

ECOGRAPHY

Software note

'phyloraster': an R package to calculate measures of endemism and evolutionary diversity for rasters

Gabriela Alves-Ferreira^{1,2,3}, Flávio Mariano Machado Mota^{1,3}, Daniela Custódio Talora^{1,3},
Cynthia Valéria Oliveira^{1,3}, Mirco Solé^{1,2,4} and Neander Marcel Heming^{1,3}

¹Programa de Pós-Graduação em Ecologia e Conservação da Biodiversidade, Universidade Estadual de Santa Cruz, Ilhéus, BA, Brazil

²Tropical Herpetology Lab, Universidade Estadual de Santa Cruz, Ilhéus, BA, Brazil

³Laboratório de Ecologia Aplicada à Conservação, Universidade Estadual de Santa Cruz, Ilhéus, BA, Brazil

⁴Zoologisches Forschungsmuseum Alexander Koenig, Herpetology Section, Bonn, North Rhine-Westphalia, Germany

Correspondence: Gabriela Alves-Ferreira (gabriela-alves77@hotmail.com)

Ecography

2024: e06902

doi: [10.1111/ecog.06902](https://doi.org/10.1111/ecog.06902)

Subject Editor: Brody Sandel

Editor-in-Chief: Miguel Araújo

Accepted 14 November 2023



The spatial exploration of richness, endemism, and evolutionary diversity patterns has become an important part of biogeographic research and conservation planning. As the volume and complexity of biogeographical and phylogenetic data increase, the need for efficient tools to manipulate and analyze these datasets becomes essential. The 'phyloraster' package addresses this need by facilitating the analysis of evolutionary diversity and endemism for rasters. Our package offers a set of functions to support the linkage of species distribution models (SDMs) with phylogenies, providing then an understanding of the spatial distribution of biodiversity. It covers three main stages: pre-processing, processing, and post-processing of macroecological and phylogenetic data. During the pre-processing step, basic functions are provided to prepare the data. The processing step combines functions to calculate indices including species richness, Faith's phylogenetic diversity, phylogenetic endemism, weighted endemism, and evolutionary distinctiveness. Additionally, this step includes functions to compute the standardized effect size for each metric using spatial and phylogenetic randomization methods, ensuring proper control for richness effects. The post-processing stage includes functions to calculate the change of metrics between different times (e.g. present and future). In relation to processing in our functions, we show that 'phyloraster' takes up considerably less RAM than the other packages when computing the same metrics (weighted endemism). Lower RAM usage minimizes the hardware requirements to work with high-resolution datasets from local to global scales. This broadens user accessibility of the spatialized measures of endemism and evolutionary diversity.

Keywords: GIS, macroecology, spatial ecology, spatial patterns



www.ecography.org

© 2024 The Authors. Ecography published by John Wiley & Sons Ltd on behalf of Nordic Society Oikos

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Background

Species diversity is not constant over time and space (Weber et al. 2014), which has intrigued scientists since the time of Darwin and Wallace. Mapping endemism patterns is a crucial approach for characterizing the distribution of biodiversity (Rosauer et al. 2009). It serves as an important part of biogeographic research to identify regions with taxa that should be prioritized for global conservation efforts (Rosauer et al. 2009). The absolute concept of endemism states that a taxon is classified as endemic when it is restricted to a particular region and does not occur anywhere else (Anderson 1994). Weighted endemism (WE; Williams et al. 1994) is a widely employed measure for evaluating centers of endemism, which weight the species range by the proportion of the range of each species present in a given region (Laffan et al. 2016). This metric can be used to identify regions with a significant concentration of species with restricted distribution (Williams et al. 1994, Crisp et al. 2001, Slatyer et al. 2007, Rosauer et al. 2009). In this way, its maximization can be used as an optimization criterion for the allocation of conservation resources.

Taxon-based measures can exclude an important biodiversity facet: the spatial restriction of evolutionary diversity. The accumulation of evolutionary heritage and its variation can be incorporated into the endemism patterns through the phylogenetic relationships between species (Rosauer et al. 2009). Phylogenetic diversity (PD; Faith 1992) is a simple and broadly used measure that assesses the cumulative evolutionary history of a set of taxa distributed in a region (Faith 1992, Moritz and Faith 1998). This metric sums the branch lengths from a set of species that often share a geographic location and may reflect the contribution of each taxon to the group diversity. PD is considered a robust metric in the presence of taxonomic uncertainties because it uses branch lengths of a phylogeny as a measure of diversity, which tends to be less susceptible to change than species or nodes (Mace et al. 2003).

To evaluate the relative contribution of species to PD (i.e. the species 'originality'), researchers often use evolutionary distinctiveness (ED; Pavoine et al. 2005, Isaac et al. 2007). This metric allows the evaluation of species that are evolutionarily distinct in the community (Isaac et al. 2007) and can be applied to the conservation of unique species or entire regions (Cadotte and Davies 2010). Information about the 'originality' of each species can be combined with the risk of extinction to identify species that are evolutionarily distinct and that are globally threatened (EDGE approach; Isaac et al. 2007).

An additional metric widely used is phylogenetic endemism (PE), which uses the sum of the branch length weighted by the clade range to identify regions with a spatial concentration of evolutionary isolation (Rosauer et al. 2009, Rosauer and Jetz 2014). A region with high PE may be formed in areas that harbor taxa with long branch lengths restricted to that area (Rosauer and Jetz 2014). As PE enables the identification of areas with restricted and evolutionarily unique

biota, this metric also can be used as an index of ecological vulnerability, allowing the identification of priority regions for conservation.

