RESEARCH ARTICLE

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

Controlling for the effects of environmental availability when testing how the environment determines community compositional uniqueness

Toby P. N. Tsang[1,2] | Timothy C. Bonebrake[2] | Lauren C. Ponisio[3,4] | Marc W. Cadotte[1,5]

[1]Department of Biological Sciences, University of Toronto-Scarborough, Toronto, Ontario, Canada

[2]School of Biological Sciences, The University of Hong Kong, Hong Kong, China

[3]Institute for Ecology and Evolution, University of Oregon, Eugene, Oregon, USA

[4]Data Science Initiative, University of Oregon, Eugene, Oregon, USA

[5]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada

Correspondence
Toby P. N. Tsang
Email: paknok.tsang@utoronto.ca

Handling Editor: Steven Kembel

Abstract

1. Identifying the drivers that promote unique species compositions (i.e. ecological uniqueness) is crucial to understanding the mechanisms underpinning diversity patterns and for effective conservation planning. Environmental conditions are often sampled differentially in datasets, which can lead to rarer environments having unique species compositions, provided that environmental differences increase compositional differences. This effect, however, will be undesirable when a study aims to test the direct impact of environments rather than their availability.

2. We developed an approach to reduce the effects of environmental availability in ecological uniqueness analyses through calculating expected pairwise compositional dissimilarities for each unique environmental condition. We further used simulations to assess the performance of our methods by randomly generating communities from two hypothetical environments with non-overlapping species pools. Additionally, we used a dataset of 50 tree communities to demonstrate how environmental availability could impact relationships between environmental conditions and ecological uniqueness in empirical studies.

3. Our simulations revealed that uniqueness metrics based on observed values are sensitive to environmental availability, while our approach correctly concluded that there were no differences among environments under the unbalanced design. Our analysis of tree communities produced divergent conclusions between the two approaches, as increasing slope reduced ecological uniqueness after controlling for their low availability in the dataset only. This suggests that low environmental availability inflated the ecological uniqueness of sites with high slope, which opposed its direct negative effects on ecological uniqueness, leading to a weak relationship based on observed values.

4. To achieve a more mechanistic understanding of ecological uniqueness patterns, the effects of environmental availability must be considered. We recommend that

studies use both uncorrected and corrected analyses to identify not only the direct effects of environmental conditions but also the degree to which their availability influences the observed relationships between ecological uniqueness and environment conditions. Our approach should be most necessary and applicable in unbalanced designs, which is a common characteristic in empirical studies.

1 | INTRODUCTION

Beta (β) diversity is critical to understanding how species composition varies across space and time, and for inferring the mechanisms that structure biodiversity (Baselga, 2010; Tatsumi et al., 2021). Estimating beta diversity is also used to understand the scaling relationships between alpha (α) and gamma (γ) diversity (Barton et al., 2013), the spatial distribution of ecosystem function and services (Mori et al., 2018), and the prioritisation of protected areas (Devictor et al., 2010; Socolar et al., 2016). The contribution of different communities to β diversity can vary strongly, depending on how unique their species compositions are (i.e. ecological uniqueness), which are quantified based on pairwise compositional dissimilarities (Legendre & de Cáceres, 2013; Mokany, Ware, et al., 2022). Many studies have sought to determine relationships between ecological uniqueness and environmental drivers, and citing ecological mechanisms such as increasing disturbance (Allen et al., 2019; Leão et al., 2020) and low ecological specialisation (García-Navas et al., 2022) as potential explanations. In conservation studies, ecological uniqueness has been included in assessing the effectiveness of conservation measures (Cetra et al., 2022; Ribeiro et al., 2022; So & Dudgeon, 2021) and identifying environments that should be prioritised for protection (Dubois et al., 2020; Jyrkänkallio-Mikkola et al., 2018; Mokany et al., 2020). One way to quantify ecological uniqueness is by averaging pairwise compositional dissimilarity, also known as pairwise beta diversity (e.g. Marion et al., 2017), between the focal sites and other sites (hereafter as β_{pair}) (Mokany, Ware, et al., 2022). Ecological uniqueness can also be measured as the distance to the centroid within a multivariate space, with the centroid representing the 'average' of communities within a sampling pool (Anderson et al., 2006). Legendre and de Cáceres (2013) have further developed the index Local Contribution to Beta Diversity (LCBD), which is the ratio between the squared distance to the centroid of each site and the total squared distance to the centroid of all sites (see Table S1 for their calculations).

Some studies have noted that uniqueness can be affected by the number of sampling units of different environmental conditions (hereafter referred to as 'environmental availability') within the sampling pool, with a rarer environment containing more unique compositions (Baidya & Bagchi, 2022; Dubois et al., 2020; Ejrnæs et al., 2018; Mokany, McCarthy, et al., 2022). This mechanism, however, is not always considered when assessing uniqueness patterns

across space (Schneck et al., 2022). In some cases, the effects of environmental availability can be desirable, such as when the objective is to illustrate or predict the ecological uniqueness of different sites within landscapes and therefore identify conservation priorities (Dansereau et al., 2022; Dubois et al., 2020; Heino et al., 2022). However, if the objective is to examine how environmental drivers affect ecological uniqueness through specific mechanisms, such as whether human disturbance favours generalists and therefore reduces the uniqueness of communities (Allen et al., 2019; Leão et al., 2020), or isolation promotes speciation, and therefore unique communities (Perbiche-Neves et al., 2019), the effects of environmental availability should be controlled to minimise its confounding effects.

Conflating the effects of environmental availability and the environment can also misguide conservation efforts. For example, incorrectly attributing uniqueness patterns to direct environmental impacts, such as habitat degradation, can lead to inappropriate management recommendations (Legendre & de Cáceres, 2013; Schneck et al., 2022). As one example, García-Navas et al. (2022) found that the ecological uniqueness of bird communities started to decline rapidly when surrounding olive grove cover exceeded 50%, interpreting this as a signal of biotic homogenisation driven by agricultural expansion. Subsequently, management recommendations were proposed, such as restoring and protecting natural and semi-natural habitats. While this may be plausible, sites with >50% surrounding olive grove cover are more common in the sampling pool, thus the results could be driven by their higher environmental availability rather than agricultural expansions driving the homogenisation of communities. Additionally, sampling bias can create mismatches in environmental availability between the sampling pool and field conditions (Ejrnæs et al., 2018). As a result, such patterns would reflect sampling designs rather than ecological processes and have limited relevance to underlying drivers.

Previously, removing the effects of uneven environmental availability in uniqueness analyses has been achieved through resampling (Baidya & Bagchi, 2022). However, such approaches do not apply to numeric variables without binning, which can lead to other issues, including reduced power, arbitrary cut-offs, biased estimates, underestimated variance and inflated type-I error (Royston et al., 2006). Furthermore, applying resampling approaches can be challenging when multiple environmental variables are of interest, especially when these variables have very different evenness distributions.

