

UC Merced

Frontiers of Biogeography

Title

PCPS - an R-package for exploring phylogenetic eigenvectors across metacommunities

Permalink

<https://escholarship.org/uc/item/9z9639dh>

Journal

Frontiers of Biogeography, 6(3)

Authors

Debastiani, Vanderlei Julio
Duarte, Leandro da Silva

Publication Date

2014

DOI

10.21425/F5FBG22943

Copyright Information

Copyright 2014 by the author(s). This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

PCPS – an R-package for exploring phylogenetic eigenvectors across metacommunities

Vanderlei J. Debastiani* and Leandro D. S. Duarte

Programa de Pós-Graduação em Ecologia, CP 15007, Universidade Federal do Rio Grande do Sul, Porto Alegre, 91501-970, Brazil. www.ufrgs.br/leff

* vanderleidebastiani@yahoo.com.br

Abstract. PCPS is an R package for exploring phylogenetic eigenvectors across metacommunities. It offers a set of functions for analyzing principal coordinates of phylogenetic structure (PCPS), allowing analysis of phylogenetic signal in ecological traits of species at the metacommunity level, and the association between each PCPS and environmental, spatial and historical factors. The package is a flexible solution for exploring the distribution of major phylogenetic lineages across ecological or biogeographic gradients. The package is freely available on the CRAN official web server for R.

Keywords. analysis; ecophylogenetics; null model; phylogenetic patterns; phylogenetic signal.

Introduction

Several factors can affect species' distributions; one of them is the phylogenetic relationship among the clades that comprise the metacommunities. The profusion of studies on phylogenetic patterns for different lineages has made it possible for ecologists to explain the evolutionary basis of species' assembly into biological communities and to investigate the major mechanisms underlying it. Early work focused mainly on assessing whether species assembled in local communities are more or less phylogenetically similar than expected given a regional species pool (Webb et al. 2002). More recently, the distribution of different phylogenetic clades across ecological gradients has been explored using different approaches, giving birth to the new discipline of metacommunity phylogenetics (Leibold et al. 2010), a research field interested in unveiling the ecological and historical determinants of phylobetadiversity patterns (Duarte 2011, Peres-Neto et al. 2012, Duarte et al. 2014).

Principal coordinates of phylogenetic structure (PCPS - Duarte 2011, Duarte et al. 2012) constitute a useful tool for exploring phylogenetic patterns across a set of ecological communities. The method involves decomposing the phylogenetic information at the metacommunity level, which is defined using phylogenetic fuzzy weighting (Pillar & Duarte 2010) in several orthogonal

eigenvectors. Each eigenvector is a phylogenetic gradient for the set of communities, capturing the variations in the entire phylogeny, from basal to terminal nodes (Duarte et al. 2012). The advantage of this method lies in the possibility of exploring each phylogenetic gradient independently of the others; therefore, it is possible to evaluate which clades are related to each phylogenetic gradient across the metacommunity and to explore the identity of these clades driving phylobetadiversity patterns among the sites (Duarte et al. 2012). The PCPS are obtained from a species-composition matrix weighted by phylogenetic distances among species, a methodological approach that employs fuzzy set theory to scale pairwise phylogenetic distances among species up to the metacommunity level (Pillar and Duarte 2010, Debastiani and Pillar 2012). PCPS is an R package¹ released under open-source license, and freely available from CRAN². The package has a set of functions for the analysis of PCPS.