The importance of these metrics (WE, PD, ED and PE) for conservation is extensively recognized, leading to many studies that investigate macroecological and biogeographical patterns linked with phylogenetic data (Faith et al. 2004, Barratt et al. 2017, Faith 2018). To explore these research questions, a wide array of existing R packages (www.r-project.org), such as 'phyloregion' (Daru et al. 2020), 'picante' (Kembel et al. 2010), and 'pez' (Pearse et al. 2015), offer calculations of evolutionary diversity using a diversity of data structures like vectors, large matrices, sparse matrices, and even presence-absence rasters – as demonstrated by 'EcoPhyloMapper' (Title et al. 2022). Additionally, there are packages outside the R ecosystem that address these metrics, including 'Biodiverse' (Laffan et al. 2010), 'SDMToolbox' (www.sdmtoolbox.org), and 'lifemapper' (<https://lifemapper.github.io/>). However, some of these packages do not efficiently handle large data sets or they take a long time to perform calculations.

Here, we introduce 'phyloraster', an R package (www.r-project.org) designed to compute measures of endemism and evolutionary diversity using presence-absence rasters and phylogenetic information as input. Our package offers a range of functions, including calculations for species richness, Faith's phylogenetic diversity (Faith 1992), phylogenetic endemism (Rosauer et al. 2009), evolutionary distinctiveness (Isaac et al. 2007), and weighted endemism (Williams et al. 1994). Moreover, 'phyloraster' includes functions to generate null models through various spatial randomization methods, allowing researchers to control for richness effects (Gotelli and Groves 1996, Gotelli and Ulrich 2012). With these comprehensive tools, our package aims to enhance the analysis of spatial patterns of endemism and evolutionary diversity providing valuable insights for conservation and ecological research.

Novelty

The increasing size and complexity of biogeographical and phylogenetic data highlight the need to provide functions that allow efficient and fast manipulation of these datasets (Daru et al. 2020). The package 'phyloraster' provides a set of functions to support the results derived from species distribution models (SDMs) or distribution polygons with phylogenetic data. Currently, many researchers are using SDMs to predict the potential distribution of species. Despite the usefulness of these models for biogeographical studies, the results generated by SDMs are provided in raster format, usually for large regions and at high resolution. Therefore, analyzing the rasters generated by the SDMs can generate scaling issues and easily exhaust the available RAM, even using sparse matrices to store the site-by-species data. Our package efficiently handles large datasets as we provide functions to calculate diversity measures directly from geospatial data such as raster objects (details in 'Raster implementations' section).

One of the main advantages of using rasters as input is that only the information about the data structure (e.g. row numbers, extent, resolution, cell numbers) is loaded in the RAM, without necessarily loading all raster cells at the same time (Hijmans 2022).

Our package also differs from others in the calculation of phylogenetic endemism and weighted endemism because the code corrects non-equal areas for a geographic coordinate system. Most existing packages calculate the range size of each species assuming that all cells have equal sizes. In this case, if the user has projected the data onto a coordinate system that does not assume that the area size is equal, there may be bias in the range size estimation and all related metrics (e.g. WE, PE) to large extents. For example, two species occupying the same number of cells in a geographic coordinate system, one at the equator and another at a subtropical latitude, will differ in the area occupied, with the tropical species having a larger range size than the subtropical one. Yet, naive calculations will erroneously consider both as having the same range size. Our function addresses this problem by calculating the range size of each species considering the actual size of raster cells (examples of `cellSize()` function of 'terra' package; for more information about projections see <https://proj.org/operations/projections/index.html>).

Finally, 'phyloraster' performs spatial randomization procedures implemented in the 'SESRaster' R package (Heming et al. 2023). 'SESRaster' currently has six algorithms for randomizing binary species distribution rasters (details in 'Spatial and phylogenetic randomizations' section) and allows the use of custom randomization algorithms. This implementation represents a novelty, as this breadth of randomization methods is not available in similar packages that calculate evolutionary diversity and endemism metrics. The available randomization functions allow the creation of communities that vary in distribution patterns for comparison with the observed patterns of evolutionary diversity (Kembel et al. 2010). In cases where patterns of evolutionary diversity and species richness are closely related, it can be very interesting to apply these randomization methods to test hypotheses about the phylogenetic community structure (Kembel et al. 2010), such as Li et al. (2015) and Mazel et al., (2016). More details about each randomization method can be obtained in the 'SESRaster' vignettes.

Raster implementations

One of the main advantages of using rasters as input is that if there is enough RAM available to store and process the raster data, it can be entirely loaded in RAM, otherwise the rasters are saved on the disk and only the information about the data structure (e.g. row numbers, extent, resolution, number of cells) is loaded in the RAM (Hijmans 2022). Furthermore, the calculations are applied to one cell at a time preventing filling up the RAM during raster processing. Function implementation in 'phyloraster' also ensures that a minimal number of temporary raster files are created during processing and that these temporary files are

automatically cleaned up after use. Finally, randomization procedures implemented in 'phyloraster' derive from the 'SESRaster' R package (Heming et al. 2023), which are specifically designed for raster data (details in 'Spatial and phylogenetic randomizations' section available in the 'SESRaster' vignette).

