

Robustness of trait distribution metrics for community assembly studies under the uncertainties of assembly processes

MASAHIRO AIBA,^{1,5} MASATOSHI KATABUCHI,² HINO TAKAFUMI,^{3,6} SHIN-ICHIRO S. MATSUZAKI,⁴ TAKEHIRO SASAKI,^{1,7}
AND TSUTOM HIURA³

¹*Graduate School of Life Sciences, Tohoku University, 6-3 Aoba, Aramaki, Aoba-ku, Sendai 980 8578 Japan*

²*Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Mengla, Yunnan 666303 China*

³*Tomakomai Research Station, Field Science Center for Northern Biosphere, Hokkaido University, Takaoka, Tomakomai 053 0035 Japan*

⁴*Center for Environmental Biology and Ecosystem Studies, National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba 305 8506 Japan*

Abstract. Numerous studies have revealed the existence of nonrandom trait distribution patterns as a sign of environmental filtering and/or biotic interactions in a community assembly process. A number of metrics with various algorithms have been used to detect these patterns without any clear guidelines. Although some studies have compared their statistical powers, the differences in performance among the metrics under the conditions close to actual studies are not clear.

Therefore, the performances of five metrics of convergence and 16 metrics of divergence under alternative conditions were comparatively analyzed using a suite of simulated communities. We focused particularly on the robustness of the performances to conditions that are often uncertain and uncontrollable in actual studies; e.g., atypical trait distribution patterns stemming from the operation of multiple assembly mechanisms, a scaling of trait–function relationships, and a sufficiency of analyzed traits.

Most tested metrics, for either convergence or divergence, had sufficient statistical power to distinguish nonrandom trait distribution patterns without uncertainty. However, the performances of the metrics were considerably influenced by both atypical trait distribution patterns and other uncertainties. Influences from these uncertainties varied among the metrics of different algorithms and their performances were often complementary.

Therefore, under the uncertainties of an assembly process, the selection of appropriate metrics and the combined use of complementary metrics are critically important to reliably distinguish nonrandom patterns in a trait distribution. We provide a tentative list of recommended metrics for future studies.

Key words: assembly rules; biotic interactions; community assembly; competition; environmental filtering; functional diversity; functional traits; habitat filtering; limiting similarity.

INTRODUCTION

Trait-based analyses of ecological communities have enhanced our understanding of community assembly processes (HilleRisLambers et al. 2011, Weiher et al. 2011). Trait distribution patterns within a local community can be either convergent or divergent, as a result of the two major mechanisms of community assembly, environmental filtering and biotic interactions (Weiher and Keddy 1995, Webb et al. 2002). Environmental

filtering refers to an ecological process where the persistence of species in a given abiotic environment is determined by functional traits (Keddy 1992, Mayfield and Levine 2010). This process should produce a convergent trait distribution as long as responses to the environment are a monotonic or unimodal function of the traits. Alternatively, biotic interactions often lead to a divergent trait distribution via competitive interaction, niche differentiation, facilitation, or enemy-mediated interaction (MacArthur and Levins 1967, Kursar et al. 2009, Spasojevic and Suding 2012; but see Cavender-Bares et al. 2009, Mayfield and Levine 2010). The assumption that convergence or divergence in a trait distribution pattern can be a sign of the contrasting assembly rules has led to an increasing number of studies seeking nonrandom trait distribution patterns for various kinds of communities (e.g., Arita 1997, Weiher et al. 1998, Mason et al. 2007, Kraft et al. 2008). Either convergent or divergent trait distributions have

Manuscript received 12 February 2013; revised 10 June 2013; accepted 13 June 2013. Corresponding Editor: N. J. Gotelli.

⁵ E-mail: mshiro5@gmail.com

⁶ Present address: Department of Environmental and Symbiotic Sciences, Rakuno Gakuen University, Bunkyo-daimidorimachi 583, Ebetsu, Hokkaido 069 8501 Japan.

⁷ Present address: Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277 8653 Japan.

been detected in many studies, suggesting that both environmental filtering and biotic interactions play an important role in community assembly (Götzenberger et al. 2011, HilleRisLambers et al. 2011, Weiher et al. 2011).

However, the metrics used for detection of these patterns vary among studies. Despite the challenges of some pioneer studies (Kraft and Ackerly 2010, Mouchet et al. 2010, Mason et al. 2012), there is no consensus regarding the choice of metrics to maximize the possibility of detecting a nonrandom trait distribution pattern. A primary problem when establishing trustworthy guidelines for the choice of metrics are the uncertainties in the assembly processes. Simulation studies have demonstrated that many metrics have sufficiently high statistical power (i.e., low risks of type II error) as long as the trait distribution pattern is typical (i.e., generated by a single assembly mechanism) and every detail of the assembly process is apparent (Kraft and Ackerly 2010, Mouchet et al. 2010, Mason et al. 2012). However, in actual studies, the essential conditions of the assembly process that potentially influence the performance of the metrics are often unknown.

Among these uncertainties, the influence of assembly scenarios has been a focus of earlier studies (Kraft and Ackerly 2010, Mason et al. 2012). These simulation studies have shown that the simultaneous operation of multiple processes blurs both convergent and divergent trait distribution patterns and thus makes it more difficult to detect significant patterns. Combinations of the three processes of community assembly (i.e., stochasticity, environmental filtering, and biotic interactions) generate various atypical patterns of trait distribution. Based on the three processes, previous studies often subtractively assembled simulated communities (Kraft and Ackerly 2010, Mouchet et al. 2010). Species were sequentially removed from a species pool to arrive at a specified richness based on one of the three processes. However, it is also natural to build simulated communities by adding species in a vacant community to a specified richness (Mason et al. 2012). Trait distribution patterns generated from additive assembly simulation may be more challenging, especially for the metrics of convergence, because species addition based on different criteria would often introduce outliers in the pattern. The robustness of the performance in these cases is practically important because assembly processes such as mass effects (Shmida and Wilson 1985) seem to introduce such outliers (cf. Mason et al. 2012). In most cases, the true assembly scenario operating in a target community is unknown. Therefore, ideally metrics that are effective in practical studies should be robust to all of these alternative assembly scenarios.

