

Majority-Rule Reduced Consensus Trees and Their Use in Bootstrapping

Mark Wilkinson

School of Biological Sciences, University of Bristol

Bootstrap analyses are usually summarized with majority-rule component consensus trees. This consensus method is based on replicated components and, like all component consensus methods, it is insensitive to other kinds of agreement between trees. Recently developed reduced consensus methods can be used to summarize much additional agreement on hypothesised phylogenetic relationships among multiple trees. The new methods are "strict" in the sense that they require agreement among all the trees being compared for any relationships to be represented in a consensus tree. Majority-rule reduced consensus methods are described and their use in bootstrap analyses is illustrated with a hypothetical and a real example. The new methods provide summaries of the bootstrap proportions of all n -taxon statements/partitions and facilitate the identification of hypotheses of relationships that are supported by high bootstrap proportions, in spite of a lack of support for particular components or clades. In practice majority-rule reduced consensus profiles may contain many trees. The size of the profile can be reduced by constraints on minimal bootstrap proportions and/or cardinality of the included trees. Majority-rule reduced consensus trees can also be selected a posteriori from the profile. Surrogates to the majority-rule reduced consensus methods using partition tables or tree pruning options provided by widely used phylogenetic inference software are also described. The methods are designed to produce more informative summaries of bootstrap analyses and thereby foster more informed assessment of the strengths and weaknesses of complex phylogenetic hypotheses.

Introduction

Bootstrap analyses, introduced by Felsenstein (1985), are now widely used to evaluate phylogenetic inferences, especially in molecular phylogenetics. Bootstrapping involves resampling with replacement of characters or sites from the original data set so as to produce a series of bootstrap replicate data sets with the same number of sites as the original. The replicates are then subject to phylogenetic analysis, yielding a series of bootstrap trees. Bootstrap proportions describe the frequency with which a component or clade (or a partition in unrooted trees) is encountered in the bootstrap trees. These results are conventionally summarized with a majority-rule component consensus (Margush and McMorris 1981) of the bootstrap trees.

The recently developed reduced cladistic consensus (RCC) method (Wilkinson 1994) for rooted trees and its analogue, the reduced partition consensus (RPC) method (Wilkinson 1995) for unrooted trees, were introduced because they have desirable properties of improved sensitivity and reduced ambiguity, compared with other widely used consensus methods. Wilkinson (1994) developed the RCC method after a consideration of the kinds of information that could be shared by two or more rooted trees, and its cladistic or phylogenetic interpretation. An n -taxon statement is an assertion that

Abbreviations: RCC = reduced cladistic consensus; RPC = reduced partition consensus.

Key words: Consensus, reduced consensus, bootstrapping, caecilians, Gymnophiona, phylogeny.

Address for correspondence and reprints: Dr. Mark Wilkinson, School of Biological Sciences, University of Bristol, Bristol BS8 1UG, England. E-mail: Mark.Wilkinson@bris.ac.uk.

Mol. Biol. Evol. 13(3):437–444, 1996
© 1996 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

some set of taxa are more closely related to each other (i.e., that they share a more recent common ancestor) than they are to some other taxa. The RCC method is designed to provide an unambiguous graphical representation of all n -taxon statements that are shared by a set of fundamental trees.

If an n -taxon statement applies to all taxa under consideration then it delimits a component (clade or cluster). Thus components represent a subset of the n -taxon statements that could be shared by trees. Component consensus methods (strict, semistrict, and majority-rule) represent agreement in this subset. In contrast, the RCC method also represents n -taxon statements of lower cardinality than components, i.e., ones that assert relationships among only a subset of the taxa under consideration. For example, consider an analysis of the relationships of the six taxa A–F producing multiple trees. If the trees agree only that A and C are more closely related to each other than they are to D and F, symbolically (AC)DF, but they show no agreement on the relationships of B or E, then this agreement would not be represented in any component consensus tree. The RCC method reflects this agreement through trees that do include those taxa to which the represented n -taxon statements do not apply.

The RPC method is an analogue of the RCC method for unrooted trees. This method represents shared n -taxon partitions rather than n -taxon statements (Wilkinson 1995). Reduced consensus methods may produce more than one consensus tree, with the collection of consensus trees constituting a consensus profile. For rooted trees, the RCC/RPC profiles will include unambiguous representation of all n -taxon statements/parti-

tions that are common to a set of trees. Because components are n -taxon statements of maximal cardinality (they apply to all taxa under consideration) the strict component consensus will be a member of the RCC profile unless it is completely unresolved and uninformative.

