10} = 2$, $x_8 = 2$, $x_9 = 5$, $x_{11} = 1$, yielding the reconstruction shown in Fig. 6A. The absence of reversals is apparent in that all character transformations are monotonic as we move from the root of the tree to any tip. All of the “extra” change or homoplasy comes in the form of parallelisms.

In other cases, one might prefer reversals over parallelisms in choosing between alternative MPRs. If so, the original optimization method of Farris

[7], as described above, may be applied. With this algorithm, the state assigned to any particular node k is contained in the Farris interval S_k , and the length required by the subtree rooted at k , $L(R_k)$, takes its minimum possible value. Since $L(R_k) + |x_k - x_{f(k)}|$ must be minimized in any MPR, if $L(R_k) = L^*(T_k)$, then $|x_k - x_{f(k)}|$ takes its maximum possible value [given the state already assigned to node $f(k)$]. Thus, as we move from the root toward the tips in the final pass, the Farris method always assigns the maximum possible amount of change to each branch, given the state assignments made thus far. When homoplasy is present, it is more likely to be reflected in reversals of earlier changes in the same lineage than in parallel changes in different lineages. The name *accelerated transformation optimization* (ACCTRAN) makes the operating characteristics of Farris's [7] method more explicit; hence we prefer it. To apply the ACCTRAN algorithm to the example data of Fig. 4A, we first calculate the Farris intervals shown in Fig. 5A. Then we assign $x_{12} = 3$, $x_{10} = 4$, $x_8 = 4$, $x_9 = 5$, and $x_{11} = 3$, in that order. The resulting reconstruction is shown in Fig. 6H. Note that a considerable amount of reversal is implied by this reconstruction, as exhibited by the $1 \rightarrow 3 \rightarrow 4 \rightarrow 2$ and $1 \rightarrow 3 \rightarrow 0$ transformation sequences.

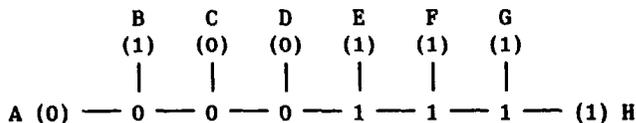
Maximization of parallelism (DELTRAN) and reversal (ACCTRAN) in the choice of optimal MPRs is appropriate only when the OTU designated ρ (the ancestor of the entire tree) exhibits the putative ancestral condition for each character; the evolutionary interpretation of these algorithms is lost when ρ is selected arbitrarily. The Wagner method is often applied in cases where the ancestral conditions (polarities) are unknown; the result is simply an undirected tree (which may or may not be converted subsequently into a rooted tree by incorporating additional information).

In these situations, the DELTRAN and ACCTRAN algorithms would have the undesirable property of being dependent on the arbitrary selection of ρ . An optimality criterion that is invariant to the selection of ρ is the f -value of Farris [8], a measure of the goodness-of-fit of the matrix \mathbf{P} of path-length distances implied by a reconstruction to the original distance matrix \mathbf{D} . Specifically, let p_{ij} represent the sum of the lengths of all branches in the path between terminal nodes i and j , and let d_{ij} be the Manhattan distance between i and j as defined by (1). Then

$$f = \sum_{i > j} (p_{ij} - d_{ij}). \quad (18)$$

(Note that d_{ij} can never exceed p_{ij} .) The f -value represents the sum of the elements of the upper (or lower) triangle of the homoplasy matrix \mathbf{H} , where $h_{ij} = p_{ij} - d_{ij}$, and therefore sometimes is referred to as the total (pairwise) homoplasy. Other measures, such as the F -value of Prager and Wilson [26] and the "deviation ratio" (DR) used in Farris's Wagner 78 program are

A



B

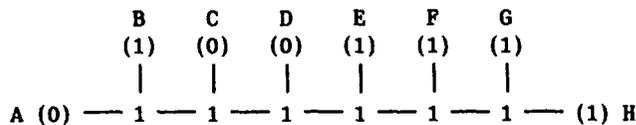


FIG. 7. An example in which the reconstruction that minimizes f is not an MPR. A: The only MPR for OTU states indicated in parentheses ($L=2$, $f=4$). B: A less parsimonious reconstruction with a lower f ($L=3$, $f=3$).

