# The relative roles of in situ diversification and lineage dispersal underlying diversity patterns at the assemblage level

Arthur Vinicius Rodrigues[1,2], Gabriel Nakamura[2,3], and Leandro Duarte[2]

[1] Research Centre for Ecological Change, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland.

[2] Programa de Pós-Graduação em Ecologia, Departamento de Ecologia, Universidade Federal do Rio Grande do Sul, Brazil

[3] Universidade de São Paulo, Instituto de Biociencias, Brazil

Corresponding author: Arthur Vinicius Rodrigues (rodrigues.arthur.v@gmail.com)

## Abstract

Speciation, extinction, and dispersal are the historical processes influencing the spatial distribution of lineages and strongly influence diversity patterns. Here, we apply a recently developed methodological approach to quantify the relative legacies in situ diversification history (i.e. diversification occurring in the biogeographical region) and historical dispersal (inferred from ex-situ diversification) on current diversity patterns of the plant genus *Myrcia* (Myrtaceae) in assemblages distributed across the Brazilian Atlantic Forest. To distinguish among these processes, we proposed a framework that characterized the assemblages based on the mean colonization age of the assemblages, phylogenetic structure, species richness and proportion of in situ diversification. Our results suggest that environmental dynamics have played important role in shaping of diversity. In the Southern Atlantic Forest, environmentally stable areas appear to have served as cradles for cold-adapted lineages. In the central region, environmental instability in this montane area seems to be a unstable area acting as cradle, with assemblages showing moderate to high in situ diversification and intermediate colonization ages. In the Northern Atlantic Forest, older and richer assemblages with high phylogenetic diversity suggest climatic stability and a refugium role. In contrast, the Central-West region, with younger colonization ages and a higher proportion of lineage dispersal, appears to act as an unstable sink for lineages. By combining community phylogenetics and diversification metrics, we infer evolutionary legacies at the assemblage level and disentangle the roles of in situ diversification and lineage dispersal. In some regions, particularly outside Evoregion A, lineage dispersal appears more relevant, possibly due to weaker selection pressures. In contrast, colder areas may have favored lineages with broader climatic tolerance. Our findings reveal distinct biogeographic dynamics across the Atlantic Forest, including areas acting as cradles, sinks, and refugia, and demonstrate the value of assemblage-level diversification metrics in

35 understanding within-biome evolutionary processes and their influence on current diversity
36 patterns.

37 **Introduction**

38 The historical processes driving the diversity and distribution of biological diversity
39 across the geographic space are, ultimately, speciation, extinction, and dispersal (Ricklefs
40 1987; Wiens and Donoghue 2004). Environmentally-driven ecological dynamics might also
41 influence each of these processes by promoting ecological drift and selection-driven
42 speciation along the evolutionary time, causing extinction (local or globally), allowing
43 dispersal (colonization) and secondary sympatry (Mittelbach and Schemske 2015).
44 Altogether, these processes contribute to determine the distribution of species and the
45 spatial diversity patterns, placing the evolutionary history of a clade as an important piece
46 to explain its current geographic distribution (Cavender-Bares et al. 2016; Gerhold et al.
47 2018).

48 In the context of historical processes shaping biodiversity, the stability (or
49 instability) of a region can influence diversification and lineage dispersal, thereby affecting
50 present-day diversity patterns. In environmentally unstable regions with heterogeneous
51 landscapes, populations are more likely to have their distributions fragmented across space,
52 which increases the chances of both speciation and extinction. Mountainous regions are
53 good examples of such areas, and it can act as both cradles and museums of biodiversity
54 (Rahbek et al. 2019; Vasconcelos et al. 2020; Vasconcelos, O'Meara, and Beaulieu 2022).
55 Cradle areas can be harder to detect because environmental instability drives both high
56 rates of speciation and extinction, leading to high species turnover leaving fragmentary
57 evidence of past dynamics (Vasconcelos et al. 2022). Museums areas, in turn, are easier to
58 recognize since they environmental stability allows the persistence of ancient lineages,
59 though this same stability can also give the false impression that they were centers of origin
60 (Vasconcelos et al. 2022). As stability allows species dispersing from to adjacent areas to
61 prevent extinction, speciation rates are expected to be lower as well as extinction rates.
62 Over time, this source-sink dynamics between unstable and stable areas could generate
63 present-day assemblages with high phylogenetic diversity, moderate to low proportion of in
64 situ diversification, consequently, higher lineage dispersal and old colonization age of
65 assemblages on stable areas (hereafter age of assemblages, for simplicity), the so called
66 museums. On the other hand, unstable areas should present more recent colonization
67 events (low assemblage ages), due to high levels of both speciation and extinction rates,

moderate to high proportion of in situ diversification and high species richness where environmental heterogeneity is high. Stable areas can also act as sources of lineages. In these regions, we can expect high in situ diversification, with varying levels of phylogenetic relatedness and relatively recent colonization events (i.e., older assemblages). Therefore, a comprehensive investigation of historical processes shaping present-day diversity patterns should integrate community phylogenetics with diversification analyses and colonization/dispersal timing. These expectations are summarized in Table 1.

**Table 1**. Scenarios and the expected patterns that can be detected using assemblage-level metrics to assess historical biogeographical processes and current phylogenetic (expressed as MPD) and taxonomic alpha diversity (expressed as species richness).

| Scenario | Biogeographical process | | Current diversity | |
|---|---|---|---|---|
| | Proportion of in situ Diversification | Colonization age of assemblage | MPD | Species richness |
| Stable area acting as refugium | Moderate to low | Old | High | High |
| Unstable area acting as a cradle | Moderate to high | Moderate to young | Low | High |
| Stable area acting as cradle | High | Old | Either low or high | Intermediate |
| Unstable area acting as sink of lineages | Low | Young | Either low or high | Low |

*Note: MPD is mean pairwise phylogenetic distance.*

Species-level diversification metrics, species age estimates (or colonization time estimates), and ancestral area reconstruction have been widely used to uncover the historical biogeographic processes underlying current diversity patterns (Pinto-Ledezma et al. 2019; Velasco and Pinto-Ledezma 2022; Villalobos et al. 2020; Nakamura et al. 2024). Recently, a new method was proposed to integrate ancestral area reconstruction (Matzke 2013; Ree and Smith 2008), which estimates the biogeographical history of a clade across regions, with diversification and colonization age estimation, commonly used to describe evolutionary patterns at the assemblage level (McGill et al. 2019; Velasco and Pinto-Ledezma 2022). Combining these approaches allows quantifying the relative role of in situ diversification (diversification occurred within a biogeographical region) and lineage

dispersal (diversification occurred in another biogeographical region, or ex situ) for the species assemblage (Nakamura et al. 2024; see Box 1 and Fig. 1).

The original method relies on estimating the most likely ancestral range at the nodes of the phylogeny and does not account for uncertainty in ancestral estimates or for range shifts along branches (anagenetic range evolution) (Nakamura et al. 2024). In this study, we address these limitations by using Biogeographical Stochastic Mapping (BSM) to estimate ancestral area (Dupin et al. 2017). BSM allows propagation of uncertainty in ancestral range area estimates into the calculation of in situ diversification and colonization age of assemblages. Additionally, both anagenetic and cladogenetic range-shift processes are incorporated, unlike the original method, which considered only cladogenetic shifts, resulting in a more reliable assemblage-level metric estimates.

Our aim is to investigate the roles of in situ diversification and lineage dispersal in shaping the species assemblages in the Atlantic Forest in South America (Figure 2). Atlantic Forest harbors high species diversity and present turnover in taxonomic and phylogenetic composition that are consistent among different taxonomic groups (Carnaval and Moritz 2008; Costa 2003; Fiaschi and Pirani 2009). The major composition turnover distinction is between Northern and Southern portions of the Atlantic Forest (but see Brown et al., 2020), which has been attributed to current and historical differences in climatic conditions between regions (Carnaval et al. 2014; Saiter et al. 2016). Climatic stability since the last interglacial period (ca. 120 ky) seems to better explain the phylogenetic endemism in the northern portion, while the current climate seems to better explain the phylogenetic endemism of the southern portion (Carnaval et al. 2014).

The combination of high species diversity with the legacy of climatic influence makes Atlantic Forest a good system to investigate the interplay of historical processes acting on current macroecological and diversity patterns. As a model group we used *Myrcia* (Myrtaceae), which is a genus of trees and shrubs. *Myrcia* present several characteristics that make it suitable as a model group for the Atlantic Forest (Lucas and Bünger 2015; Murray-Smith et al. 2009). *Myrcia* has high diversity in Atlantic Forest (Lucas et al. 2018; Rodrigues and Duarte 2024) and is an important component of tree assemblages across the Atlantic Forest (Bergamin et al. 2021; Lucas et al. 2018; Oliveira-Filho and Fontes 2000; Rodrigues and Duarte 2024). Furthermore, the evolutionary history of *Myrcia* is linked to the Atlantic Forest history (Amorim et al. 2019; Rodrigues and Duarte 2024; Santos et al. 2017). Phylogenetic evidence suggests that *Myrcia* originated in the Atlantic Forest, probably in the

Serra do Mar mountains range (Amorim et al. 2019; Santos et al. 2017). However, a recent study (Rodrigues and Duarte 2024), presents a probable origin of the group in the northern Atlantic Forest, while the southern Atlantic Forest and adjacent highlands harbor the most recent diversification of several *Myrcia* lineages within the biome. Our study builds upon the work of Rodrigues and Duarte (2024) and Nakamura et al. (2024) by proposing a general framework that uses both methods from historical biogeography (Nakamura et al, 2024) with phylogenetic metrics from community ecology to shed light on historical processes acting on present-day assemblage diveristy.

**BOX 1**

Species-level phylogenetic metrics, such as diversification and species colonization age, are calculated using only phylogenetic information (topology and divergence ages). However, when these metrics are summarized at assemblage-level, they fail to account for the biogeographical context of species origination and dispersal. One way to address this limitation is to integrate these metrics with ancestral area estimates. By incorporating information about the geography of speciation, we can refine phylogenetic metrics to reflect the biogeographical history of a lineage.

Using this approach, we can quantify the portion of tip's diversification rate that occurred within a region (in situ diversification). Since our focus is on historical process affecting local assemblages, we attribute the *ex situ* diversification (when the ancestors of a species have diversified in another region) to the lineage dispersal process because the lineage had to disperse at some point to the assemblage before starting to diversify in situ. In addition, we can quantify for how long in the evolutionary history the species is present in a region (colonization age). This framework allows us to disentangle the contributions of in situ diversification (when a lineage diversified within the assemblage's biogeographical region) and lineage dispersal (ex situ diversification). A key advantage of this approach is that tip-based metrics can be aggregated at the assemblage level, enabling the spatial quantification of in situ diversification and its impact on phylogenetic metrics, such as diversification rates or phylogenetic endemism, across large spatial scales.

