6

PARSIMONY AND PARSIMONY ANALYSIS

In the preceding chapter, we explored the concepts of characters and homology. In this chapter, we will use these concepts to demonstrate how phylogenetic problems can be analyzed using the principle of parsimony. This will be followed in the next chapter with a discussion of likelihood and Bayesian methods, which are usually described as statistical methods. We begin with a general consideration of parsimony, but will make mention of those aspects of parsimony analysis that are similar to likelihood, as the two are closely connected. We will then place this general discussion within the context of parsimony analysis.

PARSIMONY

The usual definitions of parsimony one encounters in English dictionaries concern money: extreme stinginess, extreme care in spending money, reluctance to spend money unnecessarily. Scientists and philosophers use a different version, usually attributed to William of Ockham (1285–1347) but in fact found in the works of Aristotle. The principle of parsimony is a methodological principle that posits simpler explanations of data relative to hypotheses are to be preferred over more complex explanations. This idea of simplicity relative to scientific hypotheses has been explored in some depth by Sober (1975), and the link between simplicity and parsimony has been discussed extensively in the phylogenetics literature (e.g., Wiley, 1975; Beatty and Fink, 1979; Farris, 1983). Both Farris (1983 and earlier works cited therein) and Sober (1983a) have linked simplicity with phylogenetic parsimony. Farris (1983), in particular, provides an extensive discussion of phylogenetic parsimony as a principle that leads to greater explanatory power of the resulting

Phylogenetics: Theory and Practice of Phylogenetic Systematics, Second Edition.
E. O. Wiley and Bruce S. Lieberman.

phylogenetic hypotheses and argued (correctly in our view) that the role of parsimony was to minimize the number of ad hoc explanations embedded in the preferred hypothesis in the form of hypotheses of homoplasy. This provides a direct link with Hennig's auxiliary principle. We leave the philosophical justification of phylogenetic parsimony to the end of this chapter. For now, we are concerned with how phylogenetic parsimony works.

Parsimony: Basic Principles

1. Among the many possible phylogenetic trees that graphically portray the descent of three or more taxa, only one of these trees is correct, given that the taxa are natural entities. This principle is shared with other approaches (e.g., statistical approaches) and is a simple statement that the problem is historical.
2. Characters originate and become fixed over evolutionary time such that it is possible for ancestral species to pass on such characters to descendant species. Again, the principle is shared with other approaches.
3. The sharing of character states is always evidence that those taxa that share a character state are related unless the weight of other evidence dictates that they are of independent origin. This principle is unique to parsimony approaches, although it is possible to assign different confidence in the ability of other characters to influence the decision by assigning relative weight to the character.
4. Once a character appears and is fixed, there is no reason to postulate that it will change unless the weight of other characters dictates that change is necessary because tree topology changes. This principle differs from statistical approaches where there is always a probability of change built into a model of character evolution for each class of characters.
5. Characters (data columns, transformation series) are treated as independent in any analysis (Kluge's auxiliary principle; Brooks and McLennan, 2002) for purposes of testing character hypotheses. This principle is shared by most approaches for computational reasons.
6. The result of parsimony analysis consists of placing character states on a tree where they are thought to have originated or become fixed. Parsimony analysis is particularly "transparent" in this principle, but we can place states on trees using statistical approaches if we wish to do so.
7. The tree with the fewest number of independent origins of shared characters is the preferred solution. This is the maximum parsimony principle. Parsimony differs from other approaches because trees are evaluated based on minimum length—the minimum number of changes in characters that are hypothesized to have occurred for any particular tree hypothesis. Trees of minimum length fulfill the principle.

Parsimony, then, is built around the proposition that the "best tree" is the tree that describes the evolution of any particular set of characters using the smallest number of evolutionary changes of the characters analyzed. The question is: how do we do this? Below we will review parsimony methods, beginning with classic Hennigian

argumentation and proceeding to current optimality-driven algorithms. The progression of the chapter reflects the historical development of the methods used today.

Kinds of Parsimony

In Chapter 5, we discussed the various relationships that might obtain between character states (ordered, unordered, etc.). Two common forms of parsimony are directly related to how we treat the relationships between character states within a transformation series. Characters with only two states are treated the same in both common forms of parsimony, Fitch and Wagner parsimony, but may be treated differently in the uncommon forms of parsimony, Dollo and Camin-Sokal parsimony. We will review each briefly.

Fitch parsimony (Fitch, 1971). Fitch parsimony takes all characters as unordered (see Fig. 5.16a, d). When three or more character states exist (Fig. 5.16d), a reversal from two to zero or a transformation from zero to two is counted as a single step. This type of parsimony is commonly implemented when analyzing DNA base-pair data and multistate morphological characters.

Wagner parsimony (Farris, 1970). Wagner parsimony treats binary character states identically to Fitch parsimony. However, characters with more than two states are considered ordered (Fig. 5.16e). Thus a transformation from two to zero is counted as two steps because the only route from two to zero is to pass through state one.

"General" parsimony (Swofford and Olsen, 1990). General parsimony allows mixing of different kinds of parsimony in a single analysis following a generalized Sankoff approach. For example, Fitch parsimony might be used for some characters, Wagner parsimony for others, and a step matrix for others. "Informed parsimony" (Goloboff, 1998) is a form of general parsimony.

Two uncommon forms of parsimony are also recognized. Camin-Sokal parsimony (Camin and Sokal, 1965) imposes the constraint that evolution is irreversible. Once state 1 has appeared, subsequent transformation from 1 to 0 is not allowed. However, state 1 can evolve as many times as needed. Dollo parsimony (as implemented by Farris, 1977) allows a single change from state 0 to state 1 and as many reversals from 1 to 0 as needed to explain the data. Both of these methods work on rooted trees, in contrast to Fitch and Wagner parsimony, which can work on either rooted or unrooted trees.

With these distinctions in mind, we will examine different analytical approaches to parsimony analysis. The first, classic Hennigian argumentation, does not speak directly to the different forms of parsimony outlined above because the classical approach was performed in the absence of the data matrix and a specific numerical algorithm. Nevertheless, it is directly connected to more modern methods by Hennig's auxiliary principle.

CLASSIC HENNIGIAN ARGUMENTATION

Classic Hennigian argumentation was practiced long before the advent of computer-assisted analysis. It is founded on the proposition that the investigator makes a priori decisions on relative synapomorphy and groups based on those decisions. As

such, it is in the class of algorithmic approaches. Although issues such as parsing out different kinds of parsimony were not part of the discussion prior to the advent of computer-assisted analyses, we can think of Hennig Argumentation as a form of nonexplicit general parsimony. It is based on three rules (see Brooks and McLennan, 2002).

1. *The Grouping Rule.* Characters deduced as synapomorphies are evidence of unique common ancestry while symplesiomorphies and homoplasies are of no use in determining unique common ancestry (Hennig, 1966). How one determines which of two homologous characters is apomorphic is the process of "polarizing the transformation series" and is discussed in the next section.
2. *The Inclusion/Exclusion Rule.* The information from two transformation series can be combined into a single hypothesis of relationship (unique common ancestry) if the valid evidence (i.e., the synapomorphies) implies the identical grouping or allows for the complete inclusion or exclusion of groups implied by the valid evidence. This rule is illustrated in Fig. 6.1.
3. *The Homoplasy Rule.* If the information from two transformation series results in groupings that overlap or conflict, then one and possibly both putative hypotheses of synapomorphy are false at the level used. Either one or both are not homologs, or one or both are incorrectly polarized.

It is rare to see a phylogenetic paper these days that employs classical, precomputer analysis. This rarity does not mean that classical analyses are an invalid approach, but it does signal that complexities of analysis are usually greater than the ability of an investigator to consider all of the possible phylogenetic hypotheses that might be inferred from the data. In spite of this, it is worthwhile to understand phylogenetics at a level where we can consider the meaning of such practices as character polarization and phylogenetic argumentation on a character-by-character basis. Whether one uses computer-assisted analyses or not, the quality of the initial

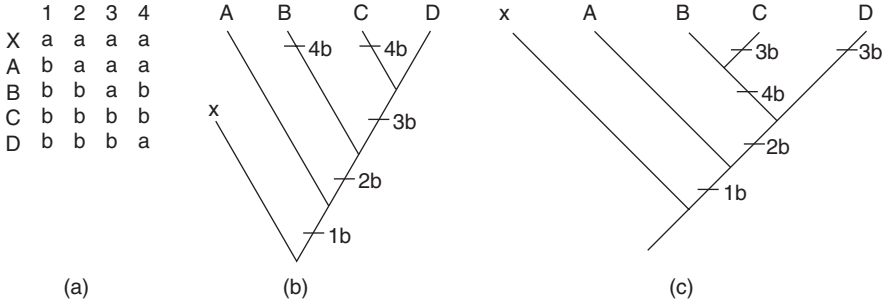


Figure 6.1. Simple examples of the inclusion/exclusion and homoplasy rules. (a) A data matrix with states coded "b" polarized as apomorphic based on the presence of states coded "a" in the outgroup X. (b) The argument that character states 2b and 3b confirm the monophyletic groups BCD and CD but excludes BC as a monophyletic group confirmed by 4b (4b in B and C is homoplastic). (c) If we accept the argument that 4b confirms the monophyletic group BC, then we must exclude 3b as a synapomorphy confirming CD and consider 3b homoplastic. Note that in both (b) and (c) the state 2b confirms the monophyletic group BCD and thus can include either hypothesis.

characters brought to the analysis is critical to the final result and a priori methods depend on such quality.

Polarization

Two character states are said to be polarized when we have determined which evolved first and which one came after. Thus, polarization refers to determining which one of two or more hypothesized states is plesiomorphic and which one (or ones) is apomorphic. The initial assumption is that all instances of the states are actually homologous, but we may find, using congruence, that they may not be homologous when we accept the homologies of other characters and their states. In fact, we may find nonhomology among instances of a state; parallel or convergent appearance of characters that share a nonhomologous identity. Thus, we are not committed to claiming that the characters and their states will turn out to be homologous, only that we will assume so for purposes of testing that very proposition.

In the example presented in Fig. 6.1, we assumed that we knew the homology and polarity of the states in advance, and then applied our three rules. In this section, we will explore ways of polarizing character states. This activity lies at the heart of the phylogenetic method, whether it is done by hand, a priori, or by rooting, a posteriori. Phylogeneticists rarely polarize characters a priori these days. Instead, an investigator relies on various computer algorithms to polarize characters by using one or more outgroups designated by the investigator to perform this task. However, understanding the reasoning behind polarization is important relative to the history of the discipline and also because it shows why computer-assisted analysis can arrive at a robust hypothesis only when given the best outgroup information possible. Further, if one is going to order more than two character states, one is performing a priori polarization, and thus the principles are vital. Many criteria for polarizing character states have been proposed, including several by Hennig (1966) himself. There is a general consensus (with some significant dissenters) that there is only one general criterion, outgroup comparison. We will discuss this criterion and then some of the alternatives.

Polarization by Outgroup Comparison. Consider a character with states distributed such that there is variation in the group under analysis. For example, among land plants, mosses and tracheophytes have xylem tissue while hornworts have undifferentiated parenchyma cells. Which is the apomorphic character? The closest relatives of hornworts, mosses, and tracheophytes are the liverworts. Liverworts have undifferentiated parenchyma cells. If certain assumptions, detailed below, are met, this observation leads to the conclusion that xylem is apomorphic relative to undifferentiated parenchyma cells. This is reinforced by the observation that the closest relatives of land plants, groups such as stoneworts (e.g., *Chara*), also have undifferentiated parenchyma cells rather than xylem. If we examine a tree of land plant evolution, we see that the most likely hypothesis of polarity is that xylem evolved sometime after the origin of hornworts but before the origin of mosses (Fig. 6.2).

This kind of reasoning is deductive, and the validity of the conclusions depends on certain assumptions. First, one must accept the monophyly of the group comprising hornworts, mosses, and tracheophytes, establishing the ingroup, which is the group one wishes to analyze. Second, one must accept the monophyly of all land plants (liverworts and above) to establish a rational sister group. Third, one must

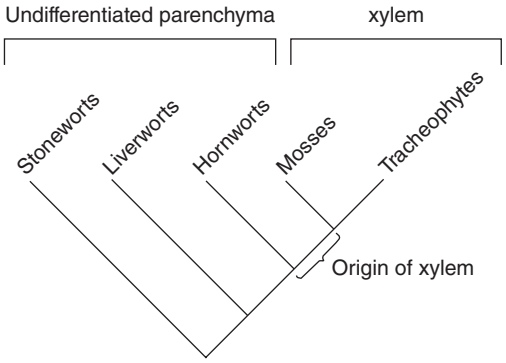


Figure 6.2. A hypothesis of plant relationships. Given this topology, the transformation of undifferentiated parenchyma cells to form xylem tissue is more parsimonious than the transformation of xylem to undifferentiated parenchyma because the close relatives of mosses and tracheophytes have undifferentiated parenchyma cells.

accept the hypothesis that stoneworts (and brittleworts and other filamentous green algae) are related to land plants. In other words, such deductive reasoning is accomplished with the acceptance of prior information. Because it depends on prior information, the conclusions will be valid only if the prior information is correct. With this in mind, the outgroup rule of polarization may be simply stated.

The Outgroup Rule. Given two (or more) homologous character states within a group studied, the state found outside this group in close relatives is the plesiomorphic state and the character found only within the group is the apomorphic state.

An explicit statement of the outgroup rule is, curiously, missing from Hennig (1966). However, and at least for binary characters, it is apparent to us that Hennig used outgroup comparison, as evidenced by the following quotes.

> Recognition that species or species groups with common apomorphous characters form a monophyletic group rests on the assumption that these characters were taken over from a stem species that only they have in common, and which already possessed these characters prior to the first cleavage (Hennig, 1966:90).

> [I]f it is a question of determining the relationships between different species groups, then it is of primary importance to show that each group has apomorphous characters, characters that are present only in it (Hennig, 1966:90).

Both of these statements imply Hennig used a comparative outgroup method. One could hardly reach the conclusion that a character was only found in the stem species of a group without examining species outside of the group. And no one could claim that a character is unique to a group without looking at other groups. Hennig (1966:95–116) discusses "accessory criteria" when considering "morphoclines," characters of more than two states. The fact that he characterizes them as "accessory" relative to the "scheme of argumentation of phylogenetic systematics" suggests that what Hennig considered strong evidence of monophyly were characters unique to a given group, which can only be deduced if one looks outside the group. This emphasis on sister groups sharing unique homologies is common in early

phylogenetic literature (e.g., Brundin, 1966). Uniqueness can only be accessed by looking broadly across groups. The criterion of what we now know as outgroup comparison was also well understood by early quantitative phylogeneticists, forming one of three criteria used by Kluge and Farris (1969).

The logic of implementing the outgroup rule was discussed by Wiley (1975) within a Popperian framework, but Wiley did not characterize looking outside the group with the formal designation of "outgroup" for those taxa consulted, but characterized the addition of new taxa (those we now think of as outgroups) as raising the level of universality of the problem (which is exactly what outgroups do). The actual origin of the term *outgroup* is not of particular consequence, because it is the principle, not the name, that is important. It is possible that the term originated in print with Wiley (1976:11):

> Hennig's (1966) method differs fundamentally from a purely phenetic method in that all the shared characters are not used to refute a given relationship; only synapomorphous characters are used. Such testing can only be accomplished in an open system, that is, by considering taxa outside the three (or more) taxon system. Such considerations may be termed *outgroup* [emphasis added] comparisons. The one condition placed on this procedure is that the three (or more) taxa must form a monophyletic group. The designation of outgroups for comparison permits an investigator to sort out which of the observed characters are unique to the three-taxon system and which characters have a more general distribution. The *outgroup* [emphasis added] comparison automatically raises the level of universality of the phylogenetic hypothesis to a new level. And, it allows the investigator to put his three-taxon statement in context with a hypothesis of a higher level of universality.

By the early 1980s, specific descriptions of character argumentation using the term *outgroup* were appearing (c.f., Eldredge and Cracraft, 1980) and specific forms of the Outgroup Rule were published (Wiley, 1981a; Watrous and Wheeler, 1981). However, the complexities of polarization using outgroups were best demonstrated by Maddison, Donoghue, and Maddison (1984). They demonstrated that the simple rules formulated in earlier works were not adequate. Their work also demonstrated that criteria such as "common is primitive" could be dismissed as fallacies. Maddison et al. (1984) begin by defining terms, illustrated in Fig. 6.3.

Ingroup. The group under analysis. In Fig. 6.3a, the ingroup is shown as a polytomy, suggesting unresolved relationships.

Ingroup Node. The trees used by Maddison et al. (1984) are node-based trees (vertexes are taxa), not stem-based trees (edges are taxa), so the internal nodes are ancestral species and the edges are relationship statements. The ingroup node represents the character states of the ingroup ancestor, that is, the ancestor of the group under analysis. Some of these character states will be synapomorphies for the ingroup, others will be plesiomorphies, and some cannot be polarized. Polarization depends on the distribution of the character state and its homolog(s) among the outgroups.

Outgroup and Sister Group. Any clade that is attached to the edge leading to the ingroup node is an outgroup. That clade immediately below the ingroup node is the relative sister group of that particular analysis.