Factors other than assembly scenarios can also influence the statistical powers of the metrics but have rarely been considered in earlier studies. Such factors include the potential mismatch of the scaling of trait–function relationships between a true assembly process

and a calculated assumption. It is often practically difficult to know whether a relationship between a trait value and the strength of environmental filtering or biotic interactions is linear or log-linear. This uncertainty leads to a potential mismatch in the scaling of traits because researchers often decide to (or not to) log-transform trait values before analysis without firm evidence. For multi-trait metrics, the choice of traits for an analysis leads to another mismatch. Usually, it is virtually impossible to know the exact set of traits that influence community assembly (cf. Petchey and Gaston 2006). Therefore, in most cases, the analyzed suite of traits will inevitably omit some important traits and/or include some unimportant traits. In addition to these two problems, previous studies have shown that local species richness often affects the statistical power of metrics, and that the direction and extent of the influence differs among metrics (Kraft and Ackerly 2010, Mouchet et al. 2010). Thus, robustness to these problems should be the second requisite of ideal metrics.

In this study, we aimed to construct practical guidelines for choosing robust metrics under multiple uncertainties of the assembly processes based on simulated communities. For this purpose, we comparatively examined the performance of five metrics of convergence and 16 metrics of divergence under two types of uncertainty that have concerned researchers in actual studies: (1) alternative assembly scenarios in which multiple mechanisms were mixed in various ways and (2) mismatches between a true assembly process and an assumption for analysis in the scaling of trait–function relationships and determining a set of responsible traits.

MATERIALS AND METHODS

Tested metrics of trait distributions

Metrics of a trait distribution pattern can be categorized into four groups based on the trait distribution pattern that they detect: single-trait metrics for convergence, multi-trait metrics for convergence, single-trait metrics for divergence, and multi-trait metrics for divergence. In this study, two single-trait metrics of convergence (range and variance), three multi-trait metrics of convergence (functional richness [FRic], functional dispersion [FDis], and functional diversity based on a dendrogram [FD]), eight single-trait metrics of divergence (mean nearest-neighbor distance [MNND], minimum nearest-neighbor distance [MinNND], standard deviation of nearest-neighbor distance [SDND], standard deviation of nearest-neighbor distance [SDNN], SDND divided by the range [SDNDR], SDNN divided by the range [SDNNr], variance, and kurtosis) and eight multi-trait metrics of divergence (MNND, MinNND, ratio of the shortest distance to the longest distance in a minimum spanning tree [Min/Max MST], FRic, FDis, FD, functional evenness [FEve], and functional divergence [FDiv]) were considered as all have been used frequently in past studies (Appendix A).

Both of the two single-trait metrics of convergence are well-known basic statistics. Some studies have examined multi-trait convergence by using some multidimensional metrics of functional diversity; e.g., FRic, FDis, and FD (Mouchet et al. 2010, Mason et al. 2012). FRic is a convex hull volume in the multidimensional space of traits that are occupied by species in a community (Villegger et al. 2008). FDis is defined as the mean distance of a species from the centroid of trait values in a multidimensional space (Laliberte and Legendre 2010). FD is the sum of the branch length of a dendrogram that is constructed by hierarchical classification of species based on traits (Petchey and Gaston 2002). Of the eight single-trait metrics of divergence, six (MNND, MinNND, SDND, SDNN, SDNDR, SDNNr) are based on neighbor distance or nearest-neighbor distance between species that are sorted along the trait values. While MNND and MinNND are defined as the mean and minimum nearest-neighbor distances, respectively (Findley 1976, Ricklefs and Travis 1980), SDND and SDNN are defined as the standard deviation of neighbor distances and nearest-neighbor distances, respectively (Ricklefs and Travis 1980, Cornwell and Ackerly 2009). SDNDR and SDNNr are SDND and SDNN standardized by the range, respectively, to decrease the risk of a type I error occurring when environmental filtering operates at the same time as biotic interactions (Kraft and Ackerly 2010). While these metrics focus on the distances between species, variance and kurtosis focus on a whole distribution form, with the assumption that biotic interactions lead to more variable and/or platykurtic trait distributions, respectively. MNND and MinNND, which we have introduced as single-trait metrics, have also been used as multi-trait metrics. Min/Max MST is also a metric that focuses on the nearest distance between species (Stubbs and Wilson 2004). Additionally, measures of functional diversity; e.g., FRic, FEve, FDiv, FDis and FD, have also been used as multidimensional metrics of trait divergence (Mouchet et al. 2010, Mason et al. 2012). FEve is the evenness of branches of a minimum spanning tree in multidimensional trait space and FDiv is the species deviance from the mean distance to the centroid of the space (Villegger et al. 2008).

Community and trait data

Although we assessed the performance of the metrics based on simulated communities, the size of species pools, trait values in the species pools, and local species richness for a simulation were obtained from data for three actual communities that possessed different characteristics. One was the understory herb community of a cool temperate forest in northern Japan (HERB). All understory plants, which were found in 60 1-m² quadrats located in the 2715-ha Tomakomai Experimental Forest, were recorded. The species pool size was 75, median local species richness was 11, and minimum and maximum richness values were 5 and 21, respec-

