

## Using rarefaction to isolate the effects of patch size and sampling effort on beta diversity

ADRIAN C. STIER,<sup>1,†</sup> BENJAMIN M. BOLKER,<sup>2</sup> AND CRAIG W. OSENBURG<sup>3</sup>

<sup>1</sup>*Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, California 93106 USA*

<sup>2</sup>*Departments of Mathematics & Statistics and Biology, McMaster University, Hamilton, Ontario L8S 4K1 Canada*

<sup>3</sup>*Odum School of Ecology, University of Georgia, Athens, Georgia 30602 USA*

**Citation:** Stier, A. C., B. M. Bolker, and C. W. Osenberg. 2016. Using rarefaction to isolate the effects of patch size and sampling effort on beta diversity. *Ecosphere* 7(12):e01612. 10.1002/ecs2.1612

**Abstract.** Beta diversity describes how species composition varies across space and through time. Current estimators of beta diversity typically ignore the effects of within-patch sample size, determined jointly by local abundance and sampling effort. Many ecological processes such as immigration, predation, or nutrient limitation affect abundance and asymptotic beta diversity concurrently; thus, existing metrics may confound changes in asymptotic beta diversity with changes that result from differences in abundance or sampling. Results from a stochastic simulation model illustrate how decreasing within-patch sample size may either increase or decrease observed beta diversity, depending on the type of metric, sample size, and community properties; these changes are easy to understand, and predict, by considering the effects of sampling on variance. A modified, patch-level form of rarefaction controls for variation in within-patch sample size; two case studies illustrate the utility of this approach. Studies seeking a mechanistic link between ecological process and beta diversity will continue to benefit from explicit consideration of sampling effects.

**Key words:** beta diversity; biodiversity; rarefaction; sampling effects.

**Received** 6 October 2016; **accepted** 10 October 2016. Corresponding Editor: Debra P. C. Peters.

**Copyright:** © 2016 Stier et al. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

† **E-mail:** adrian.stier@lifesci.ucsb.edu

### INTRODUCTION

Within-patch (alpha) and among-patch (beta) diversities combine to produce regional patterns of diversity. Traditionally, ecologists have focused on local processes and their influence on alpha diversity. More recently, ecologists have turned to studying the drivers of beta diversity; patterns of beta diversity provide insight into the role of factors that generate species boundaries (Condit et al. 2002), facilitate regional diversity via environmental heterogeneity, and produce alternate community states (Fukami and Nakajima 2011). From a conservation perspective, beta diversity is arguably as important as alpha diversity (Socolar et al. 2016): For example, quantifying the spatial scale of beta diversity can allow

resource managers to design reserve networks that maximize species complementarity across sites (Kattan et al. 2006). Similarly, beta diversity can be used to describe the rate at which species turn over through time in the presence of human impacts (Dornelas et al. 2014). Knowledge of the factors maintaining diversity among patches can also facilitate biodiversity preservation or recovery of diversity following the restoration of a degraded ecosystem (McKnight et al. 2007).

Beta diversity is typically examined at two spatial scales: (1) large-scale variation in beta diversity that links species compositional differences across gradients (e.g., with latitude or altitude) and encompasses multiple species pools (Condit et al. 2002, McKnight et al. 2007, He and Zhang 2009), and (2) local-scale variation in beta

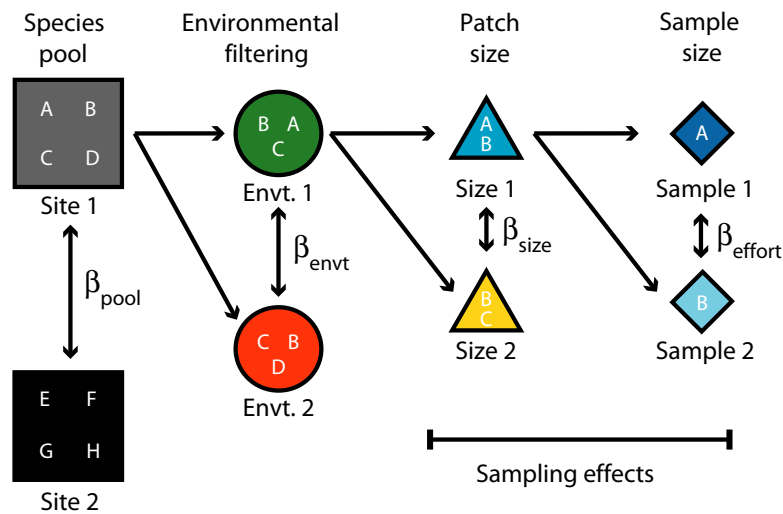


Fig. 1. Four different mechanisms of beta diversity where each letter (A–H) represents a species. Beta diversity can be produced by (1) differences in the species pool supplying two sites (gray vs. black,  $\beta_{\text{pool}}$ ), (2) differences in the environment (green vs. red,  $\beta_{\text{envt}}$ ), (3) differences in patch size (blue vs. yellow,  $\beta_{\text{size}}$ ), or (4) differences in sampling effort (light blue vs. dark blue,  $\beta_{\text{effort}}$ ).

diversity among patches representing a community drawn from a single pool. At the local scale, beta diversity is determined by four different, scale-dependent phenomena (Fig. 1): (1) changes in the species pool, which affect beta diversity primarily at large spatial scales (and which we do not consider further); (2) environmental heterogeneity driven by biotic or abiotic factors (i.e., environmental filtering); (3) local patch size (i.e., the total number of individuals found within a given location): For example, small patches may have high beta diversity because each patch has only a limited capacity to sample the species pool; (4) sampling effort: For example, the use of a few, small quadrats within a patch will reduce the number of individuals counted within patches, which could lead to high apparent beta diversity even when patches contain identical communities. Here, we focus on the latter two phenomena, patch size and sampling effort, as the roles of the species pool and environmental heterogeneity have been thoroughly discussed elsewhere in the literature (e.g., Soininen et al. 2007).

The effects of patch size and sampling effort (which we collectively call *sampling effects*) are well known in studies of alpha diversity. Well-established tools, such as rarefaction or multinomial extrapolation, facilitate the comparison of

alpha diversity among communities that differ in sampling protocols (e.g., number of samples, individuals, or coverage [fraction of the species list sampled]; Gotelli and Colwell 2001, 2011, Chao and Jost 2012, Chao et al. 2012, 2014b, Colwell et al. 2012). The importance of sampling effects in biasing estimates of beta diversity was initially demonstrated in the 1950s (Nash 1950, Morisita 1959), and has since received additional attention (Ricklefs and Lau 1980, Wolda 1981, Plotkin and Muller-Landau 2002, Cardoso et al. 2009, Tuomisto 2010a, Beck et al. 2013). Initial attempts to address sampling bias in beta diversity estimators have included identifying beta diversity metrics that are less sensitive to sampling effects (Cardoso et al. 2009) and developing metrics that adjust beta diversity for unsampled species (Chao et al. 2005). Despite these advances, ecologists have not yet developed tools that clearly separate sampling effects from other phenomena driving beta diversity. For example, ecological processes such as predation, competition, and nutrient limitation can modify beta diversity both via environmental filtering and via sampling effects (i.e., by altering the number of individuals surviving within a patch; e.g., Chase 2007, 2010, Chase et al. 2009, Stier et al. 2013). As a result, sampling effects and the influences of environmental drivers of

beta diversity may be confounded. At present, we know of no effective technique that adjusts observed beta diversity for differences in abundance to separate sampling effects from environmental filtering and facilitate comparison of communities (as is standard practice in studies of alpha diversity).

We focus on the consequences of sampling effects (i.e., phenomena 3, patch size, and 4, sample size: Fig. 1) for beta diversity estimation in scenarios with a fixed species pool. We examined these two phenomena together, as their consequences are functionally equivalent (we have the same information about the composition of a patch from which only a few individuals have been sampled whether it was the patch or the investigator that did the sampling). We first developed a stochastic simulation model to explore how sampling affects beta diversity under different situations. We then developed a rarefaction technique that adjusts observed beta diversity for sampling effects and allows the separation of sampling effects from higher-level ecological drivers (environmental filtering, species pool compositions) of beta diversity.

## METHODS

### *The model*

In contrast to recent work that generates simulated communities from a complex spatial model (Beck et al. 2013), we used a relatively simple stochastic simulation. Our approach sacrificed realism to facilitate interpretation of the processes driving the resulting patterns in beta diversity. We defined the beta diversity of a community as the variation in species composition among samples obtained from patches within that community. Variation is measured as distance in  $N$ -state space (where  $N$  is the species richness of the regional pool), where each species is either represented by presence-absence data (e.g., Jaccard index) or relative abundance (e.g., Manhattan index). There is a plethora of beta diversity definitions (for a recent review, see Legendre and De Cáceres 2013); we focused on metrics based on distance in  $N$ -state space (i.e., multivariate dispersion sensu Anderson et al. 2006), because we were specifically interested in comparing beta diversity between multiple distinct communities; however, we expect that our results will be qualitatively

similar for other metrics (see *Discussion*). Instead, we focused on metrics that quantify the variability among patches within a community.

We first used a simulation model to explore how observed beta diversity for a single community is affected by sampling (due to changes in either sample size or patch size) and species abundance distributions (SADs). All code is available in Appendix S1; community simulator is detailed in Appendix S2: Figs. S1–S7. Source code, R Package, and the code corresponding to each figure is archived at: <https://zenodo.org/record/162251#.WEoYJHeZPOQ>. In particular, we asked how beta diversity changes as sampling increases and how this effect is influenced by patterns of endemism and rarity. Hereafter, we refer to the expected beta diversity under large samples as *asymptotic beta diversity* to distinguish it from the observed (i.e., sampled) beta diversity.

We also used the model to investigate the situation in which two communities are compared (e.g., to determine differences in beta diversity between sites with and without predators). These communities can represent different locations, habitats, or treatments (each containing multiple sampled replicates or plots, “patches” in our terminology). The questions then become: Do different habitats differ in beta diversity? How is the observed magnitude of difference in beta diversity affected by sample size, patch size, or the species abundance distributions? Does a rarefaction-based approach effectively separate the effect of sampling from differences in asymptotic beta diversity?

To explore these issues, we simulated species relative abundances for each of  $P$  patches within a community pool of  $N$  species. For the first step, we defined the species abundance distribution (assumed the same for each patch). We set the number of discrete abundance categories (e.g., three, for [common, intermediate, rare]), the number of species per abundance category, and the relative abundance of successive abundance categories (defined by the rank-abundance parameter  $\phi$ , which is the abundance of the  $x$ th category relative to the  $x + 1$ st category). We normalized the results (divided by the total patch abundance) so that species proportions summed to 1.0. For example, with two abundance categories, three species per category, and  $\phi = 2$ , the relative species abundances would be  $(1, 1, 1, 0.5, 0.5, 0.5)/4.5$ ; with two species in each of three abundance

Table 1. Example eight species community matrices generated from three-step process of the stochastic simulation model.

	Species							
	A	B	C	D	E	F	G	H
Pre-mixing ( $r_i$ )								
Patch 1	1/3	1/3	1/6	1/6	0	0	0	0
Patch 2	0	0	0	0	1/3	1/3	1/6	1/6
Post-mixing ( $r_i \times p_{\text{mix}}/S$ )								
Patch 1	1/3	1/3	1/8	1/8	0	0	1/24	1/24
Patch 2	0	0	1/24	1/24	1/3	1/3	1/8	1/8
Numbers								
Patch 1	40	40	15	15	0	0	5	5
Patch 2	0	0	5	5	40	40	15	15
50% Poisson sample	22	19	5	6	0	0	2	3
sample	0	0	4	5	21	17	8	6

Notes: Columns (A–H) represent species and fractions represent relative abundance of each species. From top to bottom, the three steps include pre-mixing relative abundance ( $r_i$ ), post-mixing relative abundance ( $r_i \times p_{\text{mix}}/S$ ), numbers of individuals per patch (i.e., multiplying relative abundance by patch size), and a 50% sample of the mixed community. Model parameters:  $S = 2$ ,  $\phi = 2$ ,  $N = 120$ , and  $p_{\text{mix}} = [0, 0.5]$ .

categories and  $\phi = 10$ , the relative abundances would be (1,1,0.1,0.1,0.01,0.01)/2.22.

