

Phylogenetics

pez: phylogenetics for the environmental sciences

William D. Pearse^{1,2,3,*}, Marc W. Cadotte⁴, Jeannine Cavender-Bares¹, Anthony R. Ives⁵, Caroline M. Tucker⁶, Steve C. Walker⁷ and Matthew R. Helmus⁸

¹Department of Ecology, Evolution, and Behavior, University of Minnesota, Saint Paul, MN, USA, ²Department of Biology, McGill University, Montréal, QC, Canada, ³Département des Sciences Biologiques, Université du Québec à Montréal, Succursale Centre-ville, Montréal, QC, Canada, ⁴Department of Biological Sciences, University of Toronto Scarborough, Scarborough, ON, Canada, ⁵Department of Zoology, University of Wisconsin, Madison, WI, USA, ⁶Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, USA, ⁷Department of Mathematics and Statistics, McMaster University, Hamilton, ON, Canada, and ⁸Department of Animal Ecology, Vrije Universiteit, Amsterdam, The Netherlands

*To whom correspondence should be addressed

Associate Editor: David Posada

Received on February 13, 2015; revised on April 7, 2015; accepted on April 25, 2015

Abstract

Summary: *pez* is an *R* package that permits measurement, modelling and simulation of phylogenetic structure in ecological data. *pez* contains the first implementation of many methods in *R*, and aggregates existing data structures and methods into a single, coherent package.

Availability and implementation: *pez* is released under the GPL v3 open-source license, available on the Internet from CRAN (<http://cran.r-project.org>). The package is under active development, and the authors welcome contributions (see <http://github.com/willpearse/pez>).

Contact: will.pearse@gmail.com

1 Introduction

Community phylogenetics (or eco-phylogenetics) combines ecology and evolutionary biology, linking ecological phenomena with the evolutionary processes that generate species and their traits (see Webb *et al.*, 2002). This growing field has produced a number of statistical tools and software code to implement them (e.g. Kembel *et al.*, 2010), but this code is disparate and handles data differently, making routine data analyses challenging. In many cases, published methods are not formally implemented in a software package and are available only as supplementary materials to papers. Without active (public) maintenance, these valuable techniques are effectively lost to the scientific community.

Here, we present the *pez* *R* (R Core Team, 2015) package which provides an intuitive framework that merges existing tools and supports the development of the next generation of eco-phylogenetic methods. It is based around an *R* class (*comparative.comm*) that

links phylogenetic, community, environmental and trait data in a single object. With a *comparative.comm* object, one can presently calculate over 30 phylogenetic and trait diversity metrics (described in the documentation for *pez.metrics*), including more than 10 that were previously unavailable in *R*. *pez* additionally implements statistical models (Cavender-Bares *et al.*, 2004; Ives and Helmus, 2011; Rafferty and Ives, 2013) to infer the processes that underlie patterns in community phylogenetic and trait data. Users can also perform simulations under various community assembly and evolutionary processes, to help test hypotheses about the structure of biodiversity.

2 Description

2.1 Data manipulation and storage

pez provides a unified class to contain eco-phylogenetic data, and provides wrapper functions to manipulate the species and site

composition of community, trait, environmental and phylogenetic data. Using the *phy.build* family of functions, users can generate phylogenies for their dataset (making use of methods previously implemented in *phyloGenerator*; Pearse and Purvis, 2013). *pez* integrates and makes use of much existing R code (Bortolussi *et al.*, 2013; Genz *et al.*, 2013; Koenker, 2015; Labiérté *et al.*, 2014; Oksanen *et al.*, 2015; Paradis *et al.*, 2004), and its *comparative.comm* class is directly compatible with all *caper* (Orme *et al.*, 2013) code, easing comparative analysis of species trait data.

2.2 Metrics

Following the classification of Pearse *et al.* (2014), *pez* simplifies the calculation and comparison of over 30 metrics by grouping them into four categories: *shape*, *evenness*, *dispersion*, and *dissimilarity*. Shape metrics measure the structure of a community phylogeny, while evenness metrics additionally incorporate species abundances. Dispersion metrics are used to examine whether phylogenetic biodiversity in an assemblage differs from the expectation of random assembly from a given set of species—a species pool. Finally, dissimilarity metrics measure the pairwise difference in phylogenetic biodiversity between assemblages. This classification links directly onto the kind of data the investigator has at hand: assemblage occupancy (*shape*), assemblage abundance (*evenness*), source pool (*dispersion*) and assemblage comparisons (*dissimilarity*).

Other packages (notably *picante*; Kembel *et al.*, 2010) allow for the calculation of some of the metrics contained within *pez*. *pez* unifies these metrics with other metrics not previously implemented under a common interface that can be extended to trait and transformed phylogenetic matrices (*sensu* Letten and Cornwell, 2015). For example, *pez* provides the first implementation of the *traitgram* framework (Ackerly, 2009; Cadotte *et al.*, 2010), allowing direct comparison of the extent of phylogenetic and functional trait community structure. *pez* also contains a flexible metric estimation and null-model generation suite (the *generic.metric* family), that can be used to easily compare different metrics and perhaps generate new *dispersion* metrics. New null models based around species' trait values (using *trait.asm*), and expected mean pairwise phylogenetic and trait distances (using *ConDivSim*) can also be simulated. The speed and ease by which *pez* allows for the user to compare many metrics is important because different metrics reveal separate aspects of eco-phylogenetic structure (Cadotte *et al.*, 2010).

