

***adephylo*: new tools for investigating the phylogenetic signal in biological traits**

Thibaut Jombart^{1,*}, François Balloux¹ and Stéphane Dray^{2,*}

¹Department of Infectious Disease Epidemiology, Imperial College - Faculty of Medicine, MRC Centre for Outbreak Analysis and Modelling, St Marys Campus, Norfolk Place, London W2 1PG, UK and ²Université de Lyon, F-69000, Lyon; Université Lyon 1; CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France

Associate Editor: David Posada

ABSTRACT

Summary: *adephylo* is a package for the R software dedicated to the analysis of comparative evolutionary data. Phylogenetic comparative methods initially aimed at accounting for or removing the effects of phylogenetic signal in the analysis of biological traits. However, recent approaches have shown that considerable information can be gathered from the study of the phylogenetic signal. In particular, close examination of phylogenetic structures can unveil interesting evolutionary patterns. For this purpose, we developed the package *adephylo* that provides tools for quantifying and describing the phylogenetic structures of biological traits. *adephylo* implements tests of phylogenetic signal, phylogenetic distances and proximities, and novel methods for describing further univariate and multivariate phylogenetic structures. These tools open up new perspectives in the analysis of evolutionary comparative data.

Availability: The stable version is available from CRAN: <http://cran.r-project.org/web/packages/adephylo/>. The development version is hosted by R-Forge: <http://r-forge.r-project.org/projects/adephylo/>. Both versions can be installed directly from R. *adephylo* is distributed under the GNU General Public Licence (≥2).

Contact: t.jombart@imperial.ac.uk; dray@biomserv.univ-lyon1.fr

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on April 12, 2010; revised on May 28, 2010; accepted on May 31, 2010

1 INTRODUCTION

Evolutionary comparative data consists of biological traits observed in several taxa forming the tips of a phylogenetic tree. Phylogenetic proximity most often engenders non-independence in the distribution of traits between closely related taxa, challenging the assumption of independence among observations required by classical statistical tools used for modeling relationships between traits. This phylogenetic autocorrelation, often referred to as *phylogenetic signal*, has initially been perceived as a source of nuisance in the analysis of life histories (Harvey and Pagel, 1991). Consequently, phylogenetic comparative methods (PCMs) first focused on partialling out the effects of phylogenetic signal in the correlative analysis of biological traits (e.g. Felsenstein, 1985; Grafen, 1989; Rohlf, 2006). However, precious information can be

gained from studying the phylogenetic signal, e.g. for identifying different selective regimes (Butler and King, 2004; Hansen, 1997; Hansen *et al.*, 2008), detecting shifts in correlation patterns on a phylogeny (Revell and Collar, 2009) or inferring changes in rates of evolution (O'Meara *et al.*, 2006). Moreover, phylogenetic signal can also indicate interesting evolutionary processes such as divergence of life histories, trade-offs or niche conservatism (Jombart *et al.*, 2010). The R package *adephylo* aims to complement existing PCMs by implementing tools for describing the phylogenetic signal and uncovering phylogenetic patterns from evolutionary comparative data.

2 FUNCTIONALITIES

This package extends and replaces implementation of phylogeny-related methods in *ade4* (Dray and Dufour, 2007), a package dedicated to multivariate analysis. As an R package (R Development Core Team, 2009), *adephylo* benefits from an incomparable diversity of statistical and computational tools, along with specific packages devoted to phylogenetics (see <http://cran.r-project.org/web/views/Phylogenetics.html>). In particular, the *ape* package (Paradis, 2006; Paradis *et al.*, 2004) implements a wide array of tools including tree reconstruction, graphics, trait simulation and modeling of phylogenetic data, whereas the *phylobase* package (Bolker *et al.*, 2010) has been developed for handling, plotting and manipulating phylogenetic comparative data.

To maximize data interoperability, *adephylo* uses objects both from *ape* and *phylobase*. Phylogenies can be reconstructed or read in from Newick format in *ape*, and data can be matched with tips, nodes or even edges using *phylobase*. *adephylo* implements different tools for visualizing, testing and describing the phylogenetic signal in multivariate datasets. The function `table.phylo4d` offers various options for plotting phylogenies and traits (Fig. 1), allowing for a direct visual assessment of phylogenetic structures. The package proposes four different metrics of phylogenetic dependency including the one underlying Abouheif's test (Abouheif, 1999), which seems particularly efficient for detecting phylogenetic signal (Pavoine *et al.*, 2008). These metrics allow integrating phylogenetic information in tests of phylogenetic autocorrelation (Abouheif, 1999; Gittleman and Kot, 1990) or autoregressive models (Cheverud and Dow, 1985). As they require intensive computations, these metrics have been implemented using compiled C code: this permits to compute phylogenetic dependencies between all pairs of tips even

*To whom correspondence should be addressed.