simple functions of f with $\sum d_{ij}$ used as a normalizing factor. Although the f -value and related measures are used more frequently for trees constructed from distance matrices, some authors (e.g., Donoghue [5]) have reported only total pairwise homoplasy (deviation ratio) statistics for trees constructed from character-state data, with no mention of total tree length. Jensen (in [23]) has even suggested the use of f as a possible criterion in choosing among tree topologies. However, it makes little sense to compare f -values for different trees unless this criterion has been minimized for each topology [11, 28]. A serious difficulty with the use of f as the sole optimality criterion is that it often can be fully optimized only at the expense of parsimony—the reconstruction that minimizes f is not necessarily an MPR. The example in Fig. 7 illustrates this point. The reconstruction of Fig. 7A is the only MPR for the data, requiring a length of two steps, whereas the less parsimonious reconstruction of Fig. 7B requires three. However, there are only three nonzero elements in each triangle of the homoplasy matrix resulting from the latter reconstruction, so that $f = h_{AC} + h_{AD} + h_{CD} = 3$. In contrast, each triangle of the homoplasy matrix resulting from the reconstruction of Fig. 7A contains four nonzero elements, with $f = h_{BE} + h_{BF} + h_{BG} + h_{BH} = 4$. This property of the f -value effectively disqualifies it as a criterion for selecting tree topologies unless we are willing to abandon the parsimony criterion.

Despite the above caveats, it might nonetheless be desirable to optimize f under the constraint that we hold the length of the reconstruction to its minimum possible value. That is, from the set of all MPRs, we will choose

the one(s) with the smallest f . We can rewrite (18) as

$$f = \sum_{i > j} p_{ij} - \sum_{i > j} d_{ij}.$$

Since $\sum d_{ij}$ is constant for a particular data set, minimizing $\sum p_{ij}$ is equivalent to minimizing f . Furthermore, $\sum p_{ij}$ is simply a weighted function of the branch lengths, where the weights are given by the number of times each branch length is included in the calculation of $\sum p_{ij}$ over all distinct pairs of terminal nodes (i, j) . Since a branch length is included in the calculation of p_{ij} if and only if it lies on the path between i and j , we can define the weighting function $w: E \rightarrow \mathbb{N}^+$ as

$$w(e_k) = n_k(n - n_k),$$

where $e_k = (k, f(k))$, n_k is the total number of OTUs included in k 's subtree ($=1$ if k is terminal) and n is the total number of OTUs. We will abbreviate $w(e_k)$ to w_k . An algorithm for finding all MPRs that minimize f (subject to the restriction that states assigned to interior nodes must have been observed in at least one OTU) follows:

MINIMUM F-VALUE (MINF) ALGORITHM

1) Compute \mathcal{S}_k for each interior node k using the two-pass algorithm described above.

2) Locate an interior node, denoted u , such that the sum of the weights of its three incident branches, $w_{g(u)} + w_{h(u)} + w_u$, is maximal. This node corresponds to a *centroid* of the tree. (Basic results in graph theory obtain that a tree has at least one and at most two centroids, and that two centroids must be adjacent.) Reroot the tree at u and define new ancestor and descendant functions f' , g' , and h' in accord with this redirection.

3) For each state $z \in \mathcal{S}_u \cap \{x_i | 1 \leq i \leq n\}$:

- a. Let $x_u = z$. Complete a reconstruction for each of the three descendant subtrees according to steps 3–4 of the delayed transformation algorithm, where $k = u$ initially.
- b. Compute the weighted sum of branch lengths:

$$L_w = \sum_{k \neq \rho} w_k |x_k - x_{f'(k)}|.$$

4) Any reconstruction obtained in step 3 that minimizes L_w over all reconstructions evaluated is optimal.

