

Sphenodontid phylogeny and the problems of multiple trees

MARK WILKINSON¹ AND MICHAEL J. BENTON²

¹*School of Biological Sciences, University of Bristol, Bristol BS8 1UG, U.K.*

²*Department of Geology, University of Bristol, Bristol BS8 1RJ, U.K.*

SUMMARY

Application of optimality criteria, such as parsimony, for choosing among phylogenetic hypotheses sometimes yields multiple optimal solutions, necessitating the application of additional methodology. Phylogeneticists may adopt a conservative strategy of presenting only a summary of the multiple trees and/or they may attempt to further discriminate among them. Parsimony analysis of data for *Sphenodon* and its fossil relatives yields multiple trees and is used to illustrate and compare a variety of approaches for summarizing agreement and selecting among trees. Multiple trees may result from a lack of evidence or character incongruence. Only with the latter can techniques to discriminate among optimal trees be effective. Adoption of secondary optimality criteria depends on the same rationale as differential character weighting and is a poor substitute for weighting. Neither approach will necessarily solve the problem of multiple trees or remove the need for methods that summarize agreement among them. Many consensus methods produce summaries of optimal trees that are far from ideal. Reduced consensus methods and disqualifier faithful subtrees may have advantages over the alternatives. Safe taxonomic reduction may also reduce problems of multiple trees due to lack of evidence, and the experimental pruning of taxa provides a powerful technique for investigating agreement among multiple trees. Some consensus and weighting methods are sensitive to the treatment of arbitrary resolutions which presents some unsolved difficulties in their use and interpretation. Multiple analyses using modern techniques provide a more resolved hypothesis of sphenodontid interrelationships and a better understanding of remaining problems than previous studies.

1. INTRODUCTION

For better or worse, parsimony has become the most widely used criterion for choosing among phylogenetic hypotheses, particularly in studies based upon morphological data. In common with some other methods, parsimony analyses may identify multiple optimal hypotheses. Given that the aim of phylogenetic inference is to reconstruct the singular phylogenetic history, multiple most parsimonious trees (MPTs) pose problems that require additional methodology, and there are two main strategies. Firstly, a further selection from among the MPTs may be made. Such a selection will be arbitrary unless based upon secondary optimality criteria or differential weighting strategies and there is a small but growing literature concerned with such techniques (Carpenter 1988; Sharkey 1989; Rodrigo 1992; Sang 1995). Secondly, relationships that are common to all the *fundamental* trees can be represented by one or more consensus trees. However, consensus trees are often poorly resolved and the degree of resolution of consensus trees is often, though not universally, correlated with the number of MPTs. Thus phylogeneticists have sometimes attempted to increase the resolution of consensus trees by reducing the numbers of fundamental trees. If the multiple MPTs reflect alternative placements of particularly unstable

or 'underdetermined' taxa, removing such taxa and analysing only the interrelationships of the better known taxa, an approach termed 'taxonomic reduction' (Wilkinson 1995*a*), can sometimes lead to dramatic reductions in the number of trees. For example, in an analysis of rhynchosaur interrelationships Wilkinson & Benton (1995) were able to reduce the number of MPTs from over 20000 to just two by elimination of problematic taxa.

The problem of multiple MPTs present phylogeneticists with a range of methodological options and choices. They must decide which of the general approaches they will adopt and they must further decide among the variant methods available for implementing these approaches. In some cases, especially the use of consensus methods and the application of successive approximations character weighting, the range of variants and alternatives is potentially bewildering. Ideally the choices made in practice should be well informed and based on critical evaluation of the alternatives. However, there have been few comparative studies of these alternatives, so that much remains to be learnt about them and there is likely to be scope for further methodological developments.

Fraser & Benton (1989) attempted to resolve the relationships of the famous 'living fossil' *Sphenodon* and

various fossil sphenodontids. Parsimony analyses of their complete data yielded numerous MPTs which they summarized with a strict component consensus tree. Although they were able to resolve the more basal branches of the sphenodontid tree this was not possible within a crown group of eight taxa. Subsequent analyses after removal of the most incompletely known fossil taxa also failed to resolve crown-group relationships. However, Fraser & Benton's failure to resolve these relationships is not due entirely to deficiencies in their data. In addition, the consensus method and strategy of taxonomic reduction employed have deficiencies that limit their suitability to the task. Here we use Fraser & Benton's (1989) data with the aim of resolving more fully those relationships within the crown group that are unambiguously supported by the parsimonious interpretation of the data. We also use these data to illustrate some of the problems of multiple MPTs and to compare, evaluate, and further develop methods that can be applied to these problems.

2. MATERIALS AND METHODS

All parsimony analyses were done with PAUP 3.1.1 (Swofford 1993), using the exact branch and bound method that is guaranteed to find all MPTs (Hendy & Penny 1982). PAUP was used to generate Adams consensus trees as well as strict, semi-strict and majority-rule component consensus trees and to prune taxa from trees. COMPONENT 2.0 (Page 1992) was used to construct a greatest agreement subtree and Nelson consensus trees. MacClade 3.0.1 (Maddison & Maddison 1992) was used to examine the effects of selected manipulations of tree topologies on tree statistics. REDCON was used to produce a reduced cladistic consensus profile, TAXEQ2 was used to identify taxa that could be deleted under the constraints of safe taxonomic reduction, and PICA to rank characters through 'boil-downs' using coefficients of character state randomness and character normal deviates. REDCON, TAXEQ2, and PICA are DOS programs written by the senior author and available from him upon receipt of a disk. A reduced Adams

consensus tree, and a disqualifier faithful tree were constructed by hand. Explanations of some terms are provided in table 2, with some common consensus methods summarized in table 3. Further details of the phylogenetic analyses are presented below. All reference to 'the original analyses' is to Fraser & Benton's (1989) parsimony analysis. Consensus terminology follows that of Wilkinson (1994a).

3. REVIEW OF THE ORIGINAL ANALYSES

Fraser & Benton's (1989) data comprise 29 binary characters and 14 ingroup taxa plus the plesiomorphic lepidosauromorph *Youngina* which was used to root the tree (see table 1). Fraser & Benton analysed the data with both parsimony and clique methods. Here we are concerned only with the former. Results of clique analysis are similar, but are complicated by problems caused by missing data in the discovery of cliques of mutually compatible characters (Wilkinson 1994b).

Fraser & Benton's parsimony analysis, using PAUP 2.4.1 (Swofford 1985), was performed in four stages. A first analysis included all taxa and yielded 82 MPTs which they summarized with a strict component consensus tree (see figure 1a). Although many relationships are resolved in this consensus tree, relationships within a crown group of eight taxa are poorly resolved. Subsequent analyses excluded taxa with: (i) less than 50%; (ii) less than 80%; and (iii) less than 100% of their character states known (see figures 1b-d). Only the last of these analyses provided any additional resolution of the crown group, indicating that *Sphenodon* is more closely related to *Clevosaurus* than to *Palaeopleurosaurus*.

4. REANALYSES

If arbitrary resolutions (zero-length branches) are not suppressed, analysis of the complete data yields 82 binary (i.e. dichotomous or fully resolved) MPTs with lengths of 33 steps, consistency indices of 0.879 (0.84 ignoring uninformative characters) and retention in-

Table 1. Character data for 14 ingroup sphenodontid taxa and a single outgroup (*Youngina*)

taxa	characters					
	1-5	6-10	11-15	16-20	21-25	26-29
<i>Youngina</i>	00000	00000	00000	00000	00000	0000
<i>Gephyrosaurus</i>	111?0	10000	00000	00000	00000	000?
<i>Diphydontosaurus</i>	11111	11100	00000	00000	00000	000?
<i>Planocephalosaurus</i>	11111	11101	01100	00000	00000	0000
<i>Polysphenodon</i>	111?1	0?111	1?11?	11000	?0000	00?1
<i>Homoosaurus</i>	11111	01111	1111?	11000	00000	?011
<i>Brachyrhinodon</i>	111?1	0?111	1?11?	11100	0?000	00?0
<i>Clevosaurus</i>	11111	01111	11111	11110	00000	0010
<i>Kallimodon</i>	11111	01111	1111?	10111	00000	?010
<i>Sapheosaurus</i>	?111?	0??1?	?????	?1111	000?0	????
<i>Piocormus</i>	?1111	0????	?????	?1111	0?0??	???0
<i>Palaeopleurosaurus</i>	11111	01111	11110	10111	11100	0010
<i>Sphenodon</i>	11111	01111	11111	11111	00011	1000
<i>Eilenodon</i>	11???	?1???	11???	?????	???11	?10?
<i>Toxolophosaurus</i>	11???	?1???	11???	?????	???1?	?10?