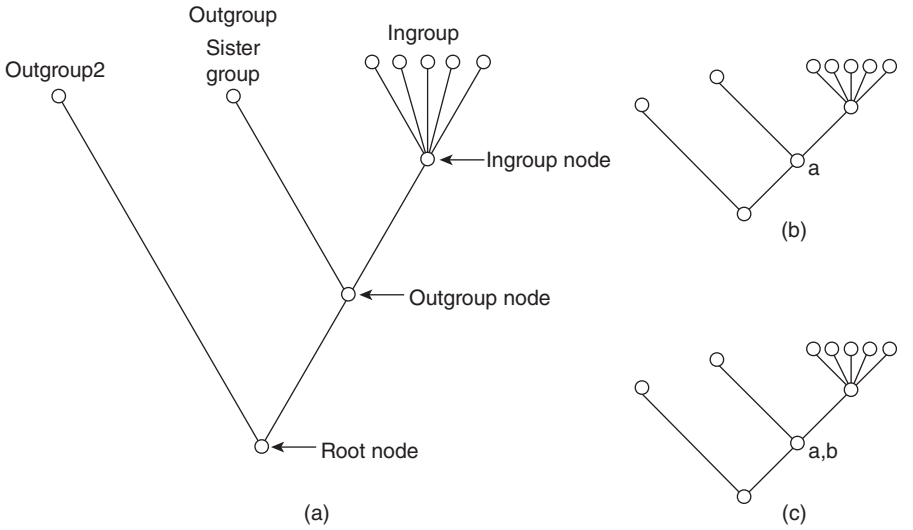
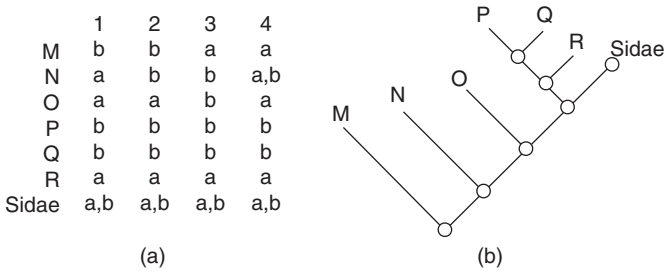


Figure 6.3. Basic terminology of parts of a Hennig tree following Maddison et al. (1984). (a) The ingroup node represents the ancestral species of all members of a group under analysis, the ingroup. The outgroup node is the node that refers to the ancestor of the ingroup node while the root node represents the most basal, or ancient, ancestral species. The sister group constitutes the closest outgroup to the group analyzed that is known or included in the analysis and constitutes the first outgroup (labeled outgroup 1). It is composed of one to many species, and its character states represent the character states inferred for the ancestral species of all members of the sister group. The outgroup 2 is simply the next known closest relative. In general, a minimum of two outgroups are needed to polarize a character *a priori* unless the sister group is considered entirely plesiomorphic (a bad assumption to make). (b) A decisive character decision in favor of state "a" of a character. (c) An equivocal decision for a two-state character. Note that these decisions are made at the outgroup node, not the ingroup node.

Outgroup Node. The node immediately below the ingroup node is the outgroup node. A character assigned to the outgroup node would be the character hypothesized to be present in the ancestor of the ingroup and its sister group.

Root Node. The most basal node in the tree.

Maddison et al. (1984) frame the polarity problem as a quest for the assignment of characters to the outgroup node. Why this is so is immediately apparent if we give it a bit of thought. We wish to arrive at two classes of hypotheses in our analysis. First, we seek evidence that the ingroup is monophyletic. Second, we wish to uncover evidence for monophyly of subgroups within the ingroup. Evidence that the ingroup is monophyletic can only be gained by accessing the character states present in the immediate common ancestor of the ingroup and its sister group. By determining the character states present at the outgroup node, we are able to either make this decision or know that the information is not adequate to make this decision, as we shall see below. Decisions at the outgroup node can be of two kinds. Given two homologous character states, if only one is assigned to the outgroup node, the polarity



	1	2	3	4
M	b	b	a	a
N	a	b	b	a,b
O	a	a	b	a
P	b	b	b	b
Q	b	b	b	b
R	a	a	a	a
Sidae	a,b	a,b	a,b	a,b

(a)

(b)

Figure 6.4. The Maddison et al. (1984) method of character polarity I. (a) The character matrix. (b) The Hennig tree of relationships of the outgroups to the ingroup. Note that this tree topology is given *a priori*; confirming characters for this topology may be totally missing in the character matrix. Also note that the ingroup is polymorphic for each character, a necessary condition for analyzing relationships among the ingroup. Redrawn from Wiley et al. (1991), used with permission, Biodiversity Institute, University of Kansas.

decision is **decisive** (Fig. 6.3b). If neither can be confidently assigned to the outgroup node, the decision is **equivocal** (Fig. 6.3c).

A simplified example of how the algorithm works is shown in Figs. 6.4 and 6.5 (taken from Wiley et al., 1991). We will use the binary transformation series, although Maddison et al. (1984) provide a general (and more complicated) algorithm for more than two character states.

Consider the Sidae, its outgroups, and character variation (Fig. 6.4a). Prior knowledge from other studies hypothesized a specific outgroup structure (Fig. 6.4b). This tree is not justified by the characters in the matrix because the taxon of interest is Sidae and the analysis of Sidae to its sister group and other outgroups is not a matter for testing (this may or may not always be a wise choice). Proceed in the following manner.

1. Proceeding from the most distant branches, label nodes on the tree according to the following rules. If all terminal taxa have the state "a," then label the node decisive "a." If all terminal taxa have the state "b," then label the node decisive "b." If one or more terminal taxon has "a" and one or more terminal taxon has "b," then label the node equivocal "a, b."
2. Proceeding, again, toward the outgroup node, label the next node with the majority character derived from the state that lead to that node. For example, if a node has the equivocal decision "a, b" and a terminal has "a," then the assignment of the next node is decisive "a." If both have "a, b", then assign "a, b."
3. Proceeding from all parts of the tree to the outgroup node, make decisions for each node until the outgroup node is reached.

These calculations are carried out for the first character in Fig. 6.5a and for the second character in Fig. 6.5b. If you perform these operations on a sufficient number of trees, you will notice that the sister group is the most influential group in the entire analysis, unless it is polymorphic. If the sister group has a single character, or if the decision at the most basal node within the sister group is decisive, then this character will appear at the outgroup node (Maddison et al., 1984). Maddison et al. (1984)

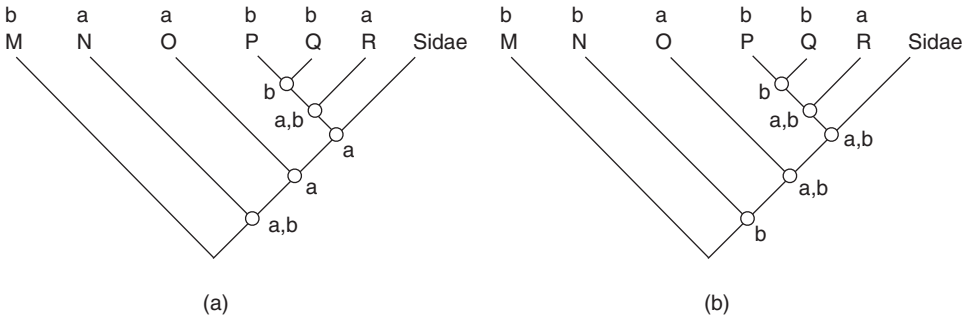


Figure 6.5. The Maddison et al. (1984) method of character polarity II. (a) An example of a decisive decision for character 1 of Fig. 6.4a. (b). An equivocal decision for the states of character 2 of Fig. 6.4a. Redrawn from Wiley et al. (1991), used with permission, Biodiversity Institute, University of Kansas.

noted two other general outcomes. First, if the sister group and the next sequential outgroup have the same state, then that state will always be decisive at the outgroup node (the *First Doublet Rule*). Second, if states alternate down the outgroup topology and if the sister group has the same state as the most basal outgroup, then that state will be decisive, but if the most basal group has a different character, then the decision will always be equivocal (the *Alternating Outgroup Rule*).

You can use such reasoning in a traditional phylogenetic analysis by preparing a matrix of ingroup and outgroup taxa, adopting an outgroup tree topology, reasoning through each polarity decision and following the grouping, inclusion/exclusion, and homoplasy rules. As we shall see in later sections, computer-assisted phylogenetic analysis does not make a priori polarity decisions. So, why in the modern age do we cover this topic? Although computer-assisted analysis does not make a priori character decisions, phylogenetic computer-assisted studies call for a priori designation of the outgroup(s) to be used to polarize the states once direction of transformation is specified through designation of an outgroup (always included in the analysis). Thus, the Maddison et al. (1984) paper is very applicable to general phylogenetic reasoning using more modern techniques of phylogenetic analysis (parsimony and statistical algorithms) for three reasons:

1. It demonstrates the importance of careful attention to identifying or discovering the sister group.
2. It calls attention to the fact that a single sister group is not sufficient to unambiguously polarize a character; the minimum for analysis is the sister group and one additional relevant outgroup, hopefully the next sister group down the tree.
3. It destroys the notion that if you do not know the sister group you can simply make a decision that the character state commonly found in some array of possible outgroups is the plesiomorphic character. This third point may call for the imposition of a particular outgroup topology prior to analysis or to the inclusion of characters that are not particularly relevant to the ingroup problem per se, but that give structure of the relationships of the outgroups to each other and to the ingroup and are possible synapomorphies of the ingroup

itself (as suggested by both J. S. Farris and D. L. Swofford to Maddison et al., 1984:99).

What if you don't know the relationships of the outgroups to the ingroups? Indeed, what if you have no a priori evidence that the ingroup is even monophyletic? Such a case calls for the solving of a larger problem and may call for a community global approach. For example, the ichthyological community has been working on the teleost tree of life in a phylogenetic framework for some forty years. Although many studies of smaller clades have been successfully pursued, the emphasis has been on working from the root of the teleost tree toward the tips. This approach creates outgroup structure *with* the flow of evolutionary time rather than *against* the flow (e.g., Greenwood at al., 1973; Stiassny et al., 1996, 2004).

The analysis of *Leysera* presented below illustrates a relatively simple application of classic Hennigian argumentation, but with considerable attention paid to the problem of identifying a suitable outgroup. This will be followed by an account of more current approaches of analyses using computer algorithms where character polarity is determined *a posteriori* using an optimality criterion.

Example 1. The Phylogenetic Relationships of *Leysera*

Leysera is a small group of four species of composite shrublets. Three species (*L. gnaphalodes*, *L. tenella*, and *L. longipes*) are found in southern Africa. One species (*L. leyseroides*) is found in the Mediterranean region. As a continuation of his study on other closely related genera, Bremer (1978a) analyzed this group.

Background Information. *Lysera* (Fig. 6.6) is a member of Compositae, tribe Inuleae. Merxmüller et al. (1977) placed *Leysera* into the *Athrixia* genus group (eight genera) within the subtribe Athrixiinae (23 genera total).

Bremer (1978a) supported the monophyly of four of the eight genera of the *Anthrixia* group on the basis of leaf and involucre characters: all have ventrally furrowed and pubescent leaves and wide, yellowish brow, and scarious involucral bracts. These characters are "a most uncommon feature" in Athrixiinae, uniting *Leysera*, *Antithrixia*, *Relhania*, and *Rosenia*. Of the four, *Antithrixia* has a pappus with many barbellate bristles compared to the three other genera, which have a reduced number of bristles as well as a complete loss of bristles on the ray-floret pappus (Fig. 6.7). Finally, Bremer (1978a) observed that only species of *Leysera* have a solitary capitula on a long peduncle (Fig. 6.6) whereas the other three genera have sessile capitulas with the exception of some species of *Relhania* (which Bremer interpreted as homoplasy based on the monophyly of *Relhania*). At this point, Bremer has established the following background information (Fig. 6.8):

1. *Antithrixia*, *Leysera*, *Relhania*, and *Rosenia* comprise a monophyletic group. Justification is via outgroup comparison and character rarity.
2. *Antithrixia* is the sister genus to the remaining genera, which form a monophyletic group (outgroup comparison).
3. *Leysera* is monophyletic (outgroup comparison).

Bremer has assumed that the *Anthrixia* genus group is monophyletic. He has also assumed that rarity in morphological characters among outgroups is evidence of

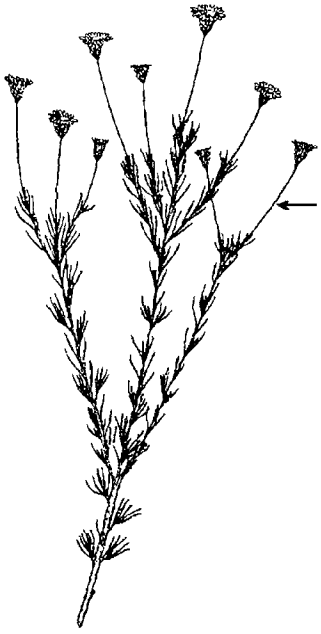


Figure 6.6. The composite plant *Leysera gnaphalodes*, illustrating the long peduncle typical of the genus (arrow). From Bremer, 1978a. Used with permission of Botaniska Notiser.

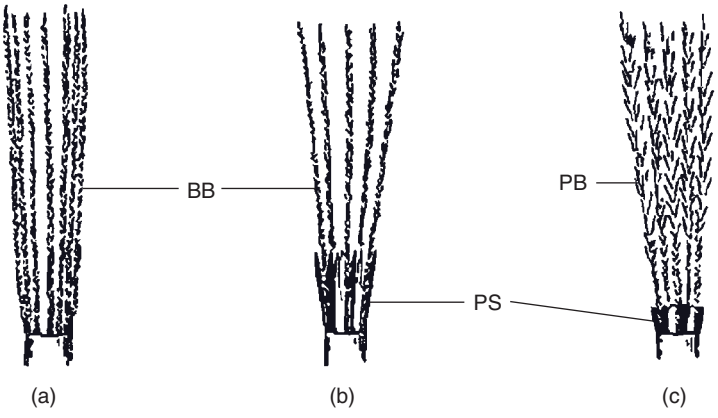


Figure 6.7. Features of the disc-floret in (a) *Antithrixia*, (b) *Leysera longipes*, and (c) *Leysera tenella*. Abbreviations: BB, barbellate bristles; PB, plumose bristles; PS, scales. Transformation of states in both characters proceed left to right. From original drawings by Kåre Bremer included in Bremer, 1978a. Used with permission of the author and Botaniska Notiser.

homoplasy for certain characters. (Note: this is pre-Maddison et al., 1984, and such assumptions were common.)

Bremer's (1978a) analysis involves two possible sister groups (*Relhania* and *Rosenia*). Further, there is a problematic taxon, "*Leysera*" *montana*. This species has solitary capitulae on long peduncles, like *Leysera*, but has a pappus with many

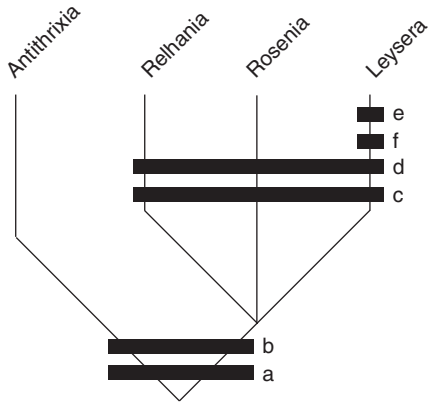


Figure 6.8. The phylogenetic relationships of *Leysera* and closely related genera. Synapomorphies are (a) leaves ventrally furrowed and pubescent; (b) involucral bracts wide, yellowish brown, and scarious; (c) floret pappus with scales but no bristles; (d) disc-floret pappus with reduced bristles and no scales; (e) solitary capula on long peduncle; and (f) chromosomes 2N = 8. Adapted from Bremer, 1978a.

barbellate bristles, like *Antithrixia* and other Athrixiinae outside the clade. Bremer removed this species from *Leysera* and later described the monotypic *Oreoleysera* for it (Bremer, 1978b) because *O. montana* did not have the synapomorphies that would place it in the monophyletic clade containing the three genera, even though it had the character that unites *Leysera*, forcing Bremer to conclude that the match was a homoplasy.

Given that there are four species of *Leysera*, a total of 15 rooted bifurcated trees are possible. Bremer (1978a) analyzed 13 characters for the four species using what is now referred to as outgroup comparison relative to the two possible sister groups, *Relhania* and *Rosenia*. Among the states were five autapomorphies, four in *L. longipes* and one in *L. leyseroides*, which will not be discussed further. In addition to the analysis presented above that demonstrates the monophyly of *Leysera* and its relationships to its relatives, two additional levels of synapomorphy analysis are required to complete the analysis.

1. Establishing the basal member of the species group.
2. Breaking up the remaining trichotomy.

One character from each level will be discussed. The complete table of characters is shown in Table 6.1, and the tree of relationships is shown in Fig. 6.9.

Level 1. Character 1; Receptacle smooth versus rough. Within *Leysera* there are two character properties of the receptacle. In *L. longipes* the receptacle is more or less smooth, without scale-like growths. In the remaining three species the receptacle is rough, and this roughness is caused by scale-like outgrowths. Receptacles with scalelike growths are not known in species of *Relhania* or *Rosenia*, nor found in any other members of the *Athrixia* genus group. They are known from less closely related genera of composites. To argue that the rough receptacle of the three species

TABLE 6.1. Characters used by Bremer (1978a) to analyze the phylogenetic relationships of *Leysera*. Autapomorphies are not listed. Characters are shown in the hypothesis presented in Fig. 6.9. All determinations were made by outgroup comparison.

