Geography and ecology shape the phylogenetic composition of Amazonian tree communities


1Deceased (Dairon Cárdenas López – January 5, 2022; Nállarett Dávila – November 30, 2022; Cid Ferreira – September 21, 2023).

For Affiliation refer page on 16

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Authors. Journal of Biogeography published by John Wiley & Sons Ltd.
Correspondence
Bruno Garcia Luize, Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas – UNICAMP, CP 6109, Campinas, SP, 13083–970, Brazil.
Email: luize.bg@gmail.com
Kyle Dexter, The University of Edinburgh, School of GeoSciences, 219 Crew Building, King’s Buildings, Edinburgh, EH9 3FF, United Kingdom.
Email: kyle.dexter@ed.ac.uk

Funding information
Conselho Nacional de Desenvolvimento Científico e Tecnológico; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior; Fundação de Amparo à Pesquisa do Estado de São Paulo; HORIZON EUROPE Marie Skłodowska-Curie Actions

Abstract
Aim: Amazonia hosts more tree species from numerous evolutionary lineages, both young and ancient, than any other biogeographic region. Previous studies have shown that tree lineages colonized multiple edaphic environments and dispersed widely across Amazonia, leading to a hypothesis, which we test, that lineages should not be strongly associated with either geographic regions or edaphic forest types.

Location: Amazonia.

Taxon: Angiosperms (Magnoliids; Monocots; Eudicots).

Methods: Data for the abundance of 5082 tree species in 1989 plots were combined with a mega-phylogeny. We applied evolutionary ordination to assess how phylogenetic composition varies across Amazonia. We used variation partitioning and Moran’s eigenvector maps (MEM) to test and quantify the separate and joint contributions of spatial and environmental variables to explain the phylogenetic composition of plots. We tested the indicator value of lineages for geographic regions and edaphic forest types and mapped associations onto the phylogeny.

Results: In the terra firme and várzea forest types, the phylogenetic composition varies by geographic region, but the igapó and white-sand forest types retain a unique evolutionary signature regardless of region. Overall, we find that soil chemistry, climate and topography explain 24% of the variation in phylogenetic composition, with 79% of that variation being spatially structured (R^2 = 19% overall for combined spatial/environmental effects). The phylogenetic composition also shows substantial spatial patterns not related to the environmental variables we quantified (R^2 = 28%). A greater number of lineages were significant indicators of geographic regions than forest types.

Main Conclusion: Numerous tree lineages, including some ancient ones (>66 Ma), show strong associations with geographic regions and edaphic forest types of Amazonia. This shows that specialization in specific edaphic environments has played a long-standing role in the evolutionary assembly of Amazonian forests. Furthermore, many lineages, even those that have dispersed across Amazonia, dominate within a specific region, likely because of phylogenetically conserved niches for environmental conditions that are prevalent within regions.

KEYWORDS
community assembly, dispersal limitation, environmental selection, evolutionary principal component analysis, indicator lineage analysis, Moran’s eigenvector maps, neotropics, Niche conservatism, tropical rain forests
1 INTRODUCTION

A major goal of tropical biologists is to understand the processes involved in the origin, maintenance and coexistence of diverse species in ecological communities. Processes that play out over multiple spatial and temporal scales influence the organization of ecological communities (Ricklefs & Schluter, 1993; Vellend, 2010), leaving an imprint on phylogenetic and functional composition. Detailed characterization of compositional patterns can help to infer the ecological and evolutionary processes at play (Carlucci et al., 2017; Dray et al., 2012; Graham & Fine, 2008; Kraft & Ackerly, 2010; McIntire & Fajardo, 2009). The assembly of tree communities in Amazonian tropical rain forests is structured by ecological drift (Hubbell, 2001), the local and regional dispersal of lineages (Dexter et al., 2017), differential recruitment into heterogeneous forest habitats and environments (Fine & Kembel, 2011; ter Steege et al., 2006; Tuomisto et al., 2003), biotic interactions (Fine et al., 2004; Kursar et al., 2009) and historical contingency (Baker et al., 2014; Dexter et al., 2012; Figueiredo et al., 2022).

Tree lineages that are iconic of present-day tropical rain forests have existed since the mid-Cretaceous (Davis et al., 2005; Jaramillo et al., 2006; Wing et al., 2009), but angiosperms only came to dominate Amazonia after the Cretaceous–Paleogene (K-Pg) boundary (c.a. 66 Ma) (Carvalho et al., 2021). Amazonian tropical rain forests cover a huge area (approximately 39% of South America) and there is a substantial turnover of species across its extent (Baraloto et al., 2021; Luize et al., submitted; ter Steege et al., 2006; Terborgh & Andresen, 1998). It also comprises heterogeneous geologic, edaphic and climatic conditions (Hoorn et al., 2010) and hosts the highest number of tree species of any biogeographic region on Earth (Raven et al., 2020; Ulloa-Ulloa et al., 2017). Locally, small topographic variation is enough to cause edaphic differences (e.g., physical, chemical and hydrologic soil variation) leading to major shifts in tree species composition (Bohlman et al., 2008; Koponen et al., 2004; Marca-Zevallos et al., 2022; Milton et al., 2022; Phillips et al., 2003; Sabatier et al., 1997; ter Steege et al., 1993). Such high species turnover can make it challenging to understand large-scale patterns as many tree communities share no species at all. Evolutionary approaches based on analysing the shared evolutionary heritage of communities (e.g., Pavoine, 2016; Webb, 2000), may provide a path forwards, as all tree communities in Amazonia are dominated by species derived from the same overarching clade, the angiosperms.

Tropical tree communities tend to have clustered phylogenetic structures (Baldeck et al., 2016; Baraloto et al., 2012; Kembl & Hubbell, 2006; Webb, 2000), which is interpreted as the signal of evolutionarily conserved preferences of tree lineages for specific environments, although other factors, such as biotic interactions and disturbances, could also result in phylogenetic clustering. Phylogenetic clustering has been found not only in local and regional studies (Baraloto et al., 2012; Fine & Kembl, 2011; Guevara et al., 2016; Kraft et al., 2007), but also at the scale of all of Amazonia (Carlucci et al., 2017; Honorio Coronado et al., 2015). This broad-scale phylogenetic clustering in Amazon-wide studies (Honorio Coronado et al., 2015) may result from a deep phylogenetic conservatism for edaphic environments. Conversely, edaphic regime switching (i.e., the shift from one forest edaphic habitat to another), associated with restricted inter-habitat gene flow, has been suggested to be a major driver of plant specialization in Amazonia (Brousseau et al., 2021; Fine et al., 2005; Gentry, 1981). If edaphic regime switching is frequent, lineages should not show strong associations with forest edaphic regimes and tree communities should not show a phylogenetic composition organized within forest types.

At the spatial scale of Amazonia, the clustering of closely related species within communities and regions can also result from speciation within geographic regions, be it sympatric, parapatric or allopatric, with limited subsequent dispersal (Fine et al., 2004; Misiewicz et al., 2020). Phylogenetically, dispersal of Amazonian trees can be limited (Prance, 1982; ter Steege et al., 2003, 2006; Terborgh & Andresen, 1998). Nevertheless, some tree lineages have dispersed widely across Amazonia over their evolutionary history (Dexter et al., 2017) and do not present any region-specific phylogenetic clustering. A recent study from a theoretical perspective also suggests widespread dispersal of lineages across Amazonia (Pos et al., 2023). Finally, while the largest Amazonian rivers can provide effective geographic barriers limiting the distribution of some organisms, particularly animals, rivers do not seem to be effective barriers for most tree species (Dexter et al., 2012; Luize et al., submitted; Nazareno et al., 2017, 2019).

