A Spatial Signal of Niche Differentiation in Tropical Forests

Mihir S. Umarani,[1,2,*,†] Dianzhuo Wang,[3,†] James P. O'Dwyer,[4] and Rafael D'Andrea[1]

1. Department of Ecology and Evolution, Stony Brook University, Stony Brook, New York 11794;   2. Institute for Cross-Disciplinary Physics and Complex Systems, University of Balearic Islands, E-07122 Palma, Spain;   3. Department of Physics and John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138;   4. Department of Plant Biology, School of Integrative Biology, University of Illinois, Urbana, Illinois 61801

ABSTRACT: Explaining diversity in tropical forests remains a challenge in community ecology. Theory tells us that species differences can stabilize communities by reducing competition, while species similarities can promote diversity by reducing fitness differences and thus prolonging the time to competitive exclusion. Combined, these processes may lead to clustering of species such that species are niche differentiated across clusters and share a niche within each cluster. Here, we characterize this partial niche differentiation in a tropical forest in Panama by measuring spatial clustering of woody plants and relating these clusters to local soil conditions. We find that species were spatially clustered and the clusters were associated with specific concentrations of soil nutrients, reflecting the existence of nutrient niches. Species were almost twice as likely to recruit in their own nutrient niche. A decision tree algorithm showed that local soil conditions correctly predicted the niche of the trees with up to 85% accuracy. Iron, zinc, phosphorus, manganese, and soil pH were among the best predictors of species clusters.

Keywords: tropical forests, spatial clusters, niche differentiation, Barro Colorado Island, soil nutrients, machine learning.

Introduction

The role of niche differentiation in maintaining high species diversity in tropical forests remains unclear. Given the difficulty of performing experiments in these systems, attempts to glean niche differentiation largely rely on associating species' spatial distributions with local environmental variables (Clark et al. 1998; Phillips et al. 2003; Paoli et al. 2006; John et al. 2007; Baldeck et al. 2013a). However, these analyses are complicated by (i) a lack of clear theoretical expectations linking niche differentiation in its various forms to community-level patterns and (ii) the scale dependence of the underlying mechanisms (Chase 2014). For example, competition for local resources may drive com-

munity composition at local scales, while dispersal limitation and selective recruitment by the environment (habitat filtering) may dominate at larger scales. Indeed, habitat filtering is often considered most relevant at regional scales where environmental heterogeneity is large (Cody 1974; Connor and Simberloff 1979). However, tropical soils can be highly heterogeneous at ~10-m scales (John et al. 2007; Baldeck et al. 2013a), suggesting that habitat filtering may also be an important process at local scales if species are differently adapted to different soil conditions.

At local scales, where the impact of competition can be strong, classical theory establishes that coexistence requires niche differentiation (Dayan and Simberloff 2005). This runs into the problem of explaining coexistence among hundreds of species that share the same few limiting resources and thus presumably the same few niches. On the other hand, species similarities slow down competitive exclusion (Chesson 2000; Scheffer and van Nes 2006; Holt 2009), which can be enough for immigration and other mechanisms to stabilize coexistence (Barabás et al. 2013; D'Andrea et al. 2019). Combining these two ideas has led to the proposal that species in diverse communities may coexist via partial niche differentiation (i.e., by being different enough to partition ecological niches but similar enough within each niche to indefinitely prolong competitive exclusion; D'Andrea and Ostling 2017; D'Andrea et al. 2019). Thus, competing species in diverse communities should segregate into several clusters along the niche axes related to ecological determinants (Hutchinson 1957; Holt 2009), such as environmental conditions or functional traits. For example, a cluster may consist of species sharing similar abiotic preferences or having similar functional traits associated with said preferences. Recent reports of such clusters in the empirical literature include tropical tree species clustering in terms of height (D'Andrea et al. 2020) and other functional traits (Rubio and Swenson 2022), phytoplankton species clustering in body mass (Graco-Roza et al. 2021),

* Corresponding author; email: mihir.umarani@gmail.com.
† These authors contributed equally to this work.
   ORCIDs: Umarani, https://orcid.org/0000-0001-5509-8170; D'Andrea, https://orcid.org/0000-0001-9687-6178.

and temperate trees clustering in space and functional traits related to growth (Ulrich et al. 2021).

In a spatial context, the partial niche differentiation paradigm predicts that (i) species sharing environmental niches in a heterogeneous landscape will be spatially clustered, (ii) these species clusters can be associated with environmental niches by matching the spatial distribution of the clusters to that of environmental variables, and (iii) the biological mechanisms underlying those niches can be ascertained by analyzing their association to species' functional traits (D'Andrea et al. 2020).

Empirical testing of this hypothesis faces additional challenges. First, the spatial scale of the analyses critically affects the inference (Chase 2014). At regional scales, it can be difficult to infer species interactions and assess their influence based on co-occurrence data alone (Barner et al. 2018; Blanchet et al. 2020). At local scales, low levels of environmental heterogeneity conditions might let dispersal limitation drive spatial aggregation of species (Dalling et al. 2002; Muller-Landau et al. 2008). Additionally, inference of niche partitioning may be sensitive to the choice of environmental variables and their respective levels of spatial heterogeneity. For example, previous studies found little support for niche differentiation of species across slope and elevation in a Panamanian forest (Harms et al. 2001), whereas recent analyses showed that many species in the same forest are nonrandomly associated with soil nutrient concentrations (John et al. 2007), which are strongly heterogeneous across the plot and explain beta diversity in 30%–40% of plant species (John et al. 2007; Baldeck et al. 2013*a*; Wolf et al. 2015). Although these analyses show that soil nutrients explain community-level aggregation patterns to some extent, the species composition of each niche, the level of niche differentiation, and how soil variables characterize those niches remain unclear.

Here, we test the spatial clustering hypothesis in a lowland semideciduous moist tropical forest in Panama. We address four sets of questions.

1. *Species clustering in space.* How strong is spatial clustering among woody plant species? How sensitive is the pattern to the spatial scale of analysis? How robust is the pattern over time?

2. *Strength of niche partitioning.* To what extent do individuals preferentially recruit within their inferred niches?

3. *Association with soil variables.* How are distinct species clusters spatially associated with soil nutrients?

4. *Association with functional traits.* How do species clusters differ in terms of their functional traits?

We find strong spatial clustering that is robust to the spatial scale of analysis. The four spatial clusters are associated with different levels of soil variables, mainly phosphorus, zinc, nitrogen, and manganese, and also with soil

pH and water potential. This association indicates that those spatial clusters reflect the presence of spatial niches—ecological niches characterized by spatially segregated environmental variables. Furthermore, species recruited within their own niche almost twice as frequently as in other areas. Species were also clustered by functional traits, but those trait clusters did not associate strongly with the spatial clusters. We corroborated those results in an Andean cloud forest in Colombia, where we found three spatial clusters associated with soil nutrients.

Methods

Datasets

We used the tree census dataset from field surveys conducted on a 50-ha tropical forest dynamic plot (FDP) on Barro Colorado Island (BCI), a lowland tropical moist forest in Panama (Condit 1998, 2019). The plot shows high tree species richness with ~300 woody plant species recorded and tagged. There were eight censuses available for BCI dating from 1985 to 2019, where all freestanding trees larger than 1 cm in diameter at breast height (dbh) are mapped and identified to species, and their dbh is measured. We also used data from a single census (from 2002) from a 25-ha FDP in La Planada, a high-elevation Andean cloud forest in Colombia, which had ~200 woody species (Vallejo-Joyas et al. 2005).