Methods and features

'phyloraster' is written in R (ver. ≥ 2.10 , www.r-project.org) language and environment. The R packages used as dependencies are: 'terra' (ver. ≥ 1.6) (Hijmans 2022), 'ape' (ver. ≥ 5.6) (Paradis and Schliep 2019), 'SESRaster' (ver. ≥ 0.7) (Heming et al. 2023), and 'phylobase' (ver. ≥ 0.810) (Hackathon 2020). Users are free to suggest improvements and report issues through the topic 'Issues' on the GitHub repository. The package can be installed from CRAN and loaded running the following code:

```
install.packages("phyloraster")
library(phyloraster)
```

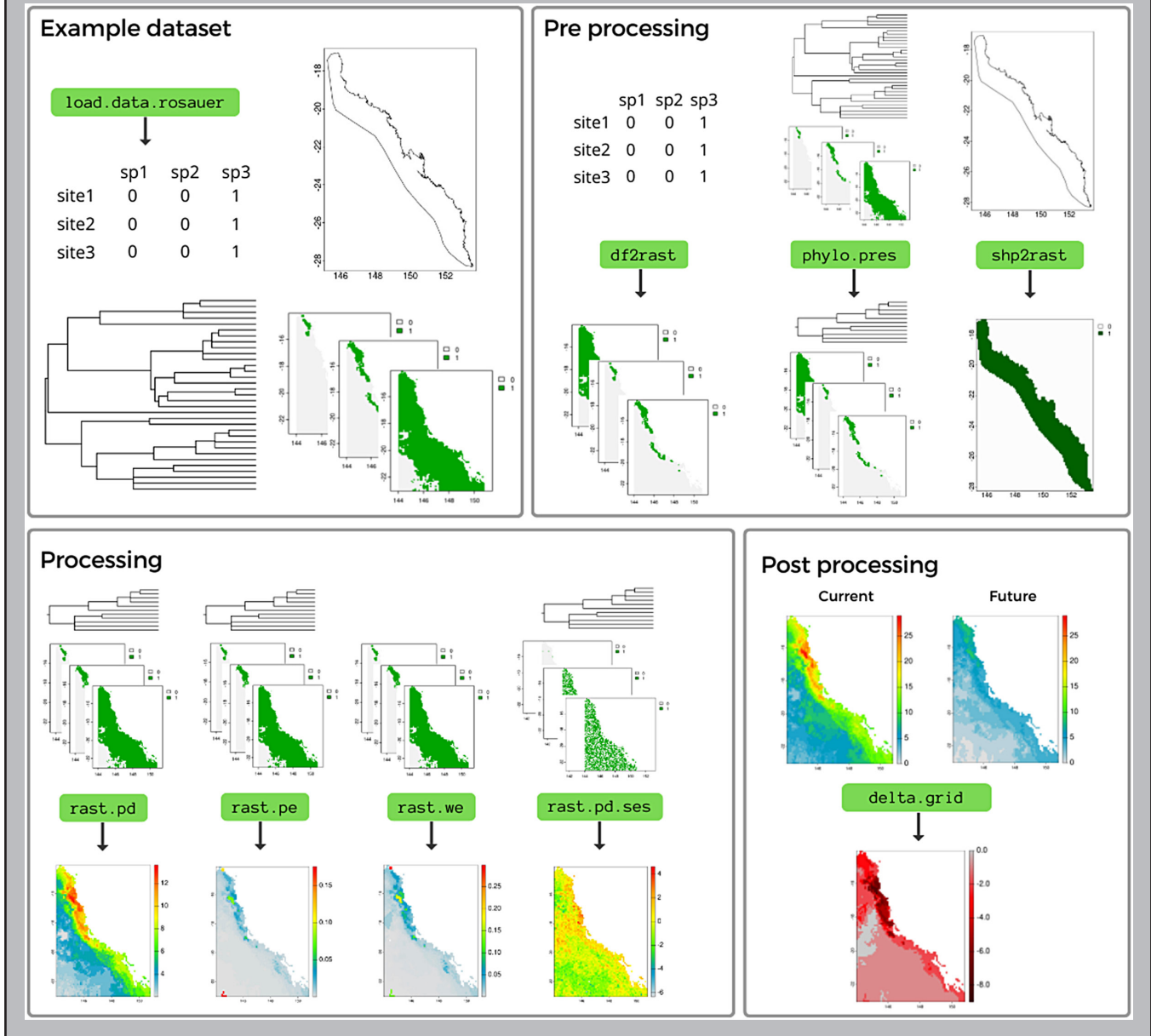
Data preparation

The functions of the 'phyloraster' package encompass pre-processing, processing and post-processing of macroecological and phylogenetic data (Box 1). In the first step, we offer support to manipulate matrices, shapefiles, rasters, and phylogenetic trees, including functions to generate the required data structures for performing subsequent analyses.

The function `df2rast()` converts traditional community matrices (i.e. species in columns and sites in rows, with coordinates in the two first columns) into binary distribution rasters. The package contains one dataset that allows visualizing the structure expected to matrices, with species in the columns and sites in the rows. This dataset contains presence records for 33 Australian tree frogs with coordinates for each site (Rosauer 2017). Another dataset widely used in macroecological analyses are the shapes of species distribution provided by the International Union for the Conservation of Nature's Spatial Database (www.iucnredlist.org/resources/spatial-data-download). The `shp2rast()` function allows working with vectorized distribution data by converting shapefiles to raster stack.

