

# ***REDCON 3.0***

## **REDUCED CONSENSUS PROGRAMS**

Mark Wilkinson  
Department of Zoology  
The Natural History Museum  
London, SW7 5BD  
England, U.K.

## PREFACE

**REDCON** was first developed in 1994 to implement the strict reduced cladistic consensus (RCC) and reduced partition consensus (RPC) methods (Wilkinson, 1994, 1995a). A second release, included additional programs for use in the implementation of majority-rule RCC methods (Wilkinson, 1996). **REDCON 3.0** features some additional analytical capabilities and other changes designed to improve ease of use. Previous versions used edited only PAUP 3.1.1 (Swofford, 1993) (Nexus) treefiles for input and these required some editing. The programs now accept PAUP 3.1.1 and PAUP 4.0 (Swofford, 1998) treefiles, and treefiles produced by Hennig86 (Farris, 1988) and these do not need editing. An important limitation of previous releases was that the programs did not output standard tree descriptions. The programs now write tree descriptions to a nexus format tree file that can be read into tree viewing software. A module to determine a recently developed measure of leaf stability (Thorley and Wilkinson, 1999) has been incorporated into the REDBOOT.EXE program, and additional features that can be used to filter and reduce output have been added to REDBOOT.EXE and MAJORITY.EXE. An additional program, SUCO.EXE has been added that allows support (as measured by bootstrap or jackknife proportions) to be determined for user-specified splits and identifies all conflicting full splits and their levels of support.

The programs suffer some severe limitations on the numbers of taxa that can be included in the fundamental trees and in the number of trees. Users should be aware of the existence of **RadCon** (Thorley & Page, 2000) which is a user-friendly consensus program for the Macintosh. RadCon can be used to produce strict reduced consensus trees, as well as a number of other techniques and measures that are unavailable in any other phylogenetic software. **RadCon** is also my preferred option for viewing trees output by **REDCON 3.0** and in my view any user of this package should also be using **RadCon**. It is available free of charge from:

<http://taxonomy.zoology.gla.ac.uk/~jthorley/radcon/radcon.html>

As far as I am aware, **REDCON 3.0** includes the only implementations of majority-rule reduced consensus trees and its main utility may be the potential for this method to enhance bootstrap or jackknife analyses (Wilkinson, 1996).

Your package should contain:

four turbo pascal executables:	example files and manual
STRICT.EXE	EXAMPLE.TRE (treefile)
MAJORITY.EXE	BOOT.TRE (jackknife tree file)
REDBOOT.EXE	EXAMPLE.SPL (splitfile)
SUCO.EXE	REDCON.PDF (manual)

## Restrictions and Citation

**REDCON 3.0** is distributed for the sole purpose of facilitating and promoting research and is a non-commercial product. The programs and this manual may be freely copied and distributed. The recommended citation is:

Wilkinson, M. 2001 REDCON 3.0: software and documentation. Department of Zoology, The Natural History Museum, London.

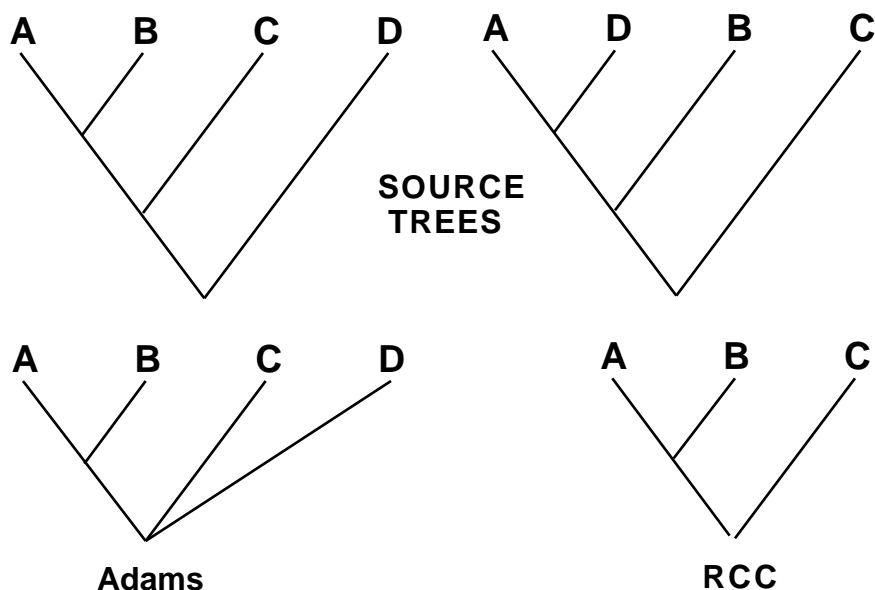
Please register as a user and report any bugs to me:

e-mail: [marw@nhm.ac.uk](mailto:marw@nhm.ac.uk) - Tel: +44 (0)20 7942 5164 - Fax: 7942 5433