Assuming that environmental differences increase compositional differences, which is typically true (Graco-Roza et al., 2022), we expect sensitivity of ecological uniqueness metrics to environmental availability. Imagine a study interested in examining the effects of habitat differences on ecological uniqueness, with 15 and five communities from habitats A and B, respectively. The average β_{pair} value for each site of habitat A would be based on 14 and five within- and between-habitat comparisons, while for each site of habitat B, the metric would be from 4 within- and 15 between-habitat comparisons. Note that the numbers of within-habitat comparisons are always one fewer than the number of communities within the group since any community will not compare metrics with itself. Thus, if within-habitat compositional differences are smaller than between-habitat differences, the rarer habitats (i.e. habitat B in the hypothetical example) should always exhibit higher uniqueness. For distance to centroid and LCBD, the centroid represents the average community within the species pool in a multivariate space. Under an unbalanced design, the average community should be more representative of habitat A due to their higher availability. Thus, again rare habitats would show higher distance to centroid and LCBD. While our hypothetical examples are based on a nominal variable, the effects should apply to continuous variables as long as environmental differences increase compositional differences. In the rare cases where environmental differences decrease compositional differences (Graco-Roza et al., 2022), the opposite would occur: the rare habitat should show lower uniqueness than the common habitats.

Here, we use simulations to illustrate how the uneven availability of different environments can alter relationships between ecological uniqueness and environmental conditions for different metrics. We then propose a statistical method to remove the effects of uneven environmental availability for continuous, ordinal, and nominal variables. We further used an empirical dataset to demonstrate the importance of our analyses to identify causes of ecological uniqueness.

2 | MATERIALS AND METHODS

2.1 | Steps to remove the effects of environmental availability

Our approach (Figure 1) is based on modelling β_{pair} with different environmental variables, including ordinal, nominal, and continuous, followed by averaging expected β_{pair} across each unique environmental condition. We developed an R function for this analysis, which is available on GitHub (https://github.com/tpaknok/Beta-diversity). Future updates to the function will also be available on the repository. We focused on β_{pair} because other metrics (e.g. distance to centroid) are additionally influenced by matrix size (Marion et al., 2017; Stier et al., 2013), and β_{pair} is widely employed in ecological analyses. While the focus on β_{pair} is similar to generalised dissimilarity modelling (GDM), our approach does not assume β_{pair} increases with environmental dissimilarity (Mokany, Ware, et al., 2022), as negative relationships have been found in some empirical studies

(a) Parametrize how β_{Pair} varies across environments for each site.
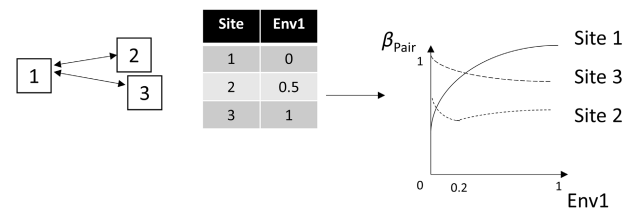
(b) Create a niche space and generate equally-spaced virtual sites.
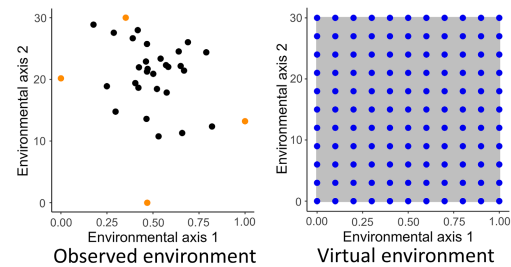
(c) Calculate pairwise β_{Pair} between actual and virtual sites, and eventually ecological uniqueness in the niche space through averaging (U_{niche})

FIGURE 1 The proposed steps to remove the effects of environmental availability in uniqueness analyses. (a) Relationships between β_{pair} and environmental conditions are modelled for each site separately, represented by different lines in the figure. (b) Generate a niche space defined as the hyperrectangle bounded by the minima and maxima of each environmental variable of interest before simulating hypothetical sites evenly spread across the niche space. Orange points represent the minima and maxima of the hypothetical variables used in this example, while black points are not used for constructing niche space. Blue points represent hypothetical sites generated. (c) For each actual site, extract the expected β_{pair} at each hypothetical site based on the model built in step 1, before averaging them to obtain the ecological uniqueness of each site.

(Graco-Roza et al., 2022). Additionally, GDM assumes the same β_{pair} for all communities compared to other communities sharing identical environmental conditions (Mokany, Ware, et al., 2022). However, this is unlikely true, as many empirical studies have shown that within-habitat β_{pair} differs across habitat groups (Newbold et al., 2016).

Briefly, the first step of our proposed analysis is to construct one model between β_{pair} and environmental variables for each focal site (Figure 1a). Then we project all models to each unique environmental condition sampled in the dataset, thus obtaining the expected

β_{pair} when comparing with each unique condition (Figure 1b). For each site, we calculate expected ecological uniqueness by averaging expected β_{pair} with each unique condition (Figure 1c). All environmental conditions are only represented once in the projection, thus removing the effects of environmental availability effects in calculating ecological uniqueness. Details of each major step of our analyses are provided below.

For each site, we obtain β_{pair} with other communities using a dissimilarity index. The R function calculates the Jaccard dissimilarity by default, as it is less sensitive to other issues (e.g. taxonomic error) and therefore more likely to derive a correct inference (Schroeder & Jenkins, 2018). Nevertheless, any dissimilarity can be used in the function. We then use generalised additive modelling (GAM) for each site separately, regressing β_{pair} against environmental conditions but not environmental similarity (Figure 1a), since the same dissimilarity can indicate different conditions and therefore different β_{pair}. For example, a natural site compared with agricultural and urban sites would both have environmental dissimilarity at 1, while at 50% urban land cover, the dissimilarity with 0% and 100% would be 0.5, with the sign being neglected (Ferrier et al., 2007). The use of GAM also allows non-linearity in how environmental differences drive β_{pair}. Random effects and correlation structures can also be added in our function, in which case it will instead fit a generalised additive mixed model (GAMM) using R package *mgcv* (Wood, 2017).

After forming the relationships between β_{pair} and environmental variables for each site, we project dissimilarity in niche space to obtain ecological uniqueness measures (Figure 1b). The niche space is a hyper-rectangular box bounded by the minima and maxima of each environmental variable observed within the sampling pool to reduce (but not eliminate) extrapolation (Figure 1b). For nominal variables all unique levels are included to form the boundaries. This approach is comparable to previous studies projecting dissimilarity in real world environments to identify ecologically unique areas (Harwood et al., 2022; Mokany et al., 2020; Mokany, McCarthy, et al., 2022; Shaw et al., 2022; Willis et al., 2012), yet here the projection is applied toward a hypothetical environment to remove the influence of environmental availability. Within this hypothetical environment, we conduct a gridded sampling (Perret & Sax, 2022) and calculate the expected ecological uniqueness for each site. We obtain expected β_{pair} at all levels for nominal and ordinal variables, but the latter was treated as a numeric variable in the models (Ferrier et al., 2007). We generated a fixed number of evenly spaced points across each environmental gradient for numeric variables.

