Common Cladistic Information and its Consensus Representation: Reduced Adams and Reduced Cladistic Consensus Trees and Profiles

Mark Wilkinson

COMMON CLADISTIC INFORMATION AND ITS CONSENSUS REPRESENTATION: REDUCED ADAMS AND REDUCED CLADISTIC CONSENSUS TREES AND PROFILES

MARK WILKINSON

Department of Zoology, University of Glasgow, Glasgow G12 8QQ, Scotland[1]

Abstract.—Consensus trees provide a means of representing cladistic information that is common to a set of fundamental trees. An analysis of the logical relations that pertain between different classes of cladistic information was undertaken. A general class of information, *n*-taxon statements, is recognized as a form of cladistic information in terms of which triplets (three-taxon statements), components, and subset nestings can be described. Disqualifiers (the negations of *n*-taxon statements) comprise an additional class of cladistic information. As is widely recognized, strict and Adams consensus methods have problems of insensitivity and ambiguity, respectively. New consensus methods were developed that solve these problems. The reduced Adams consensus method was developed to facilitate the unambiguous interpretation of Adams consensus trees. The reduced cladistic consensus method provides a solution to the problem of graphically representing all *n*-taxon statements that are common to a set of fundamental trees. A means of representing common disqualifiers using disqualifier-faithful subtrees is also suggested. The utility of consensus methods is discussed in the light of the development of the new methods, including the use of consensus trees in the measure of support and in support-based randomization tests. [Cladistics; information; consensus; phylogeny; randomization; support; bootstrapping.]

Yet the single most important principle in cladistics is that diverse fundamental cladograms may be combined to form a single general cladogram.
—G. Nelson (1979:7)

A common problem in phylogenetic systematics is that the application of a criterion of hypothesis preference, such as maximal clique size or parsimony, may be insufficient to distinguish between alternative phylogenetic hypotheses that meet the particular criterion equally well. Where there are equally well-supported hypotheses, it is desirable to be able to summarize information that is common to all, or to a majority, of the competing hypotheses. The need to summarize information common to a set of cladograms or trees may also arise in more conservative approaches to phylogenetic inference, where both optimal solutions and alternatives that are only slightly suboptimal are considered (e.g., Smith, 1989; Cracraft and Helm-Bychowski, 1991) and where series of phylogenetic hypotheses for the same group are based on the analysis of different sets of data.

Adams (1972) introduced the idea of a consensus tree, formed from a set of trees (termed the fundamental trees), as a representation of information common to the fundamental trees. He also described two consensus techniques. Only the second of Adams's techniques, which is designed for trees with unlabeled internal nodes (i.e., all OTUs are terminal taxa), has been at all widely used and thus "Adams consensus" refers here only to this method (the Adams-2 of Rohlf, 1982). A great variety of consensus methods have since been described and discussed in the literature (Adams, 1972, 1986; Nelson, 1979; Gordon, 1980, 1986, 1987; Margush and McMorris, 1981; Sokal and Rohlf, 1981; Rohlf, 1982; Day, 1983; McMorris et al., 1983; Neumann, 1983; Stinebrickner, 1984; Finden and Gordon, 1985; Mickevich and Platnick, 1989; Page, 1989; Bremer, 1990; Swofford, 1991; Wilkinson, 1992). The most widely used computer packages for parsimony analysis, PAUP (Swofford, 1990) and Hennig86 (Farris, 1988), implement one or more consensus methods.

In this paper, I describe problems associated with Adams and strict consensus trees, which are the most widely employed

[1] E-mail: gbza58@udcf.gla.ac.uk.

consensus methods in systematics. These problems motivated the development of two new consensus methods. The new methods are intended to facilitate the unambiguous representation of cladistic information that is common to a set of fundamental trees. Algorithms for the production of the new kinds of consensus trees are also presented. My prime concern is with phylogenetic inference, and throughout, the examples and terminology used are phylogenetic rather than, for example, biogeographic. Thus, I refer to terminal taxa throughout and use this as a general term for any kind of terminal entity represented in a tree or cladogram.

CLADISTIC INFORMATION

Consensus trees are intended to represent information common to a set of fundamental trees. As we shall see, what kind of information is compared across fundamental trees is an important feature of a consensus method. Here I distinguish among several types of cladistic information and examine their interrelations. Cladistic relationships are usually defined in terms of common ancestry. A set of taxa are more closely related to each other than to any taxa that are not members of the set if they share a more recent common ancestor than any member of the set does with any nonmember. A statement conveys cladistic information precisely when it describes some subset of the possible cladistic relationships among taxa or their common ancestors. That is, information is conveyed when some logically possible relationships are prohibited or conversely when some restricted subset of relationships is asserted.

Components

A common focus of several consensus methods is upon clades or components that are common to fundamental trees. Components (=clades, monophyletic groups, holophyletic groups, clusters) are statements of relationships that apply, through inclusion or exclusion, to all taxa under

consideration. Thus the statement that A, B, and C are a component, symbolically [A,B,C], implies that all other taxa under consideration are not members of that group. The cladistic information conveyed is that A, B, and C are more closely related to each other than any is to any other taxon (e.g., E, F, or G) included in the analysis but not included in the group. Components are the most inclusive unambiguous single statements of cladistic information that might be shared by a set of fundamental trees.

Triplets

Nelson and Platnick (1991) described how a cladistic character can be recoded as a set of three-taxon statements. Similarly, the component information of a cladogram can be expressed as a series of three-taxon statements or triplets (Dobson, 1975; Nelson and Platnick, 1981:238–257; Page, 1988; Nelson and Ladiges, 1991). Triplets correspond to the triads of Adams (1986:302, emphasis in original): "if two leaves, *a* and *b*, separate further from the root than they do from a third leaf, *c*, then we can say that *a* and *b* discriminate against *c*, or the triad $t(a.b.c) = c$." Adams pointed out that triads are analogous to the quartet concept for unrooted trees developed by Estabrook et al. (1985) (see also Estabrook, 1992) and that any rooted tree can be thought of as a set of triads. Given the analogy to the quartet concept, I prefer the term triplet and use this below.

When a cladogram includes more than three terminal taxa, each triplet is less inclusive or complete than the corresponding component statement. For example, the statement A and B are more closely related to each other than either is to D, symbolically (A,B)D, is a triplet that represents some of the information represented by [A,B,C] but that conveys no information concerning the relationships of taxa that are not explicitly included in the triplet, such as C, E, and F. Clearly triplets are ambiguous with respect to the relationships of taxa that are not included in the triplet, but they are unambiguous with re-

gard to relationships among the included taxa.

n-taxon Statements

Both components and triplets can be subsumed under a more general class of cladistic information, of which they represent opposite extremes, that may be called n-taxon statements. An n-taxon statement is the simplest unambiguous descriptive statement that may be true of the cladistic relationships of a set of n taxa postulated on a fundamental cladogram. An n-taxon statement is a comparison of two sets of taxa such that the taxa of the *inside set* are all more closely related to each other than they are to any members of the *outside set.* The terminology refers to the symbolic representation of n-taxon statements. In addition, I refer to the union of inside and outside sets as the *complete set.* Thus, for the n-taxon statement (A,B,C)D,E,F,G, A–G comprise the complete set, A–C comprise the inside set, and D–G comprise the outside set. The sizes of the various sets can be represented by n_i, where n is the cardinality of the statement and i is the cardinality of the inside set. Thus the above example is a 7_3-taxon statement. Triplets are 3_2-taxon statements.

Components are n-taxon statements, where n is equal to all the taxa included on a fundamental tree. Triplets (three-taxon statements) are the least inclusive yet informative n-taxon statements. Where $i <$ 2, the statement conveys no cladistic information. Any informative n-taxon statement where $n > 3$ will imply some set of informative n-taxon statements of lower cardinality. The relationship is symmetrical, so that the latter set of n-taxon statements will jointly imply the former single statement. For example, (A–C)D implies (A,B)D, (A,C)D, and (B,C)D and vice versa. Note also that any pair of these triplets jointly implies the third. In the context of consensus trees, n-taxon statements other than components are of significance because although a component may not be replicated in all (or indeed any) fundamental trees, some subset of the corresponding n-taxon statements of lower cardinality may yet represent cladistic information that is common to the fundamental trees.

Nestings

Adams (1986) described a relation of *subset nesting* that may be common to a set of fundamental trees. If A and B share a node in common that is further from the root than the node including A, B, C, and D, then A and B are a subset nesting of the larger group, symbolically {A,B}C,D. Subset nestings, or just nestings, have their analogues to components, triplets, and other n-taxon statements because they can vary in their inclusiveness. However, if it is true that {A,B}C, then it must also be true that {A,B}C,D, so that any nesting can always be stated in a form that applies to all the taxa included in a fundamental cladogram.

The simplest cladistic interpretation of the nesting {A,B}C,D is that A and B share a more recent common ancestor than do A, B, C, and D. Note that it is never possible for A and B to share a less recent common ancestor than do A, B, C, and D. Nestings can also be interpreted in terms of relative closeness of relationships of nonoverlapping sets of taxa (n-taxon statements) but only through ambiguous conjunctions of them. Any n-taxon statement entails a corresponding nesting. Thus (A,B)C entails {A,B}C and {A,B}C,D, etc. Similarly, (A,B)D also implies {A,B}C,D. Note that different n-taxon statements may imply the same nesting, so that there is a crucial asymmetry in the relation of n-taxon statements and nestings. In other words, a given nesting does not unambiguously entail a specific n-taxon statement. For {A,B}C,D to be true of a particular cladogram, either (A,B)C,D could be true or just one of (A,B)C and (A,B)D: {A,B}C,D = (A,B)C and/or (A,B)D. Any tree can be unambiguously described with a series of components or a series of nestings. However, a single nesting cannot be interpreted unambiguously in terms of taxa that are more closely related to each other than to some other taxa. Considering the information provided by

a single node on a tree, the nesting relation provides a less complete description of this information than the corresponding n-taxon statement. Nestings are less informative than n-taxon statements because they prohibit less.

CONSENSUS METHODS AND THEIR PROBLEMS

Strict Consensus Trees and Sensitivity

The strict consensus (Sokal and Rohlf, 1981) is the most widely used consensus method in systematics. It includes only those components that are replicated in all fundamental cladograms. Other components are collapsed to polytomies. Margush and McMorris (1981) developed the majority-rule consensus method, which like the strict includes only replicated components. As the name suggests, this method has the less stringent requirement that components only be replicated across a majority of fundamental trees to be included in the consensus. A further elaboration was suggested by Bremer (1990). Like the strict consensus, his combinable component (or semistrict) consensus includes all components that are replicated in all fundamental trees. In addition, components that appear only in some subset of the fundamental trees will also appear in the consensus if they are uncontradicted in those fundamental trees that do not include them. In the latter trees, the component must be a possible resolution of an incompletely resolved (polytomous) tree. My focus here is upon the strict consensus method, but my comments also apply to other consensus methods that are based upon replicated components.

Consider the two fundamental trees in Figures 1a and 1b. There is no component that is present in both fundamental trees. Consequently, the strict consensus is an unresolved bush (Fig. 1e). However, the two fundamental trees do share some cladistic information, namely the triplet (A,B)C. Because the common cladistic information of the fundamental trees in Figures 1a and 1b is not component information, it is not represented in the strict consensus. Insensitivity of the strict consensus to some common cladistic information has been noted previously by Adams (1986), Mickevich and Platnick (1989), and Funk and Brooks (1990). Swofford (1991) considered this insensitivity to be a serious liability and noted that it has led to the view that the strict consensus is too strict. In the development of the reduced consensus methods introduced in this paper, an emphasis has been placed upon the desideratum of greater sensitivity to common cladistic information (expressed as n-taxon statements of lower cardinality than components) than that shown by component consensus methods.

Adams Consensus, Ambiguity, and Error

The Adams consensus technique also compares replicated components, but where these groupings differ it bases the consensus upon the intersections (rather than unions) of the sets of taxa that are being compared. As discussed by Neumann (1983), the Adams consensus is one of a family of possible consensus methods based on intersection rules, of which he provided several other examples. An Adams consensus tree includes all those nestings that are common to a set of fundamental trees (Adams, 1986). The method parallels the strict consensus in the requirement of agreement among all fundamental trees, but semistrict and majority-rule varieties of Adams consensus trees are also conceivable.

