Herodotools: An R package to integrate macroevolution, community ecology, and biogeography

Gabriel Nakamura[1*]; Arthur V Rodrigues[2]; André Luís Luza[2,3]; Renan Maestri[2]; Vanderlei Debastiani[2] and Leandro da Silva Duarte[2]

1- Texas A&M University-Corpus Christi, Department of Life Sciences, 6300 Ocean Dr. 78412

2 – Universidade Federal do Rio Grande do Sul, Departamento de Ecologia

3 – Departamento de Ecologia e Evolução. Universidade Federal de Santa Maria, Santa Maria - RS

*corresponding author: gabriel.nakamura.souza@gmail.com

Running title: Integrating macroevolution and ecology

Acknowledgments

G.N. was supported by Texas A&M University-Corpus Christi. LD research was supported by a CNPq Productivity Fellowship (grant 307527/2018-2). R.M. was supported by CNPq (150391/2017-0) and UFRGS. G.N. and L.D. are members of the National Institute for Science and Technology (INCT) in Ecology, Evolution, and Biodiversity Conservation, supported by Ministério da Ciência, Tecnologia, Inovações e Comunicações/ CNPq (proc. 465610/2014-5) and FAPEG (proc. 201810267000023). AVR was supported by CNPq graduate fellowship, and ALL was supported by CNPq (#164240/2021-7, #151228/2021-3, #152410/2020-1).

Abstract

Historical processes like speciation, extinction and historical dispersal are the ultimate factors generating and maintaining biodiversity. Therefore, understanding how these factors affect the distribution of biodiversity is of great importance. To do so, it is necessary to integrate information from ancestral state reconstructions and current species distribution data and traits. Studies that integrated both information proved effective in unveiling questions in the intersection of macroecology, macroevolution, and community ecology. However, up to now, numerical methods that perform this integration are scattered, making integration difficult and hampering advances in these research fields. Here we developed Herodotools, an R package that integrates the macroevolutionary models with the distribution of species occurrence in assemblages to provide metrics that represent historical information, such as in-situ diversification, historical dispersal, and age of assemblages. We described the main functions and illustrated the use of our new package by analyzing the historical biogeography of the genus *Akodon*, a South American small rodent. Our package provides the first platform to investigate questions that require the integration of macroevolutionary information with ecological data (as species occurrence) and streamline analysis of historical biogeography, leveraging the investigation of the effects of historical processes in different levels of organization, from local assemblages to bioregions.

Keywords: Historical biogeography; macroevolutionary dynamics; ancestral state reconstruction

Introduction

Evolutionary processes such as speciation, extinction, and historical dispersal are the ultimate factors promoting the distribution of biological diversity across space and time (Ricklefs, 1987; Ricklefs & Jenkins, 2011; Wiens & Donoghue, 2004a). Despite the importance of those processes, they are usually acknowledged to be predominant on the spatial macro scale or affect macroevolutionary dynamics of lineages through time, but at the regional and local scales, they are less often properly assessed (Mouquet et al., 2012) or, when investigated, patterns are only interpretable by adopting simplified premises as, for example, that phylogenetically clustered communities (i.e., communities predominantly composed by close relatives) are the result of local diversification (Crouch et al., 2019). Even in cases in which historical variables are explicitly modeled (e.g., by analyzing mean values of tip-based metrics of diversification in assemblages across space, Jetz et al. 2012), the temporal dynamics in ancestral states during evolutionary time, which is the basis of phylogenetically clustered or overdispersed communities, is not considered or quantified. Consequently, we have only a limited understanding of deep past processes in generating and maintaining biodiversity patterns at assemblage scales (Maestri et al., 2019; Mouquet et al., 2012; Ricklefs & Jenkins, 2011).

When macroevolution and ecological variation (represented mainly by variation in current species occurrence) are approached separately, three general problems emerge. First, when macroevolutionary dynamics are ignored at a local scale, we assume that only contemporary factors are important, ending up with a limited picture regarding the role of evolution in shaping local communities (Cavender-Bares et al., 2009). Second, by assuming unreliable premises regarding macroevolutionary dynamics and by adopting a pattern-to-process

65 approach of trait evolution to interpret community phylogenetic patterns (e.g., static occurrence
66 area was static during lineage evolution or phylogenetic clustering as a proxy of in-situ
67 diversification), we may reach wrong conclusions about the imprints of diversification, and
68 historical dispersal as neither occurrence area is likely to be static nor phylogenetic clustering be
69 produced by diversification (Van Dijk et al., 2021). Finally, we can reach wrong estimates of
70 assemblage characteristics such as the age of assemblages or the role of in situ diversification
71 and historical dispersal (e.g., as shown by Van Dijk et al., 2021). These three general problems
72 limit our ability to reach reliable results and conclusions regarding the importance of
73 macroevolutionary events shaping the distribution of biodiversity at multiple spatial scales.
74 A way to circumvent these three problems would be using methods that integrate/reunite
75 macroevolutionary models with different spatial scales of organization (Mouquet et al., 2012).
76 For example, Van Dijk et al. (2021) coupled a macroevolutionary model of ancestral area
77 reconstruction with species occurrence information to estimate assemblage age and test two
78 concurrent macroecological mechanisms of biodiversity assembly (Out of the tropics and Niche
79 conservatism). Another example is the study by Luza et al. (2021) that coupled a
80 macroevolutionary model of trait evolution to estimates of tip-based metrics to understand how
81 the macroevolutionary dynamics of diet evolution are affected by different environmental
82 contexts in rodent assemblages. Despite only a few, those methods proved to bring reliable
83 information to test hypotheses that no longer can be tested using macroevolutionary models
84 separated from current species occurrence data. However, those methods are scattered in
85 different studies, with no single platform that allows us to perform analysis that integrates the
86 macroevolutionary dynamics with ecological scale, which limit their usage and, consequently,

87 our ability to move forward towards the understanding of deep past on ecological communities
88 (Gerhold et al., 2015).