© Mark Wilkinson, 2001

## INTRODUCTION

Consensus methods are used to provide graphical representations, consensus trees, of agreement among a set of *fundamental* or *source* trees. Not all of the desiderata we might have for the properties of a consensus method or consensus tree are readily satisfied by a single method, and there are many methods spanning a wide range of properties (see, Swofford, 1991; Wilkinson and Benton, 1996). In particular, consensus methods may differ with respect to the type of agreement among the fundamental trees and the level of the agreement (Page, 1992). The best known consensus method, usually referred to as the ‘*strict*’ consensus (Sokal and Rohlf, 1981) returns a single consensus tree that: (1) includes all the *leaves* (terminal taxa, OTUs) that are common to all the source trees, and (2). all the *components* (splits, clades if trees are rooted) that are common to all the source trees. I refer to this method as the *strict component* consensus method. It is ‘*strict*’ in the sense that it demands complete agreement, and it is a *component* consensus method because it represents only replicated components rather than any other information that might be shared by the source trees. Thus the name includes two qualifiers that relate firstly to the level of agreement, and secondly to the type of agreement (Wilkinson, 1994; Wilkinson & Thorley, 2001a). In contrast, the well known ‘*majority-rule*’ consensus method of Margush and McMorris (1981) has a less stringent requirement of relationships for them to be included in the consensus: that they be present in a majority of the source trees. Like the strict component method its focus is also upon components and it is referred to more accurately as the *majority-rule component* consensus method. Similarly the method commonly referred to as the ‘*semi-strict*’ or combinable component consensus method (Bremer, 1990) which produces a single tree including all components that are asserted by some and contradicted by no source trees, is here referred to as the *semi-strict component* consensus method to highlight the type of relationships or shared information that it represents. The first consensus methods were introduced by Adams (1972), including the method now generally referred to as the Adams consensus method (Adams, 1986). This method is also strict in the sense that it only represents relationships that are true of all the source trees. It differs crucially from component consensus methods in the type of information that it represents. Rather than representing common components, Adams consensus methods (the strict version and as yet uncharacterised majority-rule and semi-strict varieties) focus upon nestings. In phylogenetic terms, a group nests within another group if the last common ancestor of the first group is a descendant of the last common ancestor of the second group (see Adams, 1986; Wilkinson, 1994).



**Fig. 1. Two source trees and their (strict) Adams and reduced cladistic consensus trees. The strict component consensus tree is unresolved.**

It is widely recognised that the strict component consensus method has problems of insensitivity. Specifically, it may fail to represent relationships that are common to the source trees but which cannot be expressed as shared components. For example, the two source trees in Figure 1 agree that A and B are more closely related to each other than they are to C, i.e. (A,B)C and appear to differ only with respect to the position of D, but this information is not included in their strict component consensus tree. In contrast, their (strict) Adams consensus tree includes the nesting of A and B within A – D, which occurs in both the source trees. However, although the strict Adams consensus tree is more sensitive than the strict component consensus, it and the nestings it represents do not have an unambiguous cladistic interpretation (in terms of unequivocal statements that some taxa are or are not more closely related to each other than to some other taxa). Imagine that we swapped taxa C and D in the two source trees in Figure 1 so that it is now C that is unstable relative to the hypothesis that A and B are more closely related to each other than they are to, that is (A,B)D. This makes no difference to the nesting of A and B within A – D and no difference to the Adams consensus tree. Thus given only the Adams consensus tree we do not know whether the source trees agree that A and B are more closely related to each other than they are to C or whether they agree that A and B are more closely related to each other than they are to D. We can only conclude that A and B are more closely related to each other than they are either to C and/or D, that is (A,B)C and/or (A,B)D, and that further, this condition may be satisfied differently in different source trees (Wilkinson, 1994). Thus despite their greater topological sensitivity, strict Adams consensus trees also fail to provide an unambiguous representation of the cladistic information that is shared by the source trees.

Reduced consensus methods have been designed to provide more sensitive alternatives to component consensus methods that also do not suffer from the ambiguities inherent in the polytomies of Adams consensus trees. The basic units of information represented in *reduced cladistic* consensus (RCC) trees are *n*-taxon statements. An *n*-taxon statement expresses some cladistic relationships, namely that some specified set of leaves (e.g. A and C) are more closely related to each other (i.e. have a more recent common ancestor) than any of them are to some other specified set of leaves (e.g. D and F). Importantly, *n*-taxon statements need not include all the leaves that are common to the source trees. This *n*-taxon statement can be represented symbolically as (A,C)D,F, or as binary character data or an entry in a partition table. With the latter two representations, leaves that are included in the trees being compared but which are not referred to by the *n*-taxon statement (e.g. B and E) are represented with the '?' symbol. Component consensus trees include only shared *n*-taxon statements of maximal cardinality, those that refer to all of the leaves shared by the fundamental trees. In contrast, RCC methods produce trees of shared components and also any shared *n*-taxon statements of lower cardinality (those that refer to a subset of the taxa included in all the source trees). Representation of the latter requires that leaves not included in the *n*-taxon statements are not included in corresponding reduced consensus trees. An RCC *profile* may include a single tree or a number of trees, each of which has a different leaf set and represents different collections of *n*-taxon statements that are common to all (strict) or to a majority (majority rule) of the source trees. It is perhaps natural to prefer consensus trees that include all the leaves that are common to all the source trees and to prefer methods that produce a single consensus tree rather than a profile of consensus trees. All other things being equal, I share these general preferences. However, in many cases RCC profiles include consensus trees that are more informative than corresponding Adams or component consensus trees, and that provide a better representation of the agreement regarding cladistic relationships among a set of source trees (see also Thorley et al., 1998, and Wilkinson and Thorley, 2001b). That such trees may not include all the leaves that are common to a set of source trees is the price that must be paid for the greater informativeness of reduced consensus trees.