Character	Plesiomorphic	Apomorphic
1. Receptacle	Smooth	With scalelike growths
2. Floret tubules	Glands present	Hairs present
3. Pappus	Barbellate	Plumose
4. Achenes surface	Smooth	Cells imbricated
5. Pappus scales	Subulate	Wide and flat
6. life cycle	Perennial	Annual

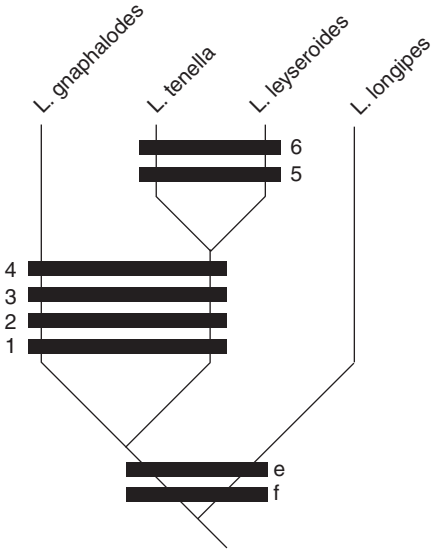


Figure 6.9. Bremer's (1978a) hypothesis of the relationships among species of *Leysera*. Character numbers correspond to the apomorphic state of transformations in Table 6.1. The two synapomorphies of the genus correspond to the states in Fig. 6.8. Autapomorphies for each species are not shown.

was convergent because a similar condition was found in distant relatives is a violation of the auxiliary principle. To argue that these three species are not members of the *Athrixia* genus group would require rejection of all synapomorphies that place *Leysera* within the group and united to *Relhania* and *Rosenia*. Bremer (1978a) concluded that the rough receptacle was a synapomorphy uniting *L. leyseroides*, *L. tenella*, and *L. gnaphalodes* by outgroup comparison and parsimony.

Level 2. Character 5; pappus scales (Fig. 6.7b, c) subulate versus wide and flat. Wide and flat pappus scales are found in *L. tenella* and *L. leyseroides* while subulate scales are found in *L. longipes* and *L. gnaphalodes*. Given the four synapomorphies that unite *L. gnaphalodes* with *L. tenella* and *L. leyseroides*, and given the monophyly of *Leysera*, we can use *L. longipes* as a "functional outgroup" to polarize the

transformation series. The use of functional outgroups is discussed by Watrous and Wheeler (1981). In short, once the monophyly of a group is established, an investigator can create "functional outgroup comparisons" by comparing the basal member(s) of the group with more apical members of the group (see also Wiley, 1981a:175–176).

A POSTERIORI CHARACTER ARGUMENTATION

There is another way to argue characters, and it is a basic aspect of more modern phylogenetic analyses. If you consider that all characters are freely reversible, and that they are not fated to ratchet ever forward, then it turns out that we can assemble a tree without a root and without polarization whose topology is logically consistent with a rooted topology determined by a priori character argumentation. The characters on a rootless tree have no phylogenetic interpretation, but we can give them such an interpretation if we specify the starting point, an activity termed *rooting the tree*. This is useful, because it provides a bridge between computer-assisted phylogenetic analysis and traditional phylogenetic analysis. Examine the character matrix and unrooted tree in Fig. 6.10a, b. Note that we have made no judgments of character polarity, we have simply plotted the characters coded "b" along branches where they occur. If we root the tree along the edge leading to E (Fig. 6.10c), then all of the characters coded "b" appear on the rooted tree as apomorphies. But if we root on the edge leading to D (Fig. 6.10d), then most of the apomorphies are those characters coded "a." If you count the total number of possible changes, you will note that both trees are the same length, TL = 7 steps. Rooting changes both the topology and the character polarity interpretations, but it does not change the tree length.

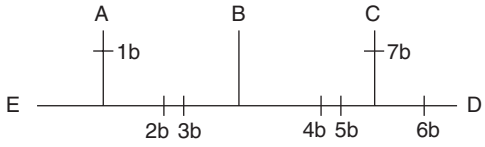
ALGORITHMIC VERSUS OPTIMALITY APPROACHES

Swofford and Olsen (1990) and Swofford et al. (1996) discuss a useful distinction between two approaches to phylogenetic inference. Algorithmic approaches combine tree inference and definitions of the preferred tree into a single operation that defines a sequence of steps that lead to the determination of a tree. Evolutionary assumptions are embedded in the algorithm and used for the analysis. In contrast, optimality approaches define an objective function such as minimum tree length (parsimony) or maximum likelihood (ML), use an algorithm to generate trees, and then sort trees with a preference for that tree that meets the objective function. In this approach, the objective function embodies the evolutionary assumptions and any algorithm that generates trees can be used because the result is not dependent on evolutionary assumptions embodied in the algorithm, but rather, the result is dependent on whether one tree (or set of trees) meets the objective criterion better than another tree (or set of trees).

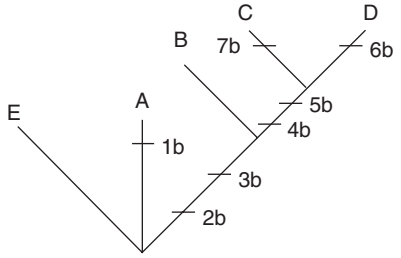
In algorithmic approaches, the algorithm is important because it defines the selection criterion and combines the inference and the criterion for the preferred tree into a single operation. Examples include classic Hennigian argumentation (Hennig,

	1	2	3	4	5	6	7
A	b	a	a	a	a	a	a
B	a	b	b	a	a	a	a
C	a	b	b	b	b	a	b
D	a	b	b	b	b	b	a
E	a	a	a	a	a	a	a

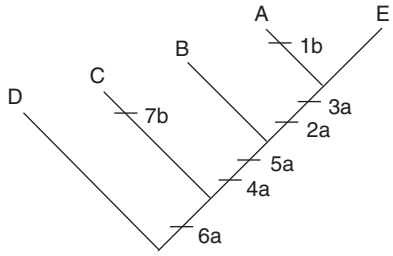
(a)



(b)



(c)



(d)

Figure 6.10. Rooting an unrooted tree. (a) A hypothetical data matrix. (b) The unrooted tree that minimizes the number of transformations needed to account for the changes shown in the matrix. (c–d) Two rooting decisions. Note that while the character polarities are much different, the lengths of the trees are the same (7 steps).

1966), Wagner ground plan divergence analysis (Wagner, 1961), most first-generation computer algorithms such as the Wagner Algorithm (Kluge and Farris, 1969), and some distance algorithms in current use such as neighbor joining (Saitou and Nei, 1987; Studier and Keppler, 1988). Algorithmic approaches are fast in terms of computation time and are likely to find trees that are close to optimal (or even optimal if the data are fairly clean). Their speed and efficiency make them excellent for building a tree hypothesis, but they can become stuck in local optima depending on the starting conditions and the nature of the data.

In optimality approaches, the investigator specifies an objective function and then uses an algorithm to compute that function for a particular tree topology. It then computes that same function for another tree and compares the trees. In parsimony the shortest tree "wins." How one obtains trees to compare is not relevant to the process, but simply to the efficiency of the search. Examples of computer packages that implement the optimality approach include all modern parsimony programs; several generations of Wagner programs (i.e., Hennig86, Farris, 1989a; PAUP, Swofford, 2001; PHYLIP, Felsenstein, 2007; NONA, Goloboff, 1999a; TNT, Goloboff et al., 2000).

Although fast, algorithmic-driven programs suffer from a problem. They compute a tree well enough, but the investigator might miss other trees that are just as good or very close to the tree computed. And the investigator has no ready way to compare the robustness of the results relative to other possible outcomes. In contrast, optimality-driven programs are slower, but can search many trees and return the results for all of the trees that fit the objective function and even those trees that might not meet the objective but are some specified distance from it. For example, if the objective function was maximum parsimony in terms of steps and there were 5 trees of length 100 steps, an exhaustive optimality search would return all 5 trees. Further, if the investigator wishes to also examine trees within 10 steps of the shortest trees, then all trees from 90–100 steps would be returned.

OPTIMALITY-DRIVEN PARSIMONY

In most current computer packages, one might begin the analysis by constructing a tree using, for example, the Wagner Algorithm or neighbor joining. However, most of the actual computing time is spent evaluating different tree topologies (branching patterns) to recover the tree(s) that meet a criterion of optimality given the data. How the tree is actually generated may be irrelevant. For example, you can evaluate all of the possible trees for a three-taxon problem by simply mapping the character distributions on the four possible trees in the most efficient manner (i.e., maximizing the number of synapomorphies and minimizing the number of homoplasies needed given the tree). You don't have to build a tree; all of the possible trees are given. Under the criterion that the shortest tree is the optimal tree (the objective function in parsimony analysis), all you have to do is count the changes and pick the shortest tree(s) among the four possibilities. Polarity is not determined a priori, but a posteriori through the designation of one or more outgroups. This is because the algorithm first computes an unrooted tree and then roots the tree at the point designated by the investigator. Although this may sound strange to classical phylogeneticists, the trick is to understand that solutions involving freely reversible characters yield a network that is logically consistent with a rooted tree that could be found using a priori character polarization in reference to the same outgroup. In essence, this is why parsimony analysis using computer algorithms is the same research program as parsimony analysis using classical Hennigian argumentation. The difference is this: as more and more taxa are analyzed and as homoplasy levels increase, the less the chance that classical Hennigian argumentation will yield all of the equally parsimonious solutions, or even the single most parsimonious solution. The order of taxa added to the analysis might lead the investigator into a local optimum. Some possible solutions for dealing with suspected homoplasy might be missed. There can also be problems for optimality approaches using computers, but they cover more ground in the hunt.

In parsimony analysis, the optimality criterion is tree length. The tree topology that minimizes the number of evolutionary steps needed to explain the evolution of characters in the matrix is the optimal tree, given the data. Other sorts of optimality criteria are possible. For example, ML also has an optimality criterion: the tree topology and evolutionary model applied that maximizes the probability of observing the data is preferred (see Chapter 7). Any particular algorithm is a method for

estimating the optimal tree given a particular criterion. As Swofford et al. (1996) stress, algorithms change and improve while optimality criteria may not. Thus, modern parsimony methods concentrate on:

1. Fitting characters on particular tree topologies such that the number of evolutionary steps is minimized within each transformation series.
2. Comparing the results obtained among tree topologies to determine which tree (or set of trees) is the shortest.
3. Visiting many possible trees in an effort to avoid locally optimal solutions.

DETERMINING TREE LENGTH

Kluge and Farris presented the algorithm for tree length with ordered characters. Fitch (1971) presented the algorithm for determining tree length in the case of unordered characters. Sankoff (1975) generalized the algorithm for general parsimony. There are two basic algorithms for determining tree length in the absence of a step-matrix or other weighting schemes. Each follows one of the two common parsimony approaches: Wagner parsimony (ordered characters) and Fitch parsimony (unordered characters). Each requires a single down-pass through the tree. Note that tree length is computed on unrooted trees. Also, although we begin with taxon A in our example, tree length can be computed from any starting taxon, as discussed above.

Tree Length under Ordered (Wagner) Parsimony. We will present the example used by Swofford et al. (1996) informally; that is, avoiding as much set theory and formal algorithms as possible to show the general method for a single transformation series. Consider an unrooted tree with five taxa, and a single ordered transformation series (Fig. 6.11a).

1. Root the tree with a terminal node. For each terminal node, assign the characters it has based on the input matrix (Fig. 6.11b). This is the taxon's character set.
2. Proceed from the tips toward the terminal node, and assign characters to each interior node (labeled X, Y, and Z in Fig. 6.11c) according to two rules.
   2a. If the intersections of the character sets of descendants is not empty, then let the character set of the ancestor equal the intersection as a closed interval.
   2b. If the intersection of the state sets is empty, let the character set of the ancestor equal the smallest closed interval containing an element from each set. Increase tree length by the length of the interval (the difference between the end points of the interval).
3. If the internal node is adjacent to the root node of the tree (immediate descendant of the root node), then go to step four, otherwise return to step two.
4. If the character of the root node is not contained in the character state of its descendant node, then increase tree length by the shortest distance between them.

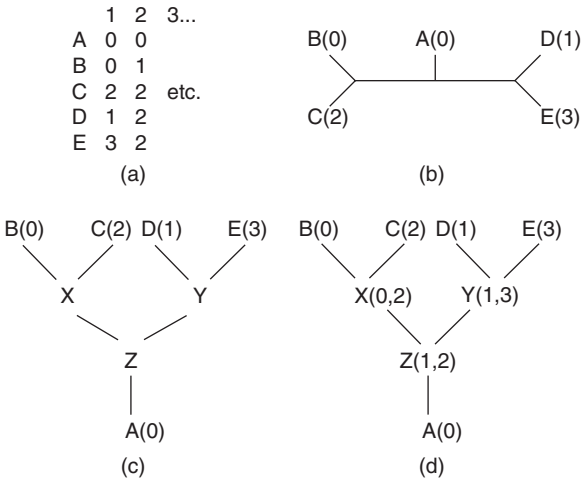


Figure 6.11. Calculating tree length for a single character. (a) A polarized transformation series with "0" as the plesiomorphic state. (b) An unrooted tree showing the distribution of transformations of character 1. (c) The tree rooted with taxon A. (d) Assignment of states to the interior nodes.

For the interior nodes in Fig. 6.11d, the values shown are computed below. Note that Z is the node at the basal fork and A is the root node.

1. X: $[0] \cap [2] = \emptyset$. Thus X = [0, 2]. English translation: the intersection, $\cap$, of "0" and "2" is empty, thus the state set of X is the interval [0, 2]. This follows Rule 2b. Increase the length by two steps.
2. Y: $[1] \cap [3] = \emptyset$. Thus Y = [1, 3]. This follows Rule 2b. Increase the length by two steps. Tree length now equals 4.
3. Z: $[0, 2] \cap [1, 3] \neq \emptyset$. Thus Z = [1, 2]. This follows Rule 2a.
4. The state set of A = [0], while that of Z = [1, 2]. Thus we increase tree length by one step.

Tree length under Wagner parsimony is TL = five steps. We would then perform the same operations on the next column of data and add the results to our count, adding columns until we reach the end of the matrix.

Tree Length under Unordered (Fitch) Parsimony. If you wish to calculate tree length under Fitch parsimony, we modify the algorithm slightly:

2a. If the intersections of the state sets of descendants is not empty, then let the state set of the ancestor equal the union of the intersection.

2b. If the intersection of the state sets is empty, let the state set of the ancestor equal the union of the state sets and increase tree length by one step.

4. If the state set of the root node is not contained in the state assigned to the basal fork of the tree, then increased length by one step.

In our example, X is assigned the character set [0, 2] and the tree length is increased by one step. Y is assigned the character set [1, 3], and the tree length

is increased by one step. Finally, Z is assigned the intersection of [0, 2] and [1, 3], which is [1, 2] and because this does not intersect with A (with character 0), the tree length is increased by one step. Thus, tree length under Fitch parsimony is TL = three steps.

Tree length may be further modified if weight is given to an entire transformation series or if a step-matrix is used for one or more transitions within a transformation series (although the use of step matrices requires a general parsimony procedure because algorithms for both ordered and unordered characters do not apply). As a simple example, if we assigned a weight of 100 to the first data column in our example, then the length would be 500 under Wagner parsimony and 300 under Fitch parsimony.

FINDING TREES

Fitting characters to a tree is a relatively easy procedure. Equip yourself with a program, and input a tree for any particular matrix of characters. Then have the program optimize the characters on the specified tree using one of the optimization criteria discussed later in this chapter. It takes almost no time to accomplish this task. The harder trick is to find the optimal tree (in parsimony, the shortest tree). There are several strategies to do this, depending on the number of taxa and the size and complexity of the data matrix.

Strategy 1: Exhaustive Search. For fewer than 12 taxa, one can simply optimize the characters on all possible tree topologies and pick the shortest tree(s). This strategy is preferred given a small number of taxa. It guarantees that the shortest tree(s) will be found.

Strategy 2: Branch-and-Bound. For up to about 20–22 taxa, one can employ a branch-and-bound algorithm that will be guaranteed to find the shortest tree(s). Above 20 taxa, the algorithm, as implemented on most computer platforms, is too slow.

Strategy 3: Heuristic Search. Above around 20–22 taxa, or in situations where the data matrix is "messy" (i.e., contains a high level of homoplasy), the number of possible tree topologies becomes so great that exact solutions are no longer possible. In such cases, heuristic search routines must be implemented.

Heuristic searches are common to all mathematical problems for which an exact solution is unobtainable or impractical. We will meet "heuristic searches" in parsimony analysis, ML analysis and Bayesian analysis under different names and with different algorithmic strategies. In parsimony and likelihood analyses, the investigator is equipped with an optimality criterion and what may be described as a "landscape" of trees with different values for that criterion.

In the case of parsimony, we might imagine a landscape where a plane surface is defined by the average length of all possible trees. This surface is interrupted by valleys and hills. The valleys are filled with trees of longer-than-average length while the hills are filled with trees of shorter-than-average length. Other metaphors describe a sea with islands of shorter trees (Maddison, 1991). Any rational metaphor works if one gets the idea that some hills are higher than others, or some islands

contain shorter trees. Hills are separated from each other by the "inhospitable" landscape of average trees or by valleys of long trees. Or the islands are separated by long stretches of the barren ocean of long trees.