The RCC method was considered by Wilkinson (1994) to be a "strict" consensus method in the sense that it demands agreement across all the fundamental trees if an element of cladistic information is to be included in the RCC profile. It was also conjectured that (1) a majority-rule version of the RCC method (analogous to the majority-rule component consensus) is possible, and that (2) such a consensus method would inherit some of the desirable properties of the "strict" RCC method that might usefully be exploited in bootstrapping. In particular, majority-rule reduced consensus methods would make it possible to determine and represent bootstrap proportions for all n -taxon statements/partitions, some of which might be well supported even when no components/bipartitions are well supported. Here, I introduce majority-rule RCC/PC methods, and use a hypothetical and a real example (caecilian DNA sequence data) to illustrate the potential utility of this approach to summarizing bootstrap analyses. I also describe two surrogate methodologies that can be implemented with current versions of widely used phylogenetic inference software in lieu of majority-rule reduced consensus methods.

Materials and Methods

Any phylogenetic inference method that is amenable to bootstrap analysis is also amenable to the kinds of analyses introduced here, and I have used parsimony analyses as an example. All parsimony analyses were done with PAUP 3.1.1 (Swofford 1993) using exact searches. Bootstrap analyses used the same analytical options for 100 replicate data sets. Partition tables and consensus trees were constructed using PAUP 3.1.1 and REDCON 2.0 (M. Wilkinson, University of Bristol) and by hand.

Theory

Majority-rule RCC/PC trees can be defined with respect to a set of fundamental trees by modification of the definitions of their "strict" counterparts (Wilkinson 1994, 1995) so that agreement is required only across a majority of, rather than across all, the fundamental trees. Thus a majority-rule RCC/PC tree is one that satisfies the three conditions: (1) *unambiguity*—that all n -taxon statements/partitions implied by the tree are found in a majority of the fundamental trees; (2) *nonredundancy*—that it includes all n -taxon statements/partitions that apply to all the included taxa in a majority of the funda-

mental trees and is not a subtree of some other tree satisfying (1), and (3) *informativeness*—that at least one informative n -taxon statement/partition is included in the tree (i.e., bushes are prohibited).

For brevity I restrict my treatment to the majority-rule RCC method and do not spell out analogous statements that apply to the majority-rule PC method for unrooted trees. As with strict RCC trees, there may be more than one majority-rule RCC tree for a set of fundamental trees, and the set of such trees constitutes the majority-rule RCC profile. The majority-rule component consensus tree will be a member of the majority-rule RCC profile if it satisfies condition (3) above, i.e., if it is not completely unresolved.

Substituting a specific (and greater than 50%) minimum frequency of replicated n -taxon statements (e.g., 90%) for "a majority" in the above definitions yields the definitions of "stricter" consensus trees and profiles. An informative n -taxon statement must apply to at least three taxa (four for an n -taxon partition) and the profiles will identify and include all informative n -taxon statements occurring with the required minimum frequency. However, phylogeneticists may also wish to restrict their attention to only those n -taxon statements that apply to some minimum number of taxa. Thus, somewhat analogously to specifying a minimum frequency of occurrence of groups for their inclusion in a reduced consensus, majority-rule RCC profiles may be further restricted by specification of a minimum cardinality for the included n -taxon statements. Such a cardinality constraint can be indicated within brackets. Thus, for example, a 50[100]% majority-rule RCC tree will include all n -taxon statements that occur in more than 50% of the fundamental trees and which apply to all (100%) of the taxa, i.e., it will be the majority-rule component consensus tree.

As with strict RCC trees, majority-rule RCC trees may be basic or derivative (Wilkinson 1995). The algorithm described for constructing basic strict RCC profiles (see Wilkinson 1994 for details) can be used to construct the basic majority-rule RCC profiles also, with only minor modification. The algorithm uses an iterative intersection method to identify all nonredundant n -taxon statements that are common to all fundamental trees. For majority-rule RCC trees only the condition of redundancy is modified. An n -taxon statement is redundant if it is entailed by another n -taxon statement of greater cardinality that occurs in equal or greater frequency. The technique of combining basic RCC trees into derivative trees (Wilkinson 1995) can be applied to majority-rule RCC trees without any modification.

