



A new method for quantifying the phylogenetic redundancy of biological communities

Carlo Ricotta¹ · Giovanni Bacaro² · Marco Caccianiga³ · Bruno E. L. Cerabolini⁴ · Sandrine Pavoine⁵

Received: 27 June 2017 / Accepted: 22 November 2017 / Published online: 5 December 2017
© Springer-Verlag GmbH Germany, part of Springer Nature 2017

Abstract

The increasing use of phylogenetic methods in community ecology recognizes that accumulated evolutionary differences among species mirror, to some extent, ecological processes. The scope of this work is thus to propose a new method for the measurement of community-level phylogenetic redundancy, which takes into account the branching pattern of the underlying phylogeny. Like for functional redundancy, a measure of phylogenetic redundancy can be described as a normalized measure in the range (0–1) that relates the observed community-level phylogenetic diversity to the value of a hypothetical assemblage with the same abundance distribution of the focal community in which all species had independent evolution. Therefore, phylogenetic redundancy can be interpreted as the diversity decrease that is obtained by taking into account the evolutionary relationships among species in the calculation of diversity. The behavior of the proposed method, for which we provide a simple *R* function called ‘phyloredundancy’, was evaluated with published data on Alpine plant communities along a primary succession on a glacier foreland in northern Italy. As shown by our results, the method accounts for the length of shared branches in the phylogeny, producing a coherent framework for describing the evolutionary relationships within a species assemblage. Being based on classical diversity measures, which have been used in ecology for decades, it also has a great potential for future research in phylogenetic community ecology.

Keywords Branch length · Lineage abundance · Phylogenetic diversity · Species features · Ultrametric trees

Communicated by John Dwyer.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00442-017-4026-x>) contains supplementary material, which is available to authorized users.

✉ Carlo Ricotta
carlo.ricotta@uniroma1.it

¹ Department of Environmental Biology, University of Rome ‘La Sapienza’, Piazzale A. Moro 5, 00185 Rome, Italy

² Department of Life Sciences, University of Trieste, Via L. Giorgieri 10, 34127 Trieste, Italy

³ Department of Biosciences, University of Milano, Via Celoria 26, 20133 Milano, Italy

⁴ Department of Theoretical and Applied Sciences, University of Insubria, Via J.H. Dunant 3, 21100 Varese, Italy

⁵ Centre d’Ecologie et des Sciences de la Conservation (CESCO UMR7204), Sorbonne Universités, MNHN, CNRS, UPMC, CP51, 43-61 rue Buffon, 75005 Paris, France

Introduction

Measures of biodiversity are central for exploring the complex ecological processes that drive species coexistence and community assembly. Early work on biodiversity attempted to summarize different aspects of community structure based on either presence and absence scores or on species abundances, thus assuming that all species are equally and maximally distinct. More recently, the recognition that the species ecological differences constitute an important facet of biodiversity has received increasing attention (Webb 2000; Petchey and Gaston 2002). As stressed by Cadotte et al. (2013): “Mechanisms influencing patterns of community assembly act on the ecological similarities (or redundancies) and differences (or complementarities) of organisms and not on the number of species”. In this light, the ecological differences between species are usually summarized based either on their functional or phylogenetic distances. The pairwise species differences can be then combined at the community level to summarize the average degree to which the species differ in terms

of their evolutionary history or functioning (Cadotte et al. 2013).

When several species perform similar functions, functional redundancy may enhance community stability, thus providing insurance against the loss of ecosystem functions induced by external factors, such as disturbance or biotic erosion (Yachi and Loreau 1999). Likewise, evolutionary relationships may affect ecological processes because the branching patterns of phylogenies mirrors the accumulation of multidimensional phenotypic differences among species (Felsenstein 1985; Harvey and Pagel 1991; Vane-Wright et al. 1991; Faith 1992), which may in turn describe the overall complexity of community functioning better than a limited set of functional traits (Cadotte et al. 2010; Tucker et al. 2016).

However, while several methods have been proposed for summarizing functional redundancy (Micheli and Halpern 2005; Mouillot et al. 2013, 2014; Guerrero et al. 2014; Ricotta et al. 2016), the calculation of phylogenetic redundancy has received far less attention. Like for functional redundancy, phylogenetic redundancy is defined as the diversity decrease that is obtained by taking into account the evolutionary relationships among species in the calculation of diversity. This definition makes phylogenetic redundancy clearly distinct from other notions of redundancy, such as, e.g., sample redundancy (sensu Garcillán et al. 2003).

The aim of this paper is thus to propose a new method for the measurement of community-level phylogenetic redundancy taking into account the branching pattern of the underlying phylogeny. The performance of the proposed method was evaluated with published data on Alpine plant communities along a primary succession on a glacier foreland in Italy.

Methods

Imagine an assemblage (or community, plot, sample, relevé, quadrat, etc.) composed of N species with relative abundances p_i ($i = 1, 2, \dots, N$) with $0 < p_i \leq 1$ and $\sum_i p_i = 1$, together with a symmetric $N \times N$ dissimilarity matrix whose elements d_{ij} summarize the pairwise functional dissimilarities between species i and j with d_{ij} in the range (0–1), such that $d_{ij} = d_{ji}$ and $d_{ii} = 0$.

Ricotta et al. (2016; see also de Bello et al. 2016) proposed to summarize functional redundancy R as the complement of community uniqueness U , such that $R = 1 - U = 1 - Q/S$ where Q is Rao's (1982) quadratic diversity $Q = \sum_{ij} p_i p_j d_{ij}$ (i.e., the expected dissimilarity between two individuals of the assemblage selected at random with replacement) and S is Simpson's Diversity

Index $S = 1 / \sum_i p_i^2$ (i.e., the probability that two individuals randomly selected from the assemblage will belong to different species).

Considering all species as equally and maximally distinct (i.e., for $d_{ij} = 1$ for all $i \neq j$ and $d_{ii} = 0$) we have that $Q = S$. Hence, according to Ricotta et al. (2016), community-level functional uniqueness (the complement of redundancy) can be summarized as the normalized ratio of two diversity measures with different levels of 'ecological abstraction': the Rao index, which takes the species dissimilarities into account for the calculation of functional diversity and the Simpson index, which considers all species equally and maximally dissimilar. For details, see Ricotta et al. (2016).

In principle, the same approach can be used for calculating community-level phylogenetic redundancy after rescaling the species phylogenetic distances d_{ij} to the unit range. However, a primary difference between functional and phylogenetic methods is that phylogenetic methods are fundamentally related to a tree-like structure, which represents the evolutionary relationships among species (Pavoine 2016). Therefore, suitable measures of phylogenetic redundancy should take into account the branching pattern of phylogenetic trees.

Faith (1992) suggested that branch lengths can be expressed in terms of species features, which can be interpreted as "an evolutionarily conserved trait that may be unique to a particular species or shared by many species" (Nipperess et al. 2010). As a result, the total branch length of the phylogeny is basically a measure of the total amount of evolutionary history embodied in an assemblage. In practice, the branch lengths may be set proportional to some measure of evolutionary change, such as the amount of functional or morphological differentiation from a common ancestor, or the number of base changes at a given locus (Chao et al. 2010). If the branch lengths are proportional to time of divergence, all species (terminal nodes) are at the same distance from the root node and the tree is called ultrametric.