89 In this work, we present Herodotools, an R package that wraps up functions designed to
90 integrate models of macroevolution in analysis of biogeography and community phylogenetics to
91 detect imprints of historical processes and the effects of macroevolutionary dynamics of traits
92 and ancestral occurrence areas into different spatial scales (from communities to bioregions).
93 Herodotools overcomes the three general problems abovementioned by integrating the current
94 approaches used in macroevolution, community ecology, and biogeography in a single
95 framework that allows projecting the effects of macroevolutionary dynamics in different spatial
96 scales (Figure 1). Herodotools go beyond the visual interpretation of
97 macroevolutionary/ecological dynamics by presenting metrics that explicitly quantify historical
98 components (e.g., age, in-situ diversification, dispersal) and can be used to test concurrent
99 hypotheses producing patterns of biological diversity.

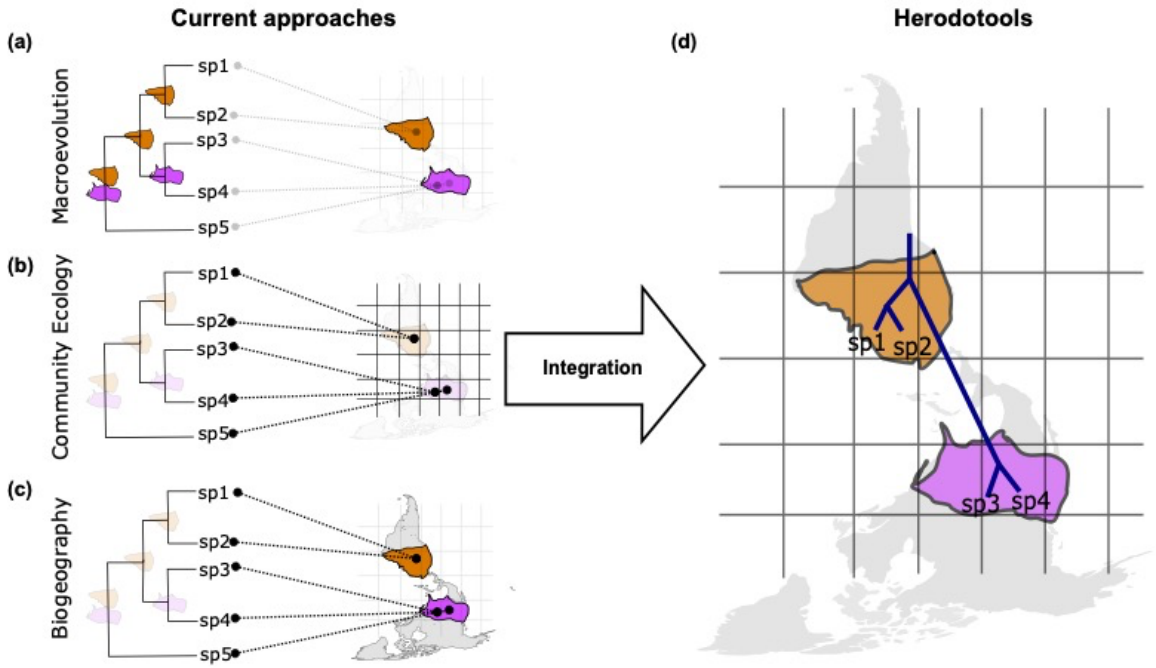


Figure 1: Conceptual representation of the integration implemented in Herodotools package. Macroevolutionary dynamics (a) usually focus on trait evolution and diversification considered at biogeographical scales (orange and purple polygons reconstructed across phylogeny nodes), ignoring the variation in local assemblages (grids). Current ecological methods at assemblage (b) and regional scale (c) usually ignore the macroevolutionary dynamics in space and/or time or approach it using biodiversity proxies. More specifically, macroevolution often ignores spatial processes. Community ecology ignores that assemblages result from dynamics in ancestral occurrence areas and historical processes that build the regional pool of species, and biogeography ignores dynamics in ancestral occurrence area. Herodotools (d) present functions that fill these gaps by integrating the macroevolutionary models in different spatial scales, from assemblage to bioregions.

Here we demonstrate the basic functionalities of Herodotools by analyzing a dataset of genus *Akodon*, a species-rich south american genus of sigmodontine rodents. We aim to show the analytical details behind core functions in the package and exemplify a general pipeline of analysis to investigate the following questions: What is the importance of in situ diversification and historical dispersal to determine the structure of assemblages? How to estimate the age of assemblages? How to quantify trait evolutionary dynamics at assemblage scale (this last using species from the Sigmodontinae family)? These questions represent just a few that can be answered by integrating macroevolution with ecology (McGill et al., 2019) by using Herodotools R package.

Methods

General description of Herodotools package

Herodotools rely on the integration of two different types of data: one that comes from macroevolutionary analysis (*e.g.,* ancestral area and trait reconstruction/mapping), and other from occurrence records of species in spatial units such as assemblages or regions (*e.g.,* biomes, ecoregions, evoregions). Specifically, functions implemented in Herodotools allows for manipulation of data from common macroevolutionary analyses (e.g., ancestral area reconstruction models in BioGeoBears (Matzke, 2013), and ancestral traits reconstruction (Bollback 2006)), converting the output of these analyses to matrices and data frames, which allows calculating macroevolutionary metrics at assemblage level (Table 1). These metrics can be used to map the effects of historical factors at different scales or as variables in common modeling frameworks allowing to test hypotheses in ecology, macroevolution, biogeography,

and community ecology (e.g., Luza et al., 2021; Van Dijk et al., 2021). Additionally, Herodotools perform phylogenetic regionalization methods, map transition zones (Maestri & Duarte, 2020), and detect source and sink regions (functions *evoregions*, *affiliation,* and *dispersal_from*, respectively).