The MINF algorithm shares with DELTRAN the property of choosing x_k so as to minimize the length of the branch $(k, f(k))$, given the state

already assigned to $f(k)$. Since the weights of the branches are nonincreasing as we move from node u toward the tips, it is clear that any performance of step 3 minimizes L_w , given the state assigned to u , and that if we try all permissible choices of u we will obtain the globally optimal reconstruction(s). To apply the MINF algorithm to the example of Fig. 4A, we first calculate the branch weights

$$w_i = \begin{cases} 6 & \text{for } i = 1, 2, 3, 4, 5, 6, 12 \\ 10 & \text{for } i = 8, 9, 11 \\ 12 & \text{for } i = 10 \end{cases}$$

and then observe that letting $k = 10$ maximizes $w_{g(k)} + w_{h(k)} + w_k$. The state assigned to node 10 must be chosen from $[2, 4] \cap \{0, 1, 2, 3, 4, 5, 6\} = \{2, 3, 4\}$. If we let $x_{10} = 2$, then we also assign state 2 to nodes 8, 11, and 12, obtaining the reconstruction shown in Fig. 6D (x_9 must always equal 5). For this reconstruction, $L_w = 10|5 - 2| + 6(|4 - 2| + |6 - 5| + |0 - 2| + |3 - 2| + |2 - 1|) = 72$. Choosing $x_{10} = 3$ yields the reconstruction of Fig. 6G, with $L_w = 68$. Finally, the reconstruction of Fig. 6H is obtained when we let $x_{10} = 4$, with a corresponding L_w of 70. Consequently, the reconstruction in which $x_{10} = 3$ is optimal. Since $\sum d_{ij} = 56$, the f -values for the three reconstructions are 16, 12, and 14, respectively.

ARBITRARY RESOLUTIONS

While the use of path-length fit as an optimality criterion may seem somewhat ad hoc, the MINF algorithm has an alternative justification. A partition $\{V_1, V_2\}$ of the set of OTUs V_0 is said to be *supported* by a reconstruction if there exists an interior node k such that (i) either V_1 or V_2 equals the set $\{v_i \in V_0 | v_i \leq k\}$ and (ii) $x_k \neq x_{f(k)}$. When $\min(\text{card}(V_1), \text{card}(V_2)) \geq 2$ and the direction of the tree is not arbitrary, the transformation $x_{f(k)} \rightarrow x_k$ represents a *synapomorphy* [19] of the OTUs descendant from k , and the partition comprises a *resolution* of the terminal taxa (OTUs).² When more than one MPR exists, some of these may support a particular resolution whereas others do not; in such cases the resolution is said to be *arbitrary*. For example, compare the two reconstructions in Fig. 8. Reconstruction 8A assigns a change to the branch leading to the common ancestor of OTUs A and B; this change supports an $\{\{A, B\}, \{C, D, E\}\}$

²We should clarify our notion of *support*. A character supports a resolution if there is an unambiguous change along the branch leading to the common ancestor of a group of taxa. For instance, in some situations if the character subsequently reverses in one taxon within the group, it would be inappropriate to claim that the character supports the group's monophyly *except* in the context of other characters that place the reversing taxon within the group.

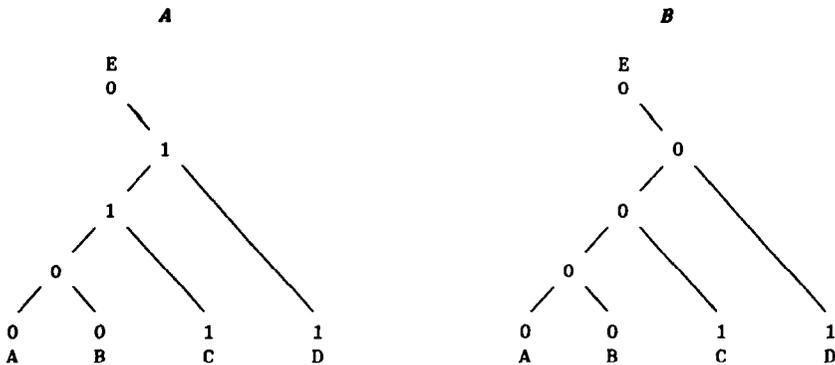


FIG. 8. **A**: An MPR in which a character change occurs along the branch leading to the {A,B}-subtree. **B**: Another MPR for the same data (see text).