An important step in macroecological and biogeographic analyses paired with evolutionary hypotheses is to ensure that phylogenetic and distribution data match. In this sense, 'phyloraster' implements the function `phylo.pres()`. This function reorders the raster stack to match the order of the tips of the tree, extracts a sub-tree containing only species present in the raster stack, and gets the branch length and descendant number for each species. The user also has the option to compute branch length and descendant number using the full supplied tree or the tree subsetted by the species present in the raster. Notice the implications of using the full or the subsetted tree. Consider, for instance, a scenario where a clade comprises

Box 1. Schematic examples of the functions available in the package. The functions of ‘phyloraster’ are focused on pre-processing, processing, and post-processing of macroecological and phylogenetic data. The example dataset includes shapefiles from IUCN, matrices of presence-absence, phylogeny, and raster of presence-absence for 33 tree frog species from Australia.



three species (A–C), and the particular area of study involves two of these species (A, B). Furthermore, let's assume that species A and B share a branch, denoted as D (the ‘phyloraster’ vignette). Using the full phylogenetic tree will estimate the whole length of branches for these two species, including the branch shared between them (D), that connects them with the ancestor shared with the species absent from that specific region. On the other hand, when using the subsetted tree, branch D will be disregarded and only the terminal branches will be used to calculate branch length, so that the calculated branch lengths of the species A and B will be shorter.

Endemism measures

The ‘phyloraster’ R package implements functions for calculating spatial patterns of endemism based on the weighted endemism method (WE; Williams et al. 1994, Crisp et al. 2001) through the function *rast.we()*. WE weight the species range by the proportion of the range of each species present in a given region Eq. 1 (Laffan et al. 2016),

$$WE = \sum_{\{c \in C\}} \frac{r_c}{R_c} \quad (1)$$

where r is the local range (in our case, the cell area) of taxon c , R_c is the total range size of the taxon c and C is the subset of taxa that occur in a given region (Eq. 1). Most implementations usually account for the size of the distribution area of the species based on the number of cells that the species occupies and consider all cells to be the same size (Laffan et al. 2016). However, our function calculates the range size of each species treating all cells as the actual cell size. The *rast.we()* function inputs a presence/absence raster for a given community and returns a raster with the values of weighted endemism by each pixel for the extent of the input rasters.

Evolutionary measures

The ‘*phyloraster*’ package implements three types of measures to calculate evolutionary diversity through the functions *rast.pd()*, *rast.pe()* and *rast.ed()*. The first is Faith’s phylogenetic diversity (PD), calculated as the sum of the branch lengths for all species occurring in a given region (Eq. 2) (Faith 1992),

$$PD = \sum_{\{c \in C\}} L_c \quad (2)$$

where L_c is the branch lengths of species c and C is the set of branches in a given region (Eq. 2).

The second metric is evolutionary distinctiveness (ED; Isaac et al. 2007) or ‘fair proportion’ (Redding et al. 2014), which is calculated dividing the total phylogenetic diversity of a clade among its members (Isaac et al. 2007). The calculation is done using both branch lengths and the number of descendants (Eq. 3),

$$ED_j = \sum_{b \in C(j, \text{root})} \frac{L_b}{N_b} \quad (3)$$

where L_b is the edge length of branch b , N_b stands for the number of species that descend from branch b and $C(j, \text{root})$ for the set of branches between species j , the tip of the tree and the root of the tree.

The third measure of diversity is phylogenetic endemism (PE), which calculates the degree to which PD is restricted to a specific region (Rosauer et al. 2009, Laffan et al. 2016). Therefore, to calculate PE for a given region we consider both ranges size and branch lengths for each species (Eq. 4),

$$PE = \sum_{\{c \in C\}} L_c \frac{r_c}{R_c} \quad (4)$$

where L_c is the branch length of taxon c , r_c is the local range (in our case, the cell area) of branch c , and R_c is the range sizes of the clade. C is the set of branches in a given region.

Standardized effect size

Null models are a widely used method to compare patterns against random processes, allowing for example, to evaluate richness effects in diversity measures (Gotelli and Groves 1996, Gotelli and Ulrich 2012), and to test hypotheses about the community structure. The standardized effect size (SES) is widely used in the community structure literature and can be calculated from null models (Gotelli and McCabe 2002). The SES is commonly used to measure the deviation from the null expectation in standard deviation units (Eq. 5), and enables an estimation of the relative position of the observed value with respect to the null distribution generated by the randomization (Mazel et al. 2016).

$$SES = \frac{\text{Metric}_{\text{obs}} - \text{Mean}(\text{Metric}_{\text{null}})}{\text{SD}(\text{Metric}_{\text{null}})} \quad (5)$$

where $\text{Metric}_{\text{obs}}$ represents the observed value for the metric, $\text{mean}(\text{Metric}_{\text{null}})$ represents the mean of randomizations calculated for n times and $\text{SD}(\text{Metric}_{\text{null}})$ represents the standard deviation of the randomizations.