Each patch initially contained a unique set of species, but with identical SADs (i.e.,  $N/P$  species in each patch). We then mixed the species among patches to achieve more complex assemblages. We allowed for a separate mixing parameter ( $p_{\text{mix}}$ ) for each abundance category. The mixing parameter specified the proportion of individuals of each species to be redistributed uniformly across all patches: 0 corresponded to no mixing and 1 corresponded to complete mixing. Thus, for a patch without species  $i$ , the new relative abundance of species  $i$  became  $r_i \times p_{\text{mix}}/P$  (where  $r_i$  refers to the initial relative abundance), while the relative abundance in the original patch became  $r_i(1 - p_{\text{mix}}) + (r_i \times p_{\text{mix}}/P)$ . For example, for a simulation with two abundance categories (common and rare) and  $p_{\text{mix}} = [0, 0.5]$ , the common species remained completely endemic to patches (i.e., no mixing), while half of the individuals from rare species were distributed homogeneously across all patches (including a fraction  $p_{\text{mix}}/P$  that were distributed back to their original patch): Table 1 gives an example.

Once the species distributions across patches were set, we took finite samples from each patch. For this sampling problem, there are (at least) two

possible approaches: (1) We could sample a fixed number of individuals, in which case, each sample would be based on a multinomial distribution; or (2) we could sample from the expected number of individuals for each species using a Poisson distribution. While both procedures give the same expected proportions, in the first case the total number of individuals per sample is identical, while in the second case the total number of individuals varies across samples according to a Poisson distribution. Table 1 gives one example for the second (Poisson) approach, which we follow for the remainder of the paper. While zero-inflated Poisson distributions better match the species distributions for some real communities (Zhang et al. 2014), we use the Poisson approach here for simplicity. We concentrate on the application of the Jaccard index based on presence-absence data because it is most commonly used in the literature, but we also report results for an abundance-based metric, the Manhattan index using a Hellinger transformation before calculating the distance matrix (sensu Legendre and De Cáceres 2013; Appendix S1: Fig. S6).

#### A beta rarefaction solution across multiple communities

The effects we describe for a single community can affect comparisons of beta diversity between two or more communities. We therefore developed and implemented a three-step approach to rarefaction to quantify rarefied beta diversity and compare communities (i.e., to compare beta diversity conditional on sampling the same total numbers of individuals per patch in each community). We focus on individual-based rather than sample-based rarefaction here, because community studies frequently take a single sample per patch (e.g., Chase 2007, 2010, Chase et al. 2009, Stier et al. 2013); in this case, sample-based rarefaction would be useless, as it would effectively resample patches, which would not mitigate the effects of reduced numbers of individuals *within* patches. While multilevel rarefaction has been previously suggested as a way to account for under-sampling bias in beta diversity (e.g., Tuomisto 2010a, b), ours is to our knowledge the first attempt to flesh out the details of how such a procedure would work in practice. In contrast to rarefaction for alpha diversity, multilevel rarefaction for beta diversity requires some extra attention to

detail. If, in order to match within-patch sample sizes across communities, researchers simply subsampled all patches down to the minimum sample size found in any patch in *any* community, they would often discard almost all of their data (because sample sizes are often highly variable among patches). (Horner-Devine et al. (2004) did successfully use this protocol, in a molecular ecological context with large sample sizes.) Instead, we follow a slightly more sophisticated rank-matching algorithm.

1. Rank samples (i.e., patches) by the number of individuals sampled (i.e., sample size). Match patches by rank across treatments. For all patches in the same rank, subsample all patches down to the minimum sample size in those patches. For example, assume we have three samples from each of two communities with samples sizes of [10,7,3] and [8,6,5]. We then subsample the two communities down to [8,6,3] producing equal sample sizes in the two communities. If there are more than two communities, we find the minimum number of individuals per patch rank across all communities; for unbalanced designs with different numbers of patches, we group multiple patches into patch ranks. For example, consider three communities with [10,7], [30,7], and [20,8,5,3]. We would create two groups and then subsample down to [8,3] for the first two communities and [8,8,3,3] for the third community. Matching groups becomes slightly trickier when the numbers of patches per community are not evenly divisible (e.g., five patches in community 1, seven patches in community 2, and 11 patches in community 3). Extending this algorithm to sample-based rarefaction is possible in principle, but the bookkeeping for equating total numbers of individuals sampled per patch across communities while reducing the total numbers sampled as little as possible becomes extremely complex.
2. For each treatment (e.g., *Control*), repeat this subsampling process 1000 times. For each subsample, compute beta diversity. In our case, we calculated diversity using Jaccard dissimilarity and computed the median bootstrapped distance from each patch to the respective treatment centroid (Anderson

et al. 2006), after applying a bias correction (see below).

3. Compare these rarefied beta diversity distributions for each treatment group using a two-tailed test.

Because unequal numbers of patches among communities can affect comparisons of multivariate distance, we applied a correction proportional to  $\sqrt{P/(P-1)}$  (where  $P$  is the number of patches) when estimating the distances to centroid. This bias correction accounts for the fact that the distances to centroid are slightly correlated with each other (because the centroid is itself determined by combining all of the individual community values); it is analogous to the bias correction typically used in estimating sample variances, that is,  $s^2 = \text{SSQ}/(n-1)$  rather than  $\text{SSQ}/n$ . It is novel, although Anderson (2006) mentions the possibility of such bias and O'Neill and Mathews (2000) discuss this type of correction in more detail, albeit in a different context. An option for applying this bias correction is included in the `betadisper` function in `vegan` package (versions  $\geq 2.0-7$ ) for community analysis (Oksanen et al. 2016).

#### Power analysis

To test whether our procedure works to distinguish differences in observed beta diversity due to sampling effects from those due to filtering, we ran a power analysis with two treatments in which we varied average patch size ( $N = 10$  or  $20$ ); proportional difference in patch size between treatments (0%, 5%, 10%, and 50%); difference in number of patches per treatment ([20,20]; [10,30]; and [5,35]); and, as our measure of effect size, the between-treatment difference in the mixing parameter  $p_{\text{mix}}$  (an average value of 0.5, with differences ranging from 0 to 0.5, i.e.,  $p_{\text{mix}} = [0.25, 0.75]$ ). We simulated 250 cases for each parameter combination, ran the hierarchical rarefaction and beta diversity comparison test procedures described above for each simulation, and counted the proportion of time that the null hypothesis of no difference in beta diversity was rejected. For the null case ( $p_{\text{mix}}$  difference = 0), the resulting proportion is a measure of the type I error rate, which should be equal to the nominal  $\alpha = 0.05$ ; otherwise, the proportion estimates the power to reject the null hypothesis.

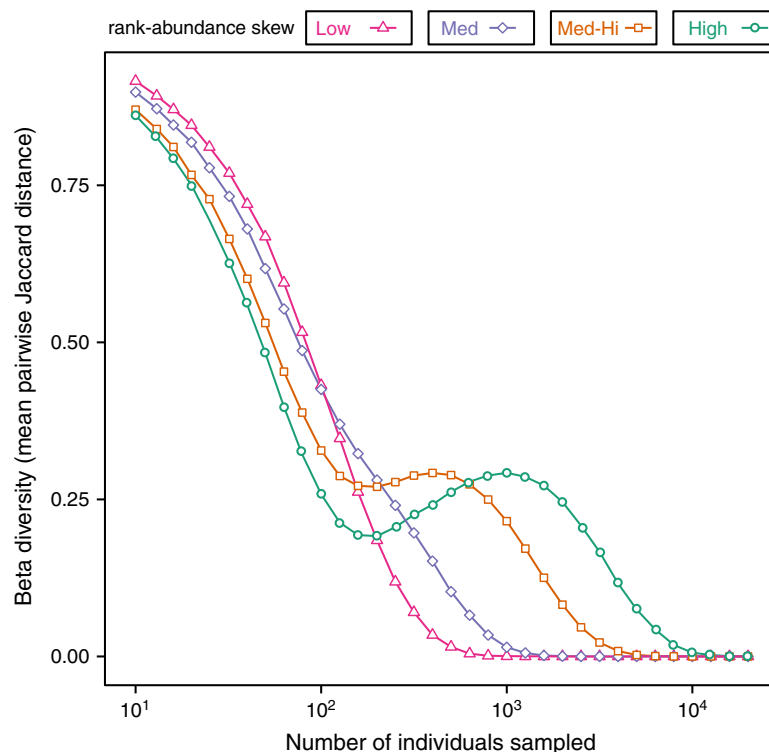


Fig. 2. The effect of local population size on observed beta diversity quantified as the mean pairwise Jaccard distance for four different rank-abundance distributions. Note the interactive effects of the rank-abundance parameter and local population size on the shape of the rarefaction curves. As the rank-abundance parameter becomes smaller, communities become more skewed in abundance with a few common individuals and a larger number of rare species. Intermediate maxima emerge for medium-high and high levels of skew. Population size on  $x$ -axis is plotted on a  $\log_{10}$  scale. See Appendix S3, Fig. S3 for similar patterns with other beta diversity metrics (e.g., Gower, alt-Gower, Bray, Canberra, Chao-Jaccard, Gower, Horn, Manhattan, Morista, and Raup-Crick indices).

### Case studies

To illustrate how rarefaction can facilitate inferences about the ecological processes driving beta diversity, we used data from two recently published experiments that examined the effect of two different predator species on the beta diversity of their prey, coral reef fish (Stier et al. 2013, 2014). In each experiment, a matrix of structurally similar and evenly spaced reefs was constructed from a combination of cinderblocks and live coral. Prey fish then naturally colonized these reefs in the presence and absence of experimentally manipulated predatory fish. We show the effects of predators on abundance, richness, Jaccard index of beta diversity, and the proposed rarefied Jaccard index.

## RESULTS

### Effects of sampling intensity on beta diversity in simulated communities

Increasing sample size had large and context-dependent effects on the observed beta diversity. Beta diversity generally declined as the number of individuals sampled per patch increased (Fig. 2). Similar patterns with other beta diversity metrics include Gower, Alt-Gower, Bray-Curtis, Canberra, Chao-Jaccard, Gower, Manhattan, Morista, and Raup-Crick indices (Appendix S3: Figs. S1-S3). However, when very rare species existed in the species pool—achieved in this case by making the rank-abundance parameter,  $\phi$ , small—the Jaccard-based beta diversity curve had an

intermediate maximum. This pattern arose from the effects of the variance of a binomial sample. When the probability of a species occurrence was very small, that species occurred nowhere or almost nowhere, and the diversity due to that species was low; the diversity also was low when the occurrence probability was very high and the species was present (almost) everywhere. Diversity was highest when the species occurrence probability was near 50%. However, if the probability of species occurrence was already at or above 50% in the smallest sample sizes considered (which was always true for the common species, and true for the rare species when the rank-abundance parameter was large), then beta diversity always declined with increasing sampling intensity. Looking at a broader range of SADs and endemism (Fig. 3), we only saw intermediate maxima when the rank-abundance parameter was small and beta diversity was driven by both common and rare species. A more pronounced intermediate maximum occurred at larger local sample sizes when the common species were better mixed than the rare species. In this model, beta diversity was 100% when there was no mixing (upper left panel) or when sample sizes were very small (left edge of each panel). Similar intermediate maxima in the relationship between sampling intensity and beta diversity have been documented empirically (Cardoso et al. 2009), emphasizing the importance of rarefaction. Beta diversity decreased to zero as sample size increased when each species had a non-zero probability of occurring in every patch.

This phenomenon of variance initially increasing with sampling effort should apply only for presence-absence metrics such as the Jaccard. For abundance-based metrics such as the Manhattan index (Appendix S3: Fig. S1), the relevant sampling distribution is the Poisson, and beta diversity—driven by the coefficient of variation across patches—decreased continuously with increasing within-patch sample size. We also evaluated a number of other common beta diversity indices and demonstrated the ubiquity of sampling issues (Appendix S3: Fig. S3); the pattern of continuous decrease in observed beta diversity with increased within-patch sample size holds generally, although the effects of different SADs and numbers of patches sampled vary across abundance-based metrics.

Beta diversity studies, and especially the development of new ways to quantify and partition diversity, have exploded in recent years (Chao et al. 2005, 2014a, Crist and Veech 2006, Anderson et al. 2011, Chao and Chiu 2016). For example, Tuomisto (2010a, b) emphasized metrics that represent ratios between global (gamma) and local (alpha) diversities, rather than directly quantifying among-community diversity. We have not evaluated the performance of the new generation of beta diversity metrics, focusing instead on the more classical approaches implemented within the vegan package (and still widely used in empirical studies). However, given that our qualitative results are driven by the basic properties of sampling variances, we expect that they will generalize to any characterization of beta diversity that attempts to measure among-patch diversity in some form.

