

PROGRAM NOTE

Phylomatic: tree assembly for applied phylogenetics

CAMPBELL O. WEBB and MICHAEL J. DONOGHUE

*Department of Ecology and Evolutionary Biology, Yale University, Box 208106, New Haven, CT 06520–8106***Abstract**

Phylomatic (<http://www.phylodiversity.net/phylomatic>) is an online phylogenetic query tool where users submit a list of taxa (e.g. from an ecological community), with modern family and genus names, and which returns a phylogenetic hypothesis for the relationships among taxa. Any set of stored phylogenies, or a user-supplied one, can be chosen as the basis for the returned phylogeny, and several output formats for the tree can be selected. Currently, the source databases cover seed plants.

Keywords: Supertree, pruned phylogeny, angiosperms, community ecology

Received 18 May 2004; revision received 18 August 2004; accepted 17 September 2004

As phylogenetic thinking reaches all branches of biology, the number of potential users of phylogenetic information is greatly increasing. Researchers have begun to apply phylogenies beyond the traditional area of comparative biology, moving into biogeography, community ecology, and even ecosystem studies (Webb *et al.* 2002). For many years, the only way to obtain phylogenies for taxa of interest had been to perform phylogenetic analyses oneself, or in collaboration with systematists. However, as the amount of phylogenetic information in the literature and online has increased, so has the opportunity to use the phylogenies produced by others, especially when the user is interested in the 'deep' relationships among a set of taxa sparsely sampled from a large clade. For example, with the increasing resolution available for angiosperm plants (Chase *et al.* 1993; Soltis *et al.* 2000; APG 2003), ecologists can quite easily estimate the phylogenetic relationships among distantly related taxa in ecological communities (Prinzing *et al.* 2001; Silvertown *et al.* 2001). Most authors have produced such trees 'by hand', using the most comprehensive 'backbone' phylogenies available (e.g. Soltis *et al.* 2000), in combination with published or unpublished phylogenies of subgroups. This procedure is labour-intensive, especially with numerous taxa. We have produced a tool for the rapid construction of phylogenetic relationships, which offers an instant phylogenetic viewpoint for any set of higher plant taxa. While it can be argued that 'a little knowledge is a dangerous thing', we feel the democratization of phylogenetic information can only have a net beneficial effect on biology.

Phylomatic takes as input a list of taxa with family and genus information, matches the taxa to the most resolved position possible in any of a set of master trees in the database (the 'megatrees'), and returns the phylogeny in one of a number of alternative formats. We first developed this working model in 2000 as a tool for our own research in phylogenetic community ecology (Webb *et al.* 2002), but it has been receiving an increasing number of 'hits' ever since (currently six search events per day), indicating that it is fulfilling a need. Its goal is to be an explicitly evolving implementation of what is possible, rather than to be the definitive repository of angiosperm phylogenies that may eventually be subsumed into a larger project (e.g. the Tree of Life project, Maddison *et al.* 2001; TreeBASE, Sanderson *et al.* 1994; or APweb, Stevens 2001).

Before running a search, users are advised to compare their input list with a list of names of monophyletic family clades included in the megatrees using tools provided. It is up to the user to resolve synonyms with the names used in the master tree. Taxa with nonmatching family names submitted to a search are excluded from the returned tree. The user pastes a list of taxa (family, genus, species) into a text box. An output format is chosen: graphical, Newick (the standard, parenthetical phylogenetic format), Nexus (with or without internal node names), or a tabular format listing ancestral nodes as factors, useful for contrast tests. The user then chooses an internal 'megatree' to run the search over, or an optional user-supplied tree. The results are then displayed in the browser window, and can be saved to disk.

For each input taxon, a match in the megatree is first sought for the genus name. If this fails, a match is sought

Correspondence: Campbell O. Webb. Tel.: 925-229-3655; E-mail: campbell.webb@yale.edu

for the family name. Congeners are attached to a polytomous genus node. If all genera are matched within a family, the internal topology of the genera in the megatree is retained. However, if one or more genera are not found, or if no internal phylogeny is available for a family, the genera nodes are connected directly to a polytomous family node. Finally, family clades are connected using the super-familial resolution in the megatree, and the resultant, rooted phylogeny is returned.