tively. Four continuous traits were measured; i.e., plant height, total aboveground dry mass, specific leaf area (SLA), and leaf nitrogen content. For further details of these data, see Aiba et al. (2012). The second was a hyperdiverse tropical rain forest tree community on a 52-ha plot located in the Lambir Hills National Park, Sarawak, Malaysia (TREE). The data contains more than 350 000 mapped trees (>1 cm diameter at breast height [dbh]) belonging to 1195 species. We focused on only 549 species whose traits were available. The entire plot was divided into 1300 quadrats (20 × 20 m), each of which was considered to represent a local community. The median local species richness was 100 and minimum and maximum richness were 61 and 131, respectively. Four continuous traits; i.e., SLA, leaf size, leaf toughness, and leaf tannin content, were measured for one to eight saplings per species. The final community comprised strictly freshwater fish in Japan (FISH). The four main islands of Japan were divided into 27 geographic regions (mean area was 13 900 km²) and each region was considered to represent a local community (Watanabe 2012). The species pool size was 90, median, minimum, and maximum local species richness were 36, 5, and 59, respectively. The functional traits were one continuous 12 binary, 9 categorical, and 3 ordered. The total of 25 functional traits consisted of 3 morphological, 13 dietary, and 5 reproductive traits and 4 habitat preferences. These traits were obtained from the literature and an electronic database, FishBase. Further details of this data set are available in Matsuzaki et al. (2013). Note that species composition in the local communities from these three data sets was never used in the following analyses.

Simulation of community assembly

Ten alternative assembly scenarios were considered for each of convergence and divergence. These scenarios were built by combining six simple steps, i.e., the random removal of a species, environment-based removal, trait similarity-based removal, random addition of a species, environment-based addition, and similarity-based addition. The three removal steps have also been used in former studies (Kraft and Ackerly 2010, Mouchet et al. 2010). For the random removal step, a species to be removed was randomly selected irrespective of its trait values. For the environment-based removal, an optimum trait value was randomly determined within the observed trait range and then a species whose trait value was most distant from the optimum was removed. For the similarity-based removal, trait dissimilarities of all species pairs were calculated and then one of the most similar pairs was randomly removed. The manner of species selection was generally mirrored in multi-trait simulations but Euclidean (for HERB and TREE data) or Gower (for FISH data, which included binary, categorical, and ordered trait variables) distances were used as measures of distance.

The subtractive assembly scenarios, which were constructed from the three removal steps, assume that community assembly is a process in which species are removed from the species pool by local extinction due to stochasticity, environmental filtering, and/or biotic interactions. However, it is also reasonable to consider community assembly as a process where species are sequentially selected from the species pool to fill a vacant local community (Mason et al. 2012). Thus, we designed the three addition steps in a similar manner to the removal steps. Random addition is a step in which one species is randomly chosen to be added into the local community, irrespective of the trait values. For the environment-based addition, species whose trait value(s) were closest to the randomly determined optimum were chosen. For the similarity-based addition, trait dissimilarity to the most similar species in the preexisting community was calculated for all species in the remaining species pool, and then the most distant species were chosen. If the local community was vacant (i.e., the first step in a simulation), the first species was randomly chosen from the species pool. In additive assembly scenarios, which were constructed from the three addition steps, stochasticity, environmental filtering, and/or biotic interactions function as a determinant of the recruitment and establishment of species. Although these processes do not necessarily prevent the recruitment of a species, they can affect a trait distribution pattern by excluding some species shortly after recruitment. In the simulation process, we handled recruit limitation and exclusion shortly after recruitment as an identical process because the difference is not critical to the simulation results. In typical cases; i.e., where only one of the three assembly rules was used, species addition produces communities that are quite similar to those to which species removal has been applied. However, in atypical scenarios in which multiple rules were combined, species removal and species addition often produce quite different patterns.

The 10 assembly scenarios that were designed for the metrics of convergence were as follows: E–, E+, ER–, ER+, EB–, EB+, EBR–, EBR+, RREE+, and EERR+. In these scenarios, E, B, and R represent environmental filtering, biotic interactions, and randomness, respectively. Plus and minus signs indicate subtractive and additive simulation, respectively. E– and E+ are the typical scenarios of trait convergence, which were constructed only from environment-based removal and addition, respectively. ER– and ER+ is a combined model in which the environment-based step and the random step were randomly selected with an equal probability for each step of a simulation. Similarly, in EB– and EB+, the environment- and similarity-based steps were randomly selected with an equal probability. In EBR– and EBR+, the random, environment- and similarity-based steps were randomly selected with an equal probability and therefore the contribution of environmental filtering decreased to one-third of the

total removal/addition of species. In RREE+, the first half of the total addition was random and the second half was environment-based and in EERR+, the first half of the total addition was environment-based and the second half was random. Similarly, we considered the following 10 scenarios for trait divergence: B–, B+, BR–, BR+, EB–, EB+, EBR–, EBR+, RRBB+, and BBRR+, where the environment-based steps of the 10 scenarios for trait convergence were replaced by similarity-based steps. These scenarios were based at least partly on existing theories of community ecology. For example, ER+ may mirror community assembly under a mass effect (Shmida and Wilson 1985) where a pattern created by environmental filtering is blurred by random recruitments. RREE+, EERR+, RRBB+, and BBRR+ were designed with temporal shifts of an assembly mechanism through a successional process in a community (Weiher and Keddy 1999).

In the analysis of alternative scenarios, other miscellaneous conditions; i.e., the scaling of trait–function relationships, the sufficiency of analyzed traits, and local species richness were fixed as follows (control [LL] condition in Appendix B). Trait values were log-transformed before both the assembly simulation and the calculation of metrics. In multi-trait cases, all traits were used for both the simulation and calculation of metrics. For local species richness, the median diversity of each data set was used. In the second step of our simulation test, the assembly model was fixed to E– for metrics of convergence or B– for metrics of divergence and the conditions were then changed as follows (Appendix B). Trait values were optionally untransformed before either the assembly simulation or the calculation of metrics. This procedure led to three alternative situations: traits were untransformed before the simulation but were log-transformed before the calculation of metrics (NL), traits were log-transformed before the simulation but were untransformed for the calculation of metrics (LN), and traits were untransformed before both the simulation and the calculation of metrics (NN). For multidimensional analysis, the sufficiency of the analyzed trait number was modified. An “excess” model was created by randomly selecting half of all traits for community assembly while the metrics were calculated by using all traits. In contrast, a “deficit” model was created by randomly selecting half of all the traits for the calculation of metrics while all of the traits were used for community assembly. Local species richness was changed to the minimum (min) or maximum (max) value of each data set. Unfocused conditions were set as in the control scenario. See the R code in our supplementary materials for details of both the calculation of metrics and the simulation algorithm (Supplement).