#### Power analysis

The results of the power analysis (Fig. 4) were encouraging. If our rarefaction method successfully accounted for differences in number of individuals within a patch, then for simulations with differing numbers of individuals within a patch but the same true asymptotic beta diversity (i.e.,  $p_{\text{mix}} = 0$ ), the rarefied comparison should have rejected the null hypothesis 5% of the time (because we used  $\alpha = 0.05$ ). When the number of patches did not differ too radically between treatments (i.e., [20,20] or [10,30]), our method was actually conservative: The rejection rate for null simulations was about 1% rather than 5% (averaging across all other parameters: rejection rate = 0.95% for [20,20], 1.5% for [10,30], standard error of 0.3%). (We note that the implementation of Anderson's (2006) PERMDISP2 procedure in the vegan package, betadisper, is also conservative in this case.) For large differences in the number of patches ([5,35]) and a patch size of 10 individuals, our rarefaction method gave a reasonable rejection rate of 5.9% (SE 0.4%); however, it became anticonservative (rejection rate = 13%, SE 0.9%) for patch size of 20 (Fig. 4; see also Appendix S4: Fig. S1). Thus, we caution against using our method to compare treatments with radically different numbers of patches, although we would expect that such unbalanced designs would only occur for opportunistic and/or observational data sets.

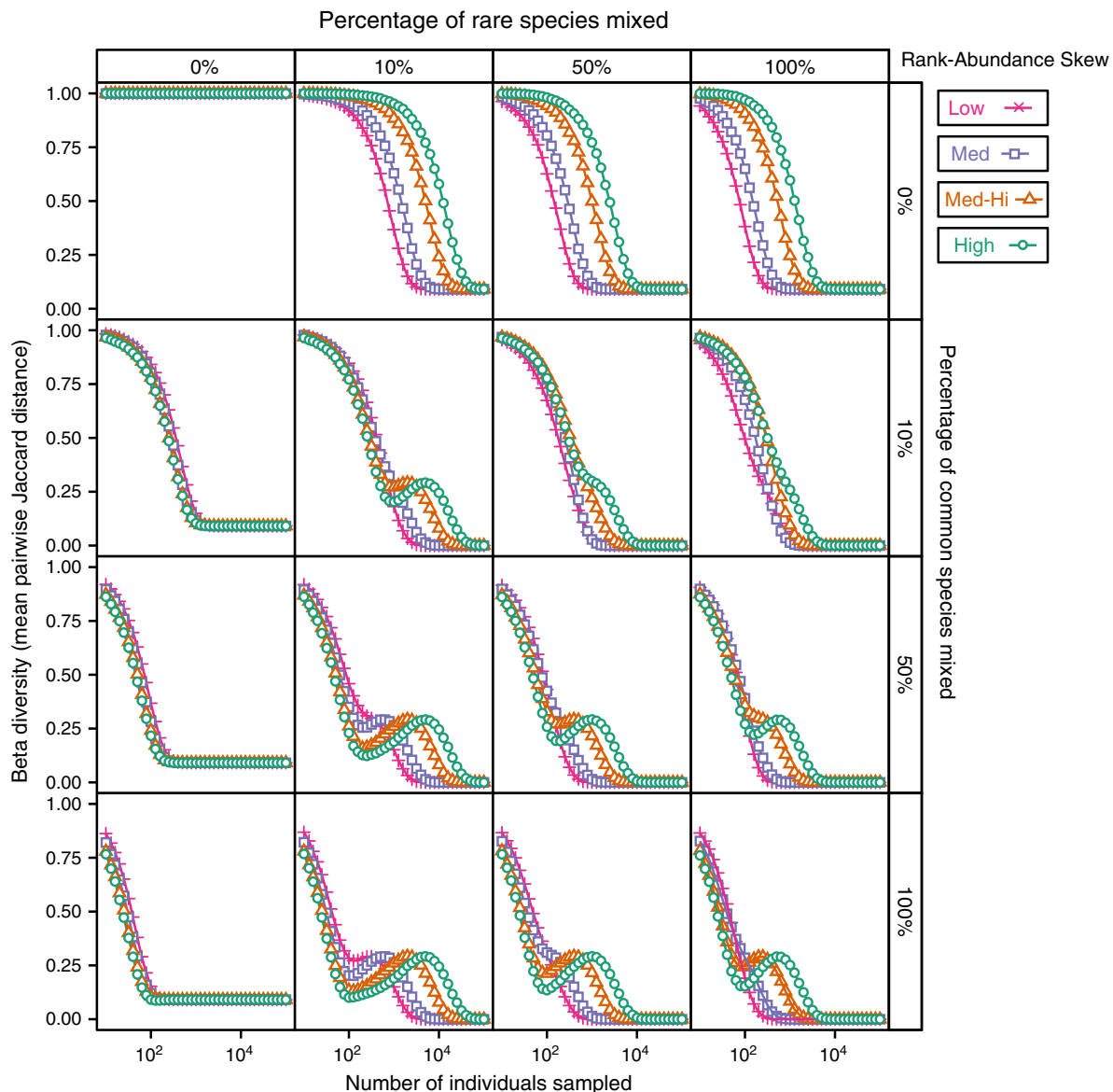


Fig. 3. Beta diversity can emerge from spatial variation in both common and rare species. Shown here is beta diversity measured as mean pairwise Jaccard distance for a response surface of beta diversity originating from rare species (from left to right) and common species (top to bottom) for communities of variable rank-abundance skew. Note the effects of rank-abundance skew are on the intermediate maxima are present only when beta diversity originates from some fraction of both rare and common species in the community, with the most pronounced intermediate maxima for communities with high rank-abundance skew and a large fraction of the beta diversity coming from both common and rare species (i.e., bottom right panel).

When there was a true among-treatment difference in asymptotic beta diversity, our method successfully captured the effects. Overall, a difference in mixing rates of approximately 0.2 (i.e., species are mixed across patches at rate 0.4 for one

treatment and rate 0.6 for the other) was sufficient to attain 80% power, with higher power being obtained in balanced designs and with larger population sizes (and correspondingly lower power for unbalanced designs and small patch sizes).

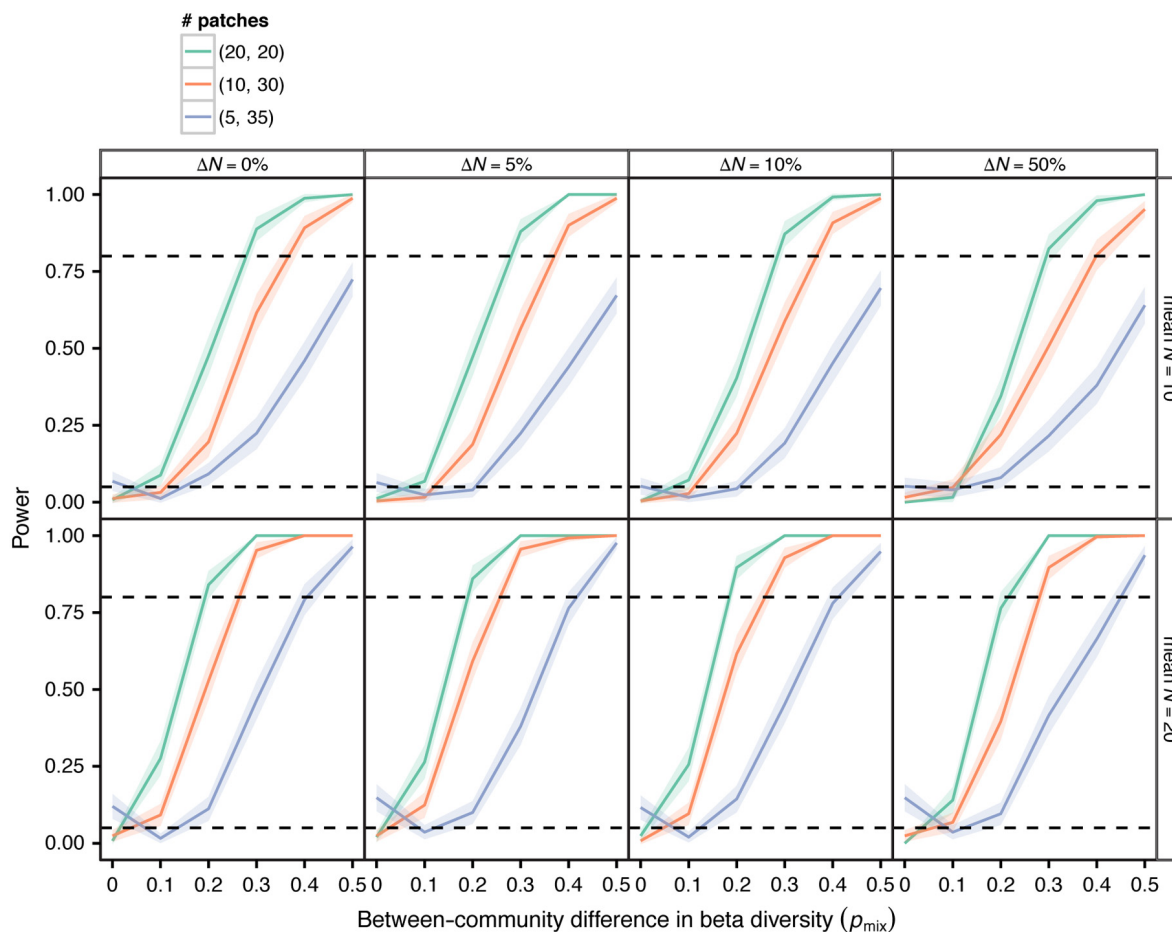


Fig. 4. Power analysis results for varying number of patches in each of two communities (colors): average per-patch population size (rows) and proportional difference in per-patch population size between communities (columns). Horizontal dashed lines show power = 0.05 (nominal  $\alpha$  level) and power = 0.8 (a standard cutoff for “adequate” power).

#### Using rarefaction to interpret the effects of predators on beta diversity

The two case studies showed that predators differed substantially in their effects on prey abundance and alpha diversity. The first predator, the arc-eye hawkfish (*Paracirrhites arcatus*, hereafter hawkfish), decreased prey abundance (Fig. 5A) and alpha diversity (Fig. 5B). Based on the decline in prey density (i.e., patch size), the model described above and the recent work presented by Chase et al. (2011), we predicted that hawkfish would cause a corresponding increase in beta diversity measured as Jaccard dissimilarity. As expected, beta diversity increased in the presence of hawkfish by 38% (Fig. 5C). Importantly, this

increase in Jaccard dissimilarity persisted even after rarefaction (Fig. 4D; albeit with a smaller effect size of a 26% increase), suggesting that predators increased beta diversity through two mechanisms: a reduction in the number of individuals (a sampling-driven effect) and a reduction in sampling-adjusted beta diversity. In this case, sampling effects inflated the positive effects of predators on beta diversity; however, in other cases, sampling effects might mask (or reverse) the apparent effects of a given ecological process on beta diversity. This was exemplified by our second example. The peacock grouper (*Cephalopholis argus*, hereafter grouper) strongly decreased prey abundance (Fig. 5E) and species

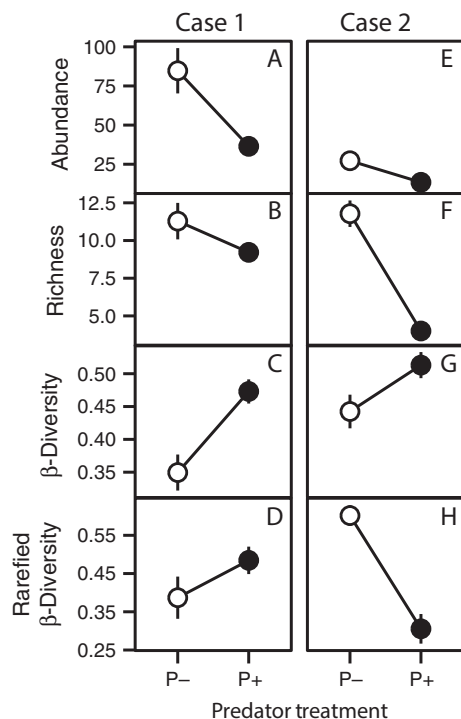


Fig. 5. Abundance (A, E), richness (B, F), beta diversity (C, G), and rarefied beta diversity (D, H) in the presence (P+, black circles) and absence (P-, white circles) of predators in two separate experiments on coral reef fish communities in French Polynesia. In the first case study (left column; Stier et al. 2013), positive effects of predators (hawkfish) on observed beta diversity (Jaccard index) persist for rarefied beta diversity (rarefied Jaccard index). In contrast, the second case study (right column; Stier et al. 2014) shows that positive effects of predators (peacock grouper) on observed beta diversity are reversed to a negative effect following rarefaction. These two case studies demonstrate that sampling effects can inflate effect sizes when estimating the effect of a given ecological process that also decreases abundance (left column). Furthermore, sampling effects can be so strong that they mask the direction of the effect of a process on beta diversity (right column). Data for these analyses are available in Appendices S6 and S7.

richness (Fig. 5F), and *increased* (non-rarefied) beta diversity by 16% (Fig. 5G). However, rarefied beta diversity *decreased* in the presence of groupers by 50% (Fig. 5H). The decline in (rarefied) beta diversity in the presence of groupers was likely driven by predators acting as

environmental filters, removing poorly defended species from the species pool (Stier et al. 2014); however, this effect was masked in the analysis of unrarefied beta diversity because of the stronger sampling effect mediated through the decline in prey density.