A general approach for taking into account the tree-like structure of phylogenies in the calculation of community diversity has been recently proposed by Pavoine et al. (2009) and Chao et al. (2010). Here, we will consider only ultrametric trees reporting time of divergence among species, which usually represent the major source of phylogenetic information in community ecology. For details on how to extend this approach to non-ultrametric trees, see Appendix 1. Given a species assemblage and the associated ultrametric tree in Fig. 1, let T be the age of the root node. We can define non-overlapping evolutionary periods slicing the tree at each node so that there are no internal nodes between two slices. Starting from the root node, the node ages can be labeled t_k ($k = 0, 1, \dots, K$) so that $t_0 = T$ (the age of the root node) and $t_K = 0$ (the present-day time).

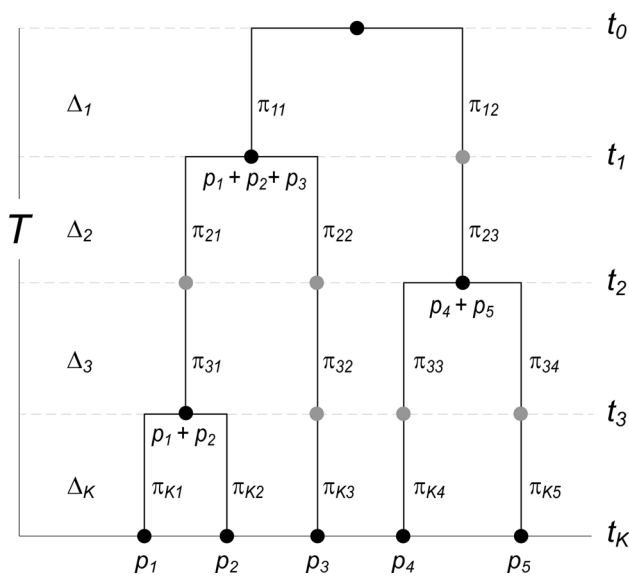


Fig. 1 Schematic ultrametric phylogeny for five species with relative abundances p_1, \dots, p_5 . Four non-overlapping evolutionary periods Δ_k can be shown by slicing the tree at each node so that there are no internal nodes between two slices. Node ages are labeled t_k ($k = 0, 1, \dots, K$) such that $t_0 = T$ (age of the root node), $t_K = 0$ (present-day), and $\Delta_k = |t_k - t_{k-1}|$. Each evolutionary period contains M lineages with relative abundances π_{km} ($m = 1, 2, \dots, M$). Lineage abundances are the sum of the relative abundances of all descendent species in the phylogenetic tree

Two subsequent slices delimit an evolutionary period $\Delta_k = |t_k - t_{k-1}|$ such that $\sum_k \Delta_k = T$. Each evolutionary period contains M lineages ($M \leq N$) with relative abundances π_{km} ($m = 1, 2, \dots, M$) such that $0 < \pi_{km} \leq 1$ and $\sum_m \pi_{km} = 1$. These lineage abundances are obtained as the sum of the relative abundances of all descendent species in the phylogenetic tree. In terms of features, the length of a lineage reflects the number of the features associated to that lineage, while the lineage abundance reflects the abundance of these features (Pavoine 2016).

Based on the lineage abundances π_{km} we can calculate a ‘lineage diversity’ D_k for each node age t_k of the phylogeny. The lineage diversity at time t_k is the traditional present-day diversity calculated from the species abundances p_i (i.e., assuming all species are equally and maximally distinct). The community-level phylogenetic diversity D_p is then obtained averaging the diversities D_k over the corresponding time intervals of interest:

$$D_p = \sum_k D_k \times \tau_k, \quad (1)$$

where $\tau_k = \Delta_k / T$ is the duration of the k -th evolutionary period after rescaling the tip-to-root length to unit, and Δ_0 (and hence τ_0) = 0 by definition. Unlike traditional diversity measures, D_p incorporates information about the branching pattern of the phylogeny and its lineage abundances. If the

diversity measure used for calculating phylogenetic diversity is species richness, the product $D_p \times T = \sum_k D_k \times \Delta_k$ turns out to be Faith’s (1992) phylogenetic diversity (PD). That is, the sum of the branch lengths of all species in the phylogeny, which is basically a measure of the total amount of evolutionary history embodied in the assemblage since the time of the most recent common ancestor. This relationship reinforces the interpretation of D_p as a measure of ‘phylogenetic diversity’.

Due to Dalton’s principle of transfers, for any Schur–concave diversity index such as species richness, the Shannon, or the Simpson diversity and their species equivalents (Hill 1973; Jost 2007), the present-day diversity D_K is always greater than or equal to any other lineage diversity D_k . For details, see Patil and Taillie (1982). Accordingly, like in Ricotta et al. (2016), we can obtain a normalized index of phylogenetic uniqueness U taking the ratio of two diversity measures: average community-level phylogenetic diversity D_p , which takes into account the branching pattern of the underlying phylogeny, and traditional present-day diversity D_K , which considers all species as equally and maximally dissimilar, such that $U = D_p / D_K$. Likewise, phylogenetic redundancy, can be simply calculated as:

$$R = 1 - U = 1 - D_p / D_K. \quad (2)$$

Phylogenetic redundancy R thus summarizes the diversity decrease that is obtained by taking into account the branching pattern of the phylogenetic tree in the calculation of diversity. Redundancy will be small when there are many deep branches, each well represented in the present-day assemblage, attaining its minimum value $R = 0$ for a star phylogeny, where all terminal branches originate from a single polytomy at the root node. By contrast, redundancy will be large when all branches emerge recently such that the present-day assemblage is composed of phylogenetically close species (Chao et al. 2010; Tucker and Cadotte 2013). Put another way, phylogenetic redundancy tells us how much the real phylogenetic diversity departs from that of a star phylogeny in which all species had independent evolution.

Case study

Starting from the assumption that, during evolution, traits are on average more conserved than random, in community ecology phylogenetic metrics have been often considered a proxy for differences in functional traits. According to this hypothesis, phylogenetically close species are generally assumed to be more functionally similar than more distantly related species (Freckleton et al. 2002; Ackerly 2009; Srivastava et al. 2012; Gerhold et al. 2015; Letten and Cornwell 2015). Therefore, the relevance of the new index of phylogenetic redundancy was evaluated comparing phylogenetic

changes in plant communities along a primary succession on the foreland of the Rutor glacier (northern Italy) with the functional changes obtained for the same data by Ricotta et al. (2016).

The study area is located between 2400 and 3000 m on moraine deposits, which are mainly composed of metamorphic schists. Climate is continental with very cold winters and low annual precipitation. The glacier terminus had its maximum expansion during the Little Ice Age (1820–1864). In the past century, the Rutor glacier had more limited oscillations, the most relevant of which ended in 1933; for additional information on the study site, see Caccianiga et al. (2006).