BCI soil is primarily categorized as oxisol, acidic with very high levels of aluminum and iron while most of the phosphorus is fixed or stored in plant biomass (Yavitt 2000; John et al. 2007; Baillie et al. 2009). Nutrient levels were estimated at a 20 × 20-m spatial resolution from soil samples taken on a regular 50 × 50-m plot (Hubbell et al. 2005; Condit et al. 2013). The list of measured soil resources includes the concentrations of elemental Al, B, Ca, Cu, Fe, K, Mg, Mn, Zn, mineralized N, and P. Additionally, data for soil pH and soil water potential (during dry seasons of the year) were also used. Details of the geostatistical methods used for sampling and interpolating local nutrient levels in the plot are found in Hubbell et al. (2005), John et al. (2007), and Kupers et al. (2019). La Planada soil is not similar to that of BCI and is categorized as andisol, formed in volcanic ash with very poor levels of plant-available phosphorus (Baldeck et al. 2013*a*; Dalling and Norden 2023).

Data Filters

We focused our clustering analysis on juvenile and adult trees in the plot (as opposed to seedlings and saplings), reasoning that selection by the local environment largely operates before the onset of the juvenile stage. We assigned life history stage status based on species-specific dbh cutoffs:

stems with greater than 1 cm dbh and below the 56th quantile in each species were considered saplings, stems between the 56th and 88th quantiles were considered juveniles, and stems above the 88th quantile were considered adults. We selected these species-specific cutoffs based on an empirical observation that they represented the relative proportions of juveniles to adults across species in this forest (Baldeck et al. 2013a). We also removed species that may not be found near each other because of rarity rather than disparate environmental niches, by applying the following abundance cutoff. Given two random uniformly distributed species with abundances n in a plot of area A and defining neighbors as trees within a distance d of each other, we expect to find $\pi d^2 n^2 / A$ interspecific pairs of neighbors across the plot. Setting this expectation to at least one pair gives us an abundance cutoff of $\sqrt{A/\pi d^2}$. For example, for $d = 20$ m and a 50-ha plot, this corresponds to a minimum abundance of 20 trees. In short, species with population size greater than 20 can be expected to have at least one heterospecific neighbor in the absence of niche partitioning. This ensures that the rare species will not bias the estimate of niche differentiation. This left about 83,000 trees per census on the BCI plot, spanning about 320 species.

Analysis

Data analyses and simulations were performed in R (ver. 4.0.5; R Core Team 2018).

Spatial Clustering Analysis

Our analysis was based on finding species that tend to occur near each other. We first determined the number of close neighbors to each tree within the plot. A close neighbor was defined as a tree within a cutoff distance d of the focal tree. This neighborhood cutoff was selected according to the observation that trees exert competitive effects on one another's growth within a range of 5–25 m (Uriarte et al. 2004). Therefore, we used 20 m as a distance cutoff for the neighborhood. We further verified that our species clustering results were robust to the choice of neighborhood cutoff within a range of 10–30 m (figs. S1, S3; figs. S1–S14 are available online), when tested on either the empirical data or under the simulation exercise (described in the following paragraphs). We then created an adjacency matrix, a square matrix with rows and columns representing species and whose entries specify the total neighbor count between each species pair divided by the maximum possible neighbor count given the species' respective abundances. The entries of this matrix thus quantify the degree of spatial proximity between species pairs. We tested spatial clustering of the species using the maximum modularity metric. For a given division of the full set of species into groups or

modules, the resulting modularity of the matrix measures the extent to which species in the same group tend to occur in mutual proximity while being rarely observed near species in other groups. Formally, the modularity of a given grouping is calculated by summing the same-group entries of the adjacency matrix for each group, subtracting the expected sum if matrix entries were assigned at random, and averaging those differences across groups (Newman 2004). We used R package igraph to find the grouping that maximizes modularity and number of clusters in a community (Csárdi et al. 2023). Specifically, we used the Louvain algorithm, which works efficiently with large datasets (Emmons et al. 2016). We established the statistical significance of maximum modularity values by comparing them with replicates of the forest where species' identities are randomized across individuals while maintaining their geographic coordinates. This corresponds to a null hypothesis where suitable locations for tree recruitment are maintained but species have no bias for any particular location.

To each species cluster we attributed a spatial niche—a geographic region with optimal abiotic conditions—and inferred the geographic locations of such spatial niches on the FDP via kernel density estimation using R package sparr (Davies 2018). Assuming that spatial niches are distributed continuously in space on the FDP, we used the empirical local density of trees of each cluster to find the smoothed probability density of each corresponding spatial niche across the landscape. Each 20 × 20-m quadrat in the FDP was then assigned a probability of containing each of the n spatial niches. We repeated the analysis for the La Planada dataset.

Recruitment Analysis

With the multicensus data from BCI, we performed recruitment analysis to quantify the degree of niche differentiation. We defined this quantity as the degree to which plants preferentially recruit in their optimal soil condition (i.e., their niche) relative to recruitment in other soil conditions. First, we performed clustering analysis and kernel density estimation for each cluster on juveniles (classified as trees with dbh between 56th and 88th percentiles of the species' population) and adults (trees with dbh above 88th percentile of the species' population) from the first census on BCI. We then defined the geographic distributions of each cluster as putative spatial niches and recorded the cluster membership for each species. For the subsequent censuses, we tested the frequency with which unaccounted adults (which grew from individuals classified as juveniles from previous censuses) of a given species occurred within their species' putative spatial niche. The proportion of recruitment events matching an individual's niche relative to null expectations measures the degree of niche differentiation, which we

referred to as θ in the analysis. The null hypothesis is $\theta = 1$, denoting that species recruit in their own niche in proportion to the area occupied by that niche. We confirmed that dispersal limitation does not alter this null hypothesis by simulating random birth-death process events of individuals in a community of richness and composition identical to that of BCI, where each species was assigned a dispersal kernel using a truncated Gaussian function, representing strong dispersal limitation. We also tested the recruitment success across different life history stages (i.e., from seedlings to saplings and from saplings to juveniles). We did this because filtering by soil variables may not act until the late stages of individual growth, while dispersal limitation can affect the spatial aggregation patterns at early life stages. Thus, differences in recruitment across life stages can indicate the relative impacts of dispersal limitation and niche partitioning. Details about the calculation of θ are provided in appendix 1 (apps. 1–6 are available online in the supplemental PDF).

Soil Nutrient Analysis

Covariation and Spatial Clustering in Soil Nutrients. To understand the spatial structure and covariance patterns in soil resources, we performed a principal component analysis and a k-means clustering analysis on measured soil resource levels. For both analyses, we centered and scaled all resources by their standard deviation. We chose the number of principal components that cumulatively explained more than 85% of the variance. For k-means clustering analysis, we used a gap statistic to evaluate an optimal number of clusters of soil resource concentrations (D'Andrea et al. 2020). We compared the geographic distributions of these clusters with those of the species spatial clusters.

Association between Species Clusters and Soil Nutrients. If our species clusters reflect preferential recruitment in soil with different local conditions, there must be a spatial association between local resource levels and our inferred spatial niches. We tested for this association in two ways. First, we calculated the Pearson correlation coefficient between the local levels of each nutrient and the inferred density of each spatial niche. If niche differentiation by specializing in different soil conditions is a dominant process on BCI, then the species clusters should have distinctive resource correlation profiles.