Table 1: Description of the main functions present in Herodotools package

Fields	Function	Description
Data preparation	get_node_range_BioGeoBEARS()	Take BioGeoBears results to obtain a matrix of ancestral occurrence (assemblages x nodes)
	spp_nodes()	Computes a matrix of species (rows) and their respective nodes (columns)
Macroevolution + Biogeography	dispersal_from()	Compute the amount of contribution of each ancestral range to the species composition in other regions
Biogeography	evoregions()	Computes phylogenetic regionalization based on

		phylogenetic fuzzy weighted method
	affiliation_evoreg()	Computes the degree of affiliation of a cell within the region
	spp_association_evoreg()	Classify species in evoregions
	find_max_n_cluster()	Computes the maximum number of clusters to be used on evoregions() function
Macroevolution + Community Ecology	db_diversification()	Computes diversification-based ecophylogenetic metrics (PD and PE)
	age_assemblage()	Compute the age of assemblages
Macroevolution + phenotypic evolution + Community Ecology	tip_based_trait_evo()	Computes tip-based metrics that express trait macroevolutionary dynamics

141

142 The integration of macroevolutionary dynamics into community, biogeographic, and trait
143 analysis comprises two steps, first is the use of an ancestral reconstruction model to decompose
144 the evolutionary history dynamics on the phylogenetic tree in two components, namely ‘in-situ

145 diversification' and 'historical dispersal', and second, use this information to calculate tip-based
146 metrics for each lineage in the phylogenetic tree.
147 The in-situ diversification component comprises the evolutionary history that emerged
148 due to in-situ speciation, i.e., all the events that occurred since each lineage's arrival and
149 establishment time in the region where an assemblage is situated. In other words, it represents
150 the path from tip to root between the species current occurrence in an assemblage to the oldest
151 ancestor in which the range was estimated to occur in the same region as the assemblage,
152 estimated through ancestral area reconstruction (Van Dijk et al., 2021). This tree component is
153 used to calculate assemblage level metrics, for example, age of assemblages, in-situ
154 diversification, the amount of phylogenetic diversity, and endemism that emerged as a process of
155 in-situ diversification in a region. The historical dispersal component corresponds to the
156 evolutionary history that arose due to events of ex-situ diversification and historical dispersal,
157 i.e., events that occurred before the arrival and establishment of a lineage in the assemblage in
158 which the present-day species are occurring.
159 In situ diversification and historical dispersal components are illustrated in Figure 2. In
160 this hypothetical example, Fig. 2a represents a result from an ancestral area reconstruction
161 model, with ancestral regions of occurrence represented by letters A, B, and C. These areas can
162 be interpreted, for example, as biomes (e.g., Maestri et al., 2019; Van Dijk et al., 2021a; Wiens
163 & Graham, 2005). The matrix in Fig. 2b represents the current area of occurrence for each
164 species in three assemblages (comm 1, comm 2, and comm 3). By reuniting in-situ
165 diversification and historical dispersal, we can decompose the amount of macroevolutionary
166 history that emerged inside a region due to in-situ diversification and the component of
167 phylogenetic history that came from another region. Following this rationale, we can notice that

community 1 is assembled only by in-situ diversification (Fig. 2c), community 2 by in-situ diversification and historical dispersal from region B (Fig. 2d), and community 3 is mainly assembled by a historical dispersal event from region B (Fig. 2e).