Fifty-nine vegetation plots of approximately 25 m² were sampled by Caccianiga et al. (2006) on the glacier foreland above the local tree limit. Based on the age of the moraine deposits, the plots were assigned to three different successional stages: 17 early-successional plots (sampled in areas exposed after 1933), 32 mid-successional plots (sampled in areas exposed between 1820 and 1933) and 10 late-successional plots (sampled in areas exposed before 1820). For each plot, species abundances were measured with a five-point ordinal scale transformed to ranks. All data are available in Ricotta et al. (2016, Appendix S2).

Based on Grime's (1974, 2001) functional classification of plants as competitors (C), stress tolerators (S) and ruderals (R), Ricotta et al. (2016; Table 2) showed that vegetation dynamics along the succession were shaped by a progressive increase in community-level functional redundancy from the early- to the late-successional stages. Here, we will analyze the phylogenetic redundancy of the same data. First, we calculated the community-level phylogenetic diversity and redundancy of each plot according to Eqs. (1) and (2), respectively. Next, we tested for significant differences among the phylogenetic diversity and redundancy values of the three successional stages using a permutational *t* test with 9999 permutations.

The phylogenetic tree was extracted from the Daphne phylogeny (Durka and Michalski 2012), a dated phylogeny of a large European flora comprising most Alpine species of Italy. For calculating phylogenetic redundancy, we used two diversity measures with different sensitivity to common versus rare lineages: the Shannon index $H = -\sum_i p_i \log p_i$ and the Simpson index S , which is more sensitive to common lineages compared to H (Jost 2007; Pavoine 2016). Note, however, that since Ricotta et al. (2016) defined functional redundancy as $1 - Q/S$, where Q is the Rao quadratic diversity, the use of the Simpson index for calculating phylogenetic diversity provides the most direct term of comparison between functional and phylogenetic redundancy. Note also that the ordinal scale used for measuring the species abundances (see Ricotta et al. 2016, Appendix S2) is disproportionately sensitive to rare species. Therefore, the Simpson

index compensates to some extent the overweighting of low species abundances induced by the ordinal measurement scale.

Results

Figure 2 shows the phylogenetic tree of the species used in this study, together with the plots of the mean Shannon and Simpson lineage diversities D_k of each successional stage vs. node age t_k . All calculations were done with the new *R* function 'phyloredundancy' available in Appendix 2; an electronic version of the phylogeny is available in Ricotta et al. (2015, Appendix A). Table 1 shows the mean community-level diversity and redundancy values for the three successional stages identified along the primary succession on the foreland of the Rutor glacier. In good agreement with the functional redundancy values calculated by Ricotta et al. (2016) and irrespective of the diversity index used, a general tendency towards an increase in phylogenetic redundancy from the early to the late-successional stages is observed.

As shown in Fig. 2, for all successional stages, the major decrease in lineage diversity from node age t_k to t_{k-1} occurs at deep evolutionary times. For the early- and mid-successional stages, which are mainly colonized by ruderal forbs, such as *Oxyria digyna* and *Tussilago farfara*, the most important change in lineage diversity along the nodes of the phylogenetic tree is associated to diversification within the eudicots at 118 Ma. By contrast, for the late-successional stage, which is dominated by stress-tolerant graminoids, such as *Carex curvula*, or *Festuca halleri*, the main change in D_k is related to differentiation within the monocots at 81 Ma. Hence, the increase in functional redundancy along the succession observed by Ricotta et al. (2016) is accompanied by a parallel increase in phylogenetic redundancy induced by a compositional shift from ruderal early-successional forbs to stress-tolerant late-successional graminoids (Caccianiga et al. 2006).

Discussion

The growing application of phylogenetic methods in ecological studies recognizes that accumulated evolutionary differences among species may reflect ecological processes. In this paper, we thus introduced a new measure of community-level phylogenetic redundancy, which is calculated by normalizing phylogenetic diversity in the range (0–1). While phylogenetic diversity is an absolute measure that takes into account the species abundances and the branching pattern of the underlying phylogeny, phylogenetic redundancy is a relative measure that relates the observed phylogenetic diversity to the value of a maximally dissimilar model assemblage

Fig. 2 **a** Ultrametric phylogeny of the species used in this study, together with the plots of the mean **b** Shannon and **c** Simpson lineage diversities D_k of each successional stage vs. node age t_k . The root node ($T = 147$ million years Ma) and the nodes corresponding to the major changes in lineage diversity along the phylogenetic tree are shown with a black dot. For details, see the main text. Color version of this figure is available online