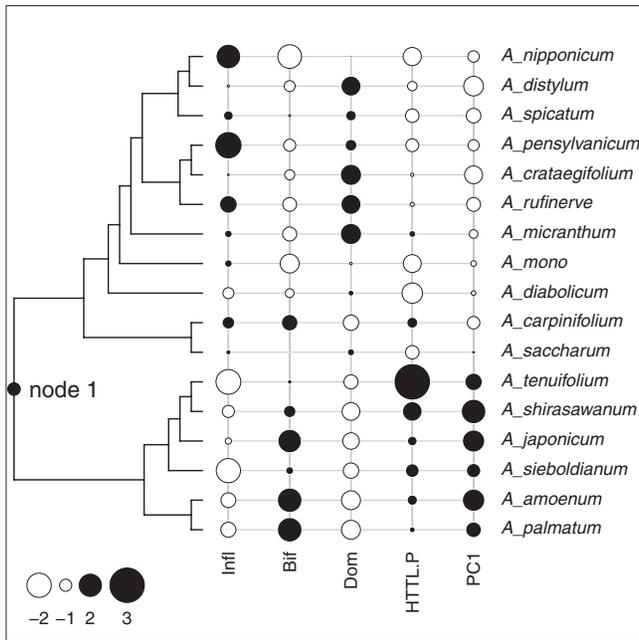


Fig. 1. Analysis of the maples data (dataset maples). This figure shows the phylogenetic tree along with the four traits identified as phylogenetically structured (Infl, Bif, Dom and HTTL.P) and the first global principal component of the pPCA of the data based on Abouheif's metric (PC1). All variables are centered and scaled. Positive and negative values are represented using black and white symbols, respectively, with a size proportional to the absolute value. See Appendix A (Supplementary Material) for details of the analysis.

in large phylogenies (i.e. including thousands of tips). These metrics can be used to code the phylogenetic structure either under the form of a matrix of distances (`distTips`), proximities (`proxTips`) or as orthonormal bases (`orthobasis.phylo`). These different representations of phylogenetic information give rise to a diversity of methods for the analysis of phylogenetic comparative data.

Phylogenetic proximities are used in the Abouheif's test (Abouheif, 1999; Pavoine et al., 2008), which provides a non-parametric procedure for detecting phylogenetic signal. Abouheif's test is implemented by the function `abouheif.moran`, which performs independent Monte Carlo tests for each trait of the analyzed dataset. This approach can be used together with other classical tests of phylogenetic signal such as Pagel's λ (Freckleton et al., 2002; Pagel, 1999) and Blomberg's K (Blomberg et al., 2003), the latter being implemented in the R package `picante` (Kembel et al., 2010).

As a complement to testing procedures, *adephylo* implements tools for describing phylogenetic structures. The decomposition of a quantitative trait onto an orthonormal basis describing the topology of the tree (Ollier et al., 2005) is implemented by the function `orthogram`. This approach proposes a series of statistical tests as well as a visual description of the phylogenetic signal, allowing to infer where phylogenetic dependence occurs in the tree. The function `orthobasis.phylo` extends this approach by proposing new orthonormal bases based on different phylogenetic metrics; therefore, allowing considerable flexibility in the description of phylogenetic relationships.

Lastly, *adephylo* also implements the phylogenetic principal component analysis (pPCA; Jombart et al., 2010), a novel method for investigating phylogenetic structures in multivariate comparative data. The function `ppca` performs the analysis, while other procedures (`summary`, `plot`, `screepplot` and `scatter`) allow for extracting useful information and plotting results of a pPCA.

adephylo is distributed with a tutorial and nine published datasets that are used to illustrate specific methodological points. In Appendix A (Supplementary Material), we illustrate some features of *adephylo* using the maples data published in Ackerly and Donoghue (1998). We first test for the presence of phylogenetic signal in the data using the function `abouheif.moran`, revealing four phylogenetically structured traits (Fig. 1). Using pPCA, we unveil a phylogenetic pattern opposing taxa from either side of node 1, which is remarkably similar to the four traits identified as phylogenetically structured, and points to the likely ancient divergence in life histories of the species considered (Fig. 1).