Statistical tests

One hundred local communities were generated for each combination of data set (HERB, FISH, and

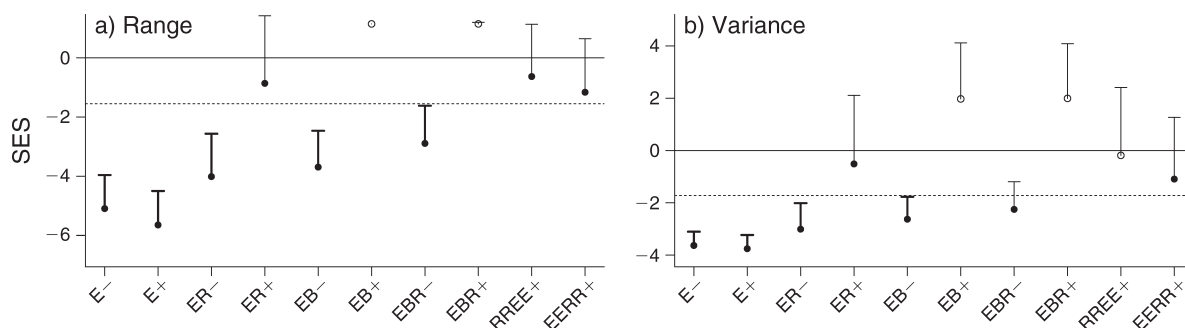


FIG. 1. The performance of the two single-trait metrics, (a) range and (b) variance, for convergence against 10 alternative assembly scenarios for plant height in the HERB data set. Symbols and error bars indicate means and one-sided 95% confidence interval (CI) of standardized effect size (SES) values, respectively. Dashed lines indicate one-sided 95% CI of SES values of the null model. If a metric had sufficient statistical power at the single local-community level, i.e., 95% CI of the SES values of metrics did not overlap with 95% CI of the null model, the results are represented by heavy error bars. Statistical powers at the metacommunity level (100 local communities), i.e., a significant shift from 0 in one-sided Wilcoxon tests, are indicated by solid symbols. In names of the tests, E, B, and R represent environmental filtering, biotic interactions, and randomness, respectively. Plus and minus signs indicate subtractive and additive simulation, respectively. See Appendix D for the full results.

TREE) $\times$ functional traits (e.g., SLA for HERB, body length for FISH and multiple traits) $\times$ simulation scenarios (alternative scenarios and other conditions). In total, 32 400 communities for 324 different situations were generated. Values of the metrics were calculated for each of these simulated communities. These values were then standardized by the mean and standard deviation of the values for the null model; i.e., 100 simulated communities that were only produced from random removal (R–), to obtain the standardized effect sizes (SESs; Gotelli and McCabe 2002). This standardization enabled comparison of the performance among different metrics and simulation scenarios. The statistical power of the metrics was evaluated in two ways based on Kraft et al. (2007) and Kraft and Ackerly (2010). First, if the one-sided 95% confidence interval (CI) of SESs did not overlap with the one-sided 95% CI of the SESs of the null model, a nonrandom pattern could be detected for 95% of the local communities following a permutation test (quadrat-level power, sensu Kraft and Ackerly

[2010]). Second, the possibility of detecting a nonrandom pattern at the metacommunity level (i.e., 100 local communities) was determined using a one-sided Wilcoxon signed rank test, which examines the shift of SES values from 0 (plot level power sensu Kraft and Ackerly [2010]). The distributions of SESs for the 10 alternative scenarios are shown as examples for some traits (Figs. 1–4) but for the other results only the statistical powers have been summarized in table form. All statistical analyses were performed in the statistical environment R 2.15.0 (R Development Core Team 2012). We summarized the process of analysis in a schematic flow chart (Appendix C).

RESULTS

Responses to the alternative assembly scenarios

Single-trait metrics for convergence.—For the typical assembly scenarios (E– and E+), both the range and variance could be used to detect trait convergence, even in a single local community, for all traits of all data sets

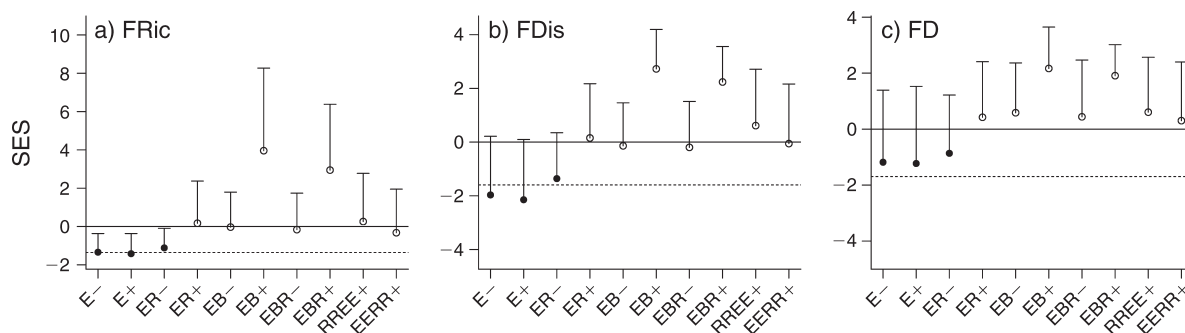


FIG. 2. The performance of the three multi-trait metrics, (a) functional richness [FRic], (b) functional dispersion [FDis], and (c) functional diversity based on a dendrogram [FD], for convergence against 10 alternative assembly scenarios for the HERB data set. Symbols and error bars indicate means and one-sided 95% confidence interval (CI) of standardized effect size (SES) values, respectively. Figure components are as in Fig. 1. The CIs of SESs were redundant with the CI of the null model in all cases, indicating that the metrics did not have sufficient statistical power at a single local-community level. See Appendix E for the full results.