Assuming that dispersal allows tree lineages to achieve widespread ranges and that migration across landscapes allows tree lineages to achieve broad niche breadths, a reasonable hypothesis might be that neither geography nor ecology are major drivers of tree-lineage distribution. If so, the phylogenetic composition of Amazonian tree communities will not align closely with geographic or environmental conditions. There is mixed evidence for the nature of the phylogenetic structure of Amazonian tree communities, including communities which do not show any clear phylogenetic structure, although most forest types tend to be phylogenetically clustered at the landscape to regional scales (Baraloto et al., 2012; Fine & Kembel, 2011; Guevara et al., 2016). The strongest evidence of phylogenetically clustered communities tends to be found in habitats with the most stressful ecological conditions, such as seasonally dry forests, white-sand forests and seasonally flooded forests (Baraloto et al., 2021; Guevara et al., 2016; Honorio Coronado et al., 2015). However, to date, most phylogenetic analyses of tree communities have focused on limited geographic areas (e.g., Aldana et al., 2017; Cárdenas et al., 2017; González-Caro et al., 2021) or relied on phylogenies with relatively low phylogenetic resolution (Honorio Coronado et al., 2015). The growing availability of DNA sequences allows the production of phylogenetic hypotheses with a better sampling and resolution of relationships (Baraloto et al., 2021; Chave et al., 2020; Dexter & Chave, 2016), which in turn enables a deeper exploration of variation in the phylogenetic composition of Amazonian forests.
In this study, we characterize the phylogenetic composition of tropical rain forests across the entire Amazonian region, quantifying the roles of spatial and environmental variables in shaping the phylogenetic composition of tree communities, with the aim to discern the key evolutionary and ecological processes structuring these hyperdiverse forests. We test the null hypothesis that widespread geographic dispersal of lineages and frequent edaphic-regime switching have erased phylogenetic signal for geographic regions and forest edaphic regimes. Previous studies that have found significant phylogenetic clustering may have done so because of poorly resolved phylogenies or restricted spatial and environmental scales. In addition, we evaluated if there were significant indicator lineages for specific geographic regions or edaphic forest types, and when found, we determined their total numbers, identities and ages.

2 | MATERIALS AND METHODS

2.1 | Study region, forest-inventory dataset and environmental variables

The focus of our study is the Amazonian rain forest, including the catchment of the Amazon River, the middle-lower course of the Araguaia-Tocantins River basin, the Atlantic coastal basins surrounding the Amazon River estuary and the rain forests of the Guianas (RAISG, 2012). Our forest-inventory dataset (ATDN_20220919) results from the combined effort of thousands of botanists and field ecologists over the last 90 years. The dataset consists of 1989 forest-inventory plots with measurements for the abundance of all trees \( \geq 10 \text{ cm DBH} \) (i.e., diameter at breast height, at 1.3 m above the ground), with a few plots (26) sampling all trees \( \geq 5 \text{ cm DBH} \); most of the plots are of standard size (1 ha) (minimum and maximum size = 0.1 and 9 ha; 533 plots <1 ha, 1350 plots of 1 ha and 106 plots >1 ha). The inclusion of few plots with non-standardized tree inclusion criteria or sample size did not affect the estimations of evoPCA scores (see below and Figure S2). Of the 1,099,810 measured trees, 89% were identified to a valid species name and the other 11% of individuals were excluded from further analyses. The 979,801 individual trees identified to a valid species name encompass a total of 115 species-level families, 754 genera and 5082 species of angiosperms. Checking of valid species names followed the previous name harmonization of the ATDN database (ter Steege et al., 2019).

Inventory plots were grouped into six geographic regions (Central Amazonia [CA = 506 plots], Guiana Shield [GS = 459], Northwestern Amazonia [NWA = 304], Southern Amazonia [SA = 273], Southwestern Amazonia [SWA = 224] and Eastern Amazonia [EA = 223]), for description of geographic regions (Feldpausch et al., 2012; ter Steege et al., 2013). These geographic regions largely follow a broad-scale classification of Amazonia based on geological, pedological and forest structural attributes (Feldpausch et al., 2012; Quesada et al., 2011) and which have been used in many subsequent large-scale biogeographic analyses in Amazonia (e.g., González-Caro et al., 2021; Guevara et al., 2016; ter Steege et al., 2013). The inventory plots were also classified by the researchers who established the plots, according to five major Amazonian forest types that develop on distinct edaphic regimes (Prance, 1979; ter Steege et al., 2013; Terborgh & Andresen, 1998): (1) terra-firme forests on clay and/or brown sands [TF = 1326], (2) várzea forests on eutrophic seasonally flooded terrains [VA = 315], (3) igapó forests on oligotrophic seasonally flooded terrains [IG = 193], (4) terra-firme forests on white-sands (podzols) [PZ = 114] and (5) swamp forests on permanently flooded terrains [SW = 41]. An initial set of 19 environmental variables capturing climatic, topographic and edaphic aspects crucial to tree ecology were extracted for each plot location from digital databases and applied as environmental predictors of phylogenetic composition (Table 1). For details regarding characteristics of forest types on distinct edaphic regimes and the environmental variables used in the analyses, see the Supplementary material (Methods extended).

2.2 | Phylogeny of Amazonian tree species

We generated a phylogeny using the GBOTB phylogeny as the base topology (Smith & Brown, 2018), using Scenario 3 in the V.phylomaker v.2 R package (Jin & Qian, 2022), which binds unsampled species (those without any usable DNA sequence data available) to the phylogeny and then prunes the phylogeny to species present in the overall sample. At the time of the production of the GBOTB phylogeny (Smith & Brown, 2018), roughly 74% of the genera and 30% of the tree species found in the largest tree plot database for the Amazonian region (the ATDN dataset) had at least one usable DNA sequence in public data sources (e.g., GenBank). The 3519 species that were not found in the GTOB phylogeny were bound to the phylogeny based on their taxonomic identity (i.e., genus identity). For unsampled species where the algorithm failed to find a genus relative, we indicated the closest genus to that species according to the most recently published phylogeny for the given family. The phylogeny of Amazonian tree species, produced from Scenario 3 of the V.phylomaker v.2, comprises 7062 lineages, of which 5082 are lower-level lineages (i.e., tips in the phylogeny and represent species) and 1980 are higher-level lineages (i.e., internal nodes in the phylogeny that may represent genera, families or other lineages above the species-level).

Previous studies of the effects of phylogenetic uncertainty on standard ecological analyses have shown that uncertainty in results is primarily due to variation in the placement of unsampled taxa, rather than uncertainty in topological relationships or node ages (Rangel et al., 2015). Scenario 3 of the V.phylomaker v.2 R package binds a given unsampled taxon on a specific branch at a determined position on that branch (e.g., halfway along the branch subtending the genus of the given species). There is an alternative scenario in the V.phylomaker package (Scenario 2) which randomly places taxa within their given taxonomic clade. Thus, in order to explore variation in results due to uncertainty in the placement of taxa, we additionally constructed a set of 100 phylogenies using Scenario 2.
of the V.phylomaker v.2 (Jin & Qian, 2022) and conducted some of our downstream analyses on this set of phylogenies.

### 2.3 Data analysis

We have opted to apply a raw data approach (i.e., ordination of community composition matrix) instead of a distance-based approach (i.e., ordination of the matrix of pairwise dissimilarities) to analyse the data. The raw data approach has been shown to be the appropriate choice for analysing spatial and environmental patterns of community composition (Legendre et al., 2008), including the spatial and environmental influences on the phylogenetic compositional variation of tropical tree communities (e.g., Rezende et al., 2021), which is a key focus of our study. Analytical approaches based on distance metrics generally address different questions (Legendre et al., 2008), such as providing insights on the influence of more basal or recent clades to phylogenetic beta diversity patterns (e.g., Carlucci et al., 2017; Duarte et al., 2014), but these are not our focus here. We also did not focus on approaches based on null models that aim to disentangle phylogenetic beta diversity patterns from taxonomic beta diversity patterns (cf. Duarte, Debastiani, Freitas & Pillar, 2016). The phylogenetic approaches we employed here consider terminal nodes in our phylogeny (i.e., species), similar to deeper nodes in the phylogeny; they both represent evolutionary units, or lineages, of interest. We aimed for general results that integrate across lineages at all evolutionary levels (in both deep and recent evolutionary time). Furthermore, the raw data approaches allowed the full use of advanced and versatile spatial eigenvector-based methods (such as Moran’s eigenvector maps and Moran Spectral Randomization), which in turn allowed the detection and quantification of broad to fine-scaled complex spatial structures.