If our problem is simple and there is only one hill, then we can find and climb it. Perhaps a simple step-wise parsimony analysis or classic Hennigian argumentation will be efficient. If our problem is complex and there are many hills, some taller than others, we may ascend a low hill and feel we have found the most parsimonious tree when we have only found a locally optimal solution, not a globally optimal solution.

Local Optimum. A local optimum is achieved when the search finds the shortest tree(s) at the top of a particular hill in the parsimony landscape. Because searches always accept shorter trees and reject longer trees, it is possible to achieve a local optimum that is globally unparsimonious if there is no mechanism for exploring other hills. Locally optimal and globally unparsimonious means that you are on the top of a hill but there are higher hills that you have not found.

Global Optimum. A global optimum is achieved when the search finds the shortest tree(s) on the highest peak(s) on a particular parsimony landscape. Global optimality is never guaranteed in a heuristic search, but may be approached if strategies are adopted that allow exploration of the landscape and in a manner that allows discovery of multiple hills.

Searches, however implemented, are designed to keep the analysis from being trapped in locally optimal solutions by random perturbations of a given tree topology. If the perturbations result in a shorter tree, that tree is retained and it is perturbed; the process continues until the program cannot find any shorter trees. There are a variety of strategies to accomplish such perturbations, ranging from modest to radical, and we describe each of these more fully.

1. Random addition searches.
2. Rearranging tree topologies and analyzing isolated parts of a larger tree.
3. Parsimony ratchet.
4. Simulated annealing.

Random Addition Searches

All modern phylogenetic methods begin with a starting tree, built by some method (or randomly assembled). In parsimony analysis, the order in which taxa are added to the tree can affect the initial tree topology (as was the case with sequence alignment discussed in Chapter 5) and this, in turn, affects all subsequent manipulations. The usual strategy is to employ random addition searches (RASs). An RAS is a strategy of running the analysis many times (10s to 100s) and varying the initial tree by adding taxa randomly during the initial tree-building process. In the metaphor of the plane of parsimony, an iteration of the RAS algorithm allows the tree to land on a different part of the landscape. If the initial tree generated is close to optimal (close to the shortest tree possible), then it may (remember the search is heuristic) land on or near a hill and quickly find the most parsimonious tree on that hill. Increasing the number of RAS iterations increases the possibility of finding more hills. Using an initial algorithm to obtain a starting tree that is close to optimal

ensures a more efficient search. Combining this strategy with rearrangement of tree topologies allows the program to explore many of the trees on a hill or even jump to a new hill.

Rearranging Tree Topologies

The idea behind rearrangement of tree topologies is the exploration of tree topology space. For any particular phylogenetic problem, there is a large number of alternative trees. As the number of taxa increases, the number of possible topologies increases (see Felsenstein, 1978a). The purpose of the exploration of tree topology space is to visit as many tree topologies as possible, compare the new rearrangement to the older result(s) and determine if the new rearrangement results in a shorter tree or group of trees. If it does then the new result is accepted and another round of rearrangements is performed in an attempt to find another group of shorter trees. The process continues until no shorter trees are found or until the investigator terminates it.

One implementation of rearrangement is *branch-swapping* (Fig. 6.12). Most programs allow the investigator to perform one of a variety of branch-swapping routines, in concert with RAS (i.e., RAS + branch swapping). Three common branch-swapping routines are listed below. The terminology is that used in PAUP (Swofford, 2001), but the routines are available in all modern parsimony programs.

1. Nearest-neighbor interchanges (NNI).
2. Subtree pruning and regrafting (SPR), global rearrangements (Felsenstein, 2007).
3. Tree bisection and reconnection (TBR), branch-breaker (Farris, 1988).

Branch swapping, when used in concert with RAS, is an efficient method for finding short trees when the number of taxa is relatively small, say 100 taxa or less. However, if a large number of taxa or "messy" data are analyzed, the computer time used to find the shortest set of trees may be prohibitive. For example, the "Zilla" data set of 500 plants and 1428 DNA base pairs (bp) (Chase et al., 1993) ran for 3.5 months on three Sun workstations without finding shortest trees using RAS + TBR (Rice et al., 1997). Soltis et al. (1998) found similar problems using a 2800 bp data set—a set they had hoped would cut computation time because fewer equally parsimonious trees are likely to be present if the data set is larger.

One reason for the long computation time was the fact that then current implementations of heuristic search routines concentrated on finding all of the most parsimonious trees on each island of trees. Each RAS attempts to find all of the most parsimonious trees. If there are many islands of short trees, each RAS might go through a great number of trees (hundreds of thousands or millions). As Farris et al. (1996) and Goloboff (1999b) point out, when data sets are large and complex, finding a significant number of shortest trees from *different* islands of optimality will result in a consensus that is likely to be identical to that produced by finding *all* of the most parsimonious trees and then computing a consensus. Farris et al. (1996) used the jackknife to discover strongly supported groups—groups that would appear well supported in any analysis, while eliminating poorly supported groups that appear only on a minority of shortest trees. In essence, they reasoned that

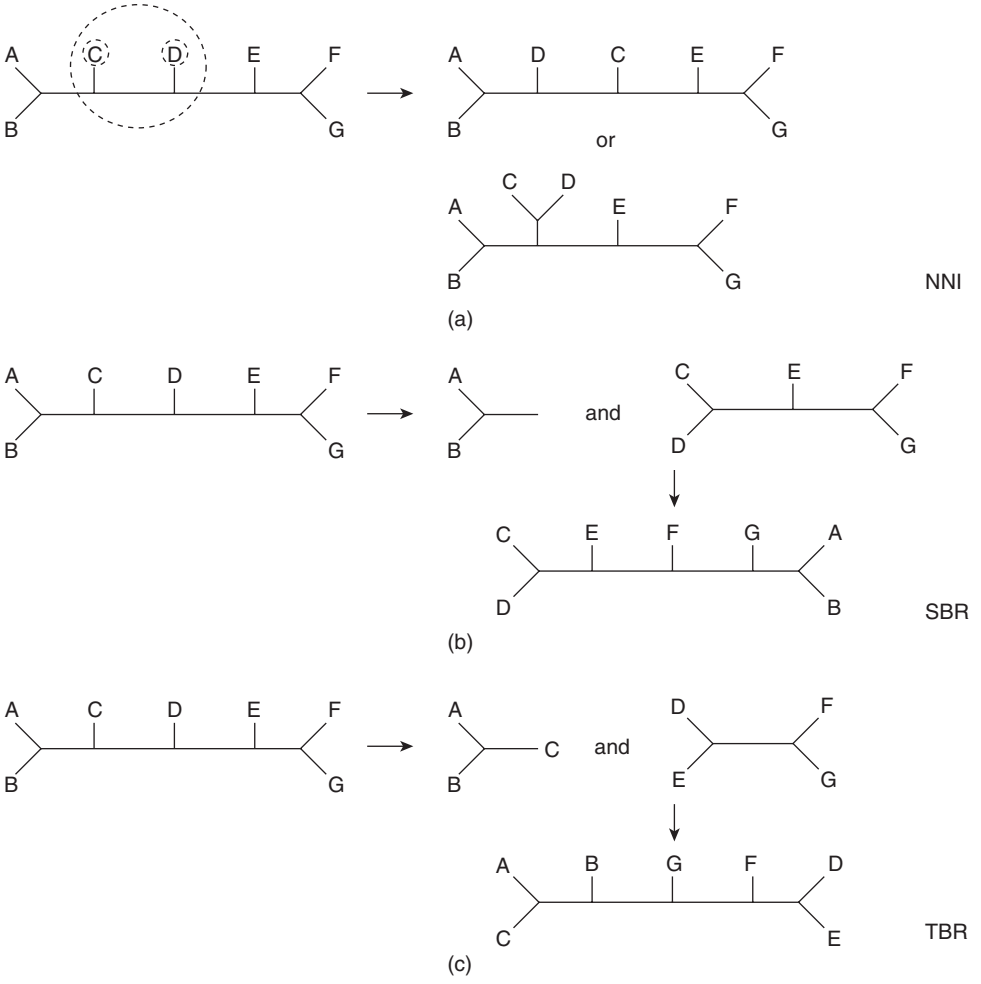


Figure 6.12. Branch swapping. (a) Nearest-neighbor interchanges, NNI. (b) Subtree bisection and regrafting, SBR. (c) Tree bisection and reconnection, TBR.

laboriously computing trees that would contain groups that disappeared when a consensus was computed was a waste of computer time.

Goloboff (1999b) demonstrated that simply increasing the number of RASs, using TBR, and keeping only a few trees for each search could dramatically decrease computation time. Using NONA, he was able to find shortest trees for the "Zilla" data set in 24–48 hours compared to 2.5 months of exhaustively finding the shortest tree on each island of trees. However, as Goloboff (1999b:417) pointed out, large data sets have "composite optima" that interfere with the quest for globally optimal solutions. Large trees of more than 50 taxa tend to have sectors, defined as local groups of taxa. A tree of 500 species, such as the "Zilla" tree, might comprise 10 sectors of 50 taxa each. The problem is: each sector may have its own local optima and whether it is placed on the tree may be partly independent of the placement of the other sectors. (If they are truly independent, then the problem is simplified; if

very dependent, then the problem is much harder because changing one affects the other(s).) This is exactly what RAS + TBR does; each iteration might break up a local optimum, and unless a great number of RASs are performed, no globally optimal solution will be found. The problem, then, is to find a tree where the sectors are in proper configuration with each other. Goloboff (1999b:417) states:

> Thus, the solution requires sectors be improved separately, one at a time—that those sectors which are suboptimal are improved without worsening the ones that are already optimal. For this, there are four basic methods: ratchet, tree fusing, tree-drifting, and sectorial searches. These methods do not attempt to find multiple trees during swapping, but simply concentrate on finding trees as short as possible.

Goloboff (1999b) suggested a number of swapping techniques built around the central idea stated above that the consensus tree produced by brief visits to many islands of most parsimonious trees would be identical to the consensus produced by laboriously calculating all of the most parsimonious trees. One strategy was simple: even with RAS + TBR, it is possible to cut computation time by saving only a few trees with each RAS and performing many RASs.

Speeding Up Rearrangements. SPR and TBR rearrangements can be speeded by recalculating only the part of a tree (a sector/window) that has been changed (Goloboff, 1999b). This uses a method outlined by Ronquist (1998b) to cut down on computation time by looking at the sectors nearest the connection point of the recalculated sector.

Tree Fusing. Tree fusing consists of exchanging subgroups of the same taxa between trees. The subgroups that are exchanged are present in the consensus of both trees and not dichotomously resolved in a consensus of the two trees. (Exchanging subtrees of dichotomously resolved taxa between consensus trees is unproductive because both trees have the same dichotomous relationships for the taxa exchanged.) If an exchange results in a shorter tree, then this tree is saved. This strategy is built around the idea that the formed subgroups might be optimal, but that relationships within them and to other subgroups might not be optimal for a particular tree.

Sectorial (Tree Window) Searches. Sankoff et al. (1994) and Goloboff (1999b) suggest that isolating certain clades and then performing an analysis might improve the resolution of the isolated subclade. Because fewer taxa are involved, the analyses are faster, and thus, the computational burden of attempting to escape local optima is less. Sankoff et al. (1994) isolated subclades of 20 or fewer nodes and performed branch and bound analyses on the isolated subclade. Goloboff (1999b) prefers the quicker method of TBR and thus analyzes more nodes (35–55). If the result improves the length of the tree, then the analysis moves to another subclade (another window or sector) and performs an analysis on the new subtree. Felsenstein (2004) suggests that if the purpose is to escape local optima, then the less exact method of Goloboff (1999b) might be preferable.

The Parsimony Ratchet

Nixon's (1999) parsimony ratchet is a technique that escapes local optima by emphasizing a limited number of characters within the data matrix to see if these characters

lead to shorter trees. From the original data matrix, some percentage (5–15 percent) of the characters are selected and weighted more heavily than the other characters. An analysis is then performed, and this will favor the weighted characters. The resulting tree topology is then evaluated using the entire data matrix with all characters equally weighted to determine the length of the tree. If a shorter tree results, the tree is saved. Many reweightings, searches, and evaluations are carried out, and the shortest trees are retained. Ratcheting is related to techniques that explore tree space by analyzing only part of the data (e.g., the Jackknife as used by Farris et al., 1996). Although Nixon (1999) implemented the ratchet specifically for parsimony analysis, Felsenstein (2004) calls attention to the fact that ratcheting can be used on any number of other approaches.

Simulated Annealing

Parsimony analysis of large data sets is one of many kinds of complex combinatorial problems for which exact solutions are not possible in practice. The solutions for such problems can be estimated using simulations of statistical mechanics using the Metropolis algorithm (Metropolis, 1953; Kirkpatrick et al., 1983). We shall hear much more about this approach in the chapter on statistical phylogenetics where it is used in Bayesian and (rarely) likelihood analyses. In short, the algorithm usually accepts a shorter tree, but it might accept a longer tree under certain, specified conditions. As the simulation proceeds, it wanders through the tree landscape usually favoring shorter and shorter trees until it settles on a peak (valley/island) from which it cannot escape. As implemented by Goloboff (1999b) under the name *tree-drifting*, suboptimal trees may be accepted during branch-swapping if they meet a criterion based on the relative fit difference between the trees.

OPTIMIZING CHARACTERS ON TREES

Character optimization is an initial step in understanding the evolution of characters. It can be applied to any tree, not just the shortest tree(s). You will find this useful if, for example, you wish to see the interpretation of the evolution of a particular character state on your preferred tree as compared to rival, less parsimonious, trees.

Both Farris (1970) and Fitch (1971) suggested strategies for optimizing characters on trees, each based on their own algorithms and neither providing formal proofs. Swofford and Maddison (1987) provided a proof for ordered optimization routines, and we use this to give an example of how to calculate the length of a tree in an earlier section. They also provided a proof for finding other equally parsimonious interpretations of character evolution based on most parsimonious resolutions (MPR) sets. The results are alternative ways of interpreting character evolution when one has more than a single most parsimonious character reconstruction. If we accelerate character transformation, then the effect is to push the time of transformation down the tree. This is commonly called ACCTRAN (accelerated transformation). If we delay character transformation, then the effect is to push transformation up the tree. This is commonly called DELTRAN (delayed transformation). These alternatives are easy to visualize with some examples (see also Wiley et al., 1991). We will begin with ACCTRAN, which is the original Farris (1970) optimization.

ACCTRAN Optimization

In Fig. 6.13a, we have a small matrix of two characters and four taxa. We will optimize the first character on the tree in Fig. 6.13b.

For a group of taxa, select the outgroup and root the tree with the outgroup and label all the terminal nodes/labeled taxa (Fig. 6.13b). Unlike computing tree length, you cannot pick any taxon to root the tree; it must be the outgroup because character polarity will vary with outgroup selection. Proceeding from the tips to the root, apply the following rules to the internal nodes.

Rule 1a. If the intersection of the state set is empty, then let the character set of the ancestor be the smallest interval from each set. (For binary transformations, simply label the ancestor with both characters, [a, b], [0, 1], etc.; Fig. 6.13c.)

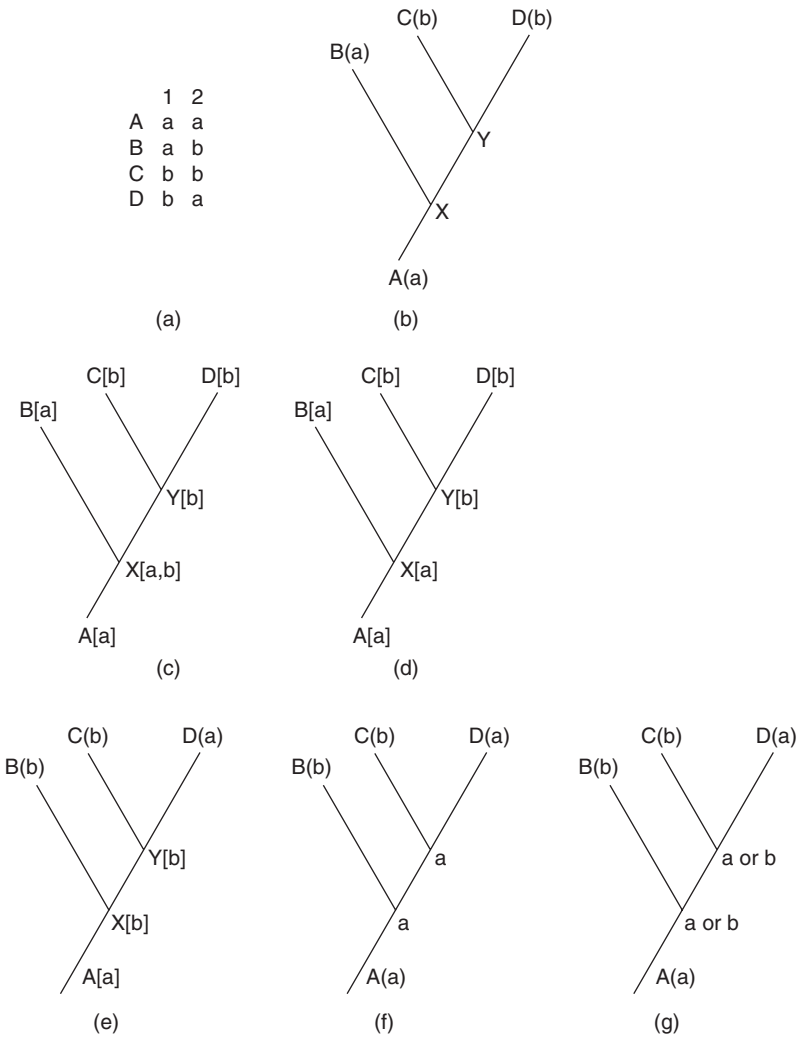


Figure 6.13. ACCTRAN and DELTRAN optimization. (a) A data matrix. (b–d) ACCTRAN optimization of character 1. (e–f) Two different but equally parsimonious optimizations of character 2 on the tree. (g) The MPR sets for all ancestors on the tree.