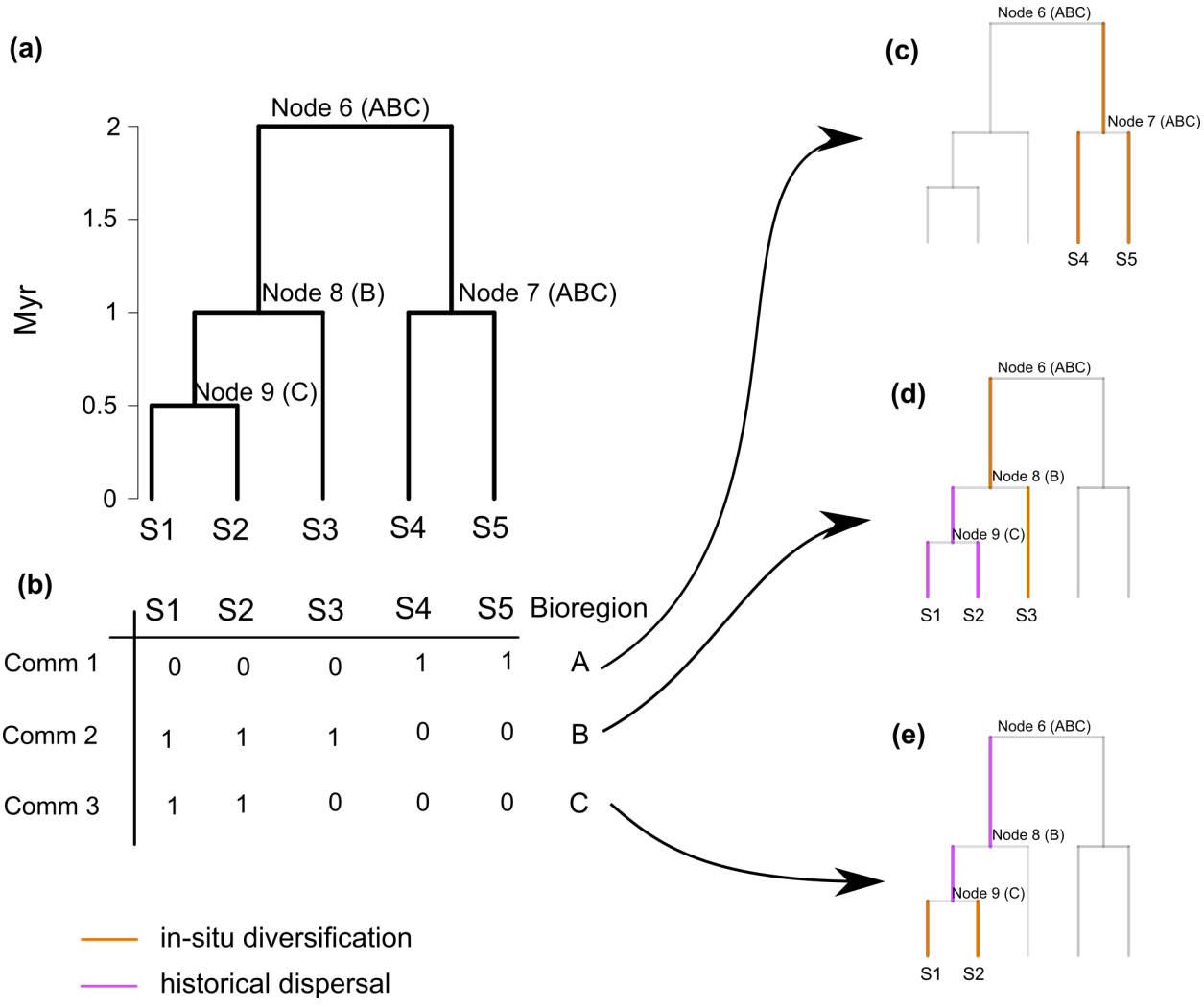


Figure 2: Schematic Figure illustrating the decomposition of macroevolutionary history dynamics performed by Herodotools. Purple branches in the tree correspond to the evolutionary history that emerged from historical dispersal events, and orange branches emerged from in-situ diversification. (a) represents a phylogenetic hypothesis with an ancestral area reconstruction (letters in each node); (b) illustrates the occurrence of species in each assemblage and the biome

of the assemblage; (c), (d), and (e) represents three different hypothetical scenarios of macroevolutionary history for each assemblage.

In the next section we explain in more detail the main functions present in Herodotools package and some specific functions that are only possible due to the integration between macroevolutionary information with assemblage data.

Phylogenetic regionalization and shifts in phylogenetic turnover across bioregions

Methods aiming to define biogeographic regions based on either taxonomic (Edler et al., 2016; Holt et al., 2013; Kreft & Jetz, 2010; Olivero et al., 2013; Vilhena & Antonelli, 2015) or phylogenetic relationships among species of a given biological group (Daru et al., 2020; Holt et al., 2013; Maestri & Duarte, 2020) has been intensively developed over the last decade, using different site resemblance and clustering methods. While all methods are valuable as classification tools for historical biogeography and evolutionary macroecology, bioregions defined from either species composition or the Simpson index of phylogenetic beta diversity (Holt, et al. 2013; Daru et al., 2020) might lead to the detection of evolutionarily unreal biogeographic regions, as regions arising from classifications might lack a coherent, shared history of diversification. It occurs because site resemblance and clustering methods neglect the detection of /can not identify transition zones, i.e., regions where sites show low phylogenetic affinity to their respective biogeographic regions (Maestri & Duarte 2020). On the other hand, classifying biogeographic regions based on evoregions (Maestri & Duarte 2020) enables mapping biogeographic transition zones in addition to core biogeographic regions, better showing intricate species distributions and facilitating the interpretation of biogeographic regions.

As an interesting development, evoregions also allow interpreting the historical development of each biogeographic region directly along with the diversification history of a lineage represented as a phylogenetic tree (e.g., Fig 2 in Duarte and Maestri, 2018). Thus, evoregions is a useful methodological approach for historical biogeography and evolutionary macroecology whenever unveiling the geographical history of diversification is a primary goal. Phylogenetic classification with evoregions can be performed using the function *evoregions()*, and detecting phylogenetic turnover zones can be done by using the function *affiliation_evoreg()* in Herodotools package.

Metrics for inference of historical processes at assemblage level

One of the main drawbacks in ecology and evolution is the integration of historical processes into the assemblage level (Mouquet et al., 2012). Herodotools fill this gap by implementing a set of metrics that can be calculated at the assemblage level, thereby showing historical processes at assemblage level. The first metric is the age of assemblage, explained in the previous section and calculated with the function *age_arrival()*. We also implemented tip-based metrics of diversification that account for macroevolutionary history. For example, the function *db_diversification()* modified the commonly used Diversification Rate metric (DR; Jetz et al., 2012) calculated as the inverse of the mean equal-splits measure (Redding & Mooers, 2006) as follows:

$$DR_i = \left(\sum_{j=1}^{N_i} l_j \frac{1}{2^{j-1}} \right) Equation\ 1$$

In our modification, instead of using all the paths from species *i* to the root, being l_j the length of the edge *j*, we used for calculation only the edges *j* that emerged after the arrival of species

lineage in the regions where the assemblage is placed. With this modification, we obtained a DR metric that accounts only for in-situ diversification.

We also implemented popular ecophylogenetic metrics, such as Phylogenetic Diversity (PD Faith, 1992) and Phylogenetic Endemism (Rosauer et al., 2009), that account for in-situ diversification by applying the same rationale. For PD and PE, we modified the original metrics by using only the branch lengths that emerged after the arrival and establishment of the species lineages in an assemblage's region. We then obtained what we called a 'diversification-based PD' and 'PE'.