Second, we used supervised machine learning to measure the extent to which local nutrients predict the species clusters. High predictive power would signify a tight link between nutrients and species clusters and thus support the idea that species clusters indeed represent distinct environmental niches. First, we labeled each 20 × 20-m quadrat with the niche with highest local probability as per

our kernel density estimator. Next, we trained a decision tree/rule-finding algorithm to predict the species clusters based on local nutrient levels. We used the C5.0 classifier in R package caret (Kuhn 2021). This algorithm balances interpretability of results with accuracy and predictive power (e.g., Hidayat et al. 2017; Abdullah et al. 2021). The algorithm also downweighs explanatory variables that do not increase the information gain significantly, thus indicating the soil nutrients with which species are associated most strongly (see further details about the C5.0 algorithm in app. 2). We assessed quality of predictions using Cohen's κ, $\kappa = (p_0 - p_e)/(1 - p_e)$, an index of classifier performance that considers both its observed (p_0) and expected (p_e) accuracy levels. For example, a random classifier choosing between four equiprobable categories would be correct 25% of the time. An observed accuracy of 40% would result in $\kappa = 0.2$, suggesting poor agreement, whereas 90% accuracy would yield $\kappa = 0.87$, considered excellent agreement (Landis and Koch 1977; Warrens 2015). For this analysis, we used functions from the R packages C50 and caret (Kuhn 2021; Kuhn and Quinlan 2022). Because both the species clusters and the nutrient clusters are spatially autocorrelated, we assessed significance of the result by comparing it against autocorrelated randomized data (see details in app. 2). For both analyses, we used the 2010 BCI census (census 7), since it was the closest following the soil survey, taken in 2005–2006. We repeated the analysis for the La Planada dataset.

Validation Analysis

To verify that our method successfully captures the niche structure of species segregating by preferences for local environments, we validated our method on simulation outcomes of a spatially explicit birth-death process on a simulated heterogeneous landscape where recruitment of plants within a certain group is more likely to occur in sites with their preferred soil type, or niche, than in mismatching environments. We then ran our spatial clustering analysis and recruitment analysis to verify that the resulting number of clusters and the degree of niche differentiation θ matched the true values. Further details of the model are provided in appendix 3.

Trait Analysis

We examined the link between our inferred clusters and plant functional traits using two sets of plant traits from BCI: morphological/growth rate traits (Wright et al. 2010) and foliar elemental concentrations (Wright 2023). The first set of traits represented five key trait axes: sapling growth and mortality rates, wood density, leaf mass per area, maximum

plant height, and seed size (see the description in app. 4). Since sapling growth and mortality are strongly correlated (Wright et al. 2010), we combined them into a single axis given by the first principal component of those rates. We performed pairwise Mann-Whitney tests of the distribution of each trait across spatial clusters. Similarly, we tested how foliar concentrations of key elements (Al, Ca, K, Mg, Mn, Zn, C, N, P, δ^{15}N, and δ^{13}C) differed across spatial clusters.

In addition, for both morphological/growth traits and foliar stoichiometry, we tested whether species were clustered in trait space. In each case, we first reduced the dimensionality of trait space by retaining the first two principal components (which explained 64% of the variance in morphological/vital rate traits and 41% in foliar stoichiometry traits). We then determined whether clustering occurred using a combination of the gap statistic (Tibshirani 2001; D'Andrea et al. 2019) and k-means clustering weighted by log abundance. To run k-means clustering, we used a modified Hartigan-Wong algorithm from R package FactoClass (Pardo and DelCampo 2007). We then checked whether the spatial distribution of those trait clusters matched that of the clusters found in the spatial clustering analysis. Overlaps in the geographic distribution of trait clusters and the spatial clusters can help understand what role key traits play in environmental filtering (Rubio and Swenson 2022). The R codes for our validation model and our clustering method and data analysis are available on figshare (https://doi.org/10.6084/m9.figshare.24539419; Umarani et al. 2023).

Results

Species Spatial Clustering

We found four species clusters on BCI, whereby species within each cluster tend to occur within 20 m (neighborhood distance cutoff) of each other compared with other clusters. The number of clusters and their spatial distribution were largely consistent across all censuses and different neighborhood distance cutoffs (figs. 1, S1). The number of species within each cluster was not uniform, but no cluster contained a majority (>50%) of the species (app. 5). Our recruitment analysis showed that an adult recruits in its own spatial niche 1.7 times more often than expected from random birth-death events and dispersal limitation (fig. S2). This recruitment bias was strong and consistent across life stages (saplings, juveniles, and adults; fig. S3). Our finding of a limited number of robust spatial clusters on BCI was corroborated by the analysis on the La Planada forest, where the species were split into three clusters (fig. S4).

Soil Nutrients

Soil resource levels were highly heterogeneous in the BCI plot. The first three principal components explained more than 70% of the variation in soil resource concentrations in all three plots. This, along with the loadings of nutrients across the principal components axes, indicates strong covariation in spatial distribution of these nutrients (fig. S5). The k-means clustering analysis suggests four contiguous
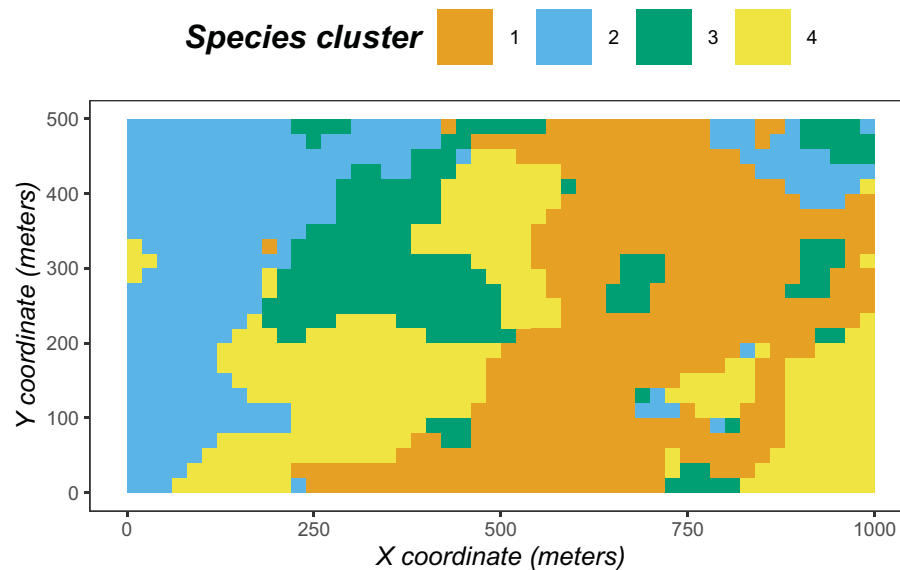

Figure 1: Spatial distribution of inferred species clusters in Barro Colorado Island (census 7). The unit across both *x*- and *y*-axes is meters. The figure reflects the total area of the forest dynamic plot, which is a rectangular 1,000 × 500-m plot with the following coordinates: SW 9.152096, −79.854984, NW 9.157010, −79.846745 (Condit 2019). A color version of this figure is available online.
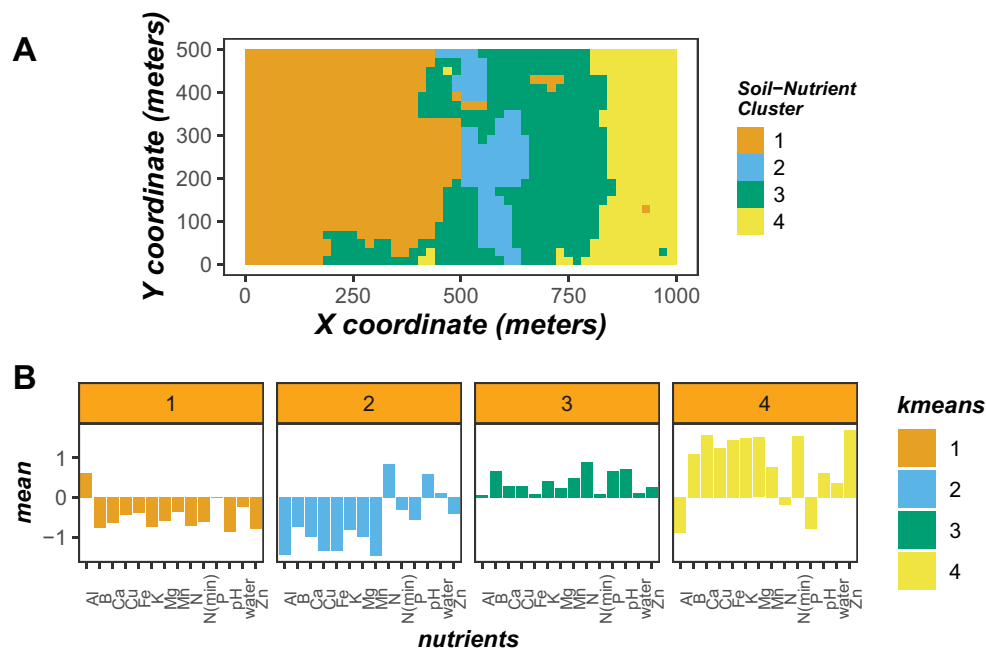
Figure 2: Soil nutrient clustering patterns in Barro Colorado Island. *A*, Spatial distribution of soil nutrient clusters. *B*, Mean concentration coefficients of soil nutrients in each soil cluster. A color version of this figure is available online.

