

## BIO::PHYLO – PHYLOINFORMATIC ANALYSIS USING PERL

Rutger A. Vos, University of Reading, [R.A.Vos@Reading.ac.uk](mailto:R.A.Vos@Reading.ac.uk)

Recent years have seen the emergence of the field of *phyloinformatics*<sup>1</sup>. In the course of a phyloinformatic analysis, data and metadata are generated, transformed, filtered, analyzed and summarized before they can be interpreted to answer meaningful biological questions. Based on first principles of good science such steps should be reproducible; and, in practice, analysis steps often need to be redone by the researcher multiple times anyway<sup>2</sup> and are too error-prone, tedious and time-consuming to do by hand. Hence, phyloinformatic analyses benefit from increased automation.

The Bio::Phylo toolkit promotes this by giving easy access to phylogenetic data objects (trees, taxa, character state matrices) read from a variety of commonly used data formats. In addition, it allows researchers to read and write data formats for which no Perl support has existed so far (e.g. NeXML, “Pagel format”), compute heretofore unimplemented topological indices (various measures for tree balance<sup>e.g.3</sup>, branchiness<sup>e.g.4</sup> and phylogenetic distinctiveness<sup>e.g.5</sup>), apply heretofore unavailable sampling and resampling algorithms<sup>6</sup> and visualize the results in publication-ready (PDF) or web (SVG, SWF, PNG) graphics. Because of its compatibility with BioPerl<sup>7</sup> it can be easily integrated in larger bioinformatic workflows.

The toolkit is available from the Comprehensive Perl Archive Network<sup>8</sup> under the same licensing conditions as perl itself, i.e. the dual Artistic License / GPL scheme. The toolkit is thoroughly documented, with usage examples for all public methods and classes and a 15-page “birds eye view” manual.

---

<sup>1</sup> Cracraft, J. (2002) The Seven Great Questions of Systematic Biology: An Essential Foundation for Conservation and the Sustainable Use of Biodiversity, *Annals of the Missouri Botanical Garden*, **89**, 127-144.

<sup>2</sup> Noble, W.S. (2009) A quick guide to organizing computational biology projects, *PLoS computational biology*, **5**, e1000424.

<sup>3</sup> Mooers, A. and Heard, S. (1997) Inferring Evolutionary Process from Phylogenetic Tree Shape, *The Quarterly Review of Biology*, **72**, 31-54

<sup>4</sup> Rohlf, F., Chang, W., Sokal, R. and Kim, J. (1990) Accuracy of Estimated Phylogenies: Effects of Tree Topology and Evolutionary Model, *Evolution*, **44**, 1671-1684.

<sup>5</sup> Redding, D., Hartmann, K., Mimoto, A., Bokal, D., Devos, M. and Mooers, A. (2008) Evolutionarily distinctive species often capture more phylogenetic diversity than expected, *Journal of Theoretical Biology*, **251**, 606-615.

<sup>6</sup> Hartmann, K., Wong, D. and Gernhard, T. (In press) Sampling trees from evolutionary models, *Systematic Biology*.

<sup>7</sup> Stajich, J., Block, D., Boulez, K., Brenner, S., Chervitz, S., Dagdigian, C., Fuellen, G., Gilbert, J., Korf, I., Lapp, H., Lehvaslaiho, H., Matsalla, C., Mungall, C., Osborne, B., Pocock, M., Schattner, P., Senger, M., Stein, L., Stupka, E., Wilkinson, M. and Birney, E. (2002) The Bioperl Toolkit: Perl Modules for the Life Sciences, *Genome Res.*, **12**, 1611-1618.

<sup>8</sup> <http://search.cpan.org/dist/Bio-Phylo>