Wilkinson (1994) pointed out that strict RCC trees can be ranked on the basis of the degree of resolution (number of nodes) and their cardinality (number of in-

Table 1
Hypothetical Character Data for 11 Ingroup Taxa (A–J, X) and a Single Outgroup Used to Root the Trees

TAXA	CHARACTERS										
	1–5	6–10	11–15	16–20	21–25	26–30	31–35	36–40	41–45	46–50	51–55
Outgroup	00000	00000	00000	00000	00000	00000	00000	00000	00000	00000	00000
A	11111	11111	11111	11111	11111	11111	00000	00000	00000	00000	00000
B	11111	11111	11111	11111	11111	00000	00000	00000	00000	00000	00000
C	11111	11111	11111	11111	00000	00000	00000	00000	00000	00000	00000
D	11111	11111	11111	00000	00000	00000	00000	00000	00000	00000	00000
E	11111	11111	00000	00000	00000	00000	00000	00000	00000	00000	00000
F	11111	00000	00000	00000	00000	00000	11111	00000	00000	00000	00000
G	11111	00000	00000	00000	00000	00000	11111	11111	00000	00000	00000
H	11111	00000	00000	00000	00000	00000	11111	11111	11111	00000	00000
I	11111	00000	00000	00000	00000	00000	11111	11111	11111	11111	00000
J	11111	00000	00000	00000	00000	00000	11111	11111	11111	11111	11111
X	11111	11111	11111	11111	11111	11111	11111	11111	11111	11111	11111

cluded taxa). These features will frequently be negatively correlated because additional resolution is obtained by eliminating problematic taxa. Wilkinson regarded maximally resolved strict RCC trees as of particular interest to phylogeneticists and defined *primary* strict RCC trees as those that maximize resolution and that secondarily maximize cardinality, with other mem-

bers of the profile designated as *secondary*. With majority-rule RCC trees, the frequencies of occurrence or bootstrap proportions for the included nodes represent a complicating factor in determining which consensus trees will be of greatest interest to phylogeneticists and I do not seek to draw this, or any analogous distinction between primary and secondary trees, in the context of majority-rule RCC profiles.

A Hypothetical Example

Table 1 gives hypothetical data designed to illustrate the potential of the majority-rule RCC method. The data support two most-parsimonious trees (MPTs) which differ only in the placement of the “rogue” taxon X (fig. 1a and b). Taxon X is uniquely responsible for all the incongruence in the data and, ignoring taxon X, there is abundant evidence supporting a single unambiguous set of relationships among taxa A–J. The strict component consensus for the two MPTs is a completely unresolved bush that fails to represent any of the considerable agreement between the two MPTs. In contrast, the strict RCC profile comprises the single tree of figure 1c. This achieves its greater sensitivity to the agreement among the MPTs at the cost of the exclusion of the “rogue” taxon X. The exclusion of taxon X is necessary because it cannot be joined to the RCC tree at any point without producing a tree that either misrepresents the agreement among the MPTs or has an ambiguous interpretation (Wilkinson 1994). The strict component consensus is not a member of the profile because it is completely uninformative.

The hypothetical data set was subjected to a bootstrap parsimony analysis. The results of this analysis are predictable. Each replicate is expected to yield one or both of the MPTs supported by the original data, or similar but less resolved trees that result from the chance absence in bootstrap replicate data sets of characters that

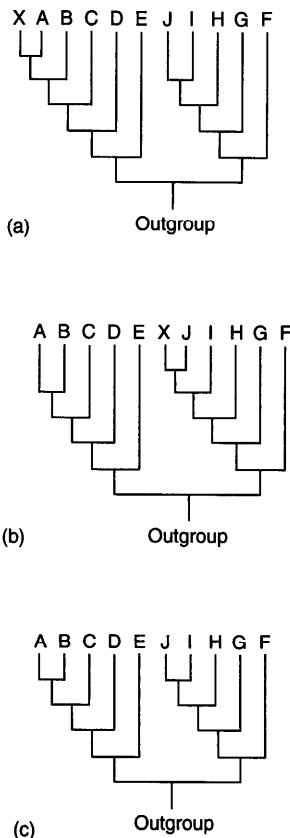


FIG. 1.—Two MPTs for the hypothetical character data of table 1 (a and b) and the single tree in the strict RCC profile (c).

Downloaded from https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/mbe/mz013/3/437/15511030 by National Institute of Standards and Technology user on 11 April 2020

Table 2
Partition Table Showing Frequencies of Components Found in the Bootstrap Analysis