3 CONCLUSION

The *adephylo* package provides tools for a new approach of evolutionary comparative data, which complements usual PCM by quantifying, testing and describing the phylogenetic signal. Being embedded in the R environment, *adephylo* benefits from the access to a wide range of computational, statistical and graphical tools along with very complete packages devoted to phylogenetics, freely available to the scientific community. The advantages of such an integrative approach have already been acknowledged (e.g. Holland, 2007), and we hope this new package will contribute to building a unifying platform for the analysis of biological data within the R software.

ACKNOWLEDGEMENTS

We are very grateful to four anonymous reviewers for their useful comments on a previous version of the manuscript. We wish to thank Sébastien Devillard, Wilfried Thuiller and Julien Truchan for testing the first versions of *adephylo* and providing feedback on the package.

Funding: BBSRC and MRC (to T.J. and F.B.); ANR Diversitalp (ANR 07 BDIV 014 to S.D.).

Conflict of Interest: none declared.

REFERENCES

- Abouheif, E. (1999) A method for testing the assumption of phylogenetic independence in comparative data. *Evol. Ecol. Res.*, **1**, 895–909.
- Ackerly, D.D. and Donoghue, M.J. (1998) Leaf size, sapling allometry, and Corner's rules: phylogeny and correlated evolution in maples (Acer). *Am. Nat.*, **152**, 767–791.
- Blomberg, S.P. et al. (2003) Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution*, **57**, 717–745.
- Bolker, B. et al. (2010) *phylobase: Base package for phylogenetic structures and comparative data*. R package version 0.5.10. Available at: <http://r-forge.r-project.org/projects/phylobase/>.
- Butler, M. and King, A. (2004) Phylogenetic comparative analysis: A modeling approach for adaptive evolution. *Am. Nat.*, **164**, 683–695.
- Cheverud, J.M. and Dow, M.M. (1985) An autocorrelation analysis of genetic variation due to lineal fission in social groups of Rhesus macaques. *Am. J. Phys. Anthropol.*, **67**, 113–121.
- Dray, S. and Dufour, A.-B. (2007) The *ade4* package: implementing the duality diagram for ecologists. *J. Stat. Software*, **22**, 1–20.
- Felsenstein, J. (1985) Phylogenies and the comparative method. *Am. Nat.*, **125**, 1–15.

- Freckleton, R.P. *et al.* (2002) Phylogenetic analysis and comparative data: a test and review of evidence. *Am. Nat.*, **160**, 712–726.
- Gittleman, J.L. and Kot, M. (1990) Adaptation: statistics and a null model for estimating phylogenetic effects. *Syst. Zool.*, **39**, 227–241.
- Grafen, A. (1989) The phylogenetic regression. *Philos. Trans. R. Soc. Lond. Series B - Biol.*, **326**, 119–157.
- Hansen, T.F. (1997) Stabilizing selection and the comparative analysis of adaptation. *Evolution*, **51**, 1341–1351.
- Hansen, T.F. *et al.* (2008) A comparative method for studying adaptation to a randomly evolving environment. *Evolution*, **62**, 1965–1977.
- Harvey, P.H. and Pagel, M. (1991) *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford, UK.
- Holland, B.R. (2007) Analysis of phylogenetics and evolution with R. *Syst. Biol.*, **56**, 694–696.
- Jombart, T. *et al.* (2010) Putting phylogeny into the analysis of biological traits: a methodological approach. *J. Theor. Biol.*, **264**, 693–701.
- Kembel, S.W. *et al.* (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, **26**, 1463–1464.
- Ollier, S. *et al.* (2005) Orthonormal transform to detect and characterize phylogenetic signal. *Biometrics*, **62**, 471–477.
- O’Meara, B.C. *et al.* (2006) Testing for different rates of continuous trait evolution using likelihood. *Evolution*, **60**, 922–933.
- Pagel, M. (1999) Inferring the historical patterns of biological evolution. *Nature*, **401**, 877–884.
- Paradis, E. (2006) *Analysis of Phylogenetics and Evolution with R*. Springer, Heidelberg.
- Paradis, E. *et al.* (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.
- Pavoine, S. *et al.* (2008) Testing for phylogenetic signal in life history variable: Abouheif’s test revisited. *Theor. Popul. Biol.*, **73**, 79–91.
- R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Revell, L.J. and Collar, D.C. (2009) Phylogenetic analysis of the evolutionary correlation using likelihood. *Evolution*, **63**, 1090–1100.
- Rohlf, F.J. (2006) A comment on phylogenetic correction. *Evolution*, **60**, 1509–1515.