clusters showing clear geographic segregation and matching the number of species clusters (figs. 2, 3), although the gap statistic failed to indicate statistical significance (see the detailed discussion in app. 6). Those putative nutrient clusters are also geographically associated with some of the major topographic habitat types on BCI, where the stream and slope between high and low plateau described the boundaries between nutrient clusters (fig. 1; Harms et al. 2001). Similarly, soil nutrients in La Planada showed three *k*-means clusters exhibiting strong spatial segregation (fig. S6) and roughly matching north-south split of spatial clusters.

Association between Soil Nutrients and Species Clusters

The correlations between species cluster densities and soil resource levels revealed that each cluster had unique associations with various nutrients (figs. 3, S7). On BCI, cluster 1 differed from the rest of the clusters in its association with the highly correlated group of elements (B, Ca, Cu, K, Mg, Mn, Zn, and mineralized N) and Al. Cluster 1 further differentiated from clusters 2 and 4 in its association with pH, P, and N. Clusters 2 and 3 differed in their association with Fe and standing water column. Clusters 3 and 4 differed in their association with Mg and P. The C5.0 analysis revealed strong explanatory power of soil nutrients in predicting spatial clusters. Cohen's κ was high as well (0.77), which was significantly higher than in the autocorrelated null datasets (~0.3; table 1). The average depth (i.e., num-

ber of rules) of resulting decision trees was very high (~67) even after applying winnowing of attributes, indicating complex relationships between nutrients and species clusters. Once we manually limited the depth of the decision trees by controlling the algorithm presets, a small subset of the soil nutrients successfully classified a large portion of the data. On BCI, those were Fe, Zn, N, P, and water levels. On La Planada (Cohen's κ, 0.61), Zn, Mg, K, and soil pH were the best predictors of spatial clusters.

Traits

Morphological and Growth Traits. Spatial clusters of BCI species did not differ strongly in any of the morphological and growth-related traits (seed size, plant maximum height, leaf mass per area, wood density, and vital rates; see fig. S8). However, the gap statistic *k*-means clustering analysis revealed four groups of species spanning a gradient from acquisitive species (low leaf mass per area and wood density, small seeds, and fast growth/high mortality) to conservative species (opposite traits; fig. 4). The spatial distribution of those trait-based groups matched that of the soil-based spatial clusters well (compare fig. S9*A* with fig. S9*B*): quadrats where particular trait clusters had high density relative to the other trait clusters also tended to harbor high density of particular spatial clusters (Cramer's $V = 0.40$, 95% confidence interval [CI] = 0.37–0.43), indicating good spatial correspondence between the trait and spatial clusters.
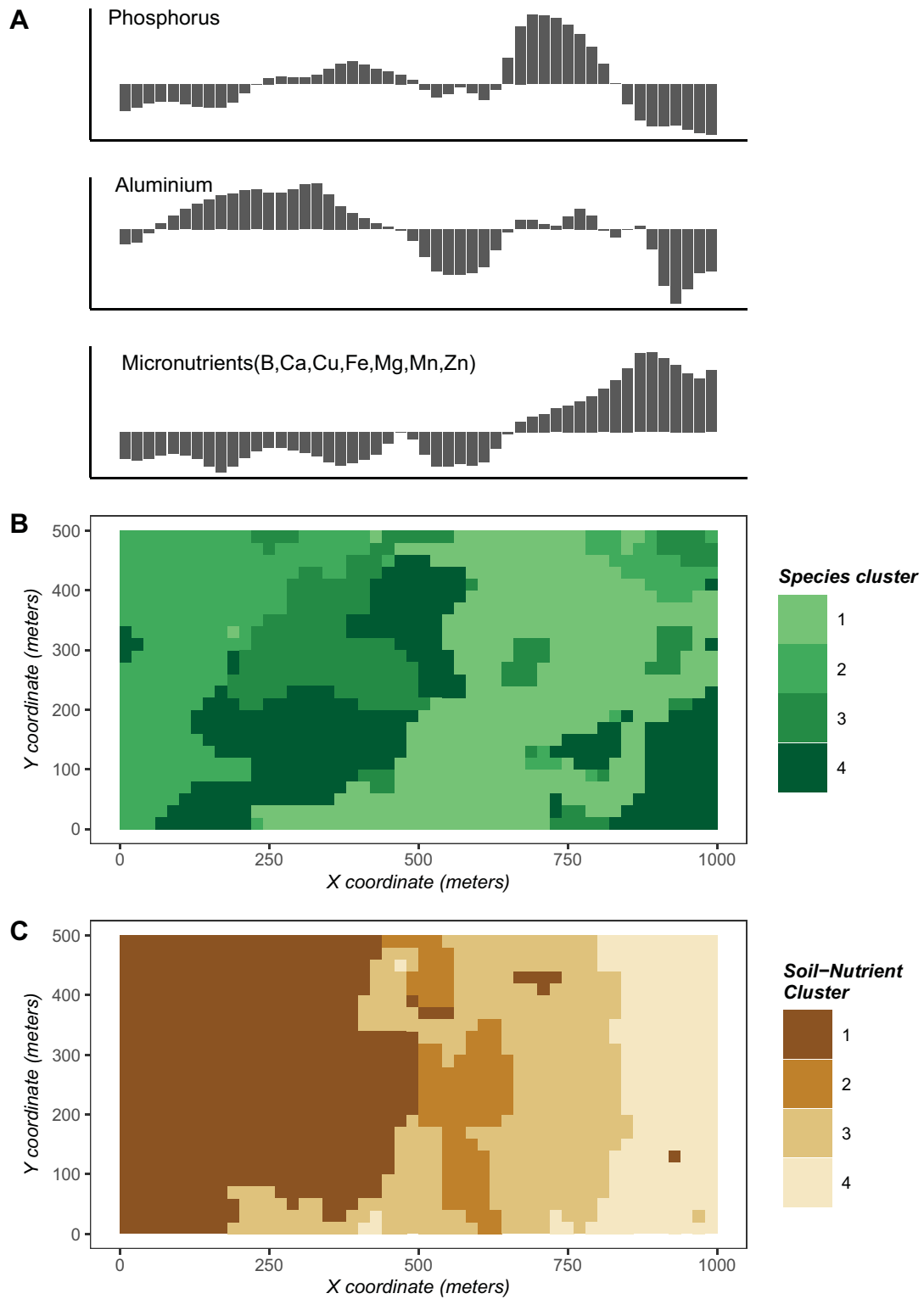
Figure 3: Spatial aggregation patterns of soil nutrient clusters and species clusters in Barro Colorado Island (census 7). *A* shows three sets of soil nutrients that strongly discriminated between species clusters and exhibit gradients with distinct peaks along the east-west axis (identical to the explicitly mentioned *x*-axis in *B* and *C*). These gradients are reflected in the boundaries of soil nutrient clusters (*C*). Species clusters (*B*) are identical to those from figure 1. A color version of this figure is available online.