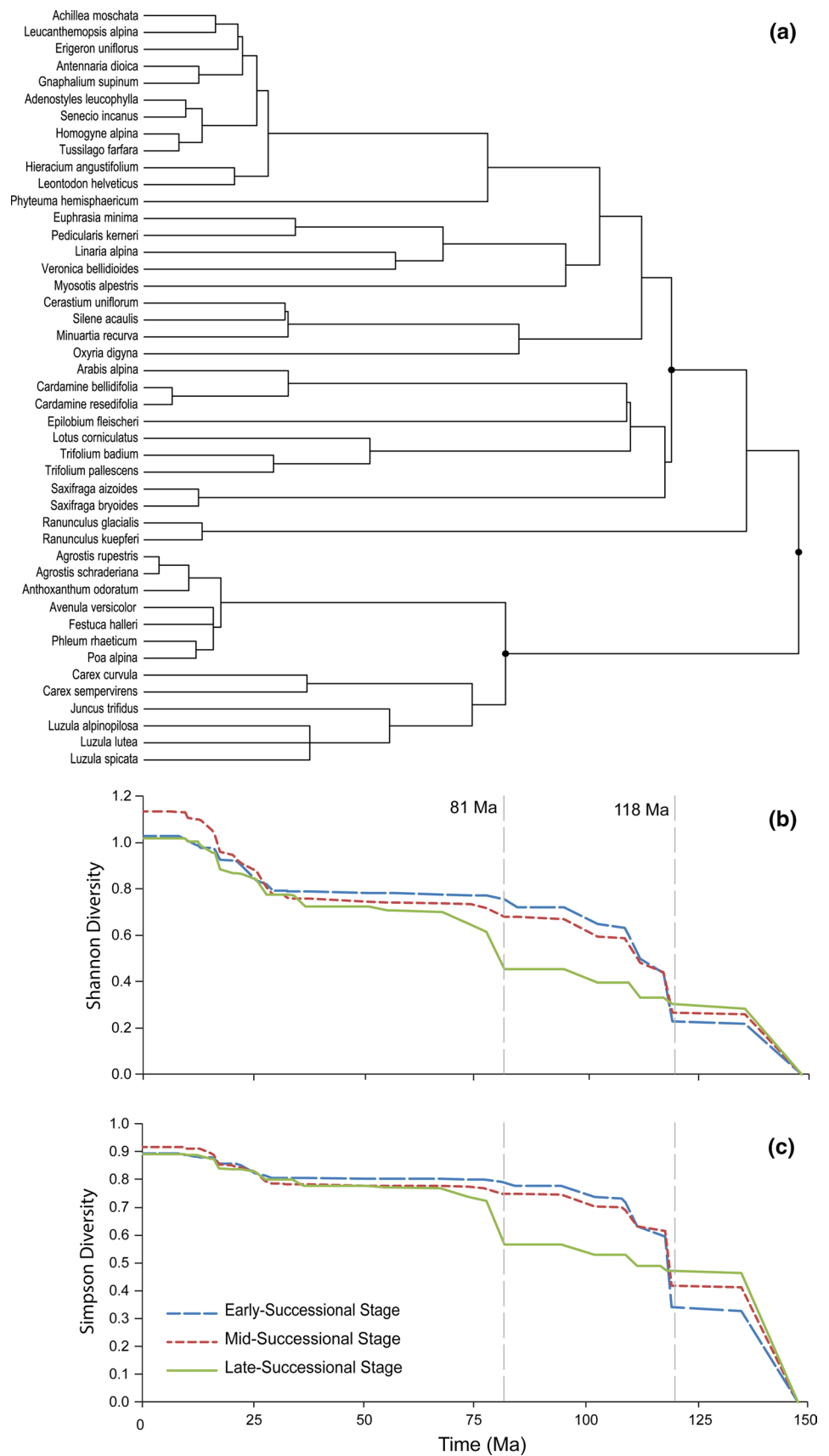


Table 1 Mean (standard deviation) phylogenetic diversity and redundancy values for the three successional stages identified on the foreland of the Rutor glacier calculated both with the Shannon and the Simpson diversity coefficients

	Early-successional stage (17 plots)	Mid-successional stage (32 plots)	Late-successional stage (10 plots)
Shannon diversity			
Phylogenetic diversity D_p	0.675 (0.064)a	0.673 (0.060)a	0.600 (0.069)
Present-day diversity D_K	1.030 (0.154)a	1.134 (0.094)	1.018 (0.136)a
Phylogenetic redundancy R	0.336 (0.086)	0.405 (0.047)a	0.408 (0.050)a
Simpson diversity			
Phylogenetic diversity D_p	0.706 (0.035)a	0.710 (0.025)a	0.671 (0.039)
Present-day diversity D_K	0.893 (0.039)a	0.918 (0.018)	0.891 (0.035)a
Phylogenetic redundancy R	0.208 (0.040)a	0.226 (0.025)ab	0.247 (0.035)b

Pairwise comparisons of index differences between the successional stages were done with one-way permutational t tests (9999 permutations)

For each index, numbers followed by the same letter do not differ significantly ($P > 0.05$; Holm-corrected values)

with the same abundance distribution of the focal community. In this sense, phylogenetic redundancy differs from the classical notion of ‘phylogenetic evenness’ sensu Villéger et al. (2008) because the latter has been used to quantify the regularity of the distribution of species dissimilarities and abundances in phylogenetic space (see also Tucker et al. 2016).

Unlike other indices of phylogenetic diversity, which are based on pairwise species distances, the proposed measures account for the length of shared branches in the phylogeny. On one hand, this is a more natural approach to the calculation of phylogenetic diversity, since a tree representation is the most logical way for describing the evolutionary relationships within a species assemblage. On the other hand, this method may become problematic for comparing phylogenetic and functional diversity, since a tree-based hierarchical representation of the functional relationships within a set of species (see Petchey and Gaston 2002) is not biologically justified (Pavoine 2016; Villéger et al. 2017).

The proposed method for calculating phylogenetic diversity and redundancy has a great potential for future research in phylogenetic community ecology as it is based on classical diversity measures, which have been used in ecology for decades. Therefore, due to their well-known properties, the practitioner can select the measure that best matches the requirements of his/her specific question. For example, at broad spatial scales, species abundances are often unknown or irrelevant, such that species occurrences are usually represented as presence/absence scores. By contrast, at local scales, the ecological interactions among species are thought to reflect both their abundances and evolutionary relationships. In this case, data on species abundances become a relevant piece of ecological information.

In this framework, unlike most tree-based measures of phylogenetic diversity (but see Cadotte et al. 2010; Tucker et al. 2016), the proposed method enables the incorporation of information on species abundances in a very flexible

manner. At the same time, since D_p is an additive quantity, it is possible to decompose phylogenetic diversity into the contributions of the single evolutionary periods which shape the phylogeny. For example, in our case study, the increased evolutionary relatedness of the late-successional species is most likely due to the increasing dominance of graminoids which progressively substituted early-successional ruderal forbs. On the other hand, one might assume that the increasing phylogenetic redundancy along the successional process is simply a ‘sampling effect’ due to larger species richness in late-successional stages. However, for the alpine vegetation of the Rutor glacier, the largest species richness is associated to the mid-successional plots (see Ricotta et al. 2016, Table 2), while the index used for calculating phylogenetic redundancy is generally independent of species richness (for a detailed discussion on the sensitivity of R to species richness, see Appendix 3). Therefore, at least for our case study, phylogenetic redundancy is not directly related to species richness.

From a more technical perspective, since the calculation of redundancy is done after normalization of the tip-to-root length T of the phylogeny to the unit interval (see Eq. 1), we recommend adapting the way the data are normalized to the question at hand. For example, if our aim consists of comparing the redundancy of sites within a region, the normalization of T in the range (0–1) should be done at the regional level. In this way, all sites are compared using the same phylogeny, but different vectors of species abundances. On the other hand, if normalization is done separately on each local phylogeny, the resulting redundancy values are not expressed in the same units, thus hampering the comparison among datasets. From a biological viewpoint, if we imagine that branch lengths are related to the number of features shared by two species (Faith 1992; Pavoine et al. 2016), the way of normalizing the height of the phylogeny can be interpreted as normalization by the theoretical maximum number of features shared by two species in our assemblage.

This raises the more general question of developing methods for scaling branch lengths to increase the ecological significance of phylogenies (Clarke and Warwick 1999; Blomberg et al. 2003; Cadotte et al. 2013; Letten and Cornwell 2015).

To conclude, in the last decades, ecologists have recognized that ecological similarities between species are driven to some extent by patterns of common ancestry (Cadotte et al. 2010). As a consequence, the concept of biodiversity has increasingly become a complex multifaceted notion that incorporates functional and phylogenetic differences. Therefore, to be really useful in ecological research, the properties of biodiversity indices must be unambiguously understandable by ecologists. We believe that the proposed method provides a relatively simple way for summarizing a key metric of community structure, such as phylogenetic redundancy, thus offering new insights into the mechanisms that drive community composition through a variety of spatial scales.

Acknowledgements We acknowledge financial support from the University of Rome ‘La Sapienza’ (Grant no. C26A15CCWF).

Author contribution statement CR formulated the idea. CR, GB and SP developed the methodology. GB and SP wrote the R code. CR, MC and BC analyzed the data. CR wrote the manuscript; other authors provided editorial advice.