The Adams consensus is sensitive to common cladistic information other than replicated components inasmuch as other common cladistic information will affect the topology of the Adams consensus. For example, the Adams consensus (Fig. 1f) for the two fundamental trees in Figures 1a and 1b clearly contains more information than does the corresponding strict consensus (Fig. 1e). However, care is required in the interpretation of the Adams consensus. The grouping of A and B in the Adams consensus reflects the fact that {A,B}C,D is the only nesting common to both the fundamental trees. In this case the nesting is implied by the common triplet (A,B)C. However, the same nesting, and thus the

same Adams consensus tree, can result from different relationships that are common to the fundamental trees. Thus the two fundamental trees in Figures 1c and 1d share the different triplet (A,B)D but have the same Adams consensus (Fig. 1f).

Any n-taxon statement, including components, entails some nesting. If any n-taxon statement is common to a set of fundamental trees then, because this implies a common nesting and because the Adams consensus includes all common nesting, the n-taxon statement will be reflected in the Adams consensus tree. This fact gives the Adams consensus the useful property of displaying topological sensitivity to common cladistic information represented by replicated n-taxon statements of lower cardinality than components. However, because different n-taxon statements may imply the same nesting, polytomies in Adams consensus trees are necessarily ambiguous with respect to common n-taxon statements of the fundamental trees. Whether a particular n-taxon statement that might be reflected in an Adams consensus tree is true of all fundamental trees cannot be determined from an inspection of the consensus tree alone. Nestings may be true, and thus appear in the Adams consensus, when no corresponding n-taxon statement is true of all fundamental trees. For example, if the trees in Figures 1a and 1c are the fundamental trees, then {A,B}C,D is true for both and is reflected in the Adams consensus (Fig. 1f), although there are no common n-taxon statements.

Interpreting Polytomies

Nelson and Platnick (1980, 1981:257–265) discussed polytomies in cladograms, although not specifically in the context of consensus trees. They distinguished two possible interpretations of nonterminal polytomies, which they described in relation to a cladogram with the topology of the Adams consensus in Figure 1f. Interpretation 1 is that "taxa A and B are more closely related to each other than either is to taxa C *and* D" (1980:87, emphasis in original), i.e., (A,B)C,D. This interpretation holds for polytomies in strict consensus
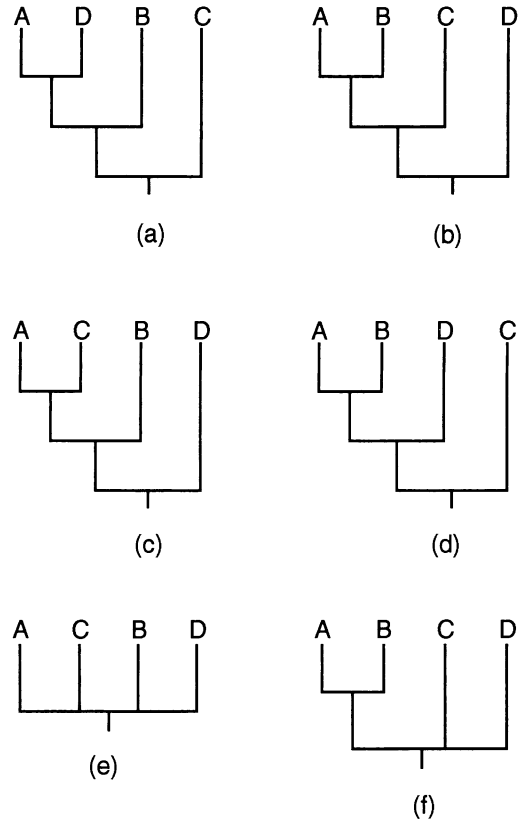

FIGURE 1. Four fundamental cladograms (a–d), the strict consensus tree (e), and the Adams consensus tree (f).

trees, and as such the interpretation is true also of relationships in each of the fundamental trees. It is also the cladistic interpretation of dichotomies.

Under interpretation 2, the cladogram indicates that "taxa A and B are more closely related to each other than either of them is to taxon C *and/or* taxon D" (1980:87–88, emphasis in original). This is the n-taxon statement interpretation of the nesting relation {A,B}C,D described by Adams (1986). Nelson and Platnick drew attention to the relation between this interpretation and (Adams) consensus trees. When the tree is an Adams consensus, interpretation 2 can itself be interpreted in two ways. The combined statement (A,B)C and/or (A,B)D might be thought to extend across all of the fundamental trees so that either (A,B)C

is true of all fundamental trees and/or (A,B)D is true of all fundamental trees collectively (interpretation 2a). Alternatively, it might be taken to imply that each fundamental tree individually satisfies the condition (A,B)C and/or (A,B)D (interpretation 2b). Of these, only 2b is a legitimate interpretation of the Adams consensus. Each of the fundamental cladograms may satisfy the condition (A,B)C and/or (A,B)D in a different way (e.g., in some (A,B)C is true and in others (A,B)D is true, but neither is true of every fundamental tree, as in Figs. 1a, 1c). It follows that the application of interpretations 1 and 2a to Adams consensus trees may lead to erroneous conclusions.

Terminal polytomies can also be interpreted in different ways. They may be taken to represent a lack of resolution or as a statement about multiple speciation (no subset of the included taxa being more closely related to each other than to the other taxa). These are soft and hard interpretations, respectively (Maddison, 1989). I assume the first of these interpretations for the purposes of this paper.

A Three-way Trade-off

Consider again the pairs of fundamental trees given in Figure 1. One pair (Figs. 1a, 1b) share the common triplet (A,B)C, but no informative n-taxon statements can be formulated regarding the relations of D. A second pair (Figs. 1c, 1d) share the common triplet (A,B)D, but no informative n-taxon statement can be formulated describing the relations of C. A third pair (Figs. 1a, 1c) share no informative n-taxon statements. All three pairs of trees have the same Adams consensus tree (Fig. 1f), and in each case the strict consensus (Fig. 1e) is completely unresolved. Interpreted correctly, the polytomy in the Adams consensus is ambiguous as to which relations actually pertain in the fundamental trees. If the ambiguity is ignored and interpretation 1 of Nelson and Platnick (1980) is applied to the Adams consensus tree, then it will be taken to imply the n-taxon statement (A,B)C,D, which is not true for any of the pairs of fundamental trees. If interpreta-

tion 2a is applied, then the Adams consensus is taken to imply (A,B)C or (A,B)D, or both. This is true for the first two pairs of fundamental trees but not for the third pair. This third pair of trees share no n-taxon statements but do share the common nesting {A,B}C,D of the Adams consensus, which highlights the difficulty of interpreting the polytomies of Adams consensus trees in terms of the closeness of relationships between taxa.

For each pair of fundamental trees there is simply no way of representing all and only those n-taxon statements that are true for all the fundamental trees on a single consensus tree that includes all the taxa. Thus with the first pair, no n-taxon statement can be formulated for taxon D that is true of both the fundamental trees. Placement of this taxon in any consensus tree leads to the three potential pitfalls: (1) ambiguity of interpretation (Adams consensus trees under interpretations 2a and 2b), (2) possible loss of information represented by true n-taxon statements that are not component statements (strict consensus trees), and (3) erroneous implications of relationships (Adams consensus under interpretations 1 or 2a). Strict consensus trees are unambiguous and contain no errors, but they achieve this at the cost of insensitivity. Adams consensus trees achieve greater sensitivity but only at the expense of either ambiguity or potential error, depending upon their interpretation. The ambiguity of the Adams consensus represents an alternative form of insensitivity to common n-taxon statements because its topological sensitivity is not accompanied by an interpretive sensitivity.

The reduced consensus methods described below are intended to achieve a greater sensitivity to cladistic information common to a set of fundamental trees, as compared with the strict consensus, without any of the ambiguities or errors of interpretation that afflict Adams consensus trees. These methods avoid the trade-off among ambiguity, error, and sensitivity by allowing consensus trees to represent relationships among a subset of the taxa included in the fundamental trees. The cost

is ambiguity concerning the relationships of taxa that are not included in the reduced consensus. An important consequence of this approach is that there may not be a unique reduced consensus for a set of fundamental trees. However, appropriate selection of a set or *profile* of reduced consensus trees provides a means of avoiding the ambiguity that results from the exclusion of taxa. Thus, in contrast to Nelson (1979), a multiplicity of different trees may be required to represent cladistic information that is common to diverse fundamental cladograms and that cannot be combined into a single consensus or general cladogram without ambiguity, error, or loss of information.

REDUCED ADAMS CONSENSUS TREES AND PROFILES

One way of avoiding the ambiguity inherent in the polytomies of Adams consensus trees is to selectively remove (prune) branches that contribute to those polytomies. If a tree can be formed by the removal of branches from another tree then the former is a *subtree* of the latter. Similarly, if a tree can be formed by the removal only of branches that contribute to polytomies in another tree then the former may be termed a *polytomy subtree* of the latter. I adopt the conventions that any tree is a subtree of itself and that any tree containing polytomies is a polytomy subtree of itself. A reduced Adams consensus (RAC) tree is defined as a polytomy subtree of an Adams consensus that satisfies the three conditions: (1) unambiguity—all implied relationships are true for all fundamental trees under interpretation 1 of Nelson and Platnick (1980) (simply referred to as *true* hereafter); (2) nonredundancy—it is not a subtree of some other tree satisfying condition 1; and (3) informativeness—at least one informative n-taxon statement is included in the tree.

RAC trees are thus Adams consensus trees that have been stripped of their ambiguity by the selective pruning of branches that imply relationships, under interpretation 1, that are not true. An important

consequence of the desire for unambiguity (condition 1) is that this condition may be satisfied in more than one way, leading to a collection of consensus trees rather than a single consensus tree. I shall call the set of RAC trees the RAC profile.

Conditions 2 and 3 relate to the composition of the RAC profile. For example, if we apply interpretation 1 to all the relationships indicated by the branching of an Adams consensus tree and find that these are true for all fundamental trees, then there is no need to remove any branches from the Adams consensus to achieve unambiguity (condition 1). The corresponding single RAC tree will therefore be topologically identical to the Adams consensus tree. However, it will differ in its information content because the Adams consensus allows only ambiguous interpretation of polytomies, whereas the RAC allows interpretation 1. Any polytomy subtree of the Adams consensus tree would also satisfy condition 1. However, such a subtree would not satisfy condition 2 and would not qualify as an RAC tree. Furthermore, such a subtree would convey only information that is conveyed by the RAC tree of which it is a subtree. Thus condition 2 serves to eliminate redundancy. Condition 3 disqualifies uninformative trees from the RAC profile because they are of no interest. The different RAC trees comprising the RAC profile will each include one or more n-taxon statements that are not implied by any other RAC trees. (An algorithm for the construction of RAC profiles is described and illustrated in the Appendix.)

Profiles and Primary Trees

An RAC profile for a pair of fundamental trees, together with the strict and Adams consensus trees, is given in Figure 2. Individually, each of the three RAC trees (Figs. 2e–g) included in the RAC profile includes more cladistic information that is true of the fundamental trees than does the corresponding strict consensus tree (Fig. 2d). As a whole, the profile includes far more information. The ambiguity of polytomies suffered by the Adams consensus
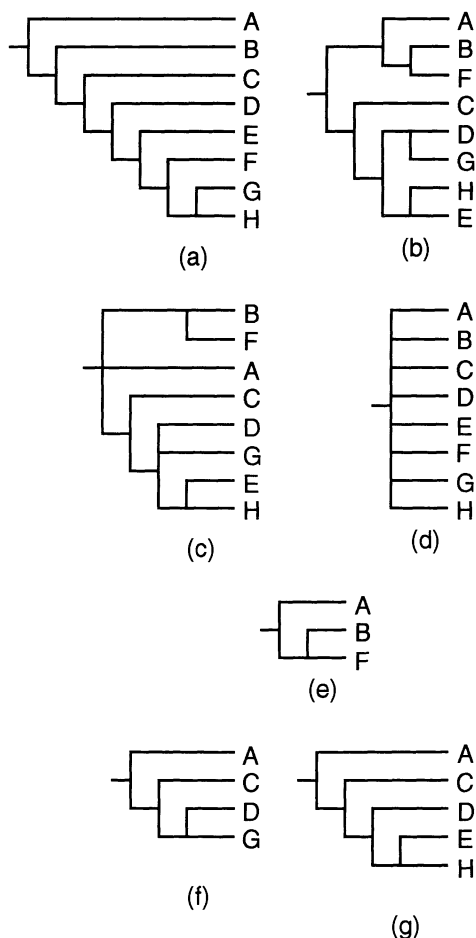
FIGURE 2. Two fundamental cladograms (a, b), the Adams consensus tree (c), the strict consensus tree (d), and the three RAC trees (e–g) that constitute the RAC profile. [Modified and extended from Adams, 1972.]

mining the RAC profile is a potentially useful way of summarizing and representing n-taxon statements that are common to a set of fundamental trees with greater sensitivity than the strict consensus method. The profile can be used alone or simply as an aid to the interpretation of an Adams consensus tree including all the taxa.