Spatial and phylogenetic randomization

The randomization procedure for the calculation of SES is done internally in the functions *rast.we.ses()*, *rast.pd.ses()*, *rast.ed.ses()*, *rast.pe.ses()* and *geo.phylo.ses()* through the package ‘SESRaster’ (Heming et al. 2023). ‘SESRaster’ currently implements six algorithms to randomize binary species distribution with several levels of constraints: SIM1, SIM2, SIM3, SIM5, SIM6 and SIM9 (*sensu* Gotelli 2000). The methods implemented in ‘SESRaster’ are based on how species (originally rows) and sites (originally columns) are treated (i.e. fixed, equiprobable, or proportional sums) (Gotelli 2000). The randomization algorithms currently available in ‘SESRaster’ are: SIM1 (species occurrence equiprobable and site richness equiprobable), SIM2 (species occurrence fixed and site richness equiprobable), SIM3 (species occurrence equiprobable and site richness fixed), SIM5 (species occurrence proportional and site richness fixed), SIM6 (species occurrence proportional and site richness fixed) and SIM9 (species occurrence fixed and site richness fixed, similar to the preserved model of Laffan and Crisp 2003). In addition, ‘SESRaster’ (consequently ‘*phyloraster*’) supports user’s custom randomization algorithms for SES calculation, as long as the function returns objects of class *SpatRaster*. This allows complete flexibility for using any algorithm not yet implemented by the package. To see more details about the randomization methods cited above, review the documentation of the ‘SESRaster’ package.

By default, the ‘*phyloraster*’ uses the function *bootspat_str()* from the ‘SESRaster’ package to conduct the randomizations, but the user is free to choose any of the other methods mentioned above through the *spat_alg* argument in the **.ses()* functions of the ‘*phyloraster*’ package. The function

bootspat_str() is equivalent to the SIM5 (proportional-fixed) method of [Gotelli \(2000\)](#), which partially relaxes the spatial structure of species distributions, but keeps the spatial structure of the observed richness pattern across cells. This method differs from [Laffan and Crisp \(2003\)](#) because their implementation shuffles the species presences across the raster, while *bootspat_str()* samples presences based on probabilities computed from their frequencies ([Heming et al. 2023](#)). Although species frequencies are not exact, the variation is relatively small, not compromising species range size patterns. The randomization will not assign values to cells with no data. The preserved model of [Laffan and Crisp \(2003\)](#), for retaining the richness pattern and the range size of the species (SIM9, fixed-fixed, [Gotelli 2000](#)), is available in the *bootspat_ff()* function of ‘SESraster’.

Note that although species range sizes are estimated using the size of the raster cell, the currently available randomization methods do not take this information into account, as shuffling is made based on the number of occupied pixels. In a geographic coordinate system, the area of the smaller polar pixels (above 60° N and 60° S – higher latitude) is nearly five times smaller than the larger equatorial pixels (close to 0° – lower latitudes). So, if a strictly equatorial species that occupies ten pixels is assigned to ten polar pixels, its range area will be nearly five times smaller than it actually is. This affects the randomized metrics especially on large latitudinal extents. We are not aware of any randomization algorithm implemented in R that explicitly overcomes this limitation and we are sure that further attention is needed to solve this shortcoming. Phylogenetic randomization can also be done using the package ‘SESraster’. The randomization can shuffle taxon branch lengths prior to PD and ED calculations.

Post-processing analysis

The ‘phyloraster’ package offers a function to evaluate the change in the community over time. This function can be applied to the results obtained from the functions *geo.phylo()* or *rast.sr()*, *rast.we()*, *rast.pd()*, *rast.ed()* and *rast.pe()*, which can represent different time points such as baseline, past and future scenarios. The *delta.grid()* function allows to evaluate the change in the community diversity metrics through time. By comparing present and future diversity patterns, *delta.grid()* reveals any variations across regions, highlighting diversity shifts resulting from environmental changes.

Implementation examples

To demonstrate how ‘phyloraster’ can be used, we developed a study case where we calculated endemism and evolutionary diversity patterns for 33 tree frog species of the subfamily Pelodyadinae from Australia ([Rosauer 2017](#)). This dataset can be accessed through the function *load.data.rosauer()*. To perform the calculations, first, we transform the presence and absence matrix into a raster using the *df2rast()* function. The function maintains the original resolution of the data. In this case, the grid cells have a resolution of 0.1°. Then,

we use the *phylo.pres()* function to sort the raster according to the tree order, extract the branch length for each species, and subset the phylogenetic tree maintaining only the species that are present in the raster. With the raster sorted and the branch lengths in hand, we can calculate Faith’s phylogenetic diversity using the *rast.pd()* function. Because richness is positively correlated with Faith’s phylogenetic diversity ([Tucker and Cadotte 2013](#)), we calculated the SES through the function *rast.pd.ses()*, using the argument *random = ‘spat’* and the argument *spat_alg = ‘bootspat_str’* (Supporting information, Fig. S1). We used 999 simulations, defined through the argument *aleats* of the *rast.pd.ses()* function.

To calculate the WE, we weight the species range by the proportion of the range of each species present in a given region. This calculation is done internally in the *rast.we()* function (Supporting information, Fig. S1). The spatial patterns of weighted endemism can also be calculated using the *geo.phylo()* function. After that, we calculate PE through the *rast.pe()* function using the raster stack with presence-absences and a phylogenetic tree for a set of species. The clade range is calculated internally in the function *rast.pe()*.

The results can be seen in the Supporting information. PD was highest in the northeast and shows a decrease towards the region south (Supporting information, Fig. S1). Meanwhile, PE has the highest values concentrated in the extreme north of the region above latitude –16, and moderate PE values between latitude –16 and –18 (Supporting information, Fig. S1). A small fraction of PE can also be found in the south of the region, between longitudes 144 and 146 (Supporting information, Fig. S1). WE patterns are congruent with PE (Supporting information, Fig. S1).