### 2.4 Evolutionary principal component analysis (evoPCA)

We applied evolutionary principal component analysis (evoPCA, Pavoine, 2016) to summarize the key dimensions of phylogenetic compositional variation. The evoPCA performs a PCA on the abundance matrix of lineages by sites (Pavoine, 2016), ordinating, at the same time, the distributions of lineages (i.e., terminal and internal nodes) among sites and the phylogenetic composition of the sites. The phylogenetic branch lengths (i.e., evolutionary units) within communities are used to weigh and position the communities according to the structure of the phylogeny and the lineage composition of the sites. To down-weight the influence of abundant lineages, we applied a Hellinger transformation to the abundance matrix of lineages by sites that form the basis of the analyses (Legendre & Gallagher, 2001; Pavoine, 2016). We used

<table>
<thead>
<tr>
<th>Environmental predictor</th>
<th>Variable</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Climate – Temperature</td>
<td>Mean annual temperature (°C)</td>
<td>Karger et al. (2017)</td>
</tr>
<tr>
<td>Climate – Temperature</td>
<td>Temperature range (°C)</td>
<td>Karger et al. (2017)</td>
</tr>
<tr>
<td>Climate – Temperature</td>
<td>Temperature seasonality (SD - Standard deviation of monthly mean temperatures)</td>
<td>Karger et al. (2017)</td>
</tr>
<tr>
<td>Climate – Precipitation</td>
<td>Total annual precipitation (kg.m⁻²)</td>
<td>Karger et al. (2017)</td>
</tr>
<tr>
<td>Climate – Precipitation</td>
<td>Precipitation seasonality (CV - Coefficient of Variation)</td>
<td>Karger et al. (2017)</td>
</tr>
<tr>
<td>Climate - Water deficit</td>
<td>Maximum Climatological Water Deficit (MCWD)</td>
<td>Funk et al. (2015); Hirota et al. (2011); Malhi et al. (2009)</td>
</tr>
<tr>
<td>Topography</td>
<td>Elevational position (height above sea level) – ALOS World 3D</td>
<td>Tadono et al. (2014); Takaku et al. (2016)</td>
</tr>
<tr>
<td>Topography</td>
<td>Topographic Diversity Index – TopoDiver</td>
<td>Theobald et al. (2015)</td>
</tr>
<tr>
<td>Topography</td>
<td>Multiscale Topographic Position Index – mTPI</td>
<td>Theobald et al. (2015)</td>
</tr>
<tr>
<td>Edaphic – Water table depth</td>
<td>Groundwater table depth (m below surface) – WTD</td>
<td>Fan et al. (2013)</td>
</tr>
<tr>
<td>Edaphic – Soil chemistry</td>
<td>Nitrogen (N)</td>
<td>This study</td>
</tr>
<tr>
<td>Edaphic – Soil chemistry</td>
<td>Phosphorous (P)</td>
<td>This study</td>
</tr>
<tr>
<td>Edaphic – Soil chemistry</td>
<td>Potassium (K⁺)</td>
<td>This study</td>
</tr>
<tr>
<td>Edaphic – Soil chemistry</td>
<td>Calcium (Ca²⁺)</td>
<td>This study</td>
</tr>
<tr>
<td>Edaphic – Soil chemistry</td>
<td>Magnesium (Mg²⁺)</td>
<td>This study</td>
</tr>
<tr>
<td>Edaphic – Soil chemistry</td>
<td>Sodium (Na⁺)</td>
<td>This study</td>
</tr>
<tr>
<td>Edaphic – Soil chemistry</td>
<td>Aluminium (Al³⁺)</td>
<td>This study</td>
</tr>
<tr>
<td>Edaphic – Soil chemistry</td>
<td>Cation exchange capacity (CEC)</td>
<td>This study</td>
</tr>
<tr>
<td>Edaphic – Soil chemistry</td>
<td>Potential of hydrogen (pHₐq)</td>
<td>This study</td>
</tr>
</tbody>
</table>
the absolute values of abundances prior to the Hellinger transformation and the ordinary weighted mean to centre the PCA matrix (arguments w=’evoba’ and option=’centered’ in ‘evopalsheller’ function, ‘adiv’ R package, Pavoine, 2020). The evoPCA results in a multidimensional evolutionary space where tree communities located closer together (i.e., with similar axes scores) tend to have a more similar phylogenetic composition. To determine the number of axes to keep for further analysis, we used parallel analysis implemented with the ‘apa.parallel’ function in the ‘psych’ R package (Revelle, 2022), which identifies those axes explaining more variation in the ordination than expected by chance. This analysis indicated 13 axes explaining more variation than expected by chance, which together account for 51% of the total phylogenetic compositional variation described by the evoPCA (Figure S1).

To visualize the multidimensional evolutionary space occupied by Amazonian forests, we produced biplots for the first three evoPCA axes. For each edaphic forest type in each geographic region, we visualized which area of the biplots contained 50% of the sampled plots, essentially representing the ‘central tendency’ of plots, that is, those found between the 1st and 3rd quartiles in two dimensions, for a given forest type in a region. We used a two-dimensional kernel density estimation to draw these 50% contour lines (‘geom_density_2d’ function in the ggplot2 R package (Wickham, 2016)).

2.5 Spatial and environmental influences on phylogenetic composition of tree communities

We used Moran’s eigenvector maps (MEMs; Dray et al., 2006) to test for visualize and quantify multiscale spatial patterns of variation in the phylogenetic composition of tree communities across Amazonia focusing on the 13 significant evoPCA axes. The MEM consists of the double diagonalization of a spatial weighting matrix (SWM) obtained from a connectivity and weighting matrix calculated from the spatial coordinates of the studied plots (details in SM – Methods Extended). The resulting spatial eigenvectors (or MEM variables) each consist of an orthogonal spatial pattern and together encompass broad to fine spatial-scaled patterns, whose linear combinations allow capturing multiscale patterns in the response of interest. We followed Bauman, Drouet, Fortin, and Dray (2018) to create the spatial eigenvectors, by optimizing the selection of a spatial weighting matrix (SWM) from a limited set of contrasting graph-based connectivity matrices by adjusting the initial significance threshold of 0.05 for the total number of compared spatial weighting matrices using a Sidak correction (to avoid type I error rate inflation). Graph-based MEMs were preferred given the irregular spacing of the sampling design. Significance tests were performed through permutation of the residuals of a redundancy analysis (RDA) of the evoPCA axes against the MEM variables (Anderson & Legendre, 1999). The three graph-based connectivity matrices considered were a Gabriel graph, relative neighbourhood graph and minimum spanning tree, each weighted by a function decreasing linearly with the distance between plots, hence assuming that sites further apart will be less connected and have exchanged fewer lineages. As we aimed to capture spatial patterns in phylogenetic composition with maximum power and accuracy, we applied forward selection with a double-stopping criterion method (Blanchet et al., 2008) to select a final subset of spatial eigenvectors (Bauman, Drouet, Dray, & Vlemminckx, 2018; Bauman, Drouet, Fortin, & Dray, 2018) and we only retained eigenvectors capturing spatial patterns of positive autocorrelation, thereby focusing on aggregation patterns. The generation of SWMs was done with the function ‘listw.candidates’ and the optimization and selection of an SWM and subset of MEM variables were performed using the function ‘listw.select’, both functions in the ‘adespatial’ R package (Dray et al., 2021). The fitting process of the selected spatial predictors (MEMs) to the evoPCA axes was performed through redundancy analysis (RDA) (SM – Methods Extended), such that the resulting constrained ordination axes correspond to a unique linear combination of the selected spatial eigenvectors and therefore to a spatial pattern present in the phylogenetic composition. We tested each of these orthogonal spatial patterns (i.e., each RDA axis) separately by permutation (999 permutations) and the scores of the forest plots over the significant RDA axes were mapped for visualization of the corresponding spatial patterns (Figure S8).