Table 1: Results of C5.0 algorithm for the Barro Colorado Island data

	50	25	10	5	2
Average number of rules	8	10.2	25.3	56.3	67.8
Average accuracy	.70	.75	.79	.81	.81
Cohen's κ	.61	.68	.74	.76	.77
Importance of resources:					
Al	0	0	89	0	0
B	0	0	77	0	0
Ca	0	0	81	0	0
Cu	0	99	92	100	100
Fe	100	88	97	100	99
K	0	73	79	88	100
Mg	0	24	50	86	99
Mn	70	90	96	84	99
N	100	99	91	100	99
N(min)	91	93	98	91	99
P	97	99	100	100	100
pH	0	0	96	91	100
Water	97	90	100	100	100
Zn	100	100	100	100	100

Note: A preset of minimum number of samples per split (50, 25, 10, 5, 2) was used to impose limits on the length of resulting trees (using more samples per split leads to fewer rules). The importance of resources was measured as a percentage of iterations that included a given element to build a decision tree.

Foliar Stoichiometry. Spatial clusters did not differ strongly in terms of foliar stoichiometry and displayed high levels of within-group variation in foliar nutrient levels (fig. S10). As with morphological and growth-related traits, there was strong clustering among species in terms of their foliar stoichiometry, with species sorting into three clusters (figs. S10, S11). The association between clusters of leaf elemental composition and spatial clusters was weak (Cramer's $V = 0.26$, 95% CI = 0.23–0.29).

Discussion

The framework of partial niche differentiation stipulates that niche strategies will be reflected in distinct species clusters (D'Andrea and Ostling 2016), which in a spatial context should be associated with autocorrelated environmental variables. In this study, we detected those clusters among tropical woody species and found that they were associated with soil resources. We then used a machine
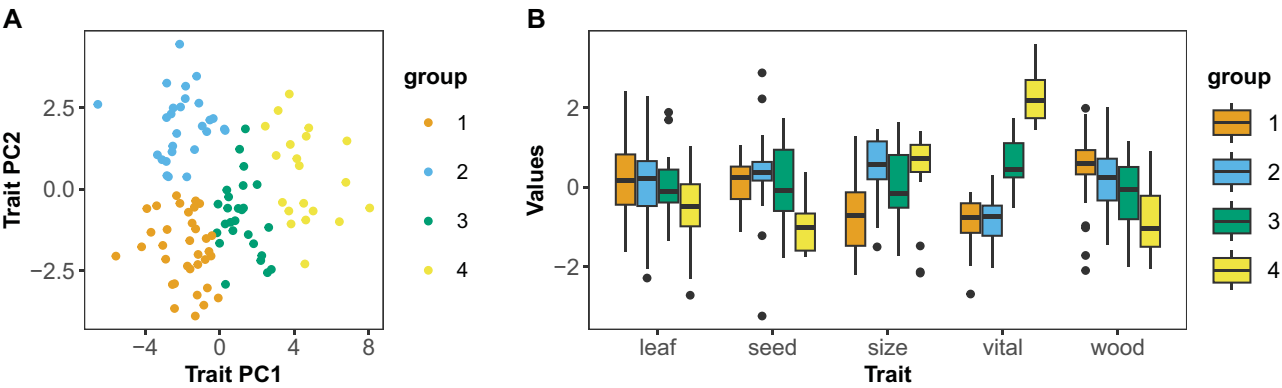

Figure 4: Species clustering by morphology/growth traits in Barro Colorado Island. *A*, Species, shown as points on the plane formed by the first two principal components of the measured traits, sorted into four groups. *B*, Distribution of trait values across groups show a gradient from conservative species in group 1 with low leaf mass per area (leaf), large seeds (seed), high wood density (wood), and slow growth and low mortality (vital) to acquisitive species in group 4 with the opposite traits. A color version of this figure is available online.

learning method to characterize these niches in terms of distinct soil resource concentrations. Our approach to infer the impacts of habitat filtering and species interactions is different from alternative approaches, such as niche modeling (e.g., Bahn and McGill 2007), in two important ways: (1) it analyzes spatial scales where species interactions plausibly influence community composition, thus offering advantage over methods such as joint species distribution models that infer the interactions statistically with varying degrees of success (Barner et al. 2018; Blanchet et al. 2020; König et al. 2021), and (2) it identifies and quantifies the niche differentiation apart from species' correlations with environmental variables. We used associations between species clusters and the environmental variables to characterize the niches. In contrast, the inference of habitat filtering in niche-modeling approaches is sensitive to the strength of statistical association between individual species and the environment.

Our analyses revealed high modularity in adjacency matrices compared with randomized data, indicating that groups of species are much more likely to occur near each other than near others, forming distinct species clusters. We then showed that these species clusters have distinct geographic distributions, thus justifying their interpretation as spatial clusters. The number, composition, and spatial distribution of inferred spatial clusters in BCI were highly consistent across eight censuses despite considerable temporal turnover (~10% of total individuals between consecutive censuses). Furthermore, adults of the species were, on average, 1.7 times as likely to recruit in their preferred soil environment as in the other soil environments, demonstrating the strength of niche differentiation by local abiotic conditions (fig. S2). This recruitment ratio was significantly higher than the null expectation from neutral community dynamics under dispersal limitation. This trend did not differ across life stages, indicating that habitat filtering acted at very early growth stages (see Baldeck et al. 2013b; fig. S3). These results establish that niche differentiation strongly drives the spatial distribution of species and can be observed as clustering in the spatial aggregation of species. There was also a strong spatial clustering of species in the La Planada plot that corroborated the observation that species tended to cluster into a small number of groups.

Since soil nutrients showed strong spatial heterogeneity on BCI, we expected a strong signature of niche differentiation by nutrients there, which was indeed the case. Species clusters showed distinct correlations with soil nutrients, indicating that soil spatial heterogeneity had a strong impact in the spatial clustering of species. The four nutrient clusters inferred from our k-means clustering analysis align well with the distributions of broad habitat types reported in previous surveys (Harms et al. 2001; Baillie et al. 2009). The slopes between high plateau and low plateau as well as streams

and swamp areas aligned well with the boundaries of soil nutrient clusters. These distinct spatial patterns are reflected in the strong predictive success of the C5.0 machine learning algorithm for BCI, which also identified the most important nutrients in discriminating between the clusters. Phosphorus, nitrogen, aluminum, and manganese, commonly considered important limiting nutrients or toxic elements in tropical forests, were among the better predictors of different species clusters (table 1).

Our finding of a relatively small number of species clusters implies that each spatial niche is shared by multiple species. In other words, species in the same cluster share similar adaptations to abiotic factors varying at the spatial scales examined in this study: soil nutrient and toxin concentrations, water availability, topology, and so on. This is consistent with predictions of emergent neutrality under competitive dynamics (D'Andrea and Ostling 2017). Under that paradigm, competitive exclusion is slow among species in the same cluster, allowing even low levels of immigration to counterbalance population losses, thus leading to a coexistence equilibrium (D'Andrea et al. 2019). It is also likely that those species further niche differentiate by environmental conditions varying at smaller spatial scales, such as light levels, and along non-spatial-niche axes, such as enemy defense. These additional sources of stabilization can also contribute to the permanence of otherwise transient clusters (Barabás et al. 2013).