Features

The starting point of PCPS analysis is to arrange a set of data matrices: the community matrix, which may contain either species' presence-absence data or abundances; the pairwise phylogenetic distances between species; environmental/spatial/historical variables for each community (optional); and traits describing the species (optional). The PCPS package op-

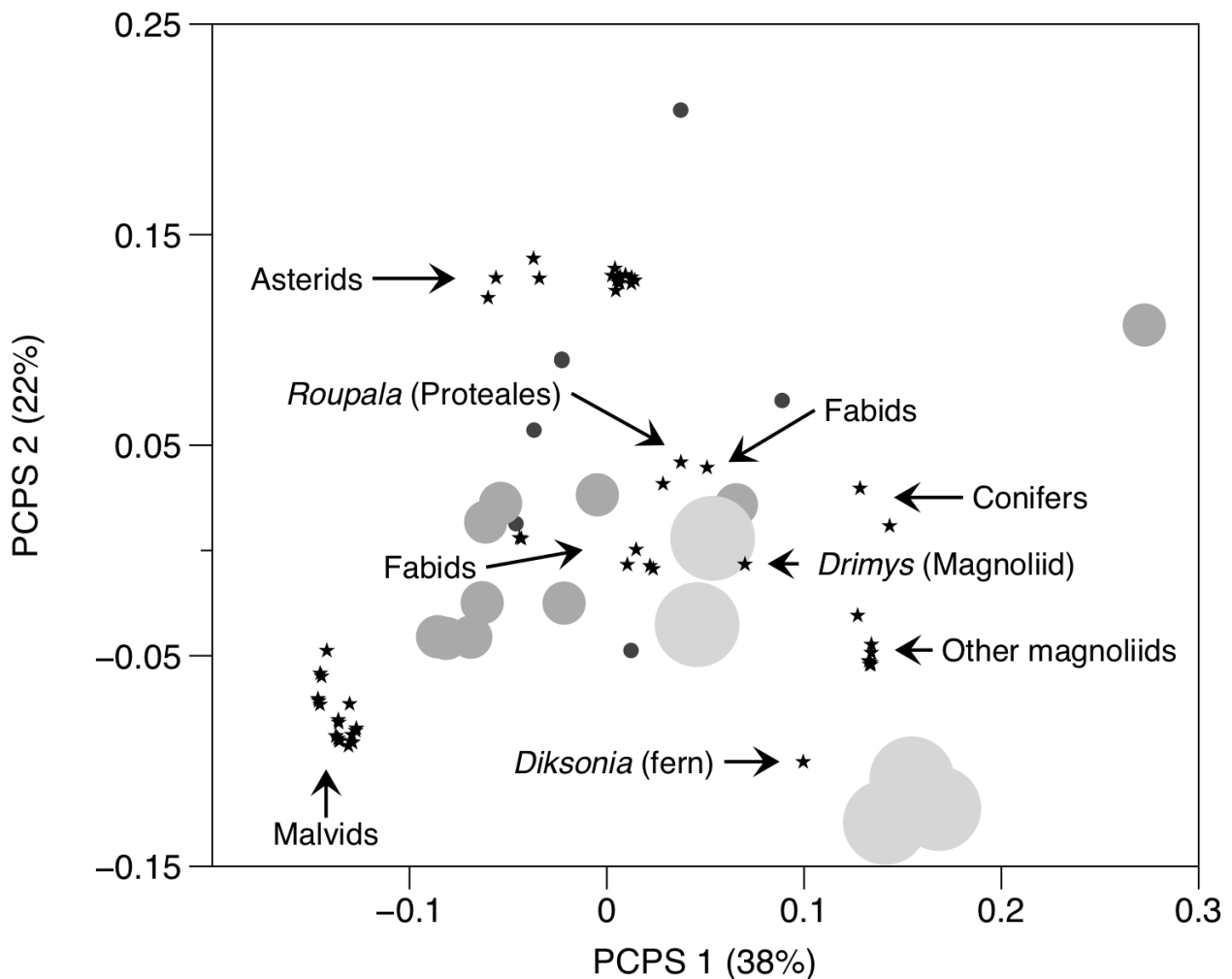


Figure 1. Scatter diagram of the first two PCPS axes computed for woody species occurring in forest patches of different sizes in Southern Brazil (modified from Duarte 2011). Circle size represents patch size (small, medium and large). Stars represent species grouped in monophyletic clades in the diagram. The nearness between the clades and the patches belonging to different sizes classes shows the association between them. For example, large patches were associated with *Dicksonia*, conifer trees and magnoliid angiosperms, whereas small patches were related to Asterids.

erates in an integrated manner with the SYNCOSA package (Debastiani and Pillar 2012), which is used to compute the matrix describing phylogeny-weighted species composition (function `matrix.p`). It can also be used to organize the data matrices (function `organize.syncosa`), given that the species and community sequence in the data matrices must be the same for all data matrices.