We then calculate the expected β_{pair} with all hypothetical-actual site pairs based on the GAMM results (Figure 1c). We subsequently calculate the unweighted mean of these expected β_{pair} to obtain ecological uniqueness of each actual site within the niche space, termed as U_{niche}. Ecological uniqueness based on the observed dataset, which is the unweighted mean of β_{pair} between each actual site pair, would be termed as U_{obs}. These variables can then be regressed against environmental conditions to examine their effects on ecological uniqueness. Comparing results between the U_{obs} and U_{niche} models provides insights into how environmental availability affects uniqueness patterns. The U_{niche} model offers a more robust conclusion on the direct effects of environmental conditions because environmental availability effects are minimised.

We also note here that, under designs with even environmental availability, our proposed analysis should provide similar, but not identical, results, with analyses based on observed data. Consider an example with an equal number (n) of communities from two habitats. The U_{obs} of any communities would be based on n between-habitat comparisons and $n-1$ within-habitat comparisons (recall that communities do not compare metrics with themselves). For U_{niche}, the expected β_{pair} to a community from different habitats would first be obtained, before further averaging them across habitats. Thus, U_{obs} would be slightly biased toward between-habitat comparisons ($n > n-1$). In U_{niche} both within- and between-habitat comparisons would have equal weighting, leading to slight differences in the results.

2.2 | Simulation

To illustrate how environmental availability alters the results of ecological uniqueness, we first simulated communities from two habitats, A and B (Figure 2). To generate ecologically realistic communities, we used the tree count data (diameter at breast height ≥ 10 cm) of a 50-ha sample plot in Barro Colorado Island of Panama obtained from the R package *vegan* (Oksanen et al., 2022), which contains 50 1-ha quadrats. We used the null model algorithm r2dtable from the same package to randomly generate species composition for each community, while preserving other community properties such as species abundance across all quadrats, and alpha (α) and gamma (γ) diversity within and across plots. Next, we generated two sets of communities of 50 quadrats with non-overlapping species pools and assigned them to habitats A and B. Specifically, we used the r2dtable algorithm twice and generated two sets of communities, but considered distinct species identities, such that no species overlapped across the two habitats. The aim was to create a habitat gradient with between-habitat variations stronger than within-habitat variations. We quantified the observed uniqueness for each quadrat from both habitats based on the full dataset, before running a linear regression with uniqueness as the response and habitat identity as predictor. The coefficient obtained is assumed to represent the "true" differences across habitats. Due to their popularity, we used three uniqueness metrics for our analyses—distance to the centroid, LCBD and U_{obs}. We used the function betadisper in *vegan* (Oksanen et al., 2022) to calculate the distance to centroid, and the function LCBD.comp in *adespatial* (Dray et al., 2022) to calculate LCBD. The Jaccard dissimilarity matrix was square root transformed before computing the distance to centroid and LCBD of each plot (Legendre & de Cáceres, 2013).

We then randomly subsampled the landscape with different sampling patterns to represent balanced and unbalanced habitat
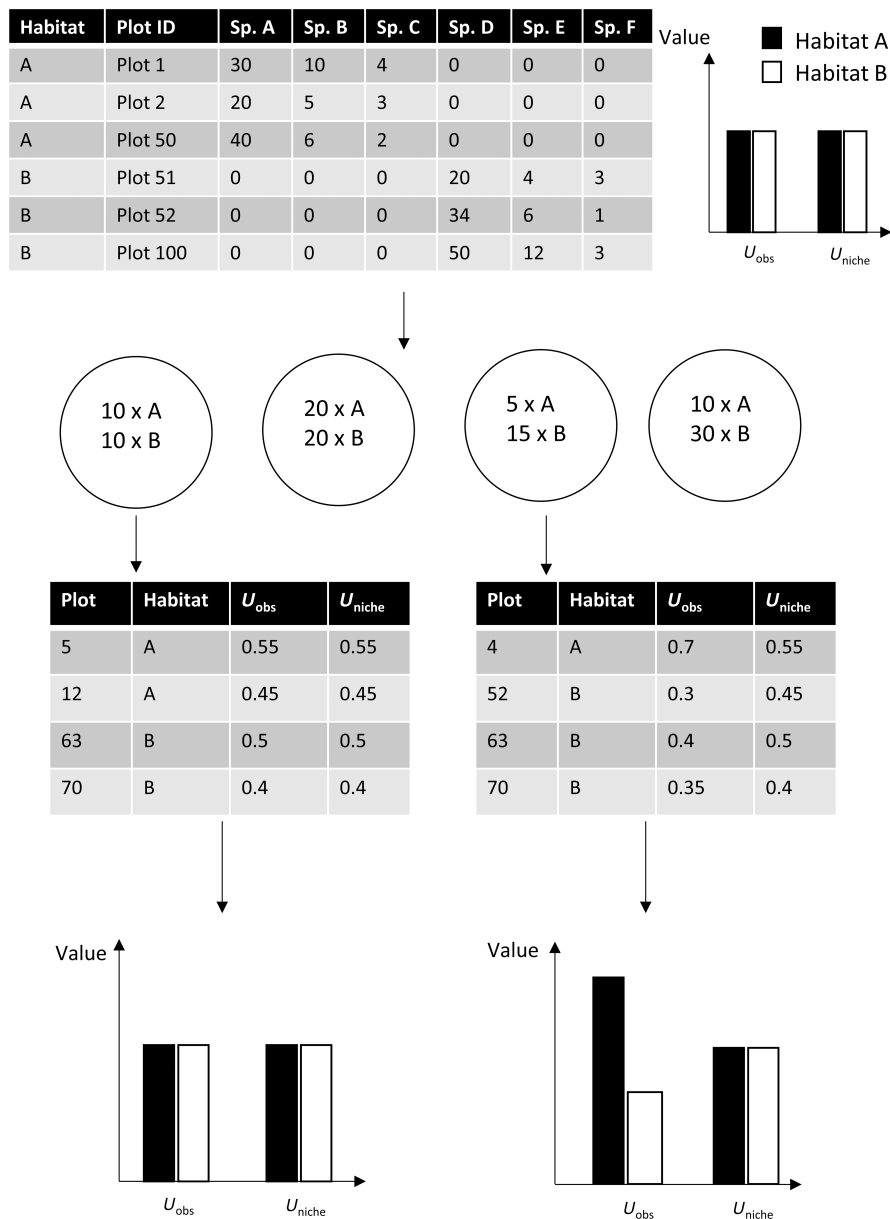
Habitat	Plot ID	Sp. A	Sp. B	Sp. C	Sp. D	Sp. E	Sp. F
A	Plot 1	30	10	4	0	0	0
A	Plot 2	20	5	3	0	0	0
A	Plot 50	40	6	2	0	0	0
B	Plot 51	0	0	0	20	4	3
B	Plot 52	0	0	0	34	6	1
B	Plot 100	0	0	0	50	12	3

Circles: 10 x A / 10 x B — 20 x A / 20 x B — 5 x A / 15 x B — 10 x A / 30 x B

Plot	Habitat	U_{obs}	U_{niche}
5	A	0.55	0.55
12	A	0.45	0.45
63	B	0.5	0.5
70	B	0.4	0.4

Plot	Habitat	U_{obs}	U_{niche}
4	A	0.7	0.55
52	B	0.3	0.45
63	B	0.4	0.5
70	B	0.35	0.4