Previous analyses have shown that spatial variation in beta diversity is explained to a large extent by soil nutrients, suggesting a strong influence of habitat filtering (John et al. 2007; Baldeck et al. 2013b). Our study complements those results in showing that the clustered spatial structure of tropical trees is consistent with a theoretical framework where habitat filtering, competition for nutrients, and dispersal drive hyperdiverse species assemblages into a state of so-called emergent neutrality, or partial niche differentiation, characterized by niche differentiation into relatively few clusters of species with mutually similar ecological strategies.

The nutrient niches in BCI can be characterized as follows: Soils on the easternmost area of the plot have relatively high pH and water levels, are low in aluminum (toxic to plants), and rich in various macro- and micronutrients except for nitrogen and phosphorus. Soils in a large area on the plot's west side are highly acidic, nutrient poor, and high in aluminum. The central area shows a sharp east-to-west gradient of decline in nutrient levels but has the highest levels of phosphorus and nitrogen. Species clusters broadly reflect these patterns. Species cluster 2 occurs exclusively in the western part of the plot associated with very low nutrient levels and high aluminum concentrations, while cluster 3 is localized on the eastern side of the plot with opposite soil conditions. Clusters 2 and 4 occur in the central

areas and show positive correlation with nitrogen and phosphorus levels but weak associations with all other nutrients. The boundaries of the clusters were located toward the central areas of the plot and coincided with sharp gradients of soil nutrients, thus indicating the important role of nutrients in niche differentiation. Indeed, the C5.0 machine learning algorithm also identified phosphorus, nitrogen, aluminum, and manganese—important limiting nutrients or toxic elements in tropical forests—as the variables that were most successful in separating the clusters.

While we show a strong characterization of spatial niches in terms of soil nutrients, we did not examine the interaction effects between different nutrients. One clear example of such interactions is that pH and water can modulate the ionization of many cations and affect their bioavailability. Water availability on BCI can fluctuate strongly between years and has been shown to induce spatially distinct patterns of tree mortality (Kupers et al. 2019). Multiple soil surveys would help with robust conclusions and can help clarify the role of temporal variability in soil nutrients levels. Furthermore, spatial niches of species in other plant communities may be characterized by different spatially heterogeneous abiotic variables beyond soil cations. For instance, different sets of nutrients were better predictors of the spatial clusters in La Planada where the soil characterization is different than that of BCI.

It is generally expected that niche differentiation between species would be reflected in their relevant functional traits (Kitajima and Poorter 2008; Kraft and Ackerly 2010). The traits analyzed in our study represent different facets of plant performance and their interaction with the environment (vital rates, size traits, seed traits, wood traits, leaf traits) and can potentially drive recruitment success in specific soil conditions. Previous analyses have also shown that foliar concentrations of some elements are conserved (see Metali et al. 2015), indicating that spatial and temporal variability in these traits should reflect species turnover in space and time. Indeed, we found strong clustering by those functional traits, and those trait clusters were geographically associated with our spatial clusters. However, we did not observe significant trait dissimilarities across spatial clusters of species (fig. S13). This could be due to several reasons. The filtering effect of soil nutrients may act on physiological or belowground traits that were not included in this analysis (see Chrispeels et al. 1999). It is also possible that the observed association between chosen traits and the species clusters could be affected by a phylogenetic signal in some traits (e.g., Sedio et al. 2012). Finally, in addition to specializing on different soil conditions, plants further niche differentiate via mechanisms such as defense (Sedio et al. 2018) and competition for light (D'Andrea et al. 2019) that may operate at finer or coarser spatial scales than the 20 × 20-m resolution in our study. The observed patterns

in the traits we analyzed may thus reflect niche differentiation at different spatial scales.

While our results support habitat filtering by local nutrient levels, alternative explanations are possible. Strong positive conditioning of soil nutrients by plants can in principle explain the spatial association between species clusters and soil nutrients (see Waring et al. 2015). Nutrient cycling of many elements is indeed driven by feedback through litterfall decomposition. Long-term litterfall manipulation experiments in the Gigante Peninsula of BCI preserve showed that N, Ca, Mg, and Zn levels in soil are influenced by litterfall (Sayer et al. 2020). However, such feedbacks are not likely to produce variation in soil nutrient levels at high spatial resolution because of high local tree species diversity, since the latter will also interfere with the consistency of soil conditioning (John et al. 2007). Additionally, higher leaching rates, erosion, and the transport of fine soil from higher to lower elevation is also common in BCI and can strongly interfere with local plant-soil feedbacks (Yavitt 2000). Large groups of species within a cluster sharing a common dispersal mode (e.g., animal dispersal) may create such covariation patterns, but a preliminary observation that species within clusters do not seem to have similar diaspore traits indicates otherwise.

In this analysis, we demonstrate the utility of our approach in distilling clear and interpretable patterns in hyperdiverse plant communities. It is evident that niche differentiation for soil nutrients drives the spatial structure of tropical trees. Our findings corroborate the idea of partial niche differentiation, according to which niche partitioning may be observed as a tendency for species to aggregate into a relatively small number of distinct clusters along the observed niche space. Our method can be used to quantify spatial signatures of niche differentiation in any censused community of sessile organisms and characterize the niches in term of the properties of a spatially heterogeneous habitat.

La Planada and plant trait data in BCI. We are thankful to James Dalling for his valuable insights and suggestions.

Statement of Authorship

M.S.U. and D.W. contributed equally to the project. R.D., D.W., and J.P.O. conceived the study; R.D., D.W., and M.S.U. contributed to the research and analysis; M.S.U. and R.D. wrote the manuscript, and all authors contributed to manuscript revisions.

Data and Code Availability

Data and code are available on figshare (https://doi.org /10.6084/m9.figshare.24539419; Umarani et al. 2023).

Literature Cited

Abdullah, A. Z., B. Winarno, and D. R. S. Saputro. 2021. The decision tree classification with C4.5 and C5.0 algorithm based on R to detect case fatality rate of dengue hemorrhagic fever in Indonesia. Journal of Physics: Conference Series 1776:012040. https://doi.org /10.1088/1742-6596/1776/1/012040.

Bahn, V., and B. J. McGill. 2007. Can niche-based distribution models outperform spatial interpolation? Global Ecology and Biogeography 16:733–742.

Baillie, I., H. Elsenbeer, F. Barthold, R. Grimm, and R. Stallard. 2009. Semi-detailed soil survey of Barro Colorado Island, Panama. https://striresearch.si.edu/bci-soil-map/.

Baldeck, C. A., K. E. Harms, J. B. Yavitt, R. John, B. L. Turner, R. Valencia, H. Navarrete, et al. 2013a. Habitat filtering across tree life stages in tropical forest communities. Proceedings of the Royal Society B 280:20130548. https://doi.org/10.1098/rspb.2013.0548.

Baldeck, C. A., K. E. Harms, J. B. Yavitt, R. John, B. L. Turner, R. Valencia, H. Navarrete, et al. 2013b. Soil resources and topography shape local tree community structure in tropical forests. Proceedings of the Royal Society B 280:20122532. https://doi.org /10.1098/rspb.2012.2532.

Barabás, G., R. D'Andrea, R. Rael, G. Meszéna, and A. Ostling. 2013. Emergent neutrality or hidden niches? Oikos 122:1565–1572. https:// doi.org/10.1111/j.1600-0706.2013.00298.x.

Barner, A. K., K. E. Coblentz, S. D. Hacker, and B. A. Menge. 2018. Fundamental contradictions among observational and experimental estimates of non-trophic species interactions. Ecology 99:557–566. https://doi.org/10.6084/m9.figshare.

Blanchet, F. G., K. Cazelles, and D. Gravel. 2020. Co-occurrence is not evidence of ecological interactions. Ecology Letters 23:1050–1063. https://doi.org/10.1111/ele.13525.