We used a variation partitioning framework (Peres-Neto et al., 2006) to test and quantify the total, unique (also referred to as ‘pure’) and joint contributions of (a) environmental variables including attributes for climate, topography and soil chemistry and (b) the spatial predictors (MEMs) in explaining phylogenetic compositional variation. Explanatory power was quantified through the adjusted coefficient of determination ($R^2_{adj}$) (Legendre & Legendre, 2012). The total and unique explanatory fractions of the variation partitioning ($R^2_{adj}$) were tested through permutation of the residuals of the corresponding RDA and partial RDA, respectively (999 permutations) (Anderson & Legendre, 1999; Peres-Neto et al., 2006). The joint space–environment fraction – that is, the spatially structured effect of climate, topography and soil chemistry on phylogenetic composition – was tested with Moran Spectral Randomization (Wagner & Dray, 2015), following Bauman et al.’s (2019) method (SM – Methods Extended). We also visualized the ‘pure’ spatial patterns of phylogenetic compositional variation by mapping the constrained scores of a partial RDA (i.e., RDA constrained by MEMs for the residuals of the RDA constrained by environmental variables – Figure S9).

We conducted two main sensitivity analyses with respect to the variation partitioning analysis. Firstly, we tested the robustness of the variation partitioning fractions, by rerunning the above analyses based on the 56 evoPCA axes capturing 75% of the variation of the community phylogenetic composition (instead of the 13 axes capturing 51% of the variation). Secondly, we tested the robustness of results with respect to phylogenetic uncertainty (here, focusing on the first 13 evoPCA axes) by repeating the full workflow (including evoPCA and MEM selection) using the 100 phylogenetic trees produced from the Scenario 2 algorithm of V.phylomaker v.2.
2.6 Indicator value of lineages for geographic regions and forest types

We tested whether lineages (lower and higher-level evolutionary lineages, that is, species and internal nodes, respectively) were indicators of certain geographic regions or edaphic forest types by performing indicator analysis (IA) based on the IndVal statistic (Dufrene & Legendre, 1997). The IndVal statistic defines the most characteristic lineages for each geographic or forest group and is calculated for each lineage independently, so the pattern observed for a given lineage does not influence the indicator value of another lineage (Cáceres & Legendre, 2009). To run the indicator analysis, we used a community matrix containing abundances of the tips (i.e., species) and the internal nodes (representing evolutionary lineages at higher taxonomic levels, e.g., orders, families and genera) within sampled communities. As the indicator analysis was based on abundances and incidence, an indicator lineage meant both that the lineage achieved markedly higher abundance in that region or forest type as compared to other regions or forest types and that the lineage would generally be found in the group in which it is an indicator. The indicator analysis was performed for geographic regions and forest types using the function `indval` of the `labdsv` R-package (Roberts, 2019). IndVal significance levels were tested with 999 permutations of sites among groups and all those lineages achieving p-values ≤0.01 were considered as indicators of their corresponding group (i.e., assuming 1 false positive over 100 indicators). To test if there was a consistent pattern of lineage association with a given geographic region or forest type over evolutionary time, we mapped the indicator lineages onto the Amazonian tree species phylogeny. We then assessed the proportion of significant indicator lineages in different time slices by sub-setting the phylogeny into time periods defined by the following breakpoints: 0, 2.5, 5, 10, 20, 30, 45, 65, >65 Million years (Ma).

3 RESULTS

3.1 The phylogenetic composition of Amazonian forests

The first two evoPCA axes described 21.6% of the variation in the phylogenetic composition of Amazonian tree communities and no lower-order axes described more than 4.7% (Figure S1). A visual assessment of the evoPCA biplots shows some compositional overlap but mostly a gradual turnover in the phylogenetic composition of Amazonian tree communities over forest types and geographic regions (Figures 1, S3 and S4). There is a differentiation between terra-firme, white-sand and wetland forests although the level of differentiation depends on the geographic region (Figure 1). In Western Amazonia, terra-firme and várzea forests have a similar phylogenetic composition, while the phylogenetic composition of these forest types diverges in Central and Eastern Amazonia. On the Guiana Shield, the evolutionary space occupied by terra-firme forests overlaps with that occupied by wetland and white-sand forests (Figure 1). Overall, though, the phylogenetic composition of terra-firme forests tends to differ from that of both wetland forests (várzea, igapó and swamp forests – Figure S5) and white-sand forests, even within the same geographic region (Figures 1, S5 and S6). The evolutionary space occupied by wetland forests overlaps with that occupied by white-sand forests with great overlap between white-sand forests and várzea forests in Central Amazonia (Figure 1). In addition, the evolutionary space occupied by igapó forests is remarkably close to that occupied by white-sand forests (Figure S5). Arecales and Ericales were the lineages contributing most to positive and negative scores on the first evoPCA axis, respectively (Figure S3). For the second evoPCA axis, Fabales is the lineage contributing most for communities with negative axis scores and Burseraceae the lineage contributing for communities with positive axis scores (Figure S3).

Maps of the first two evoPCA axes highlight the widespread distribution of tree communities with central values for phylogenetic composition (i.e., with values between −0.2 and 0.2), embedded among communities that occupy the extremes of the phylogenetic compositional gradient (i.e., with values less than −0.2 or greater than 0.2) (Figures 2 and S6). Values below −0.2 on the 1st evoPCA axis (Figure 2a) tended to be located east of the Negro River, whereas values above 0.2 tended to be more widespread around Amazonia. There are two geographic clusters of high values for the 1st evoPCA axis, one in western Amazonia in the Andean forelands and the other in eastern Amazonia surrounding the Amazon River mouth. For the 2nd evoPCA axis (Figure 2b), the tree communities with extreme values for phylogenetic composition are mixed within those tree communities attaining central axis values. However, the scores of the 2nd evoPCA axis still show a geographic trend with values below −0.2 mostly east of the Negro River and values above 0.2 mostly located close to the Andean forelands.

3.2 Multiscale spatial patterns of phylogenetic composition across Amazonia

The optimized MEM analysis highlighted significant spatial structures at multiple scales ($R^2_{adj} = 0.62$) indicating a strong spatial component to the phylogenetic composition of trees across Amazonia ($R^2_{adj} = 0.47$ based on a reduced subset of 207 selected MEM variables; see Methods Extended). The first 12 dimensions of constrained RDA axes were significantly spatially structured revealing contrasting orthogonal phylogenetic spatial structures (Figure S8). The first and most significant spatial pattern captures a gradual differentiation in phylogenetic composition from communities in the central part of Amazonia towards two regions of internally similar phylogenetic composition, one located in southwestern Amazonia close to the Andean forelands and the second located in the Guiana Shield (Figure S8a; very similar map as evoPCA axis 1, Figure 2a). Subsequent phylogenetic spatial patterns tended to occur over comparatively shorter distances (finer-scaled spatial patterns; Figure S8b-j).
FIGURE 1 The evolutionary space is occupied by Amazonian tree communities. Scatterplots of the first two evoPCA axes mapping the tree communities' scores according to geographic regions and forest types. The first axis of the evoPCA recovered a longitudinal gradient in the differentiation of the phylogenetic composition of Amazonian tree communities that mirrors terra-firme and the wetland forest differentiation. The second axis mainly separates terra-firme from white-sand and wetland forests. The contour lines show the central tendency (50% of plots) for each forest type in each geographic region. Contour lines could not be calculated for subsets of plots with low sample size (e.g., in wetland and white-sand forests of the southern Amazonia). Forest types: TF – Terra-Firme; PZ – White-Sand; VA – Várzea; IG – Igapó; SW – Swamp.