References

- Ackerly D (2009) Conservatism and diversification of plant functional traits: evolutionary rates versus phylogenetic signal. *Proc Natl Acad Sci USA* 106:19699–19706. <https://doi.org/10.1073/pnas.0901635106>
- Blomberg SP, Garland T, Ives AR (2003) Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57:717–745
- Caccianiga M, Luzzaro A, Pierce S, Ceriani RM, Cerabolini B (2006) The functional basis of a primary succession resolved by CSR classification. *Oikos* 112:10–20. <https://doi.org/10.1111/j.0030-1299.2006.14107.x>
- Cadotte MW, Davies TJ, Regetz J, Kembel SW, Cleland E, Oakley TH (2010) Phylogenetic diversity metrics for ecological communities: integrating species richness, abundance and evolutionary history. *Ecol Lett* 13:96–105. <https://doi.org/10.1111/j.1461-0248.2009.01405.x>
- Cadotte MW, Albert CH, Walker SC (2013) The ecology of differences: assessing community assembly with trait and evolutionary distances. *Ecol Lett* 16:1234–1244. <https://doi.org/10.1111/ele.12161>
- Chao A, Chiu CH, Jost L (2010) Phylogenetic diversity measures based on Hill numbers. *Philos Trans R Soc B* 365:3599–3609. <https://doi.org/10.1098/rstb.2010.0272>
- Clarke K, Warwick R (1999) The taxonomic distinctness measure of biodiversity: weighting of step lengths between hierarchical levels. *Mar Ecol Prog Ser* 184:21–29
- de Bello F, Carmona CP, Lepš J, Szava-Kovats R, Pärtel M (2016) Functional diversity through the mean trait dissimilarity: resolving shortcomings with existing paradigms and algorithms. *Oecologia* 180:933–940. <https://doi.org/10.1007/s00442-016-3546-0>
- Durka W, Michalski SG (2012) Daphne: a dated phylogeny of a large European flora for phylogenetically informed ecological analyses. *Ecology* 93:2297. <https://doi.org/10.1890/12-0743.1>
- Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61:1–10
- Felsenstein J (1985) Phylogenies and the comparative method. *Am Nat* 125:1–15
- Freckleton RP, Harvey PH, Pagel M (2002) Phylogenetic analysis and comparative data: a test and review of evidence. *Am Nat* 160:712–726. <https://doi.org/10.1086/343873>
- Garcillán PP, Ezcurra E, Riemann H (2003) Distribution and species richness of woody dryland legumes in Baja California, Mexico. *J Veg Sci* 14:475–486. <https://doi.org/10.1111/j.1654-1103.2003.tb02174.x>
- Gerhold P, Cahill JF, Winter M, Bartish I, Prinzing A (2015) Phylogenetic patterns are not proxies of community assembly mechanisms (they are far better). *Funct Ecol* 29:600–614. <https://doi.org/10.1111/1365-2435.12425>
- Grime JP (1974) Vegetation classification by reference to strategies. *Nature* 250:26–31
- Grime JP (2001) Plant strategies vegetation processes and ecosystem properties. Wiley, Chichester
- Guerrero I, Carmona CP, Morales MB, Oñate JJ, Peco B (2014) Non-linear responses of functional diversity and redundancy to agricultural intensification at the field scale in Mediterranean arable plant communities. *Agric Ecosyst Environ* 195:36–43. <https://doi.org/10.1016/j.agee.2014.05.021>
- Harvey PH, Pagel M (1991) The comparative method in evolutionary biology. Oxford University Press, Oxford
- Hill MO (1973) Diversity and evenness: a unifying notation and its consequences. *Ecology* 54:427–431
- Jost L (2007) Partitioning diversity into independent alpha and beta components. *Ecology* 88:2427–2439. <https://doi.org/10.1890/06-1736.1>
- Letten AD, Cornwell WK (2015) Trees, branches and (square) roots: why evolutionary relatedness is not linearly related to functional distance. *Methods Ecol Evol* 6:439–444. <https://doi.org/10.1111/2041-210X.12237>
- Micheli F, Halpern BS (2005) Low functional redundancy in coastal marine assemblages. *Ecol Lett* 8:391–400. <https://doi.org/10.1111/j.1461-0248.2005.00731.x>
- Mouillot D, Bellwood DR, Baraloto C, Chave J, Galzin R, Harmelin-Vivien M, Kulbicki M, Lavergne S, Lavorel S, Mouquet N, Paine CET, Renaud J, Thuiller W (2013) Rare species support vulnerable functions in high-diversity ecosystems. *PLoS Biol* 11:e1001569. <https://doi.org/10.1371/journal.pbio.1001569>
- Mouillot D, Villéger S, Parravicini V, Kulbicki M, Arias-Gonzalez JE, Bender M, Chabanet P, Floeter SR, Friedlander A, Vigliola L, Bellwood DR (2014) Functional over-redundancy and high functional vulnerability in global fish faunas on tropical reefs. *Proc Natl Acad Sci USA* 111:13757–13762. <https://doi.org/10.1073/pnas.1317625111>
- Nipperess DA, Faith DP, Barton K (2010) Resemblance in phylogenetic diversity among ecological assemblages. *J Veg Sci* 21:809–820. <https://doi.org/10.1111/j.1654-1103.2010.01192.x>
- Patil GP, Taillie C (1982) Diversity as a concept and its measurement. *J Am Stat Assoc* 77:548–561
- Pavoine S (2016) A guide through a family of phylogenetic dissimilarity measures among sites. *Oikos* 125:1719–1732. <https://doi.org/10.1111/oik.03262>
- Pavoine S, Love MS, Bonsall MB (2009) Hierarchical partitioning of evolutionary and ecological patterns in the organization of phylogenetically-structured species assemblages: application to rockfish (genus: *Sebastes*) in the Southern California Bight. *Ecol Lett* 12:898–908. <https://doi.org/10.1111/j.1461-0248.2009.01344.x>

