



Consensus trees and tree support

In this article I will look at two separate issues; consensus trees and support for the nodes on your tree. There is a tenuous link between these as we will see.

Consensus trees

Often, after we have carried out our analysis, the tree building routine (whichever algorithm we use) will report more than one parsimonious tree. In other words the data used is compatible with more than one cladogram/tree. In such circumstances there are two things that we can do. We can choose one of the trees as the one we favour (the criteria by which we do this are varied and usually based on biological/geological arguments). Or we can establish the common elements between the trees – the lowest common denominator if you like. For the second route we make consensus trees. There are several kinds of consensus trees that summarise different pieces of information. PAUP* reports four types, so we will deal with these here (you might like to be aware that there are more – see Kitching *et al.* 1988). Figure 1 steers you to the relevant part of the PAUP* program.

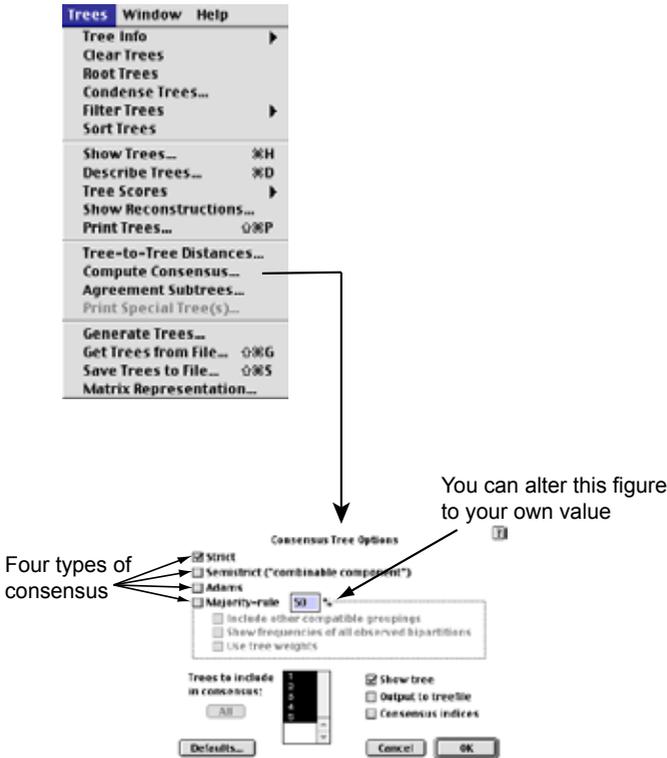
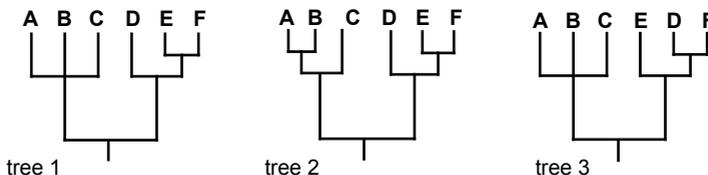


Figure 1. Control for calculating consensus trees is found under the trees menu. The control box appears for you to set options. You would normally include all fundamental trees.

Starting (fundamental) trees



Consensus trees

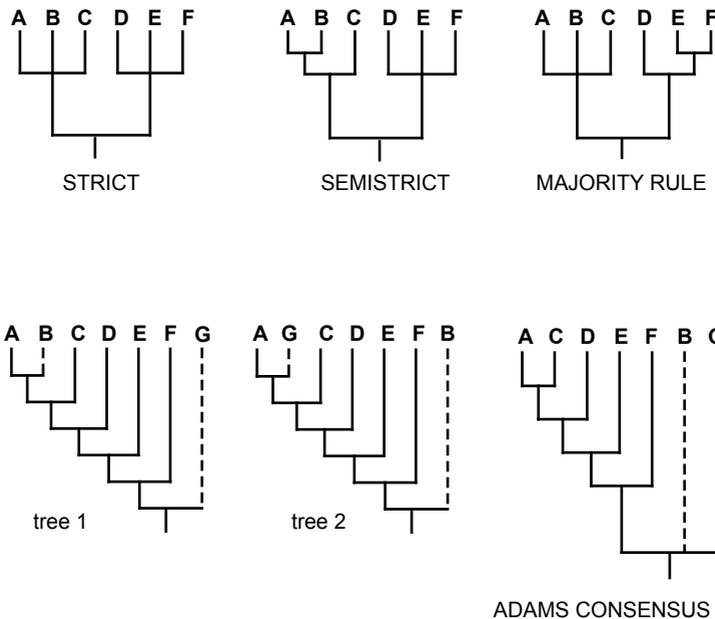


Figure 2. The types of consensus trees calculated in PAUP*. See text for explanation

Figure 2 illustrates the four kinds of consensus tree considered here. Let us assume that as a result of analysis we ended up with three equally parsimonious cladograms shown in the top row in Figure 2. These are called the starting or fundamental trees because they are the three alternatives derived from the analysis of the data.