2.3 Statistical models

pez implements the Cavender-Bares *et al.* (2004) regression framework (*fingerprint.regression*), and extends it to include more metrics of trait evolution. While community phylogenetics has been criticised for relying on the assumption of niche conservatism to explain ecological assembly (e.g. Mayfield and Levine, 2010), this early (yet previously unavailable) approach does not rely on niche conservatism. Instead, it regresses the correlation between species' trait similarity and observed degree of co-occurrence against summary statistics of the evolution of those traits. The approach goes beyond simply describing the phylogenetic or ecological structure of an assemblage, and instead draws links between how traits have evolved and how they play-out in their present-day ecological context.

The *pglmm* family of functions permit regression modelling of community (Ives and Helmus, 2011) and interaction network data (Rafferty and Ives, 2013) using the Phylogenetic Generalised Linear Mixed Model (PGLMM) framework. PGLMMs permit tests of the

mechanisms that underlie ecological and evolutionary structure in multiple communities. By fitting random effect terms that account for phylogenetic and trait co-variance, the user can statistically estimate how likely similar and dissimilar species are to (co-)occur as a function of their traits, environmental conditions, and the presence of interacting species.

2.4 Simulations

Hypothesis testing and the development of new eco-phylogenetic methods require simulation of data under known conditions, and *pez* contains functions to facilitate this. The *scape* family of functions (following Helmus and Ives, 2012) simulate species' distributions across environmental gradients. Species' environmental tolerances, range size and negative species interactions can be simulated to contain phylogenetic signal and generate patterns of phylogenetic structure at different spatial scales. The *sim.meta.phy.comm* family of functions simulate phylogeny, trait evolution and assemblage structure simultaneously across a landscape. Species' abundances across the landscape are proportional to the fit of each species' trait value to the conditions in each assemblage, and individuals migrate at random across the landscape. The *sim.meta.phy.comm* functions do not require the user to specify a phylogeny, and combined with the *phy.sim* family, provide a toolkit for the development of new simulation scenarios.

Acknowledgements

F. Boivin, E. Lind, B. Waring, two anonymous reviewers, and M. Pennell provided useful feedback.

Funding

This work was supported by NSF grant EF-1339441.

Conflict of Interest: none declared.

References

- Ackerly, D. (2009) Conservatism and diversification of plant functional traits: evolutionary rates versus phylogenetic signal *Proc. Natl. Acad. Sci.*, **106**, 19699–19706.
- Bortolussi, N. *et al.* (2013) The ecology of differences: assessing community assembly with trait and evolutionary distances. *Ecol. Lett.*, **16**, 1234–1244.
- 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.* (2004) Phylogenetic overdispersion in Floridian oak communities. *Am. Nat.*, **163**, 823–843.
- Dray, S. and Dufour, A.B. (2007) The ade4 package: implementing the duality diagram for ecologists. *J. Stat. Softw.*, **22**, 1–20.
- Genz, A. *et al.* (2013) *mutnorm: Multivariate Normal and t Distributions*. R package version 1.0-2.
- Helmus, M.R. and Ives, A.R. (2012) Phylogenetic diversity-area curves. *Ecology*, **93**, S31–S43.
- Ives, A.R. and Helmus, M.R. (2011) Generalized linear mixed models for phylogenetic analyses of community structure. *Ecol. Monogr.*, **81**, 511–525.
- Kembel, S.W. *et al.* (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, **26**, 1463–1464.
- Koenker, R. (2015) *quantreg: Quantile Regression*. R package version 5.11.
- Labiérté, E. *et al.* (2014) FD: measuring functional diversity from multiple traits, and other tools for functional ecology. R package version 1.0-12.
- Letten, A.D. and Cornwell, W.K. (2015) Trees, branches and (square) roots: why evolutionary relatedness is not linearly related to functional distance. *Methods in Ecology & Evolution*, **6**, 439–444.

- Mayfield, M.M. and Levine, J.M. (2010) Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecol. Lett.*, **13**, 1085–1093.
- Oksanen, J. et al. (2015) *vegan: Community Ecology Package*. R package version 2.2-1.
- Orme, D. et al. (2013) *caper: comparative analyses of phylogenetics and evolution in R*. R package version 0.5.2.
- Paradis, E. et al. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.
- Pearse, W.D. and Purvis, A. (2013) phyloGenerator: an automated phylogeny generation tool for ecologists. *Methods Ecol. Evol.*, **7**, 692–698.
- Pearse, W.D. et al. (2014) Metrics and models of community phylogenetics. *Modern Phylogenetic Comparative Methods and their Application in Evolutionary Biology—Concepts and Practice* (ed. Garamszegi L.Z.), Springer-Verlag, Berlin.
- R Core Team (2015) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rafferty, N.E. and Ives, A.R. (2013) Phylogenetic trait-based analyses of ecological networks. *Ecology*, **94**, 2321–2333.
- Webb, C.O. et al. (2002) Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.*, **33**, 475–505.