Principal coordinates of phylogenetic structure

The core function in the PCPS package is the function called `pcps`. This function performs a principal coordinates analysis (PCoA, Gower 1966) on the matrix describing phylogeny-weighted species

composition, thus generating the phylogenetic eigenvectors called PCPS (Duarte 2011). Each eigenvector represents a single phylogenetic gradient across the metacommunity, which is orthogonal to all other eigenvectors. The PCPS with higher eigenvalues describe wide phylogenetic gradients related to deeper nodes in the phylogeny, while other eigenvalues describe phylogenetic gradients related to shallower nodes (Duarte et al. 2012). Furthermore the `pcps` function computes correlations between each PCPS axis and phylogenetically weighted species abundances/frequencies, thus allowing biplots relating communities and species/clade scores to be built (Fig. 1).

¹ R Development Core Team, <http://www.r-project.org/>, last accessed 19/09/2014

² CRAN package repository, <http://cran.r-project.org/>, last accessed 19/09/2014

Phylogenetic signal at metacommunity level

The function `pcps.curve` estimates the phylogenetic signal at the metacommunity level (Pillar and Duarte 2010) for an ecologically relevant species trait. The first step for this analysis consists of scaling species trait information up to the metacommunity level, which is done by averaging trait values across the set of communities (Garnier et al. 2004). Then, the community-averaged trait is taken as the response variable in sequential linear regressions, in which an increasing number of PCPS are taken as predictor variables. Accordingly, in the first regression only the first PCPS axis is used, in the second regression the first two PCPS axes are taken as predictors, and so on. Finally, a curve representing the phylogenetic signal at the metacommunity level is drawn using the proportional accumulation of eigenvalues as new PCPS axes are incorporated into regression (x-axis) and the coefficient of determination of regressions (y-axis) (Diniz-Filho et al. 2012). That curve shows the degree of association between the community-averaged trait and the PCPS axes. A deviation of the curve under or over the 1:1 line, where there is a perfect match between the cumulative phylogenetic variability expressed by the PCPS and the R^2 of the community-averaged trait model, would indicate phylogenetic signal weaker or stronger, respectively, than that expected from the Brownian motion model of evolution (Diniz-Filho et al. 2012), according to which trait variance increases linearly with time.

Nonetheless, PCPS capture not only the phylogenetic signal across the metacommunity, but also the species composition. Therefore, to evaluate whether the curve is representing phylogenetic signal stronger or weaker than expected by Brownian trait evolution, we should control for the influence of species compositional variation across the metacommunity. For this, the function `pcps.curve` draws null curves, which are generated by shuffling the terminal tips across the phylogenetic tree (Bryant et al. 2008, Kembel et al. 2010) to compute a set of null PCPS. The null PCPS axes are taken as predictors of the linear regressions on the community-averaged trait, and generate curves under the scenario of a random dis-