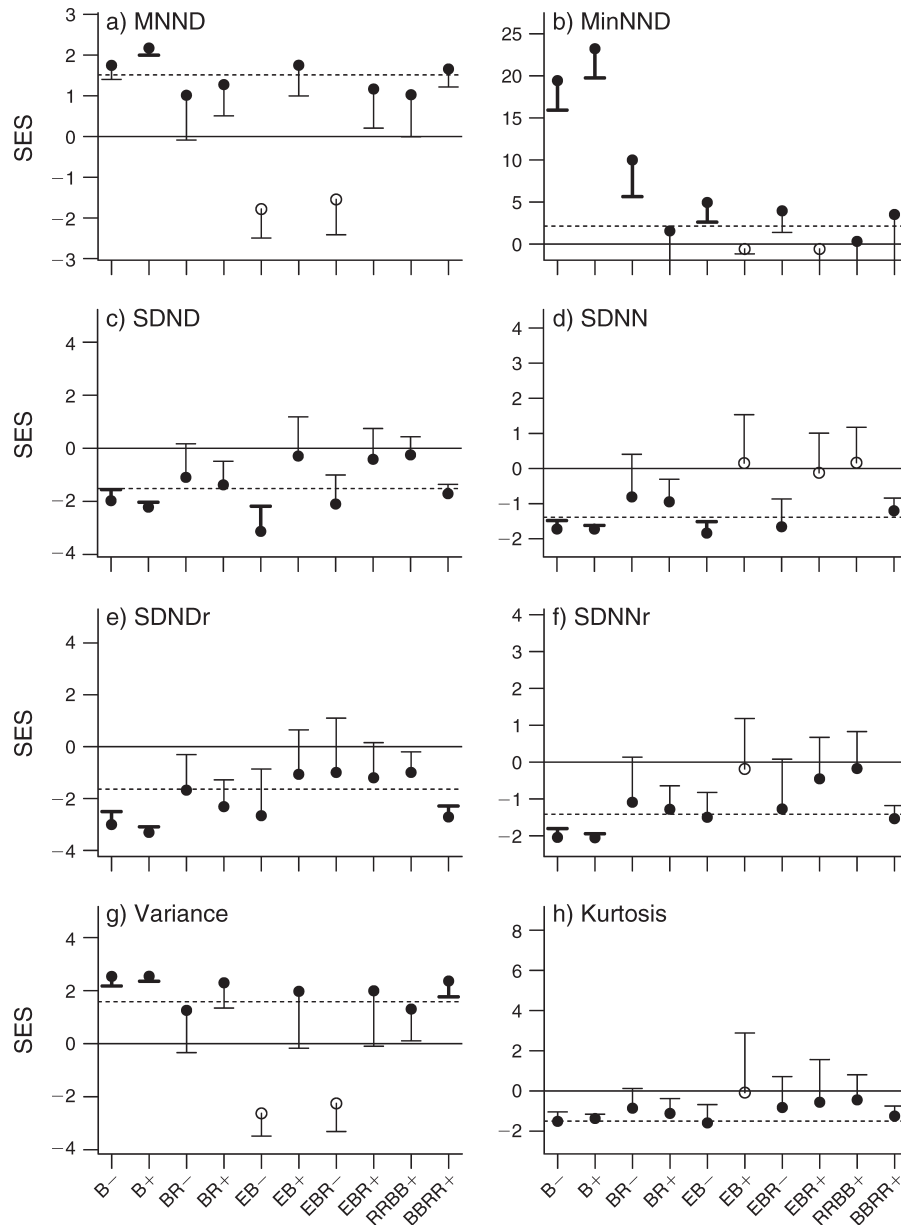


FIG. 3. The performance of the eight single-trait metrics: (a) mean nearest-neighbor distance [MNND], (b) minimum nearest-neighbor distance [MinNND], (c) standard deviation of neighbor distance [SDND], (d) standard deviation of nearest-neighbor distance [SDNN], (e) SDND divided by the range [SDNDR], (f) SDNN divided by the range [SDNNr], (g) variance, and (h) kurtosis). The metrics were tested for divergence against 10 alternative assembly scenarios for the HERB data set. Figure components are as in Fig. 1. See Appendix F for the full results.

(Fig. 1, Appendix D). The SESs of the range tended to be larger than those of the variance in the typical scenarios. For the atypical scenarios, in many cases, convergent trait distribution patterns that were generated through ER⁻, EB⁻, and EBR⁻ were significantly detected even for a single local community by both metrics. However, their statistical power was lessened (i.e., a risk of type II error was increased) in the ER⁺, RREE⁺, and EERR⁺ scenarios. For these scenarios, the range could be used to detect significant patterns more

often than the variance. Neither the range nor variance could be used to significantly detect trait convergence for EB⁺ and EBR⁺, even at the metacommunity level. The performance of both metrics was generally higher for TREE data when the same assembly scenario was compared. The performances of both metrics also varied among traits even in the same data set and were worst for SLA in HERB data.