Figure 1 illustrates our approach, showing how a phylogenetic tree can be decomposed into in situ diversification and lineage dispersal components using a hypothetical example. In this example, a clade with 7 species occurs exclusively across two regions, West and East regions (Figure 1A), where there are four sites with species composition. The evolutionary history of the clade was reconstructed using

biogeographical stochastic mapping and DEC model. For simplicity, we considered only one region occupied by each species currently, the region with more sites occupied and show one realization of all possible stochastic maps (Figure 1B). The phylogenetic tree in Figure 1C shows the evolutionary history only for the species occurring in the assemblage. To separate the importance of the evolutionary history for the local assembly into two components - in situ diversification and lineage dispersal - we trace the path from each tip to the root of the phylogeny. In situ diversification is defined as the continuous portion of the phylogenetic tree connecting a tip species to its most recent ancestor occurring in the same region as the site in which the species is found (occupied sites are represented by a color filled square in Figure 1B). In the site 1, all species had part of their history in situ. In the site 2, however, two species had no in situ component of its history in West region. These make recent historical dispersal more important for the assemblage of site 2 than for that of site 1. It is important to note that the interpretation of evolutionary process for a species is dependent on the site's biogeographic region. In other words, a species has higher diversification rate (or colonization age) in a site within the region where it diversified. Consider sites 2 and 3, as an example. Site 2 is in the West region, where species *b* and *c* had part of their evolutionary history in the region, while species *a* and *g* diversified entirely in the East region. Site 3 contains the same species as site 2 but it is located in a different region. As a result, the partitioning of species history changes: species *a* and *g* are now classified as having in situ diversification, whereas species *b* and *c* are considered as recent dispersers, since their lineages evolved in the West region. Thus, in situ diversification plays a greater role in the assemblage of site 3 than in that of 2. Finally, in site 4 (East region) all species shared the evolutionary history within the region, making in situ diversification the sole evolutionary process determining assemblage.

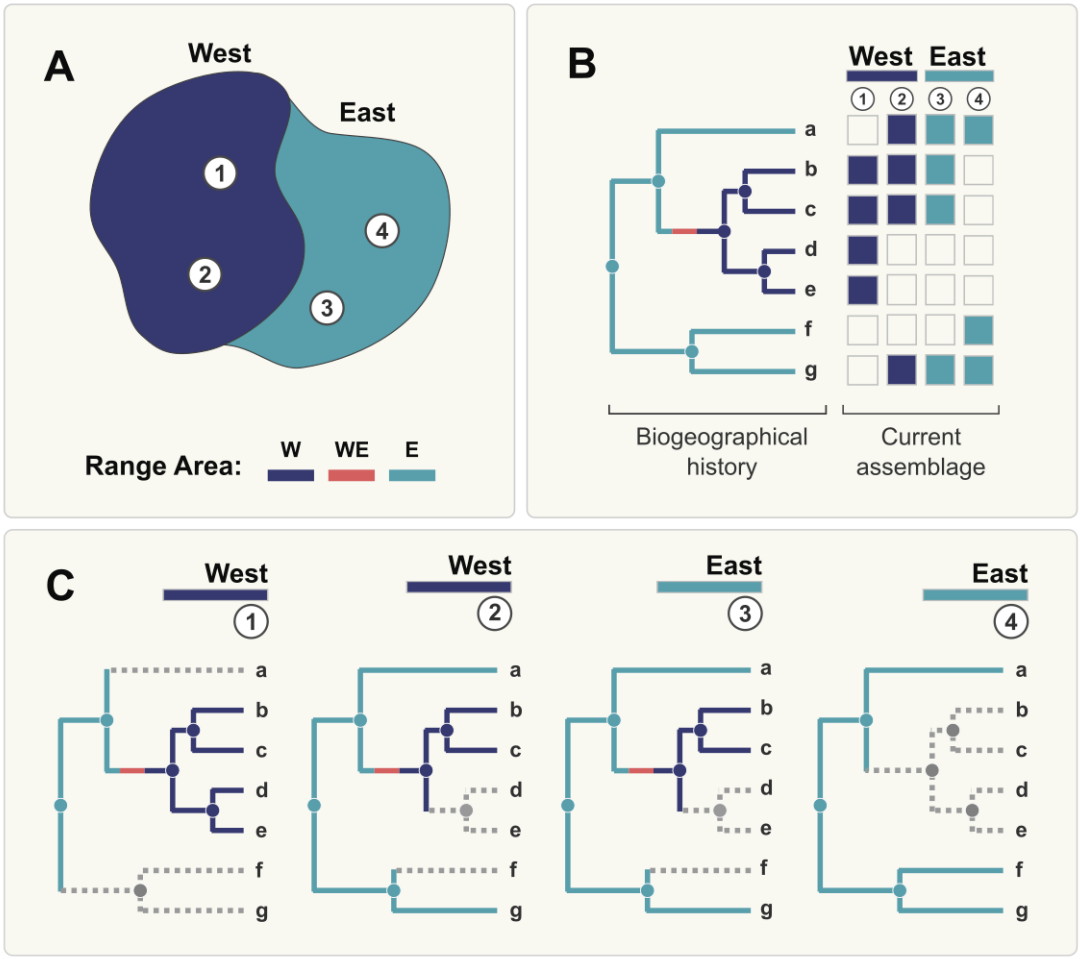



**Figure 1.** Schematic representation of the framework to quantify the relative role of in situ diversification processes and lineage dispersal in current assemblages. Panel A shows the assemblages at the West and East regions (dark blue and light blue, respectively). Panel B shows the biogeographical history of the clade, represented by the colors (regions) along the branches and nodes of the phylogenetic tree; and the current occurrence of species in assemblages is represented in the matrix of squares in which filled squares represent the presence of a species in the assemblage. Finally, in panel C, we highlight the biogeographical history of the species present in the assemblage; dotted branches are species not occurring in the assemblage. The delimitation of in situ processes is done following the path of the lineage from the tip to the root of the phylogeny, while the area estimated for an ancestral node is required to be within the same region of the current site of the assemblage. Note that the species could have different timing of in situ diversification depending on the site's region.

## Material and Methods

We used a methodological framework that integrates biogeographical history of the lineages estimated by ancestral area reconstruction and assemblage-level phylogenetic metrics (in situ diversification, colonization age of assemblages and mean phylogenetic

196 pairwise distance) to provide assemblage-level descriptors that captures the biogeographic
197 history of current assemblages (Nakamura et al. 2024). Here we extend the original
198 framework described by Nakamura et al. (2024) by using biogeographic stochastic mapping
199 (Dupin et al. 2017). Biogeographic Stochastic Mapping (BSM) is a method that samples
200 from the distribution of possible biogeographical histories (100 samples in this study),
201 estimated by a model and conditioned on model parameters, the phylogenetic tree, and the
202 data. This approach allows quantification of the uncertainty associated with reconstructing
203 the biogeographical history of species. Importantly, it enables us to propagate uncertainty
204 in ancestral area estimates and the assemblage-level phylogenetic metrics we use.

205 We applied BSM to reconstruct the biogeographical history of the genus *Myrcia* using a
206 phylogenetic tree. We also use the information on current species distribution and the
207 biogeographical region in which the assemblages occurs. These data were combined to
208 calculate the proportional in situ diversification rate ($DR_{prop}$), colonization age of
209 assemblage, mean phylogenetic distance and species richness. Using these metrics, we
210 investigated the relative roles of in situ diversification and lineage dispersal in shaping
211 *Myrcia* assemblages in the Atlantic Forest, following the framework illustrated in Table 1.
212 All analyses were conducted in R version 4.2.2 (R Core Team 2022) and the main packages
213 used are mentioned in *italic*. To define the boundary of the Atlantic Forest (our study area),
214 we used the integrative limit proposed by Muylaert et al. (2018).

215 *Genus Myrcia and phylogenetic tree*

216 The genus *Myrcia* comprise 794 species distributed across the Neotropics, subdivided in 10
217 taxonomic clades (hereafter sections) (Lucas et al. 2018; POWO 2025). A total of 264
218 Myrcia species are found in the Atlantic Forest, the region with the highest *Myrcia* richness
219 (Amorim et al. 2019; Lucas et al. 2018). The phylogeny was produced by Amorim et al.
220 (2019) using internal and external transcribed spacer (ITS and ETS) of the ribosomal
221 nuclear region and seven plastid markers. It presents strong statistical support for the
222 topology of the 10 sections (Amorim et al. 2019). Time-calibration used two calibration
223 points one to constrain the maximum age (the age of Neotropical tribe Myrteae) and
224 another to the *Myrcia* crown node (Vasconcelos et al. 2020). Here, we used the consensus
225 time-calibrated phylogenetic tree available in Vasconcelos et al. (2020). The available
226 phylogenetic tree has 195 taxons (24,5% of *Myrcia*'s species) (Amorim et al. 2019;
227 Vasconcelos et al. 2020).

228 *Species assemblages*

The species assemblages used here are based on binary predictions obtained from models of stacked species distribution across the Neotropics produced by Rodrigues and Duarte (2024). We used raster for 307 *Myrcia* species available (38.6% of *Myrcia*'s species) with resolution of 0.5 decimal degrees. Species distributions were constructed using buffers (50 km radius for n <4 records, convex-hull with 50 km buffer for 4-5 records) and species distribution models using Maxent algorithm (n < 5 records) (Phillips et al. 2017). Maxent models were fitted using a tuning approach and six environmental variables: four climatic variables (annual mean temperature, temperature seasonality, precipitation of wettest and driest quarters) and two soil variables (cation exchange capacity and clay percentage). The binarization process employed a site-specific threshold approach (Scherrer et al. 2018), where model predictions were converted to presence/absence by first estimating site richness through summing species probabilities, then ranking species by probability and selecting the highest-ranked species up to the estimated richness (Rodrigues and Duarte 2024). Here we use species distribution predictions for 96 species (12% of *Myrcia*'s Species, representing the 10 sections) that are also present in the phylogenetic tree. Although the number of species is low related to the genus richness, a comparison of the richness using all the species with distribution maps and the ones present in the phylogeny shows a high correlation (rho = 0.958) and the visual pattern is very similar across the Neotropical region, especially in the Atlantic Forest, which is the focus of this study (Figure S1).