## DISCUSSION

Recent studies exploring the ecological processes driving beta diversity have by and large neglected the biases introduced by limited sampling. The sampling effects we illustrate here highlight the need for accessible and accurate tools to quantify beta diversity within communities or compare beta diversity among communities. Our results corroborate recent empirical and theoretical studies that demonstrated the importance of sampling effects in driving beta diversity (Cardoso et al. 2009, Beck et al. 2013). We extend the current understanding of sampling effects in three distinct ways. (1) We conceptually distinguish among environmental filtering and two sources of sampling effects—patch size and sampling effort; (2) we show that the effects of sampling on beta diversity depends on the shape of the rank-abundance curve and whether beta diversity originates from common or rare species; and (3) we develop a new method to calculate rarefied beta diversity for two or more sets of patches, and use two cases that demonstrate its utility. This method is implemented in an R package, *betararef* (<https://github.com/bbolker/betararef>; Appendix S1).

### Effects of sampling effort vs. patch size

We have discussed the effects of patch size and sampling effort on observed beta diversity in a single framework because they are computationally similar. The horizontal axes in Figs. 2 and 3 could be changed from “number of individuals sampled” to “patch size” or “sample size,” changing their interpretation without changing their shape. What matters from an estimation point of view is only the expected abundances or incidences in a sample; it does not matter whether 10 individuals counted represent a 10% sample of a population of size 100, or a complete sample of a population of 10. However, the interpretation of the results may be quite different. Changes in beta diversity driven strictly by

incomplete sampling will disappear if communities are exhaustively sampled. In contrast, a set of patches can exhibit beta diversity because a limited number of individuals can persist within a given patch: That is, in these cases, the patches themselves have incompletely “sampled” the potential residents. Therefore, processes such as predation, drought, or resource limitation that drive local patch size can produce distinct patterns of beta diversity. Such effects cannot be overcome by increased sampling effort. To illustrate this issue, consider predator–prey interactions that are influenced by the availability of predator-free space (Jeffries and Lawton 1984), such that the maximum number of prey in a patch is a function of the number of refuges. In the presence of predators, the number of available refuges sets patch size, a limitation that disappears in the absence of predators. As we have shown (Figs. 2, 3, 5), communities with predators may therefore have different observed beta diversity relative to those without predators.

#### *Observed, rarefied, and asymptotic beta diversity*

It is useful to distinguish among *observed* (i.e., unrarefied) beta diversity (beta diversity as computed directly from observed samples, without any form of sample size adjustment), *rarefied* beta diversity (beta diversity adjusted for sampling effects by appropriate subsampling methods), and *asymptotic* beta diversity (the expected beta diversity from a large sample, drawn from a large local community); similar distinctions have been made in the context of alpha diversity estimation (Colwell et al. 2012, Chao et al. 2014b). While rarefaction is a useful tool for understanding and adjusting for sampling effects in beta estimators, rarefaction procedures necessarily involve throwing away data. For example, to quantify rarefied beta diversity in the case studies described above, we rarefy a group of patches (i.e., reduce the sampling effort in some patches) and then quantify beta diversity; when there is wide variation in local patch size or sampling effort so that some patches have very small numbers of individuals, the resulting loss of information can be catastrophic. This loss of data is important, because it can limit the power we have to detect differences in beta diversity between two communities. In particular, at very small sample sizes, the difference between two

rarefaction curves can be very hard to detect, as they will converge to nearly 100% dissimilarity as the number of individuals per patch approaches a single individual.

In the future, we would ideally like a strategy for estimating asymptotic rather than rarefied diversity, which would adjust for differences in sampling effort and/or local patch size without discarding data and losing power. Such solutions for alpha diversity are far ahead of current methods for beta diversity (e.g., Chao et al. 2013), although asymptotic alpha diversity estimates have their own set of constraints. In particular, Haegeman et al. (2013) have established that asymptotic species richness cannot be reliably estimated for strongly undersampled communities such as environmental or host-associated microbiomes (also see Mao and Colwell 2005, Chao et al. 2014b). For stable conclusions, researchers must either use more robust metrics associated with higher Hill numbers (i.e., Simpson diversity corresponding to the Hill number with  $q = 2$ ; Chao et al. 2014b) or compare richness at a common rarefied reference size (Navas-Molina et al. 2013). We hypothesize a similar set of limitations may be true for asymptotic beta diversity.

Despite these limitations, analogous methods for estimating asymptotic beta diversity would be extremely useful. Although such methods do not yet exist, recent advances in quantifying species rank-abundance distributions (RADs) may provide a starting point. Chao et al. (2015) highlight that the observed relative abundance for any species is an overestimate of its true value, due to the failure to detect species that are actually present in the community. This results in a biased RAD and the well-known underestimation of true alpha diversity. Importantly, Chao et al. provide a framework to estimate the number and relative abundance of undetected species inferred to be present in the assemblage, thus closing the gap between observed and asymptotic alpha diversity. This approach might be extended to beta diversity, but the significant challenge is that the undetected species in a sample are not actually identified, and thus, asymptotic beta diversity cannot be clearly delineated. There may be indirect ways to infer (dis)similarity between the undetected species, but this vexing problem does not have an apparent (or necessarily viable) solution.

Given that we cannot yet estimate asymptotic beta diversity for patches of finite size, we suggest empiricists quantify both observed and rarefied beta diversity. Researchers should also consider beta diversity metrics such as the Chao-Jaccard index, which adjusts beta diversity for the number of unseen species (Chao et al. 2005) and the modified Raup-Crick index, which adjusts beta diversity for the interdependency of alpha diversity and beta diversity (i.e., assuming that beta diversity is estimable as the difference between alpha, beta, gamma diversity on an additive or multiplicative scale; Crist and Veech 2006, Chase et al. 2011). Such a multimetric approach, while dense, allows researchers to use the best available tools, and to evaluate the robustness of their conclusions, the contribution of sampling effects to those patterns, the importance of unseen species, and the role of interdependency between alpha diversity and beta diversity.

Our case studies demonstrated how calculating observed and rarefied beta diversity can provide insight into the role of a given process in driving beta diversity. In both experiments, predators caused substantial reductions in abundance. This effect of predators on abundance drove differences in observed beta diversity between predator and no-predator treatments, in one case but not the other, by limiting the degree to which a given patch sampled the species pool. In circumstances where communities can be completely sampled (e.g., as was true with the case studies described here), rarefaction can be used to gain additional insight into whether shifts in beta diversity are driven solely by changes in effective patch size, or whether the observed beta diversity is driven by a combination of patch size effects and other effects such as disproportionate effects of a given process on certain species. For example, if predators reduce patch size through generalist predation (i.e., by affecting all species in proportion to their abundance), then observed beta diversity will be lower in the presence of predators (due to reduced sampling), but rarefied beta diversity will be similar in the two treatments (Almany and Webster 2004). In contrast, if predators exhibit positive frequency dependent foraging (e.g., by disproportionately consuming common species), or promote the persistence of certain

species by modifying competitive dynamics, then the effects of predators on beta diversity should persist even after rarefaction.

The two case studies presented here show that sampling effects can mask the qualitative and quantitative effects of predators on beta diversity. In our first case study, we showed that the positive effect of hawkfish on observed prey diversity persists for rarefied beta diversity, suggesting hawkfish positively affected beta diversity beyond affecting patch size. The authors hypothesized that hawkfish increased rarefied beta diversity by increasing the strength of competitive priority effects (Stier et al. 2013). Sampling effects inflated the positive effects of hawkfish on beta diversity, but the qualitative effect of hawkfish remained positive. In the second study, however, sampling effects masked the qualitative effect of predators on asymptotic beta diversity with a positive effect of groupers on observed beta diversity, and a negative effect of groupers on rarefied diversity. Rarefaction showed that predators reduced beta diversity more than expected based on their effects on patch size. These negative effects of groupers on beta diversity were likely driven by a homogenization of prey communities through the elimination of a number of species from the species pool.

#### *Individual- vs. sample-based rarefaction*

Our proposed method invokes individual-based rarefaction rather than sample-based rarefaction (sensu Gotelli and Entsminger 2001). A recent paper by Dornelas et al. (2014) implemented sample-based rarefaction when quantifying temporal changes in beta diversity at a global scale. Relative to rarefying by individual, sample-based rarefaction preserves species aggregation characteristics; however, the bias between rarefied and asymptotic beta diversity we detail above and in Appendix S5 can still persist when using sample-based rarefaction. For example, imagine a nutrient addition experiment where there are two treatments with and without nutrient additions (N+ and N−, respectively), each with three replicates. At the end of the experiment, the N+ plots have 30, 40, and 45 individuals and the N− plots have 5, 10, and 15 individuals. Because there is no difference in the number of plots, sample-based rarefaction is not warranted, yet an asymmetry in abundance

remains. Therefore, sample-based rarefaction will not address all sampling issues when estimating beta diversity.

## CONCLUSION

Our stochastic sampling model predicts a negative or occasionally uni- or multimodal relationship between within-patch sample size and beta diversity. These results corroborate recent studies that demonstrate the importance of sampling effects in driving beta diversity (Cardoso et al. 2009, Beck et al. 2013). Our study identifies the classes of communities most prone to sampling effects and evaluates two different types of sampling effects: (1) the effects of patch size and (2) the effects of sampling effort. The historical absence of rarefaction tools in beta diversity literature may limit the inference we can take from previous studies that have examined the ecological drivers of beta diversity. The direction and strength of bias in previously published studies is difficult to discern given the intermediate maxima present for the relationship between population size and beta diversity. Acknowledging sampling effects and addressing them using rarefaction is key to empiricists attempting to understand the ecological processes that drive beta diversity and community assembly.

## ACKNOWLEDGMENTS

We thank four anonymous reviewers and R. Colwell, H. Stevens, J. Oksanen, M. O'Connor, J. Samhouri, and O. Shelton for constructive feedback. B. Bolker was funded by SHARCnet and NSERC Discovery grant no. 386590-2010.

## LITERATURE CITED

Almany, G. R., and M. S. Webster. 2004. Odd species out as predators reduce diversity of coral-reef fishes. *Ecology* 85:2933–2937.

Anderson, M. J. 2006. Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* 62:245–253.

Anderson, M. J., K. E. Ellingsen, and B. H. McCauley. 2006. Multivariate dispersion as a measure of beta diversity. *Ecology Letters* 9:683–693.

Anderson, M. J., et al. 2011. Navigating the multiple meanings of beta diversity: a roadmap for the practicing ecologist. *Ecology Letters* 14:19–28.

Beck, J., J. D. Holloway, and W. Schwanghart. 2013. Undersampling and the measurement of beta diversity. *Methods in Ecology and Evolution* 4:370–382.

Cardoso, P., P. A. Borges, and J. A. Veech. 2009. Testing the performance of beta diversity measures based on incidence data: the robustness to undersampling. *Diversity and Distributions* 15:1081–1090.

Chao, A., R. L. Chazdon, R. K. Colwell, and T. J. Shen. 2005. A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecology Letters* 8:148–159.

Chao, A., and C. H. Chiu. 2016. Bridging the variance and diversity decomposition approaches to beta diversity via similarity and differentiation measures. *Methods in Ecology and Evolution* 7:919–928. <https://doi.org/10.1111/2041-210X.12551>

Chao, A., C.-H. Chiu, and T. Hsieh. 2012. Proposing a resolution to debates on diversity partitioning. *Ecology* 93:2037–2051.

Chao, A., C.-H. Chiu, and L. Jost. 2014a. Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through Hill numbers. *Annual Review of Ecology, Evolution, and Systematics* 45:297–324.

Chao, A., N. Gotelli, T. C. Hsieh, E. Sander, K. H. Ma, R. K. Colwell, and A. M. Ellison. 2014b. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs* 84:46–67.