Rule 1b. If the character set is not empty, let the character set of the ancestor be the intersection as a closed interval. (For binary characters, this is simple: [0] or [1], [a] or [b], etc.; X in Fig. 6.13c.)

Once you reach the root, you then traverse up the tree toward the tips and apply these rules.

Rule 2a. If the descendant node has a character set with a single element, then it remains unchanged.

Rule 2b. If a descendant node has a closed interval, assign the interval with the smallest distance from the ancestor to the descendant.

For our very simple tree (Fig. 6.13b), we can see that the down-pass results in Y being assigned [b] and X being assigned [a, b] (Fig. 6.13c). As we move from root (A) to the tip, because A[a] and X[a, b], we change X to [a] (Fig. 6.13d). Because Y[b], we do not change it, even though its ancestor has the state set [a].

DELTRAN Optimization

For some character distributions, there are other possibilities besides ACCTRAN. Let us look at the second character column in Fig. 6.13a. ACCTRAN interprets character evolution as the accelerated transformation of A[a] to X[b]. Then it interprets another transformation from Y[b] to D[a] (Fig. 6.13e). Note that tree length is two steps.

However, there is an equally parsimonious tree with a different optimization, shown in Fig. 6.13f. In this interpretation, transformation from [a] to [b] is delayed; state [b] evolves independently in taxa B and C. This tree is also two steps in length.

Obviously, X and Y have two possible elements in their character sets, [a, b] for this transformation series (Fig. 6.13g), but ACCTRAN finds only a single element [b].

Swofford and Maddison (1987) presented a formal proof for finding all of the possible elements of the node's character set, not just some as found in ACCTRAN. This character set is termed the MPR set. We will illustrate the process of finding the MPR set using the binary characters in Fig. 6.13a and show that the MPR set is exactly those states shown in Fig. 6.13g. DELTRAN is then simply implemented with the same type of upward traversal as ACCTRAN, but results in character states optimized as in Fig. 6.13f rather than 6.13e.

1. Beginning with the unrooted tree (Fig. 6.14a), we pick an internal node and root the tree with this node (Fig. 6.14b; although we used taxon X, we could have actually used any internal node).
2. We perform a downward pass, assigning characters to the internal nodes just as we did in ACCTRAN (Fig. 6.14c).
3. We then reroot the tree with the next internal node, and we perform a downward pass, assigning character states to each internal node that has not been previously optimized (Fig. 6.14d). Note that we would not change the state set of X as it has already been determined.
4. Once we have rerooted the tree with all internal nodes, we have completed the assignment of the MPR sets for each of the nodes. We then root the tree

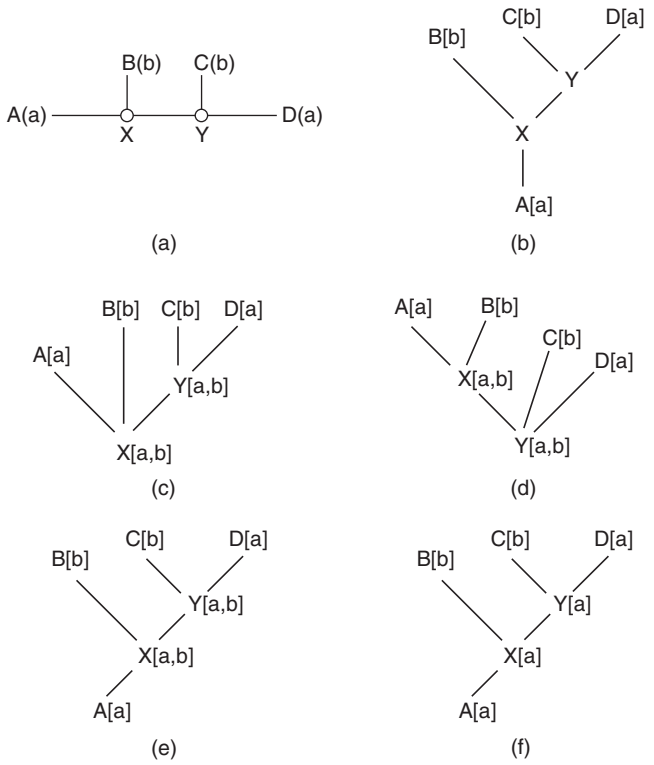


Figure 6.14. DELTRAN optimization. (a) Distribution of character states for character 2 of Fig. 6.13. (b–f) Sequential steps of optimization.

with one of the terminal taxa (presumably the outgroup) and perform an upward traversal, assigning characters to the internal nodes using the same rules we used in ACCTRAN, resulting in the optimization shown in Fig. 6.14f, which is identical to Fig. 6.13f.

The formal algorithms of Swofford and Maddison (1987) were built on informal optimization models by Farris (1970) and Fitch (1971) and applied only to dichotomous trees. W. Maddison (1989) extended MPR algorithms to polytomous trees.

SUMMARY TREE MEASURES

Once one obtains a tree or set of trees, there are various character performance measures that can be used to summarize the analysis and compare the tree(s) obtained with other possible solutions. Current computer packages provide this information, either automatically or upon request. They are not useful in evaluating the results of two data sets of the same taxa; but rather, to compare the results between trees for the same set of characters.

Tree Length. We have already described how tree length is calculated. Because the optimality criterion of a parsimony analysis is the minimum path of evolution that explains the data, tree length is a fundamental measure. Tree length is simply

measured by summing the number of changes that occur on the tree, as detailed above. There are two kinds of most parsimonious trees. First, there is the set of trees that have the same length but differ in topology, that is, parts of them contain different hypotheses of common ancestry. Second, there is the set of trees that have the same topology but differ in their interpretation of character evolution. These two types of trees have different qualities.

If an analysis results in a large number of equally parsimonious tree topologies, this reflects conflict among characters. The areas of conflict may be explored by performing a strict consensus analysis (discussed later in the chapter), which will result in polytomies where the conflict occurs. Because the number of possible trees increases quickly, it is possible to obtain many most parsimonious trees in a large matrix where conflict is confined to relatively small local regions (as, for example, among terminal species that belong to only one of many groups), but sometimes the entire consensus tree in such a situation can be unresolved. If an analysis results in a set of trees that are identical in topology but contain different interpretations of character evolution, the difference might be interesting from an evolutionary perspective. For example, it might provide possible tests of evolutionary mechanisms of character change that could be explored.

Consistency Indices. Kluge and Farris (1969) introduced measures of the performance of both individual characters and entire matrices relative to particular tree topologies. Consider a single character. If, on a particular tree, the states of this character could be mapped in such a way that there were no instances of homoplasy, then these states have "perfect" performance relative to the topology. However, if the topology was such that the only way to map the states was to invoke some level of homoplasy, then performance is less than perfect. Writ large, if an entire data matrix was composed of characters with states that required no homoplasy on a particular tree topology, then the performance of the entire data matrix would be "perfect" relative to that particular topology. The more conflict required to map the states, the greater the deviation from perfect performance. Various measures of character consistency can be generated to explore the performance, both of individual characters and entire data matrices. We can use the example provided by Wiley et al. (1991) to see how such measures are generated. We begin with data shown in Table 6.2.

Consistency index of a single transformation series (ci, or c). The ci of a single character is the ratio of the minimum number of steps or changes it might undergo

TABLE 6.2. Data matrix for the hypothetical clade A–E and its sister group OG. From Wiley et al. (1991).

Taxon	Transformation series							
	1	2	3	4	5	6	7	8
OG	0	0	0	0	0	0	0	0
A	1	0	0	0	0	1	0	1
B	1	1	1	0	1	0	1	0
C	1	0	1	1	1	0	0	0
D	1	1	1	0	1	1	1	0
E	1	1	1	1	1	1	1	1

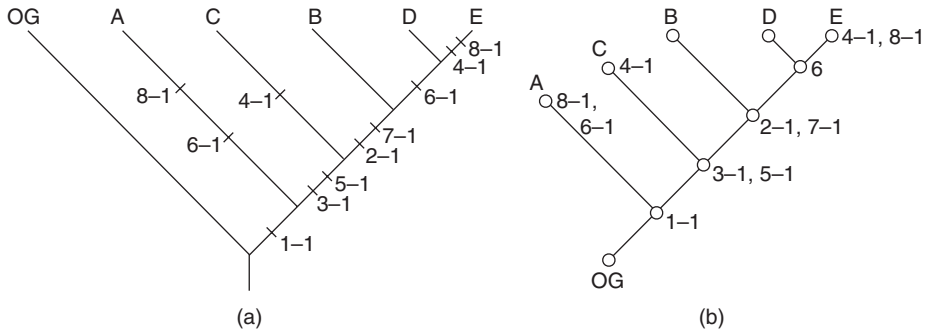


Figure 6.15. Two trees showing the distribution of synapomorphies of taxa A–E based on the matrix in Table 6.2. (a) A stem-based tree with synapomorphies mapped along inferred ancestral lineages. (b) A node-baesd tree with synapomorphies mapped at ancestral species nodes. Redrawn from Wiley et al. (1991), used with permission, Biodiversity Institute, University of Kansas.

and the number of changes or steps it actually undergoes on a particular tree topology:

$$\text{ci} = m / s$$

where m is the minimum number of steps, and s is the actual number of steps.

For binary characters, $m = 1$. For more than two states, m = the total number of steps necessary to minimally account for the evolution of the homologies (i.e., for three character states, $m = 2$; for four character states, $m = 3$, etc.).

Examine the matrix and consider character one in Fig. 6.15. Because 1 is binary, the number of minimum steps is $m = 1$. Note that there has been a single transformation from 0 to 1 at the internode leading to the clade ABCDE (labeled 1-1). Thus, $s = 1$ and the ci for this character is ci = $m/s = 1/1 = 1$.

Now consider character 8. Again, $m = 1$, but state 8-1 has evolved twice on the tree, so $s = 2$ and the ci for this transformation series is 0.5. Now consider character 6. If you do the calculations, you will find that the ci of character 6 is the same as character 8 (0.5). Note, however, that the quality of the two characters is different. In Fig. 6.15, the hypothesis that states 8-1 is a synapomorphy uniting taxa A and E is rejected in favor of the interpretation that each is an autapomorphy. As such, they contribute nothing to the resulting topology of the tree. In Fig. 6.15, the hypothesis that 6-1 is a synapomorphy is confirmed for the clade DE, but is disconfirmed for a clade containing D, E, and A. Thus, the character shared by D and E contributes to the topology of the tree, yet the ci in both 6 and 8 is identical.

The Rescaled Consistency Index (rc). To overcome this problem, Farris (1989b) introduced the rescaled consistency index (which appeared in Hennig86; Farris, 1989a). It is the product of the original consistency index and the **retention index (ri)**. We will use the characters in Fig. 6.15 to illustrate calculating the retention index and the rescaled consistency index.

The **retention index (ri)** measures the fraction of apparent synapomorphy to actual synapomorphy. To calculate the retention index, we need a new parameter, the ***g*-value (*g*)**; it is a measure of the "best of the worst" possible performance of each character relative to the actual performance of that character. The "best of the

TABLE 6.3. Some values* used to calculate rescaled consistency indices form Wiley et al. (1991).

TS	m	s	g	ci	ri	rc
1	1	1	1	1.00	0/0	0/0
2	1	1	3	1.00	1.00	1.00
3	1	1	2	1.00	1.00	1.00
4	1	2	2	0.50	0.00	0.00
5	1	1	2	1.00	1.00	1.00
6	1	2	3	0.50	0.50	0.25
7	1	1	3	1.00	1.00	1.00
8	1	2	2	0.50	0.00	0.00
Totals	8	11	18			

* m = no. changes a character might show on a tree; s = no. changes a character does show on a tree; g = minimum no. steps for each TS given a polytomy; ci = character consistency index; ri = character retention index; rc = character rescaled consistency index.

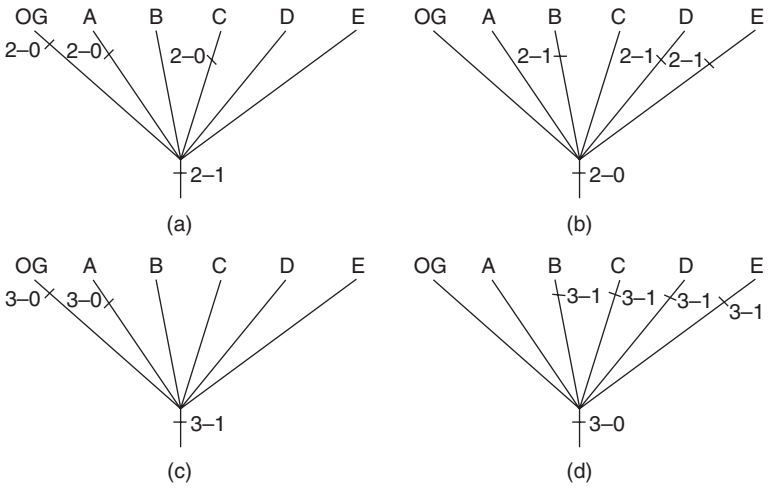


Figure 6.16. Performance of character states under the parsimony criterion. (a, b) Performance of character 2. (c, d) Performance of character 3. Table 6.3 shows the minimum number of times each is allowed to evolve independently on the polytomy. The fewest number of times is used to derive the metric "g" in Table 6.3. Redrawn from Wiley et al. (1991), used with permission, Biodiversity Institute, University of Kansas.

worst" performance would be the performance (in steps) of each transformation on a polytomy of all taxa by considering two scenarios in the binary case, one that assigns one state to the root and evaluates the performance of the other state at the tips and vice versa. The smaller value is the "best of the worst." Here is how it works.

Refer to Fig. 6.15 for the characters and Fig. 6.16 and Table 6.3 for how this is calculated. Consider character two. The worst possible performance of any character would be its performance on an unresolved tree. The "best of the worst" would be a contrast, in the binary case, between two states on a polytomous topology, with the state showing the fewest changes being better than the one showing more changes. (We are using binary transformation series for simplicity.) Now consider two cases, characters two and three (Fig. 6.16a–d). If we set the state 2-1 to the root

of a polytomous tree of all taxa, we can see that the tree requires 2-0 to evolve three times (Fig. 6.16a). Conversely, if we set 2-0 to the root, the tree requires 2-1 to evolve three times (Fig. 6.16b). It is a tie. Neither performs better in the polytomous case, so the *g*-value is $g = 3$.

Now consider character 3. If we set state 3-1 at the root, the tree requires 3-0 to evolve twice (Fig. 6.16c). If we set 3-0 to the root, we would require 3-1 to evolve four times (Fig. 6.16d). Thus, the "best of the worst" is the g-value of g = 2.

We can now define the rescaled consistency index, using the g-value and the s- and m-values that formed part of the original ci:

$$\text{ri} = (g - s)/(g - m)$$

where *g* is the best performance on the unresolved tree, *s* is the actual number of steps of a transformation series on the resolved tree, and *m* is the minimum number steps of a transformation series on the resolved tree. The rescaled consistency index, **rc**, is simply the **ri** * **ci** (see Table 6.3).

For character 2, the retention index ri = (3 – 1)/(3 – 1) = 1.0. For 3, the ri = (2 – 1)/(2 – 1) = 1.0 (Table 6.3). This makes sense; both transformation series have perfect ci-values and show no homoplasy. Thus, they are contributing the maximum possible to the tree topology.

Now, consider characters 6 and 8 (Table 6.3). They have identical consistency indices (0.5), but 8 contributes nothing to the structure of the tree while 6 acts as a synapomorphy in one place and an autapomorphy in another. If we calculate the g-value for character 8, we see that $g = 2$. Character 6 has a *g*-value of $g = 3$. Calculating ri, we find the following (Fig. 6.16e):

$$\text{ri}\ (8) = (2 - 2)/(2 - 1) = 0.0$$
$$\text{ri}\ (6) = (3 - 2)/(3 - 1) = 0.5$$

The result is straightforward. The most parsimonious tree interprets the state 8-1 as two instances of homoplasy resulting in two autapomorphies. Its contribution to the structure of the tree is zero. In contrast, the most parsimonious tree interprets the state 6-1 as one instance of homoplasy, one instance of synapomorphy, and one instance of autapomorphy. The instance of synapomorphy contributes to the structure of the tree. Incidentally, all instances of unique characters (autapomorphies or characters that map on at the ingroup node) will also have ri = 0.0, removing unique characters from contributing to the ri-value. In contrast, the ci-values of unique characters are ci = 1.0. This becomes important when we consider the next set of measures: ensemble values.

Ensemble consistency indices can be used to examine the relationship between an entire matrix and a given tree topology. One commonly reported index is the ensemble consistency index (CI) (Kluge and Farris, 1969). For a binary matrix, this index is simply calculated by taking the ratio of the number of data columns and the length of the tree. If there is no homoplasy, then the CI-value will be CI = 1.0. Deviations from this value indicate that homoplasy is present.