Mapping trait evolution dynamics over space

To be possible to scale up macroevolution to a macroecological assemblage-based level of analysis, methods should provide species-specific data. A few existent metrics are designed to gather species-specific evolutionary data directly from phylogenies, including estimates of tip-based diversification (Jetz et al., 2012; Redding & Mooers, 2006; Title & Rabosky, 2019), and tip-based trait evolutionary rates (Castiglione et al., 2018). However, these metrics cannot handle temporal variation in trait states and age/time of trait appearance in the history of a clade/phylogeny. To tackle this issue, Luza et al., (2021) formulated an analytical framework that allows analyzing species-specific rates and tempo of (discrete) trait evolution by proposing three new tip-based metrics: i) transition rates, ii) stasis time, and the iii) last transition time. Briefly, these metrics capture the evolutionary history of trait changes from the root to each current species and summarize it in species-specific number of trait state changes (transition rates), the total evolutionary time without change (stasis time), and time since the last change (last transition time). Those metrics can be projected at the assemblage level, for example, by

simply averaging species “traits” within assemblages, whereby it is possible inferring ecological and historical processes shaping the rates and tempo of trait evolution in local assemblages. These three tip-based metrics can be calculated with the function *tip_based_trait_evo*.

Historical biogeography of Akodon genus

To demonstrate the functionalities of Herodotools, we analyzed a data set of 732 assemblages of the genus *Akodon*. *Akodon* is one of the most species-rich and widely distributed genera of mammals in the Neotropics (Patton et al., 2007, Mammal Diversity Database 2022). Geographically, its 41 described species form two hotspots of richness, one in the Atlantic Forest and the other in Central Andes, dominating the more inclusive richness pattern of its tribe, the Akodontini (Maestri & Patterson 2016), and overall forming a “dumbell” richness pattern (Pardiñas et al. 2015) also due to its absence in Amazonia. Such bimodal richness peaks and the phylogenetic distribution of its species cast doubt on the geographic origins of the genus, with hypotheses along the years lending support for either an Andean or an Atlantic center of origination and main diversification of the inclusive tribe (Reig, 1987; D’Elía & Pardiñas 2015; Maestri et al. 2019).

To calculate the importance of in-situ diversification, historical dispersal events and estimate the age of assemblages, we first applied a phylogenetic regionalization method based on evolutionary turnover (Maestri & Duarte, 2020) implemented in the function *evoregion()* of Herodotools package. Based on the groups generated by the phylogenetic regionalization, we estimated species' ancestral range using BioGeoBEARS (Matzke, 2013). We built six different models implemented in BioGeoBears: DIVA, DEC, and BayArea; each with and without a jump parameter. Details of model construction and the code used can be found in the online resource

(https://gabrielnakamura.github.io/Herodotools/articles/Intro_Herodotools_vignette.html). We allowed species to belong to up to three biomes. We performed a model selection using Akaike Information Criterion (AIC) to select the best model for ancestral range estimates. Three models (DEC, DEC+J, and BayArea) presented $\Delta AIC < 2$ and were considered equivalent. We chose the DEC model for further analysis because it has the lowest AIC value and has fewer model parameters. The selected model was then used in Herodotools to calculate metrics that represent historical processes at assemblage level. Specifically, we calculated: 1) the age of each assemblage as the mean age in which the ancestors of each present-day species arrived and established in the region of a given cell (function *age_arrival*); 2) In-situ diversification rates, obtained by calculating the tip-based rate of diversification for each species as being the inverse of equal splits metric (Jetz et al., 2012; Redding & Mooers, 2006), but now considering only the branches of the lineage that emerged from an in-situ diversification process (function *db_diversification*); 3) The contribution of historical dispersal events for each assemblage, represented as the percentage of species that dispersed from a focal ancestral range for all other regions. The assemblages comprised 1x1 grids (110 x 110 km around the Equator).

Finally, we analyzed the macroevolutionary dynamics of traits by calculating transition rates, as the number of times a character change over the evolution of a lineage. Stasis time represents the maximum time span in which the current character state of lineage was maintained in the evolutionary history, and the last transition time as being the sum of branch lengths from the tip to the prior/previous node with a reconstructed character equal to the current tip-character (Luza et al., 2021). The results of trait macroevolutionary dynamics are shown only in the online supplement (https://gabrielnakamura.github.io/Herodotools/articles/Intro_Herodotools_vignette.html), as

well as other examples illustrating the use of additional functions implemented in Herodotools package.

Results

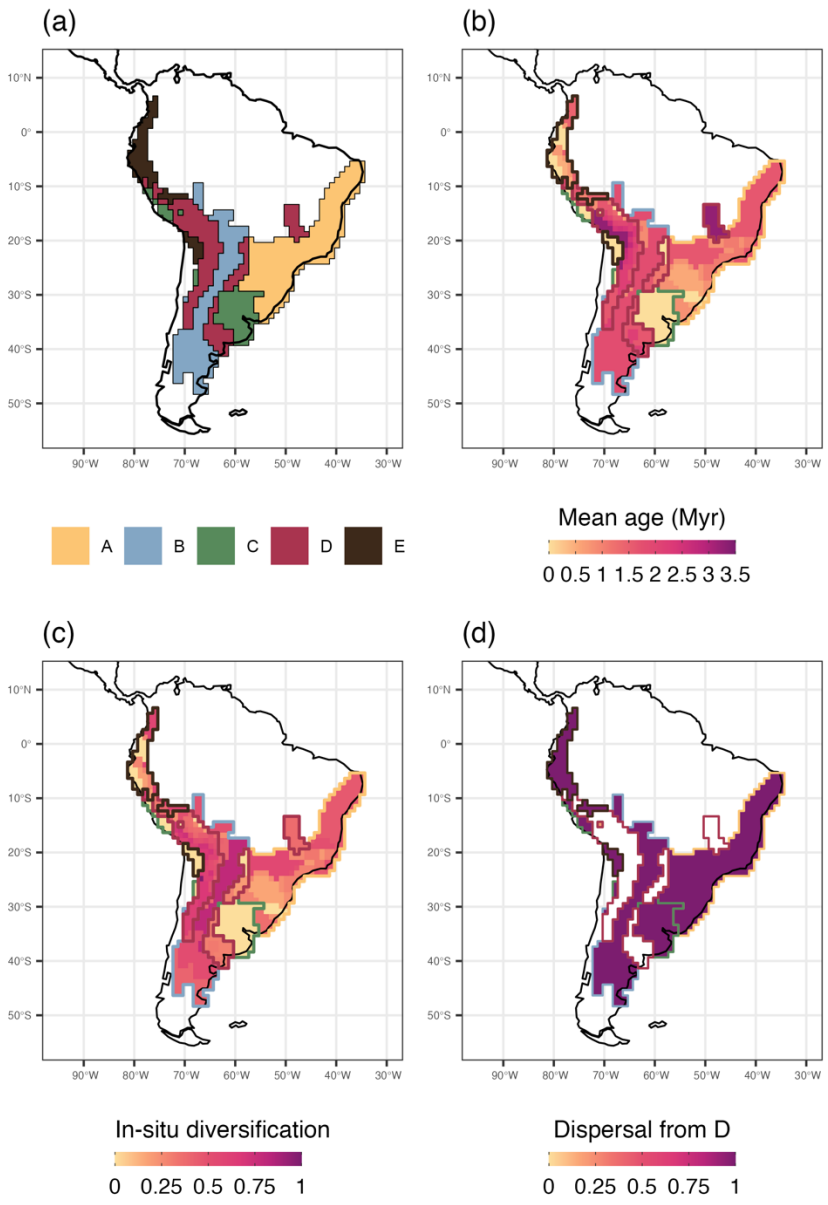
Regionalization with evoregion and assemblage ages