TF PZ IG VA SW
3.3 | Spatial and environmental influence in the phylogenetic composition

In the variation partitioning analysis, a total of 52% ($R^2_{adj}$) of variation in the phylogenetic composition (first 13 evoPCA axes) was explained by the spatial and environmental predictors. A total of 47% of the variation was significantly spatially structured into 12 orthogonal patterns (Figure S8), while over half of this spatial structure was not clearly related to spatial patterns in the measured environmental factors (28% ‘pure’ spatial fraction; $p=0.001$). After removing the influence of spatially structured environmental gradients on phylogenetic composition, we could observe spatial patterns that seem independent of measured environmental variables and demand further investigation (Figure S9). For example, the first axis of the ‘pure’ spatial pattern of phylogenetic composition shows six geographically separated clusters of forests with similar phylogenetic composition (Figure S9). The measured environmental variables explained 24% of the total variation ($p=0.001$), most of which (19%; $p=0.001$) shared a spatial structure matching the spatial patterns of the phylogenetic compositional variation (i.e., joint space-environment fraction). This 19% of variation therefore corresponded to multiple spatial patterns in the environment that shape the phylogenetic composition of forests. The ‘pure’ environmental fraction explained 5% of the phylogenetic variation. The environmental variables with the most predictive power represented a mix of edaphic (soil concentrations of K+, Na+, Mg2+, N and soil pH$_{aq}$), climatic (maximum climatological water deficit, precipitation seasonality, temperature seasonality, mean annual precipitation) and topographic variables (topographic diversity index) (Table 2).

3.4 | Sensitivity analyses

The inclusion of more of the total variation in the phylogenetic composition described by the evoPCA axes (56 evoPCA axes – 75% of total variation) did not change the relative amount of variation explained by each fraction. However, it reduced the overall amount of variation explained overall by the model ($R^2_{adj}$ = 33%), mainly due to a reduction in the variation explained by the ‘pure’ spatial (15%) and joint space-environment (12%) fractions. Compared to the model, including 51% of the variation in phylogenetic composition (i.e., 13 evoPCA axes), the phylogenetic composition explained by the ‘pure’ environmental fraction increased by 1% but remained as the fraction with the smallest adjusted coefficients of determination.

The variance partitioning analysis using 100 phylogenies produced with scenario 2 of the V.phylomaker v.2 package shows similar results to those with Scenario 3. Despite a slight change in the adjusted $R^2$, the relative amount of variation in phylogenetic composition explained by each fraction (Figure S10) was robust to uncertainty in the phylogenetic placement of taxa.

3.5 | Indicator lineages for geographic regions and forest types

A total of 5096 lineages (72% of all the 7062 lineages evaluated) were significant indicators for at least one geographic region or edaphic forest type (Table S1). There were more indicator lineages for geographic regions (IA$_{geographic regions}$ = 4070 significant indicators) than for forest types (IA$_{edaphic forest type}$ = 1026 significant indicators) (Figure 3), although much of this difference is accounted for by the very high number of species (i.e., terminal nodes) that are indicators for geographic regions (2559 species are indicators for a geographic region versus 490 species for forest type). A greater number of higher-level evolutionary lineages (i.e., above the species level) were significant indicators for geographic regions (1511 higher-level evolutionary lineages are indicators for geographic region versus 536 for forest type).

Overall, a lower proportion of the species were significant indicators of geographic regions (50% of the 5082 species tested) or forest types (9.5%), while a greater proportion of higher-level lineages are such indicators (76% and 27% of the 1980 nodes tested for geographic regions or forest types, respectively) (Figure 3). The forests of the Guiana Shield and Northwestern Amazonia both show similar proportions of indicator lineages through time (Figure 4a). However, the Guiana Shield has fewer indicator lineages overall (Figure 3a). The Central and Southwestern Amazonia have a higher proportion of significant higher-level indicators at deeper timeslices (30% and 19% respectively – Figure 4a). Terra-firme forests and white-sand forests were the forest types with greater proportions of ancient indicator lineages determined as indicators (>66 Ma) (6 and 5%, respectively – Figure 4b). Among the wetland forests, the swamp and várzea forests show a greater proportion of ancient indicator lineages compared to igapó forests (respectively 4, 2 and 0% – Figure 4b).

In order to better understand and interpret the indicator lineage results, we selected the top 10 lineages, in terms of indicator value, associated with each geographic region and forest type. Where lineages in the initial top 10 were nested, we selected the one with

---

**Figure 2** First two axes of variation in the phylogenetic composition of tree communities across Amazonia. The maps show scores for the first two evoPCA axes together describing 21.6% of the overall variation. The colour of the circles reflects values for evoPCA scores with similar colours indicating tree communities with similar phylogenetic composition. The location of the communities was slightly jittered to reduce circles overlapping. Please see Figure S7 for maps for each of the first 13 evoPCA axes and Figure S8 for spatially constrained phylogenetic composition patterns corresponding to the significantly constrained axes from a redundancy analysis of the evoPCA axes against selected spatial eigenvectors (MEMs).
TABLE 2 The contribution of selected environmental predictors for the explained variation on the phylogenetic composition of tree communities.

<table>
<thead>
<tr>
<th>Variables</th>
<th>$R^2$</th>
<th>$R^2_{Cum}$</th>
<th>Adj $R^2_{Cum}$</th>
<th>$F$</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K^+$</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
<td>189.24</td>
<td>0.001</td>
</tr>
<tr>
<td>Maximum Climatological Water Deficit (MCWD)</td>
<td>0.03</td>
<td>0.11</td>
<td>0.11</td>
<td>61.79</td>
<td>0.001</td>
</tr>
<tr>
<td>$Na^+$</td>
<td>0.02</td>
<td>0.14</td>
<td>0.13</td>
<td>49.83</td>
<td>0.001</td>
</tr>
<tr>
<td>$pH_{aq}$</td>
<td>0.02</td>
<td>0.15</td>
<td>0.15</td>
<td>37.30</td>
<td>0.001</td>
</tr>
<tr>
<td>$Mg^{2+}$</td>
<td>0.01</td>
<td>0.17</td>
<td>0.16</td>
<td>32.72</td>
<td>0.001</td>
</tr>
<tr>
<td>Precipitation seasonality</td>
<td>0.01</td>
<td>0.18</td>
<td>0.18</td>
<td>34.14</td>
<td>0.001</td>
</tr>
<tr>
<td>Topographic diversity index</td>
<td>0.01</td>
<td>0.19</td>
<td>0.19</td>
<td>27.42</td>
<td>0.001</td>
</tr>
<tr>
<td>$N$</td>
<td>0.01</td>
<td>0.20</td>
<td>0.20</td>
<td>20.22</td>
<td>0.001</td>
</tr>
<tr>
<td>Temperature seasonality</td>
<td>0.01</td>
<td>0.21</td>
<td>0.20</td>
<td>20.42</td>
<td>0.001</td>
</tr>
<tr>
<td>Total annual precipitation</td>
<td>0.01</td>
<td>0.22</td>
<td>0.21</td>
<td>20.47</td>
<td>0.001</td>
</tr>
<tr>
<td>$Al^{3+}$</td>
<td>0.01</td>
<td>0.22</td>
<td>0.22</td>
<td>16.30</td>
<td>0.001</td>
</tr>
<tr>
<td>CEC</td>
<td>0.01</td>
<td>0.23</td>
<td>0.22</td>
<td>14.29</td>
<td>0.001</td>
</tr>
<tr>
<td>Temperature range</td>
<td>0.00</td>
<td>0.23</td>
<td>0.23</td>
<td>11.84</td>
<td>0.001</td>
</tr>
<tr>
<td>$Ca^{2+}$</td>
<td>0.00</td>
<td>0.24</td>
<td>0.23</td>
<td>7.99</td>
<td>0.001</td>
</tr>
<tr>
<td>Average temperature</td>
<td>0.00</td>
<td>0.24</td>
<td>0.23</td>
<td>5.32</td>
<td>0.001</td>
</tr>
<tr>
<td>Elevation</td>
<td>0.00</td>
<td>0.24</td>
<td>0.23</td>
<td>8.40</td>
<td>0.001</td>
</tr>
<tr>
<td>Groundwater table depth</td>
<td>0.00</td>
<td>0.24</td>
<td>0.24</td>
<td>5.48</td>
<td>0.001</td>
</tr>
<tr>
<td>$P$</td>
<td>0.00</td>
<td>0.24</td>
<td>0.24</td>
<td>3.63</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Note: Forward selection for a multidimensional model (redundancy analysis RDA) of 19 environmental variables adjusting the variation in the first 13 evoPCA axes scores. The total adjusted variation was 24%, of which 19% was spatially structured environmental variation (see variance partitioning results above).