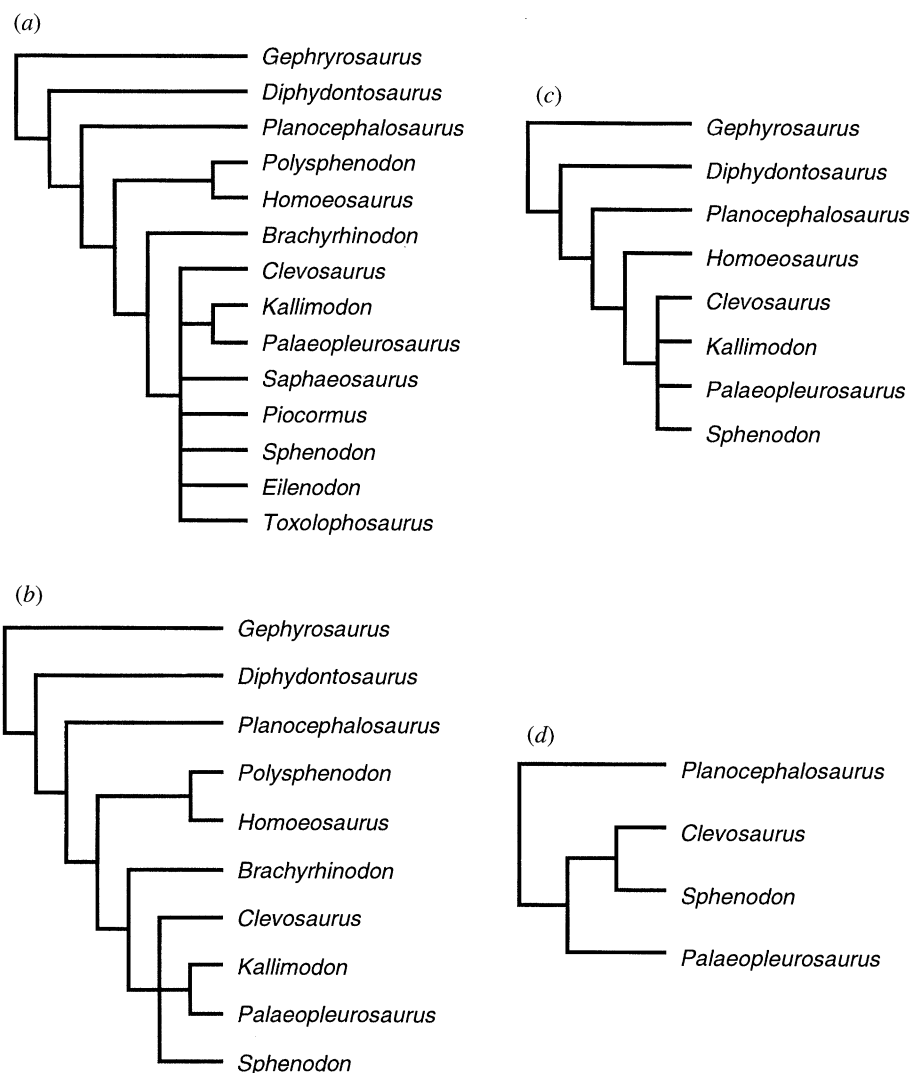


Figure 1. Strict component consensus trees (a–c) and a unique MPT (d) from Fraser & Benton’s (1989) original parsimony analysis.

Table 2. Glossary of terms used to describe relations among taxa in a tree

term	explanation
<i>N</i> -taxon statement	a statement that some specified taxa are more closely related to each other than to some other specified taxa
component (clade or cluster)	an <i>n</i> -taxon statement that applies to all taxa included in a tree
disqualifier	the negation of an <i>n</i> -taxon statement
subtree nesting	a set of taxa (X) are a subset nesting of a more inclusive set (Y) if X have a common ancestral node that is further from the root than the last common ancestral node of Y

dices of 0.922. These are the same trees as found in the original analysis of the complete data. If arbitrary resolutions are collapsed, allowing trees including polytomies, the number of distinct MPTs is reduced to 60. Consensus trees were constructed for both sets of MPTs (including or excluding arbitrary resolutions) unless the nature of the consensus method entails that consensus trees from both sets of MPTs must be identical (see table 3).

(a) **Strict component consensus (scc)**

Relationships within the crown group (see figure 1a) are poorly resolved, but the number of MPTs is far fewer than the number of possible resolutions of the crown group, suggesting that the MPTs include only a small subset of the possible relationships and may share much additional structure that is not reflected in the scc tree.

Table 3. *A summary of consensus methods and their relation to the fundamental trees*

consensus method	description
strict component ^a (Sokal & Rohlf 1981)	all and only those components that occur in all fundamentals.
semi-strict component ^b (Hillis 1987) ^d	all and only those uncontradicted components that occur in any fundamental.
majority-rule component ^b (Margush & McMorris 1981)	all and only those components that occur in more than 50% of the fundamentals.
Nelson ^b (Nelson 1979; Page 1989)	the intersection of all cliques of components that maximize the sum of their frequency of representation in the fundamentals.
Adams ^a (Adams 1972, 1986)	all and only those subset nestings that occur in all fundamentals.
reduced Adams ^{ac} (Wilkinson 1994a)	only <i>n</i> -taxon statements that occur in all fundamentals. Formed by pruning taxa from ambiguous polytomies in Adams consensus trees.
reduced cladistic ^{ac} (Wilkinson 1994a, 1995b)	all and only those <i>n</i> -taxon statements that occur in all fundamentals.
disqualifier faithful subtree ^{ac} (Wilkinson 1994a, 1995b)	all and only those <i>n</i> -taxon statements and disqualifiers common to a set of fundamentals that apply specifically to the set of taxa included in the subtree.
greatest agreement subtree ^{abc} (Kubicka <i>et al.</i> 1995) ^e	subtree formed by pruning the fewest taxa required to bring all fundamentals into agreement.

^a Strict in the sense of requiring agreement across all fundamentals.

^b Sensitive to the treatment of arbitrary resolutions.

^c May produce multiple consensus trees (a profile).

^d Also known as combinable component consensus (Bremer 1990).

^e Also known as largest common pruned tree (Gordon 1980).

(b) *Semi-strict component consensus* (sscc)

For the set of 60 MPTs the sscc tree includes a single additional group, comprising *Eilenodon* and *Toxolophosaurus*, that is not present in the scc tree. This group appears in 52 (87%) of the trees. For the set of 82 binary MPTs the sscc tree is identical to the corresponding scc tree (see figure 1a). This implies that the grouping of *Eilenodon* and *Toxolophosaurus* is contradicted in one or more of the 82 MPTs and that these contradictions are due to arbitrary resolutions that are suppressed in the set of 60 MPTs. It also illustrates Wilkinson's (1995c) observation that the sscc method may be sensitive to the treatment of arbitrary resolutions.

(c) *Majority-rule component consensus* (mrcc)

The mrcc tree for the 60 MPTs is incompletely resolved, but includes the grouping of *Eilenodon* and *Toxolophosaurus* found in the corresponding sscc tree, and also includes *Sphenodon* as their sister-taxon and *Piocormus* as the sister-taxon of these three. PAUP provides the potential to increase resolution in mrcc trees by including compatible minority components. For the 60 MPTs, the augmented mrcc tree includes two minority components and is fully resolved (see figure 2a). Applying this method to the 82 binary MPTs produces results that differ slightly in the frequencies with which particular components are encountered and more substantially in which compatible minority components are included in the consensus tree (see figure 2b). Compare the positions of *Clevosaurus* and *Sapheosaurus* in the two trees of figure 2. Thus, like the

sscc method, the mrcc method may be sensitive to the treatment of arbitrary resolutions (Wilkinson 1995c).

(d) *Nelson consensus* (nc)

NC trees are identical to the corresponding mrcc trees with compatible minority components (see figure 2). Thus the nc method is also sensitive to the treatment of arbitrary resolutions.

(e) *Adams consensus* (ac)

AC trees are topologically sensitive to kinds of agreement between fundamental trees that may not be reflected in component consensus trees, but any polytomies are ambiguous and require careful interpretation (Wilkinson 1994a). The ac tree for the sphenodontid MPTs (see figure 3) includes two groupings (*Eilenodon* and *Toxolophosaurus*; and these two plus *Sphenodon*) that are not present in the scc tree. Because these three taxa are split by a dichotomy at their last common ancestral node, we can conclude that *Eilenodon* and *Toxolophosaurus* are always more closely related to each other than they are to *Sphenodon* in all the MPTs, but not that they are more closely related to each other than to any of the other taxa or groups of taxa arising from the crown group polytomy. Nonetheless, the ac tree indicates that there is more agreement across all the fundamental trees than is apparent from the scc tree.

(f) *Reduced Adams consensus (RAC)*

The RAC profile for the sphenodontid MPTs includes two trees (see figure 4) that are produced by pruning either *Piocormus* (see figure 4a) or *Sphenodon*, *Eilenodon* and *Toxolophosaurus* (see figure 4b) from the AC tree (see

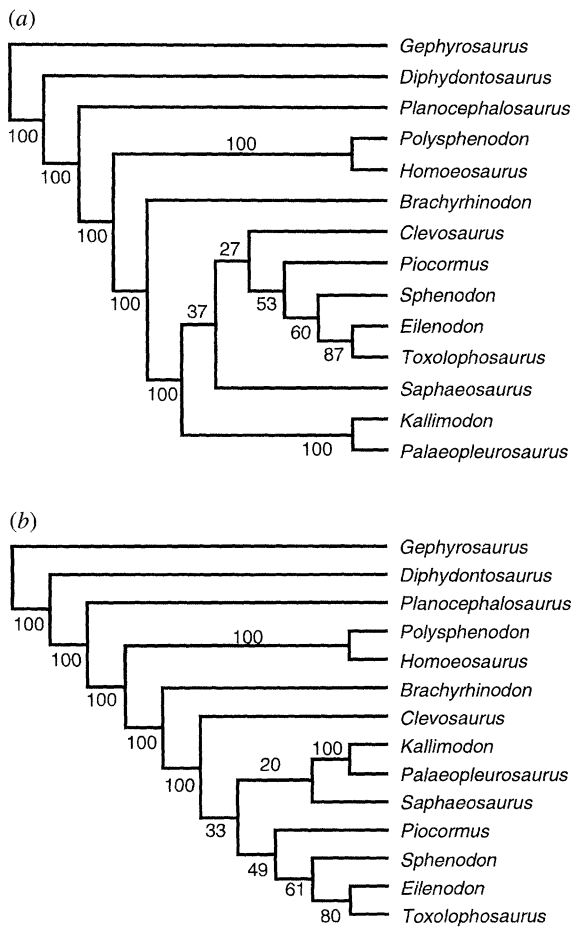


Figure 2. Majority-rule component consensus trees with compatible minority components for the 82 binary sphenodontid MPTs (a) and the 60 MPTs in which arbitrary resolutions are suppressed.

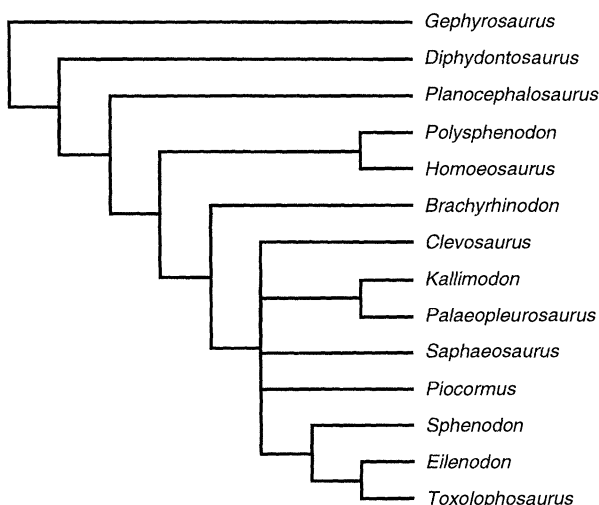


Figure 3. Adams consensus tree for the sphenodontid MPTs.