*Biogeographic regions and ancestral area reconstructions of Myrcia*

Biogeographical regionalization was based on evoregions, a regionalization method that uses phylogenetic composition of *Myrcia* across the Neotropics (Rodrigues and Duarte 2024). This regionalization divides the area occupied by the genus in the Neotropics in five regions. In the Atlantic Forest there are 2 main regions (Figure 2D), the evoregion A covering the southern part of Atlantic Forest and the mountainous area, and the evoregion B covering the northern part of Atlantic Forest and low land areas. In the extreme north of Atlantic Forest, some grid cells are defined as part of the evoregion C. This region extends mostly outside the Atlantic Forest, Neotropics, ranging from Central America, Andes and Eastern Amazonia (Figure 2B). Rodrigues and Duarte (2024) previously reconstructed the ancestral ranges of *Myrcia* using six biogeographic models in BioGeoBEARS, selecting DEC+J as the best-fitting model based on AICc. This model allowed ancestral nodes to occupy up to three areas and suggested Evoregion B as the most likely origin of *Myrcia*, with clades in Evoregion A associated with colder environments.

In this study, we build on that framework and use Biogeographical Stochastic Mapping (BSM), generating 100 realizations based on the DEC+J parameters, to infer the biogeographical history of the lineages. This approach incorporates both cladogenetic and anagenetic events and allows us to estimate colonization ages and in situ diversification metrics while accounting for uncertainty in ancestral histories. Details on the implementation of the BSM in the *Herodotools* R package are described in the Suplemmentary Material (Text S1.1).

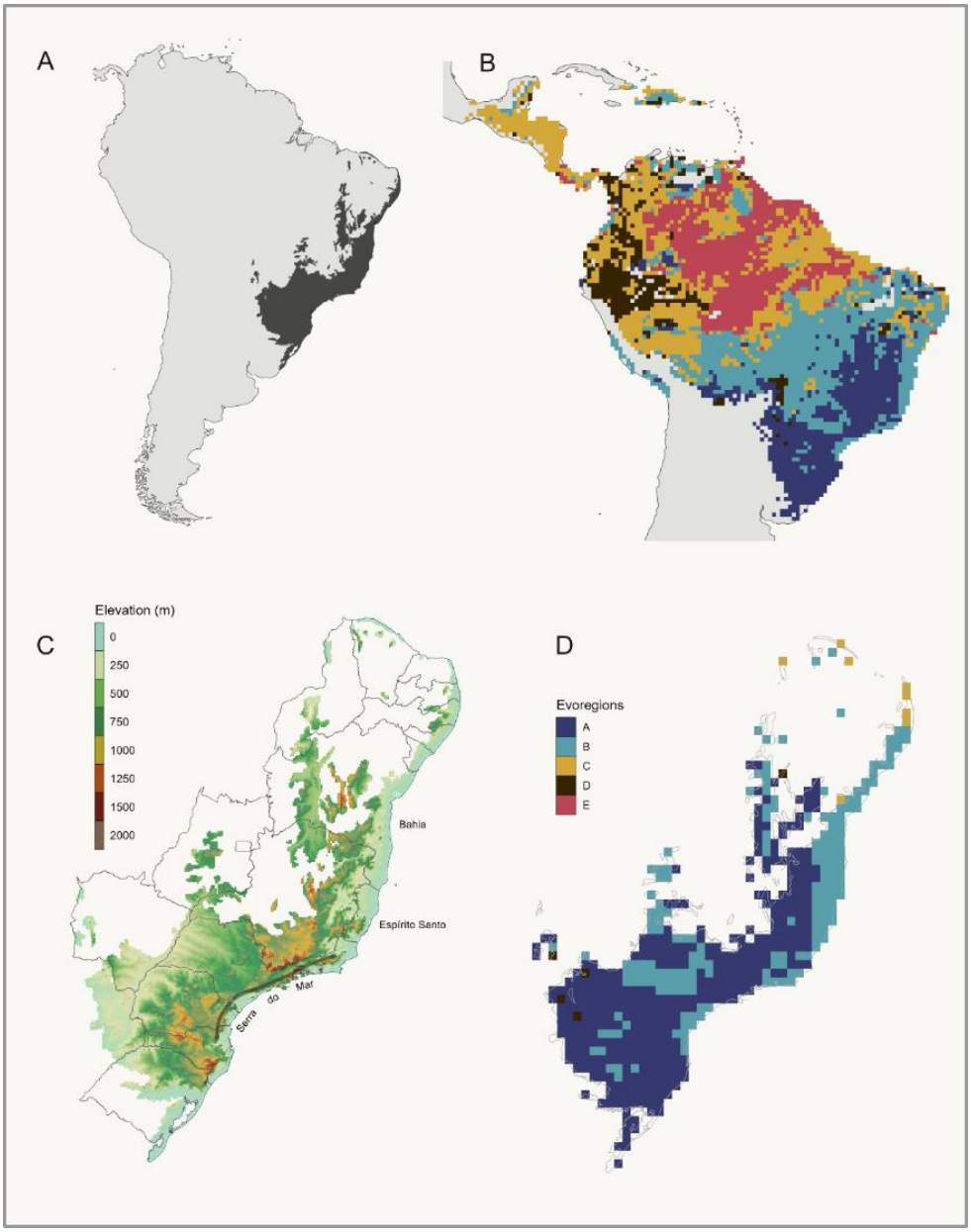



**Figure 2.** Atlantic Forest spatial context. A) Spatial distribution of Atlantic Forest (dark grey) in the South America. B) Biogeographic regions defined from phylogenetic composition (evoregions) for the genus *Myrcia* (Rodrigues and Duarte 2024). C) Elevation variation across Atlantic Forest. D) Biogeographic regions in the Atlantic Forest.

*In situ diversification and lineage dispersal*

We calculated a modified version of the Diversification Rate metric (DR; Jetz et al. 2012), to separate it into two components: in situ diversification (hereafter called $DR_{in\ situ}$) and historical dispersal., $DR_{in\ situ}$ represents the proportion of the speciation events (measured as DR) that occurred exclusively within the region where a species currently occurs. DR metric is the inverse of the equal-splits measure (ES; Redding and Mooers 2006) and is calculated as $DR_{Jetz} = 1/ES$. On a fully binary tree, the ES metric for a species *i* is the sum of the edge lengths in the path from tip to root of the tree, consecutively discounted by a factor of 0.5:

$$ES_i = \sum_{j=1}^{N_i} l_j \frac{1}{2^{j-1}}$$

where $l_j$ is the length of the edge *j*, *j=1* is the edge of the tip of the focal species, and $N_i$ is the number of nodes in the path from the tip to the root of the tree. We used the same equation to derive the $ES_{in\ situ}$ but with a modified definition of $N_i$. In $ES_{in\ situ}$ (i.e. considering only the speciation events occurring within the focal region), $N_i$ represents the number of nodes from the tip to the earliest internal node that continuously occupied the region of the focal assemblage where the species occurs. The information on node occupancy along the lineage path was derived from an ancestral reconstruction model. From this, $DR_{in\ situ}$ was calculated as $1/ES_{in\ situ}$. To calculate the community average, we use the harmonic mean for $DR_{Jetz}$ and $DR_{in\ situ}$ (ignoring zeros for $DR_{in\ situ}$). To calculate a proportion of the DR that have occurred in situ, we calculate species level as $DR_{prop} = DR_{Jetz} \times (ES_{in\ situ}/ES_{total})$ and averaged for each site as Mean $DR_{prop} = \sum DR_{prop} / \sum DR_{Jetz}$. This proportional $DR_{prop}$ (hereafter simply $DR_{prop}$) ranges from 0 to 1, where 0 indicates that the entire evolutionary history of the species in an assemblage occurred in another region (lineage dispersal) and 1 indicates that the entire the evolutionary history of the species occurred within the assemblage's region (in situ diversification process). Consequently, Mean $DR_{prop}$ capture the relative role of lineage dispersal (higher role when Mean $DR_{prop} < 0.5$) and in situ diversification history (higher role when Mean $DR_{prop} > 0.5$) for the assemblage. Diversification metrics were computed with the function 'calc_insitu_diversification' from *Herodotools* version 2.0 (Nakamura et al. 2024). Since we are using 100 BSM maps to infer biogeographical history, we compute average and coefficient of variation of mean $DR_{prop}$ and $DR_{in\ situ}$ across the BSM maps.

*Colonization Age of Assemblages*

We used the same framework described in Box 1 to calculate the colonization age of each species in the region of the assemblage, using the function 'calc_age_arrival' from the *Herodotools* package. We calculated the mean colonization age for each assemblage based on all species' colonization ages in the assemblages they currently occur. Using results from the 100 BSM maps, we compute average and coefficient of variation of community average age across the BSM maps.

*Diversity metrics*

We quantified two assemblage-level diversity metrics: species richness and phylogenetic divergence (Tucker et al. 2017). Species richness was calculated as the number of species in each grid. We quantified phylogenetic divergence of assemblages using the standardized effect size of mean pairwise distance (SES MPD) among the species in a site. The effect size was calculated by using a null model that randomizes the taxa labels in the phylogeny with 1000 repetitions of MPD calculation. To compute the SES MPD we used the package *picante* (Kembel et al. 2010). Low values of SES MPD indicate that the assemblage is assembled by phylogenetically closed-related species and high values of SES MPD indicate that the assemblage is assembled by phylogenetically distant-related species.

*Identyfying legacies in Myrtaceae in Atlantic Forest*

The last step in our methodological framework consists in providing a general overview of the different processes affecting the assemblages across the Atlantic Forest (see Table 1). We did this by performing a cluster analysis where we classified the assemblages based on four metrics: $DR_{prop}$, colonization age of assemblages, SES MPD (phylogenetic divergence) and species richness. We used Euclidian distance of the metrics' z-scores with ward clustering algorithm. In addition, to interpret the differences among groups we performed a Principal Component Analysis (PCA) on the metrics. We choose the number of groups (three to five) by visually assessing the spatial arrangement of the cluster analysis, resulting in four groups.

**Results**

The $DR_{prop}$ values were higher in the evoregion A sites than in the evoregion B sites (Figure 3A). High $DR_{prop}$ means that evoregion A showed higher contribution of in situ diversification than lineage dispersal process to the assemblage composition. On the other hand, evoregion B showed an overall even contribution of both processes ($DR_{prop}$ around 0.5), but the NE coast shows higher $DR_{prop}$ (values between 0.6 and 0.8). Most of the sites

with low $DR_{prop}$ values were located in evoregion B and close to the border of the region. $DR_{jetz}$ and $DR_{in\ situ}$ showed similar spatial patters between them (Figure S2) in which the higher diversification rates are in the central Atlantic Forest. This is a mountainous region that connects with the Cerrado biome in central Brazil. Spatial patterns of assemblage colonization ages (Figure 3) showed older colonization ages in assemblages at the southernmost Atlantic Forest, northeastern coast and in highlands of the northeast continental area. Assemblage age varied between 12-15 Ma in these areas (purple sites in Figure 3B). Most of the assemblages in the Central Atlantic Forest had ages between 9-12 Ma. It is important to highlight that *Myrcia* has an crown node estimated around 27 Ma (24.55 – 31.19 Ma; Vasconcelos et al. 2020).