FIGURE 2 Steps of the simulation analyses. First, a landscape of 100 plots was generated, with 50 plots assigned to two habitats with non-overlapping species pools and no uniqueness differences. Then two balanced and two unbalanced subsampling schemes were used, each repeated 100 times to obtain subsampled landscapes. We then calculated average uniqueness differences between habitats at plot levels for each iteration. These numbers were compared with the true effect size obtained under complete sampling.

availability in the sampling pool. In addition to the complete sampling scenario, we created two balanced and two unbalanced scenarios, with the number of quadrats as (10, 10), (20, 20), (5, 15) and (10, 30) from habitat A and B. After each subsampling, we quantified compositional uniqueness using the three metrics and U_{niche}. For factor variables no spline could be fit, and therefore the GAM in our proposed analysis is equivalent to linear regression. After obtaining uniqueness metrics for each plot, we used linear regression with habitat identity as the only variable and extracted the coefficient, which represents average habitat differences in uniqueness at plot levels. We repeated the subsampling 100 times, thus generating 100 sets of uniqueness metrics for each sampling scenario. We conducted pairwise Pearson's correlation between the four uniqueness metrics to assess the strength of correlation in each set. Because we detected almost perfect correlation between LCBD, distance to centroid, and U_{obs} ($R > 0.99$, see Section 3), we only retained U_{obs} to

compare with U_{niche} in subsequent analyses as they have more comparable scales.

To assess whether environmental availability influences the results of uniqueness analyses, we calculated the differences between the true coefficient and the ones obtained from the subsampled landscape for each metric. True coefficients for both U_{obs} and U_{niche} were set as the coefficient obtained (i.e. average habitat differences at plot levels) when using the completely sampled landscape. We further used one-sample t-tests to examine if the coefficient differences differed from zero for each metric and sampling pattern combination. We also conducted other analyses using simulated communities and numeric variables to demonstrate the influence of uneven environmental availability and obtained similar conclusions that environmental availability influenced uniqueness patterns (see Text S1 for details of simulation and Figure S1 for results).

2.3 | Empirical analyses

We again used the tree count data and topographic variables of the 50-ha sample plot in Barro Colorado Island of Panama. We selected the elevation, convexity and slope of each 1 ha quadrat (see de Cáceres et al., 2012 for details). While other environmental data such as soil nutrients are available for each quadrat, we chose these three variables as they have been shown to affect beta diversity in the system (de Cáceres et al., 2012). All environmental data were obtained from the R package *BiodiversityR* (Kindt, 2022).

We first created density plots of each topographic variable to assess the evenness of environmental availability in the data. We then calculated U_{obs} by obtaining average β_{pair} of each quadrat based on Jaccard dissimilarity, followed by applying the proposed workflow to obtain U_{niche}. We formed a hyperrectangular niche space using the minima and maxima of each topographic variable, and we generated 25 evenly spaced hypothetical sites across each gradient. In total $25^3 = 15{,}625$ hypothetical sites were distributed as a grid in the niche space. We also added a spatial exponential covariance structure based on the coordinates of each plot. We used a normal distribution in the GAMM and three splines to parameterise the relationship between β_{pair} and topographic variable for each quadrat, leading to 50 GAMM models. For both uniqueness metrics, we regressed uniqueness against the three topographic variables in a generalised least square model, again with a spatial exponential covariance structure.

3 | RESULTS

We used a simulation landscape with two habitats of a non-overlapping species pool to examine the effects of environmental availability on uniqueness patterns, expecting uniqueness to be higher in the rarer habitats under unbalanced subsampling, but similar between habitats under complete sampling and balanced subsampling. We first assessed the correlation strength between distance to centroid, LCBD, U_{obs} and U_{niche}. Pearson's correlation showed that the four uniqueness metrics were almost perfectly correlated under complete sampling and balanced subsampling in all iterations (minimum Pearson's $R > 0.99$). Under unbalanced subsampling, distance to centroid, LCBD and U_{obs} were again almost perfectly correlated in all iterations (minimum Pearson's $R > 0.99$), but their correlations with U_{niche} strongly varied across iterations, with Pearson's R ranging from -0.63 to 0.92 (Figure S2). Increasing the number of subsamples reduced the fluctuation of Pearson's R across iterations (Figure S2).

Because of the high correlations, only U_{obs} was used for comparisons with U_{niche} in subsequent analyses. Both uniqueness metrics exhibited minimal differences between habitats when the simulated landscape was completely sampled ($p \sim 0.34$ for all metrics). In all iterations, all uniqueness metrics were extremely similar between habitats under balanced subsampling irrespective of total sample size (Figure 3a), consistent with our expectations. We further compared average habitat differences under complete sampling and balanced subsampling, and found no significant difference between
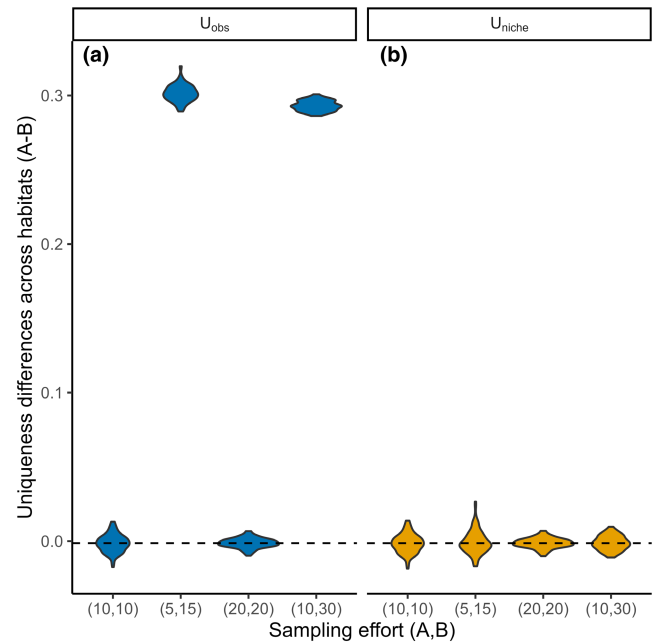

FIGURE 3 Violin plots visualizing estimated habitat differences across 100 simulations under even and uneven environmental availability based on (a) U_{obs} and (b) U_{niche}. The dashed line represents the estimated differences under a complete sampling scenario.

them ($p > 0.68$). Under unbalanced subsampling, U_{obs} was always higher in the rarer habitats (A) regardless of the total number of samples, in contrast with results obtained from complete sampling ($p < 0.001$; Figure 3a). This indicates that rarer environments in the sampling pool exhibit higher ecological uniqueness. However, habitat differences of U_{niche} based on unbalanced subsampling were similar to estimations based on complete sampling ($p > 0.41$; Figure 3b). The total number of subsamples only altered the fluctuation of habitat differences across iterations, but not the conclusions on habitat differences in uniqueness.