COM- PO- NENT	TAXA										FRE- QUENCY	
	A	B	C	D	E	F	G	H	I	J		X
1*	*	●	●	●	●	●	●	●	●	*	50.5
2*	*	*	*	*	●	●	●	●	●	*	50.5
3*	*	*	*	●	●	●	●	●	●	*	50.5
4*	*	●	●	●	*	*	*	*	*	●	50.5
5●	●	●	●	●	●	●	*	*	*	●	50.5
6*	*	*	●	●	●	●	●	●	●	*	50.0
7●	●	●	●	●	●	●	*	*	*	*	49.5
8●	●	●	●	●	●	●	*	*	*	●	49.5
9*	●	●	●	●	●	●	●	●	●	*	49.5
10*	*	●	●	●	●	●	●	●	●	●	49.5
11*	*	*	*	*	●	●	●	●	●	●	49.5
12*	*	*	*	●	●	●	●	●	●	●	49.5
13●	●	●	●	●	●	●	●	●	*	*	49.5
14●	●	●	●	●	*	*	*	*	*	*	49.5
15●	●	●	●	●	●	*	*	*	*	*	49.0
16●	●	●	●	●	●	*	*	*	*	*	49.0
17●	●	●	●	●	●	●	●	*	*	*	48.5
18*	*	*	●	●	●	●	●	●	●	●	48.0

NOTE.—Taxa marked “*” are more closely related to each other than they are to taxa marked “●”. The outgroup is not shown.

support one or more of the nodes of the original MPTs. Because there are equal numbers of characters supporting the alternative placements of the “rogue” taxon X in the original data, the bootstrap analysis is expected to produce approximately equal numbers of bootstrap MPTs with each of the two major positions of this taxon (i.e., grouping with taxa A–E, or with taxa F–J).

A partition table summarizing the frequencies of components occurring in the bootstrap trees is given in table 2, and the majority-rule component consensus tree of figure 2a provides a conventional graphical summary of the analysis. These results suggest that there is little support for any clade, and, although this conclusion is not in itself misleading, it does not provide a complete picture. If the results were interpreted more loosely as indicating that there is little support for any phylogenetic relationships, or that the data is phylogenetically uninformative, this would be wrong.

The frequencies of occurrence across the bootstrap trees of all nonredundant *n*-taxon statements of lower cardinality than components are summarized in table 3, a modified partition table in which taxa that are not specified in an *n*-taxon statement are indicated by a “?”. Note that the *n*-taxon statements that assert relationships among taxa A–F but which exclude the “rogue” taxon X have high bootstrap proportions indicating a high level of support for these relationships. The bootstrap analysis can be summarized graphically with the basic majority-rule RCC profile for the bootstrap trees. Basic ma-

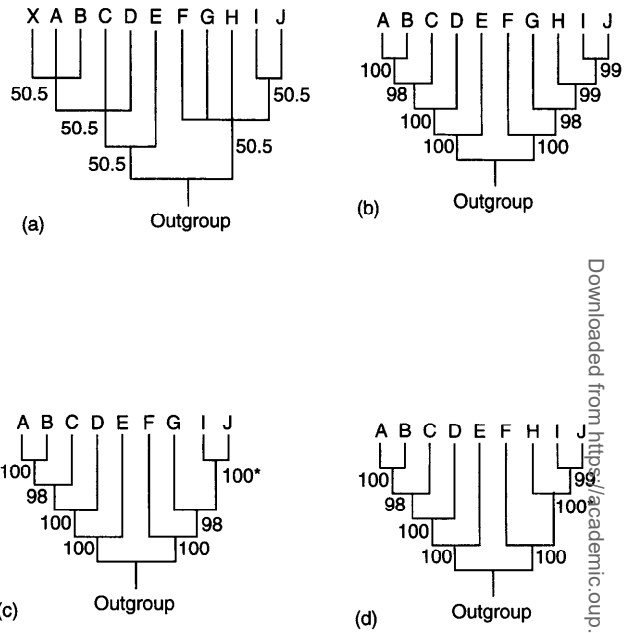


FIG. 2.—The majority-rule RCC profile for the bootstrap analyses of the hypothetical data in table 1. *a*, The majority-rule component consensus. *b*, A single 90[90]% majority-rule RCC tree. *c* and *d*, Two additional majority-rule RCC trees that complete the profile. Asterisks indicate those bootstrap proportions that are higher than the corresponding values in *b*.

majority-rule RCC trees are constructed by combining all *n*-taxon statements that (1) are informative for the same set of taxa, (2) have a frequency of occurrence greater than 50% (or some specified higher frequency), and (3) exceed or equal the minimum specified cardinality. With no cardinality constraint the 50% majority-rule RCC profile includes the majority-rule component consensus (fig. 2a) and three other basic majority-rule RCC trees (fig. 2b–d). All of the latter exclude the “rogue” taxon

Table 3
Partition Table Showing Frequencies of Nonredundant and Informative *n*-taxon Statements Found in the Bootstrap Analysis in Addition to Those in Table 2