The CI-value suffers from some problems. The CI is artificially inflated by unique characters that contribute nothing to the structure of the tree. One solution is to calculate CI after eliminating all transformation series that contain unique characters (Carpenter, 1988). This result can be calculated in most parsimony programs.

Another problem is the fact that there is a negative relationship between CI-values and the size of the data set (Archie, 1989). Large data sets have small CI-values simply as a function of the size of the data set, not the contribution of the data to the tree structure. To address these problems, Farris (1989b) suggested that ensemble values be calculated using rescaled values, not raw values. The various formulae needed are shown below.

$$\mathrm{CI} = M / S$$

where CI is the ensemble consistency index, M is the sum of m-values for each character, and S is the sum of s-values for individual characters ("totals" row, Table 6.3).

$$\mathrm{RI} = (G - S)/(G - M)$$

where RI is the ensemble retention index and G is the sum of individual g-values for all transformation series ("totals" row, Table 6.3). Then:

$$\mathrm{RC} = (\mathrm{CI})(\mathrm{RI})$$

where RC is the ensemble rescaled consistency index. Values calculated for the example (Table 6.3) are shown below.

$$\mathrm{CI} = M / S = 8/11 = 0.727$$
$$\mathrm{RI} = (G - S)/(G - M) = (18 - 11)/(18 - 8) = 0.7$$
$$\mathrm{RC} = (0.727)(0.7) = 0.509$$

EXAMPLE 2: OLENELLOID TRILOBITES

The basal lineages of trilobites, apart from the controversial Agnostida, comprise what Fortey (1997) recognized as the order Redlichiida (early to middle Cambrian). In a series of papers, Lieberman (1998, 1999, 2001, and 2002a) analyzed the relationships of this group and came to the conclusion that it is actually a grade group leading to Eutrilobita (the crown trilobites). The 2002a paper is the one we shall consider here.

Fortey (1997) recognized the paraphyletic nature of Redlichiida, and he divided the grade into two suborders: Olenellina and Redchiida. Among olenellines, Fortey recognized two superfamilies: Olenelloidea and Fallotaspidoidea. Lieberman (1998) analyzed these basal lineages and concluded that the olenelloids were monophyletic, but that Fallotaspidoidea was not; some fallotaspidoids were more closely related to olenelloids while others were more closely related to redlichiids. Further, the redlichiids are related to the rest of Eutrilobita (Fig. 6.17). Thus, olenelloids and a set of taxa traditionally assigned to the Fallotaspidoidea (specifically, judomioids and nevadioids) are the sistergroup of all other trilobites, including the other fallotaspidoids; he referred these to a monophyletic Olenellina. Lieberman (1999, 2001) subsequently analyzed relationships among the Olenellina, returning to the problem of the remaining fallotaspidoids and their relationships to other trilobites in Lieberman (2002a).

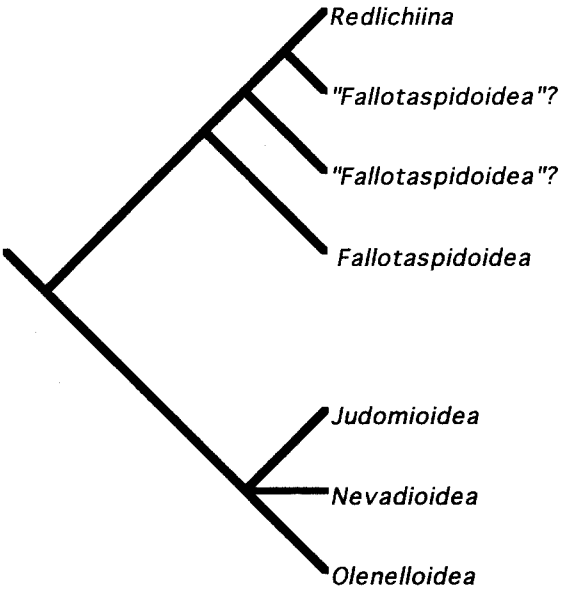


Figure 6.17. Relationships of basal trilobites as resolved in Lieberman (1998). From the *Journal of Paleontology*, used with permission of the Paleontological Society.

Lieberman (2002a) analyzed 16 species for which well-preserved specimens were available (see Table 6.4). This included all known genera of nonolenellinid fallotaspidoids and three species representing the basal members of the rest of Eutrilobita (the apical trilobites). These basal members (one species of *Bigotina* and two species of *Lemdadella*) were in effect used to root the fallotaspidoids to taxa up the tree. The analysis also added an outgroup; the outgroup used was a synthesis created by reconstructing the character vector for the inferred common ancestor of Olenellina *sensu* Lieberman (1998) using MacClade (Maddison and Maddison, 1992–2008). Twenty-nine transformation series were analyzed, all taken from the exoskeleton of holaspids. (In redlichiids and olenelline trilobites, and indeed in most trilobites, holaspids are specimens were subsequent molts do not result in an increase in the number of thoracic segments, and thus presumably represent adults.)

The resulting data matrix (Table 6.4) was analyzed using PAUP* 4.0 (Swofford, 2001). Lieberman ran 100 RASs and TBR branch-swapping. Transformations were optimized using ACCTRAN. A single most parsimonious tree (Fig. 6.18) of 106 steps was recovered. Summary tree measures were CI = 0.53, RI = 0.58, and RC = 0.30.

Character support for this tree ranges from a number of unique and unreversed transformations (e.g., data columns 3, 4, 7, 24) with ci = 1.0 to characters that are highly homoplastic (e.g., column 19, ci = 0.20). This analysis points out how ci values can be misleading. In particular, sometimes a character may have ci = 1.0 but there is missing data. For example, having a pygidium with its axis distinctly separated from the pleural field (column 27, character state 1) appears to be unique and unreversed, but there are several taxa that could not be scored for this character. Perhaps in reality these taxa possess state 0 such that the ci really should no longer = 1.0. Here we consider three character matches in greater detail.

TABLE 6.4. Character state distributions for taxa used in phylogenetic analysis of Lieberman (2002a) Selected characters and character states are discussed in text Missing data are indicated by "?". Character numbers are listed at top of table. Character states listed as "V," "W," "X," "Y," and "Z" are polymorphic, where "X" = (0&1), "Y" = (1&2), "Z" = (0&2), "W" = (0&1&2), and "V" = (1&3).

	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2
	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
Olenellina Node	0	0	0	0	0	0	0	0	0	0	0	0	W	0	0	Z	0	0	0	0	0	0	0	0	Z	0	0	0	0
Repinella sibirica	0	0	0	0	1	V	0	1	0	0	?	0	0	1	?	1	1	0	?	0	0	1	0	0	0	1	?	?	?
Profallotaspis jakutensis	0	0	0	0	1	3	1	1	0	0	1	1	1	0	1	1	1	0	1	1	2	2	0	0	?	?	?	?	?
Pelmanaspis jurii	0	0	0	0	1	3	1	1	0	0	1	1	1	0	1	1	1	0	1	1	2	1	0	0	?	?	?	?	?
Eofallotaspis tioutensis	0	0	1	0	1	3	1	1	0	0	1	1	0	2	?	0	1	1	0	1	1	1	0	0	?	?	?	?	?
Daguinaspis ambroggii	0	0	1	0	1	3	1	1	1	0	0	1	X	2	2	0	0	0	0	1	2	2	1	0	0	0	0	0	0
Choubertella spinosa	0	1	1	0	1	2	1	1	1	0	1	1	0	2	2	0	0	0	1	1	2	2	1	0	0	1	0	1	1
Fallotaspis typica	2	1	0	0	0	0	1	1	0	0	0	1	1	0	2	0	1	1	1	1	1	0	0	0	Y	1	?	?	?
F. bondoni	1	1	0	0	X	X	1	1	0	1	1	0	0	0	1	Z	1	1	1	1	1	0	0	0	2	1	0	0	0
Parafallotaspis grata	1	0	0	1	1	2	1	1	0	0	2	1	0	2	0	2	1	1	0	1	1	0	0	0	?	?	0	0	1
Archaeaspis hupei	2	0	0	1	1	1	1	0	1	0	0	1	0	0	1	0	0	1	1	1	1	1	0	0	?	?	?	?	?
A. nelsoni	1	1	0	1	1	3	1	0	0	0	1	0	0	0	2	0	?	?	1	1	1	0	0	0	0	1	?	?	?
A. macropleuron	1	1	0	1	1	1	1	0	0	1	0	0	0	0	2	0	?	?	0	1	1	0	0	0	1	0	?	?	?
Fallotaspidella musatovi	1	0	0	1	1	1	1	1	1	0	0	1	1	0	1	1	0	1	0	0	0	0	0	0	?	?	?	?	?
Bigotina bivallata	1	1	0	1	1	V	1	X	1	0	1	1	Y	2	0	2	1	0	0	0	0	1	0	1	?	?	1	1	1
Lemdadella antarcticae	1	0	0	1	1	1	1	X	0	0	1	1	2	0	1	0	0	0	0	0	0	X	0	1	?	?	1	1	1
L. linaresiae	1	1	0	1	1	1	1	1	0	0	X	1	1	2	1	2	0	0	0	0	0	1	0	1	0	1	1	1	1

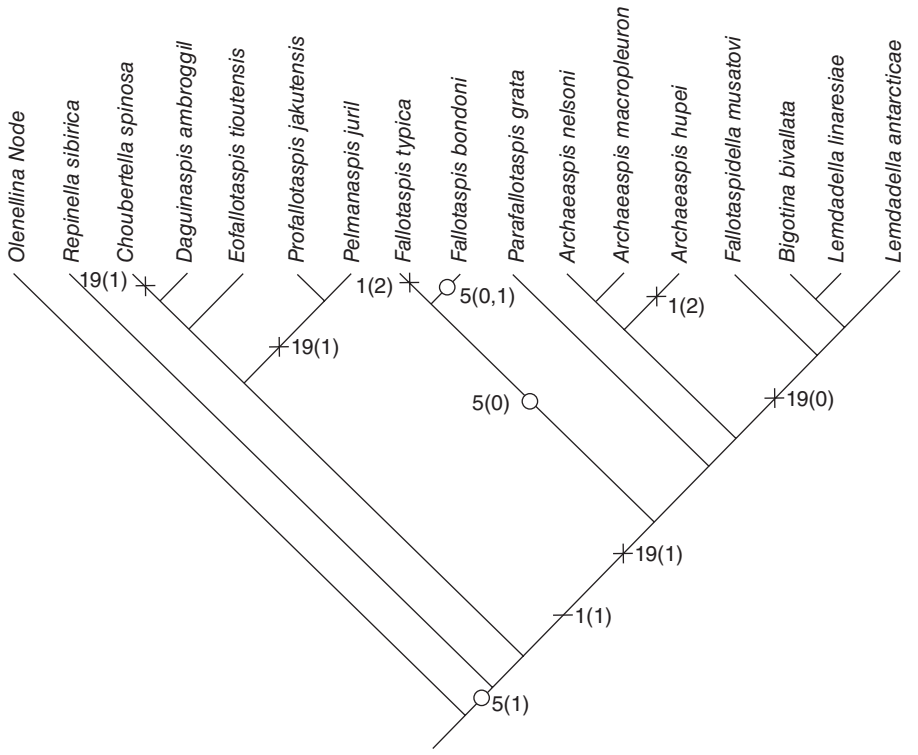


Figure 6.18. A phylogenetic tree from Lieberman (2002a). Only characters from Table 6.4 that are discussed in the text are mapped on the tree. Modified from the *Journal of Paleontology*, used with permission of the Paleontological Society.

Character 1. The relative length of the anterior border of the head shield is either shorter than the length of the occipital ring (L0) as shown in Fig. 6.19a, about the same length as the occipital lobe, or much longer than the occipital ring (Fig. 6.19b). Although this character has a ci = 0.667, the first state (short to equal) is actually unique and unreversed and diagnoses a major clade of trilobites. The homoplasy is found in the second state, where the longer length of the anterior border is homoplastic in two species formerly assigned to Fallotaspidoidea and contributes nothing to the structure of the tree.

Character 5. The cranidium (central head region) of trilobites is composed of the glabella (segmented middle part) and a complex broader shield-like structure, the fixigenae. A furrow runs across the anterior part of the shield medially, and the glabella either contacts this furrow (coded 0, Fig. 6.19a) or does not (coded 1, Fig. 6.19b). This character also has a ci = 0.667. However, this is due to a reversal in only one taxon (*Fallotaspis typica*). Another taxon has missing data. Thus the similar ci-values conceal very different kinds of character evolution.

Character 19. Returning to the occipital ring, there are two kinds of ornamentation found on the medial surface of the ring, a faint node (0, Fig. 6.19a) or a spine (1, not figured). This character presents a typical worst case scenario for characters, with a ci only = 0.20. Although it diagnoses at least two smaller clades, we can safely conclude that homoplasy makes it of little use in distinguishing between alternate

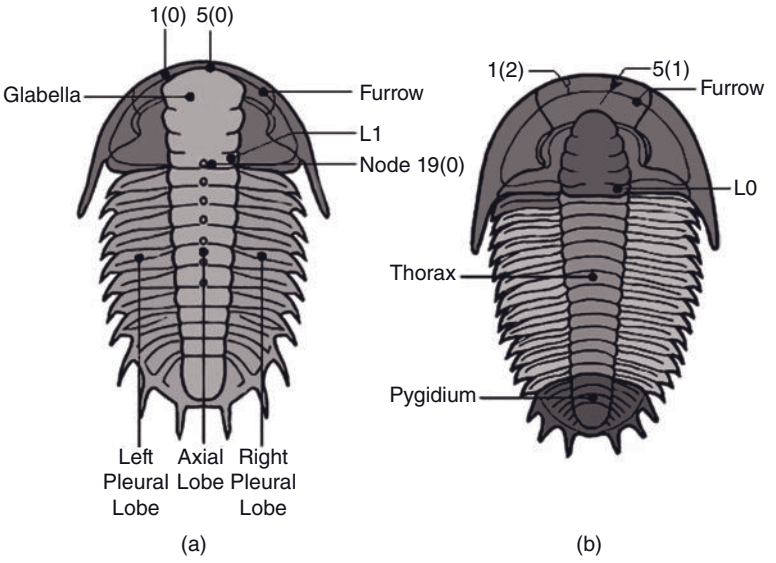


Figure 6.19. Two diagrammatic trilobites illustrating some of the characters used in the Lieberman (2002a) analysis. Key to characters: Anterior border of head shield narrow [1(0)] versus broad [1(2)]; glabella contacts furrow [5(0)] or not [5(1)]; a faint node on the occipital ring [19(0)] versus a spine [19(1)]. Used with permission of the Paleontological Institute, University of Kansas. See color insert.

tree hypotheses. When considering characters 1, 5, and 19 (in the context of all the other character data), we can see how that each character state has sorted itself out via parsimony. We are able to discern which character states ultimately provide the most phylogenetic utility and, by contrast, which are relatively uninformative.

EVALUATING SUPPORT

While the various consistency measures might yield information about the characteristics of particular transformation series on a tree and the amount of relative homoplasy, they do little to illuminate the actual support for individual monophyletic groups. There are, however, several ways of addressing this second, also important, question. At the most basic level, this question involves considering how robust is the tree(s) obtained; are all or most of the clades well supported, and which clades are suspect?

Count and evaluate the synapomorphies. The most direct method of evaluating support for a monophyletic group is to count and evaluate the synapomorphies, a largely qualitative assessment. Monophyletic groups supported by synapomorphies with high individual ci-values may be judged to be robust and the more of these kinds of characters the better. Monophyletic groups supported by synapomorphies with relatively low ci-values may be judged relatively weak and are likely to be rejected if new evidence emerges. The quality and complexity of the synapomorphies might also be judged. Highly complex synapomorphies, ones judged unlikely to have evolved twice, get high marks. Frequently specialists in a group will avoid

certain kinds of characters because they have the a priori notion that these characters are subject to homoplasy, but in principle, this should not be necessary because such characters should sort themselves out via the test of congruence when parsimony is applied.

This simple approach, as one might expect, has some problems. One problem is that the connection with some more general principles seems lacking. History is a contingent phenomenon, and evolutionary theory does not predict particular classes of characters that might fit our a priori judgments in these matters. For example, should losses be counted less than gains? Perhaps they should in the case of restriction sites, maybe not in the case of teeth. It is hard to generalize. Nevertheless, we are impressed with trees full of tick-marks that have high ci-values and suspect, with reason, that the clade is robust.

Another problem, shared by all methods of evaluation, is the problem of the independence of characters. Consider a hypothetical case in which one clade is supported by 10 unique and unreversed synapomorphies and the alternative is supported by two unique and unreversed synapomorphies. What if the 10 unique and unreversed synapomorphies corroborating the first clade are not independent of each other while the 2 synapomorphies supporting an alternative clade are independent? Then the 10 synapomorphies really represent a single synapomorphy and the alternative group would represent more support. But how do we evaluate phylogenetic independence? If synapomorphies lie on different parts of the tree and support different or nested monophyletic groups, then we can deduce independence. But if they appear as support for the same clade, then no such deduction follows and research would have to be undertaken to prove that they are truly independent (e.g., developmentally or genetically independent).