As a proof of concept, we analysed the relationships between topographic variables and the compositional uniqueness of tree communities. We found high unevenness in environmental availability for different topographic variables, with the convex, flat and higher elevational plots being more common within the dataset (Figure 4a–c). We detected no effects of convexity (Figure 4d) and elevation (Figure 4e) on uniqueness for U_{obs} and U_{niche} (Table S2). We recorded minimal effects of slope on uniqueness when effects of environmental availability were not controlled, as demonstrated by the results based on U_{obs} (Figure 4f). The effects of slope changed from minimal to negative when environmental availability was controlled using U_{niche} (Figure 4f).

4 | DISCUSSION

Our simulation shows that conventional uniqueness metrics tend to be influenced by environmental availability, with rarer
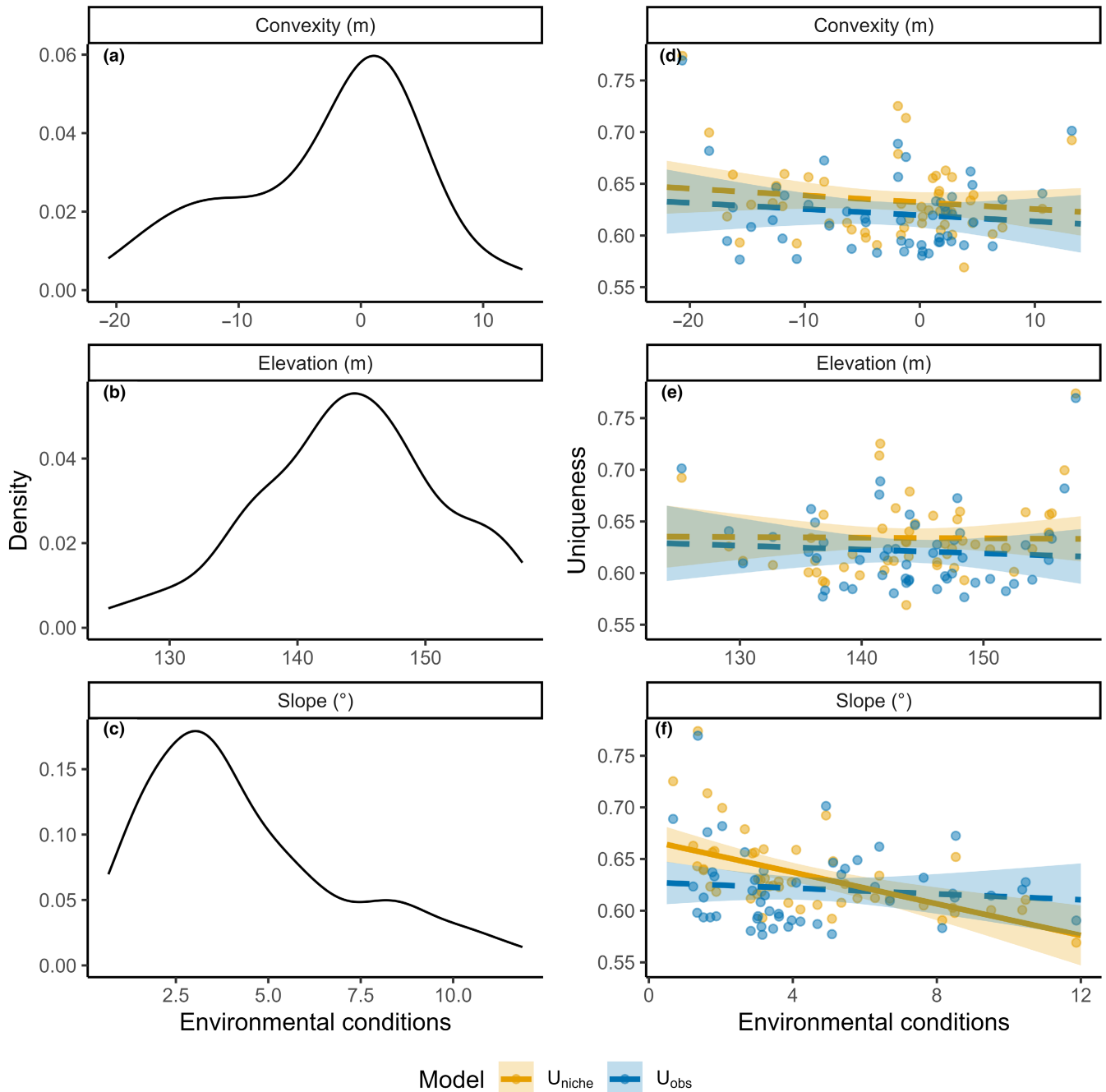
FIGURE 4 Results of empirical analyses using tree and topographical data from Barro Colorado Island of Panama. (a–c) density plots of each topographic variable illustrating their availability within the sampling pool. (d–f) predicted relationships between U_{obs} or U_{niche} with different topographic variables. Dashed and solid lines represent insignificant and significant relationships. Shaded areas represent 95% CI of predictions. Conclusions based on LCBD and distance to centroid were similar to U_{obs} and were presented in Table S2.

environments showing higher uniqueness. This result demonstrates that environmental availability can drive ecological uniqueness patterns even in the absence of a direct effect of the environment (Baidya & Bagchi, 2022; Dubois et al., 2020; Ejrnæs et al., 2018; Mokany, McCarthy, et al., 2022). Environmental availability can also obscure effects of how environmental conditions directly affect uniqueness, as demonstrated in our empirical analyses where slope emerged as an important predictor only after controlling environmental availability. This can be explained as

environmental availability promoting a positive relationship between slope and uniqueness due to flat areas being more common (Figure 4c), and subsequently opposing the negative direct effect of slope on the uniqueness of tree communities. Therefore, removing the effects of environmental availability can lead to changes in effect size or even qualitative conclusions. Thus, interpreting any uniqueness patterns requires the consideration of the role of environmental availability before attributing them to other mechanisms contributed by environmental drivers, such as

dispersal (Vilmi et al., 2017) and environmental tolerance of species associated with the environment (García-Navas et al., 2022).

Apart from statistical methods, altering the sampling design can also remove the effects of environmental availability. For nominal variables, the design should be as balanced as possible, while stratified sampling by environmental conditions (Restaino et al., 2016) or gridded sampling within environmental space (Perret & Sax, 2022) can be used for numeric variables. An advantage of directly analysing data from these sampling designs is that all errors can be retained during the analyses, whereas in our analyses the corrected ecological uniqueness is based on mean responses to each environmental condition only. Thus, a more appropriate sampling design is more preferable if environmental conditions of the study system are available, such as macroclimatic conditions, topography and landcover (Perret & Sax, 2022). However, in many systems, prior knowledge of environmental conditions is not available, and even when available such sampling design can be logistically challenging to achieve compared to random sampling (Isaac et al., 2020). Minimizing unevenness of environmental availability also becomes increasingly difficult when many environmental drivers are of interest. Studies analysing existing datasets such as long-term monitoring data (Allen et al., 2019) are unlikely to have even environmental availability. In these cases, a statistical procedure must be used to remove effects of environmental availability.

