

Picante: R tools for integrating phylogenies and ecology

Steven W. Kembel^{1,*}, Peter D. Cowan², Matthew R. Helmus³, William K. Cornwell⁴, Helene Morlon⁵, David D. Ackerly², Simon P. Blomberg⁶ and Campbell O. Webb⁷

¹Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, OR, ²Department of Integrative Biology, University of California, Berkeley, CA, USA, ³Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Kunming, Yunnan, China, ⁴Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada, ⁵Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA, USA, ⁶Faculty of Biological and Chemical Sciences, University of Queensland, Brisbane, Australia and ⁷Arnold Arboretum of Harvard University, Cambridge, MA, USA

Associate Editor: David Posada

ABSTRACT

Summary: Picante is a software package that provides a comprehensive set of tools for analyzing the phylogenetic and trait diversity of ecological communities. The package calculates phylogenetic diversity metrics, performs trait comparative analyses, manipulates phenotypic and phylogenetic data, and performs tests for phylogenetic signal in trait distributions, community structure and species interactions.

Availability: Picante is a package for the R statistical language and environment written in R and C, released under a GPL v2 open-source license, and freely available on the web (<http://picante.r-forge.r-project.org>) and from CRAN (<http://cran.r-project.org>).

Contact: skembel@uoregon.edu

Received on February 16, 2010; revised on April 9, 2010; accepted on April 12, 2010

1 INTRODUCTION

Incorporating phylogenetic information into ecology is enlightening because it allows ecological questions to be addressed in an evolutionary context, leading to a deeper understanding of the processes that give rise to patterns of biological diversity (Webb *et al.*, 2002). Ecophylogenetic research also has numerous applications such as designating conservation priorities, predicting species invasiveness and estimating the effects of biodiversity on ecosystem functioning (Cavender-Bares *et al.*, 2009).

In this rapidly expanding field, two barriers have prevented ecologists from incorporating phylogenetic information into their own work. First, many ecophylogenetic methods are not available in open-source software packages, preventing end-users from modifying methods to suit their own purposes. Second, the available tools are implemented in multiple software packages, each with unique learning curves and data formatting requirements. Here, we describe a software package designed to overcome these limitations by implementing the most commonly used ecophylogenetics tools in a single open-source software package for the R statistical language and environment (R Development Core Team, 2009).

2 DESCRIPTION

2.1 Data manipulation and visualization

Picante uses the extensive data manipulation and data import/export functions included in R. Picante includes functions for reading and writing ecological data in formats used by other software packages including Phylocom (Webb *et al.*, 2008b), allowing these tools to be integrated into data analysis workflows in the R environment. By using the basic R plotting functions along with specifically designed functions in picante and other R packages (e.g. ape: Paradis *et al.*, 2004), phylogenetic patterns of trait evolution and community structure can easily be visualized.

2.2 Phylogenetic diversity and community structure

With the growing availability of phylogenies for entire ecological communities, there has been a rapid increase in the development and application of methods to quantify phylogenetic diversity. Picante contains functions for calculating the most commonly used measures of phylogenetic α -diversity (evolutionary relatedness within ecological communities) including Faith's phylogenetic diversity (PD, the sum of all branch lengths separating taxa in a community; Faith 1992), the taxonomic distinctness index (Vane-Wright *et al.*, 1991), the mean pairwise phylogenetic distance (MPD) and mean nearest taxon distance (MNTD) within a community (Webb *et al.*, 2008b), the phylogenetic species diversity metrics of Helmus *et al.* (2007b) [phylogenetic species richness (PSR), evenness (PSE), variance (PSV), and clustering (PSC)], quadratic entropy (Rao, 1982), and species co-occurrence and phylogenetic relatedness regressions (Slingsby and Verboom, 2006). Picante also calculates widely used measures of phylogenetic β -diversity (evolutionary relatedness between communities), including the phylogenetic Sørensen index (Bryant *et al.*, 2008), the UniFrac distance metric (Lozupone and Knight, 2005), and MPD and MNTD between communities (Webb *et al.*, 2008b). The availability of this suite of tools in a single package makes it easy to conduct comparative analyses of phylogenetic diversity on the same community dataset, which is important since different diversity metrics are sensitive to different aspects of the structure of ecological communities (Cadotte *et al.*, 2010).

Measures of phylogenetic α - and β -diversity can be compared to null expectations of community assembly and evolution to estimate

*To whom correspondence should be addressed.

community structure (i.e. whether communities are composed of closely related species or distantly related species). Picante performs numerous permutation procedures not available in other software packages including generating random communities by drawing taxa from a species pool, rearrangements of the community matrix preserving sample species richnesses and species occurrence frequencies and randomizations of phylogenies. When used with the various phylogenetic diversity metrics, these permutations test different null hypotheses of community structure and allow comparisons of phylogenetic diversity among communities that vary in species richness (Kembel, 2009).

