phylogram: an R package for phylogenetic analysis with nested lists

Shaun P. Wilkinson[1] **and Simon K. Davy**[1]

1 School of Biological Sciences, Victoria University of Wellington, P.O. Box 600, Wellington, New Zealand.

Summary

The R environment continues to gain popularity as a platform for bioinformatic analysis, due to the reproducible code-based workflow and the many powerful analytical tools available in a suite of open-source packages such as `ape` (Paradis, Claude, and Strimmer 2004), `phangorn` (Schliep 2011) and `Phytools` (Revell 2012). These packages typically employ a tree structure known as the "phylo" object, whose primary element is an integer matrix with one row for each edge in the graph, and two columns giving the indices of the connecting nodes. This is a versatile and memory-efficient structure suitable for most applications encountered by evolutionary biologists, and hence a comprehensive array of tools has been developed for editing, analyzing and visualizing trees in this format.

An alternative tree structure is the "dendrogram" object, whose nodes consist of deeply nested lists. While less memory-efficient than matrix-based trees, a useful feature of this representation is its modularity, whereby the sub tree of a tree is itself a tree - a dendrogram within a dendrogram. This means that dendrograms are subsettable in the same way that standard lists are, which facilitates intuitive command-line tree manipulation. An especially powerful feature of this object type is that tree-editing operations can be carried out recursively using fast inbuilt functions in the "apply" family such as `dendrapply` and `lapply`. There is also a large and growing number of resources for manipulating and plotting dendrograms in contributed packages such as `dendextend` (Galili 2015), and hence bi-directional conversion between "dendrogram" and "phylo" class objects would expand the range of tools available for both object types.

Here, we introduce `phylogram`, an R package for developing phylogenies as deeply-nested lists, converting trees between list- and matrix-type objects, importing and exporting trees to and from parenthetic text, an editing/manipulating dendrogram objects. `phylogram` is available from GitHub (https://github.com/ropensci/phylogram) and CRAN (https://CRAN.R-project.org/package=phylogram), and version 2.1 of the package is archived to Zenodo (http://dx.doi.org/10.5281/zenodo.1293634). A full reference manual with worked examples can be found at https://cran.r-project.org/web/packages/phylogram/vignettes/phylogram-vignette.html. Bug reports and other feedback are welcomed and can be directed to the GitHub issues page at http://github.com/ropensci/phylogram/issues, or the `phylogram` google group at https://groups.google.com/group/phylogram.

Acknowledgements

References

Galili, Tal. 2015. "dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering." *Bioinformatics* 31 (22):3718–20. https://doi.org/10.1093/bioinformatics/btv428.

Paradis, Emmanuel, Julien Claude, and Korbinian Strimmer. 2004. "APE: analyses of phylogenetics and evolution in R language." *Bioinformatics* 20:289–90. https://doi.org/10.1093/bioinformatics/btg412.

Revell, Liam J. 2012. "phytools: an R package for phylogenetic comparative biology (and other things)." *Methods in Ecology and Evolution* 3 (2):217–23. https://doi.org/10.1111/j.2041-210X.2011.00169.x.

Schliep, Klaus Peter. 2011. "phangorn: phylogenetic analysis in R." *Bioinformatics* 27 (4):592–93. https://doi.org/10.1093/bioinformatics/btq706.