Unrooted trees do not include *n*-taxon statements (which are a rooted concept). Each internal branch in a tree splits the leaves into two subsets, so that unrooted trees comprise sets of *splits* or bipartitions of the included leaves. Under any rooting of the tree, at least one of the two subsets of leaves must be a clade (*n*-taxon statement of maximal cardinality), but which of the subsets are a clade depends on the position of the root. The unrooted analogue of an *n*-taxon statement, is an *n*-taxon partition, or more simply a split. Splits are *full* if all the leaves are assigned to subsets (in which case they are equivalent to components), and *partial* when only a subset of the leaves are partitioned. Reduced partition consensus (RPC) methods are the unrooted analogue of RCC methods and produce

trees summarising all  $n$ -taxon partitions (full and partial splits). Note that the component consensus of unrooted source trees can be determined by arbitrarily rooting the source trees on the same leaf, determining the consensus of the rooted trees (by focussing on clades ~ rooted components) and then unrooting the resulting consensus tree. RPC trees cannot be determined analogously through arbitrary rooting of source trees and application of the RCC method (see Fig. 2) and require a distinct algorithm that is equivalent to considering all possible rootings (Wilkinson, 1995a).

Reduced consensus profiles may comprise both basic and derivative trees. The set of basic trees jointly summarises all shared  $n$ -taxon statements/partitions whereas each basic tree includes some information that is not represented in any other combination of basic trees. Derivative trees represent fusions of basic trees. They may include more information than any single contributing basic tree, but will necessarily exclude some information present in each of the contributing basic trees (Wilkinson, 1995a). The reduced consensus programs in **REDCON 3.0** provide information only on strict reduced consensus profiles. Derivative trees can be constructed from this information but are not output by the programs.

Wilkinson (1994) further divided reduced consensus tree profiles into primary and secondary trees. Primary trees are those that have the greatest resolution, and where equally well resolved, the greatest number of leaves. All other trees are secondary. These two properties, resolution and number of leaves, relate to the information content of consensus trees. The information content of reduced consensus trees increases with both resolution and the number of leaves, and primary trees will often be the most informative trees in the profile. RadCon (Thorley and Page, 2000) includes a direct measure of the cladistic information content of strict consensus trees that can be used to select preferred trees from a reduced consensus profile (Thorley et al., 1998).

More complete descriptions of reduced consensus methods can be found in Wilkinson (1994, 1995a, 1996, and Thorley, 2000), the latter of which includes a more mathematical treatment. Some examples of their application can be found in Dyke & Thorley (1998), Gower & Wilkinson (1996), Upchurch (1998) Wilkinson (1995b, 1997, 1999), Wilkinson and Benton (1995, 1996), Wilkinson *et al.* (1996, 2000) and Wills *et al.* (1995, 1998).

## THE PROGRAMS: WHAT THEY DO

### **STRICT**

The STRICT.EXE program implements the intersection method described in Wilkinson (1994, 1995a) to identify all informative and non-redundant  $n$ -taxon statements/partitions that are common to all the fundamental trees. A statement is redundant if it is entailed by another statement that applies to more taxa. These statements are summarised in the output as a partition table. STRICT.EXE then determines how these  $n$ -taxon statements/partitions can be combined to produce all basic trees in the strict RCC/PC profile. These trees are written to a nexus format tree file called 'STRICT.TRE'. It can be used with up to 65,000 fundamental trees containing no more than 80 leaves. It accepts PAUP and Hennig86 format tree files.

### **MAJORITY**

The MAJORITY.EXE program differs from STRICT.EXE in that it identifies all non-redundant  $n$ -taxon statements and their frequencies. It treats all trees as rooted and does not implement the unrooted majority-rule PC method. With frequencies of occurrence across the fundamental trees taken into consideration the redundancy condition changes. An  $n$ -taxon statement is redundant if it is entailed by a statement of equal or greater frequency and greater cardinality (Wilkinson, 1996). Wilkinson et al. (2000) distinguished this as *numerical* as opposed to simply *structural* redundancy. The number of non-redundant  $n$ -taxon statements may be large and the investigator may not be interested in all the statements. A number of options are available for controlling and limiting output. A minimum frequency can be specified for  $n$ -taxon statements, as can a minimum number of included leaves, representing *frequency* and *cardinality* constraints respectively. It is also possible to discard statements that are non-redundant only by virtue of a difference in frequency that is less than some specified threshold using a 'collapse' option. In some cases the number of non-redundant  $n$ -taxon statements may exceed the capacity of the program. When this occurs options are provided to use only the subset of trees that have been processed to that point or to collapse the statements in memory and then continue processing additional trees. Using the collapse option in this latter context means that the reported frequencies are approximations. Subject to the various options and constraints the remaining non-redundant  $n$ -taxon statements are output in a partition table. MAJORITY.EXE then determines how these  $n$ -taxon statements/partitions can be combined to produce all basic trees in the corresponding majority-rule PC profile. These trees are written to a nexus format tree file called 'MAJORITY.TRE'. Where the number of trees is greater than 25 output of tree descriptions can be suppressed. Trees do not include compatible minority  $n$ -taxon statements. It accepts PAUP and Hennig86 format tree files and can be used with up to 65,000 fundamental trees including no more than 50 leaves.

### **REDBOOT**

The REDBOOT.EXE program is intended for use with trees produced in bootstrap or jackknife analyses. It differs from MAJORITY.EXE primarily in that each bootstrap (or jackknife) replicate has equal value. Multiple trees arising from analysis of a single replicate data set are downweighted to achieve this by a factor of  $1/m$ , where  $m$  is the number of trees produced from that replicate. As with MAJORITY.EXE trees are treated as rooted so that only the PC method is implemented. It offers the same options for restricting output, and tree descriptions are written to a file called 'REDBOOT.TRE' unless they are suppressed. REDBOOT.EXE also determines leaf stabilities and a measure of overall phylogenetic stability (Thorley and Wilkinson, 1999) from the partition table. RadCon (Thorley and Page, 2000) implements three measures of leaf stability for rooted or unrooted trees. REDBOOT.EXE reports only the simplest of these: the average bootstrap or jackknife support of the best supported resolution of each rooted triplet (3-taxon statement) including the leaf. The overall phylogenetic stability is the average of the leaf stabilities. Leaf stabilities are determined from the partition table rather than directly from the trees. Although this is more efficient, restrictions on the content of the partition table through frequency or cardinality constraints or through the use of the collapse option make leaf stabilities approximations. REDBOOT.EXE can be used with up to 65,000 fundamental trees containing no more than 50 terminal taxa, but the maximum number of bootstrap (or jackknife) replicates that can be analysed is 1000. It accepts only PAUP bootstrap or jackknife tree files.