Multi-trait metrics for convergence.—The performance of the three metrics; i.e., Fric, FDis, and FD,

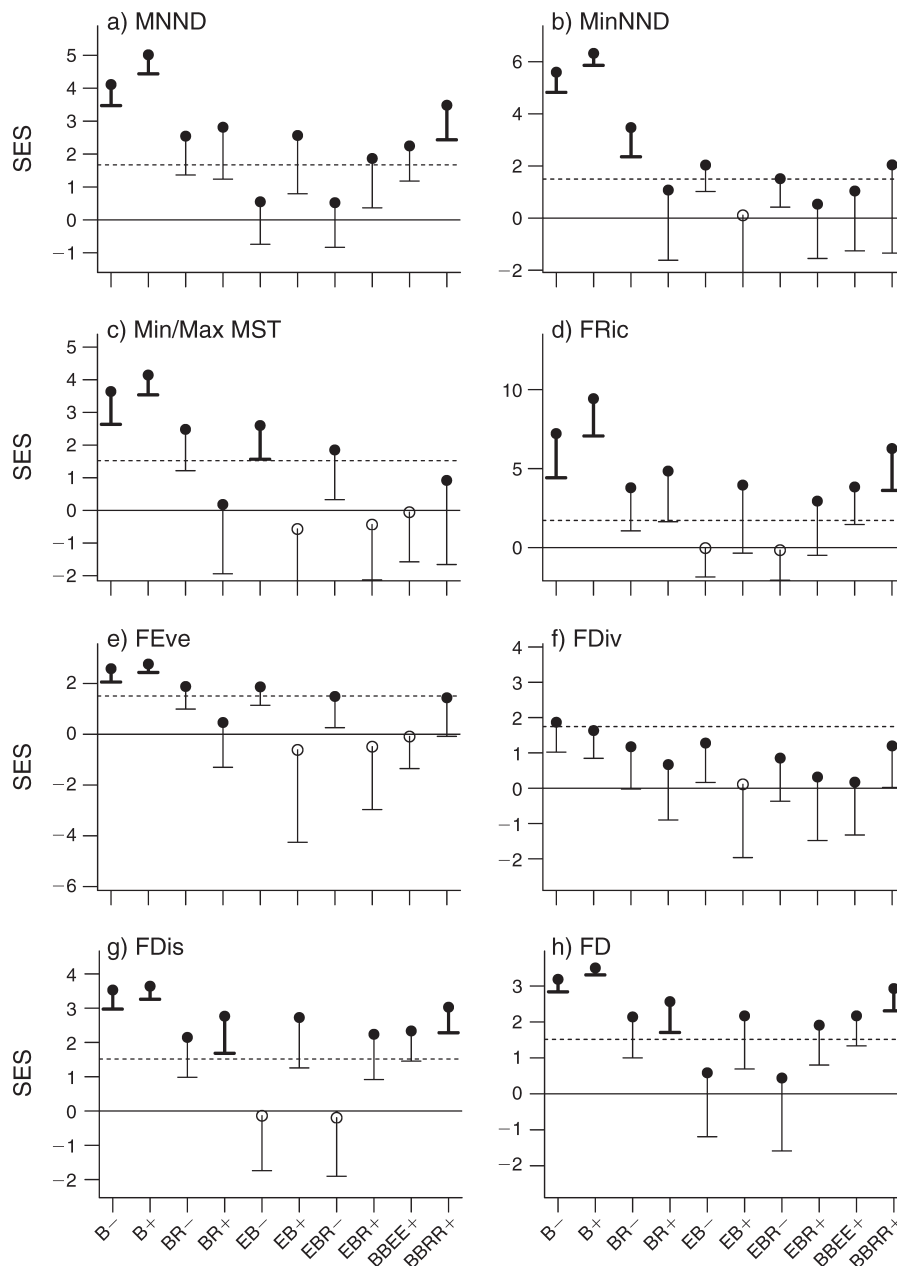


FIG. 4. The performance of the eight multi-trait metrics: (a) mean nearest-neighbor distance [MNND], (b) minimum nearest-neighbor distance [MinNND], (c) ratio of the shortest distance to the longest distance in a minimum spanning tree [Min/Max MST], (d) functional richness [FRic], (e) functional evenness [FEve], (f) functional divergence [FDiv], (g) functional dispersion [FDis], and (h) functional diversity based on a dendrogram [FD]. The metrics were tested for divergence against 10 alternative assembly scenarios for the HERB data set. Figure components are as in Fig. 1. See Appendix G for the full results.

was generally similar but FD slightly underperformed in the atypical scenarios for FISH data (Fig. 2, Appendix E). In the typical scenarios, they were able to detect convergence at least at the metacommunity level in all three data sets. All metrics could be used to detect convergence for ER- at least at the metacommunity level, whereas this was not possible for EB+ and EBR+. For the other scenarios, the performance of the metrics

depended on the data set and was lowest for HERB data.

Single-trait metrics for divergence.—The performance of the eight single-trait metrics for divergence often varied among traits even in the same data set (Fig. 3, Appendix F). SDND and SDNN had virtually no statistical power for SLA in the HERB data set. Kurtosis failed to detect trait divergence for leaf tannin

TABLE 1. The performance of the two single-trait metrics and the three multi-trait metrics for convergence against the problems of scaling relationships, the sufficiency of traits, and local species richness for the HERB data set.

Traits and metrics	Tests							
	Control	NN	LN	NL	Deficit	Excess	Min	Max
Height								
Range	†	×	×	†			†	†
Variance	†	†	†	†			†	†
Total mass								
Range	†	NS	×	×			†	†
Variance	†	†	†	×			†	†
SLA								
Range	×	×	×	×			×	×
Variance	×	×	×	×			×	×
Leaf N content								
Range	†	†	†	†			†	†
Variance	†	†	†	†			†	†
Multiple traits								
FRic	×	NS	×	×	×	×	×	×
FDis	×	NS	×	×	×	×	×	×
FD	×	NS	×	×	NS	×	×	×

Notes: Daggers (†) indicate significant results in both of the permutation tests at the single local community level and Wilcoxon signed rank tests at the meta-community level ($P < 0.05$). Multiplication symbols (×) represent significant results only in Wilcoxon signed rank tests at the meta-community level. NS represents nonsignificant results in both tests. In the NN test, traits were untransformed before both the simulation and the calculation of metrics. In the LN test, traits were log-transformed only before the simulation. In the NL test, traits were log-transformed before the calculation of metrics. In the deficit and excess tests, an “excess” model was created by randomly selecting half of all traits for community assembly while the metrics were calculated by using all traits and a “deficit” model was created by randomly selecting half of all the traits for the calculation of metrics while all of the traits were used for community assembly. These two tests were not applicable for single-trait metrics. In min and max tests, local species richness was changed to the minimum or maximum value of each data set, respectively. SLA is an abbreviation for specific leaf area. FRic, FDis, and FD indicate functional richness, functional dispersion, and functional diversity based on a dendrogram, respectively. See Appendix H and Appendix I for full results.

content in the TREE data set. For the atypical scenarios, MNND, SDNDR, and the variance had sufficient statistical power at the single-community level slightly more often than MinNND and SDNNr. MinNND was unique in its performance in the BE– and EBR– scenarios, in which statistical powers of the other metrics were often insufficient especially at the single-community level. The sizes of the SESs were comparable to, or even exceeded, those of single-trait metrics of convergence.

