

spacodiR: structuring of phylogenetic diversity in ecological communities

Jonathan M. Eastman^{1,2,*}, C. E. Timothy Paine³ and Olivier J. Hardy^{4,*}

¹Department of Biological Sciences, ²Institute for Bioinformatics and Evolutionary Studies (IBEST), University of Idaho, Moscow 83844, ³Institut für Evolutionsbiologie und Umweltwissenschaften, Universität Zürich, Zürich CH-8057 Switzerland and ⁴Evolutionary Biology and Ecology Unit, CP 160/12, Faculté des Sciences, Université Libre de Bruxelles, B-1050 Brussels, Belgium

Associate Editor: David Posada

ABSTRACT

Motivation: SPACODiR is a cross-platform package, written for the R environment, for studying partitioning of diversity among natural communities in space and time. Complementing and extending existing software, SPACODiR allows for hypothesis testing and parameter estimation in studying spatial structuring of species-, phylogenetic- and trait diversities.

Availability: Integrated with other software in the R environment and with well documented and demonstrated functions, SPACODiR is an open-source package and available at <http://cran.r-project.org>.

Contact: jonathan.eastman@gmail.com; ohardy@ulb.ac.be

Received on April 5, 2011; revised on June 10, 2011; accepted on July 1, 2011

1 INTRODUCTION

Until recently, the formalized importance of historical processes in the development of communities had been largely neglected (see Webb *et al.*, 2002). With ongoing development in the field of community phylogenetics, time-dependent processes can now be directly integrated into the study of community diversity and assembly. Yet, available software for such analyses is often idiosyncratic, requiring unique data formats or a particular platform (but see Kembel *et al.*, 2010). Our aim here is to provide a cross-platform package (SPACODiR) that can inform the processes contributing to sorting of phylogenetic-, species- and trait diversities among communities. To this end, we use the highly flexible and actively developed R environment (R Development Core Team, 2010), providing SPACODiR as an adaptation and extension of the C-coded software SPACoDi (Spatial and Phylogenetic Analysis of Community Diversity) developed by Hardy (2010). Here, we describe the central functionality of the SPACODiR package. SPACODiR is poised to study scale dependencies of diversity partitioning in space and time. Fast computation for the most demanding tasks is facilitated by compiled C-coded functions.

2 DESCRIPTION

Here, we adopt the statistical framework of Hardy and Senterre (2007) in testing pattern deviation of diversity partitioning in

communities from null expectations. Relevant measures of diversity in SPACODiR concern species identity, evolutionary relatedness and (or) trait similarity. The framework implements a variance-decomposition approach of Rao's (1982) quadratic entropy (QE), which partitions total entropy into within- and among-sample components. The measure of diversity can be explicitly phylogenetic by considering the extent of coancestry shared by two organisms in a community ecological sample. If evolutionary relatedness among organisms within a subcommunity is quite high (i.e. all sampled organisms descend from a recent ancestor), the subcommunity would be said to exhibit little phylogenetic diversity (Cadotte *et al.*, 2010; Cavender-Bares *et al.*, 2009; Faith, 1992).

Hardy and Senterre (2007) use a spatial decomposition of pairwise divergences among organisms—sampled from within and between different subcommunities—to define a balance in diversity found within versus between samples. Diversity in SPACODiR is thus measured in two primary manners: one index considers species composition (I_{ST}) and another considers evolutionary relatedness (P_{ST}) within and among samples. These indices of diversity partitioning, I_{ST} and P_{ST} , estimate parameters ranging from 0 to 1 and respectively correspond to the average deficit of species- and phylogenetic diversity within versus among samples. I_{ST} signifies the extent to which the average likelihood of species identity for two randomly drawn individuals in a local sample exceeds that at the regional scale; P_{ST} expresses the local excess of phylogenetic relatedness relative to the mean phylogenetic relatedness of species pairs drawn from the regional sample (Hardy and Senterre, 2007; Hardy and Jost, 2008). If phylogenetic structure among species is completely lacking (i.e. as for a star phylogeny), P_{ST} , as a generalized expression of the Gini-Simpson index of diversity (Gini, 1912; Simpson, 1949), simply reduces to I_{ST} . Unlike similar implementations of QE in community phylogenetics (e.g. see Dray and Dufour, 2007; Kembel *et al.*, 2010), we explicitly treat community data as statistical samples for parameter estimation, following Hardy and Jost (2008).

A key parameter of interest in community phylogenetics and which we estimate in SPACODiR is how phylogenetic diversity is partitioned among samples. Species occurrence data may be given as presence-absence, individual abundances by species, or as relative abundances. Insofar as P_{ST} involves a measure of species diversity (the phylogenetic diversity of two samples drawn from the same species is null), we can consider extracting the component of P_{ST} explained solely by phylogeny. For this purpose, Hardy

*To whom correspondence should be addressed.