The simplest way to combine the elements of all three cladograms into one is to show *only* those sister group pairings – or components – that appear in all three cladograms. Any differing solutions among the remaining taxa are shown as a single polychotomy. You will see by scanning across the three trees that relationships differ between A, B and C, and again between D, E and F. But the two groups ABC and DEF are the same in all trees. Therefore if we combine this information we end up with the tree to the left in the second row. This is known as the *Strict*



Consensus method. Many purists believe that this is the only consensus method that should be considered – all others being tainted by concessions that cannot be justified. Other practitioners think otherwise.

If we look more carefully at the starting trees we will see that in tree 1 and tree 3 there is a trichotomy between taxa A, B, C. In other words there is some ambiguity (this may result from conflicting data or perhaps no data, or alternative resolutions of question marks in the data set – palaeontologists beware!). One of the possible resolutions of that trichotomy is that taxa A and B are sister groups, with C the sistergroup of those combined. If we assumed this then all of the cladograms would be similar with respect to these three taxa and, in fact, there may be no conflict between them. [We can do nothing about taxa D, E and F since there is contradiction between the solutions seen in trees 1 and 2 on the one hand and tree 3 on the other.] Therefore another method – the *Semistrict Consensus* tree – will combine all those possible solutions that are not contradicted (this method is sometimes called the combinable component consensus).

The majority rule simply takes those solutions within the starting trees that are found in the majority of the trees. Thus the grouping (A,B,C) is found in two out of three trees and the grouping (D (E,F)) is also found in two out of three.

There is another kind of consensus we could make, and for this I have used two different kinds of starting trees, shown in the third row. This is called the Adams consensus. Let us assume that the result of analysis reported two trees that were the same shape (they need not be) but they differed in the positions of taxa B and G (dashed lines). The mutual relationships among the remaining taxa are the same. In the Adams consensus the taxa that differed in their positions (taxa B and G) are each placed at the most inclusive positions that each occupies in any of the starting trees. Since each of the taxa was positioned at the base of one or other of the starting trees, both are moved to the base of the Adams consensus tree. This type of consensus tree is useful for identifying ‘rogue’ taxa (and there are usually quite a few in palaeontological circles) – those taxa that occupy very different positions in different trees. You may think carefully about deleting such taxa from future analyses (we will return to what might be done in the final article): at the very least it would be wise to enquire as to *why* they occupied such differing positions. Although the Adams consensus may appear useful you should be aware that it is actually making a consensus of trees that were not in the starting line up. For example, one of the resolutions of the Adams consensus shown in Figure 2 is a sistergroup relationship between B and G, but that relationship was never part of the initial parsimony analysis!

Consensus trees are usually reported if more than one starting tree is obtained. BUT, they should not be used to infer anything about evolutionary pathways, rates *etc.* Remember, they are combinations of different theories of evolutionary pathways. They are used in various aspects of cladistic analysis. For instance, they are much used in vicariance biogeography, including palaeobiogeography (*Ed. there's another subject for a series of articles! – not for me though!!*) A common practice is to combine trees through consensus methods of different taxa inhabiting the same areas of the world to check for congruence and infer common explanations for common distributions. Consensus trees are also used to check the phylogenetic signal that may be given by different classes of data. There have been debates among cladists as to whether it is better to combine all the data into one large data set and analyse the lot together (character



congruence), or whether it is better to combine the trees that are produced from different data (taxonomic congruence). The most obvious situations are to use consensus methods to seek the commonality between the phylogenetic signal given by molecular data and that by morphological data, or between larval and adult morphologies. Probably this is less of an issue for palaeontologists. And they can be used for theories of co-evolution say, between hosts and parasites, or between evolutionary histories of flowers and pollinators.