FIGURE 3 The percentage of significant indicator species and higher-level evolutionary lineages for (a) geographic regions or (b) forest types. A total of 5082 species and 1980 higher-level evolutionary lineages were tested as indicators for geographic regions or forest types numbers inside bars correspond the totals of species and lineages determined as indicators for a given geographic region or forest type. Geographic regions: NWA – Northwestern Amazonia; CA – Central Amazonia; SWA – Southwestern Amazonia; GS – Guiana Shield; EA – Eastern Amazonia; SA – Southern Amazonia. Forest types: TF – Terra-Firme; PZ – White-Sand; VA – Várzea; SW – Swamp; IG – Igapó. For a complete list of significant indicator lineages, please refer to Table S1 and Figure S11 for the mapping of indicator lineages onto the phylogeny of Amazonian trees.
the highest indicator value, excluding the others, and then continued down the list of indicators sorted by indicator value, selecting additional lineages (that did not have a nested relationship with lineages already selected). As such our lists represent the top 10 non-nested indicator lineages for each geographic region and forest type (Table 3).

4 | DISCUSSION

We have provided the most comprehensive overview to date, to our best knowledge, of the phylogenetic composition of tree communities across the entirety of Amazonia leveraging detailed information on the abundance of lineages across c. 2000 forest plots, the most up-to-date phylogenetic hypothesis for Amazonian trees, numerous environmental variables and detailed spatial modelling. We found clear evidence for a strong influence of both geographic separation and environmental selection on the lineage composition of tree communities. We identified numerous evolutionary lineages as significant indicators of individual geographic regions and forest types on distinct edaphic regimes, even lineages older than 66 Ma. Given that the Amazonian flora began to take its present shape after the K-Pg mass extinction event 66 Ma (Carvalho et al., 2021; Jaramillo et al., 2006; Wing et al., 2009), it is all the more striking that the composition of tree communities today continues to bear the imprint of evolutionary associations to forest types and geographic regions that predate this event. Our results contradict the idea that dispersal has erased any signature of geography on the phylogenetic composition of Amazonian forests (Dexter et al., 2017). However, given the conflation of geography and environment in Amazonia, we suggest that the high number of indicator lineages for regions is more likely to reflect phylogenetically conserved niches than the signature of strong dispersal limitation over the long evolutionary history of individual lineages. Our results also contradict the idea that repeated transitions of lineages from one edaphic forest type to another, that is, edaphic-regime switching, have erased any signature of edaphic affiliation in lineages. Indeed, many lineages are significant indicators of individual edaphic forest types.

The first two axes of the evoPCA (Figure 1) provide a synthesis of our results showing how some forest types (white-sand and igapó forests) tend to retain a unique evolutionary signature (i.e., conserve their phylogenetic composition), whereas others are found in Amazonia (Guevara et al., 2016), whereas other forest types (terra-firme, várzea and swamp forests) shift in their phylogenetic composition across different regions of Amazonia. The broad-scale consistency of the phylogenetic composition of white-sand and igapó forests may result from the broad-scale consistency of soil structure and low-nutrient content on these older sandy habitats. Alternatively, or in addition, it may be that once a lineage specializes on the ecophysiological conditions of white-sand and igapó environments, it is difficult to then evolve adaptations to other environments, as it has been seen in drought-stressed tropical dry forests (Pennington et al., 2009). Meanwhile, the relatively younger and fertile soils of Western Amazonia host
terra-firme and várzea forests with similar phylogenetic composition, while in Central and Eastern Amazonia these forest types are more divergent in their phylogenetic composition. Such a pattern may result from the facts that: (i) terra firme forests in Western Amazonia tend to have soils of recent origin from Andean erosion, thus being more fertile soils (more similar to fertile soils in várzea forests) than those older and leached soils in Guiana Shield, Central and Eastern Amazonia and (ii) that várzea forests in Western Amazonia flood much less frequently than those in central and Eastern Amazonia (thus being more similar to terra firme forests in a lack of impact from flooding) (Terborgh & Andresen, 1998).

Our results confirm and tie together results from previous studies on the phylogenetic structure and composition of Amazonian tree communities that were carried out in more limited geographic regions or with less well-resolved phylogenies. For example, Honorio Coronado et al. (2015) found ubiquitous local phylogenetic clustering across Amazonia, which is in line with our results showing that unique lineages are hosted in different geographic regions and edaphic forest types. We have shown a clear evolutionary distinction among major types of Amazonian forests, and also along edaphic, and climatic gradients, as shown for forests in the Ecuadorian Amazonia (Guevara et al., 2017, 2021) and the distantly separated regions of Amazonian Peru and French Guiana (Baraloto et al., 2021). Overall, our results support the view that Amazonian forests show a clear ‘evolutionary signature’ that is organized at multiple spatial scales and by specific environmental conditions (Baraloto et al., 2021; Carlucci et al., 2017; Fine & Kemel, 2011).

The pervasive spatial structure we found in the phylogenetic composition of Amazonian tree communities can be interpreted as a signature of phylogenetic niche conservatism for regional and local environmental conditions. If dispersal limitation was the main process driving spatial patterns in lineage composition, we would expect the proportion of significant indicator lineages for geographic regions and edaphic forest types. We have shown a clear evolutionary distinction among major types of Amazonian forests, and also along edaphic, and climatic gradients, as shown for forests in the Ecuadorian Amazonia (Guevara et al., 2017, 2021) and the distantly separated regions of Amazonian Peru and French Guiana (Baraloto et al., 2021). Overall, our results support the view that Amazonian forests show a clear ‘evolutionary signature’ that is organized at multiple spatial scales and by specific environmental conditions (Baraloto et al., 2021; Carlucci et al., 2017; Fine & Kemel, 2011).

### Table 3: The top 10 higher-level lineages that are significant indicators for each geographic region or forest type.

<table>
<thead>
<tr>
<th>Geographic Region</th>
<th>Forest Habitat</th>
<th>Selection of top ten non-nested lineages associated with</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southern Amazonia</td>
<td>Terra-firme</td>
<td>Ocotea_subclade1; Copaifera+Hymenaea+Peltogyne; Trattinnickia_subclade1; Dialioideae; Dacryodes; Xylopia_subclade1; Inga_subclade38; Nyctaginaceae; Myrcia_subclade3; Ebenaceae</td>
</tr>
<tr>
<td>Eastern Amazonia</td>
<td></td>
<td>Carapa+Swietenia; Iryanthera+Otoha+Ostophleum; Virola_subclade2; Caesalpinioideae_mimosoid_subclade; Clusiaceae_subclade6; Papilionoideae; Arecaceae_subclade5; Gustavia+Grias; Cabralea+Guarea+Ruagea+Trichilia; Rutaceae</td>
</tr>
<tr>
<td>Central Amazonia</td>
<td></td>
<td>Salicaceae; Violacea+Goupia+Ritterae; Erycales; Euterpe; Asterales; Gentianales; Malpighiales_subclade3; Laurales_subclade2; Papilionoideae_subclade3; Myrtaceae_subclade3; Annonaceae_subclade5</td>
</tr>
<tr>
<td>Guiana Shield</td>
<td></td>
<td>Ochroma+Patinia+Septothesa; Scleronema; Cavannillesia; Dicorynia; Lecythys_subclade1; Fabales; Chrysobalanaceae_subclade2; Eschweilera+Lecythys; Aspidosperma; Pradosia</td>
</tr>
<tr>
<td>Northwestern Amazonia</td>
<td>Campanulids; Myristica+Clusiaceae_subclade1; Pereaebia; Rubiaceae_subclade18; Malpighiales_subclade6; Malvaceae; Mollia; Asterales+Dipsacales; Parkia; Detarioideae_subclade2</td>
<td></td>
</tr>
<tr>
<td>Southwestern Amazonia</td>
<td>Rosales; Arecaaceae_subclade1; Ulmaceae; Sebastiania+Sepium; Gloeospermum+Leoni+Amphirhox+Paypyrayola; Salicaceae_subclade1; Cabralea+Guarea+Ruagea; Inga_subclade3; Nectandra_subclade17; Malvales</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Terra-firme</td>
<td>Dacryodes+Trattinnickia+Protium; Eschweilera+Lecythiss+Coryphophora+Bertholletia; Pouteria+Chrysophyllum+Chromolucuma; Piptadenia_subclade5; Rinorea+Bribia; Olacaceae_subclade2; Rosales+Cucurbitales; Micropholis; Piperaceae+Canellaceae; Geissospermum</td>
</tr>
<tr>
<td></td>
<td>White-sand</td>
<td>Detarioideae; Malpighiales_subclade6; Clusiaceae_subclade1; Aspidosperma; Pachira_subclade2; Sapindaceae_subclade2; Chrysobalanaceae+Uphorbiaceae; Elaeoloma+Pouteria+Chrysophyllum; Calophyllaceae; Emmotum</td>
</tr>
<tr>
<td></td>
<td>Igapó</td>
<td>Gustavia+Grias; Campsiandra; Malouetia; Mollia; Eschweilera_subclade16; Cynometra; Crudia; Discocarpus+Amanoaa; Guatteria_subclade25; Amanoe</td>
</tr>
<tr>
<td></td>
<td>Várzea</td>
<td>Astrocaryum; Ceiba; Papilionoideae_subclade15; Polygonaceae; Caryophyllales; Canavilhesia+Scleronema+Catostemma; Batocarpus+Trophis+Bagassa+Sorocea; Spondias_subclade3; Garcinia_subclade1; Brassicales</td>
</tr>
<tr>
<td></td>
<td>Swamp</td>
<td>Mauritiella+Lepidocaryum+Mauritia; Magnoliidae; Clusiaceae_subclade6; Brassicales+Malvales; Ficus; Bignoniaceae_subclade1; Hieronyma+Richeria; Coussapoa; Guatteria_subclade32; Ilex</td>
</tr>
</tbody>
</table>