### **SUCO**

The SUCO.EXE program is a simple utility for investigating support and conflict for relationships as evident from the trees produced by bootstrapping (or jackknifing). It allows the user to specify a particular split of interest (from the keyboard or from file) and returns its bootstrap (or jackknife) proportion, and the identity of any full splits that conflict with it and their bootstrap (or jackknife) proportions. As with REDBOOT.EXE, multiple trees arising from analysis of a single bootstrap (or jackknife) replicate data set are downweighted to achieve this by a factor of  $1/m$ , where  $m$  is the number of trees produced from that replicate. Splits can be unrooted ( $n$ -taxon partitions) or rooted ( $n$ -taxon statements) and may be full or partial. It can be used with up to 65,000 trees containing no more than 100 leaves but the maximum number of bootstrap (or jackknife) replicates that can be analysed is 1000. It accepts only PAUP bootstrap or jackknife tree files. I was motivated to write this utility after many hours of examining partition tables to determine the distribution of support for particular splits and their competitors.

### INPUT FILES

The programs require PAUP 3.1.1, PAUP 4.0 (Swofford, 1998) or Hennig86 treefiles. If transferring files from Macs to PCs or vice versa, I have found it useful to use a word processor, as this prevents the appearance of extraaneous symbols in the text which can sometimes accompany direct file transfer between platforms. Load the treefile into the word processor and save the file in appropriate format. Transfer the file, open in a word processor and save as a textfile.

SUCO.EXE will read user-specified splits from a split file. An example split file is given in BOX 1, and included as the EXAMPLE.SPL file.

**BOX 1.** An example split file.

```
2
000011111
0000?111?
```

The first line of a split file must begin with the **number of splits** included in the file. Splits are represented on subsequent lines by strings of '0's and '1's, and in the case of partial splits by '?'s. The '0's and '1' represent the leaves included in the two groups specified by the split. Their order in the string corresponds to the numbering of leaves in the tree file. Where trees are treated as rooted, splits correspond to *n*-taxon statements with '1' denoting membership of the inside set and '0' denoting membership of the outside set (Wilkinson, 1994).



## RUNNING THE PROGRAMS

To run any of the programs type its name and press enter (return) at the DOS prompt, or click on its name/icon under Windows. Each program requests input from the keyboard. Some explication of the required input is given below but in most cases the response is self-explanatory and error checking helps prevent mistakes and ensure appropriate input.

Enter Minimum Frequency of Statements (0-100) :

This option allows a frequency constraint (expressed as a percentage) to be applied. This restricts attention to statements that have as high or higher a frequency of occurrence (or bootstrap or jackknife proportion) as that specified by the user. Entering 100 selects the strict RCC method. Entering 0.5 selects the standard 50% majority-rule RCC method. Entering 0.9 selects a 90% majority-rule RCC method. Entering a value below 0.5 will select a 50% majority-rule RCC consensus but  $n$ -taxon statements occurring in a minority of trees will also be included in the partition table. Entering 0 includes all non-redundant  $n$ -taxon statements in the partition table.

Enter Minimum Number of Leaves (3-X) :

This option allows a cardinality constraint to be applied. This restricts attention to statements that apply to at least as many leaves as specified by the user. Entering the number of leaves in the source trees (X) selects the component consensus, entering 3 selects the full reduced consensus, entering intermediate values selects a constrained RCC consensus and may be useful in exploring the agreement between trees when there are many basic RCC trees.

Enter Maximum Difference for Collapse (0-30) :

Many  $n$ -taxon statements may be non-redundant solely by virtue of having slightly greater frequencies of occurrence than other statements. This can lead to very complicated output and a plethora of RCC trees. Thus Wilkinson (1996) suggested that it might be useful to restrict attention to those statements that are non-redundant by virtue of at least some specified minimum level of difference in frequency. This option is my first attempt to implement a form of this idea. It allows the user to make the criteria for non-redundancy more stringent by specifying a minimum difference in frequency (expressed as a percentage) and thereby suppresses and discards those statements that offer only minor increases in frequency over other statements that otherwise entail them (structural redundancy). The collapse function is used after the non-redundant  $n$ -taxon statements are determined and discards any statement that would be redundant if its frequency were reduced by the value input by the user. Note that if there are three statements X which entails Y which entails Z, with differences of frequencies for Z-Y of 4.5 and for Y-X of 4.5, then if a level of 5% is selected as the maximum difference for collapse then both Z and Y will be discarded despite the difference between X and Z of 9%. Alternative forms of suppression are possible but have not been implemented. Entering 0 ensures that the collapse function is not invoked. This collapse function may make output more digestible but, like cardinality and frequency constraints, it changes the nature of the consensus trees, typically yielding a subset of the full profile.

Output tree descriptions (y/n) :

Where the number of basic RCC trees exceeds 25, this option allows tree descriptions and tree file output to be suppressed.