This simple approach also has an advantage, a concern for the evidence. Hennig (1966) stressed that critical phylogenetic inquiry is not simply a process of building a tree; it is a process of reciprocal illumination where the investigator is constantly questioning both the data and the results. More attention paid to the kinds and quality of evidence is never a bad goal and lays the foundation for more complicated measures of support. However, by the same token, obtaining a particular phylogenetic result should not necessarily then motivate subsequent targeted character search strategies by a worker. In particular, *a posteriori*, it would be invalid to spend a disproportionate amount of time trying to obtain additional character evidence to bolster support measures for a particular group to the expense of actually considering evidence that the group might not be monophyletic.

Bremer Support. Bremer (1988) suggested that a useful measure of support for a particular clade might be the difference in the length of a tree where it appeared as a monophyletic group and the length of the tree where it did not. We can easily run an analysis constraining the results to include the monophyletic group in question, and then run an analysis constraining the results to not include the monophyletic group in question using options commonly available in program packages. This can also be implemented using the package TreeRot (Sorensen and Franzosa, 2007). The difference in length between the two trees is a measure of how many steps longer a tree would be in order to overturn the monophyly of the group. For example, we run an analysis and obtain the group XYZ within a tree that is 100 steps in length. We can then run an analysis and have the analysis find the shortest tree that *does not* contain the group XYZ. Let us say that the resulting tree is 115

steps. The difference, 15 steps, is the Bremer support for the clade XYZ. If we do this for all groups found in our tree, we can obtain Bremer support values for each node containing a monophyletic group. Interpretation: we would have to accept a tree that is 15 steps longer than the most parsimonious tree in order to break up the clade XYZ.

Bremer (1988) argued that a way to quantify total Bremer support for a tree would be to simply sum the values at each node and then divide by the total tree length. However, partly because of the way trees are constructed, especially large trees based on complex data sets, it is not necessarily the case that different individual support values at each node should be thought of as additive across the tree (Faith and Ballard, 1994). This is because the presence of a particular clade within a tree might thereby constrain the appearance of other groups (Gatsey, 2000). Much depends on the distribution of homoplasy. Only if homoplasy is more or less evenly distributed across the tree will Bremer values be additive. If homoplasy is bunched in local regions of the tree, then they are not additive. Gatsey (2000) discusses measures of linked branch support which lead to better descriptions of tree stability than Bremer support alone.

High Bremer-values are almost always associated with strongly supported monophyletic groups and frequently correspond with other measures of nodal support such as the jackknife and bootstrap resampling, discussed below. The major problem is that no one has any idea what does or does not constitute a significant Bremer support value. Like the first method, we feel relatively confident when our clades have high Bremer support and not so confident when they have low Bremer support values. But the problem remains; what is a *significant* Bremer support value?

Jackknife and Bootstrap. Statistical measures of tree support are built on the statistical proposition that some parameter can be estimated from samples drawn from a population and that the result can be evaluated by drawing a new sample from the population. For example, if we estimate the mean body length of a population of mice, we would measure the body lengths of a sample drawn from the population and calculate the mean and standard deviation derived from the measurements. The true mean of the population would be expected to fall within some interval of length. We can check this hypothesis if we go back to the population, select another sample of mice, measure their body lengths, and find that the newly estimated mean falls within the interval we have calculated using the first sample. We can also test the hypothesis that another population of mice has mean body lengths similar to the first population. The problem in applying this strategy in phylogenetic analysis is that we have no new sample on which to draw. In such cases, we can simulate the statistical approach by subsampling the original characters and see if the subsample re-creates the result. If we draw many subsamples and reanalyze our problem with each subsample, we will create a set of trees whose number is the number of times we have subsampled. We can determine the frequency with which groups in the original analysis reappear over the course of subsampling by determining the frequency with which the groups appear in the set of trees.

For example, a strongly supported clade characterized by many synapomorphies might be expected to appear in all of the trees while a weakly supported group might appear only in a few. These are expressed as probabilities, usually by subject-

ing the set of trees to a majority consensus analysis (see below), with the probabilities expressed as the percentage that a particular clade appears on the majority consensus tree. There are two common methods for accomplishing this strategy in phylogenetic analysis, jackknifing and bootstrapping. The jackknife is the older approach and will be discussed first.

The Jackknife. The usual implementation of the jackknife is to rerun the analysis some predetermined number of times while deleting one or more observations without replacement. The trees resulting from this process are saved, and the frequency of appearance of each clade over all the trees constitutes its jackknife frequency. The variability among the trees generated by the analysis depends on the number of observations (data columns) deleted. Although any percentage of the original matrix can be subsampled, two common strategies are employed. The "half-delete jackknife" (Wu, 1986; Felsenstein, 1985a; Felsenstein, 2004) randomly samples half of the characters in each iteration, without replacement. The "parsimony jackknife" (Farris et al., 1996) deletes fewer characters. The half-delete jackknife apparently has properties similar to the bootstrap and, naturally, is favored by Felsenstein (2004) who introduced bootstrapping to phylogenetics (Felsenstein, 1985a). The parsimony jackknife, preferred by Farris et al. (1996), favors strongly supported groups and finds these groups with greater frequency when they are present. That is, such groups will have higher jackknife scores under parsimony jackknife than under half-delete jackknife.

Another approach is to jackknife taxa rather than characters (Lanyon, 1985). The question asked would take the form of seeing what effect the removal of species might have on the subsequent tree. Felsenstein (2004) points out that such a jackknife has no easy statistical interpretation. But, it might have its uses. For example, consider the scenario of the investigator who is analyzing 100 species in a group of 1000 species. It might be interesting to note that removal of 10 species drastically affects the topology, yielding groups not observed in the original analysis. Another application might be a scenario in which some groups were represented by many species and other groups, just as speciose, were represented by few species. Random subsampling would tend to pick taxa from the groups containing large numbers of species. Would this have an effect on the resulting topology that would call into question taxon sampling?

The Nonparametric Bootstrap. The bootstrap was first used in phenetic studies (Mueller and Ayala, 1982; see Felsenstein, 2004) before its introduction to phylogenetics by Felsenstein (1985a). There are two versions, parametric and nonparametric, of which the nonparametric is commonly employed to assess the fit of data to a tree. The idea behind the nonparametric bootstrap is that the matrix is a sample of the true underlying distribution of characters. If we knew the true underlying distribution of characters, then we could assess the degree of support inherent in the data for any clade. Of course, we do not know this, but bootstrapping is a way to simulate the variability in the underlying pattern of character distribution. It is a method for estimating the unknown and presumed true distribution by using the known empirical data.

A nonparametric bootstrap analysis begins by creating a number of pseudoreplicate data matrices by subsampling the original data matrix, with replacement. Each being a matrix of the same size as the original and composed of a random sample

of characters from it. Any one transformation series might be represented in any pseudoreplicate matrix once, twice, many times, or not at all. Each of these matrices is analyzed and the trees collected into a set of trees. The frequency that particular clades appear over the entire set of shortest trees generated by analyzing all of the pseudoreplicate matrices constitutes its bootstrap score (summarized using a majority consensus technique). We can even generate confidence intervals such that if we run a large number of similar bootstrap analyses, we would expect the score for a particular clade to fall within that interval. Intuitively, if a particular clade has a high number of characters supporting its monophyly, and there are few characters that refute its monophyly, then chances are that at least some of these characters will appear in each pseudoreplicate matrix and the group will appear in many or all of the sets of trees. Conversely, if evidence for the monophyly of the group is weak or if there is a high level of homoplasy in the original matrix, the group might not appear at all, or at low frequency.

There is considerable literature on biases in jackknife and bootstrap analyses as applied to phylogenetic analysis, and the consensus of opinion is that the probabilities obtained are usually low relative to the perceived reality of the clade given the data. For example, Hillis and Bull (1993) suggest that bootstrap values as low as 70 percent may indicate well-supported clades, in contrast to the usual statistical threshold of 95 percent (but see Newton, 1996; and see Felsenstein, 2004, for additional literature and discussion of various methods used to reduce bias). The probabilities obtained are not probabilities of the reality of the clades per se, but reflect the relative support of the clades in the data matrix given the assumptions of the analysis.

Permutation Tests. Permutation tests have been used in phylogenetics, but every application has been to a greater or lesser extent controversial. Both Swofford et al. (1996) and Felsenstein (2004) provide examples of the application of these tests and the controversies surrounding them. In general, permutation tests fall into two categories.

Permutation tail probability tests (Archie, 1989; Faith and Cranston, 1991) are designed to test for hierarchical structure. The test works by shuffling characters in each data column and assigning them randomly to species. We would expect that if we analyzed any single shuffling of the data the result would be much worse than an analysis of our original matrix, if our original matrix contained real hierarchical signal. Alternatively, if our original data was itself comprised of randomized data that lacked any phylogenetic signal then we would not expect to see a difference between our original data and a randomized version of our original data in such parameters as tree length or ensemble consistency indices. Of course, it is possible, by chance, that a randomized version of our data might yield a tree with as much support as or original data, but we would not expect to see this very often if our tree was supported by "good" data. We can specify how often we might expect a randomized matrix to perform as well as our original data and that expectation is the usual statistical expectation of $p = 0.05$. If we do a great number of permutations, derive a tree from each permutation, and calculate its length, we can build up a distribution of tree lengths (or other measures). If only a small percentage of these trees are as good as our tree derived from the original data, then we reject the null hypothesis that there is no difference between our original tree and a tree derived from random data. We conclude that the data contain hierarchical signal. A variant

test, the topology-dependent permutation tail probability test (T-PTP), was developed by Faith (1991) to test in a similar fashion whether specific clades are supported.

Incongruence Length Difference. The second major use of permutation tests is to test the null hypothesis that two data sets are inferring different trees. Rejection of the null hypothesis implies that both data sets infer the same or highly similar tree topologies. Data are combined and permutations are conduced by permutating the data columns (not the order of characters). This test was suggested by Farris et al. (1994a, b) and independently introduced in PAUP by Swofford in 1995 as the partition homogeneity test.

Measure of Skewness. Imagine the situation in which we could determine the length of every possible tree and plot a histogram of tree length frequency. The peak would contain many less parsimonious trees, trailing off to fewer very long trees on one side and fewer very short trees on the other side. Hillis (1991) suggested that if this frequency distribution was skewed, it suggested strong phylogenetic signal because there would be far fewer relatively short trees than long trees (see also Huelsenbeck, 1991a). By contrast, trees based on low quality or effectively randomized data should show little skewness and instead symmetry of tree-length distribution. Hillis (1991) proposed using the g_1 statistic, a measure of tree-length frequency distribution, as a way of assessing phylogenetic signal in a data set. The resultant g_1 statistic could be compared to the distribution of g_1 statistics produced from randomized data to assess the degree of significance. The proposed test is an interesting one, but it has been suggested that the tree-length frequency distributions of some data sets possessing or lacking phylogenetic signal do not always behave in such a stereotypical manner (Källersjö et al., 1992). Because of this, and for other reasons, it has been suggested that tree-length frequency skewness "may be of limited power in detecting phylogenetic signal" (Felsenstein, 2004:363).

USING CONSENSUS TECHNIQUES TO COMPARE TREES

Topologically different trees can be combined to explore their common and unique features using consensus techniques. A consensus tree is a summary of the common topological features of two or more trees that contain the same taxa and differ in details of their topology. Usually rooted trees are compared; however, consensus techniques can also be applied to unrooted trees. It is entirely possible that two unrooted trees are indistinguishable, yet two rooted trees derived from them can be in conflict, simply by specifying the root in different locations.

There are several techniques that are used to derive consensus trees. Each has its strengths and weaknesses. Three kinds of consensus trees are commonly employed, and several other kinds have been used.

Strict Consensus (Rohlf, 1982). In phylogenetic analysis, strict consensus trees contain only those monophyletic groups that are common to all of the trees compared (Fig. 6.20). Strict consensus trees are a mechanism to convey a reduced tree that highlights the common branching patterns, common speciation events for which all most parsimonious trees agree or for which trees derived from different data sets of the same taxa agree. It asks the question: what groups are always monophyletic? This is the type of consensus tree most frequently employed in phylogenetic studies.

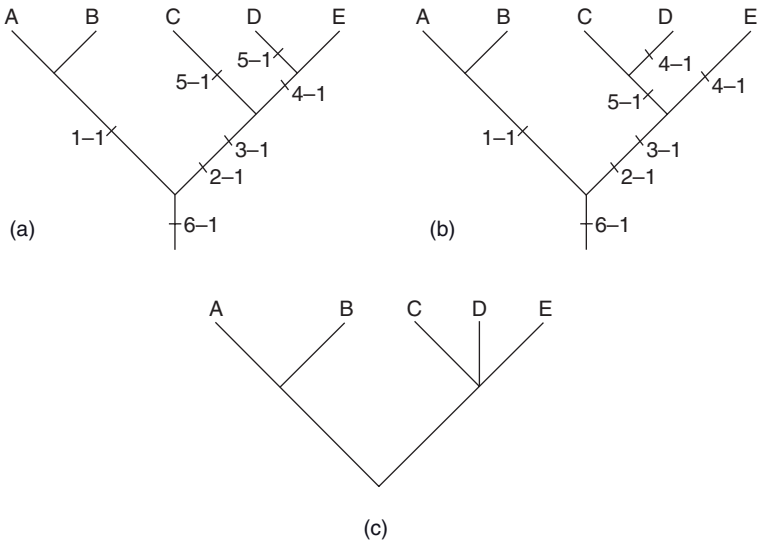


Figure 6.20. Strict consensus. (a–b) Two equally parsimonious trees. (c) The strict consensus of the two trees.

It does not give the best estimate of phylogeny and character evolution, because any of the most parsimonious trees provide a more parsimonious explanation of character change. However, the strict consensus provides an efficient summary of how the different most parsimonious trees agree. Thus, it is a way of reflecting common monophyletic groups.

Majority-rule Consensus (Margush and McMorris, 1981). Strict consensus trees are actually one extreme of the Mℓ family of consensus trees of Margush and McMorris (1981), where ℓ varies between 50 percent and 100 percent and is defined as the percentage of times a particular group appears among all of the trees compared. A commonly reported tree is the 50 percent-majority rule consensus tree, which reports all clades that appear in more than 50 percent of the trees. A strict consensus tree is simply a M_{100} tree, and a M_{75} tree would report all groups that appear in 75 percent of the trees compared, etc. The common output format is a tree with nodes that report the percentage of the occurrence of the resolved groups. The Mℓ family of consensus trees (including strict consensus) can be used either for rooted or unrooted trees.

Majority-rule consensus trees are sometimes reported in the parsimony literature, but usually only in cases where the strict consensus is poorly resolved and the scientist wants to put a more positive spin on the results. While it is tempting to consider them a "best estimate" of phylogeny, they are not so under the optimality criterion of parsimony because there remain alternative equally parsimonious trees. Further, there is no reason to suppose that the percentage of times a particular tree resolution appears in a result has any significance given that sometimes quite different results are equally parsimonious. The validity of this technique as a general way of presenting parsimony results is dubious.

These consensus techniques are more commonly used as output on statistical tests such as the jackknife and the bootstrap, as discussed above. In these cases, the

majority-rule consensus values are used to assess the fit of the data to the tree. They may be a reasonable proxy for qualitative degree of confidence in results.

Adams Consensus (Adams, 1972). Adams consensus is strictly for rooted trees (Felsenstein, 2004). Informally, Adams consensus seeks the highest resolution possible for two trees and accomplishes this by moving inconsistent taxa to basal positions where they do not conflict. The result may not necessarily reflect monophyletic groups supported in the original analysis, even those common to all trees (as in strict consensus).

Adams consensus has been used to find taxa that are "unstable" on a set of trees that are otherwise similar. They can be used to answer two kinds of questions. First, what is the most highly resolved tree that will identify "problem" taxa? An example of a problem taxon might be a species of hybrid origin with partial expression of synapomorphies of each group that contains its parental species (Funk, 1985). Second, are the trees logically consistent? Most computer packages can calculate Adams consensus trees. Wiley et al. (1991) provide some simple examples that can be calculated by hand.

There are a number of other consensus techniques, including semistrict or combinable component consensus (Bremer, 1990) and Nelson or Nelson-Page consensus (Nelson, 1979; Page, 1990; Swofford, 1991; Felsenstein, 2004), and consensus techniques that are based on branch lengths (e.g., Neumann, 1983) or path distances (e.g., Lapointe and Cucumel, 1997). Bryant (2003) provides a list of consensus techniques and their uses in phylogenetics and Felsenstein (2004) provides an overview.

STATISTICAL COMPARISONS OF TREES

A class of tests termed *paired-sites tests* can be used to test the null hypothesis that two tree topologies are statistically identical. For example, we may have the most parsimonious tree found in our own analysis and wish to see if it is statistically different from a tree of the same taxa that was previously published. Or we may wish to test the proposition that our most parsimonious tree is significantly different from a suboptimal tree that is 10 steps longer. There are parametric and nonparametric versions of the paired-sites test. Intuitively, if there is very little difference between the trees, there should be very little difference in the variation of site performance from one tree to the other and the null hypothesis would be confirmed. Alternatively, if there are large differences at many sites, then we might expect the null hypothesis to be rejected and conclude that there is a significant difference between trees.

We may wish to compare our data and results with the work of others. Or we may wish to see if our shortest tree is really that different from a tree that is almost as short. There are a number of ways of accomplishing such tasks. The most common technique is the Wilcoxon signed ranks test first introduced to parsimony analysis by Templeton (1983a, b; see also Felsenstein, 1985b), or its simplified version, the winning sites test (Prager and Wilson, 1988). Some programs, such as PAUP* and Mesquite, can be used to calculate the Wilcoxon signed ranks test and the winning sites test statistics. Other tree comparison tests are parametric and, thus, require models of evolution. We will discuss these in the next chapter.