- Pavoine S, Marcon E, Ricotta C (2016) ‘Equivalent numbers’ for species, phylogenetic or functional diversity in a nested hierarchy of multiple scales. *Methods Ecol Evol* 7:1152–1163. <https://doi.org/10.1111/2041-210X.12591>
- Petchey OL, Gaston KJ (2002) Functional diversity (FD), species richness and community composition. *Ecol Lett* 5:402–411. <https://doi.org/10.1046/j.1461-0248.2002.00339.x>
- Rao RC (1982) Diversity and dissimilarity coefficients: a unified approach. *Theor Popul Biol* 21:24–43
- Ricotta C, Bacaro G, Caccianiga M, Cerabolini BEL, Moretti M (2015) A classical measure of phylogenetic dissimilarity and its relationship with beta diversity. *Basic Appl Ecol* 16:10–18. <https://doi.org/10.1016/j.baae.2014.10.003>
- Ricotta C, de Bello F, Moretti M, Caccianiga M, Cerabolini B, Pavoine S (2016) Measuring the functional redundancy of biological communities: a quantitative guide. *Methods Ecol Evol* 7:1386–1395. <https://doi.org/10.1111/2041-210X.12604>
- Srivastava DS, Cadotte MW, MacDonald AAM, Marushia RG, Mirotchnick N (2012) Phylogenetic diversity and the functioning of ecosystems. *Ecol Lett* 15:637–648. <https://doi.org/10.1111/j.1461-0248.2012.01795.x>
- Tucker CM, Cadotte MW (2013) Unifying measures of biodiversity: understanding when richness and phylogenetic diversity should be congruent. *Divers Distrib* 19:845–854. <https://doi.org/10.1111/ddi.12087>
- Tucker CM, Cadotte MW, Carvalho SB, Davies TJ, Ferrier S, Fritz SA, Grenyer R, Helmus MR, Jin LS, Mooers AO, Pavoine S, Purschke O, Redding DW, Rosauer DF, Winter M, Mazel F (2016) A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biol Rev* 92:698–715. <https://doi.org/10.1111/brv.12252>
- Vane-Wright RI, Humphries CJ, Williams PH (1991) What to protect?—Systematics and the agony of choice. *Biol Conserv* 55:235–254
- Villéger S, Mason NW, Mouillot D (2008) New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology* 89:2290–2301. <https://doi.org/10.1890/07-1206.1>
- Villéger S, Maire E, Leprieux F (2017) On the risks of using dendrograms to measure functional diversity and multidimensional spaces to measure phylogenetic diversity: a comment on Sobral et al. (2016). *Ecol Lett* 20:554–557. <https://doi.org/10.1111/ele.12750>
- Webb CO (2000) Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. *Am Nat* 156:145–155. <https://doi.org/10.1086/303378>
- Yachi S, Loreau M (1999) Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. *Proc Natl Acad Sci USA* 96:1463–1468

Appendix 1. How to extend the calculation of phylogenetic redundancy to non-ultrametric trees.

Given an assemblage composed of N species, we will restrict our analysis to dichotomic diversity measures D expressed as the average rarity $R(i)$ of the species relative abundances $p_i = (p_1, p_2, \dots, p_N)$, such as the Shannon or the Simpson diversity:

$$D = \sum_i p_i \times R(p_i) \quad (1)$$

where $R(p_i)$ is a decreasing function of the relative abundances p_i (see Patil and Taillie 1982). Note that according to Eq. (2), the contribution of species i to overall diversity depends only on the numerical value of p_i . Patil and Taillie (1982) called this property ‘dichotomy’ because the contribution of i to D would be unchanged if the other species were grouped into a single complementary category.

Also, according to Patil and Taillie (1982), a diversity measure is said ‘schur-concave’ if D increases when abundance is transferred from a species to a another strictly less abundant species. This property is known in econometrics as Dalton’s (1920) principle of transfers. In its very essence, given two species i and j with relative abundances $p_i > p_j \geq 0$, diversity is increased if the quantity δ is transferred from p_i to p_j so long as the transfer does not reverse the ranking of the two abundances: $p_i - \delta > p_j + \delta$.

Chao et al. (2010) showed that for an ultrametric tree, the phylogenetic diversity D_P in Eq. (1) of the main text can be also calculated as follows. Consider a rooted ultrametric phylogeny with B branches and N species as tips with relative abundances p_i . Let L_j be the length (i.e. duration) of branch j ($j = 1, 2, \dots, B$), N_j be the set of species descending from branch j and p_{ij} be the relative abundance of the i -th descendant of branch j so that $a_j = \sum_{i \in N_j} p_{ij}$ is the relative abundance of all species descending from branch j . Then, phylogenetic diversity D_P can be obtained by calculating a diversity index of choice of the general form:

$$D_P = \sum_j \frac{L_j}{T} a_j \times R(a_j) \quad (2)$$

where T is the age of the root node. The so-obtained phylogenetic diversity D_P is identical to the phylogenetic diversity obtained with the procedure described in the main text, being at the same time formally similar to the weighted diversity measures proposed by Guiasu (1971) and Guiasu and Guiasu (2012). See also Chiu et al. (2014) and Pavoine (2009, 2016).

Phylogenetic redundancy can be then calculated as:

$$R = 1 - D_P / D \quad (3)$$

where D is the present-day diversity calculated from the N species relative abundances p_i . A necessary requirement to obtain a meaningful measure of phylogenetic redundancy in the range 0–1, is that $D_P \leq D$. That is, phylogenetic diversity D_P cannot be greater than the present-day diversity D , which is calculated assuming that the N species are maximally distinct from each other. From an evolutionary point of view, this is tantamount saying that D , is calculated from a star phylogeny in which all species had independent evolution.

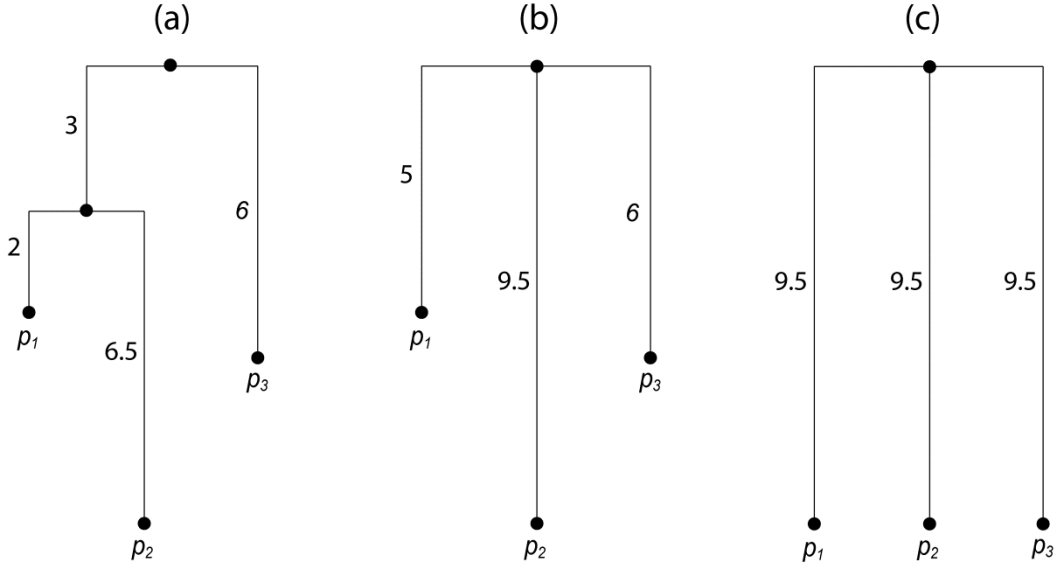


Figure 1. (a) Hypothetical non-ultrametric phylogeny with three species with relative abundances p_1 , p_2 and p_3 and unequal distances t_i from tips to root. (b) Non-ultrametric phylogeny with the same set of species of (a) in which all species evolved independently from the root node while preserving the unequal distances t_i from tips to root. (c) Ultrametric phylogeny with the same set of species of (a) in which all species assumed the maximum distance t_{max} from tips to root.