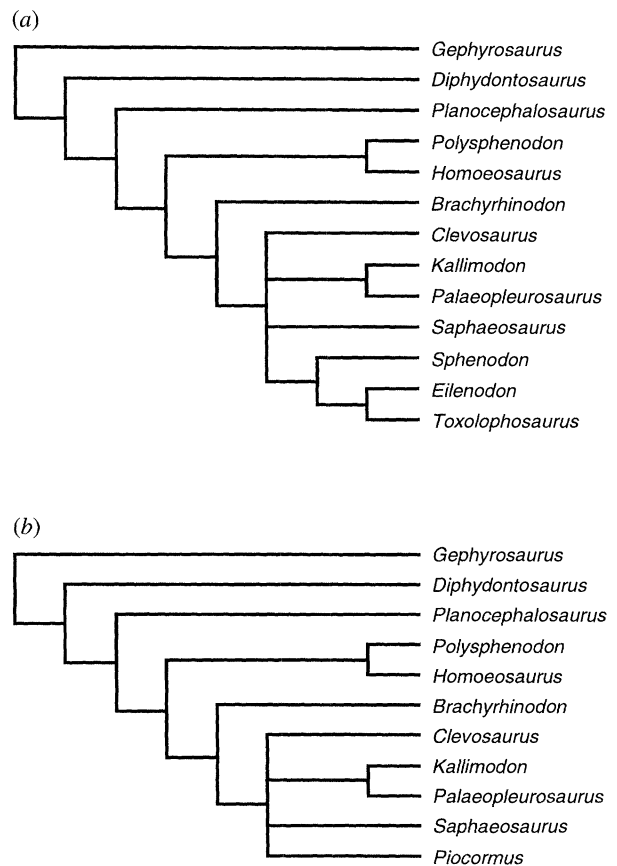


Figure 4. Reduced Adams consensus profile for the sphenodontid MPTs.

figure 3). The tree in figure 4a is a *primary* RAC tree as it contains more nodes than the *secondary* tree (see figure 4b) and it indicates that all the MPTs agree that *Sphenodon*, *Eilenodon* and *Toxolophosaurus* on the one hand, and *Kallimodon* and *Paleopleurosaurus* on the other, are more closely related to each other than they are to all other crown-group sphenodontids with the exception of *Piocormus*. The secondary RAC tree indicates that, in addition, *Kallimodon* and *Palaeopleurosaurus* are more closely related to each other than to *Piocormus* in all the MPTs, as was evident from the scc tree. Taken together, the two RAC trees imply that *Piocormus* groups within the *Sphenodon-Eilenodon-Toxolophosaurus* assemblage in some of the MPTs.

Although the AC tree might be taken as indicating these relationships also, they are not legitimate inferences from the AC tree alone: the same AC tree could be produced from sets of fundamental trees where these relationships do not hold across all the trees. The relationships of *Kallimodon* and *Palaeopleurosaurus* are represented in the scc, whereas those of *Sphenodon*, *Eilenodon* and *Toxolophosaurus* are not because *Piocormus* sometimes, but not always, clusters within this group. The variable position of *Piocormus* makes its representation in a consensus tree problematic. The scc includes *Piocormus*, but fails to represent the relationships of *Sphenodon*, *Toxolophosaurus* and *Eilenodon*, whereas the AC tree includes *Piocormus* but this produces ambiguity in the legitimate interpretation of the relationships of both it and other taxa. The RAC trees

solve these problems by eliminating taxa that cannot be represented on the same consensus tree without either ambiguity or a loss of resolution.

(g) *Strict-reduced cladistic consensus (RCC)*

Wilkinson (1994*a*) described a case where the RCC profile included common *n*-taxon statements that are not represented in the RAC profile. In the present case the profile includes the SCC tree (see figure 1*a*) and a single additional more fully resolved, and therefore primary, RCC tree that is identical to the primary RAC tree (see figure 4*b*). Note that the secondary RAC tree (see figure 4*a*) is a redundant subtree of the secondary RCC tree (see figure 1*a*). The RCC profile does not include any information that is not also included in the RAC profile and differs only in how this information is represented.

(h) *Disqualifier faithful tree or subtree (DFS)*

An ideal consensus method would provide a summary of all common disqualifiers as well as all common *n*-taxon statements. This ideal has yet to be realised with a general method, but it may be approximated, if not achieved, through the use of DFSS (Wilkinson 1994*a*, 1995*b*). A DFS can be used as a framework for the additional description of other disqualifiers that circumscribe the possible positions of taxa that are not included in the DFS relative to the taxa that are included, and Wilkinson & Benton (1995) have described a method using PAUP to prune taxa from trees for investigating these possible positions. For any set of fundamental trees there will typically be a multiplicity of DFSS. Those DFSS with the greatest number of nodes and secondarily the greatest number of taxa will typically be most useful and, following the terminology adopted for reduced consensus trees (Wilkinson 1994*a*), we distinguish these as primary DFSS.

There is a single primary DFS for the multiple sphenodontid trees (see figure 5) that differs from the primary RAC/RCC (see figure 4*a*) tree only by the exclusion of *Clevosaurus*. The DFS contains a single polytomy at node 2 and in a DFS this indicates that the polytomy itself and/or all possible resolutions of the polytomy are among the fundamental trees, and thus that there are no common disqualifiers that apply to the three groups of taxa arising from this polytomy (Wilkinson 1995*b*). In the primary RCC/RAC tree *Clevosaurus* is also included in this single crown group polytomy. This indicates that, throughout the MPTs, *Clevosaurus* is always more closely related to the other crown-group sphenodontids than it is to *Brachyrhinodon* or any of the more basally diverging taxa, and that *Kallimodon* and *Palaeopleurosaurus* on the one hand, and *Sphenodon-Eilenodon* and *Toxolophosaurus* on the other, are more closely related to each other than they are to *Clevosaurus*. However, because the polytomies of RCC/RAC trees (or those of any of the other kinds of consensus trees discussed thus far) cannot be assumed to be disqualifier faithful they do not imply that all possible resolutions are equally parsimonious. In fact,

all the MPTs agree that *Clevosaurus* is not more closely related to *Kallimodon* and *Palaeopleurosaurus* than it is to the other crown-group sphenodontids. This common disqualifier cannot be easily represented with the aid of the primary RCC/RAC tree but can be represented semi-graphically by describing the possible positions of *Clevosaurus* relative to the numbered nodes of the DFS.

We have summarized common disqualifiers as lists of possible positions of the excluded taxa (*Piocormus* and *Clevosaurus*) in the legend to see figure 5. This approach summarizes more agreement among the sphenodontid trees than any of the other consensus methods considered thus far. Indeed, we have found no simple common disqualifiers that are not represented by the combination of DFS and list of possible positions of excluded taxa. However, as Wilkinson (1994*a*) pointed out, this approach may fall short of the ideal because some disqualifiers reflecting the relationships among excluded taxa are not easy to determine or to describe relative to the DFS. Thus it is not easy to represent whether or not the two excluded taxa *Piocormus* and *Clevosaurus* may be more closely related to each other than to any other sphenodontids or not. There is also some further agreement across the sphenodontid MPTs that cannot be described by simple disqualifiers or *n*-taxon statements and that cannot be simply represented graphically in trees or semi-graphically through reference to possible positions in trees. For example, although *Clevosaurus* may be more closely related to *Sapheosaurus* than to any other crown-group sphenodontid this is only ever the case when *Sphenodon*, *Eilenodon*, *Toxolophosaurus* and *Sapheosaurus* are more closely related to each other than any one of them is to *Kallimodon* and *Palaeopleurosaurus*. This agreement could be represented by more complex conjunctions of disqualifiers and *n*-taxon statements and an ideal method would facilitate its graphical or semi-graphical representation through reference to a consensus tree. Unfortunately, it cannot be claimed that any existing method fulfils this desideratum.

(i) *Greatest agreement subtree (GAS)*

There is a single GAS for the sphenodontid trees that is produced by pruning three taxa, *Piocormus*, *Clevosaurus* and *Sapheosaurus* from these trees and which is equivalent to the DFS (see figure 5) minus *Sapheosaurus*. Wilkinson (1994*a*) considered there to be an intimate relationship between DFSS and GASS. In general, GASS are also DFSS and they can therefore be used as frameworks for the description of the possible positions of the pruned taxa. However, as suggested by Wilkinson (1995*b*), and as illustrated here, the relationship is asymmetric: there are DFSS that are not also GASS. Where primary DFSS differ from GASS, as in this case, it is clear that the primary DFS will convey more information. Here it conveys additional information about the relationships of *Sapheosaurus*. Although GASS are strict consensus trees, demanding agreement across all the MPTs, they may (unlike other strict consensus methods) be sensitive to the treatment of arbitrary resolutions. If all possible resolutions of the relationships of a set of taxa were arbitrary resolutions,

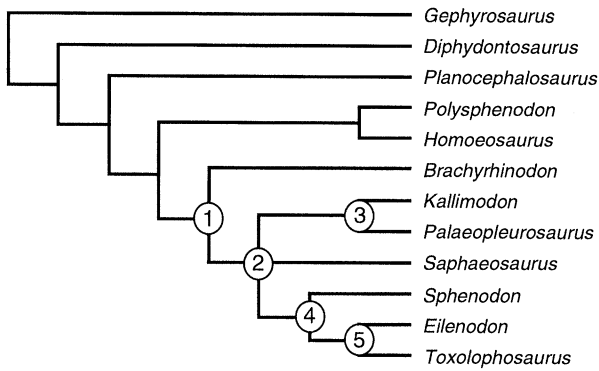


Figure 5. Unique primary disqualifier faithful subtree for the sphenodontid mPTs. The possible positions of excluded taxa are: *Piocormus* – above node 1 and below node 3, *Clevosaurus* – above node 1 and below nodes 3 and 4, and not between nodes 2 and 3.

then suppressing these would produce a polytomy that is common to all the mPTs and which could therefore be included in a dFs. In contrast, the set of binary trees would not share this polytomy and could not yield a GAS that included it.