partition and the apparent synapomorphy might be taken as support for the phylogenetic hypothesis that A and B share a more recent common ancestor than does either with the remainder of the OTUs. However, the equally parsimonious reconstruction of Fig. 8B assigns no change to the branch leading to the ancestor of A and B, indicating that support for the full resolution shown in Fig. 8A is ambiguous, at best. If unambiguous support for an $\{\{A,B\},\{C,D,E\}\}$ partition exists, it will have to come from some other character. The MINF algorithm, in transferring length onto terminal branches wherever possible, minimizes the total amount of support available for resolutions that may in fact be arbitrary. In this example, the MINF algorithm, in choosing reconstruction 8B, adopts a more "conservative" interpretation, preferring lack of resolution when this is equally parsimonious [12].³

Choosing the reconstruction that minimizes f does not guarantee the absence of arbitrary resolutions, however, as length transferred from the center of the tree may result in an arbitrary resolution further toward the tips, as illustrated in Fig. 9. The MINF reconstruction (Fig. 9B) avoids an arbitrary resolution defining the group $\{1,2,3\}$ but does support a $\{4,5\}$ group. However, the latter resolution is arbitrary (cf. Fig. 9C). The only way to be certain that a particular resolution is not arbitrary is to obtain the minimum possible length of each interior branch. If this length exceeds zero,

³Farris (see [12, p. 425]) apparently independently developed a similar method; since his algorithm is unpublished, we do not know if it is identical to the MINF procedure we describe.

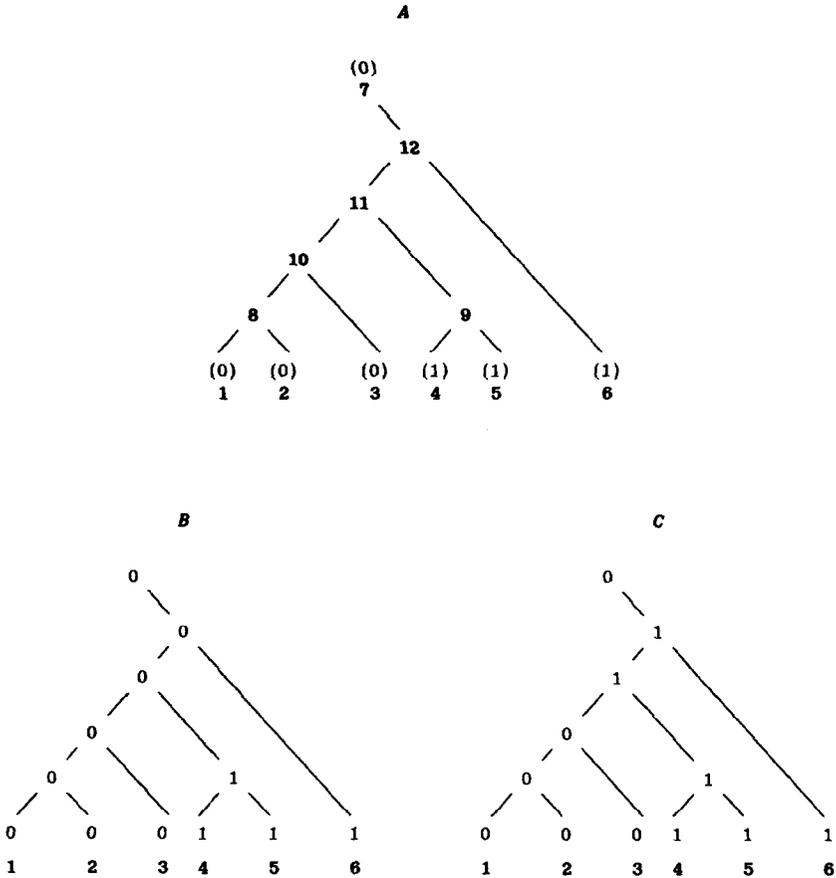


FIG. 9. **A:** Tree topology, node labels (boldface), and character-state data (in parentheses) for example discussed in text. **B:** An MPR that both minimizes f and permits an arbitrary resolution ($f = 4$). **C:** An alternative MPR in which the $\{4,5\}$ group is not supported ($f = 6$).

