

## RadCon: phylogenetic tree comparison and consensus

Joseph L. Thorley<sup>1, 2,\*</sup> and Roderic D. M. Page<sup>3</sup>

<sup>1</sup>School of Biological Sciences, University of Bristol, Bristol BS8 1UG, UK,

<sup>2</sup>Department of Zoology, The Natural History Museum, London SW7 5BD, UK and

<sup>3</sup>Division of Environmental and Evolutionary Biology, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK

Received on December 21, 1999; revised and accepted on February 2, 2000

### Abstract

**Summary:** RadCon is a Macintosh® program for manipulating and analysing phylogenetic trees. The program can determine the Cladistic Information Content of individual trees, the stability of leaves across a set of bootstrap trees, produce the strict basic Reduced Cladistic Consensus profile of a set of trees and convert a set of trees into its matrix representation for supertree construction.

**Availability:** The program is free and available at <http://taxonomy.zoology.gla.ac.uk/~jthorley/radcon/radcon.html>.

**Contact:** [j.l.thorley@bris.ac.uk](mailto:j.l.thorley@bris.ac.uk)

The holy grail of phylogenetics is the reconstruction of the one true tree of life. To overcome some of the obstacles on their quest phylogeneticists need tools for handling sets of trees. For example, uncertainties surrounding inferred relationships typically result in the analysis of a single data set yielding multiple trees. A problem that is compounded by the sensitivity of the results to both the method of analysis and the data set used. Furthermore, the computational and practical limitation of individual studies to particular parts of the tree of life means that our current knowledge of the tree is in the form of numerous, often conflicting, subtrees with partially overlapping leafsets.

A key difference between RadCon and many other tree programs is its ability to handle sets of trees with non-identical leafsets. A property that allows the implementation of subtree and supertree consensus methods.

Uncertainty in phylogenetic inference manifests itself in the form of multiple optimal or near-optimal trees. When faced with this occurrence, the most conservative approach is to use a *strict* (*sensu* Wilkinson, 1994) consensus method to summarize those relationships common to all the alternative trees. However, the two strict consensus methods most commonly used—Adams and

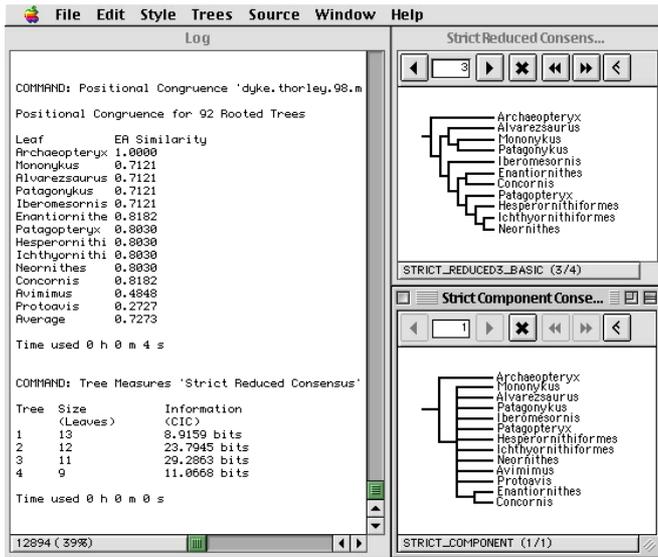
strict component—suffer from the problems of ambiguity and insensitivity, respectively (Wilkinson, 1994). Subtree consensus methods, for example, Largest Common Pruned Trees (Gordon, 1980) and strict Reduced Cladistic Consensus (RCC) trees (Wilkinson, 1994), overcome these problems through the exclusion of leaves, from the consensus, which prevent unambiguous representation of the remaining relationships. Although, these methods may output multiple consensus trees, the phylogeneticist now faces a less onerous problem as the trees are not conflicting alternatives but compatible and complementary. Should a single consensus tree be desired Thorley *et al.*'s (1998) Cladistic Information Content (CIC) provides an objective criterion for choosing between consensus trees.

RadCon is the first user-friendly implementation of strict RCC trees (Wilkinson, 1994), computing the strict basic RCC profile of sets of rooted and unrooted trees with up to 500 leaves and displaying them in a trees window. It is also the first program to calculate CIC doing this for individual strict component, Adams and strict RCC trees with up to 150 leaves.