Multi-trait metrics for divergence.—For the typical scenarios, in most cases, all of the eight metrics were able to detect trait divergence at the single-community level (Fig. 4, Appendix G). The SESs of FEve and FDiv were consistently lower than those of other metrics in all data sets. For the BE+, EBR+, RRBB+, and BBRR+ scenarios, the performances of MNND, FDis, and FD were better than those of other metrics while the performances of MinNND, Min/Max MST, and FEve were better for the BE– and EBR– scenarios. The performance of the metrics was generally better for the TREE data set than for the other two data sets. The SESs of the metrics, excluding FEve and FDiv, were

generally larger than those of the multi-trait metrics of convergence.

Responses to the other conditions of assembly

For a single-trait test of convergence, the scaling problem affected the statistical power of both the range and variance but its influence had a less frequent effect on variance (Table 1, Appendix H). The magnitude of the influence varied among the traits and was most serious for the total mass in the HERB data set. The performance of the multi-trait metrics for convergence was also influenced by the scaling problem (Table 1, Appendix I). Especially, in the HERB and TREE data sets, all of the three metrics failed to detect trait convergence under the NN condition. The problem of scaling was less serious for FISH data, for which only one of the 25 variables was continuous. FD lost statistical power when the analyzed trait number was insufficient against the number responsible for community assembly for HERB data.

For the single-trait metrics of divergence, all of the five conditions affected the results in at least one case (Table 2, Appendix J). The scaling problem weakened the statistical power of many metrics. However, MNND

TABLE 2. The performance of the eight single-trait metrics and the eight multi-trait metrics for divergence against the problems of scaling relationships, the sufficiency of traits, and local species richness for plant height, specific leaf area (SLA), and multiple traits in the HERB data set.

Traits and metrics	Tests							
	Control	NN	LN	NL	Deficit	Excess	Min	Max
Height								
MNND	×	×	×	×			×	†
MinNND	†	†	†	†			†	†
SDND	†	×	†	×			†	×
SDNN	†	×	×	NS			†	†
SDNDR	†	×	†	×			†	†
SDNNr	†	×	×	×			†	†
Variance	†	×	†	×			†	†
Kurtosis	×	×	×	NS			×	×
SLA								
MNND	†	†	†	†			†	†
MinNND	†	†	†	†			†	†
SDND	NS	NS	NS	NS			NS	NS
SDNN	NS	NS	NS	NS			NS	NS
SDNDR	×	×	×	×			×	×
SDNNr	×	×	×	×			×	×
Variance	†	†	†	†			†	†
Kurtosis	×	×	×	×			×	×
Multiple traits								
MNND	†	†	×	×	†	†	†	†
MinNND	†	†	×	×	×	×	†	†
Min/Max MST	†	†	×	×	×	×	×	†
FRic	†	†	×	×	†	×	×	†
FEve	†	†	×	×	×	×	×	†
FDiv	×	×	×	×	×	×	×	×
FDis	†	†	×	×	†	†	†	†
FD	†	†	×	×	†	†	†	†

Notes: Table components are as described in Table 1. Abbreviations are: MNND, mean nearest-neighbor distance; MinNND, minimum nearest-neighbor distance; SDND, standard deviation of neighbor distance; SDNN, standard deviation of nearest-neighbor distance; SDNDR, SDND divided by the range; SDNNr, SDNN divided by the range. FRic, FDis, FD, FEve, and FDiv indicate functional richness, functional dispersion, functional diversity, functional evenness, and functional divergence based on a dendrogram, respectively. See Appendix J and Appendix K for full results.

and MinNND were relatively robust to these conditions and they hold sufficient statistical power at the metacommunity level in many cases. The performance of multi-trait metrics for divergence was also affected by the sufficiency of analyzed traits as well as the scaling problem and the low levels of local species richness (Table 2, Appendix K). Here again the metrics, excluding FRic, were less seriously affected by the scaling problem for FISH data. Both a deficit and excess of analyzed traits weakened the statistical power in many cases. The performance of MNND, FDis, and FD was relatively robust against these conditions.

DISCUSSION

Assembly scenarios and other conditions considerably influenced the statistical powers of the metrics to distinguish nonrandom patterns in the distribution of a trait. In actual studies, the combination of these nonexclusive uncertainties would make it more difficult to detect signs of nonrandom community assembly in a trait distribution pattern. This fact, as well as the use of inappropriate metrics, can partly account for the

relatively high proportion of nonsignificant results in past studies (82% in studies of plants [Götzenberger et al. 2011]). However, our results demonstrate that these influences can be reduced by choosing the appropriate metrics.

Robustness under atypical assembly scenarios

In most cases, all of the metrics tested, for either convergence or divergence, had sufficient statistical power to detect patterns in the typical scenarios, at least at the metacommunity level. Furthermore, in many cases, if an appropriate metric was selected, the detection of the patterns at a single local-community level was also possible. However, the performance of SDND, SDNN, and kurtosis as a single-trait metric for divergence were considerably sensitive to trait distribution patterns in the species pool, and hence their use should be avoided. Although it has been argued that trait divergence is more difficult to detect than trait convergence (Kraft and Ackerly 2010), our results demonstrate that this is not always true once appropriate metric was chosen.