then at least some change will be assigned to that branch regardless of which MPR is chosen. The minimum possible length that any branch $(k, f(k))$ can take in an MPR is equal to $d(\mathcal{S}_k, \mathcal{S}_{f(k)})$ (see proof of the DELTRAN algorithm). That is, if the intersection of \mathcal{S}_k and $\mathcal{S}_{f(k)}$ is nonempty, we may assign any state in the intersection to both nodes, with a branch length of zero, and if \mathcal{S}_k and $\mathcal{S}_{f(k)}$ are disjoint, we can assign to each node the state closest to the other node's MPR-set. For the example in Fig. 9, $\mathcal{S}_9 = \{1\}$, $\mathcal{S}_{10} = \{0\}$ and $\mathcal{S}_{11} = [0,1]$, so that the minimum lengths of both the $(9,11)$ and $(10,11)$ branches are zero.

The amount of *potential* support for a given resolution is also of interest. Whenever $\mathcal{S}_k \neq \mathcal{S}_{f(k)}$, at least one reconstruction will support the resolution, as we can assign to at least one node a state from the symmetric difference of the two MPR-sets. To determine the *maximum* support available for the resolution, we can calculate the maximum possible length of the branch $(k, f(k))$ in an MPR by rerooting the tree at a point p between k and $f(k) = u$ and calculating the Farris intervals for k and u after the redirection. From Theorem 1, $L^*(T_p) = L^*(T_k) + L^*(T_u) + d(S_k, S_u)$. It follows that $d(S_k, S_u)$ is the maximum possible value of $|x_k - x_p| + |x_u - x_p| = |x_k - x_{f(k)}|$, the length of the $(k, f(k))$ branch. An easy way to compute maximum lengths for all branches of the tree is to note that the sets S_k and S_u are equivalent to S_k and S'_k as defined in the computational algorithm for obtaining all of the MPR-sets. The maximum possible length of each branch $(k, f(k))$ is simply $d(S_k, S'_k)$.

7. COMPUTER PROGRAMS

All of the algorithms described above (with the exception of exhaustive enumeration of all possible MPRs) are contained in the PAUP (Phylogenetic Analysis Using Parsimony) program, available from the first author. The program MacClade, available from the second author, contains the main algorithm finding \mathcal{S} for all nodes and allows the user to set the state at selected nodes and thus to examine any MPR. Both PAUP and MacClade contain a number of other phylogenetic algorithms and capabilities.

The authors express their thanks to Joseph Felsenstein, David Maddison, and David Penny for helpful discussions and to Ian Henderson and an anonymous reviewer for extremely useful comments on the manuscript. Travel support for WPM was provided by Harvard University.

REFERENCES

- 1 J. W. Archie, Prediction = parsimony or partitions?, in *Numerical Taxonomy* (J. Felsenstein, Ed.). NATO ASI Series Vol G1, Springer-Verlag, Berlin, 1983, pp. 102–106.
- 2 S. H. Berlocher, Insect molecular systematics, *Ann. Rev. Entomol.* 29:403–433 (1983).
- 3 J. H. Camin and R. R. Sokal, A method for deducing branching sequences in phylogeny, *Evolution* 19:311–326 (1965).
- 4 R. W. DeBry and N. A. Slade, Cladistic analysis of restriction endonuclease cleavage maps within a maximum-likelihood framework, *Syst. Zool.* 34:21–34 (1985).
- 5 M. J. Donoghue, A preliminary analysis of phylogenetic relationships in *Viburnum* (Caprifoliaceae s.l.), *Syst. Bot.* 8:45–58 (1983).
- 6 D. P. Faith, Information content and most parsimonious trees, in *Numerical Taxonomy* (J. Felsenstein, Ed.). NATO ASI Series Vol. G1, Springer-Verlag, Berlin, 1983, pp. 107–111.