Chao, A., T. C. Hsieh, R. L. Chazdon, R. K. Colwell, and N. J. Gotelli. 2015. Unveiling the species-rank abundance distribution by generalizing the Good-Turing sample coverage theory. *Ecology* 96:1189–1201.

Chao, A., and L. Jost. 2012. Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology* 93:2533–2547.

Chao, A., Y. T. Wang, and L. Jost. 2013. Entropy and the species accumulation curve: a novel entropy estimator via discovery rates of new species. *Methods in Ecology and Evolution* 4:1091–1100.

Chase, J. M. 2007. Drought mediates the importance of stochastic community assembly. *Proceedings of the National Academy of Sciences USA* 104:17430–17434.

Chase, J. M. 2010. Stochastic community assembly causes higher biodiversity in more productive environments. *Science* 328:1388–1391.

Chase, J. M., E. G. Biro, W. A. Ryberg, and K. G. Smith. 2009. Predators temper the relative importance of stochastic processes in the assembly of prey meta-communities. *Ecology Letters* 12:1210–1218.

Chase, J. M., N. J. Kraft, K. G. Smith, M. Vellend, and B. D. Inouye. 2011. Using null models to disentangle

- variation in community dissimilarity from variation in  $\alpha$ -diversity. *Ecosphere* 2:1–11.
- Colwell, R. K., A. Chao, N. J. Gotelli, S. Y. Lin, C. X. Mao, R. L. Chazdon, and J. T. Longino. 2012. Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *Journal of Plant Ecology* 5:3–21.
- Condit, R., et al. 2002. Beta-diversity in tropical forest trees. *Science* 295:666–669.
- Crist, T. O., and J. A. Veech. 2006. Additive partitioning of rarefaction curves and species–area relationships: unifying  $\alpha$ -,  $\beta$ - and  $\gamma$ -diversity with sample size and habitat area. *Ecology Letters* 9:923–932.
- Dornelas, M., N. J. Gotelli, B. McGill, H. Shimadzu, F. Moyes, C. Sievers, and A. E. Magurran. 2014. Assemblage time series reveal biodiversity change but not systematic loss. *Science* 344:296–299.
- Fukami, T., and M. Nakajima. 2011. Community assembly: Alternative stable states or alternative transient states? *Ecology Letters* 14:973–984.
- Gotelli, N. J., and R. K. Colwell. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* 4:379–391.
- Gotelli, N. J., and R. K. Colwell. 2011. Estimating species richness. *Biological Diversity: Frontiers in Measurement and Assessment* 12:39–54.
- Gotelli, N. J., and G. L. Entsminger. 2001. Swap and fill algorithms in null model analysis: rethinking the knight's tour. *Oecologia* 129:281–291.
- Haegeman, B., J. Hamelin, J. Moriarty, P. Neal, J. Dushoff, and J. S. Weitz. 2013. Robust estimation of microbial diversity in theory and in practice. *ISME Journal* 7:1092–1101.
- He, K., and J. Zhang. 2009. Testing the correlation between beta diversity and differences in productivity among global ecoregions, biomes, and biogeographical realms. *Ecological Informatics* 4:93–98.
- Horner-Devine, M. C., M. Lage, J. B. Hughes, and B. J. Bohannan. 2004. A taxa–area relationship for bacteria. *Nature* 432:750–753.
- Jeffries, M. J., and J. H. Lawton. 1984. Enemy free space and the structure of ecological communities. *Biological Journal of the Linnean Society* 23:269–286.
- Kattan, G. H., P. Franco, C. A. Saavedra-Rodriguez, C. Valderrama, V. Rojas, D. Osorio, and J. Martinez. 2006. Spatial components of bird diversity in the Andes of Colombia: implications for designing a regional reserve system. *Conservation Biology* 20:1203–1211.
- Legendre, P., and M. De Cáceres. 2013. Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. *Ecology Letters* 16:951–963.
- Mao, C. X., and R. K. Colwell. 2005. Estimation of species richness: mixture models, the role of rare species, and inferential challenges. *Ecology* 86:1143–1153.
- McKnight, M. W., P. S. White, R. I. McDonald, J. F. Lamoreux, W. Sechrest, R. S. Ridgely, and S. N. Stuart. 2007. Putting beta-diversity on the map: broad-scale congruence and coincidence in the extremes. *PLoS Biology* 5:e272.
- Morisita, M. 1959. Measuring interspecific association and similarity between communities. *Memoirs of the Faculty of Science, Kyushu University, Series E* 3:65–80.
- Nash, C. B. 1950. Associations between fish species in tributaries and shore waters of western Lake Erie. *Ecology* 31:561–566.
- Navas-Molina, J. A., J. M. Peralta-Sánchez, A. González, P. J. McMurdie, Y. Vázquez-Baeza, Z. Xu, L. K. Ursell, C. Lauber, H. Zhou, and S. J. Song. 2013. Advancing our understanding of the human microbiome using QIIME. *Methods in Enzymology* 531:371.
- Oksanen, J., F. Guillaume Blanchet, R. Kindt, P. Legendre, R. B. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. Stevens, and H. Wagner. 2016. *vegan: community ecology package*. <https://cran.r-project.org/web/packages/vegan/index.html>
- O'Neill, M. E., and K. Mathews. 2000. A weighted least squares approach to Levene's test of homogeneity of variance. *Australian & New Zealand Journal of Statistics* 42:81–100.
- Plotkin, J. B., and H. C. Muller-Landau. 2002. Sampling the species composition of a landscape. *Ecology* 83:3344–3356.
- Ricklefs, R. E., and M. Lau. 1980. Bias and dispersion of overlap indices: results of some Monte Carlo simulations. *Ecology* 61:1019–1024.
- Socolar, J. B., J. J. Gilroy, W. E. Kunin, and D. P. Edwards. 2016. How should beta-diversity inform biodiversity conservation? *Trends in Ecology & Evolution* 31:67–80.
- Soininen, J., J. J. Lennon, and H. Hillebrand. 2007. A multivariate analysis of beta diversity across organisms and environments. *Ecology* 88:2830–2838.
- Stier, A. C., S. W. Geange, and B. M. Bolker. 2013. Predator density and competition modify the benefits of group formation in a shoaling reef fish. *Oikos* 122:171–178.
- Stier, A. C., K. M. Hanson, S. J. Holbrook, R. J. Schmitt, and A. J. Brooks. 2014. Predation and landscape

- characteristics independently affect reef fish community organization. *Ecology* 95:1294–1307.
- Tuomisto, H. 2010*a*. A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography* 33:2–22.
- Tuomisto, H. 2010*b*. A diversity of beta diversities: straightening up a concept gone awry. Part 2. Quantifying beta diversity and related phenomena. *Ecography* 33:23–45.
- Wolda, H. 1981. Similarity indices, sample size and diversity. *Oecologia* 50:296–302.
- Zhang, J., T. O. Crist, and P. Hou. 2014. Partitioning of alpha and beta diversity using hierarchical Bayesian modeling of species distribution and abundance. *Environmental and Ecological Statistics* 21:611–625.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online at: <http://onlinelibrary.wiley.com/doi/10.1002/ecs2.1612/full>

## Appendix S1 Code/introduction

All the code that we use in the paper, and in the appendix, is available as an R package. It can be installed via

```
library(ghit)
install_github("bbolker/betararef")
```

You will need to install the `ghit` package from any standard CRAN repository first (if you have the `devtools` package installed, you can use that instead). Since the package contains only R code (no compiled C/C++/FORTRAN code), you should *not* need to have other development tools installed.

To begin using the package you need to load it (it needs to be installed only once per R installation, but loaded every time you start a new R session):

```
library(betararef)
```

## Appendix S2 Community simulator

The simulation first simulates a set of ‘endemic’ species in a specified number of patches, then mixes some proportion of each species randomly across all patches according to specified rules. The parameters of the simulation are as follows:

**n.abund** number of abundance categories (e.g. 2={common, rare}; 3={common, intermediate, rare})

**p.abund** proportional change between abundance categories (rank-abundance curve is geometric: overall frequencies of abundance categories are proportional to  $\{1, \phi, \phi^2, \dots\}$ )

**spcat** number of species per site per abundance category (i.e. the total number of (endemic) species per site is **spcat\*n.abund**, and the complete size of the species pool is **spcat\*n.abund\*n.site**): may be  $< 1$ , but the total number of species per abundance category (**spcat\*n.site**) must be an integer

**n.site** number of sites

**n.indiv.site** number of individuals per site

**p.mix** vector of mixing probabilities for each abundance category: 0=no mixing, 1=complete mixing, e.g. when **n.abund**=2,  $\{0,1\}$  means that common species are completely mixed and rare species are completely endemic

**rand** “none”: *expected* number of individuals per site/species is reported; “multinomial”: number of individuals per site is held constant, proportions chosen randomly; “poisson”: number of individuals per site/species is Poisson distributed

In general the simulator returns a site  $\times$  species matrix, which we can post-process with the tools from **vegan** to compute beta diversities.

Here are some typical examples. If we use all of the default values for **betasim**, we get 3 sites (rows) with a total of 30 species: 5 abundance classes  $\times$  2 species per site  $\times$  3 sites. By default there is no mixing among sites (100% beta diversity), so looking at the default plot (rows=sites, columns=species) we see that each site has blocks of 2 species from each abundance class (grayscale indicates prevalence) (Figure S1). (The default for **betasim** (but not for **simComm**) is to introduce no randomness at all, simply returning a matrix with the *expected* number of individuals of each species at each site.)

If we allow 100% mixing, then all species are present at all sites in the same numbers (Figure S2). We can also allow partial mixing (Figure S3) and Poisson variation in the output (Figure S4).

We can change the number sites or the number of abundance classes: by default, the **spcat** (number of species per site per abundance category) will be adapted to try to make the total number of species come out correctly (i.e. **spcat** will be set to **totsp/(n.site\*n.abund)**: Figure S5). If **totsp** is not an even multiple of **n.site\*n.abund**, the results may be slightly unpredictable ...

**calcbeta** calculates distances from the centroid (by default, using a Jaccard index on presence-absence data: Figure S6).

**simComm** simulates multiple communities (with possibly different characteristics) at the same time (Figure S7).

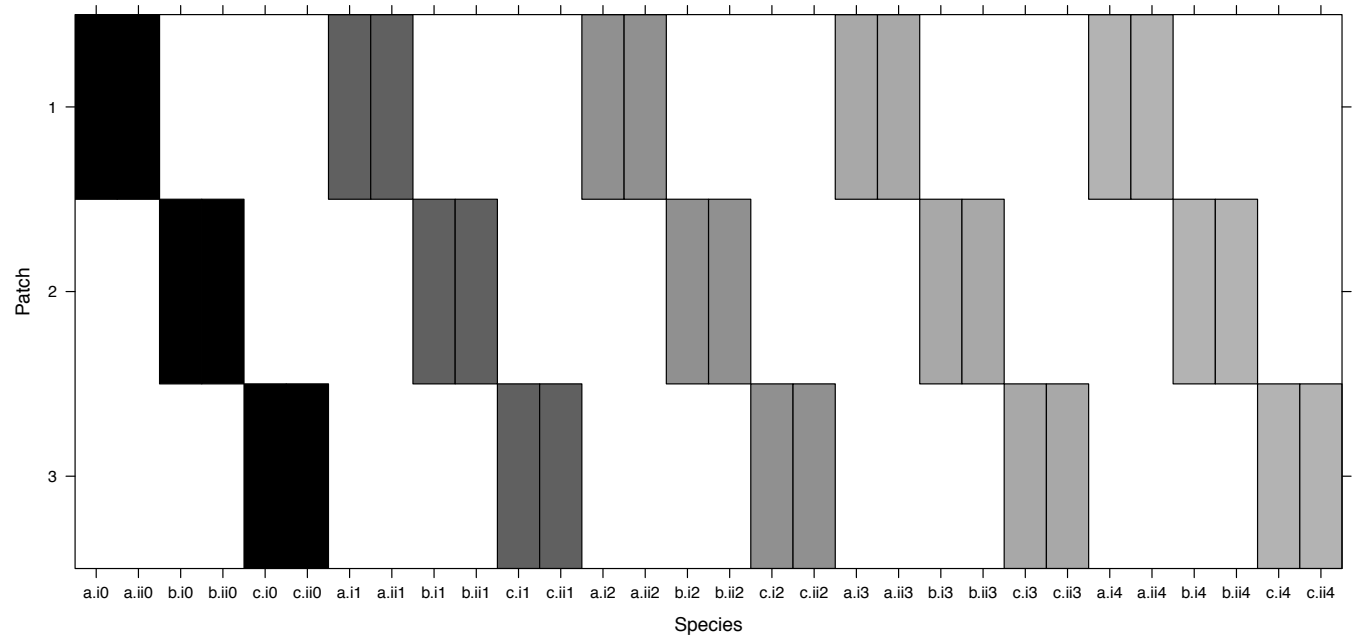


Figure S1: Species matrix for default `betasim` settings: 3 sites, 10 species per site (5 abundance classes/2 spp per abundance class), no mixing.