In our empirical example using the genus *Akodon*, Evoregion D captures a mix of species from the four species complexes within *Akodon*: *boliviensis*, *cursor*, *aerosus*, and *dolores*. We found components of all these complexes within Evoregion D, culminating in its central position where the richness peak is located, plus idiosyncrasies such as the occurrence of *A. lindberghi*, a species whose species group is not easily defined (Gonçalves et al. 2007; Jayat et al. 2010; Coyner et al. 2013). Members of the cursor group clearly define Evoregion A; evoregion C is mainly related to the group *boliviensis* plus *A. azarae*; evoregion E is primarily determined by members of the *aerosus* group; and evoregion B by members of the *dolores* group. Furthermore, regarding the affiliation of assemblages to each evoregion, we can observe that the lowest values of affiliation are found close to the boundaries of evoregions. Furthermore, evoregion D and the south portion of evoregion A presented the lowest affiliation values (i.e., when a cell has a low chance of belonging to the region in which it was classified), indicating zones of high phylogenetic turnover and multiple colonization events (Supplementary online material).

Assemblage level metrics – Age, diversification, and model-based diversification

Our estimates of assemblage age indicates that assemblages did not present high variation in the age in which ancestors arrived and colonized the assemblages, except for evoregions C and E

314 that presented the most recent assemblages. On the other hand, ancient assemblages are within
315 regions B and D (Fig. 3b).
316 The diversification and the model-based diversification metrics showed similar spatial
317 patterns, with higher values of in situ diversification in the evoregion B, with some assemblages
318 presenting almost all the diversification occurring inside this region (Fig. 3c). On the other hand,
319 evoregion C was the one that presented assemblages with lowest values of in-situ diversification.
320 Together, age of assemblages and in-situ diversification patterns reflected the explosive
321 diversification in a few million years of *Akodon* assemblages.



322

323 Figure 3: Spatial representation of evoregions (a), and historical variables (b-d). (b) represents
324 the age of assemblages, (c) the in-situ diversification as a proportion of the total diversification
325 (calculated as the DR metric) and (d) represents the proportion of contribution of region D with
326 lineages for all other evoregions.

327

Historical dispersal patterns

Our analysis of historical dispersal showed that evoregion D was the region that most contributed with lineages dispersal for other evoregions. Assemblages in evoregions A, B, C and E were almost entirely constituted of lineages from evoregion D (Fig. 3d). On the other hand, evoregions A and B presented only a little contribution regarding lineage dispersal. The contribution of other regions for lineage dispersal events are shown in the online Supplementary material (https://gabrielnakamura.github.io/Herodotools/articles/Intro_Herodotools_vignette.html).

Discussion

Herodotools fills the gap between macroevolution and ecology by providing a computational infrastructure of analysis that allows scaling historical variables at community and assemblage levels (Mouquet et al., 2012). Our package allows investigating the importance of historical processes acting on ecological communities at finer spatial grains than the usual approaches focusing on entire bioregions or single lineages. Our package provides a direct quantification of the effects of in-situ diversification and historical dispersal through the integration of macroevolutionary models of ancestral state reconstruction (area and traits) in ecophyogenetics (PD, PE) and other assemblage/community metrics like age, and trait dynamics metrics, what allows moving ecophylogenetics from the pattern-to-process inference to a more mechanistic approach.

Other approaches, like DAMOCLES (Pigot & Etienne, 2009), also allows to investigate the macroevolutionary dynamics at community level, however it relies on a null model that incorporate historical processes for testing whether there is non-randomness in community phylogenetic structure. The methods presented in Herodotools differ from DAMOCLES since

the former aims to decompose the phylogenetic metrics in components that indicates two opposite processes, in situ diversification and historical dispersal. In addition, it is worth to mention the behavior of our metric of assemblage age compared to other existing approaches. Estimates of age for bioregions are of great value to test historical competing hypotheses in macroecology, like out of the tropics (OTT) and tropical niche conservatism (Wiens & Donoghue, 2004b). Consequently, it is important to derive metrics that reliably reflect these characteristics of assemblages. Our proposition of age differs from previous studies (Wiens et al., 2011, 2011; Wiens & Donoghue, 2004b) by allowing the calculation of age for each lineage (tip-based) from an assemblage-based perspective. This means that a species can have different times of colonization depending on the assemblage and region where it occurs. This characteristic also applies for other metrics of Herodotools such as DR, PD and PE. Furthermore, instead of providing a single age for an entire region, in our proposition the age is variable within a region, since each assemblage might present different species composition. One possible topic of investigation for future studies is the implications of different age estimates to explain macroecological patterns (Wiens et al., 2011; Wiens & Graham, 2005).