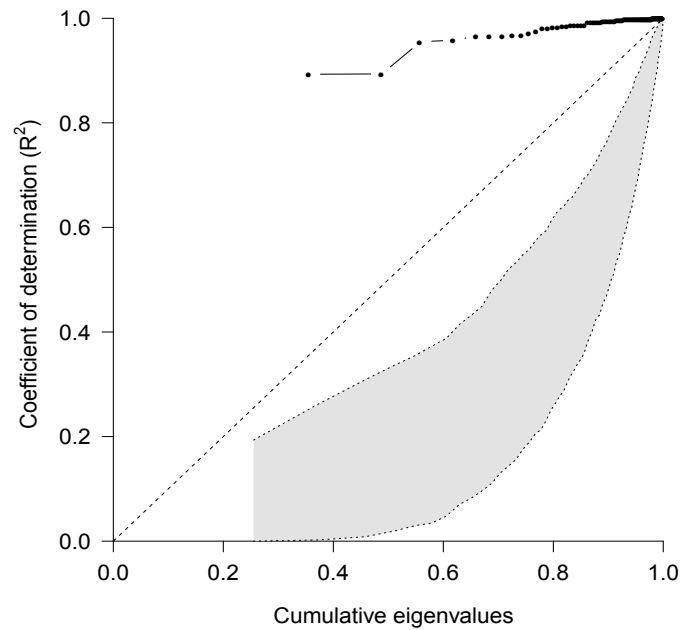


Figure 2. Hypothetical curve representing phylogenetic signal at the metacommunity level. The diagonal line represents the 1:1 ratio between cumulative eigenvalues of PCPS and coefficient of determination of the community-averaged trait regression on PCPS. The black dots represent the observed curve and the shaded area shows 95% confidence intervals based on null curves generated using a null model that shuffles the terminal tips across the phylogenetic tree.

tribution of species across phylogenetic tree. When the observed curve falls above the range of the null curves, it indicates that the phylogenetic signal at the metacommunity level is higher than expected merely by chance. On the other hand, if the observed curve falls under the range of null curves, it indicates that the association between community-averaged traits and the PCPS is lower than expected by chance (Fig. 2).

Association between PCPS axes and environmental factors

The function `pcps.sig` runs a generalized linear model (GLM) with a Gaussian error distribution (Nelder and Wedderburn 1972) to analyze the association between a single PCPS and a set of environmental and/or historical predictors (either categorical, quantitative, or both) (Debastiani et al. unpublished). The significance of the model is obtained by comparing its F-value with null F-values obtained from null models. The null models shuffle terminal tips across the phylogenetic tree

to randomize phylogenetic relationships among species, given a tree topology and branch lengths (Bryant et al. 2008, Kembel et al. 2010) and generate sets of null PCPS. The null PCPS are then submitted to a procrustean adjustment (Jackson 1995) and the fitted values between observed PCPS and null PCPS are obtained. Then, the adjusted null PCPS are taken as response variables, the model is rerun, and null F-values are generated. The fraction between the number of null F-values higher than the original F-value and the total number of null F-values computed is taken as the probability of the association between a given PCPS and a set of environmental variables being generated merely by chance.

Questions that can be answered using the PCPS approach

The package PCPS can be used to explore phylogenetic patterns in metacommunities, from finer (communities, assemblages) to broader spatial scales (biomes, continents). At finer scales PCPS analysis has been used to evaluate the association between the distribution of clades across different habitat types for woody plant communities developing over grasslands (Duarte 2011) and along natural grassland–forest ecotones (Debastiani et al. unpublished), and for avian communities distributed across a coastal gradient in Southern Brazil (Gianuca et al. 2014). At wider scales the method allows us to assess the extent to which the distribution of different phylogenetic clades along biogeographic gradients is determined by environmental conditions or spatial gradients. Some examples of this approach are available for woody plants in the Brazilian *Araucaria* forest biome (Duarte et al. 2012) and New World amphibians (Duarte et al. 2014). Furthermore, PCPS analysis enables us to evaluate turnover in phylogenetic composition through time (see Loyola et al. 2014 for an example for amphibians occurring in protected areas in the Brazilian Atlantic Forest) and also to analyze the extent to which the association between average trait values and environmental gradients is influenced by the phylogenetic composition of sites (Brum et al. 2012, 2013).