STATE- MENT	TAXA										FRE- QUENCY		
	A	B	C	D	E	F	G	H	I	J		X	
19*	*	●	●	●	●	●	●	●	●	●	?	100
20*	*	*	*	●	●	●	●	●	●	●	?	100
21*	*	*	*	*	●	●	●	●	●	●	?	100
22●	●	●	●	●	*	*	*	*	*	*	?	100
23●	●	●	●	●	●	●	?	*	*	*	?	100
24●	●	●	●	●	●	?	*	*	*	*	?	100
25●	●	●	●	●	●	●	*	*	*	*	?	99
26●	●	●	●	●	●	●	●	*	*	*	?	99
27*	*	*	●	●	●	●	●	●	●	●	?	98
28●	●	●	●	●	●	*	*	*	*	*	?	98

NOTE.—Format as in table 2 with “?” denoting taxa not included in the *n*-taxon statement.

X and record much higher bootstrap proportions for the relationships among the remaining taxa. The profile supports the accurate interpretation that, although no components (clades or clusters) are well supported, relationships among taxa A–J are well supported and the data is far from uninformative.

Comparing the four majority-rule RCC trees in the profile it can be seen that these trees represent alternative resolutions of a trade-off between the number of taxa included in the consensus tree and the bootstrap proportions for the hypotheses of relationships among the remaining taxa. Higher bootstrap proportions are bought at the cost of the exclusion of more taxa. Note also that the increase in bootstrap proportions achieved through the elimination of the “rogue” taxon X is dramatic whereas the increase achieved through the additional elimination of either A or B (indicated by asterisks in fig. 2c and d) is negligible.

An important practical problem is that majority-rule RCC profiles may contain very many consensus trees that differ only slightly in bootstrap proportions with the elimination of different combinations of taxa. Use of high minimum frequencies and/or high cardinality constraints may provide a means of getting further information from bootstrap analyses without being swamped by alternative trees. For example, the 90[90]% majority-rule RCC profile for the hypothetical example includes just the tree in figure 2b. An alternative strategy for ameliorating this problem is to make an a posteriori selection of a subset of trees from the profile based on consideration of whether the increments in bootstrap proportions achieved by further elimination of taxa are sufficient to warrant the reporting of different trees.

Use of Partition Tables

There are quite severe limitations on the numbers of taxa and numbers of trees that can be analyzed by the majority-rule RCC method as implemented in the REDCON 2.0 program. Furthermore, the method is not implemented in widely used phylogenetic inference software packages such as PAUP and PHYLIP (Felsenstein 1993). However, the partition tables output by these latter programs summarizing the frequencies of components (i.e., *n*-taxon statements of maximum cardinality) can be used to further explore bootstrap proportions for *n*-taxon statements/partitions of lower cardinality such as those that can be represented in majority-rule RCC trees. If two components are incompatible then they cannot co-occur on the same bootstrap tree. Therefore, any *n*-taxon statements/partitions of lower cardinality that are entailed by both incompatible components will occur in all those trees that include either component, and their bootstrap proportions will be equal to or greater than the

Table 4
Partition Table Showing Frequencies of Nonredundant and Informative *n*-taxon Statements Formed from the Intersections of Pairs of Incompatible Components of Table 2

STATE- MENT	SOURCE OF STATE- MENT	TAXA										FRE- QUENCY ^a	
		A	B	C	D	E	F	G	H	I	J		X
29 1 + 12	*	*	?	?	●	●	●	●	●	●	?	100
30 1 + 18	*	*	?	●	●	●	●	●	●	●	?	98.5
31 3 + 11	*	*	*	?	●	●	●	●	●	●	?	100
32 4 + 7	●	●	●	●	●	?	?	*	*	*	?	100
33 4 + 15	●	●	●	●	●	?	*	*	*	*	?	99.5
34 6 + 12	*	*	*	?	●	●	●	●	●	●	?	99.5
35 7 + 16	●	●	●	●	●	●	?	*	*	*	?	98.5
36 8 + 17	●	●	●	●	●	●	●	?	*	*	?	98

NOTE.—Format as in table 3.
^a Lower bound.

sum of the bootstrap proportions of the incompatible components.

This information can be obtained by a simple intersection method. Following the symbolism of Wilkinson (1994, 1995) any *n*-taxon statement can be written (A)B where A and B are sets of taxa, and the taxa in A are more closely related to each other than they are to any taxa in B. For unrooted partitions, A/B denotes the two sets of taxa without any implication of which are more closely related in the cladistic (rooted) sense. Given any pair of incompatible statements (A₁)B₁ and (A₂)B₂ with frequencies **f**₁ and **f**₂ respectively, these entail a third statement (A₃)B₃ with frequency **f**₃ where A₃ = A₁ ∩ A₂, B₃ = B₁ ∩ B₂ and **f**₃ = **f**₁ + **f**₂.