Our projection uses a niche space constructed as a bounding box formed by minima and maxima of all environmental variables of interest to obtain expected β_{pair}, followed by averaging to calculate ecological uniqueness. Yet in principle, any ecologically relevant environmental data can be used to obtain expected β_{pair}. One example is obtaining predictions under conditions identical to the observed sites, which effectively calculates β-diversity within the same environment. Many studies have calculated β-diversity within the same habitat group, such as natural, agricultural and urban habitats, before comparing them across groups (Newbold et al., 2016). However, our approach allows comparisons across continuous variables. Additionally, if future environmental conditions are available, our analyses can also be used to assess the extent of future changes in the species composition of different sites (Mokany, Ware, et al., 2022).

One drawback of our method is the large computational effort required when the dataset is large. We set 25 points by default evenly spaced across the gradient for any continuous variable. Therefore, the size of the niche space would be 25^n if there are n environmental variables of interest. The BCI dataset contains 50 communities with three topographic variables, and the analyses took 6s (using a Lenovo Legion 5 with AMD Ryzen 7 6800H and 16GB DDR5 memory). For large datasets with many sites and environmental variables, we recommend carefully selecting environmental variables before running analyses to minimize computational effort. Multicollinearity should be examined and, if present, some environmental variables should be combined using multivariate analyses such as PCA (Dormann et al., 2013; Zuur et al., 2010). Reducing the number of hypothetical sites is also possible, especially when the relationship between β_{pair} and environmental conditions is less complicated.

Our proposed analyses only address the averaging problem associated with uneven environmental availability. Another problem commonly associated with uneven environmental availability is that low sampling effort can result in poor estimations of community properties that influence β diversity, including gamma diversity and the shape of species abundance distributions (Brocklehurst et al., 2018; Engel et al., 2021; Lu et al., 2019; Schroeder & Jenkins, 2018), which can eventually propagate to uniqueness analyses. Such problems can also occur in studies having gridded or balanced design since sampling completeness might not be equal between environments even under identical sampling effort, which is well-known for α and γ diversity estimations (Chao & Jost, 2012) but also increasingly recognised for β diversity (Engel et al., 2021). In our study, we used Jaccard dissimilarity due to its popularity; though Jaccard can still be affected by other sampling issues but is relatively more robust compared to many other metrics (Beck et al., 2013; Cardoso et al., 2009; Schroeder & Jenkins, 2018). Recent studies have also proposed new metrics to minimise sampling effects on beta diversity patterns (Brocklehurst et al., 2018; Chao et al., 2005; Engel et al., 2021; Zou & Axmacher, 2020). More simulations are required to compare the performance of different indices under uneven environmental conditions, which is beyond the scope of this study. Selecting an appropriate metric can further enhance the performance of our analyses in identifying per-se effects of environmental conditions.

The accuracy in determining direct environmental effects on uniqueness also depends on how well other confounding factors are controlled, including the effects of distance between communities on β_{pair} (Nekola & White, 1999). As our analyses use GAMM from R package *mgcv* (Wood, 2017), correlation structures can be added to control for spatial effects, as in the case of our empirical analyses where we used a spatial exponential correlation structure. In some cases, the distance-decay relationships and therefore spatial effects can vary across environmental conditions (Newbold et al., 2016). Grouping factors can be added when specifying correlation structures in GAMM easily, such that only observations within environments (but not between environments) are correlated. For continuous environmental variables, users must manually specify a correlation structure. Because of the possibility to include correlation structures, our analyses can also be used to identify how environmental conditions drive temporal uniqueness, or both spatial and temporal uniqueness simultaneously (Tatsumi et al., 2021).

Apart from taxonomic identities, recent studies have also extended ecological uniqueness metrics to spectral, functional trait and phylogenetic data (Heino et al., 2022; Laliberté et al., 2020; Nakamura et al., 2020). Additionally, any β_{pair} metric can be decomposed into species replacement and richness difference components, followed by applying uniqueness metrics to calculate how sites contribute to these components in generating overall β patterns (Legendre, 2014). Our simulation and empirical example only focused on overall taxonomic uniqueness, but we believe this can be generalised to other uniqueness analyses, as the quantification of these metrics is based on the same principles (Heino et al., 2022; Laliberté et al., 2020; Nakamura et al., 2020). Overall, our proposed

analyses allow for comparisons of uniqueness patterns of different diversity dimensions based on observed data and after reducing the effects of uneven environmental availability, promoting a more comprehensive understanding of mechanisms generating ecologically unique communities.

AUTHOR CONTRIBUTIONS

Toby P. N. Tsang conceived the study, developed the methodology, analysed the data, and wrote the first draft of the manuscripts with inputs from all co-authors. All authors contributed to revision and editing.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14211.

DATA AVAILABILITY STATEMENT

The BCI dataset is available in R package *vegan*, while the topographic data are available in R package *BiodiversityR*. All R code is available on the GitHub repository https://github.com/tpaknok/Beta-diversity, where future updates will also be available. The version described in this paper is also archived in Zenodo at https://doi.org/10.5281/zenodo.8347047 (Tsang et al., 2023).

ORCID

Toby P. N. Tsang 🄳 https://orcid.org/0000-0002-8176-7777
Timothy C. Bonebrake 🄳 https://orcid.org/0000-0001-9999-2254
Lauren C. Ponisio 🄳 https://orcid.org/0000-0002-3838-7357
Marc W. Cadotte 🄳 https://orcid.org/0000-0002-5816-7693

REFERENCES

Allen, D. C., Bateman, H. L., Warren, P. S., de Albuquerque, F. S., Arnett-Romero, S., & Harding, B. (2019). Long-term effects of land-use change on bird communities depend on spatial scale and land-use type. *Ecosphere*, *10*(11), e02952. https://doi.org/10.1002/ecs2.2952

Anderson, M. J., Ellingsen, K. E., & McArdle, B. H. (2006). Multivariate dispersion as a measure of beta diversity. *Ecology Letters*, *9*(6), 683–693. https://doi.org/10.1111/j.1461-0248.2006.00926.x

Baidya, P., & Bagchi, S. (2022). Influence of human land use and invasive species on beta diversity of tropical ant assemblages. *Insect Conservation and Diversity*, *15*(1), 73–85. https://doi.org/10.1111/icad.12536

Barton, P. S., Cunningham, S. A., Manning, A. D., Gibb, H., Lindenmayer, D. B., & Didham, R. K. (2013). The spatial scaling of beta diversity. *Global Ecology and Biogeography*, *22*(6), 639–647. https://doi.org/10.1111/geb.12031

Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, *19*(1), 134–143. https://doi.org/10.1111/j.1466-8238.2009.00490.x

Beck, J., Holloway, J. D., & Schwanghart, W. (2013). Undersampling and the measurement of beta diversity. *Methods in Ecology and Evolution*, *4*(4), 370–382. https://doi.org/10.1111/2041-210x.12023

Brocklehurst, N., Day, M. O., & Fröbisch, J. (2018). Accounting for differences in species frequency distributions when calculating beta diversity in the fossil record. *Methods in Ecology and Evolution*, *9*(6), 1409–1420. https://doi.org/10.1111/2041-210X.13007

Cardoso, P., Borges, P. A. v., & Veech, J. A. (2009). Testing the performance of beta diversity measures based on incidence data: The robustness to undersampling. *Diversity and Distributions*, *15*(6), 1081–1090. https://doi.org/10.1111/j.1472-4642.2009.00607.x