Tree support

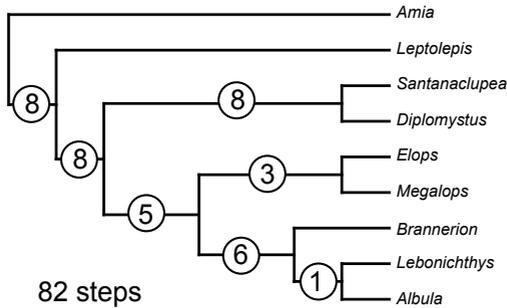
There are many measures that have been devised to try and express how good your tree is. 'Good' does not mean how accurate it is to reality but refers to several parameters of the tree itself. One class of measures estimate how much hierarchical structure there is in the tree. This means, how far away is your tree, in the number of steps, from random data. We came across one of these measures before (Fig. 15 in the *Tree Building* article) as the 'g' value. There are several others: but since they are not usually reported and even less understood we can glide quietly past them.

The other class of measures are those that estimate the support for individual nodes on the tree(s). These are usually reported and much discussed. There are two commonly used methods for morphological data: **Bremer support** and **the Bootstrap**.

Bremer support is by far the most useful for the amount of data we use as palaeontologists (we rarely have more 100 characters). Bremer support is named after the Swedish botanist Kore Bremer, who devised the method, but it is also known as the "Decay Index", for reasons that will become clear. The method asks the question: how much longer should the tree/cladogram be before a particular node collapses? The larger the number the stronger the support for that node. There are specific computer programs that will automatically calculate these numbers for you. But you can do it in PAUP*, and by doing so you will understand the method. As usual, it will be best to explain by example. In Figure 3 top (overleaf) the optimal tree is given for the interrelationships between eight teleost fishes and an outgroup. We are interested in the support for the individual nodes in the ingroup. This optimal tree is 82 steps long.

The first stage in calculating the Bremer support is to re-run the data, but this time we will keep the optimal tree *plus* all those trees *one* step longer. We do this on the tree searching menu. I have shown the Branch and bound menu here but the other searches have similar boxes. You will see that you can type in any number larger than the optimal length. In this case I have inserted 83. Re-running the data under the same conditions yielded two trees in this case. The next stage is to make a strict consensus of the two trees. This tree is shown bottom left in Figure 3. When this is done and compared to the original tree it can be seen that the original node supporting the sistergroup between *Albula* and *Lebonichthys* has collapsed, so that now there is a trichotomy between those taxa and *Brannerion*. This means that the original node supporting the sistergroup *Lebonichthys* + *Albula* collapsed after the addition of one step on the tree.

Now we repeat the process, increasing the number of trees to be saved to 84. In this case three trees were saved but there was no change in topology. At 85 steps, the node supporting *Elops* + *Megalops* collapsed. This is three steps longer than the original tree and therefore that node will be given a Bremer support of three, that we can insert back onto the original tree.



Add a value for trees of length X and less that you wish to save

Branch-and-Bound Search Options

Initial upper bound
 compute via stepwise
 set to

Keep
 minimal trees only
 all trees \leq score

Tree-score frequency distribution only:

Save all optimal trees (Multrees)

Enforce topological constraints
Constraints:
Keep trees that...
 Are Are NOT
...compatible with constraints

Addition sequence
 furthest
 as is
 simple
 max-mini
 max-mini (K)

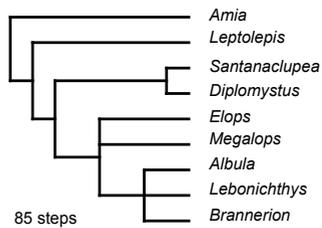
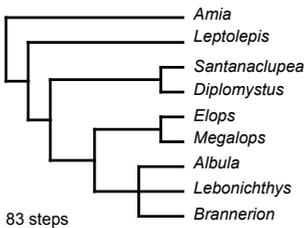
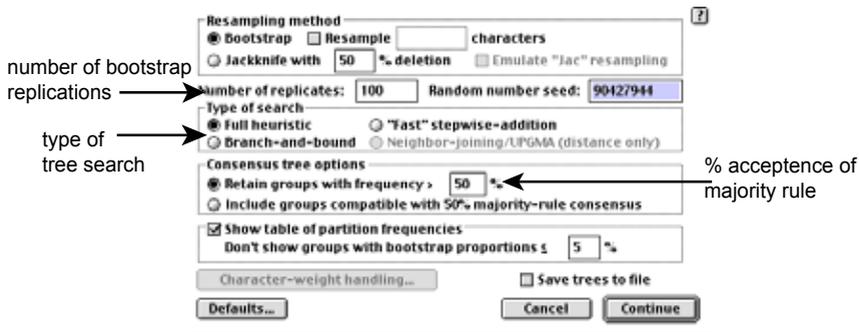