Similarly, for a pair of incompatible partitions A₁/B₁ and A₂/B₂ with frequencies **f**₁ and **f**₂ respectively, these entail the two *n*-taxon partitions A₃/B₃ with frequency **f**₃ and A₄/B₄ with frequency **f**₄ where A₃ = A₁ ∩ A₂, B₃ = B₁ ∩ B₂, A₄ = A₁ ∩ B₂, B₄ = B₁ ∩ A₂ and **f**₃ = **f**₄ = **f**₁ + **f**₂.

Table 4 is a partition table summarizing informative *n*-taxon statements that can be constructed from the intersections of pairs of incompatible components in table 2. For example, statements 1 and 12 of table 2 both entail statement 29 of table 4, and, because they are incompatible and cannot occur on the same tree, the sum of their frequencies provides a lower bound for the frequency of the statement that they both entail. As with the application of the majority-rule RCC method, this approach identifies *n*-taxon statements that exclude the “rogue” taxon X and have high bootstrap proportions.

One limitation of this approach is that any *n*-taxon statements formed from the intersections of components must exclude a minimum of two taxa so that the method cannot be used to identify *n*-taxon statements that exclude only a single taxon. Thus although partition tables,

compatibility analysis, and intersection rules provide a simple method for exploring the results of bootstrapping, they provide a more incomplete assessment than is possible using the majority-rule RCC method. However, this approach can be used to identify cases where application of the majority-rule RCC method is likely to prove useful.

Taxon Pruning

PAUP allows taxa to be pruned from trees that are stored in memory. Pruning taxa has no effect upon the relationships among the remaining taxa, but it may render initially distinct trees identical and it can be conceived of as producing trees that are summaries of n -taxon statements/partitions of lower cardinality than the original components/partitions. Thus pruning taxa from bootstrap trees and then constructing majority-rule component consensus trees for the pruned trees can be used as a surrogate for the majority-rule RCC method.

However, this approach has a number of difficulties. In order to explore the effects of pruning different taxa, or different combinations of taxa, it is necessary to store the bootstrap trees in a file and repeatedly reload the trees. This may be time consuming but, more importantly, PAUP may compute a majority-rule component consensus for trees read into memory from a file in a way that is inappropriate for bootstrap trees. Analysis of a single bootstrap replicate data set may produce one or multiple bootstrap trees. If x multiple trees are produced during a bootstrap analysis PAUP treats each as worth $1/x$; i.e., a component would need to be included in all x trees to contribute the same to bootstrap proportions as a component of a unique bootstrap tree. However, when trees are read in from a file their relation to the bootstrap replicate data sets is lost and all trees are treated equally. Thus the majority-rule component consensus trees and partition tables produced after pruning will only approximate bootstrap proportions.

By way of example, if the problematic taxon X is pruned from the bootstrap trees for the hypothetical data, the majority-rule "component" consensus of the pruned trees is equivalent in topology to the 90[90]% majority-rule RCC tree of figure 2b, but there are minor differences in the "bootstrap proportions." To be practical, the pruning of taxa should be directed rather than just "trial and error." Choice of candidate taxa for pruning might usefully be informed by examination of partition tables as described in the previous section.

A Real Example

Hedges, Nussbaum, and Maxson (1993) recently presented the first phylogenetic analysis of DNA sequence data for caecilian amphibians. Their alignment of partial 12S and 16S mitochondrial rRNA gene se-

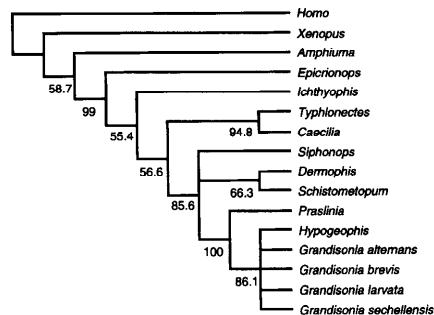


FIG. 3.—The majority-rule component consensus providing a conventional summary of the bootstrap analysis of Hedges, Nussbaum, and Maxson's (1993) caecilian sequence data.

quences included some 329 sites that are informative under parsimony, 13 caecilian taxa, plus a salamander (*Amphiuma*), a frog (*Xenopus*) and us (*Homo*) used as outgroups. My parsimony bootstrap analysis of this data, summarized in the conventional manner with a majority-rule component consensus tree (fig. 3), identifies very little strong support. Only three components are supported by bootstrap proportions in excess of 90%. These identify the clade including all caecilians, the seychellean clade (*Prasinia*, *Hypogeophis*, and *Grandisonia*), and the grouping of *Typhonectes* and *Caecilia* as well-supported components.