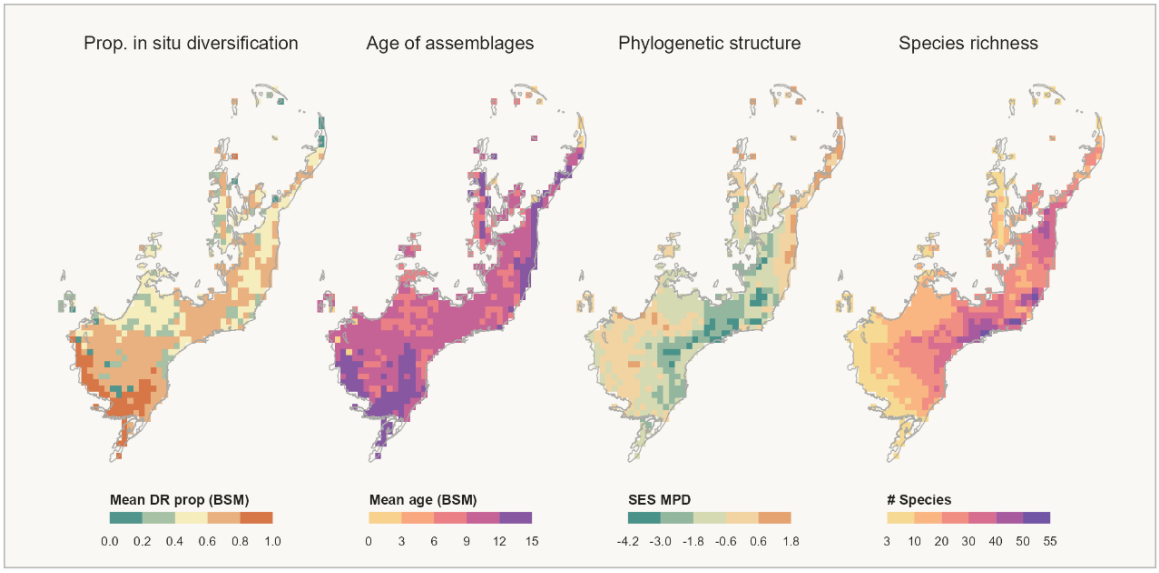



**Figure 3.** Spatial patterns of diversity and biogeographical processes influencing the assembly of *Myrcia* species in the Atlantic Forest. From left to right, proportion of in situ diversification quantifies the role of in situ diversification (high values) vs dispersal (low values) processes for the assemblage; mean age of assemblages, Phylogenetic structure, represent clustered (negative values) or overdispersed (positive values) assemblages; Species richness (D) is the number of species in each site.

The ancestral reconstruction affects the $DR_{prop}$, $DR_{in\ situ}$ and age of assemblages estimates since these metrics assumes the ancestral reconstruction is known. Using the BSM approach we were able to propagate the uncertainty in the ancestral reconstruction estimates to the assemblage level metrics. Our results show low uncertainty for $DR_{prop}$ and $DR_{in\ situ}$ which median coefficient of variation across the sites were 0.05 and 0.07, respectively. However, the age of assemblages showed higher uncertainty with median

coefficient of variation across the sites of 0.35. The spatial variation of the uncertainty is quite uniform in all the metrics (Figure S3)

Species richness, using only the species present in the phylogeny, showed at least three peaks along the coastal area of the Atlantic Forest (Figure 3). From South to North, the largest area with high species richness is in the Northern portion of the Serra do Mar, the second peak is in the Espirito Santo State, and the last one is in the Bahia State. The former two peaks in species richness also had low SES MPD values, but that in the Bahia state had high values of SES MPD. SES MPD metric revealed clustered assemblages of *Myrcia* along the Serra do Mar range (greener area in Figure 3B). By contrast, phylogenetically overdispersed assemblages predominated in the northeastern Atlantic Forest (orange region in Figure 3B).The southwestern the Atlantic Forest displayed a phylogenetic structure consistent with random expectation.

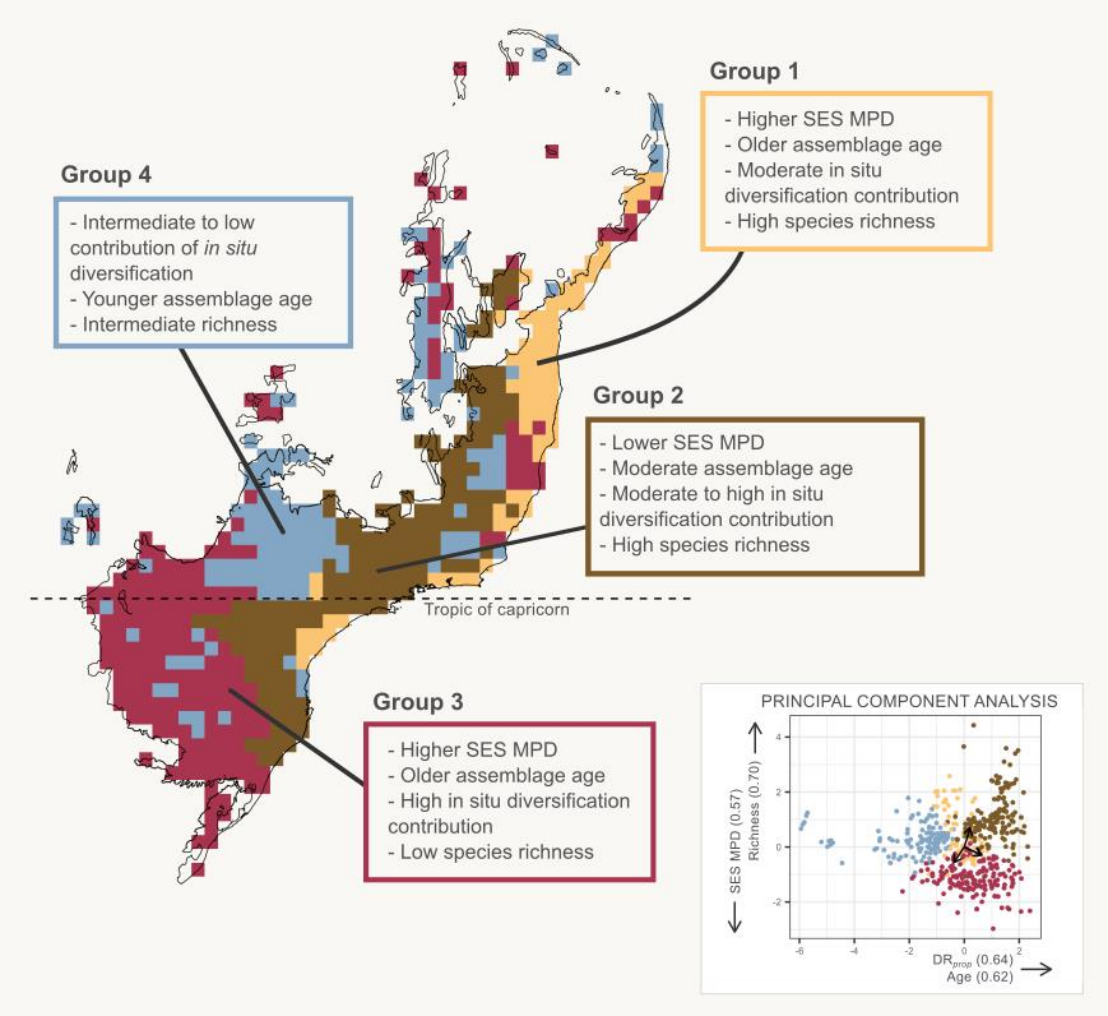



**Figure 4.** Schematic figure showing distinct historical structure of lineages for *Myrcia* assemblages in the Atlantic Forest. Groups were defined using Ward method of hierarchical clustering based on

4 metrics: $DR_{prop}$, age of assemblage, SES MPD and richness. The inset graphic represents a Principal Component Analysis (PCA) of assemblages using SES MPD, $DR_{prop}$, and richenss. The values in parenthesis represent the correlation between each variable and the PCA axis. Cluster analysis and Principal Component Analysis used to interpret differences between groups are presented in the Supplementary Material.

**Table 2**. Principal Component Analysis summary showing the relationship of the variables with the Principal Components (PC) and the variation explained by each PC axis.

| | PC1 | PC2 | PC3 | PC4 |
|---|---|---|---|---|
| Mean Colonization Age | 0.623 | -0.303 | 0.334 | 0.639 |
| Mean Drprop | 0.644 | -0.291 | -0.136 | -0.694 |
| SES MPD | -0.385 | -0.572 | 0.680 | -0.251 |
| Richness | 0.221 | 0.705 | 0.639 | -0.215 |
| Proportion of Variance | 0.495 | 0.336 | 0.141 | 0.028 |
| Cumulative Proportion | 0.495 | 0.831 | 0.972 | 1.000 |

By aggregating the information on $DR_{prop}$, age of colonization, SES MPD and richness we found four groups of sites with distinct legacies from biogeographical historical processes (Figures 4 and S4). The Principal Component Analysis using the same variables, the first tree axes explaining 97% of the variation (Table 2 and Figure S5). PC1 explains 49.5% of the variation and are positively related to mean $DR_{prop}$ and mean colonization age. PC2 explains 33.6% of the variation and are positively related to richness and negatively related to SES MPD. In contrast, PC3 (14.1% of explained variation) is positively related to both richness and SES MPD. This means that $DR_{prop}$ and colonization age are corelated but independent to SES MPD. The four groups are well separated in the PCA space (Figure 4 and S5). Following the expectations summarized in Table 1, we can classify the Group 1 (yellow in Figure 4) as a stable area acting as refugium since sites in this group present moderate history of in situ diversification, old colonization age of assemblages, overdispersed phylogenetic structure and high species richness. Group 2 (brown in Figure 4) seems to be an unstable area acting as cradle since it has moderate to high proportion of in situ diversification, moderate to high colonization ages, clustered phylogenetic structure and intermediate to high species richness. Group 3 (magenta in Figure 4) can be classified as stable area acting as cradle since it has high proportion of in situ diversification, old age of colonization, high SES MPD but with low species richness. Finally, we classify Group 4

(blue in Figure 4) as a unstable area acting as sink of lineages since it has lower proportion of in situ diversification, younger age of colonization, neutral SES MPD (similar to expected by chance) and intermediate species richness.