(j) *Safe taxonomic reduction (STR)*

STR is based on the simple idea that a necessary (but not a sufficient) condition for the inclusion or exclusion of a terminal taxon to have any effect upon the parsimonious interpretation of relationships among other terminal taxa is that the taxon must have a unique combination of phylogenetically informative character states (Wilkinson 1992, 1995*a*). Thus if any terminal taxa do not have a unique combination of character states they bring nothing ‘new’ to the analysis and can be ‘safely’ eliminated from the

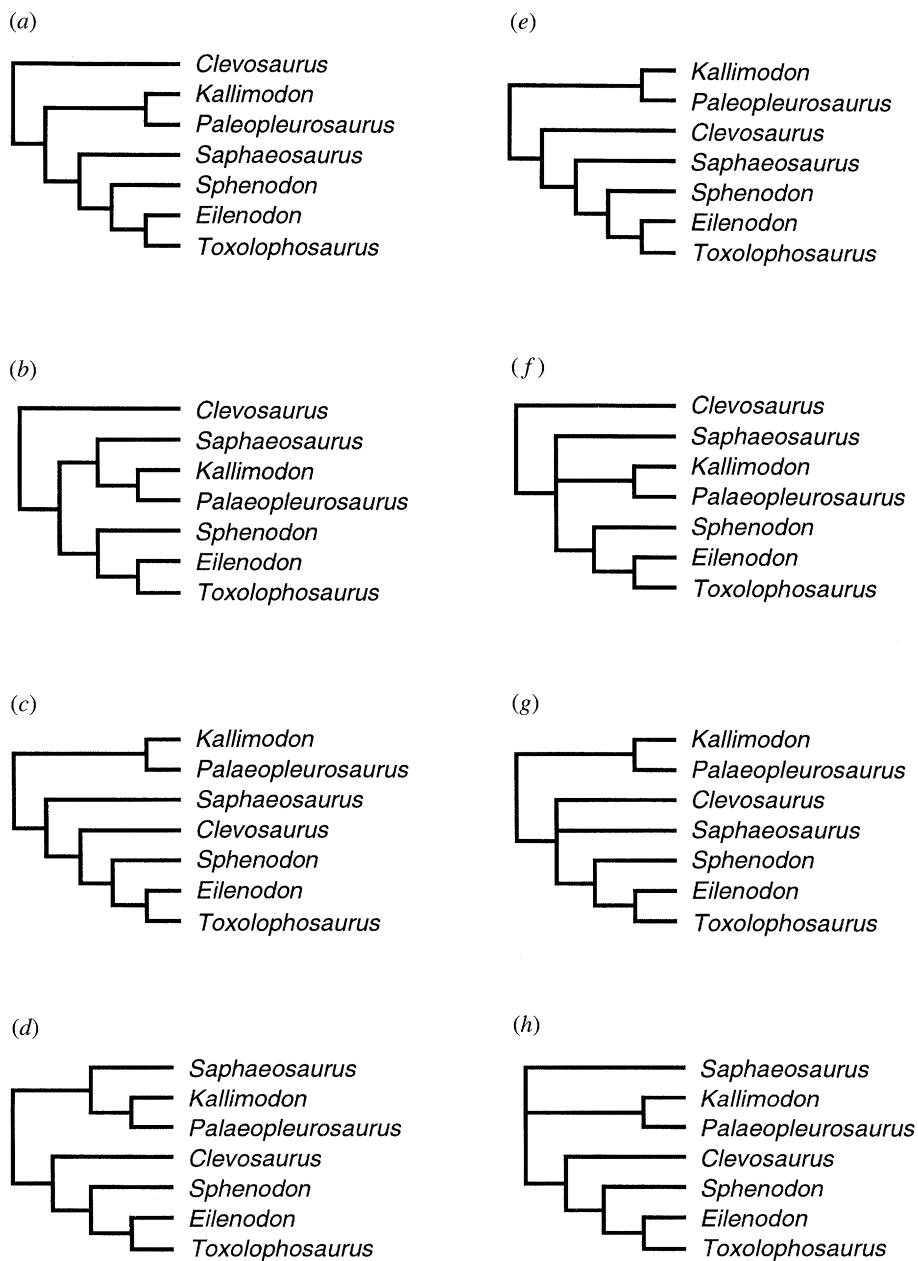


Figure 6. The five binary (a–e) and three polytomous (f–h) mPTs found in the analysis after deletion of *Piocormus* with arbitrary resolutions suppressed. Only relationships within the crown group are shown.

analysis without any danger of ignoring relevant data and without any effect upon the parsimonious interpretation of relationships among the remaining taxa (Wilkinson 1995*a*; Wilkinson & Benton 1995). The *a priori* identification of taxa that can be safely deleted may help resolve relationships that are obscured when many MPTs result from equally parsimonious interpretations of the phylogenetic relationships of poorly known taxa. If any such 'underdetermined' taxa can be safely eliminated *a priori* then this may reduce the number of MPTs and thereby increase the resolution of consensus summaries of the MPTs. Other workers have used the amount of missing data associated with taxa as a basis for taxonomic reduction, and the original analysis provides one example of this approach.

Wilkinson (1995*a*) presented several safe deletion rules based on different kinds of relations termed taxonomic equivalence, but essentially these all lead to the elimination of taxa that do not have unique combinations of phylogenetically informative character states. Determining these relations is a simple task and has been computerised in the TAXEQ2 program. Wilkinson's (1995*a*) discussion considered only the role of missing entries in affecting the scope for safe taxonomic reduction, but this method can be adapted to deal with uninformative characters or character states also. If two taxa differ only in phylogenetically uninformative characters this does not rule out safe taxonomic reduction as would differences in phylogenetically informative characters. The TAXEQ2 program treats any differences between taxa, except those involving missing entries, as informative in a first run through the data and then recodes any uninformative characters and character states as missing entries and makes a second pass through the data (Wilkinson 1995*d*).

Several of the sphenodontid taxa included in the data matrix are poorly known fossils that are coded as missing for many characters. Missing entries account for just over 21% of the data points and render just over 38% of the possible pairwise comparisons of taxa for individual characters uninformative. Without modifying uninformative characters and character states, TAXEQ2 identifies *Piocormus* and *Toxolophosaurus* as taxa that have their character states replicated in *Eilenodon* and *Sphenodon* respectively, and which can therefore be safely deleted. Recoding uninformative characters and character states as described above increases the missing data to over 43% with almost 56% of all pairwise comparisons identified as uninformative and allows TAXEQ2 to identify two additional taxa, *Polysphenodon* and *Kallimodon* that have their character states replicated in *Homoeosaurus* and *Palaeopleurosaurus* respectively and can also be safely deleted.

With just *Piocormus* deleted, parsimony analysis with arbitrary resolutions suppressed yields eight MPTs (see figure 6), five that are binary and three that each have a single polytomy. Note that under the 'soft' interpretation of polytomies, the five binary trees are all possible resolutions of the three polytomous trees and are therefore, in a sense, redundant (see below). An scc tree for the 8 MPTs is identical to the primary RAC/RCC

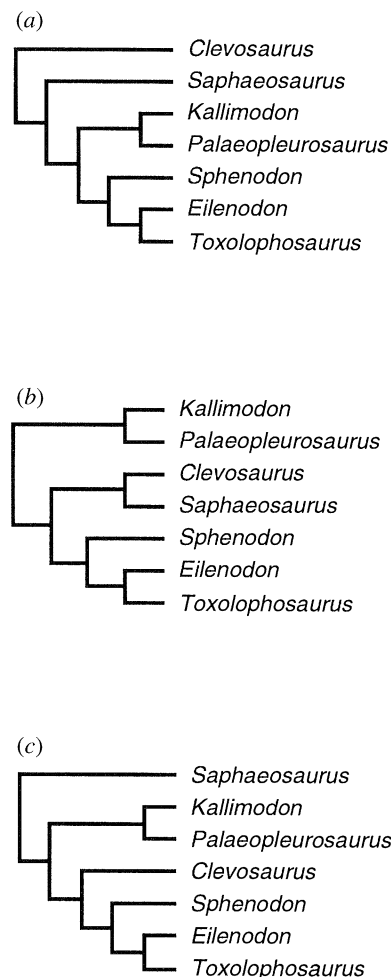


Figure 7. The three additional binary MPTs found in the analysis after deletion of *Piocormus* with arbitrary resolutions allowed. Only relationships within the crown group are shown.

tree from the multiple trees produced when *Piocormus* is included in the analyses (see figure 4*b*). In this case STR allows the same additional relationships to be resolved as does the application of the RCC and RAC methods. Analysis with arbitrary resolutions produces a second, partially overlapping set of 8 MPTs and the same scc tree. This includes the five binary trees of figure 6, and three others (see figure 7) that contain arbitrary resolutions of the three polytomous trees of figure 6. The additional deletion of *Polysphenodon*, *Kallimodon* and *Palaeopleurosaurus* has no effect upon tree statistics and produces no further reduction in the number of MPTs. Their deletion has the negative effect of simply reducing the amount of information conveyed by corresponding consensus trees.

(k) Pruning MPTs

Because STR is not expected to affect relationships among the remaining taxa, trees produced by analyses that exclude these taxa should be the same as trees produced by simply pruning the eliminated taxa from the sets of MPTs found for the complete data and discarding duplicates when trees are rendered identical by the pruning. In the present example, the *a priori*

deletion and *a posteriori* pruning of *Piocormus* are analytically equivalent and have no effect upon tree length. Importantly, simply pruning any taxa from MPTs generated for the complete data can never alter relationships among the remaining taxa in the pruned trees and will therefore always be safe, whether the eliminated taxon can be deleted under the constraints of STR or not. Thus pruning taxa from sets of MPTs and eliminating duplicates from the set of pruned trees provides a simple but potentially powerful procedure for reducing numbers of MPTs, and thereby elucidating relationships that may be obscured by a multiplicity of MPTs. STR can be useful in this regard in suggesting which taxa might be profitably pruned. Similarly, taxa that contribute to the ambiguous polytomies of AC trees are good candidates for initial pruning.