Ultimately, uncertainty is caused by weak or misleading phylogenetic signals which warrants further investigation. The stability of phylogenetic relationships is commonly quantified by identifying their frequency of occurrence across a set of bootstrap trees (Felsenstein, 1985). However, almost all studies have focused exclusively on the stability of internal branches which can cause misleading conclusions to be drawn if just a few leaves are differentially responsible for any observed instability. Thorley and Wilkinson (1999) recently developed Leaf Stability (LS), a method which allows the stability of each leaf across a set of bootstrap trees to be quantified.

RadCon implements the newly developed LS method, for sets of rooted and unrooted bootstrap trees with up to 500 and 250 leaves, respectively. Unlike most other tree programs RadCon allows differential weighting of trees and distinguishes between bootstrap replicates allowing each bootstrap tree to be appropriately weighted.

\*To whom correspondence should be addressed.



**Fig. 1.** A screenshot of RadCon. Left is the log window, top right a trees window displaying a strict RCC tree and bottom right a trees window displaying the strict component consensus tree.

The strict RCC and LS methods help reduce uncertainty moving phylogeneticists closer to the true tree. However, the exponential explosion in the number of possible trees and the incompatibility of different types of phylogenetic data restricts individual studies to particular parts of the tree. The development of supertree consensus methods (Gordon, 1986; Sanderson *et al.*, 1998) means that trees with overlapping leaf sets can be fused together, a vital step in the reconstruction of the tree of life.

RadCon implements the Matrix Representation with Parsimony (MRP) supertree consensus method (Baum, 1992; Ragan, 1992), the method that has been used in the construction of most large supertrees. The user can choose between variants of the method, including Purvis and triplet coding, the only limitation being the size of the matrix which cannot exceed  $10^6$  entries. The matrix is written to a PAUP (Swofford, 1993) file for parsimony analysis. To facilitate supertree construction, trees in different files can be efficiently read in as a single set of trees.

RadCon supports the NEXUS format (Maddison *et al.*, 1997), a standard phylogenetic tree format. The user

interface includes a log window to record results of analyses, a text editor and trees windows. The program supports the drag-drop opening of files, printing, copy and pasting of graphics, text editing of data files and Balloon Help. A manual is available from the RadCon web site.

Figure 1 is a screen shot of a typical RadCon session. For an application of RadCon see Thorley and Wilkinson (1999).

## Acknowledgements

J.L.T. was supported by a University of Bristol PhD Studentship and a NERC Short-Term Fellowship in Taxonomy.

## References

- Baum, B. (1992) Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. *Taxon*, **41**, 3–10.
- Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783–791.
- Gordon, A. (1980) On the assessment and comparison of classifications. In Tommassone, R. (ed.), *Analyse de Donnees et Informatique*. INRIA, LeChesnay, France, pp. 149–160.
- Gordon, A. (1986) Consensus supertrees: the synthesis of rooted trees containing overlapping sets of leaves. *J. Classif.*, **3**, 335–348.
- Maddison, D., Swofford, D. and Maddison, W. (1997) Nexus: an extensible file format for systematic information. *Syst. Biol.*, **46**, 590–621.
- Ragan, M. (1992) Phylogenetic inference based on matrix representation of trees. *Mol. Phyl. Evol.*, **1**, 53–58.
- Sanderson, M., Purvis, A. and Henze, C. (1998) Phylogenetic supertrees: assembling the trees of life. *Trends Ecol. Evol.*, **13**, 105–109.
- Swofford, D. (1993) Paup: phylogenetic analysis using parsimony, version 3.1.1. Computer program distributed by the Illinois Natural History Survey.
- Thorley, J. and Wilkinson, M. (1999) Testing the phylogenetic stability of early tetrapods. *J. Theor. Biol.*, **200**, 343–344.
- Thorley, J., Wilkinson, M. and Charleston, M. (1998) The information content of consensus trees. In Rizzi, A., Vichi, M. and Bock, H.-H. (eds), *Advances in Data Science and Classification*. Studies in Classification, Data Analysis, and Knowledge Organization, pp. 91–98.
- Wilkinson, M. (1994) Common cladistic information and its consensus representation: reduced adams and reduced cladistic consensus trees and profiles. *Syst. Biol.*, **43**, 343–368.