Note: Higher-level lineages were named for constituent genera or families, or if these are too numerous, as a subclade of a higher-level lineage (see Table S1 for a complete list of species and higher-level lineages determined as indicators).
regions to decrease over evolutionary time simply because older lineages would have had more time to disperse. However, we found that a high proportion of lineages remained significant indicators of geographic regions even in deeper evolutionary time (>30 Ma). While the geographic restriction of young lineages may (or may not) have resulted from dispersal limitation, it is unlikely that older lineages lacked time to disperse across Amazonia. We suggest that the geographic restriction of the older lineages (and maybe all lineages) is more likely a result of ancestral environmental niche preferences. While it might be obvious that different sub-regions across Amazonia have different environmental conditions, there is also variation across regions in the prevalence of otherwise widespread environmental conditions. The results of the variation partitioning analysis support our interpretation since 24% of phylogenetic compositional variation was explained jointly by environmental and spatial factors at multiple scales.

In addition, there is compelling evidence for a lack of long-lasting geographic barriers that could impede the dispersal of tree lineages throughout Amazonia (Dexter et al., 2017; Dick & Heuertz, 2008; Hoorn et al., 2022; Nazareno et al., 2017; Pos et al., 2023) and even if some geographic barriers have existed in Amazonia, they were likely to have been semi-permeable (Hoorn et al., 2022). Furthermore, local geomorphological changes can lessen the strength of previously existing barriers, allowing secondary contact among previously isolated lineages (Dexter et al., 2012; Honorio Coronado et al., 2019). Our results show that tree communities with similar phylogenetic composition can be found at opposite ends of Amazonia (Figure 2). This supports the hypothesis that the dispersal of lineages across space to find their optimal habitat is indeed possible (Baraloto et al., 2021).

Nevertheless, there is a substantial amount of variation (28%) in the phylogenetic composition of Amazonian tree communities that is significantly spatially structured but not related to the environmental variables in our analyses. Such spatial structure may relate to (i) the intrinsic dispersal capabilities of lineages and their ages (Dexter & Chave, 2016); (ii) extrinsic factors such as the presence of geographic and ecological barriers across Amazonia; (iii) important environmental variables not accounted for in our analyses; and/or (iv) spatially structured environmental variables that existed in the geological past but not at present. As an example, we can cite the marine incursions in Northwestern Amazonia during the Miocene (Hoorn et al., 2022), which may explain why some mangrove lineages were indicators for this region (e.g., Rhizophoraceae; Table S1). Distinguishing among these processes will require further research.

The correspondence between the phylogenetic composition of tree communities and edaphic forest types shows that ecological selection influences the distribution of Amazonian tree lineages (Figueiredo et al., 2022), also supported by the species and higher-level lineages that were indicators of forest types. Such habitat specialization seems to take place mainly in the upland forest habitats (terra-firme and white-sand forests) rather than in the flooded forests (várzea, swamp and igapó forests). For terra-firme, the high number of indicators can result, at least in part, from its much larger areal extent and higher overall diversity. White-sand forests are clearly exceptional in that a much higher proportion of lineages and species are white-sand forest indicators, even though white-sand forests cover less than 5% of Amazonia and their tree diversity is low (Adeney et al., 2016). Overall, though, the relatively low proportion of higher-level lineages (c.a. 27%) and species (c.a. 9%) that are significant indicators of forest types suggests that many Amazonian tree lineages have high plasticity, which allows them to occur in multiple forest types and on distinct edaphic regimes (Pitman et al., 1999). Other studies have reported a higher proportion of species to be habitat specialists, reaching up to 53% among the 227 most abundant Amazonian tree species (ter Steege et al., 2013), while around 20% have been reported as várzea forests specialists (Wittmann et al., 2013). In contrast, a more recent study has shown that only a small percentage of all Amazonian tree species are likely to be restricted to a single forest type (Oliveira-Filho et al., 2021), echoing earlier results for specific geographic regions (Pitman et al., 1999) and forest types (García-Villacorta et al., 2016). Regardless, the low percentage of significant indicators for edaphic forest types that we found may be due, at least in part, to variation in the environmental characteristics of forest types within and particularly among Amazonian regions. For example, várzea in the central Amazon has a very distinct flooding regime from várzea in the western and eastern Amazonia (Fassoni-Andrade et al., 2021) with different tree lineages likely specializing on these distinct regimes. Moreover, the rarity of most tree species (ter Steege et al., 2013, 2019) makes it difficult to achieve significance levels for such species. Our simplistic characterization of edaphic forest types enables a synthetic Amazon-wide analysis of indicator lineages but future work should delve deeper into the region-specific environmental associations of different evolutionary lineages of trees.

The role of speciation in driving our results also merits consideration. The fact that many lineages show strong affiliations with individual geographic regions suggests that much speciation has happened within geographic regions. Given how large the regions are that we have studied, this does not preclude a role for allopatric or peripatric speciation in Amazonian trees (Dick & Pennington, 2019). Meanwhile, the relatively low number of significant indicators for edaphic forest types suggests that at least some lineages may be speciating via edaphic-regime switching (Figueiredo et al., 2022; Fine et al., 2005; Glassmire et al., 2017). In contrast, the consistent phylogenetic composition of white-sand forests can be interpreted as a signature of lineage speciation within an edaphic regime with an archipelago-like distribution across the Amazon (Costa et al., 2020; García-Villacorta et al., 2016; Guevara et al., 2016). In any case, processes such as speciation are challenging to assess in a study of the entire angiosperm tree flora of Amazonia and where our phylogeny only genuinely samples c.a. 30% of species, with genetic markers that generally do not accurately resolve species-level relationships. Clearly, there needs to be a major focus in establishing resolved species-level phylogenies for important Amazonian tree lineages.
Overall, our analytical framework allowed us to explain just over half of the variation in the dominant axes that together explain half the variation in the phylogenetic composition of Amazonian tree communities. These dominant axes were those that represented more variation in phylogenetic composition than expected by chance. When we expanded our variation partitioning analyses to try and explain axes that represent 75% of the variation in phylogenetic composition, our total variation explained dropped by 1/3. Essentially, we could not explain any of the variation in these additional axes with the measured environmental variables and our decompositions of spatial structure. As a visual reflection of unexplained variation, we can see that some clusters of tree communities with heterogeneous phylogenetic composition are spatially proximate and in apparently similar environments (Figure 2). Overall, we suggest that the large amount of unexplained variation in our compositional data, while not usual, could be due to disturbance events, neutral ecological drift, non-linear relationships between the distribution of lineages and geographic and environmental gradients, and to other, poorly understood phenomena.