There are strong conceptual links between reduced consensus methods, STR, and selective pruning of MPTs that are illustrated by the sphenodontid data. The RCC method identifies *n*-taxon statements *a posteriori* from a set of trees, and constructs RCC trees to represent these relationships. In general, the RCC method will identify all taxa that must be excluded from a reduced consensus tree in order to specify relationships unambiguously among the remaining taxa. In the sphenodontid example many common statements do not apply to *Piocormus* and its exclusion from the primary RCC tree allows relationships among the remaining taxa to be represented. The interpretation of RCC trees, like that of SCC trees, is completely unambiguous. STR also identifies taxa that can be eliminated, but does this through *a priori* comparisons of the data rather than through *a posteriori* comparisons of trees. If STR leads to the deletion of exactly the same taxa that are excluded from an RCC tree, then the SCC tree for the MPTs produced after STR must be identical to the RCC tree for the MPTs produced from the complete data. Similarly if the taxa that are excluded from an RCC tree are selectively pruned from the MPTs, the SCC tree for the pruned trees will be identical to the RCC tree. The RCC profile can thus be conceived of as a collection of SCC trees that are produced after various combinations of taxa are selectively pruned from the MPTs.

Pruning taxa from MPTs can also be used to construct RAC trees as a practical alternative to Wilkinson's (1994a) algorithm for constructing the RAC profile which is not yet computerised. Using PAUP, the procedure is as follows.

1. Save all MPTs in a treefile.
2. Construct both the SCC and AC trees. Any groups appearing in the AC tree but not in the SCC tree suggest that there may be additional *n*-taxon statements that are shared by all the MPTs but, because they apply to only a subset of the taxa, these are not reflected in the SCC tree and are necessarily ambiguous in the AC tree.
3. Experimentally delete taxa that contribute those polytomies in the AC tree (the *p*-groups of Wilkinson 1994a) that do not also occur in the SCC tree, instructing PAUP to prune these taxa from the trees in memory. Construct the SCC and AC trees for the pruned trees. If these consensus methods produce identical consensus trees, then there are no ambiguous relationships in the

AC tree, and the consensus tree will either be an RAC tree, or a less informative subtree of an RAC tree.

4. Repeat step 3, reloading the original set of MPTs from memory as necessary, until all combinations of eliminated *p*-groups have been tried. The RAC profile includes all the non-redundant topologies produced where the SCC and AC trees are identical.

(1) *Redundant trees and arbitrary resolutions*

It is often the case that reducing numbers of trees increases the resolution that can be achieved in consensus trees, but this is not expected when reductions in the numbers of tree are achieved by the suppression of arbitrary resolutions. With the sphenodontid data this reduces the number of MPTs from 82 to 60 but has no effect upon the resolution of any strict consensus tree. In avoiding arbitrary resolutions PAUP collapses only those branches that have a length of zero under all possible character optimizations, a procedure dubbed Swofford's rule by Wilkinson (1995c). The branches that remain are those that are supported by some optimization. This approach has been criticized because the trees that remain may include branches that have zero-length under particular character optimizations, or worse, combinations of branches which cannot all have non-zero-length under any character optimization (Wilkinson 1995c). Alternative approaches to the suppression of arbitrary resolutions that remove these persistent arbitrary resolutions have been described (Wilkinson 1995c) and these make it theoretically possible to either: (i) retain only those MPTs for which all internal branches have non-zero-length under some particular optimization; or (ii) retain just those MPTs that can have no non-zero-length branches under any optimization. However, these methods are not yet computerized and the number of sphenodontid trees makes the manual application of these approaches impractical in this case (cf. below).

The polytomies produced by collapsing zero-length branches can be interpreted as soft polytomies: they imply that all possible resolutions of the polytomy will be equally parsimonious, though not that they will be equally free of arbitrary resolutions. As demonstrated by the eight MPTs found after the deletion of *Piocormus* (see figure 6) the suppression of arbitrary resolutions used in PAUP can produce sets of trees that include both polytomous trees and some of their possible resolutions, the latter being redundant. Thus eliminating any redundant trees provides another potential strategy for reducing the number of MPTs and suppressing persistent arbitrary resolutions. The simplicity of this approach makes its manual application to the 60 sphenodontid MPTs practical. Some 47, just over 75%, of the 60 MPTs, are redundant. However, some of the 13 remaining sphenodontid MPTs retain branches with a minimum length of zero. Thus applying this rule and then eliminating redundant trees, although capable of drastically reducing persistent arbitrary resolutions, will not necessarily eliminate them all.

We used MacClade to explore the effects of collapsing branches in the remaining 13 trees and found that seven of these trees were free of arbitrary

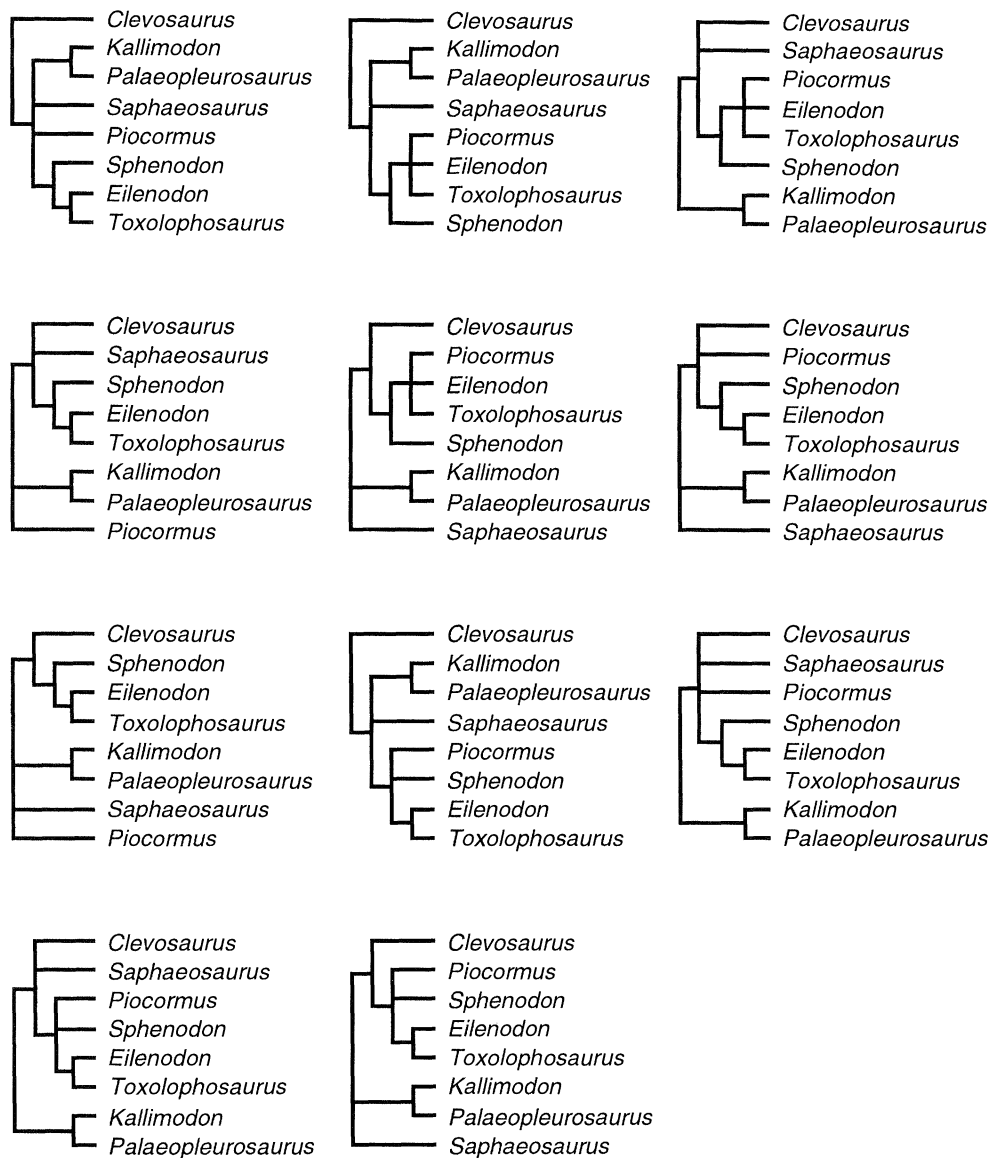


Figure 8. 11 non-redundant sphenodontid MPTs that contain no arbitrary resolutions. Only relationships within the crown group are shown.

resolutions but that six contained branches that could be collapsed without increasing tree length (calculated, as in PAUP, under the hard interpretation of polytomies). Collapsing these branches yielded only four distinct tree topologies, leaving a total of just 11 MPTs that contain no zero-length branches under any optimization (see figure 8). Strict consensus trees produced for this group of 11 trees are identical to those produced from the full set of 82 binary MPTs and the reduced set of 60 MPTs retained by PAUP after its incomplete suppression of arbitrary resolutions. Thus reducing the number of MPTs through more complete suppression of arbitrary resolutions and by eliminating redundant trees will not produce better resolved consensus summaries. However, it may reduce the number of trees sufficiently to allow inspection of all the trees and the application of methods that are not yet computerized. Also, in cases where the number of MPTs exceeds the limits imposed by computer memory on the number of trees that can be stored, phylogenetic studies are susceptible to problems of selective sam-

pling. If available memory is used to store only non-redundant trees then this difficulty may be avoided or ameliorated.

(m) *Secondary optimality criteria*

Secondary criteria can sometimes be used to choose among otherwise optimal MPTs. Rodrigo (1992) recommended selecting those MPTs that allow the largest number of characters to be free of homoplasy. His optimal character compatibility index (occi) is the ratio of the number of homoplasy-free characters to the total number of characters and is maximised with the number of homoplasy-free characters. Sang (1995) advocated his average unit character consistency (AUCC) as a secondary optimality criterion for choosing among MPTs. As its name suggests, this is the average of the individual character consistency indices. As with the occi, the AUCC increases with increase in homoplasy-free characters, but is also sensitive to the distribution of homoplasy among the homoplastic

characters, increasing as more homoplasy is concentrated in fewer characters.