Conclusions

The field of ecophylogenetics has experienced accelerated development over the last few years. The PCPS package constitutes a flexible way to explore phylogenetic gradients across metacommunities using the same data manipulation ordinarily used to perform multivariate analysis in R. PCPS allows us to describe phylogenetic eigenvectors across metacommunities and to analyze their responses to environmental factors and the links between phylogenetic patterns and community-averaged traits of species.

References

- Brum, F.T., Kindel, A., Hartz, S.M. & Duarte, L.D.S. (2012) Spatial and phylogenetic structure drive frugivory in Tyrannidae birds across the range of Brazilian *Araucaria* forests. *Oikos*, 121, 899–906.
- Brum, F.T., Gonçalves, L.O., Cappelatti, L., et al. (2013) Land use explains the distribution of threatened New World amphibians better than climate. *PLoS ONE*, 8, e60742.
- Bryant, J.A., Lamanna, C., Morlon, H., Kerkhoff, A.J., Enquist, B.J. & Green, J.L. (2008) Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 11505–11511.
- Debastiani, V.J. & Pillar, V.D. (2012) SYNC-SA-R tool for analysis of metacommunities based on functional traits and phylogeny of the community components. *Bioinformatics*, 28, 2067–2068.
- Diniz-Filho, J.A.F., Rangel, T.F., Santos, T. & Bini, L.M. (2012) Exploring patterns of interspecific variation in quantitative traits using sequential phylogenetic eigenvector regressions. *Evolution*, 66, 1079–1090.
- Duarte, L.D.S. (2011) Phylogenetic habitat filtering influences forest nucleation in grasslands. *Oikos*, 120, 208–215.
- Duarte, L.D.S., Both, C., Debastiani, V.J., et al. (2014) Climate effects on amphibian distributions depend on phylogenetic resolution and the biogeographical history of taxa. *Global Ecology and Biogeography*, 23, 213–222.
- Duarte, L.D.S., Prieto, P. V. & Pillar, V.D. (2012) Assessing spatial and environmental drivers of phylogenetic structure in Brazilian *Araucaria* forests. *Ecography*, 35, 952–960.
- Garnier, E., Cortez, J., Billes, G., et al. (2004) Plant functional markers capture eco-system properties during secondary succession. *Ecology*, 85, 2630–2637.
- Gianuca, A.T., Dias, R.A., Debastiani, V.J. & Duarte, L.D.S. (2014). Habitat filtering influences the phylogenetic structure of avian communities across a coastal gradient in southern Brazil. *Austral Ecology*, 39, 29–38.
- Gower, J.C. (1966) Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 54, 325–338.
- Jackson, D.A. (1995) Protest: A Procrustean Randomization test of community environment concordance. *Ecoscience*, 2, 297–303.

- Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P. & Webb, C.O. (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, 26, 1463–1464.
- Leibold, M.A., Economo, E.P. & Peres-Neto, P. (2010) Metacommunity phylogenetics: separating the roles of environmental filters and historical biogeography. *Ecology letters*, 13, 1290–1299.
- Loyola, R.D., Lemes, P., Brum, F.T., Provete, D.B. & Duarte, L.D.S. (2014). Clade-specific consequences of climate change to amphibians in Atlantic Forest protected areas. *Ecography*, 37, 65–72.
- Nelder, J.A. & Wedderburn, R.W.M. (1972) Generalized Linear Models. *Journal of the Royal Statistical Society Series A*, 135, 370–384.
- Peres-Neto, P.R., Leibold, M.A. & Dray, S. (2012) Assessing the effects of spatial contingency and environmental filtering on metacommunity phylogenetics. *Ecology*, 93, S14–S30.
- Pillar, V.D. & Duarte, L.D.S. (2010) A framework for metacommunity analysis of phylogenetic structure. *Ecology Letters*, 13, 587–596.
- Webb, C.O., Ackerly, D.D., McPeck, M. A. & Donoghue, M.J. (2002) Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, 33, 475–505.

Submitted: 25 June 2014

Accepted: 21 August 2014

Edited by Marcus Cianciaruso