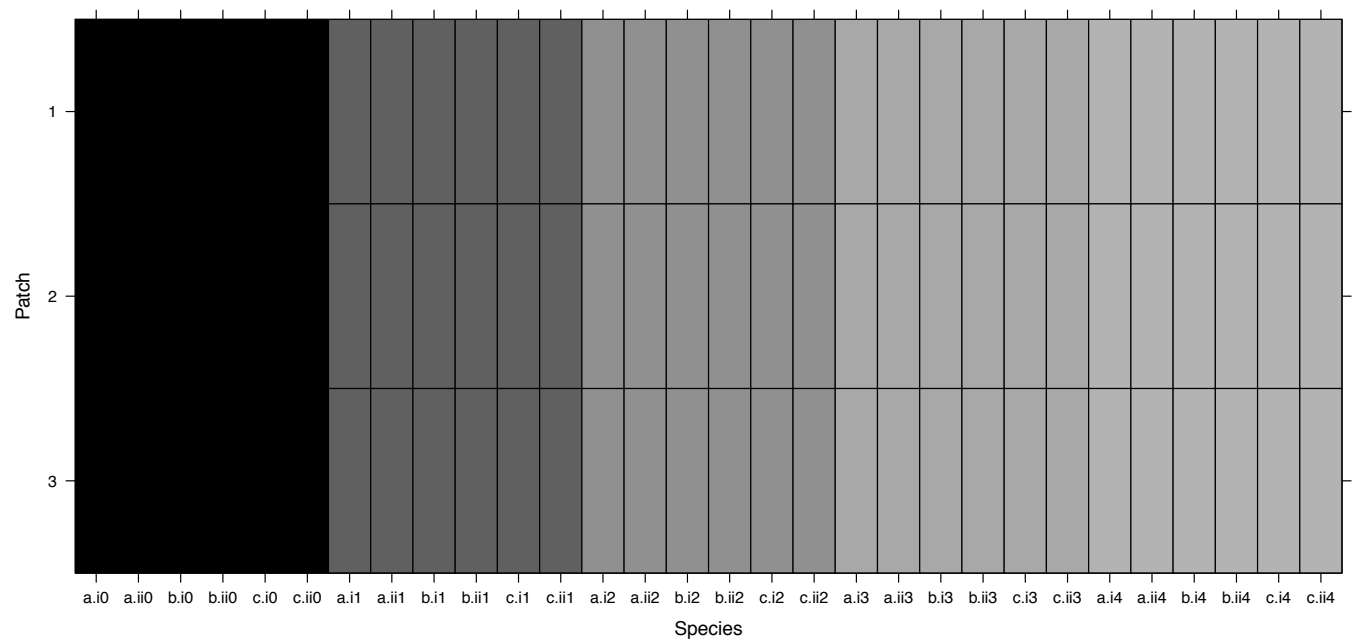


Figure S2: Species matrix for `betasim()` with default parameters (3 sites/5 abundance classes/2 spp per AC) but 100% mixing.

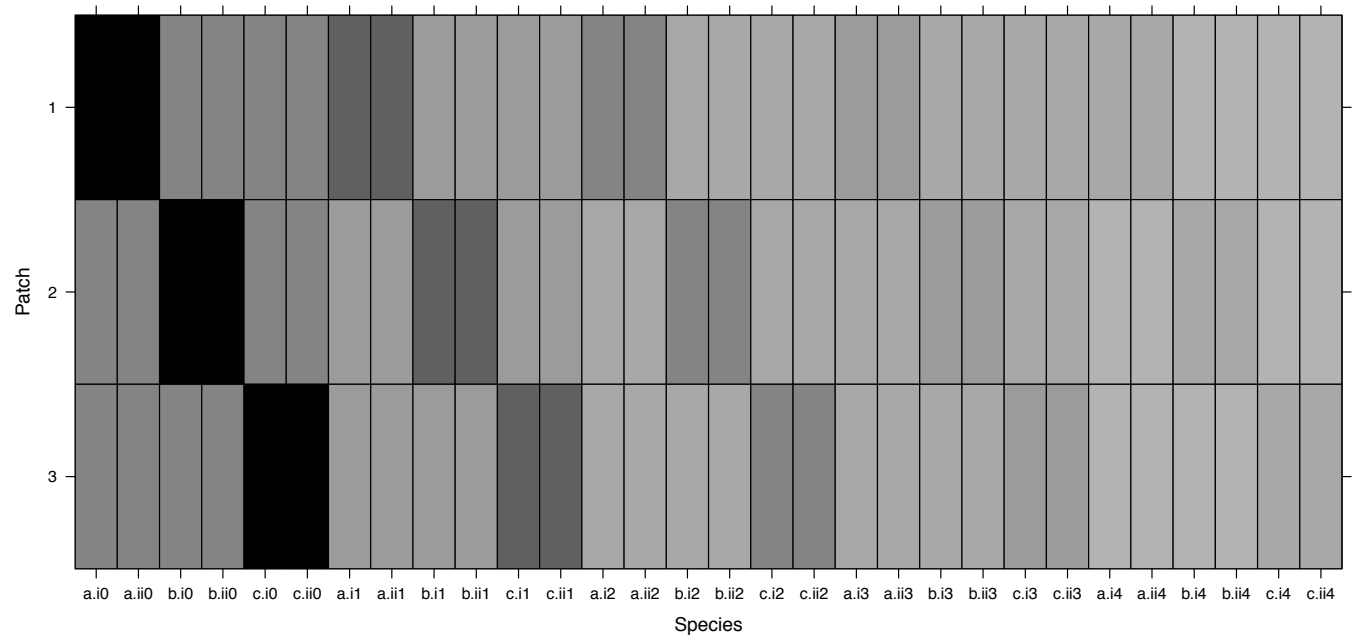


Figure S3: Species matrix for `betasim()` with default parameters (3 sites/5 abundance classes/2 spp per AC) and 50% mixing for all abundance classes.

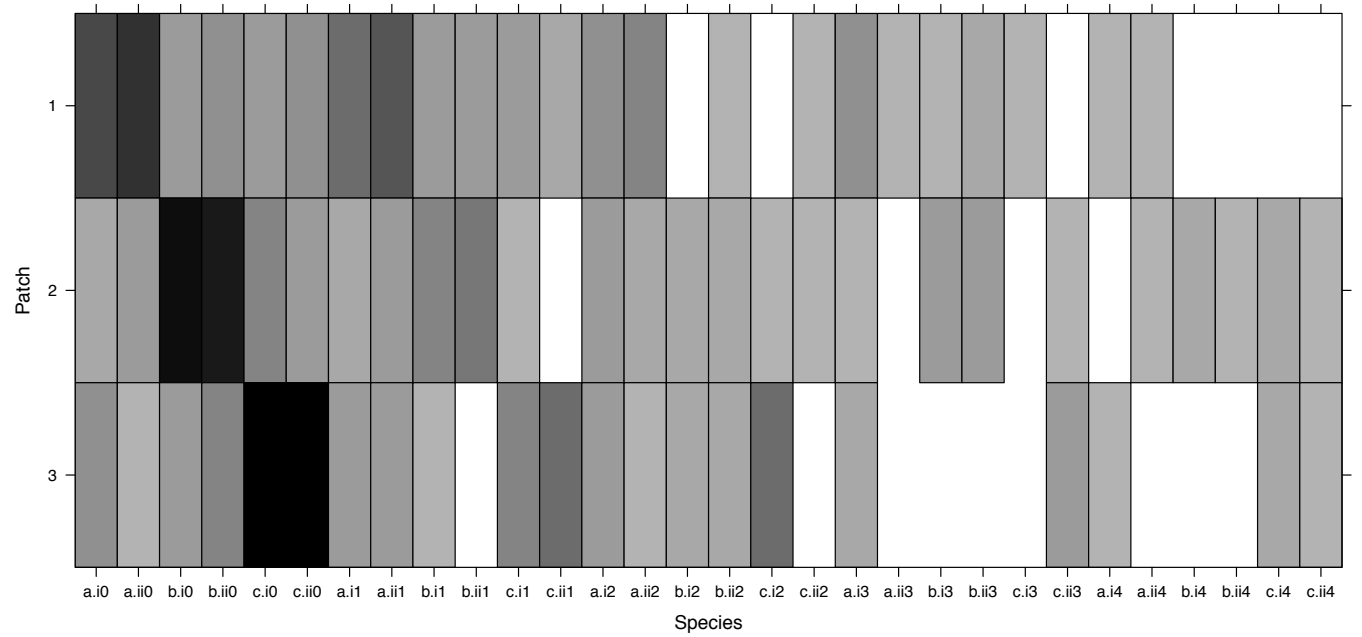


Figure S4: Species matrix for `betasim()` with default parameters (3 sites/5 abundance classes/2 spp per AC) with 50% mixing for all abundance classes and incorporating Poisson variation in the output.

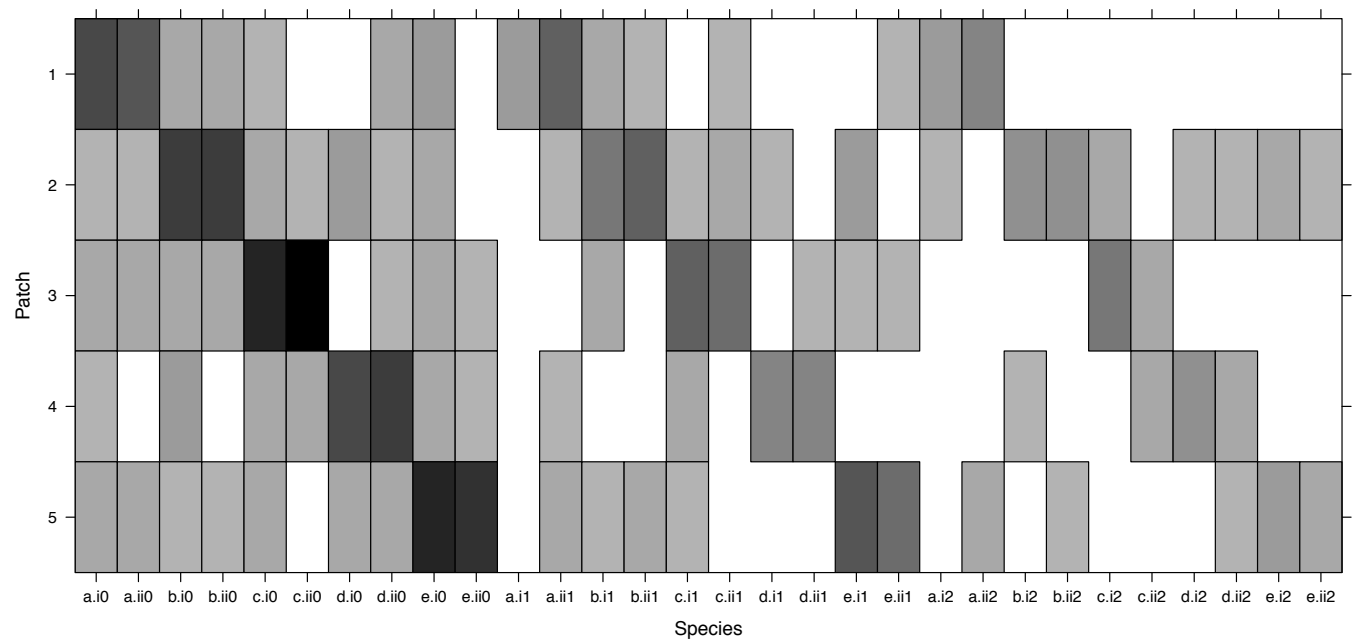


Figure S5: Species matrix for `betasim()` with 5 sites, 3 abundance classes, 2 spp per AC, 25% mixing for all abundance classes, and Poisson variation.