Complexity Greater than Capacity on Tree X

- 1 - Collapse
- 2 - Quit now

Enter your selection (1 or 2) :

If the number of non-redundant  $n$ -taxon statements exceeds the capacity of the programs MAJORITY.EXE and REDBOOT.EXE provide the option of collapsing the  $n$ -taxon statements in memory and then continuing with the remaining trees (option 1) or summarising the results and producing consensus trees from the trees that have been processed thus far (option 2). Where the maximum difference for collapse is zero then option 2 is selected automatically.

Enter split:

After requesting the names of input and output files SUCO.EXE uses this to prompt for a split to analyse. The split can be typed directly from the keyboard using the symbols '0' and '1'. Other symbols are interpreted as '?'. Type the entire split before pressing return. When a split is entered from the keyboard, SUCO.EXE writes the split to screen and asks the user to verify it which provides the option of correcting mistakes or entering a different split. If instead of entering a split you enter **'f'** (lower case) then SUCO.EXE will prompt for the name of a split file and will proceed to analyse all splits in the split file.

### **OUTPUT FILES**

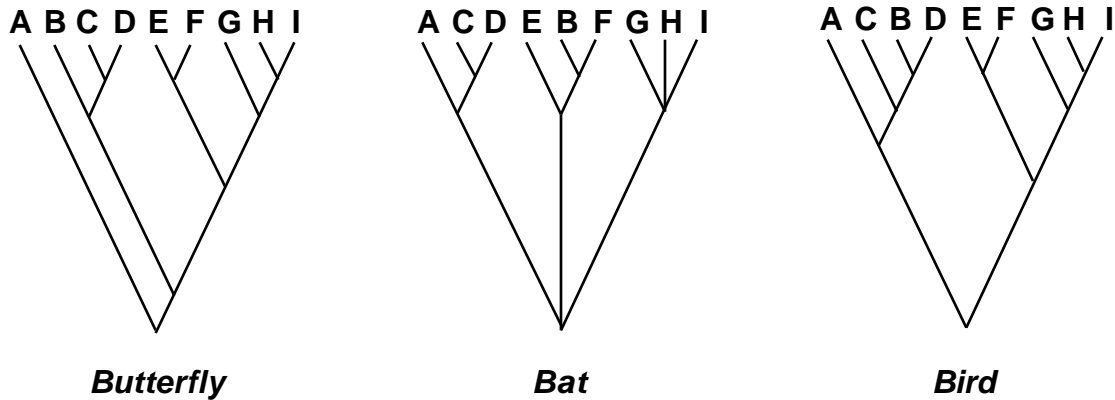
The main output files for the consensus programs summarise the options chosen for the analysis and gives a partition table of all non-redundant  $n$ -taxon statements or partitions and, where appropriate, their frequencies given as percentages.

Leaves are referred to by numbers that will typically be their order in the original data files from which the trees were produced. With PAUP treefiles these are the same numbers as used to refer to leaves in the treefile. With Hennig86 files, where the numbering begins with zero, the numbers are incremented by one compared to the treefile.

The output also includes lists of the combinations of taxa and  $n$ -taxon statements (or partitions) included in each basic RCC (or RPC) tree in the profile.

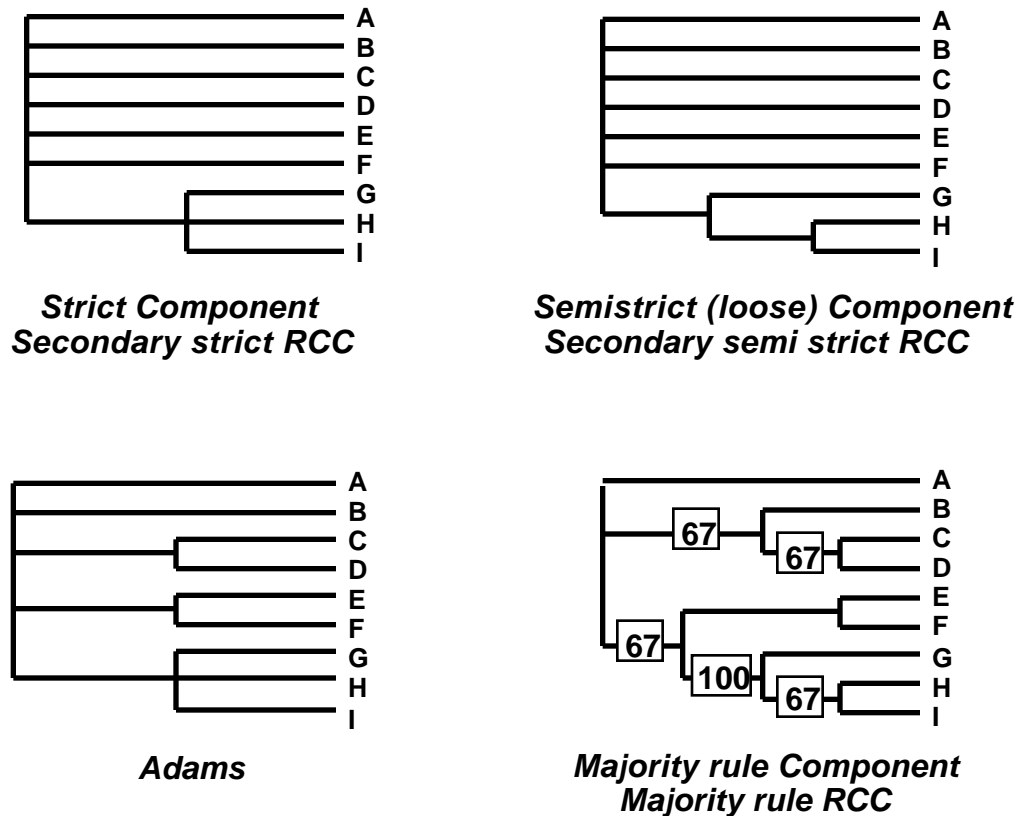
Because the treefiles may contain trees with different leaf sets they may be incompatible with some tree viewing/printing software. I recommend use of RadCon (Thorley and Page, 2000) to view trees.