RAC trees may differ both in the number of nonredundant n-taxon statements that they include (one for each nonbasal node) and in the cardinality of the included n-taxon statements (the number of included terminal taxa), and they can be ranked according to these features. With many RAC profiles, these rankings may be expected to be perfectly correlated. In general, as the number of nodes increases the minimum number of included taxa also increases. However, when RAC trees include polytomies the tree with the most nodes may not also be the tree with the greatest number of included taxa. I define primary RAC trees as those RAC trees that (1) incorporate the greatest number of nodes and secondarily (2) incorporate the greatest number of terminal taxa, subject to condition 1. There may be more than one primary RAC tree. For the above example, there is a single primary RAC tree, that given in Figure 2g, which includes both the largest number of taxa and the largest number of included nodes. RAC trees that are not primary trees are termed secondary.

Further Examples

The utility and properties of the RAC method are also illustrated by the RAC trees (Fig. 3) for the pairs of fundamental trees in Figure 1 and a comparison with the corresponding strict (Fig. 1e) and Adams (Fig. 1f) consensus trees. The RAC profile for the fundamental trees in Figures 1a and 1b comprises only the single RAC tree in Figure 3a. Similarly, Figure 3b is the single RAC tree for the fundamental trees in Figures 1c and 1d. For both sets of fundamental trees, the RAC tree unambiguously represents the single n-taxon statement that is true of the fundamental trees. The strict

(Fig. 2c) is greatly clarified by the RAC profile. For example, the basal polytomy does not allow the conclusion that (C,D,E,G,H)A is true of all fundamental trees, although it is. The Adams consensus merely allows the conclusion that this, along with other interpretations, might be true of the fundamental trees. This n-taxon statement does not appear on any single RAC tree, but it is entailed by the conjunction of the last two. If (C,D,G)A (Fig. 2f) and (C,D,H,E)A (Fig. 2g) are true, then (C,D,G,H,E)A must also be true.

The example demonstrates that deter-

(Fig. 1e) and Adams (Fig. 1f) consensus trees are identical for both these pairs of fundamental trees despite the difference in replicated cladistic information, and they either fail to represent this information or are ambiguous. The tree in Figure 3c is the single RAC tree for the pair of fundamental trees in Figures 1b and 1d and also the strict and Adams consensus trees. In this case, there is a replicated component [A,B] —equivalent to the *n*-taxon statement (A,B)C,D—that appears in all three types of consensus tree. The single RAC tree is equivalent in both topology and interpretation to the strict consensus and in topology but not interpretation to the Adams consensus.

The RAC profile for the fundamental trees in Figures 1a and 1c is empty. These fundamental trees share no *n*-taxon statements, although they share a common nesting, and thus there can be no tree that satisfies condition 3 of the definition of an RAC tree. Thus even when the RAC profile is empty, firmer conclusions may be drawn about the common cladistic information represented in the Adams consensus tree.

REDUCED CLADISTIC CONSENSUS TREES AND PROFILES

Not all *n*-taxon statements that are true of a set of fundamental trees need be implied, either individually or conjointly, by the trees comprising the RAC profile. For example, the *n*-taxon statements (C,D,E,F, G,H)B and (G,H)F are true for the fundamental trees in Figures 2a and 2b but are not implied by the corresponding RAC profile. The reduced cladistic consensus concept is a method that will produce reduced consensus trees that include all shared *n*-taxon statements and are also unambiguous and subject to interpretation 1 without any danger of erroneous interpretation.

A reduced cladistic consensus (RCC) tree is defined as a tree that satisfies the three conditions described for the RAC trees but that is not necessarily a polytomy subtree of an Adams consensus. Clearly, if an RCC tree is a polytomy subtree of an Adams consensus then it will also be an RAC tree,
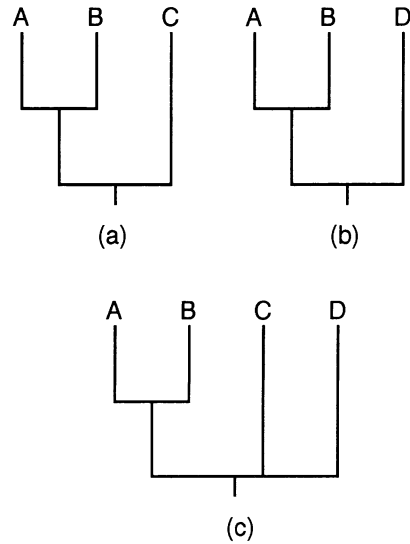

(a)   (b)

(c)

FIGURE 3.   Three RAC (and reduced cladistic consensus [RCC]) trees for pairs of fundamental trees in Figure 1. Each tree is the sole member of the RAC profile and is thus also the primary RAC (and RCC) tree.

but not all RAC trees are also RCC trees and vice versa. (An algorithm for the construction of RCC profiles is described and illustrated in the Appendix.) The RCC profile for the two fundamental trees (Figs. 2a and 2b) is given in Figure 4.

Comparison of RAC and RCC Trees and Profiles

The RCC profile (Fig. 4) can be contrasted with the RAC profile (Figs. 2e–g) and the corresponding strict and Adams consensus trees (Figs. 2c, 2d) for the same pair of fundamental trees (Figs. 2a, 2b). The RCC and RAC profiles differ most obviously in the number of included trees. There is one more RCC tree. Two of the RCC trees correspond closely to particular RAC trees (Fig. 4a to Fig. 2f; Fig. 4b to Fig. 2g). The smallest RAC and RCC trees, those in Figures 4d and 2e, are identical. The additional RCC tree (Fig. 4c) includes the triplet (G,H)F that was identified earlier as an element of common cladistic information that was not represented in the RAC profile. The other information conveyed by this RCC tree, namely (G,H)A,B,C, is also provided by a
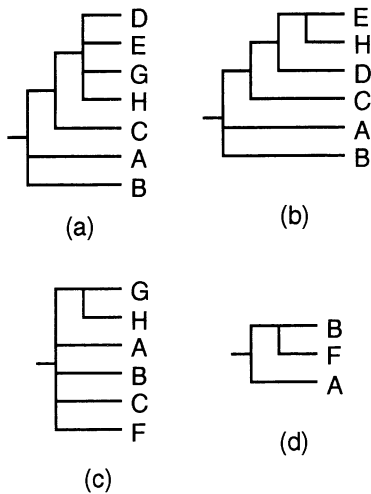
FIGURE 4. The RCC profile for the fundamental trees in Figures 2a and 2b.

different RCC tree (Fig. 4a) and is thus redundant. The RCC trees in Figures 4a and 4b differ from the corresponding RAC trees in the inclusion of extra taxa but not of extra nodes. They both convey extra information, e.g., (C,D,E,G,H)B, that is absent from the RAC profile. The statement (D,E,G,H)C is provided by the RAC profile only by the joint implication of RAC trees, whereas in the RCC profile this is indicated by a single tree (Fig. 4a) and is thus represented more directly.

The RCC profile provides a graphical solution to the problem of representing all n-taxon statements that are true of the fundamental trees. Furthermore, this solution does not depend on some statements being entailed by a combination of RCC trees; each n-taxon statement is represented directly on one or more of the RCC trees. In these respects, the RCC method is clearly superior to the RAC method. Compared with the strict and Adams consensus trees (Fig. 2c, 2d), the additional sensitivity provided by the RCC method is quite dramatic. The greater potential information content of the RCC profile over the RAC profile is achieved at the expense of potentially greater redundancy of information content in the RCC trees.

As with the RAC method, primary RCC trees can be defined on the basis of the number of nodes and number of included taxa. In this example the correlation between numbers of nodes and taxa breaks down. The primary RCC tree (Fig. 4b) has the largest number of nodes but has fewer taxa than the RCC tree in Figure 4a. For the various pairs of fundamental trees in Figure 1, the RCC profiles are identical to the RAC profiles given in Figure 3. Generally, we might expect RAC profiles to closely approximate RCC profiles in their information content, but we cannot be confident that they will capture all the information present in the RCC profile.

RAC trees have the property of representing all n-taxon statements that are true of just the taxa included in the particular RAC tree. This is true also for RCC trees when they include no polytomies, but where polytomies are included this need not hold. Thus, for example, (E,H)D is true of the fundamental trees (Figs. 2a, 2b) but is not indicated on the RCC tree in Figure 4a despite the inclusion of these three taxa. This triplet cannot be represented on any tree that includes G without erroneous implications under interpretation 1 of the relationships of G to D, E, and H. No informative n-taxon statement is true of just D, E, G, and H. Hence, in the RCC profile the triplet (E,H)D must be represented on an RCC tree that excludes G (Fig. 4b). Thus, where n is equal to the number of included taxa, an RCC tree includes all true n-taxon statements for the included taxa. However, some n-taxon statements of lower cardinality might also be true of some subset of these taxa but not be included in the RCC tree. When the number of taxa included in an RCC tree is equal to the number included in the fundamental trees (i.e., no taxa are excluded), the RCC tree will represent all true components and is equivalent to the strict consensus tree. Therefore, if the strict consensus tree is not completely unresolved, it must also be a member of the RCC profile.

NEGATIVE N-TAXON STATEMENTS AND THEIR REPRESENTATION

I return now to the subject of cladistic information and the ways in which a set of fundamental trees might agree. Con-

sider the negation of the triplet (A,B)C. This negation states that A and B are not more closely related to each other than they are to C, symbolically not(A,B)C. Any n-taxon statement has its negation, and it is useful to distinguish between such positive and negative cladistic information. Here I restrict the term "n-taxon statement" to positive cladistic information. The negation of an n-taxon statement shall be called a *disqualifier.*

Any positive n-taxon statement entails some set of disqualifiers. Thus (A,B)C implies not(A,C)B and not(B,C)A. The relationship is symmetrical only if the tree is dichotomous, because in this case one of (A,B)C, (A,C)B, and (B,C)A must be true. On a polytomous tree, there are soft and hard interpretations of disqualifiers. The soft interpretation of a polytomy including A, B, and C is that there are no disqualifiers. The hard interpretation is that the three possible disqualifiers, not(A,B)C, not(A,C)B, and not(B,C)A, are all implied.

The disqualifier not(A,B)C also implies not(A,B)C,D, so that, like nestings, any common disqualifier can always be expressed with a cardinality equal to the number of taxa included in the fundamental trees. Also, like nestings, a single disqualifier does not imply any positive n-taxon statement. Thus the disqualifier not(A,B)C implies (A,C)B or (B,C)A (dichotomous trees) or in addition that there is no relationship between A, B, and C (hard interpretation of a polytomy). Consequently, a disqualifier may be true of a set of fundamental trees, although no n-taxon statement that would have implied it is itself true. An example of this is provided by the pair of fundamental trees in Figures 1a and 1c. No positive n-taxon statements are true for this pair of fundamental trees, and thus the RCC profile is empty. However, the disqualifier not(B,D)A,C is true of both fundamental trees. Thus the fundamental trees agree in some aspect of their cladistic information content that is not reflected in either the RAC or the RCC profile, which are both empty. Similarly, this information is not represented in either the strict (Fig. 1e) or Adams (Fig. 1f) consensus trees. I have no suggestion as to how

disqualifiers might be determined or represented in general, but here I examine a method that uses primary RAC or RCC trees, or subtrees produced from them, as a framework for the representation of common disqualifiers.