## Discussion

Our study examined how evolutionary and biogeographical history shape present-day patterns of phylogenetic composition. We quantified the relative roles of in situ diversification and lineage dispersal in assembling communities, along with the colonization age and phylogenetic structure of these assemblages. Our results show that in situ diversification is proportionally higher in Evoregion A (groups 2 and 3 in Figure 4), especially in the subtropical zone, whereas lineage dispersal plays a larger role in shaping assemblages in Evoregion B. Furthermore, by integrating $DR_{prop}$, colonization age, SES MPD, and species richness, we identified four regions with distinct historical biogeographical dynamics. This demonstrates the potential of our framework to uncover historical processes at the assemblage level.

High levels of in situ diversification may provide evolutionary advantages for species in assemblages subjected to stronger environmental selection pressure. In Evoregion A, particularly in the subtropical Southern Atlantic Forest (Group 3; Figure 4), assemblages are largely composed by species that diversified within this region. In contrast, assemblages in the evoregion B, are more evenly shaped by both in situ diversification and lineage dispersal. That suggests that lineages originating in Evoregion A have likely accumulated adaptations that enhanced their success during community assembly in these areas. Since evoregion A has colder climate, which are more frequent and intense at higher altitudes and latitudes (Rodrigues and Duarte 2024), we argue that these adaptations might be associated with the ability to cope with colder climatic conditions. For instance, in Southwest Atlantic Forest (Group 3, Figure 4), assemblages are dominated by some of the oldest lineages evolving within the Evoregion A, with only a few lineages having recently dispersed into the region. This indicates that in situ diversification plays a key role in shaping the phylogenetic composition of this area, despite its relative low species richness. Supporting this pattern, previous studies provided evidence of selection for cold-adapted lineages in tree assemblages in the subtropical Atlantic Forest, where Myrtaceae species are particularly important at high altitude sites (Bergamin et al. 2021; Duarte et al. 2014; Giehl and Jarenkow 2012). In addition, the subtropical Atlantic Forest appears to have remained under stable conditions since the last ~120 kyr (Carnaval et al. 2014). Together, these

patterns suggest that Group 3 area has have provided stable environmental condition for cold adapted species, and that the long-term persistence of the lineages in this region has promoted diversification, likely driven by lower extinction rates on cold adapted lineages.

Lineage dispersal appears to play a greater role in shaping tropical assemblages, particularly outside the Evoregion A. In these areas, evolutionary history seems to be less important for successful colonization, likely due to weaker selection pressures. Within the Atlantic Rainforest, species inhabiting warmer and wetter environmental conditions tend to exhibit a narrower climatic niche, whereas those in colder and drier environments exhibit broader niche (Klipel et al. 2022). These contrasting niche characteristics likely favored the selection of lineages tolerant or adapted to colder conditions in the Evoregion A. Conversely, in areas where the selection pressure is weaker, the history of diversification is less determinant for the assembly process. The Central-West Atlantic Forest (Group 4, Figure 4) exemplify such a case: here, in situ diversification and lineage dispersal contributes equally to assemblages with young age of colonization. This suggests that the region functions as an unstable sink for lineages originating elsewhere. In such context, contemporary ecological processes, such as competition and priority effects, are likely more important in shaping community assembly than geographical diversification history (Gerhold et al. 2018; Mittelbach and Schemske 2015; Ricklefs 1987).

Northeastern Atlantic Forest is characterized by an old colonization history, intermediate proportion of in situ diversification, an overdispersed phylogenetic structure, and high species richness (Group 1, Figure 4). These patterns are consistent with the Bahia refugium hypothesis, which proposes that this region has maintained relatively stable climatic conditions since at least the last 21 ky (Carnaval et al. 2014; Carnaval and Moritz 2008; Staggemeier et al. 2015). The stable climatic condition potentially enabled the persistence of old and distant-related lineages, which in turn contributed to the high species richness found in this area. In a study focused on *Myrcia* section *Aulomyrcia*, Staggemeier et al. (2015) found a low extinction rate in the Bahia refugium and high speciation rates outside the refugium area. Therefore, this region seems to be stable for long time and act as a refugium for *Myrcia* lineages.

Central Atlantic Forest (Group 2, Figure 4) probably act as an unstable cradle area. This is a mountainous region with the higher altitude in Atlantic Forest. Mountainous regions can promote ecological speciation due to high environmental and topographic heterogeneity (Pyron et al. 2015; Rahbek et al. 2019) which can also be a source of higher

levels of extinction (Rahbek et al. 2019; Vasconcelos et al. 2022). The climate in this area has been shown to have higher levels of paleoinstability since last interglacial maximum (Carnaval et al. 2014; Carnaval and Moritz 2008). The climatic instability associated to the topographic heterogeneity of the area has resulted in phylogenetic clustered assemblages with moderate to high levels of in situ diversification, but with intermediate colonization ages in comparison to other areas. We speculate that the diversity in this area results from highly dynamic evolutionary and biogeographical processes, characterized by continuous speciation driven by the creation of novel habitats and the fragmentation of species distribution, as well as range expansion and contractions in species range in response to environmental changes, alongside the extinction of some lineages.

By integrating diversity, in situ diversification history, and phylogenetic structure of assemblages, we were able to identify areas with distinct historical dynamics that surpass biogeographical borders (defined here as evoregions) and explain compositional variation within biogeographical areas. Typically, the influence of biogeographical history and its dynamics on biodiversity patterns is examined at courser resolutions using biogeographical areas as the unity of analysis (Antonelli et al. 2018; Matos-Maraví et al. 2021). Here we increase the resolution of analysis and address historical biogeographical dynamics at local assemblages, which is an important step towards understanding the effects of past dynamics on local communities (Gerhold et al. 2018; Mittelbach and Schemske 2015). This methodological advance was possible because we incorporate biogeographical information (from ancestral area estimates models) into tip-based diversification metrics and combining these with diversity (richness) and phylogenetic metrics (MPD). Therefore, our approach allows the detection of the influence of past source-sink dynamics among evolutionary regions on present-day local assemblages that could be applied in other systems and taxonomic groups.

Our study shed light on the legacies of the biogeographical history on the current assembly of *Myrcia* species in the Atlantic Forest. We were able to capture distinct role of in situ diversification and dispersal across the Atlantic Forest and identify areas with distinct biogeographical dynamics. In line with other studies, we detected patterns of a stable area acting as a refugium in the Northeastern Atlantic Forest (Carnaval et al. 2014; Carnaval and Moritz 2008). In addition, we showed a large area acting as cradle and areas acting as sink of lineages. Finally, our findings demonstrate the ability of the new proposed metrics of in situ diversification to shed light on within biome diversification dynamics and its legacies at the local assemblages.

## Data availability statement

Code and data necessary to reproduce the analysis are available at https://doi.org/10.5281/zenodo.14171268

## References


Amorim, Bruno S., Thais N. C. Vasconcelos, Gustavo Souza, Marccus Alves, Alexandre Antonelli, and Eve Lucas. 2019. "Advanced Understanding of Phylogenetic Relationships, Morphological Evolution and Biogeographic History of the Mega-Diverse Plant Genus Myrcia and Its Relatives (Myrtaceae: Myrteae)." *Molecular Phylogenetics and Evolution* 138(July 2018):65–88. doi:10.1016/j.ympev.2019.05.014.

Antonelli, Alexandre, Alexander Zizka, Fernanda Antunes Carvalho, Ruud Scharn, Christine D. Bacon, Daniele Silvestro, and Fabien L. Condamine. 2018. "Amazonia Is the Primary Source of Neotropical Biodiversity." *Proceedings of the National Academy of Sciences of the United States of America* 115(23):6034–39. doi:10.1073/pnas.1713819115.

Bergamin, Rodrigo S., Guilherme D. S. Seger, Marcos B. Carlucci, Martin Molz, Ricardo S. P. Mello, Rafael Martins, João A. Jarenkow, Paulo Brack, Sandra C. Müller, and Leandro D. S. Duarte. 2021. "Elevational Shifts in Phylogenetic Diversity of Angiosperm Trees across the Subtropical Brazilian Atlantic Forest." *Austral Ecology* 46(3):486–95. doi:10.1111/aec.12996.

Carnaval, A. C., E. Waltari, M. T. Rodrigues, D. Rosauer, J. VanDerWal, R. Damasceno, I. Prates, M. Strangas, Z. Spanos, D. Rivera, M. R. Pie, C. R. Firkowski, M. R. Bornschein, L. F. Ribeiro, and C. Moritz. 2014. "Prediction of Phylogeographic Endemism in an Environmentally Complex Biome." *Proceedings of the Royal Society B: Biological Sciences* 281(1792):20141461–20141461. doi:10.1098/rspb.2014.1461.

Carnaval, Ana Carolina, and Craig Moritz. 2008. "Historical Climate Modelling Predicts Patterns of Current Biodiversity in the Brazilian Atlantic Forest." *Journal of Biogeography* 35(7):1187–1201. doi:10.1111/j.1365-2699.2007.01870.x.

Cavender-Bares, Jeannine, David D. Ackerly, Sarah E. Hobbie, and Philip A. Townsend. 2016. *Evolutionary Legacy Effects on Ecosystems: Biogeographic Origins, Plant Traits, and Implications for Management in the Era of Global Change*. Vol. 47.

Costa, L. P. 2003. "The Historical Bridge between the Amazon and the Forest of Brazil a Study of Molecular Phylogeography with Small Mammals." *Journal of Biogeography* 30(1):71–86. doi:10.1046/j.1365-2699.2003.00792.x.

Duarte, Leandro Da Silva, Rodrigo Scarton Bergamin, Vinícius Marcilio-Silva, Guilherme Dubal Dos Santos Seger, and Márcia Cristina Mendes Marques. 2014. "Phylobetadiversity among Forest Types in the Brazilian Atlantic Forest Complex." *PLoS ONE* 9(8):1–10. doi:10.1371/journal.pone.0105043.

Dupin, Julia, Nicholas J. Matzke, Tiina Särkinen, Sandra Knapp, Richard G. Olmstead, Lynn Bohs, and Stacey D. Smith. 2017. "Bayesian Estimation of the Global Biogeographical History of the Solanaceae." *Journal of Biogeography* 44(4):887–99. doi:10.1111/jbi.12898.

Fiaschi, Pedro, and José R. Pirani. 2009. "Review of Plant Biogeographic Studies in Brazil." *Journal of Systematics and Evolution* 47(5):477–96. doi:10.1111/j.1759-6831.2009.00046.x.

Gerhold, Pille, Marcos B. Carlucci, Şerban Procheş, and Andreas Prinzing. 2018. "The Deep Past Controls the Phylogenetic Structure of Present, Local Communities." *Annual Review of Ecology, Evolution, and Systematics* 49(August):477–97. doi:10.1146/annurev-ecolsys-110617-062348.