With the sphenodontid data, 24 of the 29 characters fit all the MPTs with no homoplasy and three characters (6, 17, 28) require two steps on each tree. Only two characters, 15 and 20, vary in their fit to the MPTs, with 15 requiring two steps and 20 being homoplasy-free or *vice versa*. Character 15 supports the hypothesis that *Clevosaurus* is more closely related to *Sphenodon* than to *Palaeopleurosaurus* and requires a single step on either the 38 or 55 MPTs that include this hypothesis, depending on the suppression or inclusion of arbitrary resolutions respectively. Character 20 supports the incompatible hypothesis that *Clevosaurus* is outside a group that includes *Sphenodon* and *Palaeopleurosaurus* as well as most other crown-group sphenodontids, and requires a single step on the remaining 22 or 27 MPTs that include this hypothesis. Thus $occi$ is 0.86 and the $aucc$ is 0.931 for all trees. Applied to this example, neither of these secondary optimality criteria allows any further choice among the MPTs and this reflects the fact that many of the MPTs differ in the positions of underdetermined taxa without any corresponding differences in the distributions of homoplasy.

Sharkey (1989) presented a method for choosing among MPTs that uses a probabilistic approach to compatibility analysis pioneered by Le Quesne (1969, 1972) and further developed by Meacham (1981) and Gauld & Underwood (1986). Essentially this method involves an iterative procedure, the boil down, that produces a ranking of characters based on their ratios of observed incompatibilities to the number expected if all characters are random, the randomness ratio (Gauld & Underwood 1986) or coefficient of character state randomness (Le Quesne 1969). The preferred MPTs are those that most concentrate homoplasy in the characters of lowest rank. Wilkinson (1992) suggested that the character normal deviate (Le Quesne 1972), a z-score, might be used in the boil down in place of the randomness ratio.

Applied to the sphenodontid data, both procedures assign character 20 higher rank than character 15 and lead to the selection of those 22 or 27 MPTs, depending on the treatment of arbitrary resolutions, in which *Clevosaurus* is outside a group that includes *Sphenodon*, *Palaeopleurosaurus* and the other crown-group sphenodontids. This is sufficient to determine the position of *Clevosaurus* relative to the DFS (see figure 5, between nodes 1 and 2) and if this choice of MPTs is adopted *Clevosaurus* can be grafted onto the DFS to give a more fully resolved summary of the reduced set of preferred trees.

(n) *Successive approximations character weighting (SACW)*

Carpenter (1988) recommended using Farris's (1969) SACW to select from among multiple MPTs. As with the application of secondary optimality criteria, SACW is expected to resolve character incongruence by concentrating homoplasy in the most unreliable, and thus low-weight, characters. However, although SACW may identify optimal trees that are a subset of an

original set of MPTs this approach may also yield optimal trees that are not among the original set of MPTs. It may also increase rather than decrease the number of trees.

Farris (1969) investigated a variety of weighting functions that related a measure of fit to a numerical character weight. In Hennig86 (Farris 1988) character weights are based on their rescaled consistency indices (Farris 1989) and where these vary across the trees, the maximum value (corresponding to the best fit) is used. In PAUP the user must choose among a variety of weighting procedures that use either best, worst, or average fits of either the character consistency, retention or rescaled consistency indices.

Application of these alternatives to the problems of sphenodontid phylogeny, using the MPTs recovered with equal weighting of characters as the first approximation produced a variety of somewhat inconsistent results. As noted above, only five characters show any homoplasy on any of the MPTs and only two of these (the mutually incompatible 15 and 20) vary in their fit across the MPTs, each requiring either no homoplasy or a single extra step. SACW using best-fits of any of the three indices does not discriminate between these characters (the best-fit indices are all unity) and produces no change in the optimal trees. Similarly, the worst-fit consistency indices are the same for both characters (0.5) and SACW using these also has no effect. In contrast, the worst-fit retention and rescaled consistency indices are greater for character 20 (0.75 and 0.375 respectively) than for character 15 (both zero indicating that it achieves the worst possible fit), and SACW using these measures discriminates in favour of character 20. This yields either 12 or 27 optimal SACW trees, depending on whether arbitrary resolutions are collapsed or not, that place *Clevosaurus* as the sister taxon of the other crown-group sphenodontids. With arbitrary resolutions not suppressed, the set of 27 optimal SACW trees is simply the subset of equal weighted MPTs in which homoplasy is attributed to character 15 and character 20 requires only a single step. With arbitrary resolutions suppressed this relationship does not hold (or it would have yielded 22 optimal trees). This points to an additional effect of SACW here that arises from the assignment of zero-weight to 15, which is thereby excluded from the analysis. In those equal-weight MPTs where character 15 is homoplastic, some branches must have a non-zero maximum length solely because of possible optimizations of character 15. With character 15 excluded, these branches become arbitrary resolutions and are suppressed.

Worst-fit and best-fit indices cannot vary with the inclusion or exclusion of arbitrary resolutions, but average fits are expected to vary with the treatment of a zero-length branches. SACW based on average fits of their consistency indices favoured character 15 over character 20, whether arbitrary resolutions were suppressed or not, because it is homoplastic in fewer of the equal-weight MPTs. In both cases, the 38 or 55 optimal SACW trees are simply the subset of equal weighted trees in which character 15 requires a single step. These trees agree that *Clevosaurus* is closer to

Table 4. *Character weights for characters 15 and 20 produced by two iterations of successive approximations character weighting using average fits and three different weighting functions*

(Arrows indicate changes in weights with successive iterations of the procedure. See text for further details.)

starting trees character	arbitrary resolutions suppressed (60 MPTS)		arbitrary resolutions included (82 MPTS)	
	15	20	15	20
consistency indices	0.817 → 1	0.683 → 0	0.835 → 1	0.665 → 0
retention indices	0.633 → 0	0.842 → 1	0.67 → 1	0.832 → 1
rescaled consistency indices	0.517 → 0	0.575 → 1	0.560 → 1	0.553 → 375

Sphenodon (and *Eilenodon* and *Toxolophosaurus*) than it is to *Palaeopleurosaurus* (and *Kallimodon*) and restrict the possible positions of *Clevosaurus* to between nodes 2 and 4 or between node 2 and *Sapheosaurus* of the DFS (see figure 5). SACW using the average retention index is also unaffected by the inclusion or exclusion of arbitrary resolutions. However, it favours character 20 and produces the same sets of optimal SACW trees found when SACW was based on the worst fits of these measures.

The results of SACW using the average fits of the rescaled consistency index depend on the treatment of arbitrary resolutions. With arbitrary resolutions collapsed, character 20 is favoured and, as with the use of the worst fit of this index, this yields 12 optimal SACW trees. With arbitrary resolutions included, character 15 is favoured and 55 optimal SACW trees are produced. Further details of the character weighting based on average fits are summarized in table 4. Note that the initial differential in the average rescaled consistency index between character 15 and 20 is very slim.

5. DISCUSSION

(a) *Secondary optimality criteria and character weighting*

All of the secondary optimality criteria illustrated above are similar in that they select hypotheses that concentrate homoplasy in the 'worst' characters and improve the fit between the phylogeny and the 'better' characters. As Sang (1995) pointed out, they are very much in the spirit of compatibility analysis and some of the approaches represent marriages of compatibility and parsimony. However, in all these approaches, the secondary optimality criteria are constrained by the primary optimality criterion of maximum parsimony, typically with equally weighted characters.

There is a close conceptual link between secondary optimality criteria used to distribute homoplasy asymmetrically across low- and high-quality characters and differential character weighting systems that down-weight low-quality characters. We suspect that, despite the prevalence of equal weighting of characters, most phylogeneticists would agree that not all characters are equally reliable. Indeed this is an inescapable conclusion of most phylogenetic analyses, and it is a prerequisite for the justification of using secondary optimality criteria to select among MPTS and for differential character weighting. Differential weighting schemes that accurately reflected the quality of characters would surely be universally acceptable and

more likely to generate accurate phylogenies. However phylogeneticists vary in their acceptance of differential weighting and can be classified into three camps. In the first camp are many practitioners who eschew differential weighting. For these practitioners we expect that the use of secondary optimality criteria to select among MPTS is also unconvincing. We also consider that there are few arguments in favour of equal character weighting in the face of evidence of differences in character quality and these are no more convincing than the arguments in favour of differential weighting. That equal weighting may be more convenient or easier to apply is surely no good reason for ignoring alternatives. In the second camp, are those practitioners that adopt or advocate secondary optimality criteria to choose among MPTS. They seem to agree that character quality can be assessed and used to discriminate among trees. However, in rejecting differential character weighting they place strong limits upon the role of assessments of character quality, allowing them to discriminate only among the original MPTS although there may be other trees that provide a better fit to the data in the light of the assessments of character quality. Thus it is hard to justify the use of secondary optimality criteria to select from a restricted set of trees rather than the use of differential character weighting based on the same estimates of character quality and an unconstrained search for optimal trees. In the third camp, are those phylogeneticists who use or advocate differential character weighting. They would seem to have no need for any secondary optimality criterion because they will have incorporated assessments of character quality into their primary optimality criterion, typically differentially weighted parsimony, and identified the optimal trees.

SACW has been advocated as a method for choosing among MPTS, and although it was not designed for this purpose it often has this effect. Application of SACW to the sphenodontid phylogeny demonstrates that choices among the various available measures of fit of characters used as weighting functions can have profound effects upon the results. Surprisingly, these methodological alternatives do not appear to have been much investigated or discussed previously. Ideally, choices among these alternatives should be justified and not determined arbitrarily by the defaults in the available software.