In addition, we calculated the temporal difference in the PD and the PE through the function *delta.grid* (Supporting information). To assess temporal changes, let’s assume that by 2050, four species *Litoria lorica*, *Litoria rheocola*, *Litoria nyakalensis* and *Litoria infrafrenata* will be completely extinct due to climate change and calculated the PD and PE for baseline and future scenarios. The results generated by the function shows that the potential loss of four species in the future will decrease the PD by up to 2.365. With the loss of these species, PE should also change, with some regions gaining up to 0.014 and others losing up to 0.082. The script and dataset used to run the example implementation and generate the figures can be accessed at: [Alves-Ferreira et al. 2022](#).

Performance comparisons

We compared the performance of ‘phyloraster’ with two packages that have similar functionalities: ‘epm’ ([Title et al. 2022](#)) and ‘phyloregion’ ([Daru et al. 2020](#)). For the performance comparison, we evaluated the patterns of weighted endemism (WE) and phylogenetic diversity (PD) using a dataset of geographical distribution (presence and absence) of 82 tree frog species of the subfamily Pelodyadinae. We tested the performance using two different resolutions (0.1° and 0.05°) and considered both the functions for data preparation from each package and the specific function for calculating the metrics

Table 1. Comparison of ‘phyloraster’ against ‘epm’ and ‘phyloregion’ for analysis of endemism and phylogenetic diversity for tree frog species of Australia.

Metric	Package	Step	Specific functions of each package	Memory allocation	Time	Spatial resolution
Weighted endemism	<i>epm</i>	Data import	(createEPMgrid)	562.07 MB	00 m 01.23 s	0.05°
	<i>phyloraster</i>	Data import	(shp2rast)	83.22 KB	00 m 00.91 s	0.05°
	<i>phyloregion</i>	Data import	(rast2comm)	210.01 MB	00 m 01.34 s	0.05°
	<i>epm</i>	Metric calculation	(gridMetrics)	338.88 MB	00 m 00.59 s	0.05°
	<i>phyloraster</i>	Metric calculation	(rast.we)	8.38 KB	00 m 00.50 s	0.05°
	<i>phyloregion</i>	Metric calculation	(weighted_endemism)	574.95 KB	00 m 00.43 s	0.05°
Phylogenetic diversity	<i>epm</i>	Data import	(createEPMgrid), (addPhylo)	113.22 MB	00 m 00.79 s	0.05°
	<i>phyloraster</i>	Data import	(shp2rast), (phylo.pres)	18.29 MB	00 m 00.50 s	0.05°
	<i>phyloregion</i>	Data import	(rast2comm)	210.01 MB	00 m 01.35 s	0.05°
	<i>epm</i>	Metric calculation	(gridMetrics)	59.06 MB	00 m 00.14 s	0.05°
	<i>phyloraster</i>	Metric calculation	(rast.pd)	44.27 MB	00 m 01.17 s	0.05°
	<i>phyloregion</i>	Metric calculation	(PD)	2.99 MB	00 m 00.32 s	0.05°

WE and PD. The script and dataset used to compare packages can be accessed from: <https://doi.org/10.7910/DVN/QSNTSG>. To run the tests we used a machine with the following specifications: ‘AMD® Ryzen 7 4800h with radeon graphics × 16 cores’ with 46.4GB RAM, 16GB SWAP, and 256GB SSD. The software used was Linux Ubuntu 20.04.6 LTS, and the R ver. is 4.3.1 (2023-06-16, www.r-project.org). For simplicity we show only the results for the high resolution dataset. The results for the low resolution dataset can be found in the Supporting information, Table S1.

According to the benchmarking, loading and preparing data at high resolution (0.05°) before WE calculation requires considerably more RAM in the ‘epm’ (562.07 MB) and in the ‘phyloregion’ (210.01 MB) than in the ‘phyloraster’ (83.22 KB) (Table 1). The WE calculation also consumes more RAM in ‘epm’ (338.88 MB) and ‘phyloregion’ (574.95 KB) packages than in ‘phyloraster’ (8.38 KB). We also found that loading data to calculate the PD in ‘phyloraster’ (18.29 MB) consumes relatively less RAM than in the ‘epm’ (113.22 MB) and in the ‘phyloregion’ (210.01 MB) (Table 1). For the PD calculation, the results show that ‘phyloraster’ (44.27 MB) and ‘epm’ (59.06 MB) uses slightly more RAM than the ‘phyloregion’ (2.99 MB) (Table 1).

The lower RAM usage in ‘phyloraster’ functions is due to the main dependency on the ‘terra’ package, as explained in the ‘Raster implementations’ section. Faster calculations and substantial low RAM usage minimizes the hardware requirements to work with high-resolution datasets from local to global scales. The functions of ‘phyloraster’ broadens user accessibility of the spatialized measures of endemism and evolutionary diversity. By allowing users without access to machines with high processing power and large amounts of RAM to perform analyses of spatial evolutionary patterns, we also hope to promote research equity for low income researchers (as discussed for other scientific topics by [Williams et al. 2023](#)).

Conclusions and future directions

The ‘phyloraster’ package aims to unite species range data with phylogenetic information and facilitate the spatial

analysis of taxon richness, phylogenetic diversity and phylogenetic endemism. The main novelty of this package is the capacity to calculate measures of diversity and endemism directly for rasters with very efficient memory usage and fast processing time. We have shown that the ‘phyloraster’ is lighter and faster than its counterparts, which may allow users to work with high-resolution datasets from local to global scales. By reducing the dependency on machines with high processing power and large amounts of RAM, we hope that research equity for low income researchers is being promoted. In addition, our package differs from others in the calculation of phylogenetic endemism and weighted endemism because it takes into account the latitudinal variation of the pixel area, which affects the estimated range size (and all subsequent analyses) of species occurring along a latitudinal gradient. In upcoming versions, we plan to expand the package functionalities, adding functions for calculating neo and paleo endemism ([Mishler et al. 2014](#)), mean pairwise distance between all species in an assemblage (MPD), and pairwise distance between the closest relatives in an assemblage (MNTD) ([Webb et al. 2002](#)).