Chase, J. M. 2014. Spatial scale resolves the niche versus neutral theory debate. Journal of Vegetation Science 25:319–322. https:// doi.org/10.1111/jvs.12159.

Chesson, P. 2000. General theory of competitive coexistence in spatially-varying environments. Theoretical Population Biology 58:211–237.

Chrispeels, M. J., N. M. Crawford, and J. I. Schroeder. 1999. Proteins for transport of water and mineral nutrients across the membranes of plant cells. Plant Cell 11:661–675. https://doi.org/10.1105 /tpc.11.4.661.

Clark, D. B., D. A. Clark, and J. M. Read. 1998. Edaphic variation and the mesoscale distribution of tree species in a neotropical rain forest. Journal of Ecology 86:101–112.

Cody, M. L. 1974. Competition and the structure of bird communities. Princeton University Press, Princeton, NJ.

Condit, R. 1998. Tropical forest census plots. Springer, Berlin.

———. 2019. Complete data from the Barro Colorado 50-ha plot: 423617 trees, 35 years, v3. Dryad Digital Repository, https://doi .org/10.15146/5xcp-0d46.

Condit, R., B. M. J. Engelbrecht, D. Pino, R. Pérez, and B. L. Turner. 2013. Species distributions in response to individual soil nutrients and seasonal drought across a community of tropical trees. Proceedings of the National Academy of Sciences of the USA 110:5064–5068.

Connor, E. F., and D. Simberloff. 1979. The assembly of species communities: chance or competition? Ecology 60:1132–1140.

Csárdi G., T. Nepusz, V. Traag, S. Harvát, F. Zanini, D. Noom, and K. Müller. 2023. igraph: network analysis and visualization in R. https://doi.org/10.5281/zenodo.7682609.

Dalling, J. W., H. C. Muller-landau, S. J. Wright, and S. P. Hubbell. 2002. Role of dispersal in the recruitment limitation of neotropical pioneer species. Journal of Ecology 90:714–727.

Dalling, J. W., and N. Norden. 2023. La Planada forest dynamics plot soils dataset. Illinois Data Bank, University of Illinois at Urbana-Champaign. https://doi.org/10.13012/B2IDB-6140727_V1.

D'Andrea, R., J. Guittar, J. P. O'Dwyer, H. Figueroa, S. J. Wright, R. Condit, and A. Ostling. 2020. Counting niches: abundance-by-trait patterns reveal niche partitioning in a Neotropical forest. Ecology 101:e03019. https://doi.org/10.1002/ecy.3019.

D'Andrea, R., and A. Ostling. 2016. Challenges in linking trait patterns to niche differentiation. Oikos 125:1369–1385. https://doi .org/10.1111/oik.02979.

———. 2017. Biodiversity maintenance may be lower under partial niche differentiation than under neutrality. Ecology 98:3211–3218.

D'Andrea, R., M. Riolo, and A. M. Ostling. 2019. Generalizing clusters of similar species as a signature of coexistence under competition. PLoS Computational Biology 15:e1006688. https://doi.org /10.1371/journal.pcbi.1006688.

Davies, T. M., and J. C. Marshall. 2018. The sparr package: spatial and spatiotemporal relative risk. https://cran.r-project.org/web /packages/sparr/index.html.

Dayan, T., and D. Simberloff. 2005. Ecological and community-wide character displacement: the next generation. Ecology Letters 8:875–894. https://doi.org/10.1111/j.1461-0248.2005.00791.x.

Emmons, S., S. Kobourov, M. Gallant, and K. Borner. 2016. Analysis of network clustering algorithms and cluster quality metrics at scale. PLoS ONE 11:e0159161.

Graco-Roza, C., A. M. Segura, C. Kruk, P. Domingos, J. Soininen, and M. M. Marinho. 2021. Clumpy coexistence in phytoplankton: the role of functional similarity in community assembly. Oikos 130:1583–1597. https://doi.org/10.1111/oik.08677.

Harms, K. E., R. Condit, S. P. Hubbell, and R. B. Foster. 2001. Habitat associations of trees and shrubs in a 50-ha Neotropical forest plot. Journal of Ecology 89:947–959. https://doi.org/10.1111 /j.1365-2745.2001.00615.x.

Hidayat, D. S., I. S. Sitanggang, and G. Semiadi. 2017. Decision support system for bat identification using random forest and C5.0.

Telkomnika 15:1215–1222. https://doi.org/10.12928/TELKOMNIKA.v15i3.3638.

Holt, R. D. 2009. Bringing the Hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. Proceedings of the National Academy of Sciences of the USA 106(suppl.):19659–19665.

Hubbell, S. P., R. Condit, and R. B. Foster. 2005. Barro Colorado forest census plot data. http://ctfs.si.edu/webatlas/datasets/bci.

Hutchinson, G. E. 1957. Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology 22:415–427.

John, R., J. W. Dalling, K. E. Harms, J. B. Yavitt, R. F. Stallard, M. Mirabello, S. P. Hubbell, et al. 2007. Soil nutrients influence spatial distributions of tropical tree species. Proceedings of the National Academy of Sciences of the USA 104:864–869. https://doi.org/10.1073/pnas.0604666104.

Kitajima, K., and L. Poorter. 2008. Functional basis for resource niche partitioning by tropical trees. Pages 160–181 *in* W. P. Carson and A. S. Schnitzer, eds. Tropical forest community ecology. Wiley-Blackwell, Oxford.

König, C., R. O. Wüest, C. H. Graham, D. N. Karger, T. Sattler, N. E. Zimmermann, and D. Zurell. 2021. Scale dependency of joint species distribution models challenges interpretation of biotic interactions. Journal of Biogeography 48:1541–1551. https://doi.org/10.1111/jbi.14106.

Kraft, N. J., and D. D. Ackerly. 2010. Functional trait and phylogenetic tests of community assembly across spatial scales in an Amazonian forest. Ecological Monographs 80:401–422.

Kuhn, M. 2021. caret: classification and regression training. R package version 6.0-90. https://CRAN.R-project.org/package=caret.

Kuhn, M., and R. Quinlan. 2022. C50: C5.0 decision trees and rule-based models. R package version 0.1.6. https://CRAN.R-project.org/package=C50.

Kupers, S. J., C. Wirth, B. M. J. Engelbrecht, and N. Rüger. 2019. Dry season soil water potential maps of a 50 hectare tropical forest plot on Barro Colorado Island, Panama. Scientific Data 6:63. https://doi.org/10.1038/s41597-019-0072-z.

Landis, J. R., and G. G. Koch. 1977. The measurement of overserver agreement for categorical data. Biometrics 33:159–174.

Metali, F., K. Abu Salim, K. Tennakoon, and D. F. R. P. Burslem. 2015. Controls on foliar nutrient and aluminium concentrations in a tropical tree flora: phylogeny, soil chemistry and interactions among elements. New Phytologist 205:280–292. https://doi.org/10.1111/nph.12987.

Muller-Landau, H. C., S. J. Wright, O. Calderón, R. Condit, and S. P. Hubbell. 2008. Interspecific variation in primary seed dispersal in a tropical forest. Journal of Ecology 96:653–667. https://doi.org/10.HH/j.1365-2745.2008.01399.x.

Newman, M. E. J. 2004. Analysis of weighted networks. Physical Review E 70:9. https://doi.org/10.1103/PhysRevE.70.056131.

Paoli, G. D., L. M. Curran, and D. R. Zak. 2006. Soil nutrients and beta diversity in the Bornean Dipterocarpaceae: evidence for niche partitioning by tropical rain forest trees. Journal of Ecology 94:157–170. https://doi.org/10.1111/j.1365-2745.2005.01077.x.