Suppose that the primary RCC (or RAC) tree is fully dichotomous. In such a tree, the relationships of the included taxa are fully resolved. As a consequence, all disqualifiers that are true of just the included taxa will also be represented in the tree. Thus the only true disqualifiers that pertain to any of the included taxa, and that are not represented in the primary tree, are those that describe relationships among taxa that are included and those that are not. Such disqualifiers not reflected in the primary RCC tree must be either those that describe the possible positions of taxa that are excluded from the primary tree relative to those that are included or those that only describe relationships among taxa that are not included in the primary RCC tree. A fully dichotomous RCC or RAC tree can be used as a framework upon which descriptions of the possible positions of the excluded taxa can be developed. The framework usefully summarizes those disqualifiers that are true of just the taxa included in the consensus. Description of possible positions is facilitated by providing the reference RCC tree with numbered internal nodes so that specific branches can be readily identified.

A simple example illustrating this kind of representation is given in Figure 5, which includes five fundamental trees (Figs. 5a–e), the Adams (Fig. 5f) and strict (Fig. 5g) consensus trees, and the RAC (Figs. 5h, 5i) and RCC (Figs. 5g, 5h) profiles. The primary RAC tree is also the primary RCC tree in this example and is provided with numbered nodes. The fundamental cladograms have a common set of relationships among taxa A–E but vary dramatically in the position of X. The trees portray a result that is encountered (but not exclusively) when taxa, such as incomplete fossils, have characters that specify some of the limits of their relationships but are missing entries for characters that would help resolve their more precise affinities within these
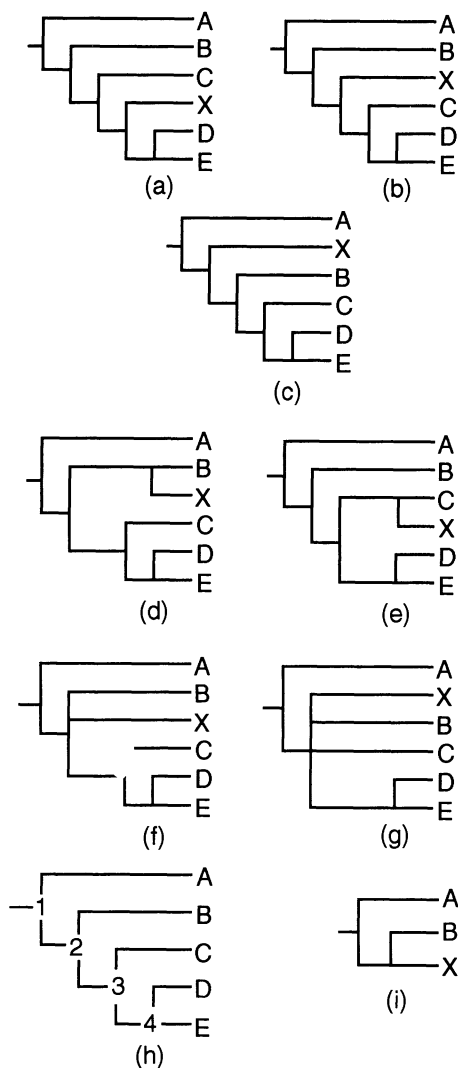
FIGURE 5. Five fundamental trees (a–e); the Adams consensus tree (f); the strict consensus tree (g), which is also the secondary RCC tree; the primary RAC tree (h), which is also the primary RCC tree; and the other member of the RAC profile (i).

the following conventions: (1) above a node—positions on any branch that is a descendant of that node, (2) below a node—positions on any branch that is not a descendant of that node, and (3) between nodes—positions upon the path that links the nodes. One of the disqualifiers, *not*(A,X)B–E, is implied by the conjunction of the two RCC trees and specifies a position above node 2 or between nodes 1 and 2. The second, *not*(D,E,X)A–C, is not represented in the RCC profile and specifies a position below node 4. These possible positions are exactly those in which excluded taxa appear in one or more of the fundamental trees.

Appending a list of possible positions of excluded taxa to a fully dichotomous primary RCC or RAC tree provides one means of representing disqualifiers that are not reflected directly in any consensus tree. In the above example, we can identify an increasing representation of common cladistic information that runs from the strict consensus tree to the RAC profile to the RCC profile and reaches its most complete state with the primary RAC/RCC tree and an appended list of possible positions of X. With this example, I can identify no cladistic information that is replicated in all the fundamental trees but is not represented using the last method. I have not incorporated the Adams consensus tree into this sequence because its fundamental ambiguity complicates any evaluation of its information content.

In many cases, the primary RCC or RAC tree will include polytomies. The inclusion of polytomies means that not all disqualifiers that are true of just the included taxa will necessarily be included in the consensus. This is a limitation of primary RCC trees because they may not include all true *n*-taxon statements (and thus implied disqualifiers) that describe relationships among the included taxa and are of lower cardinality than the number of taxa included in the consensus. It is true also of primary RAC trees, although these do include all true *n*-taxon statements that apply to just the included taxa. Thus, the primary RAC tree (Fig. 3c) for the two fundamental trees in Figures 1b and 1d includes both

limits (pers. obs.). The common relationships are captured by the primary RCC tree, which excludes only the "fossil" taxon X, whose position changes. Two disqualifiers are true of taxon X, *not*(A,X)B–E and *not*(D,E,X)A–C. These disqualifiers set limits upon the possible positions of X relative to the taxa included in the primary RCC tree.

To describe possible positions, I adopt

the true triplets (A,B)C and (A,B)D but does not include the common disqualifier *not*(C,D)A,B.

Polytomies may be present in any consensus tree for one of two reasons: either the polytomy is present in all of the fundamental trees (i.e., the fundamental trees agree) or the polytomy reflects disagreement among the fundamental trees. When a polytomy occurs in a fundamental tree (that is not also a consensus tree) then there are no disqualifiers regarding the relationships of the groups contributing to the polytomy (soft interpretation). Thus, none of the relationships implied by various possible arbitrary resolutions of the polytomy are prohibited. If the polytomy is true of all fundamental trees then the same soft interpretation could be given to the polytomy if it also appears in a consensus tree. I shall call this the disqualifier-faithful interpretation of a consensus polytomy.

In contrast, when a polytomy appears in a consensus tree because of disagreement among fundamental trees then, as the above consideration of the polytomies of RCC and RAC trees showed, it cannot be concluded that there are no disqualifiers that prohibit some relationships among the groups contributing to the polytomy and that are true of all the fundamental trees. The interpretation of the polytomy must be disqualifier neutral, in the sense that it does not prohibit any possible disqualifiers. In the absence of knowledge that a consensus polytomy is due to agreement or disagreement among a set of fundamental trees, only the disqualifier-neutral interpretation of the consensus polytomy is necessarily safe from possible error.

Just as a fully dichotomous RCC or RAC tree represents all disqualifiers that are true of just the included taxa, so will an RCC or RAC tree that includes polytomies, provided that these polytomies are ones that appear in all the fundamental trees. Furthermore, RCC or RAC trees that include one or more polytomies that do not satisfy the above criterion (i.e., they are not true of all fundamental trees) can be converted into trees that do satisfy this criterion through the elimination of some of the taxa that contribute to polytomies. The result

of such elimination will be either a tree including polytomies that can be interpreted as disqualifier faithful or a fully dichotomous tree. In either case, the resulting tree will accurately summarize both all n-taxon statements and all disqualifiers that are true of just the included taxa.

I define a disqualifier-faithful subtree (DFS) of an RAC as a polytomy subtree of the RAC that (1) summarizes all disqualifiers that are true of just the included taxa and (2) is not a subtree of any other tree satisfying condition 1. A parallel definition can be produced for a DFS of an RCC tree by substituting RCC for RAC wherever it occurs in the above definition. When the RAC or RCC tree includes no polytomies or only includes polytomies that are present in all fundamental trees, it will also be the unique corresponding DFS. When RAC or RCC trees are not also DFSs, they can be converted into corresponding DFSs by the selective removal of branches that contribute to polytomies. In such cases, there may not be a unique DFS.

The utility of DFSs, especially those that can be formed from primary RAC or RCC trees, is that they provide a suitable framework for the description of the possible positions of excluded taxa. The hypothetical example described above illustrates the potential of the DFS approach in cases where many equally parsimonious trees result from the inclusion of poorly known fossil taxa. In such cases, groups that would constitute replicated components, were it not for the uncertain inclusion of poorly known fossils, cannot be discovered using the strict consensus method or unambiguously identified from an Adams consensus tree. The application of this approach to real data sets including poorly known fossils will be discussed elsewhere. Some suggestions regarding the construction of DFSs and their use in determining possible positions are given in the Appendix.

DISCUSSION

Adams Consensus and Triplet Consensus Trees

Adams (1986) discussed the possibility of constructing a consensus tree from all
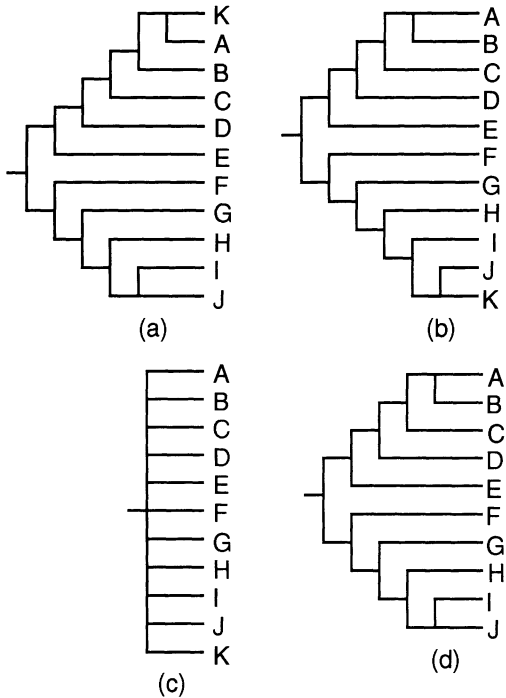
FIGURE 6. Two fundamental trees (a, b), the strict consensus tree (c), and the primary RCC tree (d). [Modified from Adams, 1986.]

triplets or three-taxon statements, such as (A,B)C, that are common to a set of fundamental trees. Given that all n-taxon statements can be represented as a series of triplets, any triplet-based consensus method can be expected to resemble the RCC method described in this paper. Adams gave an example (1986: fig. 1) (presented here in Fig. 6) that emphasizes the connection. In this example, a tree (Fig. 6d) that is built up of all the triplets that are common to two fundamental trees (Figs. 6a, 6b) is contrasted with the strict consensus tree, which shows no resolution (Fig. 6c). The purpose of the example was to demonstrate the shortcomings of cluster- or component-based consensus methods, especially the strict consensus, because there is evidently much common structure in the two fundamental trees. The triplet-based tree does not include taxon K, whose position changes in the otherwise similar

fundamental trees. Adams suggested that devising a rationale for choosing the placement of K would yield a definition for a consensus tree, but he did not discuss this problem further. Instead, his consideration of subset nesting led him to conclude that there are elements of structure (nestings) that are common to fundamental trees and that may be missed by the triplet approach.

The triplet-based tree of Figure 6d is both the unique RCC and unique RAC tree for the two fundamental trees. For Adams, a consensus tree is one that includes all the taxa. However, to satisfy the desiderata of faithfulness to common cladistic information and avoidance of ambiguity, as embodied in the polytomies of Adams consensus trees, this requirement must be relaxed. In contrast to this example, it will not always be possible to represent all triplets that are true on a single unambiguous consensus tree. When this holds, then each RAC or RCC tree in their respective profiles will summarize a subset of the common triplets.