Figure 3. Calculation of Bremer support values shown on top tree. See text for explanation.



partition table

123456789	Freq	%
..**.....	998.75	99.9%
...*****	974.67	97.5%
.....***	974.65	97.5%
..*****	966.08	96.6%
...**...	960.18	96.0%
.....**	705.40	70.5%

50% majority Rule Consensus tree

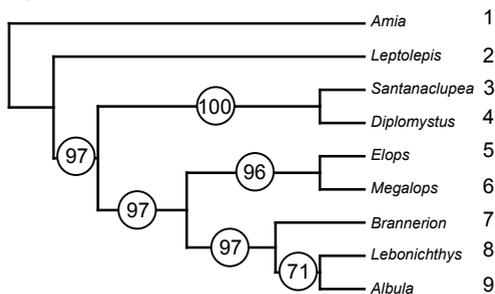


Figure 4 The Bootstrap See text for explanation

We carry on increasing the tree lengths to save by one each time and look for the nodes collapsing. By 90 steps (eight steps longer than the optimum) all nodes had collapsed to a single polychotomy (or unresolved tree). In other words we are deliberately decaying the tree to an unresolved bush – hence the Decay Index. (A common computer program used to calculate these numbers automatically is called TreeRot).

The Bremer support is certainly worth calculating and including in your papers. But it falls somewhat short of perfection because it does not tell you what kind of support each of the nodes has.



For instance, a particular node may have just one character supporting it and a Bremer index of just 1, but that character could be an unambiguous synapomorphy. Alternatively, a node with high Bremer value may be justified only by the weight of many homoplasious characters. You can really only check the quality by looking at the character change output.

Another technique that has been applied to measure node support is **the Bootstrap**. This is a statistical technique where the data matrix is sampled for characters, the data matrix is made up to its original size by duplicating some of the characters remaining, and the analysis performed again to see what groupings are found. This is repeated many times and a majority rule consensus tree computed with numbers against the nodes given from a partition table. The procedure is activated from the Analysis menu. Figure 4 shows the relevant screen where you can set how many replications you would like (I would recommend no fewer than 1000), what kind of tree search you want (Heuristic and Branch & Bound are the only options) and what set level of the majority rule consensus tree you want (you would normally leave this at 50%). After analysis two things appear: a partition table and the Majority rule consensus tree with numbers applied against the nodes. In the lower part of Figure 4 I have carried out a Bootstrap analysis of the same data set used in the previous figure. In this instance I set the number of replications to 1000. The translation between the partition table and consensus tree is obvious. The partition table tells us what percentage of the times of sampling characters particular groups were recovered. So, a grouping of taxa 8 (*Lebonichthys*) and 9 (*Albula*) was recovered in 70.5% of the samplings. A grouping of taxa 3 and 4 was recovered in all samplings, etc.

If you scan between the Bremer support values and the Bootstrap values for the same data set you will see that there is a very broad agreement, but there are anomalies. The Bremer support appears to be more discriminating than the Bootstrap. Many have criticised the Bootstrap being used for morphological data because there are relatively few characters and sampling may simply miss the inclusion of some characters more than others. Perhaps by increasing the number of replications this effect may be lessened. Bootstrap techniques are probably more effective with molecular data in which there are several thousand characters.

For completeness, some of you may notice that there is another sampling option in PAUP*. This is called the **Jackknife**: it is similar to the Bootstrap in that it is a repetitive sampling routine, but it samples taxa rather than characters. I have not seen it used with morphological data.

At the bottom line, Bremer support should be given for morphological data.

Peter Forey

REFERENCE

KITCHING, I. J., P. L. FOREY, *et al.* (1998). *Cladistics*. Oxford, Oxford University Press.

ACKNOWLEDGEMENTS

I would like to express my thanks to Sinaeur Associates, Inc., for permission to use 'screen dumps' from the PAUP* program as part of the illustrated material in this article.