- 7 J. S. Farris, Methods for computing Wagner trees, *Syst. Zool.* 19:83–92 (1970).
- 8 J. S. Farris, Estimating phylogenetic trees from distance matrices, *Am. Nat.* 106:645–668 (1972).
- 9 J. S. Farris, A probability model for inferring evolutionary trees, *Syst. Zool.* 22:250–256 (1973).
- 10 J. S. Farris, Phylogenetic analysis under Dollo's Law, *Syst. Zool.* 26:77–88 (1977).
- 11 J. S. Farris, Distance data in phylogenetic analysis, in *Advances in Cladistics. Proceedings of the First Meeting of the Willi Hennig Society* (V. A. Funk and D. R. Brooks, Eds.), New York Botanical Garden, Bronx, New York, 1981, pp. 3–23.
- 12 J. S. Farris, Simplicity and informativeness in systematics and phylogeny, *Syst. Zool.* 31:413–444 (1982).
- 13 J. S. Farris, A. G. Kluge, and M. J. Eckardt, A numerical approach to phylogenetic systematics, *Syst. Zool.* 19:172–191 (1970).
- 14 J. Felsenstein, Parsimony in systematics: biological and statistical issues, *Ann. Rev. Ecol. Syst.* 14:313–333 (1983).
- 15 W. M. Fitch, Toward defining the course of evolution: minimal change for a specific tree topology, *Syst. Zool.* 20:406–416 (1971).
- 16 R. L. Graham and L. R. Foulds, Unlikelihood that minimal phylogenies for a realistic biological study can be constructed in reasonable computational time, *Math. Biosci.* 60:133–142 (1982).
- 17 F. Harary, *Graph Theory*, Addison-Wesley, Reading, Massachusetts, 1969.
- 18 J. A. Hartigan, Minimum mutation fits to a given tree, *Biometrics* 29:53–65 (1973).
- 19 W. Hennig, *Phylogenetic Systematics*, Univ. of Illinois Press, Urbana, 1966.
- 20 R. J. Jensen, Wagner networks and Wagner trees: a presentation of methods for estimating most parsimonious solutions, *Taxon* 30:576–590 (1981).
- 21 A. G. Kluge and J. S. Farris, Quantitative phyletics and the evolution of anurans, *Syst. Zool.* 18:1–32 (1969).
- 22 J. G. Lundberg, Wagner networks and ancestors, *Syst. Zool.* 18:1–32 (1972).
- 23 J. McNeill, Report on the fifteenth annual Numerical Taxonomy Conference, *Syst. Zool.* 31:197–201 (1982).
- 24 W. P. Maddison, M. J. Donoghue, and D. R. Maddison, Outgroup analysis and parsimony, *Syst. Zool.* 33:83–103 (1984).
- 25 M. F. Mickevich, Transformation series analysis, *Syst. Zool.* 31:461–478 (1982).
- 26 E. M. Prager and A. C. Wilson, Construction of phylogenetic trees for proteins and nucleic acids: empirical evaluation of alternative matrix methods, *J. Mol. Evol.* 11:129–142 (1978).
- 27 E. Sober, Parsimony in systematics: Philosophical issues. *Ann. Rev. Syst. Ecol.* 14:335–357 (1983).
- 28 D. L. Swofford, On the utility of the distance Wagner procedure, in *Advances in Cladistics. Proceedings of the First Meeting of the Willi Hennig Society* (V. A. Funk and D. R. Brooks, Eds.), New York Botanical Garden, Bronx, New York, 1981, pp. 25–43.
- 29 D. A. Young, Are the angiosperms primitively vesselless? *Syst. Bot.* 6:313–330 (1981).