Adams (1986) also defined the full consensus as containing all the nesting information common to the fundamental trees. He then proved that the Adams consensus is the same as the full consensus. I agree with Adams (1986) that nesting is information that may be missed by some consensus methods. Furthermore, this information has a cladistic interpretation in terms of recency of common ancestry. However, as we have seen, the information is ambiguous in terms of relative closeness of relationships (n-taxon statements) and so cannot be represented on any tree that is unambiguous with respect to closeness of relationships, unless it also happens to be entailed by the unambiguous relationships specified in that tree. I suggest that the information content of Adams consensus trees is not as complete as its status as a "full" consensus implies. The RAC profile provides the extra information required to allow the Adams consensus tree to be interpreted precisely and without error. However, even when the Adams consensus is interpreted in the light of an RAC

profile, not all true n-taxon statements or disqualifiers will necessarily be represented.

Common Pruned Trees

Gordon (1980) developed a consensus approach that focuses upon largest common pruned trees. When two or more fundamental trees differ, they can be brought into agreement by the elimination of those taxa that have variable positions. The tree that remains after such elimination is a common pruned tree. Common pruned trees are thus common subtrees. Typically, there will be a multiplicity of common pruned trees for a set of fundamental trees corresponding to the elimination of different sets of taxa. Gordon (1980) suggested preference for the largest common pruned (LCP) tree, i.e., that incorporating the largest number of taxa. Exact and heuristic algorithms for discovering the LCP trees have been discussed and described by Finden and Gordon (1985) and implemented in software by Finden (1983, 1984a, 1984b, 1984c).

Common pruned trees have not received the attention that they deserve from systematic biologists. I am unaware of LCP trees ever having been used explicitly as a consensus in any phylogenetic study, although they have been used rather informally (e.g., the reduced area cladograms of Rosen, 1978). Indeed, my awareness of common pruned trees is due to their inclusion by Swofford (1991) in a review of consensus methods that I encountered subsequent to the development of the reduced consensus methods presented in this paper.

In some respects RAC and RCC trees and common pruned trees are fundamentally similar. Taxa included in the fundamental trees may be selectively eliminated from both, and both are unambiguous, the common pruned trees because they include no polytomies that are not themselves present in all the fundamental trees. In many simple cases, the LCP tree is identical to the primary RAC or RCC trees. For example, the primary RAC/RCC tree of Figure 6d is

also the LCP tree. However, the methods are distinct. RAC and RCC trees may include polytomies that are not part of any fundamental tree, so that an RAC tree need not also be a common pruned tree (e.g., the RAC tree in Fig. 3c). Gordon (1980) also discussed ways of regrafting pruned taxa onto a common pruned tree so as to produce a single consensus tree including all the taxa. Likewise, RCC and RAC trees could be formed by regrafting of pruned taxa onto a common pruned tree. RCC trees can thus be thought of as particular kinds of common pruned trees that may have additional taxa regrafted onto them under the constraints (1) of fidelity under interpretation 1 and (2) that they are not redundant subtrees of some other tree that satisfies the first constraint. These constraints clearly define the limits and content of the RCC. An additional constraint for RAC trees is that the regrafting of taxa must yield a polytomy subtree of an Adams consensus.

Primary RAC/RCC trees are selected from the corresponding profile first on the basis of numbers of nodes and only secondarily on the numbers of included taxa. These selection criteria also distinguish primary RAC/RCC trees from LCP trees. Swofford (1991) provided two examples of LCP trees that are useful here. His first example (1991: fig. 14.6) is presented here in Figure 7. The LCP tree (Fig. 7c) for the two fundamental trees (Figs. 7a, 7b) is also the primary RCC/RAC tree. Swofford (1991:304) noted that "this approach is particularly useful when only a few taxa are responsible for the incongruence among trees, thereby providing a means of identifying 'unstable' taxa."

The example also illustrates that not all common pruned trees are RCC/RAC trees and vice versa. The RAC profile is completed by a single secondary RAC tree, (B,D)E. Similarly, the RCC profile is completed by a single secondary RCC tree (Fig. 7d). The two smaller common pruned trees (Figs. 7e, 7f) satisfy condition 1 for the definition of an RCC tree (they have an unambiguous interpretation) but they are
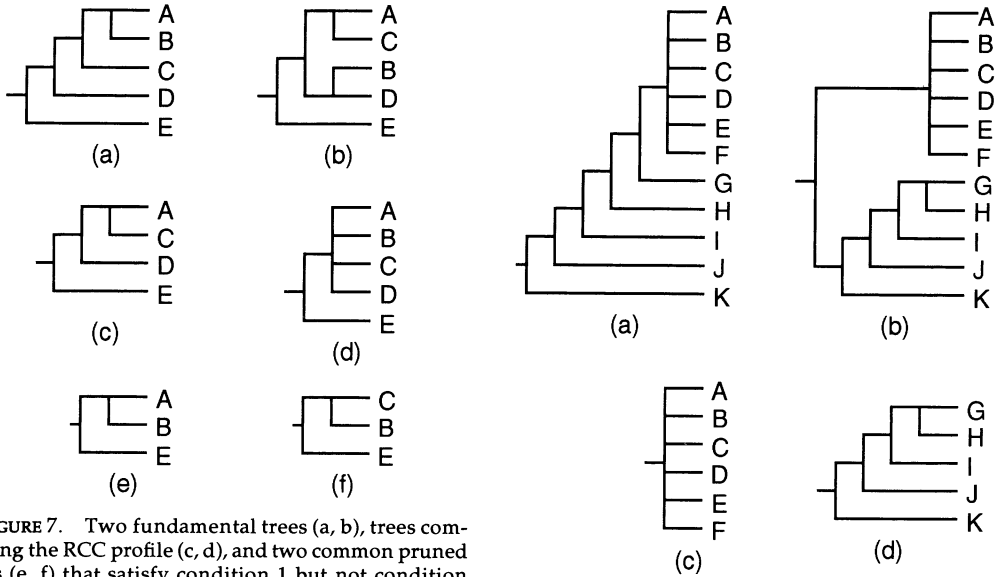
FIGURE 7. Two fundamental trees (a, b), trees comprising the RCC profile (c, d), and two common pruned trees (e, f) that satisfy condition 1 but not condition 2 of the definition of RCC trees. The largest common pruned tree (c) is also the primary RCC tree. [Modified and extended from Swofford, 1991.]
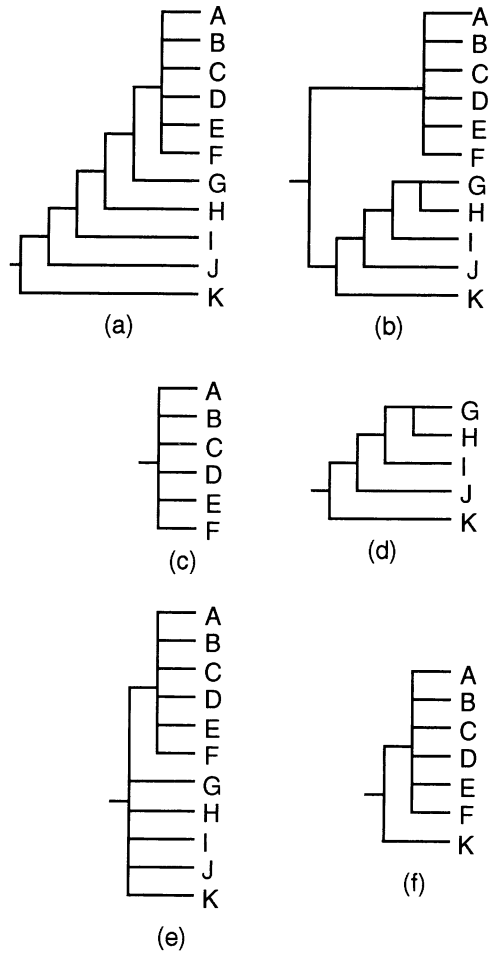

FIGURE 8. Two fundamental trees (a, b), a common pruned tree (c), the primary RCC/RAC tree (d), the unique secondary RCC tree (e), and the unique secondary RAC tree (f). [Modified and extended from Swofford, 1991.]

subtrees of the second RCC tree in the profile and so cannot themselves be RCC trees. Similarly, these trees are not members of the RAC profile because they are not polytomy subtrees of the Adams consensus.

Swofford (1991:304) noted also that "like other consensus methods the largest-common-pruned-tree approach has some undesirable properties that limit its utility." He illustrated the major problem with the second example (1991: fig. 14.7), which is presented here in Figure 8. The LCP tree identified by Swofford (Fig. 8c) for the two fundamental trees (Figs. 8a, 8b) is an unresolved bush. In contrast, Swofford pointed out that the smaller common pruned tree (Fig. 8d) conveys more information than the larger one. The better but smaller common pruned tree identified by Swofford is the primary RCC tree (it is also the primary RAC tree). The apparent greater information content of this tree over the LCP tree is due to its inclusion of a greater number of nodes. The criteria used in the definition of primary RAC or RCC trees (maximizing included nodes in preference to included number of taxa) might also be

usefully applied to the selection of preferred common pruned trees.

Polytomies that occur in DFSs are ones that are also true of all fundamental trees and thus can also be present in common pruned trees. DFSs corresponding to primary RAC or RCC trees would also be preferred common pruned trees selected on the basis of these criteria. Thus, algorithms for finding DFSs could be developed by suitable modification of the algorithms that have been described for finding LCP trees.

Swofford's example appears to include

an error. The tree in Figure 8c is not in fact the LCP tree. A set of five common pruned trees, which each include a single additional taxa, share this status. One of these is shown in Figure 8f. This tree is the same as the tree identified by Swofford except that taxon K is included also. Equally, any one of the taxa G–K could be included in this position giving five equally large common pruned trees. The illustrated tree (Fig. 8f) is also an RAC tree, and together with the primary RAC tree (Fig. 8d) it completes the RAC profile. Interpreted as an LCP tree it is but one of five, whereas interpreted as an RAC tree it is the only secondary RAC tree. Similarly, the tree in Figure 8f is the single secondary RCC. It is also the strict consensus tree and is not a common pruned tree. It includes the cladistic information that is included in the set of five LCP trees. Some of this information, (A–E)F–J, is simply not included in the RAC profile.

The Use of Consensus Methods

Barrett et al. (1991) recently distinguished five distinct uses of consensus methods:

1. to find the points of agreement among a set of trees that are equally most parsimonious, or nearly so, with respect to a single data set (this use is not really specific to parsimony analysis!);
2. to find points of agreement among hypotheses based on the separate analyses of data sets that could be combined;
3. to estimate the degree of agreement or disagreement among different data sets;
4. to examine the agreement or disagreement among the results of different methods of analysis of the same data set; and
5. to find points of agreement among hypotheses based on the separate analysis of data sets that cannot be combined.

To this list can be added

6. investigation of shared historical patterns in the evolution of different lineages or of lineages and biogeographic areas;
7. use of majority-rule consensus trees to summarize the results of bootstrapping; and
8. the recently developed use of the strict consensus in the estimation of the Bremer support for particular groups and total support for the overall hierarchical correlation between characters in a cladistic data set (Källersjö et al., 1992).

Barrett et al. (1991) provided a strong critique of the second of these uses of consensus methods and concluded that data sets should be combined into a single analysis whenever possible. They bolstered their argument with a hypothetical example of two data sets. When these were analyzed separately and a strict consensus produced, the consensus was inconsistent with the tree supported by the combined data.

Barrett et al. (1991) took pains to draw a distinction between the different uses of consensus trees and to point out the specificity of their criticisms for the second use only. The RCC method has been developed in the context of the first usage, but it may be useful in other contexts also. On this use of consensus techniques, Barrett et al. (1991:491) commented,

> Indeed, far from wishing to criticize the use of consensus methods in this context, we believe that there is much to be said for them. If the most parsimonious tree, relative to all the data, requires 50 changes in character states and the next most parsimonious tree requires 51, one may wish to take seriously only those clades endorsed by both trees. Conservatism has a role to play in such cases, one that may be well served by the method of strict consensus.

I am in almost full agreement with this sentiment. However, I suggest that the greater sensitivity to common cladistic information of the RCC method over the strict consensus renders it better suited to such examinations of the level of agreement between fundamental trees.

Carpenter (1988) argued that the extension of consensus techniques, originally designed for use with trees produced from

different data sets, to the case of equally optimal trees from a single data set (use 1) is a misapplication. He argued that because the consensus is less resolved than any of the trees from which it is calculated it also has less explanatory power. Thus, any of the optimal trees would be a better choice as a phylogenetic hypothesis than the consensus. Ignoring the fact that, contrary to Carpenter, some consensus methods yield trees that may be no less resolved than some of the equally optimal trees, the important question is how a choice can be made between equally optimal trees.