Giehl, Eduardo L. Hettwer, and João André Jarenkow. 2012. "Niche Conservatism and the Differences in Species Richness at the Transition of Tropical and Subtropical Climates in South America." *Ecography* 35(10):933–43. doi:10.1111/j.1600-0587.2011.07430.x.

Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann, and A. O. Mooers. 2012. "The Global Diversity of Birds in Space and Time." *Nature* 491(7424):444–48. doi:10.1038/nature11631.

Kembel, Steven W., Peter D. Cowan, Matthew R. Helmus, William K. Cornwell, Helene Morlon, David D. Ackerly, Simon P. Blomberg, and Campbell O. Webb. 2010. "Picante: R Tools for Integrating Phylogenies and Ecology." *Bioinformatics* 26(11):1463–64. doi:10.1093/bioinformatics/btq166.

Klipel, Joice, Rodrigo Scarton Bergamin, Adriane Esquivel-Muelbert, Renato A. F. de Lima, Alexandre Adalardo de Oliveira, Paulo Inácio Prado, and Sandra Cristina Müller. 2022. "Climatic Distribution of Tree Species in the Atlantic Forest." *Biotropica* 54(5):1170–81. doi:10.1111/btp.13140.

Lucas, E. J., B. S. Amorim, D. F. Lima, A. R. Lima-Lourenço, E. M. Nic Lughadha, C. E. B. Proença, P. O. Rosa, A. S. Rosário, L. L. Santos, M. F. Santos, M. C. Souza, V. G. Staggemeier, T. N. C. Vasconcelos, and M. Sobral. 2018. "A New Infra-Generic Classification of the Species-Rich Neotropical Genus Myrcia s.l." *Kew Bulletin* 73(1). doi:10.1007/s12225-017-9730-5.

Lucas, Eve J., and Mariana O. Bünger. 2015. "Myrtaceae in the Atlantic Forest: Their Role as a 'Model' Group." *Biodiversity and Conservation* 24(9):2165–80. doi:10.1007/s10531-015-0992-7.

Matos-Maraví, Pável, Niklas Wahlberg, André V. L. Freitas, Phil Devries, Alexandre Antonelli, and Carla M. Penz. 2021. "Mesoamerica Is a Cradle and the Atlantic Forest Is a Museum of Neotropical Butterfly Diversity: Insights from the Evolution and Biogeography of Brassolini (Lepidoptera: Nymphalidae)." *Biological Journal of the Linnean Society* 1–21. doi:10.1093/biolinnean/blab034.

Matzke, Nicholas Joseph. 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). doi:10.21425/f5fbg19694.

McGill, Brian J., Jonathan M. Chase, Joaquín Hortal, Isaac Overcast, Andrew J. Rominger, James Rosindell, Paulo A. V. Borges, Brent C. Emerson, Rampal Etienne, Michael J. Hickerson, D. Luke Mahler, Francois Massol, Angela McGaughran, Pedro Neves, Christine Parent, Jairo Patiño, Megan Ruffley, Catherine E. Wagner, and Rosemary Gillespie. 2019. "Unifying Macroecology and Macroevolution to Answer Fundamental Questions about Biodiversity." *Global Ecology and Biogeography* 28(12):1925–36. doi:10.1111/geb.13020.

Mittelbach, Gary G., and Douglas W. Schemske. 2015. "Ecological and Evolutionary Perspectives on Community Assembly." *Trends in Ecology and Evolution* 30(5):241–47. doi:10.1016/j.tree.2015.02.008.

591 Murray-Smith, Charlotte, Neil A. Brummitt, Ary T. Oliveira-Filho, Steven Bachman, Justin
592 Moat, Eimear M. Nic Lughadha, and Eve J. Lucas. 2009. "Plant Diversity Hotspots
593 in the Atlantic Coastal Forests of Brazil." *Conservation Biology* 23(1):151–63.
594 doi:10.1111/j.1523-1739.2008.01075.x.

595 Muylaert, Renata Lara, Maurício Humberto Vancine, Rodrigo Bernardo, Júlia Emi Faria
596 Oshima, Thadeu Sobral-Souza, Vinicius Rodrigues Tonetti, Bernardo Brandão
597 Niebuhr, and Milton Cezar Ribeiro. 2018. "A Note on the Territorial Limits of the
598 Atlantic Forest." *Oecologia Australis* 22(3):302–11. doi:10.4257/oeco.2018.2203.09.

599 Nakamura, Gabriel, Arthur V. Rodrigues, André Luís Luza, Renan Maestri, Vanderlei
600 Debastiani, and Leandro Duarte. 2024. "Herodotools: An R Package to Integrate
601 Macroevolution, Biogeography and Community Ecology." *Journal of Biogeography*
602 (51):533–43. doi:10.1111/jbi.14610.

603 Oliveira-Filho, Ary T., and Marco Aurelio L. Fontes. 2000. "Patterns of Floristic
604 Differentiation among Atlantic Forests in Southeastern Brazil and the Influence of
605 Climate1." *Biotropica* 32(4b):793–810. doi:10.1111/j.1744-7429.2000.tb00619.x.

606 Phillips, Steven J., Robert P. Anderson, Miroslav Dudík, Robert E. Schapire, and Mary E.
607 Blair. 2017. "Opening the Black Box: An Open-Source Release of Maxent."
608 *Ecography* 40(7):887–93. doi:10.1111/ecog.03049.

609 Pinto-Ledezma, Jesús N., Alex E. Jahn, Victor R. Cueto, José Alexandre F. Diniz-Filho,
610 and Fabricio Villalobos. 2019. "Drivers of Phylogenetic Assemblage Structure of
611 the Furnariides, a Widespread Clade of Lowland Neotropical Birds." *American*
612 *Naturalist* 193(2):E41–56. doi:10.1086/700696.

613 POWO. 2025. "Plants of the World Online. Myrcia DC. Ex Guill." *Facilitated by the Royal*
614 *Botanic Gardens, Kew.*
615 https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:30001220-2#children.

616 Pyron, R. Alexander, Gabriel C. Costa, Michael A. Patten, and Frank T. Burbrink. 2015.
617 "Phylogenetic Niche Conservatism and the Evolutionary Basis of Ecological
618 Speciation." *Biological Reviews* 90(4):1248–62. doi:10.1111/brv.12154.

619 R Core Team. 2022. "R: A Language and Environment for Statistical Computing."

Rahbek, Carsten, Michael K. Borregaard, Alexandre Antonelli, Robert K. Colwell, Ben G. Holt, David Nogues-Bravo, Christian M. Ø. Rasmussen, Katherine Richardson, Minik T. Rosing, Robert J. Whittaker, and Jon Fjeldså. 2019. "Building Mountain Biodiversity: Geological and Evolutionary Processes." *Science* 365(6458):1114–19. doi:10.1126/science.aax0151.

Redding, David W., and Arne O. Mooers. 2006. "Incorporating Evolutionary Measures into Conservation Prioritization." *Conservation Biology* 20(6):1670–78. doi:10.1111/j.1523-1739.2006.00555.x.

Ree, Richard H., and Stephen A. Smith. 2008. "Maximum Likelihood Inference of Geographic Range Evolution by Dispersal, Local Extinction, and Cladogenesis." *Systematic Biology* 57(1):4–14. doi:10.1080/10635150701883881.

Ricklefs, R. E. 1987. "Community Diversity: Relative Roles of Local and Regional Processes." *Science* 235(4785):167–71. doi:10.1126/science.235.4785.167.

Rodrigues, Arthur Vinicius, and Leandro Duarte. 2024. "Mapping Species Richness and Evolutionary Regions of the Genus Myrcia." *Journal of Biogeography* 51(5):827–39. doi:10.1111/jbi.14791.

Saiter, Felipe Zamborlini, Jason L. Brown, William Wayt Thomas, Ary T. de Oliveira-Filho, and Ana Carolina Carnaval. 2016. "Environmental Correlates of Floristic Regions and Plant Turnover in the Atlantic Forest Hotspot." *Journal of Biogeography* 43(12):2322–31. doi:10.1111/jbi.12774.

Santos, Matheus Fortes, Eve Lucas, Paulo Takeo Sano, Sven Buerki, Vanessa Graziele Staggemeier, and Félix Forest. 2017. "Biogeographical Patterns of Myrcia s.l. (Myrtaceae) and Their Correlation with Geological and Climatic History in the Neotropics." *Molecular Phylogenetics and Evolution* 108:34–48. doi:10.1016/j.ympev.2017.01.012.

Scherrer, Daniel, Manuela D'Amen, Rui F. Fernandes, Rubén G. Mateo, and Antoine Guisan. 2018. "How to Best Threshold and Validate Stacked Species Assemblages? Community Optimisation Might Hold the Answer." *Methods in Ecology and Evolution* 2018(June):2155–66. doi:10.1111/2041-210X.13041.

649 Staggemeier, Vanessa Graziele, José Alexandre Felizola Diniz-Filho, Félix Forest, and Eve
650 Lucas. 2015. "Phylogenetic Analysis in Myrcia Section Aulomyrcia and Inferences
651 on Plant Diversity in the Atlantic Rainforest." *Annals of Botany* 115(5):747–61.
652 doi:10.1093/aob/mcv005.

653 Tucker, Caroline M., Marc W. Cadotte, Silvia B. Carvalho, T. Jonathan Davies, Simon
654 Ferrier, Susanne A. Fritz, Rich Grenyer, Matthew R. Helmus, Lanna S. Jin, Arne O.
655 Mooers, Sandrine Pavoine, Oliver Purschke, David W. Redding, Dan F. Rosauer,
656 Marten Winter, and Florent Mazel. 2017. "A Guide to Phylogenetic Metrics for
657 Conservation, Community Ecology and Macroecology." *Biological Reviews* 92(2):698–
658 715. doi:10.1111/brv.12252.

659 Vasconcelos, Thais N. C., Suzana Alcantara, Caroline O. Andrino, Félix Forest, Marcelo
660 Reginato, Marcelo F. Simon, and José R. Pirani. 2020. "Fast Diversification through
661 a Mosaic of Evolutionary Histories Characterizes the Endemic Flora of Ancient
662 Neotropical Mountains." *Proceedings of the Royal Society B: Biological Sciences* 287(1923).
663 doi:10.1098/rspb.2019.2933.

664 Vasconcelos, Thais, Brian C. O'Meara, and Jeremy M. Beaulieu. 2022. "Retiring 'Cradles'
665 and 'Museums' of Biodiversity." *The American Naturalist* 199(2):194–205.
666 doi:10.1086/717412.