The majority-rule RCC profile includes the majority-rule component consensus tree (fig. 2a) and a further 81 basic majority-rule RCC trees that achieve some higher bootstrap proportions at the expense of the elimination of taxa. Many of these trees show only slight differences, and I restricted my attention to those trees that (1) excluded no more than two taxa (32 basic trees satisfy this cardinality constraint), and (2) included bootstrap proportions in excess of 90% for one or more n -taxon statements that have bootstrap proportions lower than this in the majority-rule component consensus.

There are three basic majority-rule RCC trees (fig. 4) satisfying these constraints. The first (fig. 4a) excludes two of the seychellean taxa (*Grandisonia brevis* and *Hypogeophis*) and reveals a bootstrap proportion of 93% for the hypothesis that the remaining species of *Grandisonia* are more closely related to each other than to *Prasinia* or any of the nonseychellean caecilians. The positions of the excluded taxa must be more unstable across the bootstrap trees and more uncertain. Interestingly, the three remaining species of *Grandisonia* have similar morphologies whereas the excluded taxa are morphologically dissimilar from them and from each other. Relationships among these species are difficult to determine using morphological data (R. A. Nussbaum, personal communication) and now also appear to be resisting the, as yet limited, application of molecular data.

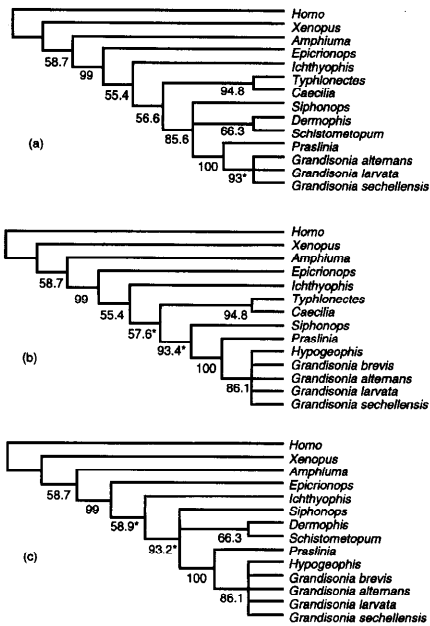


FIG. 4.—Three majority-rule RCC trees revealing additional strong bootstrap support for *n*-taxon statements not apparent from the majority-rule component consensus. Asterisks indicate those bootstrap proportions that are higher than the corresponding values in figure 3.

The second tree (fig. 4b) excludes *Dermophis* and *Schistometopum* and reveals a bootstrap proportion of 93.4% for the hypothesis that the neotropical *Siphonops* is more closely related to the Seychelles clade than to the remaining caecilians. Elimination of *Siphonops* produces only a negligible corresponding increment in the bootstrap proportion for the association of *Dermophis*, *Schistometopum*, and the Seychelles clade (to 85.8% from 85.6%). Thus the positions of *Dermophis* and *Schistometopum* must be more variable than the position of *Siphonops* across the bootstrap trees, and thus more uncertain, a conclusion that cannot be drawn from the conventional majority-rule component consensus. Elimination of these taxa also produces a negligible increment in the bootstrap proportion for the hypothesis that *Ichthyophis* and *Epicrionops* are outside a clade comprising all other included caecilians, a hypothesis that is well supported by morphological data (Nussbaum 1979; Nussbaum and Wilkinson 1989).

The third tree (fig. 4c) excludes *Typhlonectes* and *Caecilia*, one of the three groups united by high (>90%) bootstrap proportions in the majority-rule component consensus tree. Their exclusion reveals a high bootstrap proportion (93.2%) for the group including all remaining caecilians except *Ichthyophis* and *Epicrionops*, in accord with the morphological data. There is also a negligible increment in the bootstrap proportions for the hypothesis that rhinatrematid *Epicrionops* is the sister-taxon of the remaining caecilians, a relationship that is also well sup-

ported by available morphological data (Nussbaum 1977; Wilkinson 1992). Clearly the more basal relationships within the caecilian tree are only poorly resolved by Hedges, Nussbaum, and Maxson's limited sequence data, and much of the uncertainty concerns the affinities of the *Typhlonectes*-*Caecilia* clade.