Using best-fits of any index is equivalent to seeing characters in their best possible light and ignoring the fact that some characters may appear less reliable on some other trees. It is an optimistic strategy that values

any evidence of high quality over any contradictory evidence. Conversely, using worst-fits is a pessimistic strategy that values any evidence of low quality over any contradictory indications. If one considers the propositions that: (i) a character that shows no homoplasy should be judged more reliable than a character that is homoplastic on some but not all MPTs; and (ii) a character that is homoplastic on all trees should be judged less reliable than a character that is homoplastic on some but not all MPTs, then using neither best nor worst fits seems entirely sensible. Using average fits may be a better alternative but, as shown by the sphenodontid example, different treatments of arbitrary resolutions can produce contradictory results and this is far from satisfactory. One means of taking into account variation in the fits of characters whilst avoiding the effects of arbitrary resolutions would be to take the mid-point of the range between the best and worst fits of characters as a basis for weighting. This expedient ignores the relative frequencies of different fits of characters across the trees.

Farris (1969) used simulation studies to evaluate alternative weighting coefficients for use with SACW and obtained the best results with a function that assigned a greater differential between good and mediocre characters than that between mediocre and poor characters. His simulations involved adding random characters (noise) to a set of perfectly correlated characters that specified a particular tree (signal). Under these conditions the success of his preferred weighting function is perhaps not surprising, given that the recovery of the 'correct' tree depends primarily on the discrimination between the strong signal and the noise, and depends little upon differences in the relative weights of the noisy random characters which are expected to have mainly mediocre or poor fits to the approximated phylogenies. Of the character consistency, retention and rescaled consistency indices that can be used in current implementations of SACW, only the consistency and rescaled consistency indices approximate Farris's (1969) preferred weighting function. The consistency index has well known deficiencies that severely compromise its use for differential weighting. In particular, the minimum consistency index of a binary character is the inverse of the maximum number of steps. In contrast, the rescaled consistency index ranges between zero and one for all informative characters so that a character achieving its maximum number of steps has a rescaled consistency index of zero. The square or cube of the retention index would also share these useful properties and might also provide a suitable weighting function. Of the available computerized methods, we tentatively recommend the average fit of the rescaled consistency index, but with the warning that this may be sensitive to the treatment of arbitrary resolutions.

In practice, the application of secondary optimality criteria, or any weighting scheme, can leave the phylogeneticist with multiple optimal trees. Thus with the sphenodontid example, the OCCI and AUCC and several of the SACW regimes supported no further choice among the MPTs. Multiple MPTs may result from either contradictory evidence producing balancing character

incongruence or from lack of evidence. Poorly known fossil taxa with much missing data often have a multiplicative effect on numbers of MPTs. Character weighting and secondary optimality criteria can only promote further preferences among a set of optimal trees when these trees differ in the way they distribute homoplasy so as to explain contradictory evidence. They will not reduce the number of trees when the multiplicity results from absence of data. With the sphenodontid data, Sharkey's (1989) method discriminates in favour of character 20 over character 15 and in doing so determines the otherwise variable position of *Clevosaurus*. However, this still leaves a multiplicity of MPTs and much of the variation in these trees is due to the lack of data for *Piocormus*.

The sphenodontid example also shows that the available methods vary greatly in their ability to discriminate among characters and further resolve relationships that are unresolved because of balancing character incongruence. The retention and rescaled consistency indices and probabilistic compatibility measures have greater discriminatory power than the consistency index because their assessment of character quality depends on a comparison with how poorly they might perform and this depends upon the number of taxa that have the alternative character states and varies among characters. A binary character requiring two steps may fit as poorly as possible or quite well depending on the maximum number of steps, but the consistency index will be 0.5 in either case. This suggests that the average character retention or rescaled consistency indices, analogous to Sang's (1995) AUCC, might provide more discriminatory power than the latter, as does Sharkey's (1989) probabilistic compatibility approach.

An important question to be considered when applying the discriminatory power of weighting methods or secondary optimality criteria is how much better are the preferred trees over their rejected competitors. In parsimony analysis with either differential or equally weighted characters, length differences between trees provide one means of assessing relative support (Bremer 1988; Källersjö *et al.* 1992). In contrast, no methods for assessing the relative support for hypotheses preferred on the basis of secondary optimality criteria have been advanced, and this deserves further attention. Similarly, no methods for evaluating the relative support for alternative weighting strategies have been advanced and because weighting is inescapable, we consider this to be a priority area for future research.

We are concerned that some differential weighting schemes may produce large differences in character weights when there are only slight differentials in the measures of character reliability. For example, characters 15 and 20 in the sphenodontid data have very similar average rescaled consistency indices and very similar coefficients of character state randomness. A cautious worker would question whether these differences provide a sufficient basis for differential weighting and choice between phylogenetic hypotheses. We suggest that variation in phylogenetic hypotheses within a range of rational weighting strategies

can itself be used to evaluate phylogenetic hypotheses. Those relationships that are unaffected by differences in weighting regimes should be considered better supported than those that depend on a subset of the weighting schemes. Applied to the sphenodontid data, this approach leads us to prefer the hypothesis that *Clevo*saurus is the sister taxon of the other crown-group sphenodontids because character 20 appears to have the edge over character 15. However, we also stress that this slim advantage provides only relatively poor differential support for this hypothesis compared to the other relationships that can be hypothesised among the sphenodontids and that do not vary with the various weighting regimes.

(b) Consensus methods

Whatever stance is taken on the use of character weighting and secondary optimality criteria, these methods do not eliminate the need to summarize agreement among multiple trees through consensus methods and the need to choose among available techniques.

An important distinction between consensus methods separates those that summarize agreement across all the fundamental trees (strict methods) from those that may also express relationships that are not true of every fundamental tree (majority-rule and semi-strict methods). Advocates of parsimony analysis will often be interested in only those relationships that are replicated across all MPTs because they are unambiguously supported by the parsimonious interpretation of the data. If length differences are used to measure support, only relationships that are common to all MPTs can have a support value greater than zero. Thus majority-rule and semi-strict consensus methods, which may appeal because they are well resolved, achieve this additional resolution only by including poorly supported groups.

The frequency of appearance of groups in a set of MPTs is typically reported with majority-rule consensus trees, and may often be interpreted as an indicator of the support for the group. This may be a reasonable interpretation when the consensus summarizes multiple analyses of different data sets (e.g. in bootstrapping), but such an interpretation may be problematic when applied to group frequencies in MPTs from a single data set, and at the very least raises important questions that have yet to be addressed. In what sense can the frequency with which a group appears in a set of MPTs be considered a measure of its support? If we assume that all the MPTs are equally well supported and thus equally probable or plausible, then the proportion of times a particular hypothesis occurs within the trees would be a measure of the relative probability or plausibility of that hypothesis. However, as we have seen, majority-rule consensus methods may be sensitive to the treatment adopted for arbitrary resolutions. Depending on the treatment of arbitrary resolutions, individual branches on a single tree may have non-zero lengths and thus receive direct support, under some character optimizations but not under others, or they may simply be compatible with the data and not

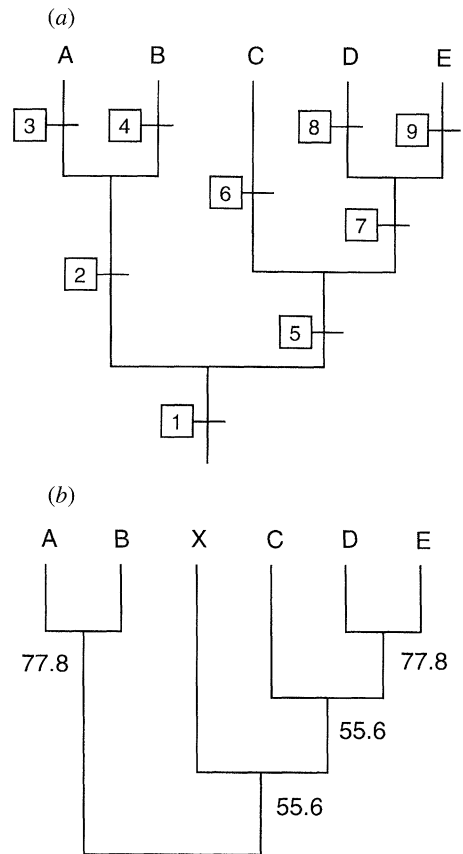


Figure 9. (a) Hypothetical phylogeny for five taxa with nine equally parsimonious positions for taxon X that lacks any data, (b) and corresponding mrcc tree.

directly supported by non-zero lengths at all. If group frequencies are to be taken as a measure of support, how should arbitrary resolutions be treated?

When data sets include poorly known fossil taxa, the position of these taxa in a majority-rule consensus tree may be largely determined by relationships among the better-known taxa. For example, suppose that the tree in figure 9a represents a set of well defined most parsimonious relationships for the five taxa, A–E. If an analysis were performed with arbitrary resolutions allowed, and an additional taxon (X) with only missing entries, this would produce nine trees differing only in the arbitrarily resolved position of X. The corresponding MRCC tree would place X as the sister group to CDE (see figure 9b) with a frequency of 5/9. Here, X is associated with CDE simply because there are more ways that it can associate with this group. Given that X has no data, any argument that this position of X is the best supported would be difficult to justify. With the sphenodontid analyses the position of *Piocormus* in the MRCC trees is similarly partly determined by tree balance rather than purely from considerations of data.

This difficulty might be avoided by suppressing arbitrary resolutions but, as we have seen, there are different methods of suppressing arbitrary resolutions and these can lead to different group frequencies in MRCC trees. Assessments of support when there are branches that may have zero length under some but not all optimizations, as may be produced by Swof-

ford's rule implemented in PAUP and by Wilkinson's (1995*c*) first method, should presumably also take into account this additional variation. Thus we could assume that all optimizations are equally plausible or possible, so that measures of the relative probabilities or plausibilities of particular groups would depend, not only upon the number of times the group occurs, but also upon the number of favourable character optimizations and the number of unfavourable ones. Computationally this would be very demanding. A simpler alternative would be to base MRCC trees upon only those MPTS that must have all non-zero-length branches under all character optimizations, but this would in turn ignore differences in the potential support for groups that are possible resolutions of polytomies in these trees. We offer no definitive solution to these problems, which apply both to MRCC trees and average measures of character quality in character weighting schemes, but we caution against the limitations of existing software.