667 Velasco, Julián A., and Jesús N. Pinto-Ledezma. 2022. "Mapping Species Diversification
668 Metrics in Macroecology: Prospects and Challenges." *Frontiers in Ecology and*
669 *Evolution* 10. doi:10.3389/fevo.2022.951271.

670 Villalobos, Fabricio, Jesús N. Pinto-Ledezma, and José Alexandre Felizola Diniz-Filho.
671 2020. "Evolutionary Macroecology and the Geographical Patterns of Neotropical
672 Diversification." Pp. 85–101 in.

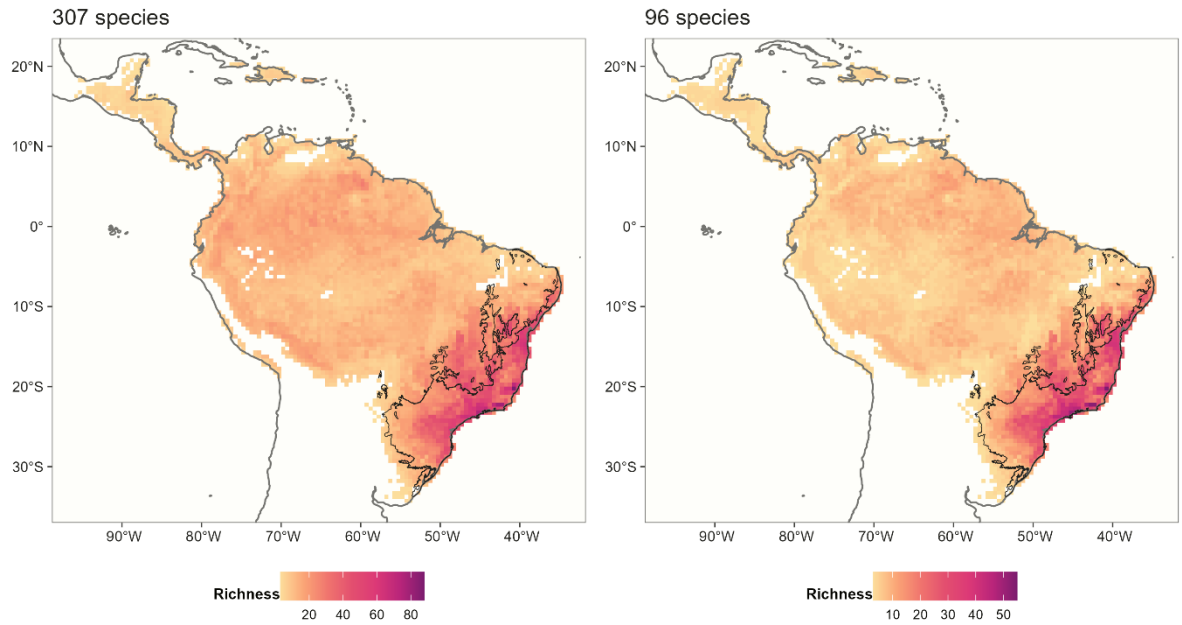
673 Wiens, John J., and Michael J. Donoghue. 2004. "Historical Biogeography, Ecology and
674 Species Richness." *Trends in Ecology and Evolution* 19(12):639–44.
675 doi:10.1016/j.tree.2004.09.011.

676

677

678 **Supplementary Material**

679



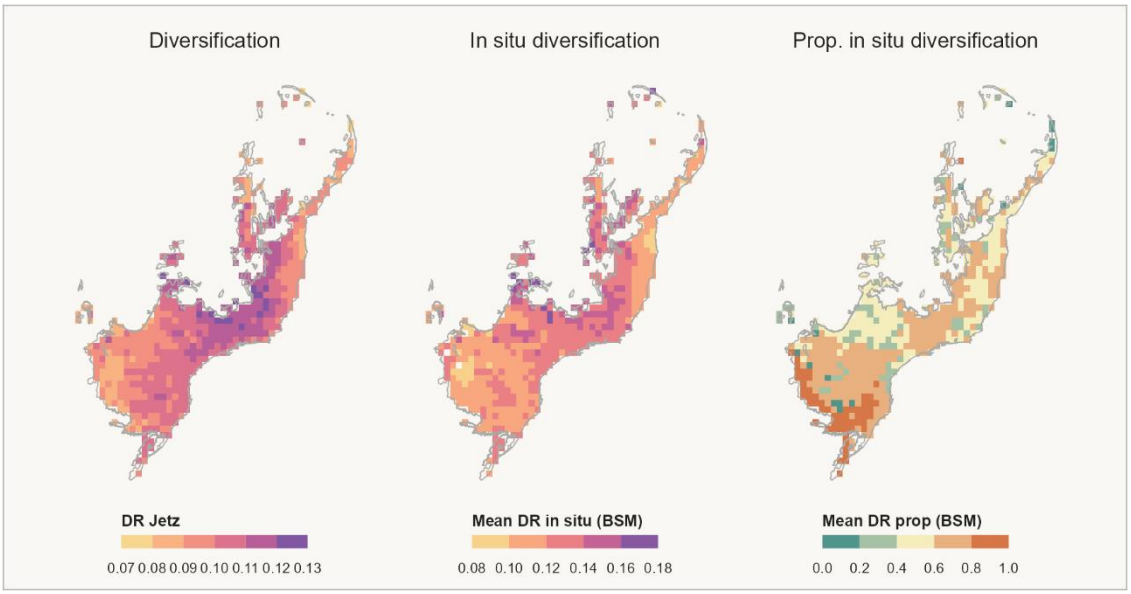

680

681 Figure S1. *Myrcia*'s species richness patterns in Neotropics. The figures compare the spatial
682 pattern between two datasets, one with all species with distribution map available (left
683 panel) and one with only the species that are present in the phylogenetic tree used in this
684 study (right panel). Pearson coefficient of correlation is 0.958. The outer lines show the
685 border of America's continent and inner lines show the border of the Atlantic Forest.

686

687

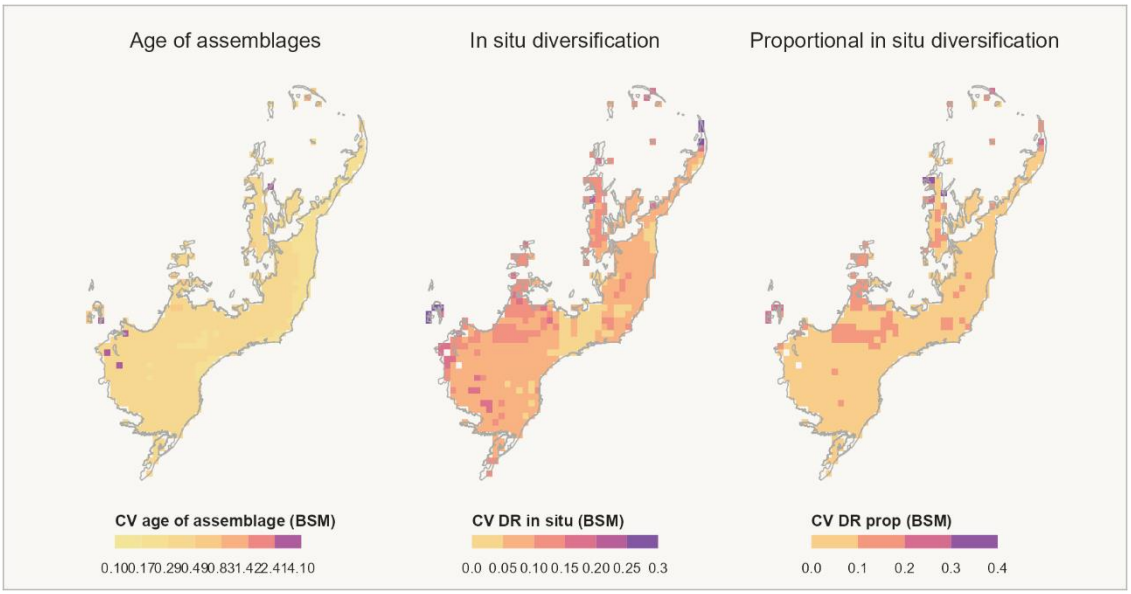
688




689

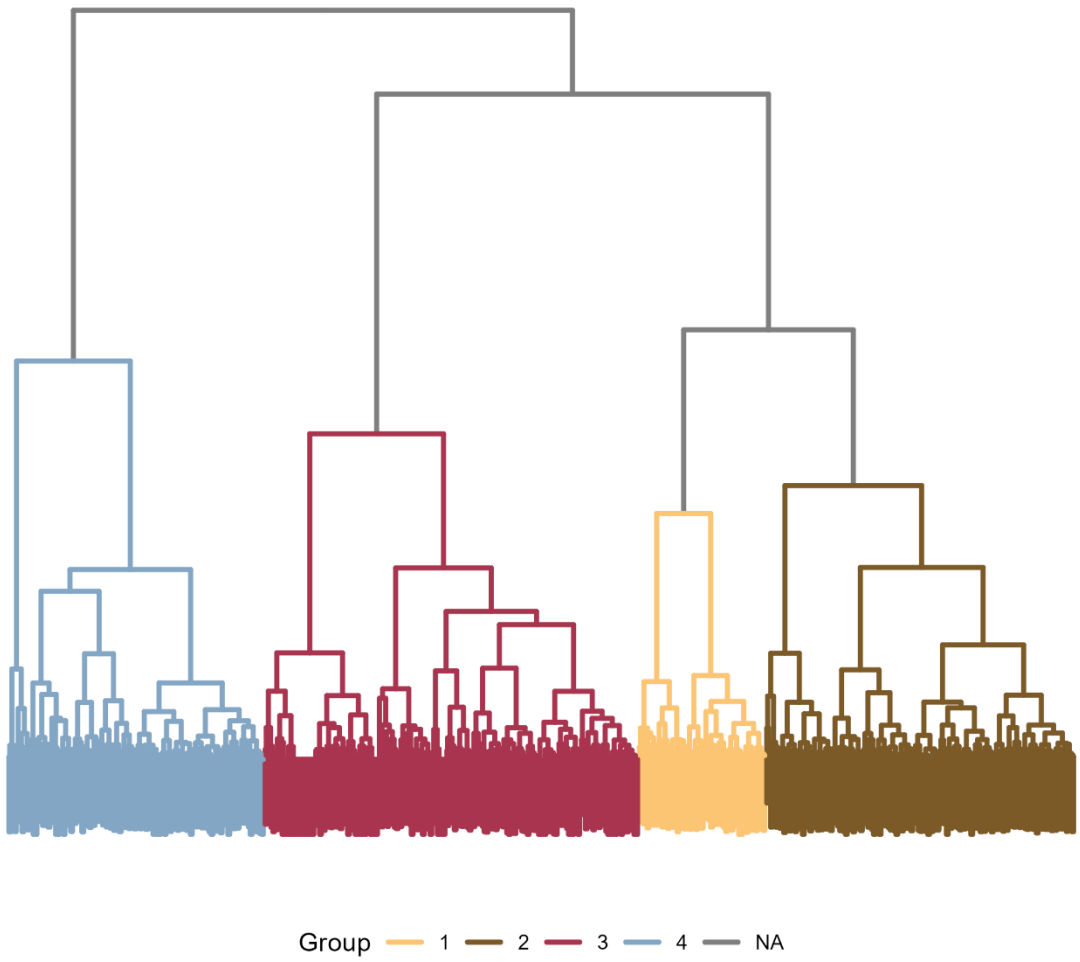
690 Figure S2. Comparison among the diversification rate (DR) metrics. Community average
691 DR as the inverse of the equal splits as proposed by Jetz et al. 2012 is shown in the left
692 panel (Diversification). Central panel shows DR in situ (In situ diversification) and the right
693 panel shows DR proportional (Prop. In situ diversification). DR in situ calculates the
694 inverse of the equal splits only for species in the community that have diversified in situ
695 and for the time it has occurred in situ since colonization. DR proportional calculates the
696 mean proportion of the diversification history that has occurred in situ. For details on the
697 calculation of DR metrics refer to the Methods section of the main text.