More generally, for any tree, be it ultrametric or not, D_P can be calculated as (Chao et al. 2010):

$$D_P = \sum_j \frac{L_j}{\bar{T}} a_j \times R(a_j) \quad (4)$$

where $\bar{T} = \sum_j L_j a_j = \sum_j L_j \sum_{i \in N_j} p_{ij} = \sum_i p_i t_i$ and t_i is the distance from tip i to root on the phylogenetic tree.

We can compare the phylogenetic diversity D_P of any non ultrametric tree, such as that in Figure 1a, to the diversity assumed by the same set of species if all species evolved independently from the root node while preserving the unequal distances t_i from tips to root (Figure 1b):

$$D_{Star} = \frac{1}{\bar{T}} \sum_j L_j \sum_{i \in N_j} p_{ij} R(p_{ij}) \quad (5)$$

Equivalently, D_{Star} can be written as follows:

$$D_{Star} = \frac{1}{\bar{T}} \sum_i t_i p_i R(p_i)$$

Because the index of diversity $D = \sum_i p_i \times R(p_i)$ is schur-concave, and because it is dichotomic *sensu* Patil and Taillie (1982), then for any j we have:

$$\left(\sum_{i \in N_j} p_{ij} \right) \times R \left(\sum_{i \in N_j} p_{ij} \right) \leq \sum_{i \in N_j} p_{ij} R(p_{ij})$$

This yields $D_P \geq D_{Star}$ (6)

With ultrametric trees, the distances t_i from tips to root are constant ($t_i = T$) and hence $D_{Star} = \sum_i p_i R(p_i)$. D_{Star} is thus species diversity (D).

With non-ultrametric trees, however, D_{Star} still contains phylogenetic information: the distances from tip to root (t_i) and their mean \bar{T} . As an alternative, to compare phylogenetic diversity with species diversity, we can consider the following formulas:

$$D_{\Pi} = \frac{1}{t_{\max}} \sum_j L_j a_j \times R(a_j) \quad (7)$$

for phylogenetic diversity, where t_{\max} is the maximum of t_i over i , and

$$D_{\Sigma} = \frac{1}{t_{\max}} \left[\sum_j \sum_{i \in N_j} L_j p_{ij} R(p_{ij}) + \sum_i (t_{\max} - t_i) p_i R(p_i) \right] \quad (8)$$

for species diversity.

D_{Σ} corresponds to the star phylogeny with equal, maximum distances from tips to roots (Figure 1c). Because the index of diversity $D = \sum_i p_i \times R(p_i)$ is schur-concave, and because it is dichotomic *sensu* Patil and Taillie (1982)

$$\sum_j L_j a_j \times R(a_j) \leq \sum_j \sum_{i \in N_j} L_j p_{ij} R(p_{ij})$$

This yields

$$D_{\Pi} \leq \frac{1}{t_{\max}} \left[\sum_j \sum_{i \in N_j} L_j p_{ij} R(p_{ij}) \right] \leq D_{\Sigma} \quad (9)$$

Note that, if the phylogeny is ultrametric, $D_{Star} = D_{\Sigma}$ and $D_P = D_{\Pi}$.

References

- Chao A, Chiu CH, Jost L (2010) Phylogenetic diversity measures based on Hill numbers. *Phil Trans R Soc B* 365:3599–3609
- Chiu CH, Jost L, Chao A. (2014) Phylogenetic beta diversity, similarity and differentiation measures based on Hill numbers. *Ecol Monogr* 84: 21–44
- Dalton H (1920) Measurement of the inequality of incomes. *Econ J* 30: 348–361
- Patil GP, Taillie C (1982) Diversity as a concept and its measurement. *J Am Stat Ass* 77: 548–561
- Guiasu RC, Guiasu S (2012) The weighted Gini-Simpson index: revitalizing an old index of biodiversity. *Int J Ecol*. Article ID: 478728. Doi: 10.1155/2012/478728
- Guiasu S (1971) Weighted entropy. *Rep Mat Phys* 2: 165–179
- Pavoine S (2016) A guide through a family of phylogenetic dissimilarity measures among sites. *Oikos* 125: 1719–1732
- Pavoine, S., Love, M.S., Bonsall, M.B. (2009) Hierarchical partitioning of evolutionary and ecological patterns in the organization of phylogenetically-structured species assemblages: applications to rockfish (genus: *Sebastes*) in the Southern California Bight. *Ecology Letters* 12: 898–908.

Appendix 2. R function 'phyloredundancy'

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License <http://www.gnu.org/licenses/>.

Disclaimer: users of this code are cautioned that, while due care has been taken and it is believed accurate, it has not been rigorously tested and its use and results are solely the responsibilities of the user.

Description: the function 'phyloredundancy' calculates community-level phylogenetic redundancy taking into account the branching pattern of the underlying phylogeny.

Dependencies: "vegan" and "ape" libraries

Usage: phyloredundancy(community, tree, index=c("richness","shannon","simpson"))

Arguments

community: a community data frame with plots as rows, species as columns, and species abundance as entries.

tree: an object of class 'phylo'. Only ultrametric trees are allowed. Phyloredundancy checks if the phylogenetic tree is ultrametric with the function "is.ultrametric" of the "ape" library. If the tree is not ultrametric, a warning message is displayed.

index: Phyloredundancy computes phylogenetic species richness ("richness") or the phylogenetic analogue of the Shannon or Simpson diversity ("shannon" or "simpson", respectively); "richness" is assumed as the default value.

Outputs:

An object of class "data.frame" is returned containing the following statistics:

D_K : the present-day diversity of all plots in the data frame.

D_P : the phylogenetic diversity D_P of all plots in the data frame obtained averaging the diversities D_k over the corresponding k -th evolutionary periods after rescaling the tip-to-root length to unit (for details, see the main text).

U : phylogenetic uniqueness calculated as the ratio between phylogenetic diversity and present-day diversity D_P / D_K .

R : phylogenetic redundancy, calculated as $1-U$.