To cite ‘phyloraster’ or acknowledge its use, cite this Software note as follows, substituting the version of the application that you used for ‘version 1.0’:

Alves-Ferreira, G. et al. 2023. ‘phyloraster’: an R package to calculate measures of endemism and evolutionary diversity for rasters. – *Ecography* 2023: e06902 (ver. 2.0.1).

Acknowledgements – GA-F thanks for the ‘Introdução ao R’ course and the ‘Ciência Replicável’ course ministered by Professor Neander Marcel Heming in the postgraduate program in Ecologia e Conservação da Biodiversidade at the Universidade Estadual de Santa Cruz. GA-F also thanks Idea Wild for donating equipment to develop the codes and the text of this manuscript. We thank Martín de Jesús Cervantes-López for testing the code.

Funding – This work was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior–Brazil (CAPES) (Finance code 001) through a Doctoral Scholarship to GA-F and CO. NMH acknowledges the post-doc fellowship (no. 88882.314922/2019-01) received from CAPES. This work was

also supported by the project approved by the Pró-Reitoria de Pesquisa e Pós-Graduação (PROPP) – UESC (no. 0731101020220030757-31).

Author contributions

Gabriela Alves-Ferreira: Conceptualization (lead); Data curation (lead); Methodology (equal); Software (equal); Validation (lead); Visualization (lead); Writing – original draft (lead); Writing – review and editing (lead). **Flávio Mariano Machado Mota:** Methodology (supporting); Software (supporting); Validation (supporting); Visualization (supporting); Writing – review and editing (supporting). **Daniela Custódio Talora:** Conceptualization (supporting); Supervision (supporting); Validation (supporting); Visualization (supporting); Writing – review and editing (supporting). **Cynthia Valéria Oliveria:** Conceptualization (supporting); Validation (supporting); Visualization (supporting); Writing – review and editing (supporting). **Mirco Solé:** Conceptualization (supporting); Supervision (supporting); Validation (supporting); Visualization (supporting); Writing – review and editing (supporting). **Neander Marcel Heming:** Conceptualization (supporting); Data curation (supporting); Methodology (equal); Software (equal); Supervision (lead); Validation (supporting); Writing – review and editing (supporting).

Transparent peer review

The peer review history for this article is available at <https://publons.com/publon/10.1111/ecog.06902>.

Data availability statement

The data and scripts used in this article are available on Harvard Dataverse Repository: <https://doi.org/10.7910/DVN/QSNTSG>.

Supporting information

The Supporting information associated with this article is available with the online version.

References

Alves-Ferreira, G., Mota, F., Oliveira, C., Talora, D., Solé, M. and Heming, N. 2022. phyloraster: an R package to calculate spatialized measures of endemism and evolutionary diversity, vol. 4. – Harvard Dataverse, <https://doi.org/10.7910/DVN/QSNTSG>.
Anderson, S. 1994. Area and endemism. – *Q. Rev. Biol.* 69: 451–471.
Barratt, C. D., Bwong, B. A., Onstein, R. E., Rosauer, D. F., Menegon, M., Doggart, N., Nagel, P., Kissling, W. D. and Loader, S. P. 2017. Environmental correlates of phylogenetic endemism in amphibians and the conservation of refugia in the Coastal Forests of eastern Africa. – *Divers. Distrib.* 23: 875–887.
Cadotte, M. W. and Davies, T. J. 2010. Rarest of the rare: advances in combining evolutionary distinctiveness and scarcity to