## AN EXAMPLE



**Fig. 2. Three fundamental cladograms. Modified from Page (1989) and Bremer (1990).**

The three trees in the EXAMPLE.TRE file are the well known Butterfly, Bat and Bird cladograms (Fig. 1) that have been used to explore biogeographic and consensus methods by a number of workers (Nelson and Platnick, 1981; Page, 1989; Bremer, 1990). Some conventional consensus trees and reduced cladistic consensus trees are illustrated in Figures 2 and 3 respectively. BOX 2 and BOX 3 show the output from STRICT.EXE and MAJORITY.EXE respectively.



**Fig. 3. Standard consensus trees for the three source trees in Fig. 2.**

**BOX 2. Main output and information on trees produced by the STRICT.EXE program.**

```

Strict Reduced Cladistic Consensus Profile 13/3/2001
3 trees input from example.tre

Partition table of 3 n-taxon statements

      123456789
-----
1  .....***
2  .?..**...
3  .?**. ....
-----

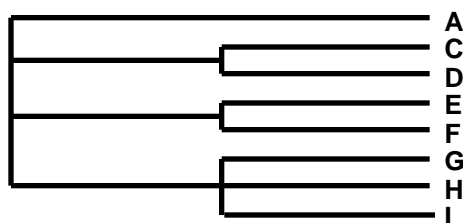
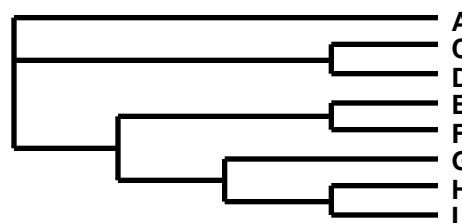
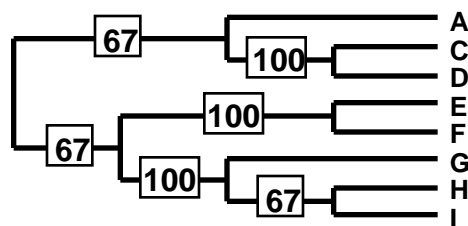
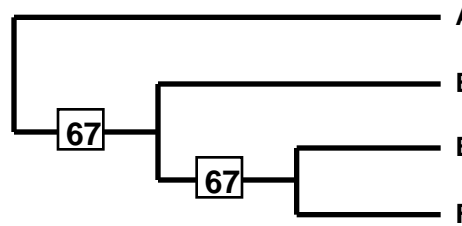
Groupings of n-taxon statements into 2 basic RCC
trees

(1) STRICT COMPONENT CONSENSUS - ALL LEAVES
Including 1 Statement: 1

(2) - Excluding leaf 2
Including 3 Statements: 1-3

```

STRICT.EXE identifies one component and two other  $n$ -taxon statements that exclude taxon 2 (B). The strict RCC profile includes two trees. One of these includes all the leaves and is the same as the strict component consensus tree (Fig. 2). The other excludes B and is a unique primary RCC tree (Fig. 3). It is more informative than the strict component consensus and the extra information it conveys is paid for through the exclusion of B.

**Primary strict RCC****Primary semistrict RCC****Primary majority rule RCC tree****Majority rule RCC tree****Fig. 4. Reduced cladistic consensus trees for the fundamental trees in Fig. 2.**

**BOX 3. Main output and information on trees produced by the MAJORITY.EXE program.**

```
Majority-Rule Reduced Cladistic Consensus Profile 13/3/2001
```

```
3 Trees input from example.tre
```

```
Minimum frequency (%) = 0.00
```

```
Minimum no. leaves = 3
```

```
Maximum difference for collapse = 0.00
```

```
Partition table of 16 n-taxon statements
```

	123456789	Frequency (%)
1	.....***	100.00
2	.?***.....	100.00
3	.?..**....	100.00
4	..**.....	66.67
5	.***.....	66.67
6	....**....	66.67
7	.....**	66.67
8	....*****	66.67
9	*?***.....	66.67
10	.*??*???	66.67
+++++		
11	.*****	33.33
12	.*.*.....	33.33
13	****.....	33.33
14	*.*.*.....	33.33
15	.*...*....	33.33
16	.*..**....	33.33

```
Groupings of n-taxon statements into 3 basic RCC trees
```

```
(1) MAJORITY-RULE COMPONENT CONSENSUS - ALL LEAVES
```

```
Including 6 Statements: 1 4-8
```

```
(2) - Excluding leaf 2
```

```
Including 6 Statements: 1-3 7-9
```

```
(3) - Excluding 5 leaves: 3-4 7-9
```

```
Including 2 Statements: 6 10
```

In this analysis, the minimum frequency was set at zero, the minimum number of taxa at three and no collapse function was employed. These settings should be used if you want the complete majority-rule RCC profile. MAJORITY.EXE identifies ten numerically non-redundant *n*-taxon statements that occur in a majority of the source trees. Six of these include all the leaves and correspond to tree 1, the majority-rule component consensus (Fig. 3) which is almost completely resolved. A second member of the majority rule RCC profile excludes leaf two and includes six statements. This tree is fully resolved and is the primary majority-rule RCC tree. It provides additional information on the relationships of leaf 1 (A) which is unresolved in the majority-rule component consensus. In addition exclusion of B changes the groups in the tree which are no longer clades or components but instead are the inside sets of *n*-taxon statements, and with this the frequencies associated with corresponding edges can increase. This happens with two of the internal edges in this tree whose frequencies increase from 66 to 100%. The majority-rule reduced consensus profile is completed by a third smaller tree that includes just five leaves but provides additional information on the relationships of leaf 1 (A) relative to leaves 2, 5 and 6 (B, E & F).