Function Syntax

```
phyloredundancy<-function(community, tree, index= "richness")
{
  require(vegan)
  require(ape)
  if (!inherits(community, "data.frame")){
    if(!inherits(community, "matrix"))
      stop("Object community must be of class data.frame or matrix")
  }
  if (any(community < 0))
    stop("Negative value in community")
  if (any(rowSums(community) < 1e-16))
    stop("Empty plots in community")
  if (!inherits(tree, "phylo"))
    stop("tree must be of class phylo")
  if (ncol(community)!=length(tree$tip.label))
    stop("Species should be the same and ordered in the same way in the 'community' and 'tree' objects")
}
```

```

else if (is.null(colnames(community)))
  warning("Species should ordered in the same way in the 'community' and 'tree' objects")else
  community <- community[, tree$tip.label]
if (is.ultrametric(tree, tol = 0.001)== FALSE)
  warning("the phylogenetic tree is not ultrametric")
if(!index[1]%in%c("shannon", "simpson", "richness"))
stop("index must be shannon or simpson or richness")
dis_phylo<-cophenetic.phylo(tree)
clusters<-hclust(as.dist(dis_phylo))
groups<-cutree(clusters, h=clusters$height)
colnames(groups) <- paste("g", 1:ncol(groups), sep="")
groups_matrix<-data.frame(groups)
norm_dist<-c(0,clusters$height/max(clusters$height))
delta<-diff(norm_dist)
all_delta<-c(delta,0)
if (index %in% c("simpson", "shannon")){
  index_base <- apply(community, 1, function(v) diversity(v, index=index))
funmerge <- function(w)
  {
    merged<- aggregate(t(community) ~ w, FUN=sum)
    cleaned <- t(merged[,-1])
    return(apply(cleaned, 1, function(v) diversity(v, index=index)))
  }
index_trees <- apply(groups_matrix, 2, funmerge)
  all_index_plot <- cbind.data.frame(index_base,index_trees)
}
if (index == "richness"){
  index_base <- apply(community, 1, function(v) specnumber(v))
  funmerge <- function(w)
  {
    merged<- aggregate(t(community) ~ w, FUN=sum)
    cleaned <- t(merged[,-1])
    return(apply(cleaned, 1, function(v) specnumber(v)))
  }
  index_trees <- apply(groups_matrix, 2, funmerge)
  all_index_plot<-data.frame(index_base,index_trees)
}
all_branch_index<- t(t(all_index_plot) * all_delta)
mean_index_tree<-rowSums(all_branch_index)
phylo_uniq<-mean_index_tree/ index_base
phylo_redundancy<-1-phylo_uniq
outputs<-data.frame(index_base,mean_index_tree,phylo_uniq,phylo_redundancy)
colnames(outputs)<-c("Dk", "Dp", "Phylogenetic Uniqueness", "Phylogenetic Redundancy")
return(outputs)
}

```

Example with simulated data - requires the 'ape' and the 'TreeSim' libraries

```

library(ape)
library(TreeSim)
phy <- sim.bd.taxa(n=10, 1, 1, 0)[[1]] #generates an ultrametric phylogenetic tree with 10 tips (assumed as
species) under a pure birth model
com_matrix<-t(data.frame(replicate(20, rpois(10, 1)))) #generates a community matrix with ten species and
20 plots. Strictly positive abundance values are randomly assigned.
colnames(com_matrix)<-phy$tip.label
phyloredundancy(com_matrix, phy, index="richness") #calculates the phylogenetic redundancy of the
simulated dataset.

```


Appendix 3. On the sensitivity of phylogenetic redundancy to species richness.

Imagine an assemblage composed of N species with species relative abundances $p_i = (p_1, p_2, \dots, p_N)$, and let R be a measure of community-level phylogenetic redundancy $R = 1 - U = 1 - D_P/D_K$, where U is phylogenetic uniqueness, D_P is phylogenetic diversity, and D_K is traditional present-day diversity defined as in the main text.

While all measures of present-day diversity are by definition sensitive to species richness, in most cases the ratio D_P/D_K removes this effect. Irrespective of the index of present-day diversity D_K and of the number of species N , redundancy attains its minimum value for a star phylogeny for which $D_P = D_K$ and hence $R = 0$. However, in some cases, the maximum value of redundancy may depend on species richness.

If present-day diversity is measured as species richness, such that $D_K = N$, phylogenetic diversity D_P becomes Faith's (1992) index after normalization of the tip-to-root length of the phylogeny to the unit interval $[0, 1]$. It follows that the average phylogenetic diversity per species D_P/D_K can range from $1/N$ (for the extreme case where a long branch of length $1 - \varepsilon$ with $\varepsilon \ll 1$ leads to a polytomy of N branches of length ε), to 1 (for a star phylogeny; see Figure 1).

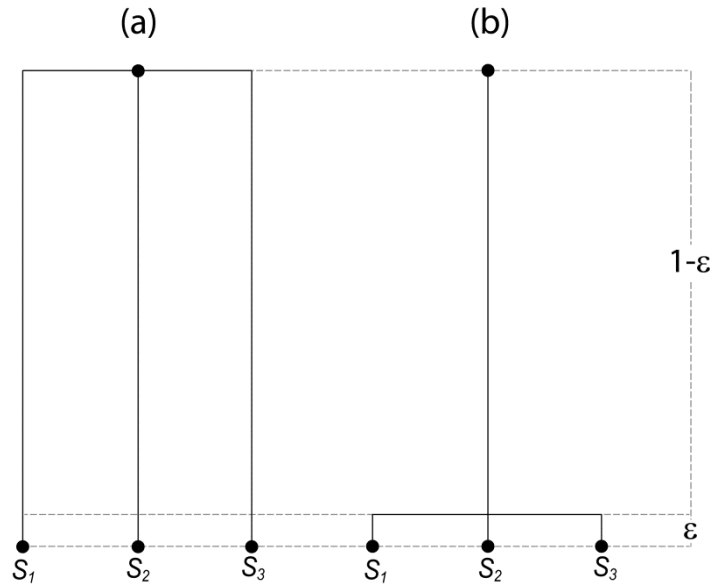


Figure 1. (a) Hypothetical star phylogeny with three species S_1 , S_2 and S_3 . (b) Phylogeny with the same species richness of (a) in which a long branch of length $1 - \varepsilon$ (with $\varepsilon \ll 1$) leads to a polytomy of three branches of length ε . When species abundances are considered for the calculation of D_P and D_K , the phylogenetic redundancy of (b) tends to 0 for ε tending to 0.

Therefore, for $D_K = N$, $0 \leq R \leq (N - 1)/N$. In this case, to make phylogenetic redundancy independent of species richness, we may rescale R to the unit interval. Alternatively, when dealing with species presence/absence data, instead of species richness N , we can use the index $N - 1$ or $\log N$ for summarizing D_P and D_K (see Hill 1973 and Patil and Taillie 1982). With index $N - 1$, the extreme scenario where a long branch of length $1 - \varepsilon$ leads to a polytomy of N branches of length ε

will lead to $D_p = (1 - \varepsilon) \times 0 + \varepsilon \times (N - 1)$ which tends to 0 when ε tends to 0. With $\log N$, it will lead to $D_p = (1 - \varepsilon) \times \log(1) + \varepsilon \times \log(N)$ which also tends to 0 when ε tends to 0.

To the contrary, when species abundances are considered, like for the Shannon or the Simpson index, redundancy R becomes independent of species richness. For a given value of D_K , the extreme scenario where a long branch of length $1 - \varepsilon$ leads to a polytomy of N branches of length ε (with $\varepsilon \ll 1$) will lead to $D_p = D_K \times \varepsilon + (1 - \varepsilon) \times 0$ and thus to $D_p / D_K = 0$ and $R = 1$. Indeed with the Shannon or the Simpson index the diversity associated to the branch of length $1 - \varepsilon$ is zero. In that case, when ε tends to 0, the range of variation for uniqueness and redundancy is $[0, 1]$.

References

- Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61: 1–10
Hill MO (1973) Diversity and evenness: a unifying notation and its consequences. *Ecology* 54: 427–431
Patil GP, Taillie C (1982) Diversity as a concept and its measurement. *J Am Stat Ass* 77: 548–561