Cetra, M., Mattox, G. M. T., Romero, P. B., & Escobar, S. H. (2022). Protected areas and compositional diversity of fish from Serranias Costeiras of the Ribeira de Iguape River basin, Southeast Brazil. *Neotropical Ichthyology*, *20*. https://doi.org/10.1590/1982-0224-2021-0130

Chao, A., Chazdon, R. L., Colwell, R. K., & Shen, T. (2005). A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecology Letters*, *8*(2), 148–159. https://doi.org/10.1111/j.1461-0248.2004.00707.x

Chao, A., & Jost, L. (2012). Coverage-based rarefaction and extrapolation: Standardizing samples by completeness rather than size. *Ecology*, *93*(12), 2533–2547. https://doi.org/10.1890/11-1952.1

Dansereau, G., Legendre, P., & Poisot, T. (2022). Evaluating ecological uniqueness over broad spatial extents using species distribution modelling. *Oikos*, *2022*(5), e09063. https://doi.org/10.1111/oik.09063

de Cáceres, M., Legendre, P., Valencia, R., Cao, M., Chang, L., Chuyong, G., Condit, R., Hao, Z., Hsieh, C., & Hubbell, S. (2012). The variation of tree beta diversity across a global network of forest plots. *Global Ecology and Biogeography*, *21*(12), 1191–1202. https://doi.org/10.1111/j.1466-8238.2012.00770.x

Devictor, V., Mouillot, D., Meynard, C., Jiguet, F., Thuiller, W., & Mouquet, N. (2010). Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: The need for integrative conservation strategies in a changing world. *Ecology Letters*, *13*(8), 1030–1040. https://doi.org/10.1111/j.1461-0248.2010.01493.x

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., & Leitão, P. J. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., Jombart, T., Larocque, G., Legendre, P., Madi, N., & Wagner, H. H. (2022). *adespatial: Multivariate multiscale spatial analysis*. R package version 0.3-20. https://CRAN.R-project.org/package=adespatial

Dubois, R., Proulx, R., & Pellerin, S. (2020). Ecological uniqueness of plant communities as a conservation criterion in lake-edge wetlands. *Biological Conservation*, *243*, 108491. https://doi.org/10.1016/j.biocon.2020.108491

Ejrnæs, R., Frøslev, T. G., Høye, T. T., Kjøller, R., Oddershede, A., Brunbjerg, A. K., Hansen, A. J., & Bruun, H. H. (2018). Uniquity: A general metric for biotic uniqueness of sites. *Biological Conservation*, *225*, 98–105. https://doi.org/10.1016/j.biocon.2018.06.034

Engel, T., Blowes, S. A., McGlinn, D. J., May, F., Gotelli, N. J., McGill, B. J., & Chase, J. M. (2021). Using coverage-based rarefaction to

infer non-random species distributions. *Ecosphere*, *12*(9), e03745. https://doi.org/10.1002/ecs2.3745

Ferrier, S., Manion, G., Elith, J., & Richardson, K. (2007). Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, *13*(3), 252–264. https://doi.org/10.1111/j.1472-4642.2007.00341.x

García-Navas, V., Martínez-Núñez, C., Tarifa, R., Molina-Pardo, J. L., Valera, F., Salido, T., Camacho, F. M., & Rey, P. J. (2022). Partitioning beta diversity to untangle mechanisms underlying the assembly of bird communities in Mediterranean olive groves. *Diversity and Distributions*, *28*(1), 112–127. https://doi.org/10.1111/ddi.13445

Graco-Roza, C., Aarnio, S., Abrego, N., Acosta, A. T. R., Alahuhta, J., Altman, J., Angiolini, C., Aroviita, J., Attorre, F., & Baastrup-Spohr, L. (2022). Distance decay 2.0—A global synthesis of taxonomic and functional turnover in ecological communities. *Global Ecology and Biogeography*, *31*(7), 1399–1421. https://doi.org/10.1111/geb.13513

Harwood, T., Love, J., Drielsma, M., Brandon, C., & Ferrier, S. (2022). Staying connected: Assessing the capacity of landscapes to retain biodiversity in a changing climate. *Landscape Ecology*, *37*, 3123–3139. https://doi.org/10.1007/s10980-022-01534-5

Heino, J., García Girón, J., Hämäläinen, H., Hellsten, S., Ilmonen, J., Karjalainen, J., Mäkinen, T., Nyholm, K., Ropponen, J., & Takolander, A. (2022). Assessing the conservation priority of freshwater lake sites based on taxonomic, functional and environmental uniqueness. *Diversity and Distributions*, *28*(9), 1966–1978. https://doi.org/10.1111/ddi.13598

Isaac, N. J. B., Jarzyna, M. A., Keil, P., Dambly, L. I., Boersch-Supan, P. H., Browning, E., Freeman, S. N., Golding, N., Guillera-Arroita, G., & Henrys, P. A. (2020). Data integration for large-scale models of species distributions. *Trends in Ecology & Evolution*, *35*(1), 56–67. https://doi.org/10.1016/j.tree.2019.08.006

Jyrkänkallio-Mikkola, J., Siljander, M., Heikinheimo, V., Pellikka, P., & Soininen, J. (2018). Tropical stream diatom communities—The importance of headwater streams for regional diversity. *Ecological Indicators*, *95*, 183–193. https://doi.org/10.1016/j.ecolind.2018.07.030

Kindt, R. (2022). *BiodiversityR: Package for community ecology and suitability analysis version 2.14-4*.

Laliberté, E., Schweiger, A. K., & Legendre, P. (2020). Partitioning plant spectral diversity into alpha and beta components. *Ecology Letters*, *23*(2), 370–380. https://doi.org/10.1111/ele.13429

Leão, H., Siqueira, T., Torres, N. R., & de Assis Montag, L. F. (2020). Ecological uniqueness of fish communities from streams in modified landscapes of eastern Amazonia. *Ecological Indicators*, *111*, 106039. https://doi.org/10.1016/j.ecolind.2019.106039

Legendre, P. (2014). Interpreting the replacement and richness difference components of beta diversity. *Global Ecology and Biogeography*, *23*(11), 1324–1334. https://doi.org/10.1111/geb.12207

Legendre, P., & de Cáceres, M. (2013). Beta diversity as the variance of community data: Dissimilarity coefficients and partitioning. *Ecology Letters*, *16*(8), 951–963. https://doi.org/10.1111/ele.12141

Lu, M., Vasseur, D., & Jetz, W. (2019). Beta diversity patterns derived from Island biogeography theory. *The American Naturalist*, *194*(3), E52–E65. https://doi.org/10.1086/704181

Marion, Z. H., Fordyce, J. A., & Fitzpatrick, B. M. (2017). Pairwise beta diversity resolves an underappreciated source of confusion in calculating species turnover. *Ecology*, *98*, 933–939. https://doi.org/10.1002/ecy.1753

Mokany, K., Ferrier, S., Harwood, T. D., Ware, C., di Marco, M., Grantham, H. S., Venter, O., Hoskins, A. J., & Watson, J. E. M. (2020). Reconciling global priorities for conserving biodiversity habitat. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(18), 9906–9911. https://doi.org/10.1073/pnas.1918373117