For the sphenodontid trees the NC trees are identical to the corresponding MRCC trees with compatible minority components. Swofford (1991) also described a case in which these consensus methods produced identical results. Swofford noted that the equivalence of the NC and MRCC methods does not hold in general, and it is easy to design hypothetical examples where these methods produce different consensus trees (F. R. McMorris personal communication). Unfortunately there is no obvious way to determine *a priori* whether these consensus methods will produce identical results. The properties and behaviour of the NC method remain rather poorly understood, but it is apparent that this method is not a strict method and that it is sensitive to the treatment of arbitrary resolutions.

Strict consensus methods, except the GAS method, are immune to the problems of arbitrary resolutions. However not all strict consensus methods are equally useful for identifying all the relationships that may be of interest to phylogeneticists, and these inequalities are well illustrated by our sphenodontid example. The SCC has been criticized by many authors because it may produce poorly resolved consensus summaries (Adams 1986; Swofford 1991) despite considerable unanimity among the fundamental trees. The problems of this most widely used consensus method has motivated strategies of taxonomic reduction and probably contributed to the perceived need for secondary optimality criteria. Fraser & Benton's reliance upon this method contributed to their inability to resolve relationships among the sphenodontid crown-group. The AC tree (see figure 3) suggests that their data supports additional resolution but, unfortunately, what AC trees gain in sensitivity they confound in ambiguity.

The RAC method solves the problem of AC tree ambiguities, but may lead to a profile that itself includes a multiplicity of consensus trees. This is true also for the RCC method which, like the RAC method, produces trees that have an unambiguous interpretation as collections of common *n*-taxon statements, but which may require a multiplicity of RCC trees to summarize all such statements. If the aim of the phylogeneticist is to identify all relationships that are

supported unambiguously by the parsimonious interpretation of the data, then the closest approximation to this ideal seems to be provided by a primary DFS in conjunction with a listing of the possible positions of the excluded taxa. Unfortunately, this method is not yet computerized and it is therefore not readily applicable when there are large numbers of MPTS. GASS may also serve this purpose but, where they differ from primary DFSs, they are less informative.

Unfortunately, there is no guarantee that a DFS will provide a suitable framework for the simple semi-graphical description (by lists of possible positions) of all relationships that are common to a set of MPTS and that are not graphically represented in the DFS itself. Those disqualifiers that describe relationships among the excluded taxa cannot be represented easily in this way and, as the sphenodontid example shows, there may be agreement over relationships that can only be described through complex conjunctions of disqualifiers and *n*-taxon statements that do not lend themselves to graphical or semigraphical representation. There thus remains the possibility that better consensus techniques can be developed.

(c) Taxonomic reduction

Taxonomic reduction is aimed at reducing the number of MPTS and thereby increasing the resolution of consensus summaries. The original analysis provides an example of a series of taxonomic reductions using the amount of missing data associated with a taxon as a criterion for the exclusion of taxa. This approach carries the danger that excluding taxa may remove unique combinations of character states and thereby change inferred relationships among the remaining taxa i.e. it may not be safe. Fraser & Benton found that *Clevosaurus* is more closely related to *Sphenodon* than to *Palaeopleurosaurus* (see figure 1*d*) when all taxa with missing entries are removed. However, analysis of the complete data shows that it is equally parsimonious for *Sphenodon* and *Palaeopleurosaurus* to be more closely related to each other than to *Clevosaurus*. These are the two alternatives supported by the incongruent characters 15 and 20, respectively. When only the taxa without missing entries are considered, character 20 becomes phylogenetically uninformative and the evidence for homoplasy in character 15 is removed, so that parsimony finds only the relationship supported by character 15. In this case, eliminating some taxa alters the range of inferred relationships among the remaining taxa. In contrast, STR operates under the constraint that elimination of taxa must not alter those relationships that are supported by the complete data. If taxonomic reduction is used in an attempt to reduce numbers of MPTS and enhance the resolution of consensus trees, the taxa to be deleted should not be selected on the amount of missing data or any other arbitrary criterion. *A priori* taxonomic reduction should be done under the constraints of STR. Alternatively, it should be done *a posteriori* by pruning taxa from MPTS after analysis of the complete data.

Although, taxonomic reduction may reduce numbers of trees, it cannot be relied upon to remove the

need for consensus trees altogether. This is especially true when the multiple trees reflect alternative interpretations of incongruent characters rather than alternative placements of poorly known taxa.

(d) *Sphenodontid phylogeny*

Applied to the problem of sphenodontid crown-group relationships, *a priori* STR, *a posteriori* pruning of MPTs, and each of the reduced consensus methods (RAC, RCC, DFS and GAS) are all effective at enhancing our understanding beyond that achieved by the original analysis. They show that *Toxolophosaurus* and *Eilenodon* are more closely related to each other than to any other sphenodontids, with the possible exception of *Piocormus*. Similarly, with the possible exception of *Piocormus*, they are the closest relatives of *Sphenodon*. The variable positions of *Piocormus* reflect absence of evidence rather than incongruent evidence. The relationships of *Sapheosaurus* and *Clevosaurus* are ambiguous, but less so than *Piocormus*. As with *Piocormus*, *Sapheosaurus* owes its variable position to lack of evidence. In contrast, incongruent characters place *Clevosaurus* as a basal taxon within the crown-group or as more closely related to the *Sphenodon-Eilenodon-Toxolophosaurus*. Assessments of character quality suggest that the latter position is better supported but only marginally so. The remaining lack of resolution is due to limitations of the data rather than methodological problems. We consider that the additional resolution afforded by the MRCC or NC methods does nothing to improve upon this summary of sphenodontid relationships.

This work was supported in part by SERC Grant GR/F 87912. R. D. M. Page, C. Patterson, F. R. McMorris and R. Powers provided useful feedback.

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. Class.* **3**, 299–317.
- 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.
- Farris, J. S. 1969 A successive approximations approach to character weighting. *Syst. Zool.* **18**, 374–385.
- Farris, J. S. 1988 *Hennig86. Version 1.5 software and documentation*. Port Jefferson Station, New York.
- Farris, J. S. 1989 The retention index and the rescaled consistency index. *Cladistics* **5**, 417–419.
- Fraser, N. C. & Benton, M. J. 1989 The Triassic reptiles *Brachyrhinodon* and *Polysphenodon* and the relationships of the sphenodontids. *Zool. J. Linn. Soc.* **96**, 413–445.
- Gauld, I. & Underwood, G. 1986 Some applications of the LeQuesne compatibility test. *Biol. J. Linn. Soc.* **29**, 191–222.
- Gordon, A. D. 1980 On the assessment and comparison of classifications. In *Analyse de Données et Informatique* (ed. R. Tomassone), pp. 149–160. Le Chesnay: INRIA.
- Hendy, M. D. & Penny, D. 1982 Branch and bound algorithms to determine minimal evolutionary trees. *Math. Biosci.* **59**, 277–290.
- Källersjö, M., Farris, J. S., Kluge, A. G. & Bult, C. 1992 Skewness and permutation. *Cladistics* **8**, 275–287.
- Kubicka, E., Kubicki, G. & McMorris, F. R. 1995 An algorithm to find agreement subtrees. *J. Class.* (In the press.)
- Le Quesne, W. J. 1969 A method of selection of characters in numerical taxonomy. *Syst. Zool.* **18**, 201–205.
- Le Quesne, W. J. 1972 Further studies based on the uniquely derived character concept. *Syst. Zool.* **21**, 281–288.
- Maddison, W. P. & Maddison, D. R. 1992 *MacClade version 3.01. Computer program*. Sunderland, Massachusetts: Sinauer Associates.
- Margush, T. & McMorris F. R. 1981 Consensus *n*-trees. *Bull. Math. Biol.* **43**, 239–244.
- Meacham, C. A. 1981 A probability measure for character compatibility. *Math. Biosci.* **57**, 1–18.
- 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.
- Page, R. D. M. 1989 Comments on component-compatibility in historical biogeography. *Cladistics* **5**, 167–182.
- Page, R. D. M. 1992 *COMPONENT for Windows*. London: The Natural History Museum.
- Rodrigo, A. C. 1992 Two optimality criteria for selecting subsets of most parsimonious trees. *Syst. Biol.* **41**, 33–40.
- Sang, T. 1995 New measurements of the distribution of homoplasy and reliability of parsimonious cladograms. *Taxon* **44**, 77–82.
- Sharkey, M. J. 1989 A hypothesis-independent method of character weighting for cladistic analysis. *Cladistics* **5**, 63–86.
- Sokal, R. R. & F. J. Rohlf 1981 Taxonomic congruence in the Leptodomorpha re-examined. *Syst. Zool.* **30**, 309–325.
- Swofford, D. L. 1985 *PAUP: Phylogenetic analysis using parsimony, version 2.4.1, documentation and software*. Champaign: Illinois Natural History Survey.
- Swofford, D. L. 1991 When are phylogeny estimates from molecular and morphological data incongruent? In *Phylogenetic analysis of DNA sequences* (ed. M. M. Miyamoto & J. Cracraft), pp. 295–333. Oxford University Press.
- Swofford, D. L. 1993 *PAUP Phylogenetic Analysis Using Parsimony, version 3.1.1 Documentation and Software*. Champaign: Illinois Natural History Survey.
- Wilkinson, M. 1992 *Consensus, compatibility and missing data in phylogenetic inference*. PhD thesis, University of Bristol.
- Wilkinson, M. 1994a Common cladistic information and its consensus representation: reduced Adams and reduced cladistic consensus trees and profiles. *Syst. Biol.* **43**, 343–368.
- Wilkinson, M. 1994b Three-taxon statements: when is a parsimony analysis also a clique analysis? *Cladistics* **10**, 221–223.
- Wilkinson, M. 1995a Coping with missing entries in phylogenetic inference using parsimony. *Syst. Biol.* **44**, 501–514.
- Wilkinson, M. 1995b More on reduced consensus methods. *Syst. Biol.* **44**, 436–440.
- Wilkinson, M. 1995c Arbitrary resolutions, missing entries and the problem of zero-length branches in parsimony analysis. *Syst. Biol.* **44**, 108–111.
- Wilkinson, M. 1995d *TAXEQ2, software and documentation*. School of Biological Sciences, University of Bristol.
- Wilkinson, M. & Benton, M. J. 1995 Missing data and rhynchosaur phylogeny. *Hist. Biol.* **10**, 137–150.

Received 11 May 1995; accepted 19 July 1995