The three trees in the majority-rule RCC profile jointly display an interesting property. In a strict RCC profile all the  $n$ -taxon statements are present in all the source trees. The trees in the profile will therefore be compatible with all the source trees also with each other. This means that they can be combined with each other using either the simple tree fusion rules given in Wilkinson (1995a) or by the fusion rules for  $n$ -taxon statements recently described by Wilkinson et al. (2000). This does not hold in the case of majority-rule RCC profiles. Tree 2 includes the rooted triplet (C,B)A. Tree 3 includes the rooted triplet (E,B)A. Both of these triples occur in two of the three source trees. Ignoring their frequencies these two triplets entail a third rooted triplet (C,E)A. However this triplet occurs in only one of the source trees and is contradicted by the other two. This is because the two triplets are not always together, only in tree 1 do they co-occur. The example illustrates that it would be a mistake to attempt to combine the relationships in different trees in the majority rule profile using fusion rules for  $n$ -taxon statements as this may entail relationships that do not occur in a majority of the source tree. If derivative majority-rule RCC trees need to be constructed the tree fusion rules of Wilkinson (1995) should be used. Exact frequencies of the internal edges ( $n$ -taxon statements) included in the derivative tree must be determined from the partition table. The highest frequency of any corresponding internal edge in the parent trees provides only a lower bound for this frequency.

The figures include semistrict RCC profile for the three source trees. These includes the semistrict component consensus (Fig. 2) and a single additional and primary semistrict RCC tree (Fig. 3). Comparing the strict and semistrict component consensus trees (Fig. 1) the single difference is the inclusion of the H-I clade in the latter, which is not present in the Bat cladogram but is not contradicted by it (Bremer, 1990). Comparing the primary strict and semistrict RCC trees (Fig. 3), it is apparent that this same difference pertains, but that, in addition the (E-I)A-C  $n$ -taxon statement is also included. This relationship is also not included in the Bat cladogram, but is uncontradicted by it. This statement is represented at the expense of excluding B from the tree. The semi-strict profile was calculated by an experimental program that I hope to develop further and include in a future release.

BOX 4 shows the some of the output from a REDBOOT.EXE analysis of the set of bootstrap trees in the BOOT.TRE file. This analysis used minimum frequency of 85% and a cardinality constraint of six. The partition table summarises all numerically non-redundant  $n$ -taxon statements that occur in 85% or more of the bootstrap replicates and that apply to at least six leaves. The seven such  $n$ -taxon statements can be graphically summarised in five basic RCC trees. Note that only two  $n$ -taxon statements that include all the leaves have bootstrap proportions greater than 85%, the other five represent relationships that would have had their support underestimated by traditional component based summaries (consensus trees and partition tables). Several of the  $n$ -taxon statements are structurally non-redundant. Use of a collapse function would have suppressed some of these. For example by specifying a minimum difference of 2, statements 2 and 4 would have been suppressed in favour of statements 3 and 5 respectively with a concomitant decrease in the number of basic trees in the consensus profile.

**BOX 4. Main output and information on trees produced by the REDBOOT.EXE program.**

```

Majority-Rule Reduced Cladistic Consensus Profile 13/3/2001
56 Trees input from boot.tre                      Leaf Stabilities
20 bootstrap replicates                          -----
Minimum frequency (%) = 85.00                    1 -    1.000
Minimum no. leaves = 6                          2 -    0.596
Maximum difference for collapse = 0.00          3 -    0.596
Partition table of 7 n-taxon statements          4 -    0.596
123456789 Frequency (%)                        5 -    0.702
-----
1 .***** 100.00                               6 -    0.705
2 ...?***? 100.00                               7 -    0.739
3 ...?***? 99.00                                8 -    0.739
4 ....?***? 97.08                               9 -    0.649
5 .....***? 95.83                               -----
6 ...?**** 93.33                               Avge.    0.702
7 ....***** 92.33                               WARNING: With selected
-----                                         options stabilities are
                                         only approximate

```

The output also includes a table of leaf stabilities and a measure of overall phylogenetic stability, with a warning that these are approximate because of the cardinality and frequency constraints employed.

Box 5 shows the basic features of output from a SUCO.EXE analysis of the bootstrap trees in the BOOT.EXE program using the splits specified in the EXAMPLE.SPL file.

**BOX 5. Output from the SUCO.EXE program.**

SUCO analysis 14/4/2001	
*** 1 ***	*** 2 ***
=====	=====
123456789	123456789
=====	=====
92.33 000011111	100.00 0000?111?
-----	-----
6.667 011111110	=====
=====	

The output comprises a partition table for each split and its bootstrap (or jackknife proportion). The first split in each block is the selected split. Subsequent splits in the same block are those full splits that conflict with the selected split. The EXAMPLE.SPL file contained the one full and one partial split. The full split has a bootstrap proportion of 92.33 and there is a single split that also occurs in the bootstrap trees and that is incompatible with the selected split. The latter has a bootstrap proportion of (with a bootstrap proportion of 6.667. The partial split occurs in all the bootstrap trees and thus there are no conflicting splits.

### Acknowledgements

Development of reduced consensus methods was supported, in part, by SERC GR/87912 and by a grant from the University of Glasgow New Initiatives Fund. Thanks are due to those who grappled with earlier versions of the programs and provided feedback.