Mokany, K., McCarthy, J. K., Falster, D. S., Gallagher, R. v., Harwood, T. D., Kooyman, R., & Westoby, M. (2022). Patterns and drivers of plant diversity across Australia. *Ecography*, *2022*(11), e06426. https://doi.org/10.1111/ecog.06426

Mokany, K., Ware, C., Woolley, S. N. C., Ferrier, S., & Fitzpatrick, M. C. (2022). A working guide to harnessing generalized dissimilarity modelling for biodiversity analysis and conservation assessment. *Global Ecology and Biogeography*, *31*(4), 802–821. https://doi.org/10.1111/geb.13459

Mori, A. S., Isbell, F., & Seidl, R. (2018). β-diversity, community assembly, and ecosystem functioning. *Trends in Ecology & Evolution*, *33*(7), 549–564. https://doi.org/10.1016/j.tree.2018.04.012

Nakamura, G., Vicentin, W., Súarez, Y. R., & Duarte, L. (2020). A multifaceted approach to analyzing taxonomic, functional, and phylogenetic β diversity. *Ecology*, *101*(10), e03122. https://doi.org/10.1002/ecy.3122

Nekola, J. C., & White, P. S. (1999). The distance decay of similarity in biogeography and ecology. *Journal of Biogeography*, *26*(4), 867–878. https://doi.org/10.1046/j.1365-2699.1999.00305.x

Newbold, T., Hudson, L. N., Hill, S. L. L., Contu, S., Gray, C. L., Scharlemann, J. P. W., Börger, L., Phillips, H. R. P., Sheil, D., & Lysenko, I. (2016). Global patterns of terrestrial assemblage turnover within and among land uses. *Ecography*, *39*(12), 1151–1163. https://doi.org/10.1111/ecog.01932

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., de Caceres, M., Durand, S., … Weedon, J. (2022). *vegan: Community ecology package version 2.6-4*.

Perbiche-Neves, G., Saito, V. S., Simões, N. R., Debastiani-Júnior, J. R., Naliato, D. A. d. O., & Nogueira, M. G. (2019). Distinct responses of Copepoda and Cladocera diversity to climatic, environmental, and geographic filters in the La Plata River basin. *Hydrobiologia*, *826*(1), 113–127. https://doi.org/10.1007/s10750-018-3722-9

Perret, D. L., & Sax, D. F. (2022). Evaluating alternative study designs for optimal sampling of species' climatic niches. *Ecography*, *2022*(1). https://doi.org/10.1111/ecog.06014

Restaino, C. M., Peterson, D. L., & Littell, J. (2016). Increased water deficit decreases Douglas fir growth throughout western US forests. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(34), 9557–9562. https://doi.org/10.1073/pnas.1602384113

Ribeiro, C., Rodrigues, M. E., Sahlén, G., & de Oliveira Roque, F. (2022). Dragonflies within and outside a protected area: A comparison revealing the role of well-preserved Atlantic forests in the preservation of critically endangered, phytotelmatous species. *Journal of Insect Conservation*, *26*(2), 271–282. https://doi.org/10.1007/s10841-022-00385-4

Royston, P., Altman, D. G., & Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: A bad idea. *Statistics in Medicine*, *25*(1), 127–141. https://doi.org/10.1002/sim.2331

Schneck, F., Bini, L. M., Melo, A. S., Petsch, D. K., Saito, V. S., Wengrat, S., & Siqueira, T. (2022). Catchment scale deforestation increases the uniqueness of subtropical stream communities. *Oecologia*, *199*(3), 671–683. https://doi.org/10.1007/s00442-022-05215-7

Schroeder, P. J., & Jenkins, D. G. (2018). How robust are popular beta diversity indices to sampling error? *Ecosphere*, *9*(2), e02100. https://doi.org/10.1002/ecs2.2100

Shaw, R. E., Spencer, P. B., Gibson, L. A., Dunlop, J. A., Kinloch, J. E., Mokany, K., Byrne, M., Moritz, C., Davie, H., & Travouillon, K. J. (2022). Linking life history to landscape for threatened species conservation in a multiuse region. *Conservation Biology*, *37*, e13989. https://doi.org/10.1111/cobi.13989

So, K. Y. K., & Dudgeon, D. (2021). When is protection not conservation? A case study of semi-natural freshwater marshes in Hong Kong. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *31*(12), 3345–3356. https://doi.org/10.1002/aqc.3724

Socolar, J. B., Gilroy, J. J., Kunin, W. E., & Edwards, D. P. (2016). How should beta-diversity inform biodiversity conservation? *Trends in Ecology & Evolution*, *31*(1), 67–80. https://doi.org/10.1016/j.tree.2015.11.005

Stier, A. C., Geange, S. W., Hanson, K. M., & Bolker, B. M. (2013). Predator density and timing of arrival affect reef fish community assembly. *Ecology*, *94*(5), 1057–1068. https://doi.org/10.1890/11-1983.1

Tatsumi, S., Iritani, R., & Cadotte, M. W. (2021). Temporal changes in spatial variation: Partitioning the extinction and colonisation components of beta diversity. *Ecology Letters*, *24*(5), 1063–1072. https://doi.org/10.1111/ele.13720

Tsang, T. P. N., Bonebrake, T. C., Ponisio, L. C., & Cadotte, M. W. (2023). Controlling for the effects of environmental availability when testing how the environment determines community compositional uniqueness (v1.0.0). *Zenodo*, https://doi.org/10.5281/zenodo.7675752

Vilmi, A., Karjalainen, S. M., & Heino, J. (2017). Ecological uniqueness of stream and lake diatom communities shows different macroecological patterns. *Diversity and Distributions*, *23*(9), 1042–1053. https://doi.org/10.1111/ddi.12594

Willis, K. J., Jeffers, E. S., Tovar, C., Long, P. R., Caithness, N., Smit, M. G. D., Hagemann, R., Collin-Hansen, C., & Weissenberger, J. (2012). Determining the ecological value of landscapes beyond protected areas. *Biological Conservation*, *147*(1), 3–12. https://doi.org/10.1016/j.biocon.2011.11.001

Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279

Zou, Y., & Axmacher, J. C. (2020). The chord-normalized expected species shared (CNESS)-distance represents a superior measure of species turnover patterns. *Methods in Ecology and Evolution*, *11*(2), 273–280. https://doi.org/10.1111/2041-210X.13333

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, *1*(1), 3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x

SUPPORTING INFORMATION

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

Table S1: Calculations of different uniqueness metrics.

Table S2: Relationships between uniqueness and all uniqueness metrics for the case study using generalized least square regression.

Text S1: Simulation with numeric variables

Figure S1: Relationships between uniqueness metrics and the simulated environmental gradient across 100 simulations. The solid lines represent the average relationship across simulations, while the transparent lines represent results from individual simulations.

Figure S2: Correlations between different uniqueness metrics with U_{niche} under two unbalanced sampling scenarios. Because of the nearly perfect correlations among the three uniqueness metrics, the three metrics exhibited similar correlation patterns.