Carpenter suggested the use of successive approximations character weighting (Farris, 1969), and Rodrigo (1992) recently developed compatibility and maximum likelihood criteria of hypothesis preference to this end. However, none of these methods will necessarily lead to a single optimal solution. For example, when many equally parsimonious trees result from the inclusion of underdetermined incomplete fossils, the effects of the fossils are unlikely to be ameliorated by any of these methods. Thus it remains useful to be able to summarize the common cladistic information of a set of optimal fundamental trees.

Carpenter's (1988) general argument also warrants further comment. If a consensus tree, such as the strict consensus tree, is less resolved than any of the corresponding fundamental trees then, provided the additional resolution of the fundamental trees is not due to zero-length branches, the strict consensus tree will indeed be less parsimonious than any fundamental tree. However, the additional length is not due to additional hypotheses of homoplasy in the normal sense. Carpenter's additional homoplasy interpretation implies that the polytomies of the consensus are hard (indicating multiple specification) rather than soft (indicating lack of resolution). This interpretation is inappropriate because it takes the consensus to assert relationships that are in direct contradiction to some of the relationships indicated by the fundamental trees. This is simply not the purpose of consensus trees. The correct interpretation of the polytomies is that they are soft and that they assert neither relation-

ships that are not true of the fundamental trees nor additional homoplasy that is not required by any fundamental tree.

Anderburg and Tehler (1990) argued for the use of consensus trees in relation to the needs of taxonomic practice. They claimed that the explanatory power of consensus trees (or more precisely the strict consensus) derives from topological consistency among the fundamental trees rather than from the amount of similarity that can be explained in terms of synapomorphy. However, they also claimed that consensus trees should not be treated as phylogenetic hypotheses. If consensus trees are not (conservative) summary statements about phylogenetic relationships (i.e., conservative phylogenetic hypotheses), then how can they be a basis for a phylogenetic classification? RAC and RCC trees are intended to be interpreted as a set of statements about cladistic relationships (i.e., as reduced phylogenetic hypotheses). Similarly, any consensus tree can be taken as a phylogenetic hypothesis that is uncontradicted by any, or by a majority, of the fundamental trees, provided that their polytomies are interpreted correctly.

The fact that the RCC profile may include more than a single RCC tree might be seen as a disadvantage of this method with respect to the needs of classification. The prime motivation behind this method is to allow the efficient graphical representation of common cladistic information. However, inasmuch as classifications are based upon phylogenetic conclusions, then classifications can be based upon an RCC profile, just as they can be based upon any consensus tree. All that is required of the classification is that it be able to accommodate uncertainty that may exist in the hypotheses of phylogenetic relationships upon which it is based.

Choice of Consensus Method

Mickevich and Platnick (1989) advocated the general use of the Adams consensus in preference to the strict consensus because it shows a better fit to the fundamental trees. Page (1992a) questioned this claim. He noted (1992:93) that choosing a consensus method requires a minimum of

two decisions: "(1) what aspect of tree structure is considered important and (2) what level of agreement between trees is to be reflected in the consensus." He also claimed that both the strict and the Adams consensus trees combine all the taxonomic statements common to the fundamental trees and that the real difference lies not in how much structure they combine but in what this structure is. Thus, choosing among consensus methods involves choosing among different kinds of structure.

These consensus methods focus on different aspects of cladistic structure; however, the different foci may reveal, and thus be sensitive to, different amounts of cladistic information. Thus, the RCC profile will summarize all common n-taxon statements, including components. In contrast, the strict consensus summarizes only components, a subset of the cladistic information present in the RCC profile. Here, the choice has not been between one kind of information or another, but between one set of information and a more inclusive set. Mickevich and Platnick's (1989) advocacy of the Adams consensus seems to be in reference to the first decision that must be made in choosing a consensus. They advocated the use of the method they believed to attach significance to the greatest amount of common cladistic information. However, their claim (1989:40) that "any taxonomic statements shared by the classifications being compared are included in the [Adams] consensus" goes too far. Not all common n-taxon statements are included in the Adams consensus; rather, they are simply not ruled out by their ambiguous polytomies.

Rohlf (1982) noted that the Adams consensus may include groups that do not appear in any of the fundamental trees and preferred the strict consensus on that basis. An example of such a *Rohlf group* is the group B + F in the Adams consensus (Fig. 2c) for the pair of fundamental trees in Figures 2a and 2b. The same Rohlf group also appears in the smallest RAC tree (Fig. 2e), which is also an RCC tree. The interpretation of the Rohlf group in the RCC/RAC tree is unproblematic, and thus the appearance of Rohlf groups in a consensus

is not necessarily a good reason for rejecting a consensus method. Avoiding Rohlf groups limits consensus methods to the representation of common component information with consequent insensitivity to common n-taxon statements of lower cardinality. The real problem of Rohlf groups in Adams consensus trees is the ambiguity of the polytomies that must accompany them and the possibility of misinterpretation. The RAC and RCC methods have been developed specifically to avoid these problems.

When consensus methods are used to summarize common cladistic information, then the strict consensus is the best method for representing just common components. The RCC appears to be the single best method for representing common n-taxon statements, including components, and renders the strict consensus redundant. If no member of the RCC profile includes all taxa (and is therefore also a strict consensus), then the strict consensus will be an unresolved bush. One or more DFSs formed from primary RCC trees, together with appended lists of the possible positions of taxa that are not included in the DFS, provide a means of representing both n-taxon statements and additional disqualifiers. The Adams consensus provides the best method of representing common nestings, and the RAC facilitates the unambiguous interpretation of an Adams consensus identifying those nestings that are true by virtue of common n-taxon statements and those that have a more ambiguous interpretation. I advocate the use of multiple methods because only through the combined application of consensus methods can a more complete understanding of the common cladistic information be obtained.

Consensus and Support

Techniques that utilize consensus methods for evaluating the support for particular groups in a cladogram have been developed in the context of parsimony analysis (Farris et al., 1982; Bremer, 1988; Källersjö et al., 1992). Evaluation of support is based upon the comparison of consensus trees for a set of optimal fundamental trees with consensus trees for sets of fundamen-

tal trees, including suboptimal ones. In this approach, "a group on a considered most parsimonious tree is supported by strong evidence when a large increase in the length of included trees is required before that group is lost in the consensus" (Källersjö et al., 1992:284). Källersjö et al. (1992) defined the Bremer support of a group as the difference in length between the considered tree and the shortest tree lacking this group. Because the strict consensus tree will not include any group that is absent from the set of fundamental trees, Källersjö et al. (1992), following Bremer (1988), suggested the use of this consensus method in the assessment of support. They noted that the minimal assessment of common structure (cladistic information) provided by the strict consensus method is perhaps much too strict but that other consensus techniques have faults also. Källersjö et al. also defined total support as the sum of the Bremer supports of each group.

Based as it is upon the strict consensus, Bremer support is a measure of the support for a particular component. Similarly, total support as defined by Källersjö et al. (1992) represents total component support. Analogous measures of support can be developed for cladistic information, other than components, that are used in other consensus methods. Further, analogous measures can be developed for alternative methods of phylogenetic inference. For example, where hypothesis preference is based upon the largest cliques of compatible characters, the analogue of Bremer support would be the difference in numbers of included characters in the largest clique supporting a particular component and the largest clique that does not support the component. The sum of the group supports would provide a clique-based total component support. Here, I define the general support measure as the difference in optimized quantity between the most optimal tree that includes a particular item of cladistic information and the most optimal tree that does not include this item. For any class of items, the total support measure will be the sum of the individual support measures for all items of that kind that are included on a particular tree.

Källersjö et al. (1992) developed their total support measure in the context of contributions to parsimony-based randomization or permutation tests of cladistic data. The randomization tests developed by Archie (1989) and Faith and Cranston (1991) compare tree lengths of most-parsimonious trees for real data to the lengths of shortest trees produced from randomly permuted data. Källersjö et al. (1992:283) advocated support-based randomization tests on the premise that total support provides a better measure of the extent to which the data show "unambiguous hierarchic structure" than does tree length. The difference is most profound when there are equally most-parsimonious trees and little resolution in the strict consensus. Those groups of fundamental trees that are not included in the strict consensus will have a Bremer support of zero and contribute nothing to total component support. Although a data set might support minimum length trees that are significantly shorter than those supported by randomly permuted data, it may yet fail a randomization test that is based upon total component support.

As we have seen, strict consensus trees can be highly unresolved despite the fact that the various fundamental trees include many common n-taxon statements of lower cardinality than the components. Thus, the support-based randomization test of Källersjö et al. (1992), which uses the strict consensus, can be expected to give spurious results. For example, a highly structured cladistic data set (containing many characters and no incongruence) may appear not to be significantly different from random data in a support-based randomization test simply because of the inclusion of poorly known taxa that cause many nodes to be collapsed in the strict consensus. Support measures that employ the RCC method, and thus do not focus just upon components, may provide a better framework for the further development of support-based randomization tests. As currently employed, bootstrapping is also a component-based procedure. An n-taxon statement of lower cardinality than the components could have high bootstrap

proportions even when all components that imply the statement have low bootstrap proportions.

Consensus Terminology

The strict consensus is one of a family of consensus methods, based on replicated components, that also includes the major-ity-rule and combinable component (or semistrict) consensus methods. The Adams consensus is analogous to the strict con-sensus in its requirement that the nestings represented upon the fundamental tree be replicated in all fundamental trees; both semistrict and majority-rule variants of the Adams method are possible. The terms strict, majority rule, and semistrict might be useful as general descriptors of consen-sus methods rather than being restricted to the names of particular component-based consensus methods. Thus, the strict con-sensus should be more properly termed a strict component consensus. Similarly, the Adams and LCP consensus methods should properly be termed the strict Adams and strict LCP consensus methods.

The RCC method described in this paper is a strict RCC method. Thus the possibility remains that majority-rule and semistrict RCC methods might also be developed. There are contrasting expectations of strict and semistrict RCC trees when the fun-damental trees being compared include overlapping but not equivalent collections of terminal taxa. A strict RCC tree can only include taxa that are common to all fun-damental trees. The n-taxon statements that include taxa that are not represented in all fundamental trees cannot be replicated in all those trees. In contrast, a semistrict RCC tree can include taxa that are not common to all fundamental trees provided that the n-taxon statements including them are un-contradicted by any of the fundamental trees (Fig. 9).

CONCLUSIONS

Consensus trees are intended to sum-marize cladistic information that is com-mon to a set of fundamental trees. To un-derstand existing consensus methods or to have informed development of useful new methods requires a careful consideration
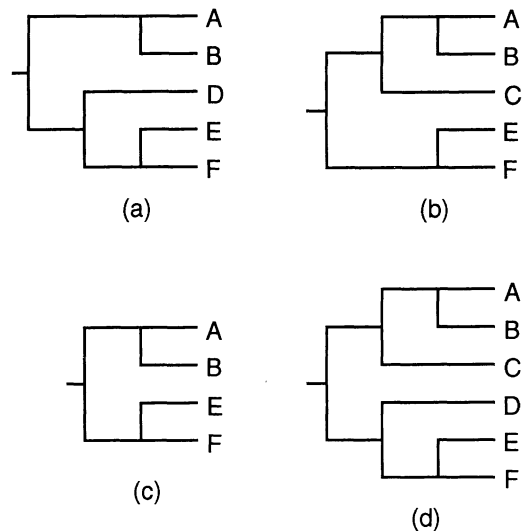

FIGURE 9. Two fundamental trees including over-lapping but different sets of terminal taxa (a, b), the strict RCC profile (c), and the semistrict RCC profile (d). Both profiles contain a single RCC tree. The strict RCC tree can only include relationships among taxa that are common to both fundamental trees. The semi-strict RCC tree can incorporate relationships that are not true of both fundamental trees if they are not contradicted by either of them.

of the nature of cladistic information. Sev-eral classes of cladistic information can be distinguished that interrelate in a variety of ways. Component consensus methods can be particularly insensitive to common cladistic information represented by n-tax-on statements of lower cardinality. The po-lytomies of Adams consensus trees are am-biguous with respect to common n-taxon statements. The RAC method enables the Adams consensus to be interpreted un-ambiguously. The RCC method provides a graphical solution to the representation of all common n-taxon statements. In addi-tion, common disqualifiers (negations of n-taxon statements) can be identified using DFSs and described by lists of their pos-sible positions relative to one or more DFSs. The RAC, RCC, and DFS approaches to the problem of consensus should provide sys-tematists with a series of tools that will enable them to compare trees more effec-tively. The identification of different kinds of cladistic information may provide a whole range of support-based measures.