The tree query application is separate from the trees themselves, and a user-supplied tree can even be chosen as the megatree. Users have the further option to manipulate the returned tree. However, for many in the target user group, the objective is to retrieve a directly useable phylogeny for their taxa, and so some thought must go into the set of megatrees on offer. We have opted for a pragmatic approach, providing two trees for the seed plants, based on the Angiosperm Phylogeny Group (APG) base tree at Peter Stevens' APweb (Stevens 2001). All monophyletic families in APG II (APG 2003) are included. To the terminal families in APweb, are adding an increasing number of family phylogenies in the literature. One tree presents a highly resolved hypothesis, the other a more conservative one, displaying uncertainty (less than 80% bootstrap support) as multiple polytomies. The users can thus 'bracket' the effects of phylogenetic uncertainty in their subsequent analyses. Full details of the decisions involved in megatree construction are given at the website. In addition, the current 'best-estimate' single supertree for the angiosperm families (Davies *et al.* 2004) is given as a megatree option. All megatrees used in Phylomatic are archived, so the original source phylogeny can be referred to. Future plans include incorporating branch-length information, adding intrageneric resolution to the megatrees, adding megatrees for fish and mammals, and distributing a stand-alone (offline) version.

For many biologists this tool might provide the first phylogenetic perspective on their taxa of interest. Such a perspective frequently stimulates a bout of excited, natural-history 'story-telling' as similarities among taxa are recognized to be based on common ancestry. This leads to the formulation of more concrete, testable hypotheses about phylogenetic conservatism (or convergence). Phylomatic can also greatly assist in testing for phylogenetic independence in studies of trait correlation and general comparative biology (Felsenstein 1985; Harvey & Pagel 1991). However, it is very important that all users recognize that the returned tree is highly 'pruned', and conclusions about ancestral states are highly provisional on the taxa included in the sample. Biogeographers have long used phylogenies in their exploration of the movement of taxa in space and time, but the phylogenies provided by Phylomatic may often be too coarse for the construction of area cladograms. However, multi-lineage comparisons of the biota of different areas (i.e. number of endemic species, genera, etc.) will strongly

benefit from a phylogenetic approach, as methods for 'comparative phylogenetic floristics' develop. Finally, community ecologists may be the largest user group for Phylomatic, dealing as they often do with assemblages of phylogenetically distantly related taxa, and beginning with little knowledge of phylogenetic biology. Because the characters that interact during community organization usually have a nonrandom phylogenetic distribution (either conserved or convergent), sometimes even at deep phylogenetic levels, the taxa that coexist often show a nonrandom phylogenetic structure relative to a pool of all possible species (Webb *et al.* 2002; Cavender-Bares *et al.* 2004). Assembling phylogenies for species pools and communities is the first step to assessing this phylogenetic structure.

Overall, we feel that all areas of biology can benefit from a move away from treating taxa as independent entities and toward recognizing explicit relationships in the tree of life. Phylomatic provides an easy way for biologists to begin to move towards this evolutionary viewpoint.

Acknowledgements

This project was partially funded by NSF grants DEB-0212873 and DBI-9876815. Many thanks to David Ackerly for extensive feedback, Rick Ree for programming advice, Kevin Livingstone for comments on the manuscript, and to all users of Phylomatic who have contacted us.

References

- APG (2003) An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society*, **141**, 399–436.
- Cavender-Bares J, Ackerly DD, Baum DA, Bazzaz FA (2004) Phylogenetic overdispersion in Floridian oak communities. *American Naturalist*, **163**, 823–843.
- Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plasmid gene *rbcL*. *Annals of the Missouri Botanical Garden*, **80**, 528–580.
- Davies TJ, Barraclough TG, Chase MW, Soltis PS, Soltis DE, Savolainen V (2004) Darwin's abominable mystery: insights from a supertree of the angiosperms. *Proceedings of the National Academy of Sciences*, **101**, 1904–1909.
- Felsenstein J (1985) Phylogenies and the comparative method. *American Naturalist*, **125**, 1–15.
- Harvey PH, Pagel MD (1991) *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford.
- Maddison DR, Maddison WP, Wheeler T, Mandel D (2001) *Tree of Life Web Project*. URL <http://tolweb.org>.
- Prinzing A, Durka W, Klotz S, Brandl R (2001) The niche of higher plants: evidence for phylogenetic conservatism. *Proceedings of the Royal Society of London, Series B. — Biological Sciences*, **268**, 2383–2389.
- Sanderson MJ, Donoghue MJ, Piel W, Eriksson T (1994) TREEBASE: a prototype database of phylogenetic analyses and an interactive tool for browsing the phylogeny of life. *American Journal of Botany*, **81**, 183.
- Silvertown J, Dodd ME, Gowing DJG (2001) Phylogeny and the

niche structure of meadow plant communities. *Journal of Ecology*, **89**, 428–435.

Soltis DE, Soltis PS, Chase MW *et al.* (2000) Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. *Botanical Journal of the Linnean Society*, **133**, 381–461.

Stevens PF (2001) *Angiosperm Phylogeny Website*. URL <http://www.mobot.org/MOBOT/research/APweb>.

Webb CO, Ackerly DD, McPeck M, Donoghue MJ (2002) Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, **33**, 475–505.