Discussion

Bootstrapping is very widely used to evaluate support for phylogenetic inferences, especially in molecular phylogenetics, and it is encouraged in the instructions to authors of manuscripts for *Molecular Biology and Evolution*. Inasmuch as bootstrapping can be usefully applied to the evaluation of particular components (clades) or partitions, so it can be applied to other elements of complex phylogenetic hypotheses that cannot be represented by components or partitions. This extension of bootstrapping requires no modification of the basic bootstrapping protocol, only the use of a suitable consensus method for summarizing the results. The hypothetical example illustrates the potential utility of bootstrapping with a majority-rule RCC consensus. Given only the majority-rule component consensus of figure 2a, a conclusion that no phylogenetic relationships are well supported by the data and that the data probably contain little or no "phylogenetic signal" might seem reasonable but would be incorrect. The majority-rule RCC profile demonstrates that the data provide strong support for an unambiguous set of relationships among taxa A-F.

Applied to real problems of phylogenetic inference, the potential for the majority-rule RCC profile to include very many trees, some of which differ but little, is a major practical drawback. This difficulty may be ameliorated by specifying high frequency and/or cardinality constraints and/or by a posteriori selection of trees from among the profile. Further work is needed to replace such a posteriori selection with a priori criteria for selecting from among majority-rule RCC trees those that will be most useful to phylogeneticists. Such criteria will have to incorporate decisions as to what level of improvement in bootstrap proportions that can be achieved by eliminating taxa is worthwhile and what can be considered negligible.

Uncertainty in phylogenetic inferences must sometimes be associated with the phylogenetic position of some subset of the taxa, with the data providing strong support for an unambiguous set of relationships among the remaining taxa. Whenever this is the case, summarizing bootstrapping with the majority-rule component consensus method may provide a very incomplete and potentially misleading assessment of the data and associated phylogenetic hypotheses. Majority-rule reduced

Downloaded from https://academic.oup.com/iob/advance-article-abstract/doi/10.1093/iob/obz010/5511010 by Uppsala University user on 05 September 2020

consensus methods provide phylogeneticists with sharper tools for dissecting complex phylogenetic relationships and investigating their differential support. I hope they will be harnessed to the bootstrapping procedures available in widely used phylogenetic software. Until such a time, REDCON 2.0 can be used to help construct strict and majority-rule RCC trees. Alternatively, the surrogate methods exploiting pruned trees and/or partition tables can be used to further investigate phylogenetic hypotheses.

Acknowledgments

I thank F. R. McMorris, R. Powers, R. Hirt and T. M. Embley for the opportunity to discuss these ideas, and R. A. Nussbaum for discussions of caecilian phylogeny. This work was supported in part by NERC grant GST/02/832. REDCON 2.0 is a PC compatible program that is available from the author upon receipt of a disk.

LITERATURE CITED

- FELSENSTEIN, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**:783–791.
- . 1993. PHYLIP, version 3.5. University of Washington, Seattle.
- HEDGES, S. B., R. A. NUSSBAUM, and L. R. MAXSON. 1993. Caecilian phylogeny and biogeography inferred from mitochondrial DNA sequences of the 12S and 16S rRNA genes (Amphibia: Gymnophiona). *Herpetol. Monogr.* **7**:64–76.
- MARGUSH, T., and F. R. MCMORRIS. 1981. Consensus *n*-trees. *Bull. Math. Biol.* **43**:239–244.
- NUSSBAUM, R. A. 1977. Rhinatrematidae: a new family of caecilians (Amphibia: Gymnophiona). *Occas. Pap. Mus. Zool. Univ. Mich.* **682**:1–30.
- . 1979. The taxonomic status of the caecilian genus *Uraeotyphlus* Peters. *Occas. Pap. Mus. Zool. Univ. Mich.* **687**:1–20.
- NUSSBAUM, R. A., and M. WILKINSON. 1989. The classification and phylogeny of caecilians (Amphibia: Gymnophiona): a critical review. *Herpetol. Monogr.* **3**:1–42.
- SWOFFORD, D.L. 1993. PAUP: phylogenetic analysis using parsimony, version 3.1.1. Computer program distributed by the Illinois Natural History Survey, Champaign.
- WILKINSON, M. 1992. The phylogenetic position of the Rhinatrematidae (Amphibia: Gymnophiona), evidence from the larval lateral line system. *Amphibia-Reptilia* **13**:74–79.
- . 1994. Common cladistic information and its consensus representation: reduced Adams and reduced cladistic consensus trees and profiles. *Syst. Biol.* **43**:343–368.
- . 1995. More on reduced consensus methods. *Syst. Biol.* **44**:436–440.

STANLEY SAWYER, reviewing editor

Accepted October 31, 1995

Downloaded from https://academic.oup.com/mbe/article/13/3/437/1057310 by U.S. Department of Justice user on 17 August 2022