WEIGHTING CHARACTERS IN PARSIMONY

Character weighting in parsimony can take several forms, including selecting a priori characters (what characters to include in an analysis) and assigning a particular cost of transformation to one or more characters. Character selection is common in morphological analyses and also present in molecular analyses in the form of gene selection. A priori weighting assigns the cost of a particular transformation prior to an analysis and usually is based on some implicit model of evolutionary changes envisioned by the investigator. A posterori weighting is performed after an analysis and is based on assumptions regarding character performance in that analysis.

Character selection. Character selection is an inevitable consequence of the inability to examine all of the characters of specimens. How characters are selected may profoundly affect the results of a phylogenetic analysis. We suspect, but cannot prove, that the reason many morphological analyses work well in a parsimony framework is that the investigator picks characters that show (1) a fairly low level of intra-taxon variation and (2) an interpretable level of inter-taxon variability. This would result in the informal adoption of a model of evolutionary change that favors parsimony analysis so long as the covariation of true synapomorphies is greater than the covariation of true homoplasies. However, because we do not know the distribution of true synapomorphies over that of true homoplasies, we are betting that the level of homoplasy is relatively low compared to that of homology. Note that this differs rather strongly from typical molecular analysis when entire contiguous regions of DNA or amino acids are analyzed and there is no a priori picking of properties that show these characteristics. In a DNA analysis, we may find characters that do not vary in their states and have no "control" over intra-taxon variability. In molecular analysis, character choice resides in the gene regions picked for the analysis.

A Priori Weighting

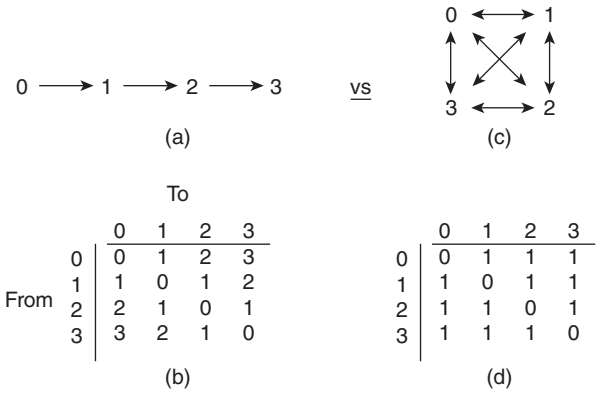
All forms of parsimony (Fitch, Wagner, etc.) are special cases of applying Sankoff optimization, with "Sankoff characters" being those characters where the cost of transformation between states is specified by the investigator (e.g., Sankoff, 1975; Sankoff and Rousseau, 1975; Sankoff and Cedergren, 1983; Swofford and Maddison, 1992; Goloboff, 1998). As Swofford has pointed out, treating all character states as equally weighted is assigning a weight of one to the weighting function. Character weighting is different from ordering. Although the investigator may think that no evolutionary assumptions are invoked when he or she decides to treat all characters as equally weighted and unordered, he or she has, in fact, made an explicit evolutionary assumption that each transformation has the same information content.

What we usually think of as weighting in parsimony is the activity of assigning different costs (in steps) to different kinds of transformation. The weights we assign are based on what we think are the probabilities of the transformation of one character state to another. Modern parsimony programs make it possible for the investigator to assign different weights to different characters and to assign different weights to the transformation of different states within a character.

A Priori Weighting: Parsimony-Specific Weighting Functions. Coding characters might imply different step-costs under certain conditions in a parsimony analysis and optimization. Some of these costs are bound up in selection of a parsimony

criterion. For example, Fitch parsimony introduces equal costs between transformations while Wagner parsimony may impose different costs (zero to one costs one step; zero to two costs two steps). Others costs fall into the class of a priori weighting as a function of general parsimony where differential cost is assigned to characters within the context of an overall parsimony analysis.

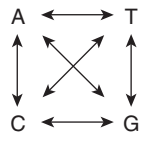
In modern computer packages, differential weighting of states may be easily accomplished by using one or more step or cost matrices and referring particular characters to these step matrices during analysis (Maddison and Maddison, 1992; and Ree and Donoghue, 1998). Maddison and Maddison (1992) discuss a simple example of two characters with five character states, one treated as with Farris optimization (Fig. 6.21a, b), and the other with Fitch optimization (Fig. 6.21c, d). The cost of transformation in each case is entered into the matrix according to the evolutionary model adopted, and these costs are added to the weight of the edges when the tree is calculated. Another example would be differential weighting of transversions and transitions among sequence data (Fig. 6.22).



From \ To	0	1	2	3
0	0	1	2	3
1	1	0	1	2
2	2	1	0	1
3	3	2	1	0

	0	1	2	3
0	0	1	1	1
1	1	0	1	1
2	1	1	0	1
3	1	1	1	0

Figure 6.21. Character relationships and cost (or Sankoff) matrices. (a) An ordered and polarized series of character states. (b) A cost matrix expressing the cost of transformation between each state in (a). (c) An unpolarized and unordered series of states. (d) The cost matrix expressing the cost of transformation between each state in (c). From Maddison and Maddison (1992).



	A	T	C	G
A	0	2	2	1
T	2	0	1	2
C	2	1	0	2
G	1	2	2	0

(b)

Figure 6.22. An example of a molecular cost matrix. In this case we have an unpolarized and unordered series of character states, but we have set the cost of transversions twice as high as the cost of transitions.

Weighting by Performance

Two common methods of performance weighting are successive approximations (Farris, 1969) and self-weighted optimization (Goloboff, 1997). Each is founded on the proposition that parsimony analysis can be refined by taking into account the performance (in units such as steps) of characters, given a topology and on the idea that the search for the final topology can be "informed" by character performance.

A successive approximation (Farris, 1969) performs an equally weighted analysis and then proceeds to a round of analyses where characters are weighted according to the consistency indices in the previous analysis. Some have suggested that this approach is circular (e.g., Cannatella and de Queiroz, 1989; Swofford and Olsen, 1990). Others have argued that it is recursive (Carpenter et al., 1993; Carpenter, 1994). Felsenstein (1981a) likens it to compatibility analysis. In such analyses, the investigator attempts to circumvent circularity problems by taking the average consistency index (for example) over all most parsimonious trees rather than from a single topology. This approach is probably the one that is most consistently applied in parsimony analyses that apply successive approximations.

Goloboff's (1997) self-weighted optimization takes a different approach based on Farris' (1969) concave fitting function (Fig. 6.23). In developing successive approximations, Farris (1969) developed a series of functions to illustrate the relationships between the relative weight of a character and its influence on the final tree. For equally weighted characters (in Fitch optimization), the relationship is linear because transformation of both homologous and homoplastic characters equally influences the tree topology. If some characters are weighted more than others, there are two possibilities. If characters with homologous states are weighted more than characters with homoplastic states, the fit function becomes concave. If the opposite obtains, then the fit function becomes convex. Goloboff's (1997) self-weighted optimization is built on his earlier work (Goloboff, 1993) of evaluating

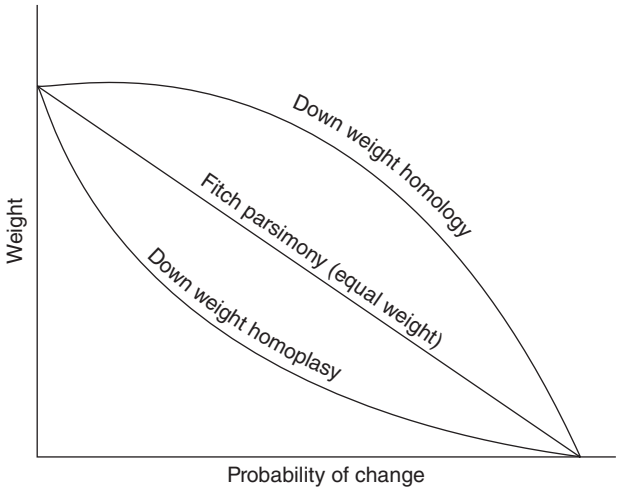


Figure 6.23. Three functions showing the effect of character weighting on the fit of probability of changes on a tree topology (from Farris, 1969).

trees relative to the weight they imply, and the weight of a tree is related to the concavity function.

Consider a matrix with no homoplasy. A tree mapping this matrix would have a linear function, and the lack of homoplasy would mean that there is no distortion. Distortion can be seen as a lack of congruence and the more the homoplasy, the greater the distortion. Now consider a matrix with some level of homoplasy. The linear fit is not perfect, creating distortion. However, if we weight the homoplasies less than the homologies for that particular tree, we minimize the distortion, minimizing the concavity that is created by the presence of homoplasy. The Goloboff (1997) method is to combine tree searches and weighting in an attempt to find the best concavity function for the tree and data as a whole.

Goloboff (1997) reviews criticisms of his technique. The most interesting charge is that the results are not parsimonious (Turner and Zandee, 1995). Goloboff's (1997:236) answer to this criticism is interesting: adopting an optimality criterion based on a concave rather than linear fit function is a "refined way to measure parsimony in trees."

Weighting by Character Elimination

In general, phylogeneticists are loath to eliminate characters once they gather them (which does not mean they are loath to not gather characters if past experience suggests they are bad performers). There are times when weighting by character elimination can be reasonably employed, and this usually involves some understanding of the strength of particular homology statements. For instance, certain regions of DNA such as some loops in 16S mitochondrial ribosomal DNA, introns between coding regions of protein-coding genes, etc. are so entropic (randomly evolving; see Brooks and Wiley, 1986) as to make homology matches impossible or meaningless. Therefore, such regions are often avoided in molecular systematic studies and this is a repeatable and valid weighting criterion.

Weighting: Concluding Remarks

Any form of weighting, including equal weighting, assumes certain things about the evolutionary process. Differential weighting appears to assume more than uniform weighting because it implies that the investigator knows something about the relative behavior of one kind of transformation relative to another kind of transformation. It is a form of model selection, albeit not a statistical form of model selection. We take up the statistical form of model selection in the next chapter.

PHYLOGENETICS WITHOUT TRANSFORMATION?

In the first edition of *Phylogenetics*, Wiley (1981a) took considerable time to discuss two alternative systems of systematic inference: Phenetics and Evolutionary Taxonomy. We largely bypassed that section in this edition because we thought that controversy was resolved. However, a new version of systematics has appeared that competes with Hennig's system that merits a short discussion, the idea that phylogenetic analysis can be performed without the assumption that characters transform

during the course of evolutionary history. This particular idea is not new; it can be traced back to what Platnick (1979) termed the "transformation of cladistics" and was first fully explicated in Nelson and Platnick (1981). What is amazing is that its recent incarnation, exemplified by a recent book by Williams and Ebach (2008), actually accuses those of us who practice traditional Hennigian principles of being "pheneticists."

The method of analysis is usually termed three-taxon analysis (3ta). It attempts to avoid the entire idea of transformation, and thus the entire idea of evolutionary descent, by analyzing presence-absence matrices where the state zero simply means the absence of the state one. Starkly, Scotland provides the following statement that exemplifies the basic idea of what he terms a "complement relationship":

> For example, paired appendages (fins+limbs) constitute a homology at the level of gnathostomes. Within gnathostomes, fins do not form a group and are therefore non-homology. Forelimbs diagnose a group (tetrapods) and are homologous at the level of tetrapods (Scotland, 2000:488).

In opposition to the kind term of *homology proposition* described above are what he terms *paired homologs* entertained by standard cladistic analysis. The unwary may be misled unless they pay careful attention to what Scotland is actually saying. Note that he uses the term *paired appendages* and includes in that category both fins and limbs. This obviates the need for transformation. Did the ancestral species of all other gnathostomes have both fins and limbs? This seems to violate Patterson's (1981) conjunction criterion: angels cannot have both arms and wings if arms and wings are homologous.

So far as we can see, paired homologs are simply hypotheses of transformational homology; for example, hyomandibular and stapes or pectoral fin and front leg. Three strong claims by Scotland are: (1) "Standard cladistic analysis" never tests the transitional proposition represented by "paired homologs." The relationship is "simply assumed" (Nelson, 1994; Pleijel, 1995; and Carine and Scotland, 1999 are cited as the source of this statement). (2) Because characters do not give rise to other characters (Sattler, 1984, 1994), the entire idea of the transformational view is in question. (3) The final claim seems to be that because characters and their states are hypotheses, they cannot be said to have participated in any real processes (Weston, 2000, is cited for this point).

Claim (1) is false when we consider the entire process of analyzing characters. Transitional propositions are tested in many systems and for day-to-day phylogenetic analysis these usually take the form of applying the various criteria discussed by Remane and by Patterson that take the final form of columns of data and the transformational hypotheses they contain (i.e., pairs of plesiomorphic and apomorphic homologs in the binary case). Further, transitional propositions can be refuted after one analysis and before another; that is one part of Hennig's reciprocal illumination idea. It is true enough that the relationship between plesiomorphic and apomorphic homology pairs (or triplets, etc.) are not directly tested during a phylogenetic analysis, but to say that they are not tested at all implies that no thought has gone into gathering empirical data concerning the plausible nature of the relationships before matrix construction and that no thought is given to the results obtained after the analysis. The idea that it is "simply assumed" that there is a his-

torical relationship between pectoral fins and forelegs or between the hyomandibular and staple is, in our opinion, not valid. We note that Scotland (2000) acknowledges what he terms "deep homology," which is simply the idea that it is information in the genome and epigenetic phenomena that are behind the structures we study, but yet he did not get the tetrapod limb correct. Yes, it is true that fin rays are not homologous to autopodial bones, but everyone already knew this, even before the work of Shubin et al. (1997), and the endochondral skeleton of the pectoral girdle includes more than radials, axials, and fin rays. It also includes a scapula and a coracoid and so do tetrapods. Further, it is a matter of rearranging the expression of certain genes that seems to be behind the transformation of fins to limbs, making the transformation understandable.

Claim (2) is dealt with in Chapter 5. But let us expand on it further. It is as nonsensical to claim that "complement relationships" (identity statement in our terms, hypothesizing that similar characters in different organisms are the "same" character) are an illusion as it is to say that paired homologs are an illusion given the reasoning of Sattler (1984, 1994). Why just single out paired homologs? Complement homologs do not give rise to complement homologs any more than plesiomorphies give rise to apomorphies. Instead, as we outline in Chapter 5, information that specifies how to build a complement is passed to each generation and the complement is built anew each generation from the previous generation. The difference between complements and pair homologs is that some of the information has changed during the transmission of that information. Nothing new here unless you wish to expunge all of biology from consideration instead of simply expunging evolution. Change, of course, is nothing but entropy in action (Brooks and Wiley, 1986), and change is as easy as falling off a log. It's stasis that is hard to explain, not change. In short, one can account for neither sameness nor transformation without attending to underlying processes. Far from providing a rationale for rejecting transformational homologies, the observation that characters do not give rise to other characters is cause for rejecting pattern analysis in general.

Claim (3) is the most curious of all. All that we recognize, whether complement or paired, are conjectures about the regularities of the world. Is Scotland arguing that there is no way of studying processes in the world at all? Characters and states are data about the world. All data about the world are hypothesis-bound. Complements are as much data (and identity-statement-theory-laden) as paired homologs. Can we not talk about the process of gravity because our data on falling objects are, in the end, data hypotheses?

Variation on methods of three-taxon analysis have been presented by Nelson and Platnick (1981), Carine and Scotland (1999), Scotland (2000), and Williams and Siebert (2000), as well as Williams and Ebach (2008). Criticisms of three-taxon analysis can be found in Farris and Kluge (1998), Farris et al. (1995), Kluge and Farris (1999), and a recent review of the Williams and Ebach volume by Farris (2010). Williams and Ebach (2008) claim that transformed cladistics and its attendant method, three-taxon analysis is the true phylogenetics, and that everyone not connected to transformed cladistics is actually practicing phenetics (!). We do not consider three-taxon analysis (or pattern analysis in general) a phylogenetic technique and rather than review its procedures in detail refer the reader to the works cited above, both pro and con. Our opinion: while the phylogenetic tent is big enough to include such diverse approaches as parsimony, likelihood, and Bayesian

approaches, it is not big enough to include three-taxon analysis with its need for a priori character ordering, reliance on irreversibility, rejection of reversals as synapomrophies, and other attendant methods and assumptions.

CHAPTER SUMMARY

- Parsimony analysis is performed under the assumption that the best estimate of phylogeny is that tree which is the shortest tree, measured by the number of evolutionary transformations among the characters.
- Phylogenetic analysis may be performed by polarizing the characters a priori and employing rules of character inclusion and exclusion. This is an algorithmic approach and the one used by early phylogenetists to reconstruct phylogenies.
- Computer-assisted phylogenetic analysis may take either the algorithmic or criterion-driven approach, but the criterion-driven approach is usually employed.
- A large number of increasingly sophisticated computer programs are available for parsimony analysis.
- Trees may be evaluated using certain data summaries such as tree length and consistency indices and fit of data to result, including Bremer support, jackknifing, and bootstrapping. Parsimony trees may also be compared using other techniques such as the Wilcoxon signed ranks test.
- Systematics without transformation and outside the evolutionary paradigm is not phylogenetic whatever its other qualities might be.