The metrics implemented in Herodotools can also be used in common hypothesis testing frameworks as, for example, linear models relating age and diversity, or diversification and biodiversity at assemblage level. Regarding our empirical example, evoregions for the genus *Akodon* depict the differential geographic distribution of its internal monophyletic groups (or subclades) according to the phylogeny used here (Maestri et al., 2017; Upham et al., 2019), which also closely match other phylogenetic propositions, such as the distribution of the four monophyletic species groups / complexes proposed for the genus, the *aerosus*, *boliviensis*, *cursor* and *dolores* species groups (Coyner et al. 2013). These species groups are not undisputed

(see Jayat et al. 2010, Pardiñas et al. 2015) but the overall pattern for the genus' bioregions is likely to be similar under different species groups and phylogenetic propositions if the backbone phylogeny is proven to be similar. Unsurprisingly, dispersal from evoregion D to others was found to be the most prominent compared to dispersal deriving from other evoregions, given the very nature of evoregion D as a high-richness region whose members are found phylogenetically widespread. Generally, the empirical analysis illustrates how Herodotools can be used to investigate questions that are in the intersection of macroevolution and macroecology (McGill et al., 2019), since our package allows to obtain variables that represent historical components of macroevolutionary dynamics (in-situ diversification, historical dispersal, age) at the assemblage scales.

Future improvements consist in implementation of functions that allows to handle with and integrate macroevolutionary models from other popular programs like RevBayes (Landis et al., 2013). We envision constantly improvements in package to make the integration among macroevolutionary models and assemblage data easier and straightforward, in a way that Herodotools can work as the main toolkit for researchers in macroevolution and macroecology to be integrated in the same endeavor: to disentangle the ecological and evolutionary processes creating and maintaining biodiversity.

Data availability statement

All data used in this work is publicly available at https://github.com/GabrielNakamura/Herodotools/tree/main/inst/extdata.

References

Bollback, J.P. SIMMAP: Stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* **7,** 88 (2006). https://doi.org/10.1186/1471-2105-7-88

Castiglione, S., Tesone, G., Piccolo, M., Melchionna, M., Mondanaro, A., Serio, C., Di Febbraro, M., & Raia, P. (2018). A new method for testing evolutionary rate variation and shifts in phenotypic evolution. *Methods in Ecology and Evolution*, *9*(4), 974–983. https://doi.org/10.1111/2041-210X.12954

Cavender-Bares, J., Kozak, K. H., Fine, P. V. a, & Kembel, S. W. (2009). The merging of community ecology and phylogenetic biology. *Ecology Letters*, *12*, 693–715. https://doi.org/10.1111/j.1461-0248.2009.01314.x

Crouch, N. M. A., Capurucho, J. M. G., Hackett, S. J., & Bates, J. M. (2019). Evaluating the contribution of dispersal to community structure in Neotropical passerine birds. *Ecography*, *42*(2), 390–399. https://doi.org/10.1111/ecog.03927

Daru, B. H., Karunarathne, P., & Schliep, K. (2020). phyloregion: R package for biogeographical regionalization and macroecology. *Methods in Ecology and Evolution*, *11*(11), 1483–1491. https://doi.org/10.1111/2041-210X.13478

Edler, D., Guedes, T., Zizka, A., Rosvall, M., & Antonelli, A. (2016). Infomap Bioregions: Interactive Mapping of Biogeographical Regions from Species Distributions. *Systematic Biology*, syw087. https://doi.org/10.1093/sysbio/syw087

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation*, *61*, 1–10. https://doi.org/10.1016/0006-3207(92)91201-3

Gerhold, P., Cahill, J. F., Winter, M., Bartish, I. V., & Prinzing, A. (2015). Phylogenetic patterns are not proxies of community assembly mechanisms (they are far better). *Functional Ecology*, *29*(5), 600–614. https://doi.org/10.1111/1365-2435.12425

Holt, B. G., Lessard, J.-P., Borregaard, M. K., Fritz, S. A., Araújo, M. B., Dimitrov, D., Fabre, P.-H., Graham, C. H., Graves, G. R., Jønsson, K. A., Nogués-Bravo, D., Wang, Z., Whittaker, R. J., Fjeldså, J., & Rahbek, C. (2013). An Update of Wallace's Zoogeographic Regions of the World. *Science*, *339*(6115), 74–78. https://doi.org/10.1126/science.1228282

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, *491*(7424), 444–448. https://doi.org/10.1038/nature11631

Kreft, H., & Jetz, W. (2010). A framework for delineating biogeographical regions based on species distributions: Global quantitative biogeographical regionalizations. *Journal of Biogeography*, *37*(11), 2029–2053. https://doi.org/10.1111/j.1365-2699.2010.02375.x

Landis, M. J., Matzke, N. J., Moore, B. R., & Huelsenbeck, J. P. (2013). Bayesian analysis of biogeography when the number of areas is large. *Systematic Biology*, *62*(6), 789–804. https://doi.org/10.1093/sysbio/syt040

Luza, A. L., Maestri, R., Debastiani, V. J., Patterson, B. D., Maria, S., & Leandro, H. (2021). Is evolution faster at ecotones? A test using rates and tempo of diet transitions in Neotropical Sigmodontinae (Rodentia , Cricetidae). *Ecology and Evolution*, *December*, 18676–18690. https://doi.org/10.1002/ece3.8476

Maestri, R., & Duarte, L. D. S. (2020). Evoregions: Mapping shifts in phylogenetic turnover across biogeographic regions Renan Maestri. *Methods in Ecology and Evolution*, *2020*(August), 1652–1662. https://doi.org/10.1111/2041-210X.13492