inform conservation at biogeographical scales. – *Divers. Distrib.* 16: 376–385.
Crisp, M. D., Laffan, S., Linder, H. P. and Monro, A. 2001. Endemism in the Australian flora: endemism in the Australian flora. – *J. Biogeogr.* 28: 183–198.
Daru, B. H., Karunarathne, P. and Schliep, K. 2020. phyloregion: R package for biogeographical regionalization and macroecology. – *Methods Ecol. Evol.* 11: 1483–1491.
Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. – *Biol. Conserv.* 61: 1–10.
Faith, D. P., Reid, C. A. M. and Hunter, J. 2004. Integrating phylogenetic diversity, complementarity, and endemism for conservation assessment. – *Conserv. Biol.* 18: 255–261.
Faith, D. P. 2018. Phylogenetic diversity and conservation evaluation: perspectives on multiple values, indices, and scales of application. – In: Scherson, R. A. and Faith, D. P. (eds), *Phylogenetic diversity*. Springer, pp. 1–26.
Gotelli, N. J. 2000. Null model analysis of species co-occurrence patterns. – *Ecology* 81: 2606–2621.
Gotelli, N. J. and Groves, G. R. 1996. *Null models in ecology*. – Smithsonian Institution Press.
Gotelli, N. J. and McCabe, D. J. 2002. Species co-occurrence: a meta-analysis of J. M. Diamond's assembly rules model. – *Ecology* 83: 2091–2096.
Gotelli, N. J. and Ulrich, W. 2012. Statistical challenges in null model analysis. – *Oikos* 121: 171–180.
Hackathon, R. 2020. phylobase: base package for phylogenetic structures and comparative data. – <https://CRAN.R-project.org/package=phylobase>.
Heming, N. M., Mota, F. M. M. and Alves-Ferreira, G. 2023. SESraster: raster randomization for null hypothesis testing. – R package ver. 0.7, <https://CRAN.R-project.org/package=SESraster>.
Hijmans, R. J. 2022. Terra, spatial data analysis (1.6.7). – <https://CRAN.R-project.org/package=terra>.
Isaac, N. J., Turvey, S. T., Collen, B., Waterman, C. and Baillie, J. E. 2007. Mammals on the EDGE: conservation priorities based on threat and phylogeny. – *PLoS One* 2: e296.
Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., Blomberg, S. P. and Webb, C. O. 2010. Picante: R tools for integrating phylogenies and ecology. – *Bioinformatics* 26: 1463–1464.
Laffan, S. W. and Crisp, M. D. 2003. Assessing endemism at multiple spatial scales, with an example from the Australian vascular flora: assessing endemism at multiple spatial scales. – *J. Biogeogr.* 30: 511–520.
Laffan, S. W., Lubarsky, E. and Rosauer, D. F. 2010. Biodiverse, a tool for the spatial analysis of biological and related diversity. – *Ecography* 33: 643–647. Ver. 1.0.
Laffan, S. W., Rosauer, D. F., Di Virgilio, G., Miller, J. T., González-Orozco, C. E., Knerr, N., Thornhill, A. H. and Mishler, B. D. and Mishler, B. D. 2016. Range-weighted metrics of species and phylogenetic turnover can better resolve biogeographic transition zones. – *Methods Ecol. Evol.* 7: 580–588.
Li, R., Kraft, N. J., Yu, H. and Li, H. 2015. Seed plant phylogenetic diversity and species richness in conservation planning within a global biodiversity hotspot in eastern Asia. – *Conserv. Biol.* 29: 1552–1562.
Mace, G. M., Gittleman, J. L. and Purvis, A. 2003. Preserving the tree of life. – *Science* 300: 1707–1709.
Mazel, F., Davies, T. J., Gallien, L., Renaud, J., Groussin, M., Münkemüller, T. and Thuiller, W. 2016. Influence of tree shape and evolutionary time-scale on phylogenetic diversity metrics. – *Ecography* 39: 913–920.

- Mishler, B. D., Knerr, N., González-Orozco, C. E., Thornhill, A. H., Laffan, S. W. and Miller, J. T. 2014. Phylogenetic measures of biodiversity and neo- and paleo-endemism in Australian *Acacia*. – *Nat. Commun.* 5: 4473.
- Moritz, C. and Faith, D. P. 1998. Comparative phylogeography and the identification of genetically divergent areas for conservation. – *Mol. Ecol.* 7: 419–429.
- Paradis, E. and Schliep, K. 2019. *ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R*. – *Bioinformatics* 35: 526–528.
- Pavoine, S., Ollier, S. and Dufour, A. B. 2005. Is the originality of a species measurable? – *Ecol. Lett.* 8: 579–586.
- Pearse, W. D., Cadotte, M. W., Cavender-Bares, J., Ives, A. R., Tucker, C. M., Walker, S. C. and Helmus, M. R. 2015. *pez: phylogenetics for the environmental sciences*. – *Bioinformatics* 31: 2888–2890.
- Redding, D. W., Mazel, F. and Mooers, A. Ø. 2014. Measuring evolutionary isolation for conservation. – *PLoS One* 9: e113490.
- Rosauer, D., Laffan, S. W., Crisp, M. D., Donnellan, S. C. and Cook, L. G. 2009. Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. – *Mol. Ecol.* 18: 4061–4072.
- Rosauer, D. F. and Jetz, W. 2014. Phylogenetic endemism in terrestrial mammals: mammal phylogenetic endemism. – *Global Ecol. Biogeogr.* 24: 168–179.
- Rosauer, D. F. 2017. Phylogenetic tree for 33 Australian tree frogs. – <https://github.com/DanRosauer/phylospatial>.
- Slatyer, C., Rosauer, D. and Lemckert, F. 2007. An assessment of endemism and species richness patterns in the Australian Anura. – *J. Biogeogr.* 34: 583–596.
- Title, P. O., Swiderski, D. L. and Zelditch, M. L. 2022. *EcoPhyloMapper: an R package for integrating geographical ranges, phylogeny and morphology*. – *Methods Ecol. Evol.* 13: 1912–1922.
- Tucker, C. M. and Cadotte, M. W. 2013. Unifying measures of biodiversity: understanding when richness and phylogenetic diversity should be congruent. – *Divers. Distrib.* 19: 845–854.
- Webb, C. O., Ackerly, D. D., McPeck, M. A. and Donoghue, M. J. 2002. Phylogenies and community ecology. – *Annu. Rev. Ecol. Syst.* 33: 475–505.
- Weber, M. M., Stevens, R. D., Lorini, M. L. and Grelle, C. E. V. 2014. Have old species reached most environmentally suitable areas? A case study with South American phyllostomid bats: have old species reached most suitable areas? – *Global Ecol. Biogeogr.* 23: 1177–1185.
- Williams, P., Humphries, C. J. and Forey, P. L. 1994. *Biodiversity, taxonomic relatedness, and endemism in conservation*, vol. 50. – Oxford Univ. Press.
- Williams, J. W. et al. 2023. Shifts to open access with high article processing charges hinder research equity and careers. – *J. Biogeogr.* 50: 1485–1489.