Of these, a support measure based upon the RCC method provides a more promising approach for support-based randomization tests than do measures based upon the strict consensus. Similarly, bootstrapping might be enhanced by a majority-rule RCC method.

This discussion is only a preliminary exploration of the new consensus methods. These methods probably will have additional applications. For example, they may provide a basis for the development of consensus indices for quantifying agreement among fundamental trees. Further exploration and the practical utility of the methods would be greatly facilitated by the incorporation of these methods into widely used phylogenetic inference and biogeographic software packages.

ACKNOWLEDGMENTS

REFERENCES

ADAMS, E. N. 1972. Consensus techniques and the comparison of taxonomic trees. Syst. Zool. 21:390–397.

ADAMS, E. N. 1986. N-trees as nestings: Complexity, similarity and consensus. J. Classif. 3:299–317.

ANDERBURG, A., AND A. TEHLER. 1990. Consensus trees, a necessity in taxonomic practice. Cladistics 6:399–402.

ARCHIE, J. W. 1989. A randomization test for phylogenetic information in systematic data. Syst. Zool. 38:239–252.

BARRETT, M., M. J. DONOGHUE, AND E. SOBER. 1991. Against consensus. Syst. Zool. 40:486–493.

BREMER, K. 1988. The limits of amino-acid sequence data in angiosperm phylogenetic reconstruction. Evolution 42:795–803.

BREMER, K. 1990. Combinable component consensus. Cladistics 6:369–372.

CARPENTER, J. M. 1988. Choosing among multiple equally parsimonious cladograms. Cladistics 4:291–296.

CRACRAFT, J., AND K. HELM-BYCHOWSKI. 1991. Parsimony and phylogenetic inference using DNA sequences. Some methodological strategies. Pages 184–220 in Phylogenetic analysis of DNA sequences (M. M. Miyamoto and J. Cracraft, eds.). Oxford Univ. Press, New York.

DAY, W. H. E. 1983. The role of complexity in comparing classifications. Math. Biosci. 66:97–114.

DOBSON, A. J. 1975. Comparing the shapes of trees. Pages 95–100 in Combinatorial mathematics. III. Lecture notes in mathematics 452. Springer-Verlag, Berlin.

ESTABROOK, G. F. 1992. Evaluating undirected positional congruence of individual taxa between two estimates of the phylogenetic tree for a group of taxa. Syst. Biol. 41:172–177.

ESTABROOK, G. F., F. R. MCMORRIS, AND C. A. MEACHAM. 1985. Comparison of undirected phylogenetic trees based on subtrees of four evolutionary units. Syst. Zool. 34:193–200.

ESTABROOK, G. F., J. G. STRAUCH, AND K. L. FIALA. 1977. An application of compatibility analysis to the Blackiths' data on orthopteroid insects. Syst. Zool. 26:269–276.

FAITH, D. P., AND P. S. CRANSTON. 1991. Could a cladogram this short have arisen by chance alone? On permutation tests for cladistic structure. Cladistics 7:1–28.

FARRIS, J. S. 1969. A successive approximations approach to character weighting. Syst. Zool. 18:374–385.

FARRIS, J. S. 1988. Hennig86, version 1.5. Distributed by the author, Port Jefferson Station, New York.

FARRIS, J. S., A. G. KLUGE, AND M. F. MICKEVICH. 1982. Immunological distances and the phylogenetic relationships of the Rana boylii species group. Syst. Zool. 31:479–491.

FINDEN, C. R. 1983. CMTREE manual, report. Department of Statistics, University of St. Andrews, St. Andrews, Scotland.

FINDEN, C. R. 1984a. FSTREE manual, report. Department of Statistics, University of St. Andrews, St. Andrews, Scotland.

FINDEN, C. R. 1984b. LMTREE manual, report. Department of Statistics, University of St. Andrews, St. Andrews, Scotland.

FINDEN, C. R. 1984c. UBTREE manual, report. Department of Statistics, University of St. Andrews, St. Andrews, Scotland.

FINDEN, C. R., AND A. D. GORDON. 1985. Obtaining common pruned trees. J. Classif. 2:225–276.

FUNK, V. A., AND D. R. BROOKS. 1990. Phylogenetic systematics as the basis of comparative biology. Smithson. Contrib. Bot. 73:1–45.

GORDON, A. D. 1980. On the assessment and comparison of classifications. Page 149–160 in Analyse de donnees et informatique (R. Tomassone, ed.). INRIA, LeChesnay, France.

GORDON, A. D. 1986. Consensus supertrees: The synthesis of rooted trees containing overlapping sets of leaves. J. Classif. 3:335–348.

GORDON, A. D. 1987. A review of hierarchical classification. J. R. Stat. Soc. A 150:119–137.

KÄLLERSJÖ, M., J. S. FARRIS, A. G. KLUGE, AND C. BULT. 1992. Skewness and permutation. Cladistics 8:275–287.

MADDISON, W. 1989. Reconstructing character evolution on polytomous cladograms. Cladistics 5:365–377.

MARGUSH, T., AND F. R. MCMORRIS. 1981. Consensus *n*-trees. Bull. Math. Biol. 43:239–244.

MCMORRIS, F. R., D. B. MERONK, AND D. A. NEUMANN. 1983. A view of some consensus methods for trees. Pages 122–126 *in* Numerical taxonomy: Proceedings of NATO Advanced Study Institute (J. Felsenstein, ed.). Springer-Verlag, Berlin.

MEACHAM, C. A. 1981. A manual method of character compatibility analysis. Taxon 30:591–600.

MICKEVICH, M. F., AND N. I. PLATNICK. 1989. On the information content of cladograms. Cladistics 5:33–47.

NELSON, G. 1979. Cladistic analysis and synthesis: Principles and definitions, with a historical note on Adanson's *Famille des Plantes* (1763–1764). Syst. Zool. 28:1–21.

NELSON, G., AND P. Y. LADIGES. 1991. Three-area statements: Standard assumptions for biogeographic analysis. Syst. Zool. 40:470–485.

NELSON, G., AND N. I. PLATNICK. 1980. Multiple branching in cladograms: Two interpretations. Syst. Zool. 29:86–91.

NELSON, G., AND N. I. PLATNICK. 1981. Systematics and biogeography: Cladistics and vicariance. Columbia Univ. Press, New York.

NELSON, G., AND N. I. PLATNICK. 1991. Three-taxon statements: A more precise use of parsimony? Cladistics 7:351–366.

NEUMANN, D. A. 1983. Faithful consensus methods for *n*-trees. Math. Biosci. 63:271–287.

PAGE, R. D. M. 1988. Quantitative cladistic biogeography: Constructing and comparing area cladograms. Syst. Zool. 37:254–270.

PAGE, R. D. M. 1989. Comments on component-compatibility in historical biogeography. Cladistics 5:167–182.

PAGE, R. D. M. 1992a. Comments on the information content of classifications. Cladistics 8:87–95.

PAGE, R. D. M. 1992b. COMPONENT for Windows. The Natural History Museum, London.

RODRIGO, A. C. 1992. Two optimality criteria for selecting subsets of most parsimonious trees. Syst. Biol. 41:33–40.

ROHLF, F. J. 1982. Consensus indices for comparing classifications. Math. Biosci. 59:131–144.

ROSEN, D. E. 1978. A vicariance model of Caribbean biogeography. Syst. Zool. 24:431–464.

SMITH, A. B. 1989. RNA sequence data in phylogenetic reconstruction: Testing the limits of its resolution. Cladistics 5:321–344.

SOKAL, R. R., AND F. J. ROHLF. 1981. Taxonomic congruence in the Leptopodomorpha re-examined. Syst. Zool. 30:309–325.

STINEBRICKNER, R. 1984. S-consensus trees and indices. Bull. Math. Biol. 46:923–935.

SWOFFORD, D. L. 1985. CONTREE (consensus tree program). Illinois Natural History Survey, Champaign.

SWOFFORD, D. L. 1990. PAUP: Phylogenetic analysis using parsimony, version 3.0. Illinois Natural History Survey, Champaign.

SWOFFORD, D. L. 1991. When are phylogeny estimates from molecular and morphological data incongruent? Pages 295–333 *in* Phylogenetic analysis of DNA sequences (M. M. Miyamoto and J. Cracraft, eds.). Oxford Univ. Press, New York.

WILKINSON, M. 1992. Consensus, compatibility and missing data in phylogenetic inference. Ph.D. Thesis, Univ. Bristol, Bristol, England.

WILKINSON, M. 1994. The permutation method and character compatibility. Syst. Biol. 43:274–277.

APPENDIX

This appendix includes algorithms for the construction of RAC and RCC profiles and some preliminary suggestions concerning the construction of DFSs and determination of possible positions. Each algorithm is illustrated by its application to the two fundamental trees in Figures 2a and 2b.

Constructing the RAC Profile

The algorithm described here will produce the complete RAC profile for a given collection of fundamental trees. It uses the Adams consensus as a starting point. The Adams consensus method is implemented in PAUP 3.0 as well as in the CONTREE (Swofford, 1985) and COMPONENT (Page, 1992b) software packages. Adams consensus trees can be converted to RAC trees by selective elimination of taxa that contribute to ambiguous polytomies when interpretation 1 of the polytomy is not true for all the fundamental trees, until this interpretation is true of relationships among the remaining taxa (condition 1). The alternative ways of eliminating taxa produce the various RAC trees that constitute the RAC profile. Adams consensus trees are not invariant under different rootings of the fundamental trees; this discussion applies only to rooted trees.

The algorithm makes pairwise comparisons of the relationships of the groups that contribute to Adams polytomies (*p*-groups) across the fundamental trees. The aim of the comparison is to determine whether each *p*-group of the pair comprises taxa that are more closely related to each other than they are to any of the taxa in the other *p*-group. Only when this condition is met by all the fundamental trees can the two groups be included in a single consensus tree without implying erroneous relationships under interpretation 1, in which case the pair can be described as being *consensus compatible*.

Suppose, for example, that there are three branches from an Adams census tree polytomy that lead to the *p*-groups (A,B,C), (D,E), and F. There are three pairwise comparisons of these *p*-groups given by the partitions ABC\DE, ABC\F, and DE\F. The condition for consensus compatibility of the two groups con-

tributing to the first pairwise comparison requires that both the n-taxon statements (A,B,C)D,E and (D,E)A,B,C are true of all fundamental trees.

Whether such a pair of n-taxon statements are consistent with a tree can be determined using a method suggested by R. Page (pers. comm.). The last common ancestor of the taxa comprising each p-group in the pair is determined (e.g., the last common ancestor of A, B, and C and the last common ancestor of D and E). Each last common ancestor identifies a component on the fundamental tree. These components may include terminal taxa other than those included in the p-group, and I distinguish them as c-groups. If the intersection of the two c-groups is empty, then the corresponding n-taxon statements for the p-groups are consistent with the fundamental tree. If this holds across all fundamental trees, the p-groups are consensus compatible. When the pairwise comparison of p-groups is extended to cases involving p-groups that contribute to different polytomies in the Adams consensus tree, then one p-group may be a subset of another (the intersection of the p-groups is not empty). Such p-groups must be considered consensus compatible, although the intersection of the corresponding c-groups will not be empty. Prior identification of such p-group pairs removes the step of determining the intersection of the corresponding c-groups. Thus, only if the intersection of the p-groups is empty and the intersection of the corresponding c-groups is not empty will the p-groups be consensus incompatible.