698




699

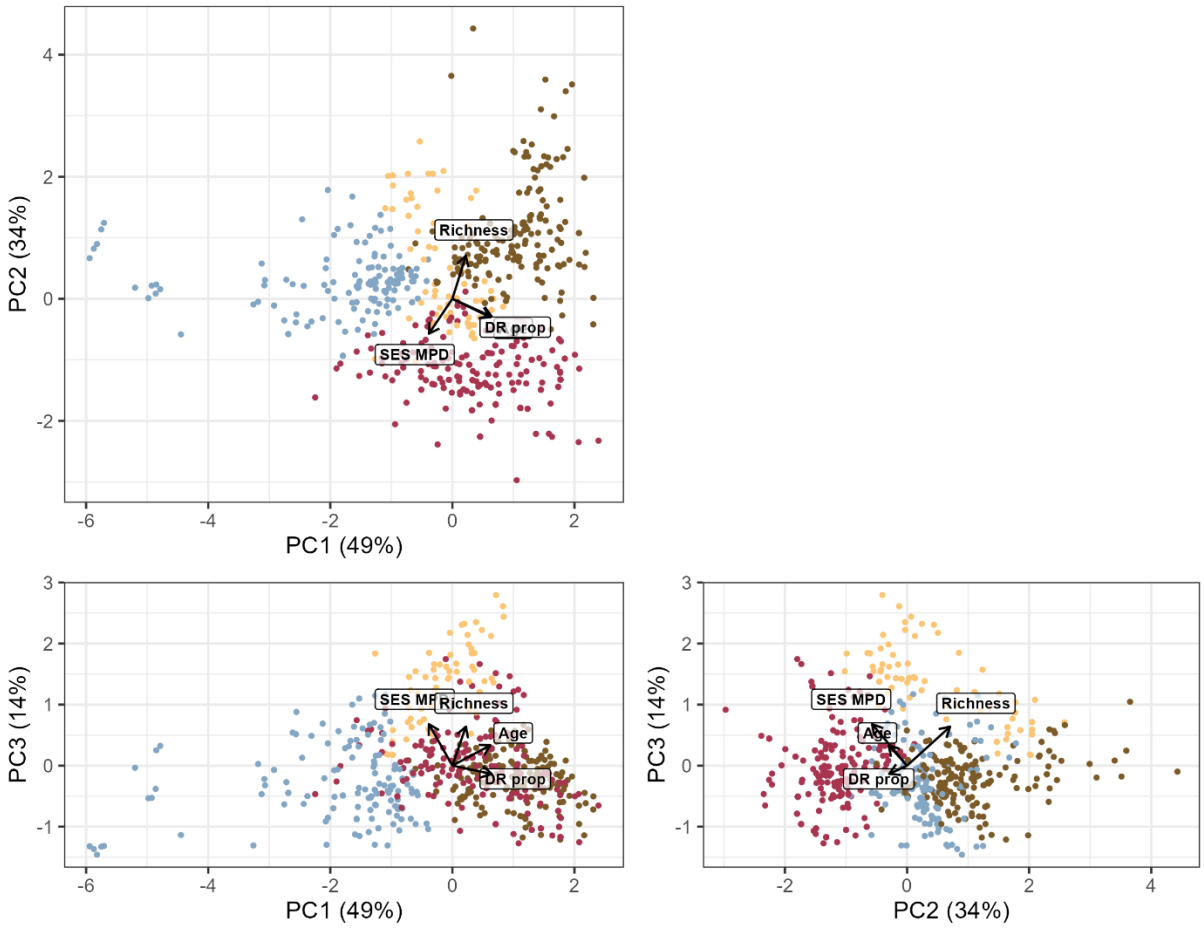
700 Figure S3. Uncertainty (calculated as the coefficient of variation) of colonization age of
701 assemblages, DR in situ (In situ diversification) and DR prop (Proportional in situ
702 diversification) based on 100 biogeographical stochastic mappings using DEC model of
703 ancestral state reconstruction with BioGeoBEARS.




704

705 Figure S4. Dendrogram of ward clustering of sites based on *in situ* diversification,

706 colonization age of assemblages, SES MPD and richness. Each color represents one group.




707

708 Figure S5. Principal Component Analysis of the site's variability on *in situ* diversification

709 (represented by DR prop), colonization age of assemblages (Age), SES MPD and richness.

710 Each color represents one group of sites obtained using ward clustering method (Figure

711 S4).

712

713

714

## Text S1 - Methods description:

### Text S1.1 - Implementation of the Biogeographic Stochastic Mapping (BSM) in Herodotools R package

The *Herodotools* package was developed to integrate biogeographical history methods of ancestral range reconstruction with phylogenetic metrics at the assemblage level. This integration uses ancestral area estimates as the basis for disentangling a species' history between the region where it currently occurs (the assemblage) and other regions (Box 1, main text). The method accepts biogeographic reconstructions from any of the models implemented in the R package *BioGeoBEARS* (Matzke, 2013).

In the first implementation (Nakamura et al., 2024), *Herodotools*, used the most likely range state estimated at each phylogenetic node to track species' biogeographical histories. A key limitation of this approach is that it cannot account for range-state changes along branches, since only node-based changes are considered. Moreover, by relying on a single most likely ancestral state, this approach does not incorporate the uncertainty inherent in ancestral range estimates.

To overcome these issues, we extended the *Herodotools* to incorporate Biogeographic Stochastic Mapping (BSM) outputs from BioGeoBEARS into its workflow. BSM generates a set of equally probable histories consistent with the phylogenetic tree, model parameters, and data. Each stochastic map represents one possible biogeographic history, assigning range changes to both nodes and branches by random sampling based on a fitted biogeographic model (e.g., the Dispersal–Extinction–Cladogenesis model, DEC). Because each map is a random draw, multiple reconstructions (e.g., 100) must be generated to capture the uncertainty in range change estimates. The advantage of using the BSM is that it tracks the range state transition along the entire phylogeny (nodes and branches) and by

740 producing many maps, allows the quantification of uncertainty in ancestral range
741 reconstructions. Within the *Herodotools* framework, assemblage-level phylogenetic metrics
742 are computed for each BSM map, thereby propagating uncertainty in biogeographic
743 reconstruction into the phylogenetic metrics. After metrics are computed across all maps,
744 summary statistics such as the mean and standard deviation can be derived.

745 **Text S1.2 - Obtaining Diversification Rates and importance of in situ diversification**
746 **at assemblage scale**

747 Based on the results of biogeographical ancestral range reconstruction with BSM, we
748 modified the diversification rate (DR) metric proposed by Jetz et al., (2012) to account for
749 in situ diversification. By in situ diversification, we refer to the speciation rate that occurs
750 for a set of lineages within a given region (in this study, an evoregion). Here, we detail how
751 we obtained the proportion of diversification attributable to in situ speciation, which we
752 call $DR_{prop}$ .

753 We first calculated $DR_{in\ situ}$—the diversification rate of each lineage within the region it
754 currently occupies. This value is defined as the inverse of the in situ evolutionary
755 distinctness, $ED_{in\ situ}$, for each lineage. For a given region, $ED_{in\ situ}$ is computed by tracing
756 each lineage only along the branches (from tip to root) where the ancestral range includes
757 that region. By definition, $ED_{in\ situ} \leq ED_{total}$. EDtotal is the ED calculation that do not
758 account for the in situ diversification, as originally proposed (Redding et al., 2008; Redding
759 & Mooers, 2006).

760 Finally, based on $ED_{toalt}$, $ED_{in\ situ}$ , and $DR_{Jetz}$, we derived the proportion of diversification
761 due to in situ speciation, denoted as $DR_{prop}$ in situ per lineage *i*.

762

$$DR_i^{prop} = DR_i \cdot \frac{ED_i^{insitu}}{ED_i^{tot}} = \frac{ED_i^{insitu}}{(ED_i^{tot})^2}$$

$DR_{prop}$ was calculated for each lineage *i*, which in this study corresponds to each tip of the phylogenetic tree. We then scaled up lineage-level $DR_{prop}$ in situ to the assemblage level by computing the community-weighted average of per-species in situ proportions, denoted as mean $DR_{prop}$, as follows:

$$Mean\ DR_{comm} = \frac{\sum_i DR_i^{prop}}{\sum_i DR_i^{Jetz}}$$

This equation corresponds to a community-weighted average of per-lineage in situ proportions, where the weights are proportional to the diversification rate of each lineage. It can be interpreted as the average fraction of diversification occurring in situ across all lineages present in the assemblage, with species exhibiting higher in situ diversification rates contributing more strongly to the overall value. In the Herodotools package, the function calc_insitu_diversification is used to calculate $DR_{Jetz}$ $DR_{in\ situ}$, $DR_{prop}$ , and their correspondents community mean values.

References


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

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

Nakamura, G., Rodrigues, A. V., Luza, A. L., Maestri, R., Debastiani, V., & Duarte, L. (2024). Herodotools: An R package to integrate macroevolution, biogeography and community ecology. *Journal of Biogeography*, *51*, 533–543. https://doi.org/10.1111/jbi.14610

Redding, D. W., Hartmann, K., Mimoto, A., Bokal, D., DeVos, M., & Mooers, A. (2008). Evolutionarily distinctive species often capture more phylogenetic diversity than expected. *Journal of Theoretical Biology*, *251*(4), 606–615. https://doi.org/10.1016/j.jtbi.2007.12.006

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