Pardo, C. E., and P. C. DelCampo. 2007. Combanacion de metodos factorials y de analisis de conglomerados en R: el paquete FactoClass. Revista Colombiana de Estadistica 30:231–245. https://www.rdocumentation.org/packages/FactoClass/versions/1.2.7.

Phillips, O. L., P. Núñez Vargas, A. Lorenzo Monteagudo, A. Peña Cruz, M. Chuspe Zans, W. Galiano Sánchez, M. Yli-halla, et al. 2003. Habitat association among Amazonian tree species: a landscape-scale approach. Journal of Ecology 91:757–775.

R Core Team. 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Rubio, V. E., and N. G. Swenson. 2022. Functional groups, determinism and the dynamics of a tropical forest. Journal of Ecology 110:185–196. https://doi.org/10.1111/1365-2745.13795.

Sayer, E. J., C. Rodtassana, M. Sheldrake, L. M. Bréchet, O. S. Ashford, L. Lopez-Sangil, D. Kerdraon-Byrne, et al. 2020. Revisiting nutrient cycling by litterfall—insights from 15 years of litter manipulation in old-growth lowland tropical forest. Advances in Ecological Research 62:173–223. https://doi.org/10.1016/bs.aecr.2020.01.002.

Scheffer, M., and E. H. van Nes. 2006. Self-organized similarity, the evolutionary emergence of groups of similar species. Proceedings of the National Academy of Sciences of the USA 103:6230–6235. https://doi.org/10.1073/pnas.0508024103.

Sedio, B. E., J. D. Parker, S. M. Mcmahon, and A. S. J. Wright. 2018. Comparative foliar metabolomics of a tropical and a temperate forest community. Ecology 99:2547–2653. https://doi.org/10.1002/ecy.2533.

Sedio, B. E., S. J. Wright, and C. W. Dick. 2012. Trait evolution and the coexistence of a species swarm in the tropical forest understorey. Journal of Ecology 100:1183–1193. https://doi.org/10.1111/j.1365-2745.2012.01993.x.

Tibshirani, R., G. Walther, and T. Hastie. 2001. Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society B 63:411–423.

Ulrich, W., P. Olszewski, R. Puchałka, P. Sewerniak, and M. Koprowski. 2021. Inter- and intraspecific spatial distributions, spatial segregation by dominants and emergent neutrality in understorey plants. Oikos 130:1813–1822. https://doi.org/10.1111/oik.08269.

Umarani, M., D. Wang, J. O'Dwyer, and R. D'Andrea. 2023. Data from: A spatial signal of niche differentiation in tropical forests. American Naturalist, figshare, https://doi.org/10.6084/m9.figshare.24539419.

Uriarte, M., R. Condit, C. D. Canham, and S. P. Hubbell. 2004. A spatially explicit model of sapling growth in a tropical forest: does the identity of neighbours matter? Journal of Ecology 92:348–360. https://doi.org/10.1111/j.0022-0477.2004.00867.x.

Vallejo-Joyas, M. I., A. C. Londoño-Vega, R. López-Camacho, G. Galeano, E. Álvarez-Dávila, and W. Devia-Álvarez. 2005. Establecimiento de Parcelas Permanentes en Bosques de Colombia. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Bogotá.

Waring, B. G., L. Álvarez-Cansino, K. E. Barry, K. K. Becklund, S. Dale, M. G. Gei, A. B. Keller, et al. 2015. Pervasive and strong effects of plants on soil chemistry: a meta-analysis of individual plant "Zinke" effects. Proceedings of the Royal Society B 282:20151001.

Warrens, M. J. 2015. Five ways to look at Cohen's kappa. Journal of Psychology and Psychotherapy 5:1000197.

Wolf, J. A., S. P. Hubbell, G. A. Fricker, and B. L. Turner. 2015. Geospatial observations on tropical forest surface soil chemistry. Ecology 96:2313.

Wright, S. J. 2023. Foliar elemental composition and carbon and nitrogen isotope values for 339 woody species from Barro Colorado

Island, Panama. Smithsonian Institution, Washington, DC. https://doi.org/10.25573/data.23463254.

Wright, S. J., K. Kitajima, N. J. Kraft, P. B. Reich, I. J. Wright, D. E. Bunker, R. Condit, et al. 2010. Functional traits and the growth-mortality trade-off in tropical trees. Ecology 91:3664–3674. https://doi.org/10.6084/m9.figshare.c.3303654.v1.

Yavitt, J. B. 2000. Nutrient dynamics of soil derived from different parent material on Barro Colorado Island, Panama. Biotropica 32:198–207.

H.H.NICHOLS. SC.
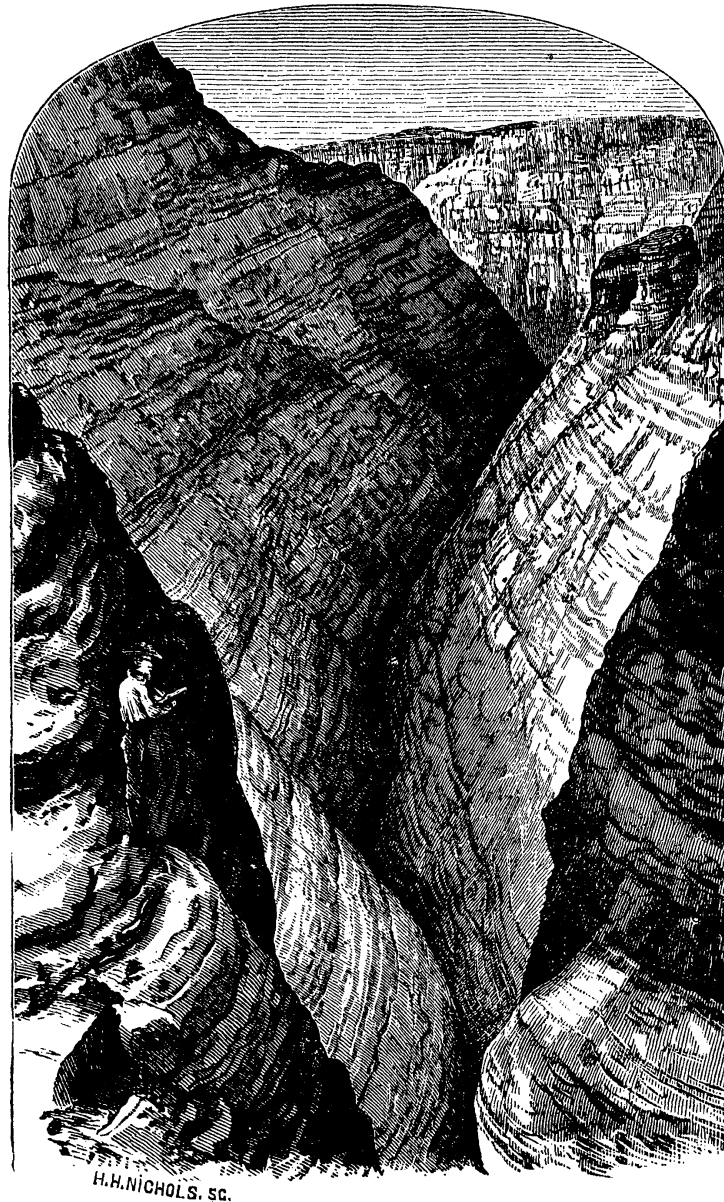
"An idea of the grandeur of these dark, solitary gorges, with vertical sides often very nearly a mile high, and with pinnacles and towers overhanging the river winding like a silvery thread below, may be gained by a glance at the figures of Mu-koon-tu-weap Cañon, of a cañon in Escalante Basin [figured], but especially of the Grand Cañon." From the review of Powell's *Exploration of the Colorado* (*The American Naturalist*, 1876, 10:102–105).