2.3 Environmental correlates of phylogenetic community structure

Picante implements several methods to test for environmental drivers of phylogenetic community structure. First, picante calculates sample-based rarefaction curves of PSR (Helmus *et al.*, 2007b), allowing statistically valid comparisons of phylogenetic diversity among communities that differ in sampling intensity, size, area, etc. (Gotelli and Colwell, 2001). Second, picante performs species-environment regressions (Helmus *et al.*, 2007a) to test if closely related species respond to environmental variation in the same way and to separate determinants of phylogenetic community structure. Third, phylogenetic β -diversity metrics can be used by methods in other R packages (e.g. vegan: Oksanen *et al.*, 2009) to perform multivariate analyses of community-environment relationships including Mantel tests, phylogenetic clustering of communities and phylo-ordinations (Webb *et al.*, 2008a).

2.4 Phylogenetic signal in species traits and interactions

Phylogenetic signal is the tendency for taxa that share a common ancestor to resemble each other in how they look, behave and with whom they interact. Phylogenetic signal in the niches and traits of species violates the assumption that data points are statistically independent; techniques such as phylogenetically independent contrasts are available to overcome this non-independence (Paradis *et al.*, 2004). However, methods to quantify the amount of phylogenetic signal in phenotypic traits are not widely available.

Picante performs two kinds of statistical analyses of phylogenetic signal. The first is the calculation of the K statistic (Blomberg *et al.*, 2003), a Brownian motion-based metric of the strength of phylogenetic signal. Statistical significance of K is estimated by permuting trait values across the tips of a phylogenetic tree. The second implemented analysis is a phylogenetic regression of interaction strengths between species (e.g. plants and pollinators: Ives and Godfray, 2006) using an Ornstein–Uhlenbeck evolutionary model. Data on species traits can be included in the phylogenetic regression to estimate the relative contribution of traits versus phylogeny in explaining species interactions.

Funding: Chinese Academy of Sciences Research Fellowship for International Young Researchers (to M.R.H.); Google Summer of Code (to P.D.C. and M.R.H.); Gordon and Betty Moore Foundation (to S.W.K.); National Center for Ecological Analysis and Synthesis (to S.W.K. and M.R.H.); National Evolutionary Synthesis Center (to P.D.C. and S.W.K.); Natural Sciences and Engineering Research Council of Canada (to S.W.K.); US National Science Foundation (to M.R.H. and C.O.W.).

Conflict of Interest: none declared.

REFERENCES

- Blomberg, S.P. *et al.* (2003) Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution*, **57**, 717–745.
- Bryant, J.A. *et al.* (2008) Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. *Proc. Natl Acad. Sci. USA*, **105**, 11505–11511.
- Cadotte, M.W. *et al.* (2010) Phylogenetic diversity metrics for ecological communities: integrating species richness, abundance and evolutionary history. *Ecol. Lett.*, **13**, 96–105.
- Cavender-Bares, J. *et al.* (2009) The merging of community ecology and phylogenetic biology. *Ecol. Lett.*, **12**, 693–715.
- Faith, D.P. (1992) Conservation evaluation and phylogenetic diversity. *Biol. Conserv.*, **61**, 1–10.
- Gotelli, N.J. and Colwell, R.K. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett.*, **4**, 379–391.
- Helmus, M.R. *et al.* (2007a) Separating the determinants of phylogenetic community structure. *Ecol. Lett.*, **10**, 917–925.
- Helmus, M.R. *et al.* (2007b) Phylogenetic measures of biodiversity. *Am. Nat.*, **169**, E68–E83.
- Ives, A.R. and Godfray, H.C.J. (2006) Phylogenetic analysis of trophic associations. *Am. Nat.*, **168**, 1–14.
- Kembel, S.W. (2009) Disentangling niche and neutral influences on community assembly: assessing the performance of community phylogenetic structure tests. *Ecol. Lett.*, **12**, 949–960.
- Lozupone, C. and Knight, R. (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.*, **71**, 8228–8235.
- Oksanen, J. *et al.* (2009) *vegan: Community Ecology Package*. R package version 1.15-4.
- Paradis, E. *et al.* (2004) Ape: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.
- Rao, C.R. (1982) Diversity and dissimilarity coefficients: A unified approach. *Theor. Popul. Biol.*, **21**, 24–43.
- R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Slingsby, J.A. and Verboom, G.A. (2006) Phylogenetic relatedness limits co-occurrence at fine spatial scales: Evidence from the schoenoid sedges (Cyperaceae: Schoeneae) of the Cape Floristic Region, South Africa. *Am. Nat.*, **168**, 14–27.
- Vane-Wright, R.I. *et al.* (1991) What to protect? Systematics and the agony of choice. *Biol. Conserv.*, **55**, 235–254.
- Webb, C.O. *et al.* (2002) Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.*, **33**, 475–505.
- Webb, C.O. *et al.* (2008a) Ecological organization, biogeography, and the phylogenetic structure of tropical forest tree communities. In Carson, W.P. and Schnitzer, S.S.A. (eds) *Tropical Forest Community Ecology*. Blackwell, Oxford, pp. 79–97.
- Webb, C.O. *et al.* (2008b) Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics*, **24**, 2098–2100.