5 | CONCLUSION

The phylogenetic composition of Amazonian tree communities varies over multiple spatial scales, as a function of geographic, edaphic and climatic gradients. Throughout Amazonia different lineages are indicative of geographic regions and forest types on distinct edaphic regimes. Such preferences imprint a spatial structure on the phylogenetic composition of tree communities. Whereas terra-firme and wetland forests show a more variable phylogenetic composition across Amazonia, white-sand forests tend to conserve a phylogenetic composition regardless of the geographic region where they were found. Our results support a role for dispersal and selection processes in structuring the evolutionary assembly of tropical tree communities. The incorporation of an evolutionary lens contributes to improving our knowledge on variation in the composition of organisms assembling the forests of Amazonia and highlights the complex evolutionary history of trees over multiple spatial scales and environmental conditions.

AFFILIATIONS
1Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas – UNICAMP, Campinas, Brazil
2AMAP, Université de Montpellier, CIRAD, CNRS, INRAE, IRD, Montpellier, France
3Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford, UK
4Laboratoire d'Ecologie végétale et Biogéochimie, Université libre de Bruxelles, Brussels, Belgium
5Naturalis Biodiversity Center, Leiden, The Netherlands
6Quantitative Biodiversity Dynamics, Utrecht University, Utrecht, The Netherlands
7Coordenação de Biodiversidade, Instituto Nacional de Pesquisas da Amazônia – INPA, Av André Araújo, Manaus, Brazil
8Programa Professor Visitante Nacional Sênior na Amazônia – CAPES, Universidade Federal Rural da Amazônia, Belém, Brazil
9Coordenação de Botânica, Museu Paraense Emílio Goeldi, Belém, Brazil
10Wetland Department, Institute of Geography and Geocology, Karlsruhe Institute of Technology – KIT, Rastatt, Germany
11Ecology, Monitoring and Sustainable Use of Wetlands (MAUA), Instituto Nacional de Pesquisas da Amazônia – INPA, Manaus, Brazil
12Centro de Pesquisa Agroflorestal de Roraima, Boa Vista, Brazil
13Departamento de Botânica, Instituto de Pesquisas Científicas e Tecnológicas do Amapá – IEPA, Macapá, Brazil
14Grupo de Investigación en Biodiversidad, Medio Ambiente y Salud-BIOMAS, Universidad de las Américas, Quito, Ecuador
15Keller Science Action Center, Chicago, Illinois, USA
16School of Geography, University of Leeds, Leeds, UK
17Coordenação de Pesquisas em Ecologia, Instituto Nacional de Pesquisas da Amazônia – INPA, Manaus, Brazil
18Anel Amapá Florestal e Celulose SA, Belém, Brazil
19Herbario Amazónico Colombiano, Instituto SINCHI, Bogotá DC, Colombia
20Science and Education, Chicago, Illinois, USA
21División de Sensoramiento Remoto – DSR, Instituto Nacional de Pesquisas Espaciais – INPE, São José dos Campos, Brazil
22Herbario Vargas, Universidad Nacional de San Antonio Abad del Cusco, Cusco, Peru
23Biological and Environmental Sciences, University of Stirling, Stirling, UK
24Centro de Biociencias, Departamento de Ecología, Universidade Federal do Rio Grande do Norte, Natal, Brazil
25Departamento de Biologia, Universidade Federal de Rondônia, Porto Velho, Brazil
26Programa de Pós-Graduação em Biodiversidade e Biotecnologia PPG-Bionorte, Universidade Federal de Rondônia, Porto Velho, Brazil
27Department of Biology and Florida Museum of Natural History, University of Florida, Gainesville, Florida, USA
28Centre for Tropical Environmental and Sustainability Science and College of Science and Engineering, James Cook University, Cairns, Queensland, Australia
29Instituto de Investigaciones de la Amazonia Peruana (IIAP), Iquitos, Peru
30School of Geography and Sustainable Development, University of St Andrews, St Andrews, UK
31Jardín Botánico de Missouri, Oxapampa, Peru
32Instituto Boliviano de Investigacion Forestal, Santa Cruz, Bolivia
33Geography, College of Life and Environmental Sciences, University of Exeter, Exeter, UK
34Agteca-Amazonica, Santa Cruz, Bolivia
35Programa de Pós-Graduação em Ecologia e Conservação, Universidade do Estado de Mato Grosso, Nova Xavantina, Brazil
36Facultad de Ciencias Agrícolas, Universidad Autónoma Gabriel René Moreno, Santa Cruz, Bolivia
37Natural History Museum, University of Oslo, Oslo, Norway
38International Center for Tropical Botany (ICTB) Department of Biological Sciences, Florida International University, Miami, Florida, USA
39Caird UMR Ecololog AgrosParisTech, CNRS, INRAE, Univ Guyane, Kourou Cedex, France
40Centro de Investigaciones Ecológicas de Guayana, Universidad Nacional Experimental de Guayana, Ciudad Bolívar, Venezuela
41Empresa Recursos Genéticos e Biotecnologia, Parque Estação Biológica, Prédio da Botânica e Ecologia, Brasilia, Brazil
42Projeto Dinâmica Biológica de Fragmentos Florestais, Instituto Nacional de Pesquisas da Amazônia – INPA, Manaus, Brazil
43Programa de Pós-Graduação em Ecologia, Instituto Nacional de Pesquisas da Amazônia – INPA, Manaus, Brazil
44Laboratório de Ecologia de Doenças Transmissíveis da Amazônia (EDTA), Instituto Leônidas e Maria Deane, Rio de Janeiro, Brazil
45Programa de Pós-graduação em Biodiversidade e Saúde, Instituto Oswaldo Cruz – IOC/FIOCRUZ, Rio de Janeiro, Brazil
46Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém, Brazil
47Programa de Pós-Graduação em Ecologia, Universidade Federal do Pará, Belém, Brazil
48Empresa Brasileira de Pesquisa Agropecuária, Embrapa Amazônia Oriental, Belém, Brazil
This work has been supported by the following Brazilian research agencies: FAPESP, CAPES, CNPq.

CONFLICT OF INTEREST STATEMENT
None to declare.

DATA AVAILABILITY STATEMENT
Data is available upon request following ATDN agreements (https://sites.google.com/naturalis.nl/amazon-tree-diversity-network/homepage?authuser=0).

ORCID
Bruno Garcia Luize https://orcid.org/0000-0002-8384-8386
Juan Ernesto Guevara https://orcid.org/0000-0002-5433-6218
John Terborgh https://orcid.org/0000-0002-8238-4583
Euridice N. Honorio Coronado https://orcid.org/0000-0003-2314-590X
Pablo R. Stevenson https://orcid.org/0000-0003-2394-447X
Vitor H. F. Gomes https://orcid.org/0000-0002-3855-5584
Joseph E. Hawes https://orcid.org/0000-0003-0053-2018
Juliana Stropp https://orcid.org/0000-0002-2831-4066
Angela Cano https://orcid.org/0000-0002-5090-7730
Maria Natalia Umaña https://orcid.org/0000-0001-5876-7720
Kyle G. Dexter https://orcid.org/0000-0001-9232-5221

REFERENCES


**BIOSKETCH**

Bruno Garcia Luize is an ecologist interested in the temporal and spatial patterns and the ecological and evolutionary processes sustaining the biodiversity on Earth. The ATDN comprises a diverse array of researchers with common interests ranging from physical geography, natural history, botany, ecology, evolution and the conservation of life on Earth, with a special focus on Amazonia.

**Author Contributions:** All co-authors conducted the fieldwork and collected data with additional material from collaborators; BGL, KGD, DB and HtS conceived the ideas and methodological approach and conducted the analyses; BGL led the writing with assistance from KGD, DB, CP and HtS; all authors provided data and had the opportunity to comment on the text; all authors approved the final version of the manuscript.

**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.