Maestri, R., Monteiro, L. R., Fornel, R., Upham, N. S., Patterson, B. D., & de Freitas, T. R. O. (2017). The ecology of a continental evolutionary radiation: Is the radiation of sigmodontine rodents adaptive? *Evolution*, *71*(3), 610–632. https://doi.org/10.1111/evo.13155

Maestri, R., Upham, N. S., & Patterson, B. D. (2019a). Tracing the diversification history of a Neogene rodent invasion into South America. *Ecography*, *42*(4), 683–695. https://doi.org/10.1111/ecog.04102

Maestri, R., Upham, N. S., & Patterson, B. D. (2019b). Tracing the diversification history of a Neogene rodent invasion into South America. *Ecography*, *42*(4), 683–695. https://doi.org/10.1111/ecog.04102

Matzke, N. J. (2013). Probabilistic historical biogeography: New models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography*, *5*(4). https://doi.org/10.21425/f5fbg19694

McGill, B. J., Chase, J. M., Hortal, J., Overcast, I., Rominger, A. J., Rosindell, J., Borges, P. A. V., Emerson, B. C., Etienne, R. S., Hickerson, M. J., Mahler, D. L., Massol, F., McGaughran, A., Neves, P., Parent, C., Patiño, J., Ruffley, M., Wagner, C. E., & Gillespie, R. (2019). Unifying macroecology and macroevolution to answer fundamental questions about biodiversity. *Global Ecology and Biogeography*, *28*(12), 1925–1936. https://doi.org/10.1111/geb.13020

Mouquet, N., Devictor, V., Meynard, C. N., Munoz, F., Couteron, P., Dalecky, A., Gravel, D., Hardy, O. J., Jabot, F., Prinzing, A., Rodrigues, A. S. L., Rohr, R. P., & Th, E. (2012). Ecophylogenetics: Advances and perspectives. *Biological Reviews*, *87*, 769–785. https://doi.org/10.1111/j.1469-185X.2012.00224.x

Münkemüller, T., Gallien, L., Pollock, L. J., Barros, C., Carboni, M., Chalmandrier, L., Mazel, F., Mokany, K., Roquet, C., Smyčka, J., Talluto, M. V., & Thuiller, W. (2020). Dos and don ' ts when inferring assembly rules from diversity patterns. *Global Ecology and Biogeography*, *July 2019*, 1212–1229. https://doi.org/10.1111/geb.13098

Olivero, J., Márquez, A. L., & Real, R. (2013). Integrating Fuzzy Logic and Statistics to Improve the Reliable Delimitation of Biogeographic Regions and Transition Zones. *Systematic Biology*, *62*(1), 1–21. https://doi.org/10.1093/sysbio/sys061

Patton, J. L., Pardiñas, U. F. J., & D'Eli?a, G. (2007). *Mammals of South America*. University of Chicago press.

Redding, D. W., & Mooers, A. O. (2006). Incorporating evolutionary measures into conservation prioritization. *Conservation Biology*, *20*(6), 1670–1678. https://doi.org/10.1111/j.1523-1739.2006.00555.x

Ricklefs, R. E. (1987). Community Diversity: Relative Roles of Local and Regional Processes. *Science*, *235*(4785), 167–171. https://doi.org/10.1126/science.235.4785.167

Ricklefs, R. E., & Jenkins, D. G. (2011). Biogeography and ecology: Towards the integration of two disciplines. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1576), 2438–2448. https://doi.org/10.1098/rstb.2011.0066

Rosauer, D., Laffan, S. W., Crisp, M. D., Donnellan, S. C., & Cook, L. G. (2009). Phylogenetic endemism: A new approach for identifying geographical concentrations of evolutionary

history. *Molecular Ecology*, *18*(19), 4061–4072. https://doi.org/10.1111/j.1365-294X.2009.04311.x

Title, P. O., & Rabosky, D. L. (2019). Tip rates, phylogenies and diversification: What are we estimating, and how good are the estimates? *Methods in Ecology and Evolution*, *10*(6), 821–834. https://doi.org/10.1111/2041-210X.13153

Upham, N. S., Esselstyn, J. A., & Jetz, W. (2019). Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLOS Biology*, *17*(12), e3000494. https://doi.org/10.1371/journal.pbio.3000494

Van Dijk, A., Nakamura, G., Rodrigues, A. V., Maestri, R., & Duarte, L. (2021). Imprints of tropical niche conservatism and historical dispersal in the radiation of Tyrannidae (Aves: Passeriformes). *Biological Journal of the Linnean Society*, *134*(1), 57–67. https://doi.org/10.1093/biolinnean/blab079

Vilhena, D. A., & Antonelli, A. (2015). A network approach for identifying and delimiting biogeographical regions. *Nature Communications*, *6*. https://doi.org/10.1038/ncomms7848

Wiens, J. J., & Donoghue, M. J. (2004a). Historical biogeography, ecology and species richness. *Trends in Ecology and Evolution*, *19*(12), 639–644. https://doi.org/10.1016/j.tree.2004.09.011

Wiens, J. J., & Donoghue, M. J. (2004b). Historical biogeography, ecology and species richness. *Trends in Ecology and Evolution*, *19*(12), 639–644. https://doi.org/10.1016/j.tree.2004.09.011

Wiens, J. J., & Graham, C. H. (2005). Niche Conservatism: Integrating Evolution, Ecology, and Conservation Biology. *Annual Review of Ecology, Evolution, and Systematics*, *36*(36), 519–539. https://doi.org/10.1146/annurev.ecolsys.36.102803.095431

Wiens, J. J., Pyron, R. A., & Moen, D. S. (2011). Phylogenetic origins of local-scale diversity patterns and the causes of Amazonian megadiversity: Phylogeny and local richness. *Ecology Letters*, *14*(7), 643–652. https://doi.org/10.1111/j.1461-0248.2011.01625.x