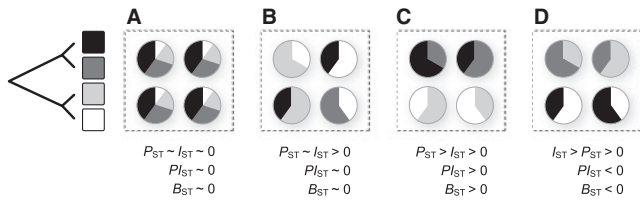


Fig. 1. Illustration of the partitioning of diversity in four situations (A–D). Shown are four distinct configurations of species assemblages (labeled A–D and enclosed by dashed boxes), where each is collection of four local communities (represented by pie-charts) that comprise up to four species (denoted by different grey tones). Evolutionary relationships of the species are depicted in the phylogenetic tree at left. (A) All local communities have the same species composition; there is no species turnover and hence no phylogenetic turnover. (B) Local communities have different species composition (species turnover) but species assemble by pairs irrespective of their phylogenetic relationship (no phylogenetic turnover). (C) Same as B except that the most related species assemble together (spatial phylogenetic clustering). (D) Same as B except that the most related species do not assemble together (spatial phylogenetic overdispersion). These situations can be distinguished using the diversity partitioning indices in SPACODiR.

and Jost (2008) constructed the index B_{ST} , which isolates the balance of evolutionary relatedness at local and regional scales from P_{ST} . The B_{ST} index, then, is closely related to $P_{ST} - I_{ST}$. Where $B_{ST} > 0$ or where $P_{ST} > I_{ST}$, distinct species within a local sample are more related on average than distinct species from different samples, a pattern referred to as spatial phylogenetic clustering (or underdispersion; e.g. Fig. 1C). Alternatively, where $B_{ST} < 0$ or $I_{ST} < P_{ST}$, distinct species within a local sample are less related on average than distinct species from different samples; phylogenetic diversity is thus spatially overdispersed (even) in distribution (e.g. Fig. 1D). A fourth index, Π_{ST} (Hardy and Senterre, 2007), is equivalent to B_{ST} but considers presence–absence rather than relative abundances of species within samples, thereby reducing the relative weight of the most abundant species. (In SPACODiR, Π_{ST} is termed ‘ PI_{ST} ’ owing to the difficulty displaying Greek letters.) Note that, as B_{ST} , Π_{ST} cannot be estimated when a community has fewer than two species because the index considers only pairs of distinct species (a typographical error in the formula given in Hardy and Senterre (2007) is corrected in the R documentation of SPACODiR). Significant departures from zero for these estimators may be used to inform processes of community assembly (Hardy and Senterre, 2007; Kembel *et al.*, 2010; Parmentier and Hardy, 2009), and several permutation tests (see Hardy 2008) are available within the software. SPACODiR provides the capability to compute each of these four indices at the global scale (i.e. considering all data) as well as for all pairwise comparisons between samples. Partial tree-randomization tests allow investigators to discern whether non-random patterns of structuring exhibit temporal- or lineage dependency (see Hardy and Senterre, 2007).

Direct assessment of spatial structuring of trait diversity may be essential to distinguishing neutral from non-neutral processes of assembly. We note that estimation of P_{ST} , B_{ST} and Π_{ST} need not only involve phylogenetic distances also but is amenable to applications involving phenotypic dissimilarities between species.

Indeed, QE can be appropriately used for all dissimilarity matrices constructed from ultrametric distances between species (Pavoine *et al.*, 2005). Structuring of phylogenetic diversity is often used as a surrogate to make inferences concerning trait-based assembly, where hypotheses for which traits are assembly relevant are absent. In essence, the phylogeny expresses the expected covariance structure among species for an idealized phenotype (see Felsenstein, 1973). Spatial clustering of phylogenetic diversity is often interpreted as evidence for local environmental filtering of (these idealized and unknown) traits (e.g. Cavender-Bares *et al.*, 2009; Ives and Helmus, 2010). The contrast between structuring of phylogenetic and phenotypic diversities should provide substantial new insights into non-neutral evolutionary and assembly processes.

Beyond constructing tests to determine whether diversity partitioning departs significantly from null expectations, SPACODiR can be readily applied to formal tests of which environmental or spatial correlates provide significant explanations of the indices of diversity partitioning. We refer readers elsewhere for examples of such tests (see Hardy and Senterre, 2007; Hardy, 2008).

Funding: F.R.S.-FNRS F.4.519.10.F to O.J.H.; BRIDGE (programme Biodiversité de l’Agence Nationale de la Recherche, France to C.E.T.P. and O.J.H.).

Conflict of Interest: none declared.

REFERENCES

- 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.
- Dray, S. and Dufour, A.B. (2007) The ade4 package: implementing the duality diagram for ecologists. *J. Stat. Softw.*, **22**, 1–20.
- Faith, D.P. (1992) Conservation evaluation and phylogenetic diversity. *Biol. Conserv.*, **61**, 1–10.
- Felsenstein, J. (1973) Maximum likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet.*, **25**, 471–492.
- Gini, C. (1912) *Variabilité mutabilité (Variability and Mutability)*. C. Cuppini, Bologna.
- Hardy, O.J. (2008) Testing the spatial phylogenetic structure of local communities: statistical performances of different null models and test statistics on a locally neutral community. *J. Ecol.*, **96**, 914–926.
- Hardy, O.J. (2010) SPACoDi 0.10: a program for Spatial & Phylogenetic Analysis of Community Diversity. Available at <http://ebe.ulb.ac.be/ebe/Software.html>.
- Hardy, O.J. and Jost, L. (2008) Interpreting and estimating measures of community phylogenetic structuring. *J. Ecol.*, **96**, 849–852.
- Hardy, O.J. and Senterre, B. (2007) Characterizing the phylogenetic structure of communities by an additive partitioning of phylogenetic diversity. *J. Ecol.*, **95**, 493–506.
- Ives, A.R. and Helmus, M.R. (2010) Phylogenetic metrics of community similarity. *Am. Nat.*, **176**, E128–E142.
- Kembel, S.W. *et al.* (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463–1464.
- Parmentier, I. and Hardy, O.J. (2009) The impact of ecological differentiation and dispersal limitation on species turnover and phylogenetic structure of inselberg plant communities. *Ecography*, **32**, 613–622.
- 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.
- Simpson, E.H. (1949) Measurement of diversity. *Nature*, **163**, 688–688.
- Webb, C.O. *et al.* (2002) Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.*, **33**, 475–505.