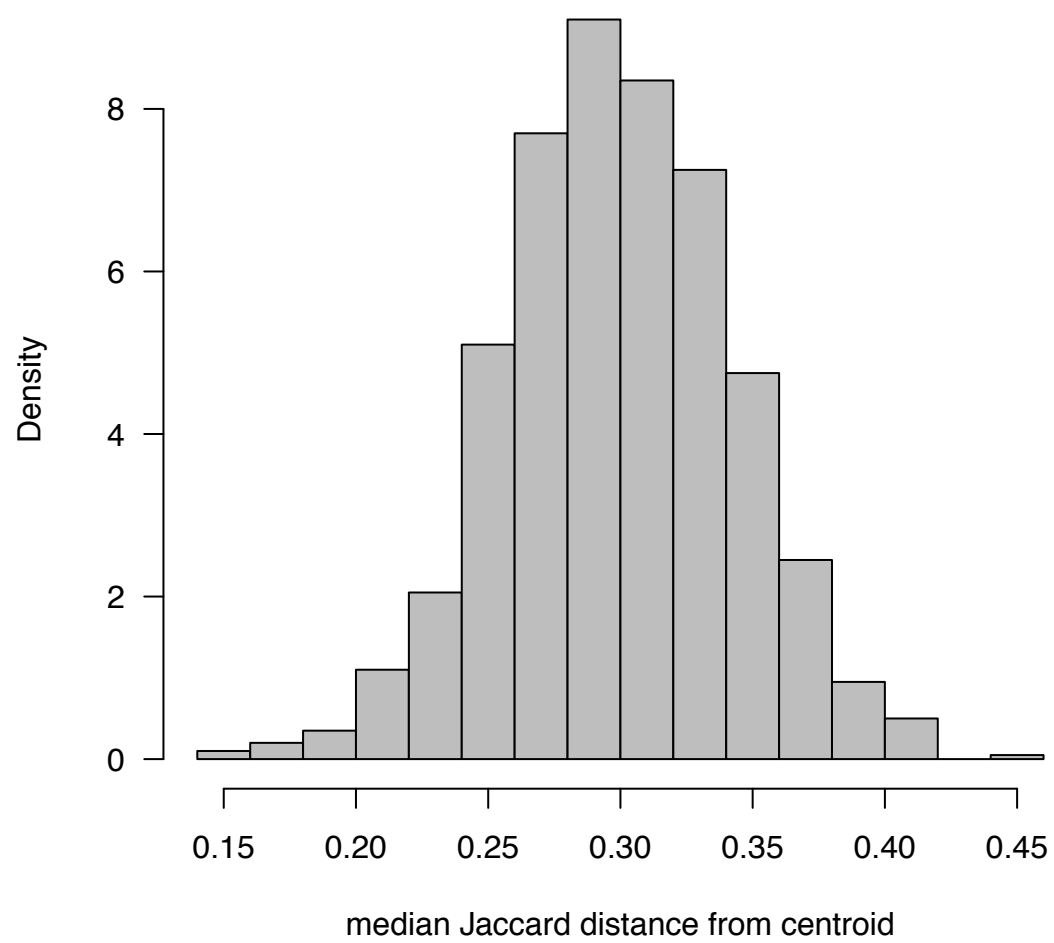


Figure S6: Histogram of `calcbeta` output for 10000 simulated communities (median Jaccard distance to centroid).

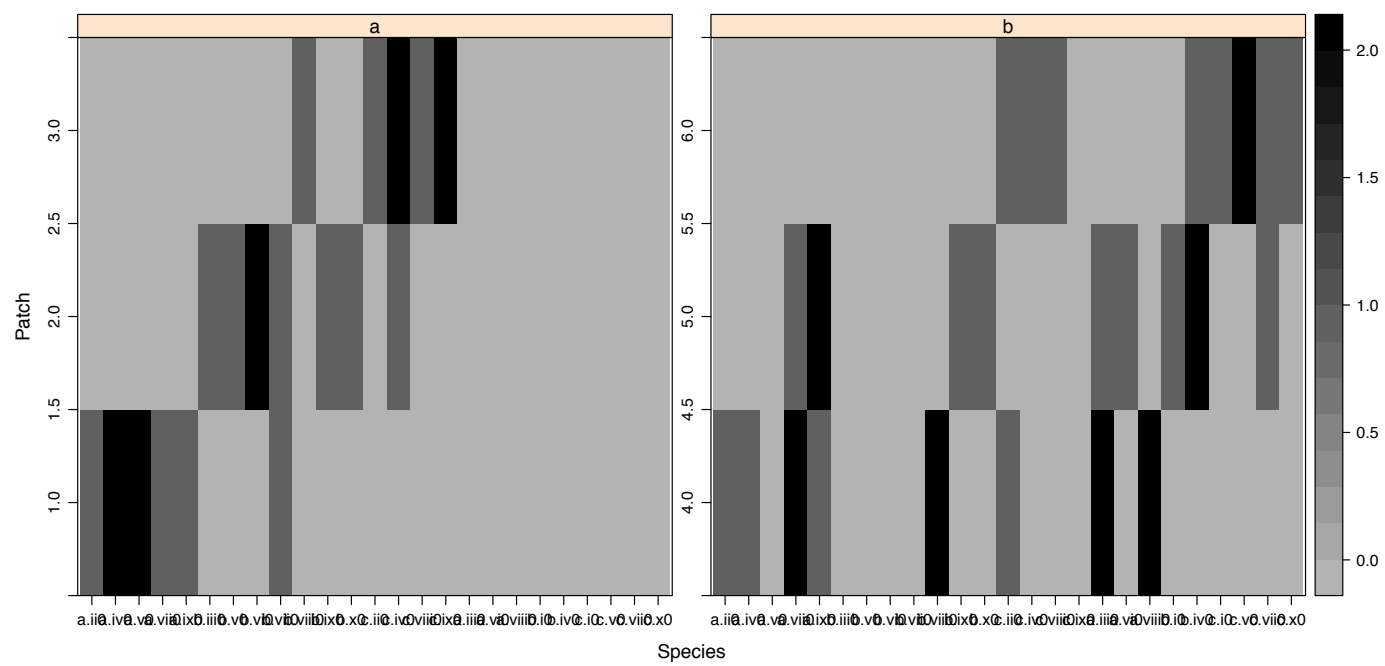


Figure S7: Output of `simComm`: two communities simulated simultaneously, with 30 species at each site and mixing proportions of 0.2 and 0.7 respectively.

## Appendix S3 Alternative diversity indices

In the main text we focus on the Jaccard index based on binary (presence-absence) data. The literature is full of alternative diversity measures; while our analysis focuses on the effect of sampling rather and hopes to be generally applicable across different indices, we should at least consider what would happen if we were to use an index based on continuous (count) data instead. Figure S1 shows results for the Manhattan index (we use either Hellinger or square root transformations as recommended by Legendre and De C  ceres (2013); however, Legendre and De C  ceres 2013 have also criticized the Manhattan index, saying that in particular it does not properly handle situations where multiple sites are missing the same pairs of species).

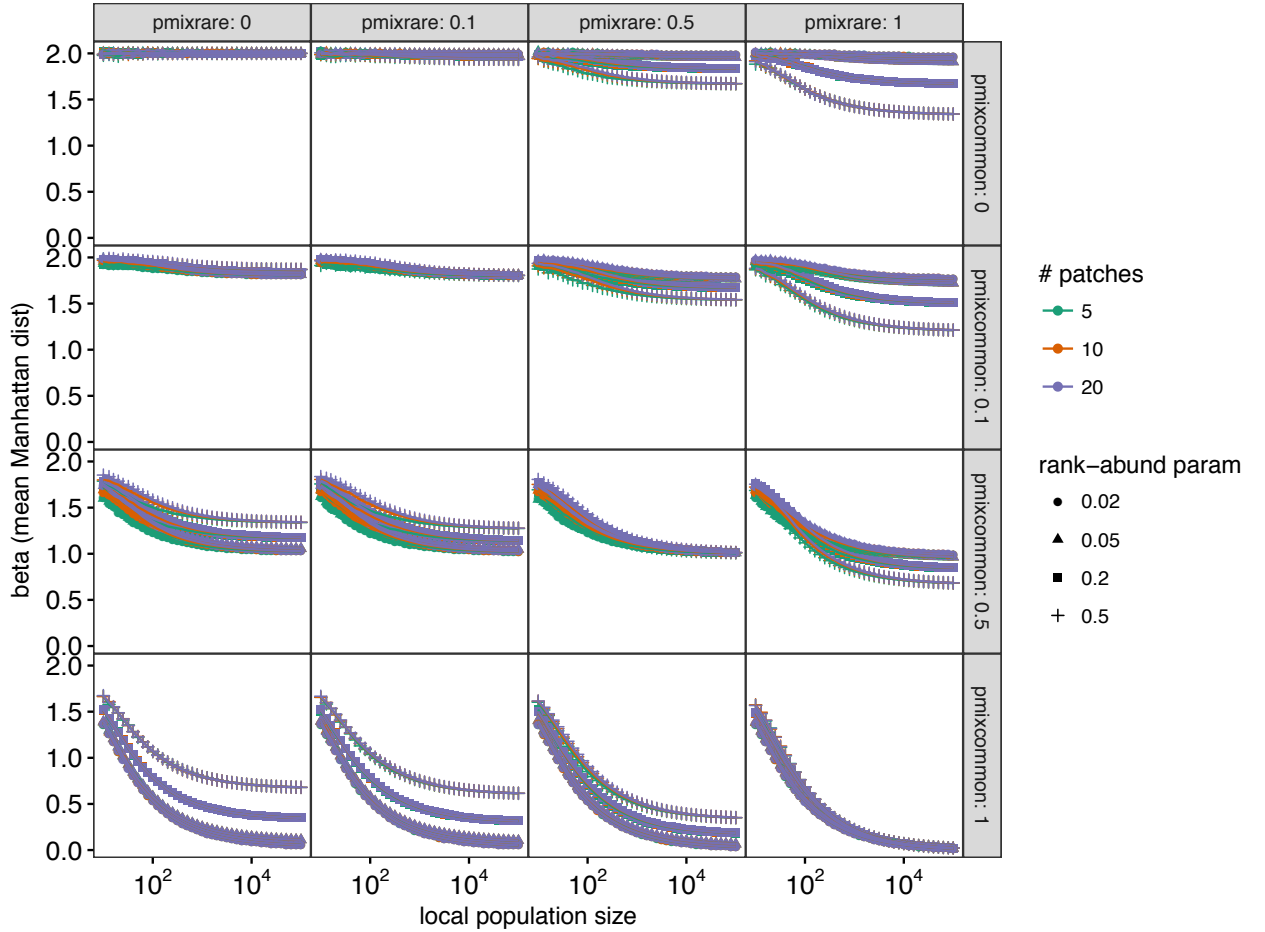


Figure S1: Diversity results for Manhattan index

This analysis extends the general principle that we can understand the effects of sampling on beta diversity by considering the expected effect of sampling intensity on variance. In this case, the responses are counts rather than binomial (binary) responses, so the expected variance of the response, and thus the estimated beta diversity, is proportional to  $1/\sqrt{N}$ . In contrast to the binary Jaccard results, there is no hint of non-monotonicity on the beta diversity curves. The number of sites has negligible effects on the computed beta diversity: the rank-abundance curve has moderate effects, although they disappear in extreme (low- or high-mixing) cases.