The results of each pairwise comparison of p-groups can be usefully summarized in the form of a consensus compatibility matrix, in which each p-group that contributes a separate branch to any of the polytomies in the Adams consensus tree is compared with every other such p-group. A "yes" or "1" is entered into the matrix if the pair of p-groups are consensus compatible. Clique analysis (Estabrook et al., 1977; Wilkinson, 1994) is performed on the consensus compatibility matrix to determine all cliques of p-groups that are mutually consensus compatible. Cliques are nonredundant, comprising the largest possible collections of compatible characters (or in this case consensus compatible p-groups), and thus the use of clique analysis guarantees that the consensus trees generated by the algorithm satisfy condition 2 for RAC trees. Meacham (1981) described a manual method of compatibility analysis that identifies cliques.

The RAC profile includes an RAC tree corresponding to each clique. Each RAC tree is formed from an Adams consensus by pruning those p-groups that are not members of the particular clique from the Adams consensus tree. Given these preliminaries, the algorithm can now be stated.

1. Determine the Adams consensus tree.
2. Compute the consensus compatibility matrix for all p-groups in the Adams consensus tree.
3. Perform a clique analysis of the consensus compatibility matrix.
4. Construct an RAC tree corresponding to each clique.

A strict consensus tree will include all those components that are replicated across all fundamental trees. Such components will also appear in the Adams consensus tree, and if they appear as a p-group in the Adams consensus they will necessarily be consensus compatible with all other p-groups arising from the polytomy. Comparison of strict and Adams consensus trees can thus provide a shortcut to the completion of some of the cells in the consensus·compatibility matrices.

An Example

The method can be illustrated with reference to the two fundamental trees (Figs. 2a, 2b) that Adams (1972) used to illustrate his method. The Adams consensus for the two trees is given in Figure 2c, and the strict consensus is given in Figure 2d for comparison. The strict consensus is completely unresolved, whereas the Adams consensus shows considerable structure. There are two polytomies in the Adams consensus, each of which yields three p-groups. The consensus compatibility matrix is given in Table A1, and the results of the clique analysis are given in Table A2. Removing the groups that are not members of each clique from the Adams consensus yields three trees (Figs. 2e–g). These three trees are RAC trees, and together they comprise the RAC profile.

Constructing the RCC Profile

The following algorithm requires discovery of n-taxon statements of highest cardinality that are true across a set of fundamental trees. These statements are then combined in specific ways to produce trees. An exhaustive approach would involve formulation of each possible n-taxon statement in turn and evaluation of each statement across each of the fundamental trees. A more efficient approach would be to formulate all n-taxon statements that are true of a particular fundamental tree and test only these against the other fundamental trees. No other n-taxon statements are true of the selected tree, and therefore they cannot be true of all fundamental trees. Further efficiency could be obtained by formulating triplets that are true of a particular fundamental tree and testing only these against the fundamental trees. True n-taxon statements of higher cardinality can then be built up through the combination of true triplets. An initial algorithm based upon this approach was rejected in favor of an alternative method based upon comparison of the components of trees, which is more efficient.

A component statement is also an n-taxon statement, so that they have inside and outside sets. When components on a pair of trees differ, they may yet agree on some informative n-taxon statements of lower cardinality. The n-taxon statement of highest cardinality that they agree upon is given by the intersections of the outside and inside sets. Thus, if two trees for taxa A–F have the components [A–C] and [A,C,D], the equivalent n-taxon statements are (A–C)D–F and (A,C,D)B,E,F. These components agree upon the n-taxon statement (A,C)E,F. Components will only agree upon an informative n-taxon statement when the intersection of their inside sets contains more than a single taxon and the intersection of their outside sets is not empty.

A pairwise comparison of all the components de-

TABLE A1. Consensus compatibility matrix for the six p-groups contributing to polytomies in the Adams consensus tree of Figure 2d.

p-groups	p-groups				
	1	2	3	4	5
1. A	—				
2. B,F	1	—			
3. C,D,E,G,H	1	0	—		
4. D	1	0	1	—	
5. E,H	1	0	1	1	—
6. G	1	0	1	1	0

termines a set of n-taxon statements that are common to both trees. This set can be summarized in an intersection matrix. These n-taxon statements can then be compared with components in the next fundamental tree, and so on. Some of these n-taxon statements may be redundant (i.e., they are implied by other n-taxon statements). An n-taxon statement is redundant when both its inside and outside sets are subsets of the inside and outside sets, respectively, of some other n-taxon statement. The algorithm requires redundant n-taxon statements to be eliminated. This can be achieved efficiently by an eliminative pairwise comparison that makes use of the fact that once an n-taxon statement is identified as redundant no more pairwise comparisons involving it need be made. Elimination of these statements occurs each time an intersection matrix is computed. The remaining nonredundant set of n-taxon statements takes the place of components, and these statements are compared with the components of the next fundamental tree to determine their intersection matrix and those n-taxon statements included in it that are nonredundant. This procedure is repeated until the components of all remaining fundamental trees have been compared with the common n-taxon statements remaining from the previous comparison of trees or until there are no common n-taxon statements. The result in the first case is a group of true n-taxon statements of the highest possible cardinality that include all n-taxon statement information that is common to the fundamental trees (an hc set).

Additional efficiency can be obtained by determining the strict consensus. All components of the strict consensus are also components in all the fundamental trees. Thus, they can be included in the hc set, and no comparisons involving them need be made. The n-taxon statements of the hc set are used to construct the RCC profile. When a group of n-taxon statements

have equivalent complete sets of included taxa, then combining these to form a tree is a trivial exercise. The n-taxon statements that do not have equivalent complete sets may yet be rendered equivalent by simply removing taxa from one or more of the n-taxon statements. This step is important for the construction of RCC trees only when equivalence of complete sets can be produced without the modification of one of the original n-taxon statements in the hc set and when any new n-taxon statement so produced is informative. This will be true when (1) one of a complete set of one n-taxon statement is a less inclusive subset of the complete set of some other n-taxon statement(s), (2) the intersection of the inside sets includes at least two taxa, and (3) the intersection of the outside sets is not empty. All n-taxon statements that can be so produced can be found by a pairwise examination of the n-taxon statements of the hc set.

The algorithm can be stated as follows.
1. Determine the hc set; if this is empty then stop.
2. For each n-taxon statement in the hc set, produce any new informative n-taxon statements with equivalent complete sets that can be formed by the deletion of taxa from other n-taxon statements in the hc set and add them to the hc set.
3. Construct a tree for each group of n-taxon statements in the hc set that has equivalent complete sets. Add the tree so produced to the RCC profile.

An Example

The operation of the algorithm is illustrated here for the same pair of fundamental trees used to illustrate the construction of the RAC profile. We begin with the two fundamental trees in Figures 2a and 2b. The intersection matrix summarizing the comparison of components and the resultant n-taxon statements are given in Table A3. Five n-taxon statements are nonredundant:
1. (E,H)A,B,C,D
2. (G,H)A,B,C,F
3. (D,E,G,H)A,B,C
4. (C,D,E,G,H)A,B
5. (B,F)A

If there were further fundamental trees, these n-taxon statements would be compared with the components of the next fundamental tree to produce a new intersection matrix, but because there are only two trees these statements comprise the hc set. Statements 1, 3, and 4 can be rendered equivalent to their complete sets by the elimination of taxon G from 3 and 4. This produces two additional n-taxon statements that are added to the hc set.
6. (D,E,H)A,B,C

TABLE A2. Cliques of consensus compatible p-groups determined from the consensus compatibility matrix in Table A1.

Tree	Cliques of p-groups	Excluded p-groups
Figure 2e	A; (B,F)	(C,D,E,G,H); D; (E,H); G
Figure 2f	A, (C,D,E,G,H); D; G	(B,F); (E,H)
Figure 2g	A; (C,D,E,G,H); D; (E,H)	(B,F); G

TABLE A3. Intersection matrix of n-taxon statements formed through the pairwise comparison of the components of the two fundamental trees in Figures 2a and 2b. The n-taxon statements of the highest cardinality (the hc set) are marked with an asterisk.

| Components of tree 2a | Components of tree 2b | | | | | |
	(E,H) A–D,F,G	(D,G) A–C,E,F,H	(D,E,G,H)A–C,F	(C–E,G,H)A,B,F	(B,F) A,C–E,G,H	(A,B,F) C–E,G,H
(G,H)A–F	—	—	(G,H)A–D,F*	(G,H)A,B	—	—
(F–H)A–E	—	—	(G,H)A–D	(G,H)A,B	—	—
(E–H)A–D	(E,H)A–D*	—	(E,G,H)A–C	(E,G,H)A,B	—	—
(D–H)A–C	(E,H)A–C	(D,G)A–C	(D,E,G,H)A–C*	(E,D,G,H)A,B	—	—
(C–H)A,B	(E,H)A,B	(D,G)A,B	(D,E,G,H)A,B	(C–E,G,H)A,B*	—	—
(B–H)A	(E,H)A	(D,G)A	(D,E,G,H)A	(C–E,G,H)A	(B,F)A*	—

7. (C,D,E,H)A,B,

No further informative n-taxon statements can be produced by eliminating taxa from one or more of the statements so that they have complete sets equivalent to those of some other statement in the original hc set. Trees are then constructed from the sets of statements with equivalent complete sets. There are four such sets: statements 3 and 4; statements 1, 6, and 7; statement 2; and statement 3. These yield the four RCC trees (Figs. 4a–d, respectively) that comprise the RCC profile.

Suggestions for the Construction of DFSs

Consider the construction of one or more DFSs from a primary RCC tree. The aim is to produce a DFS that will be most useful for representing the possible positions of taxa that are not included. To this end, we want to maximize primarily the number of nodes and secondarily the number of taxa. The following considerations also apply to DFSs constructed from a primary RAC tree. If the primary RCC tree is fully dichotomous, then it is also the DFS. Only when the primary RCC includes polytomies is there any need for further work. If any of the fundamental trees contain no polytomies, then the polytomies in the primary RCC tree cannot be disqualifier faithful. DFSs can thus be produced by pruning p-groups from the primary RCC tree until the remaining polytomy subtree includes no polytomies. To maximize the number of included nodes, those p-groups that include the fewest nodes are eliminated in preference to p-groups including greater numbers of nodes. Because the resulting DFS is dichotomous, maximizing the number of nodes will also maximize the number of included terminal taxa.

Where all the fundamental trees include one or more polytomies, then constructing DFSs becomes more complicated and requires the comparison of the polytomies in the fundamental trees with those in the primary RCC. The comparison is facilitated by converting the fundamental trees into fundamental subtrees that include only those terminal taxa that are also included in the primary RCC. The other taxa are simply pruned from the fundamental trees. If any of the fundamental subtrees include no polytomy, then a dichotomous DFS can be constructed as described above.

Suppose, for simplicity, that the primary RCC tree and each of the fundamental subtrees include just a single polytomy. Each polytomy can be characterized as a listing of p-groups. Comparing the listing of p-groups across the primary RCC and the fundamental subtrees allows the determination of the set of p-groups that contribute to the polytomy in each of the fundamental trees. If the set includes less than three p-groups, then the DFS will be dichotomous and its construction can proceed as described above. If this set includes three or more p-groups, then it comprises a common subpolytomy.

Where there is a common subpolytomy, this can be represented as such on a DFS by pruning all other p-groups from the primary RCC tree. Whether representing the common subpolytomy in a DFS will produce the most useful DFS depends on the number of nodes and taxa included in the DFS. A dichotomous DFS that includes p-groups that are not part of a common polytomy may include more nodes. Thus the polytomy-containing DFS can be compared with the most useful dichotomous DFS constructed as described above to determine which is most useful as a framework for the description of possible positions.

An algorithm for constructing one or more DFSs that could be used as a framework for representing disqualifiers would need to accommodate the possibility that a number of polytomies might be present at different nodes in the primary RCC tree and in the fundamental subtrees. This can be done by treating each polytomy in the primary RCC tree separately, beginning with those that are furthest from the root.

The relationship between disqualifiers and possible positions suggests that a method for determining replicated disqualifiers could be developed by comparing one or more DFSs with the fundamental trees and determining the positions at which excluded taxa would have to be joined to the DFS to transform it into the corresponding fundamental tree or subtree. With small numbers of fundamental trees, this information can readily be determined by inspection (Wilkinson, 1992).