### References

- Adams, E. N. 1972 Consensus techniques and the comparison of taxonomic trees. *Systematic Zoology* **21**: 390-397.
- Adams, E. N. 1986 *N*-trees as nestings: Complexity, similarity and consensus. *Journal of Classification* **3**: 299-317.
- Bremer, K. 1990. Combinable component consensus. *Cladistics* **6**:369-372.
- Dyke G. J. & Thorley, J. L. 1998. Reduced cladistic consensus methods and the inter-relationships of Protoavis, Avimimus and Mesozoic birds. *Archaeopteryx* **16**: 123-129.
- Farris, J. S. 1988. *Hennig86. Version 1.5 software and documentation*. Port Jefferson Station, New York.
- Gower, D. J. & Wilkinson, M. 1996. Is there any consensus on basal archosaur phylogeny? *Proceedings of the Royal Society B* **263**: 1399-1406.
- Margush, T. & McMorris, F. R. 1981. Consensus *n*-trees. *Bulletin of Mathematical Biology* **43**: 239-244.
- Nelson, G. & Platnick, N. I. 1981. *Systematics and Biogeography: Cladistics and Vicariance*. Columbia Univ. Press, New York.
- Page, R. D. M. 1989. Comments on component-compatibility in historical biogeography. *Cladistics* **5**:167-182.
- Page, R. D. M. 1992. Comments on the information content of classifications. *Cladistics* **8**: 87-95.
- Sokal, R. R. & Rohlf, F. J. 1981. Taxonomic congruence in the Leptodomorpha re-examined. *Syst. Zool.* **30**:309-325.
- Swofford, D. L. 1991. When are phylogeny estimates from molecular and morphological data incongruent? pp. 295-333 in M. M. Miyamoto and J. Cracraft (eds.), *Phylogenetic Analysis of DNA Sequences*. Oxford University Press, Oxford.



- Swofford, D. L. 1993. *PAUP: phylogenetic analysis using parsimony*, version 3.1.1, Computer program distributed by the Illinois Natural History Survey, Champaign, Illinois
- Swofford, D. L. 1998. *PAUP\*: phylogenetic analysis using parsimony (\*and other methods)*, version 4. Sinauer Associates, Sunderland, Massachusetts.
- Thorley, J. L. 2000. *Cladistic information, leaf stability and supertree construction*. Ph.D. Thesis. Department of Biological Sciences, Faculty of Science, University of Bristol, UK.
- Thorley, J. L. & Page, R. D. M. 2000. RadCon: Phylogenetic tree comparison and consensus. *Bioinformatics*. **16**: 486-487.
- Thorley, J. L. & Wilkinson, M. 1999. Testing the phylogenetic stability of early tetrapods. *Journal of Theoretical Biology* **200**: 343-344.
- Thorley, J. L., Wilkinson, M. & Charleston, M.A. 1998. The information content of consensus trees. pp. 91-98. In (A. Rizzi, M. Vichi, and H.-H. Bock, eds.) *Advances in data science and classification*. Springer, Berlin.
- Upchurch, P. 1998. The phylogenetic relationships of sauropod dinosaurs. *Zoological Journal of the Linnean Society* **124**: 43-103.
- Wilkinson, M. 1994. Common cladistic information and its consensus representation: reduced Adams and reduced cladistic consensus trees and profiles. *Systematic Biology* **43**:343-368.
- Wilkinson, M. 1995a. More on reduced consensus methods. *Systematic Biology*. **44**:435-439.
- Wilkinson, M. 1995b. Coping with missing entries in phylogenetic inference using parsimony. *Systematic Biology*. **44**:501-514.
- Wilkinson, M. 1996. Majority-rule reduced consensus methods and their use in bootstrapping. *Molecular Biology and Evolution*. **13**:437-444.
- Wilkinson, M. 1997. Characters, congruence and quality: a study of neuroanatomical and traditional data in caecilian phylogeny. *Biological Reviews* **72**: 423-470.
- Wilkinson, M. 1999. Choosing and interpreting a tree for the oldest mammal. *Journal of Vertebrate Paleontology* **19**: 187-190.
- Wilkinson, M. & Benton, M. J. 1995. Missing data and rhynchosaur phylogeny. *Historical Biology* **10**:137-150.
- Wilkinson, M. & Benton, M. J. 1996. Sphenodontid phylogeny and the problems of multiple trees. *Philosophical Transactions of the Royal Society.B* **351**:1-16.
- Wilkinson, M., Suter, S. J. & Shires, V. L. 1996. The reduced cladistic consensus method and cassiduloid echinoid phylogeny. *Historical Biology*, **12**:63-73.
- Wilkinson, M. & Thorley, J. L. 2001a. No compromise on consensus. *Taxon* in press.
- Wilkinson, M. & Thorley, J. L. 2001. The efficiency of strict consensus trees. *Systematic Biology* in press.
- Wilkinson, M., Thorley, J. L. & Upchurch, P. M. 2000. A chain is no stronger than its weakest link: Double decay analyses of phylogenetic hypotheses. *Systematic Biology* **49**: 754-776.
- Wills, M. A., Briggs, D. E. G., Fortey, R. A. & Wilkinson, M. 1995. The significance of fossils in understanding arthropod evolution. *Verhandlungen der Deutschen Zoologischen Gesellschaft* **88**:203-215.
- Wills, M. A., Briggs, D. E. G., Fortey, R. A., Wilkinson, M. & Sneath, P.H.A. 1998. An arthropod phylogeny based on fossil and recent taxa. pp. 33-105 in Edgecombe, G. E. (ed.) *Arthropod fossils and phylogeny*. Colombia University Press.