Chase et al. 2011 have recommended the Raup-Crick dissimilarity index, which characterizes

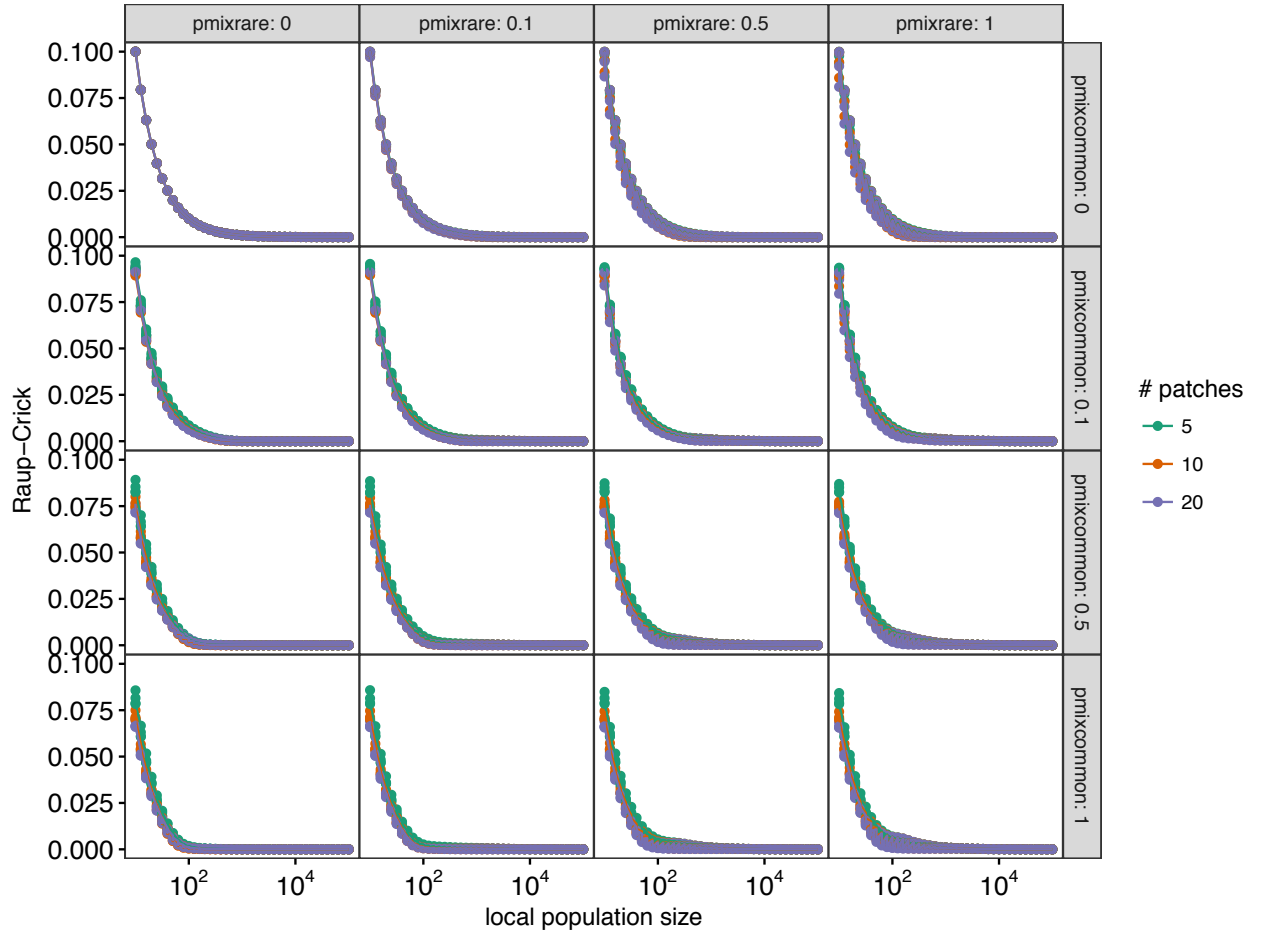


Figure S2: Diversity results for Raup-Crick index

diversity as the probability (based on permutation tests) that two communities have the same species composition. Raup-Crick indices are robust to many different aspects of community structure, and indeed when we run the same set of analyses as Figure 3 in the main text, we see that this index is robust to variation in all axes of ecological variation (number of sites, rank-abundance curve, mixing probabilities) *except* for sampling intensity/local population size (Figure S2). In other words, the figure shows the same general decrease in beta diversity with increasing sample size as other density-based metrics such as the Manhattan index, but the curves are identical across all other axes of variation.

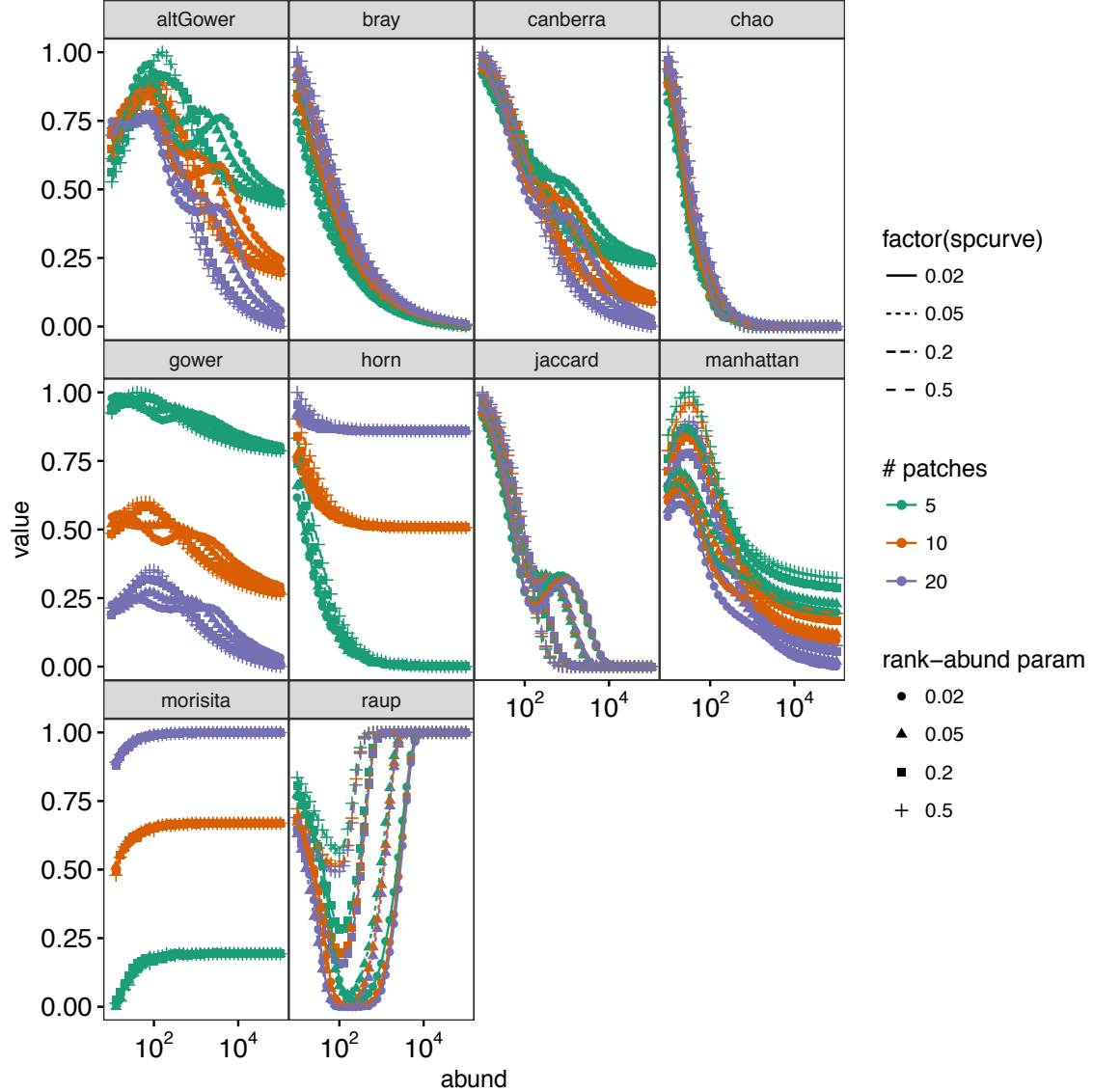


Figure S3: Beta diversity curves (scaled 0-1 for each diversity metric) as a function of sample size for a range of distribution for  $\text{pmixrare}=\text{pmixcommon}=0.5$ .

## References

- Chase, Jonathan M., Nathan J. B. Kraft, Kevin G. Smith, Mark Vellend, and Brian D Inouye. 2011. “Using null models to disentangle variation in community dissimilarity from variation in  $\alpha$ -diversity.” Art24, *Ecosphere* 2 (2): 1–11. ISSN: 2150-8925. doi:10.1890/ES10-00117.1. <http://dx.doi.org/10.1890/ES10-00117.1>.
- Legendre, P., and M. De C  ceres. 2013. “Beta diversity as the variance of community data: dissimilarity coefficients and partitioning.” *Ecology Letters* 16:951–963.

## Appendix S4 Power analyses

We ran power analyses of hierarchical rarefaction analysis, varying the true difference in beta diversity, the average population/sample size per patch, the difference in population/sample size per patch, and the difference in the number of patches. We kept the average number of patches equal to 20, and the number of species per community equal to 20.

The results show that for a per-patch population size of 20, the test is actually slightly conservative (i.e., the probability of rejecting the null hypothesis with  $\alpha = 0.05$  when it is true ( $\text{pmixdiff}=0$ ) is actually slightly less than 0.05; for a per-patch population size of 10, the results are slightly anti-conservative. The power curves (i.e., power to detect a specified level of difference in beta diversity) change little as a function of per-patch population size or population size difference, and declines slightly with increasing asymmetry in the number of patches per site, probably driven by low power in the community with the small number of patches.

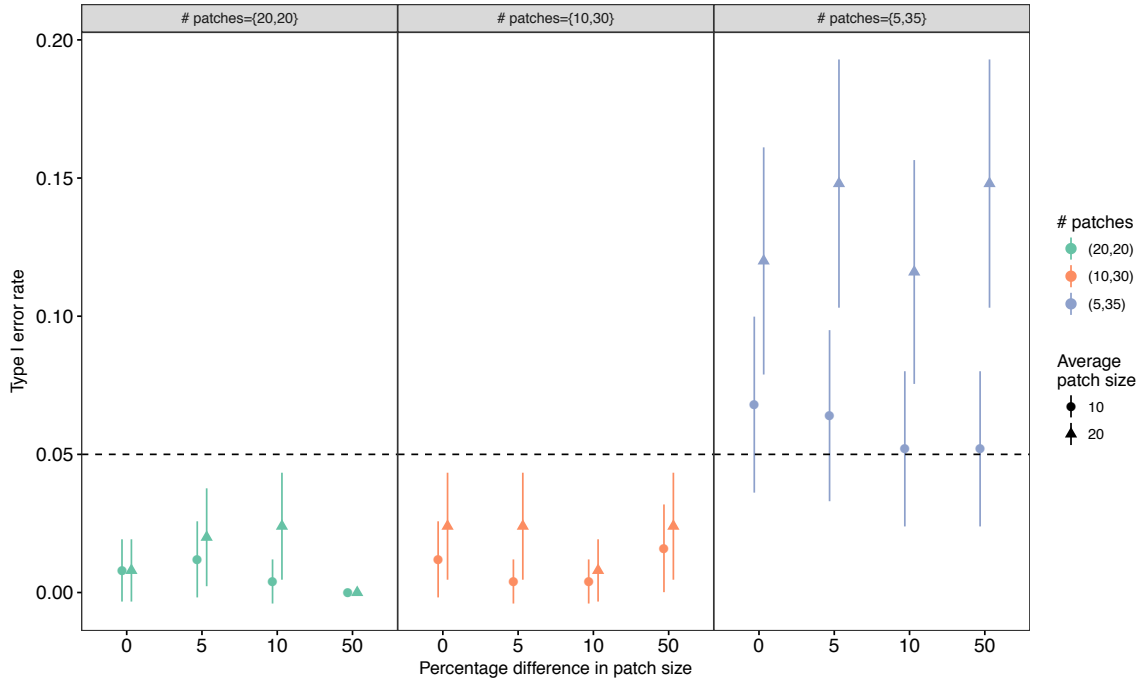


Figure S1: Type I error rates (rejection rates when  $\text{pmixdiff}=0$ ).

Mean and standard error of type I error rate (true difference in asymptotic beta diversity/mixing parameter=0), by differences in numbers of sites, *not* including patch size=20 case for 30-site {5,35} case:

sizediff	a.mean	a.se
0	0.009	0.003
20	0.015	0.003
30	0.059	0.004

For patch size=20, number of sites={5,35}:

a.mean	a.se
0.133	0.009

## Appendix S5 Sample-based rarefaction

As we mention in the paper, our approach to multi-level rarefaction is based on individual-level rarefaction, in part because of the commonness of community data with only a single sample per patch (in which case sample-based rarefaction is useless) and in part because of the complexity of bookkeeping in the multi-sample/sample-based rarefaction case.

If we stop worrying about the constraints of rarefying each patch to the minimum number of individuals found in any patch, in any treatment (which is what necessitates the complex matching strategy implemented in the package), we can come up with a reasonably simple strategy for rarefaction.

- pick the per-patch size  $N$  to rarefy to
- for each patch in each community,
  - while the accumulated patch size is  $< N$ , pick a sample from that patch without replacement
  - if the next-chosen patch makes the patch size  $> N$ , either
    - \* take only the number of individuals from the sample that brings the patch up to the target, *or*
    - \* add the sample to the patch community, then remove enough individuals from the community at random to bring it back down to the target size

We have not implemented this functionality yet.