For the atypical scenarios, statistical powers were often considerably weakened. The statistical power of the single-trait metrics for convergence; i.e., range and variance, were largely discounted for the ER+, EB+, EBR+, RREE+, and EERR+ scenarios. These are scenarios in which convergent trait distributions resulting from environmental filtering were blurred by outliers that were introduced through the similarity-based or random addition of species. These results clearly indicate that both the range and variance are sensitive to outliers, which are expected to be introduced via processes such as the mass effect in actual studies. Interestingly, the use of variance, which has been considered to be somewhat tolerant to outliers (Cornwell and Ackerly 2009), often underperformed the use of the range.

The overall performance of multi-trait metrics for convergence was not as good as the performance of single-trait metrics. FRic, FDis, and FD successfully detected convergence at the local-community level only for the TREE data set. If local species richness was set to the same level as for HERB data (11 species), the results for the TREE data set were worse but still better than results for the other data set (data not shown). This indicates that the size of both the local community and species pool are responsible for the performance of these metrics (cf. Kraft et al. 2007, Kembel 2009). All of the metrics failed to detect convergence even at the metacommunity level for the EB+ and EBR+ scenarios, indicating their sensitivity to outliers as in the case of single-trait metrics. The performance of the three metrics was generally similar but FRic and FDis performed slightly better for the FISH data set. The performance of FDis and FD, which seem to tolerate outliers from the algorithm, were not better than the performance of FRic (but see Mason et al. [2012] for an abundance-weighted case). When combined with the results for single-trait metrics, our results indicate that the detection of trait convergence for presence/absence data using these metrics is virtually impossible when a convergent pattern is modified by extreme outliers.

Of the single-trait metrics for divergence, MinNND was unique in its tolerance to the EB– and EBR– scenarios, for which the performance of the other metrics was insufficient. This is reasonable because the algorithm of MinNND is independent of the range of a trait value. For the other metrics, whose algorithms are sensitive to the range of a trait value, evidence of biotic interactions would be difficult to detect if environmental filtering operates at the same time. Although the performance of both SDND and SDNN seem comparable to that of MinNND, these results should be discounted because these metrics carry the risk of generating a type I error in these scenarios. If environmental filtering operates, SDND and SDNN values tend to be small even without biotic interactions, because the possible range of neighbor distances depends on the range of a trait value (Lake and Ostling

2009). When the RE– scenario was used in place of R– as a null model, the statistical power of both SDND and SDNN for the EB– scenario was much decreased and was no longer comparable with that of MinNND (data not shown).

The responses of the multi-trait metrics for divergence to atypical scenarios can be categorized into two groups: MNND, FRic, FDis, and FD outperformed others for the EB+, EBR+, RRBB+, and BBRR+ scenarios, while MinNND, Min/Max MST, and FEve performed better in the EB– and EBR– scenarios. EB+, EBR+, RRBB+, and BBRR+ are the scenarios in which some species with extreme traits are necessarily included via the similarity-based addition of species. Therefore, metrics that are sensitive to these extreme values (MNND, FRic, FDis, and FD) should be advantageous for the detection of divergence. In contrast for the EB– and EBR– scenarios, environmental filtering remove species with extreme traits and thus the trait range can be even narrower than the null model (R–). In this case, metrics that are insensitive to the absolute size of the occupied volume in the multidimensional space of traits but capture the arrangement of species within the occupied volume (MinNND, Min/Max MST, and FEve) should be advantageous. Mouchet et al. (2010) recommended FRic as a metric of divergence following a comparative analysis of the measures of functional diversity. However, our analysis shows that this is no longer true in the atypical scenarios

*Robustness to trait scaling, sufficiency of traits,
and species richness*

The influence of the scaling problem was widespread and often severely weakened the statistical power of the metrics. In most cases, the best performances of the metrics were observed when trait values were log-transformed before both simulation and the calculation of metrics. In the case of a log-scale assembly, the situation is rather simple because a log-transformation of trait values before the calculation of metrics always improved the statistical power. In the case of a normal-scale assembly, however, the metrics for convergence and those for divergence were affected differently following a log-transformation of trait values. In most cases, for both single and multiple traits, log-transformation improved the performance of the metrics of convergence even when a trait–function relationship was linear. This indicates that, for these metrics, a trait distribution form in a species pool, possibly symmetry, is more important than a match of the scaling between a real process and an assumption for analysis. However, the performance of the metrics for divergence sometimes deteriorated following a log-transformation when a trait–function relationship was linear, indicating that a match of scaling is important for the performance of the metrics.

Sufficiency of analyzed traits and local species richness also affected the performance of metrics for

divergence to some extent. Many multi-trait metrics for divergence were affected by both a deficit and excess of analyzed traits. Although it is difficult to know the exact set of traits that operate in a community assembly process, the inclusion of traits whose role in an assembly process are not clear should be avoided (cf. Petchey and Gaston 2006). The statistical powers of the single-trait metrics for divergence often decreased when local species richness was low. Kraft et al. (2010) demonstrated that the statistical powers of SDNDR and SDNNr were negatively correlated with local species richness in their simulation. However, the influence of local species richness in this study was dependent on the metrics, data set, and trait values, suggesting that the relationships between the statistical power of the metrics and local species richness are not simple. Additionally, a trait distribution form in a species pool can be a potential risk to the performance of the metrics. For single-trait metrics, the performance in both typical and atypical scenarios differed considerably among traits even within a single data set.

Our results indicate that magnitudes of the influences from these conditions are different among the metrics and thus can be alleviated by choosing the appropriate metrics. Of the single-trait metrics for convergence, variance was less affected than the range by the problem of scaling. The three multi-trait metrics of convergence generally displayed similar responses to these conditions. Of the single-trait metrics for divergence, the performance of both MNND and MinNND was relatively robust to these conditions and they never lost sufficient statistical power at the metacommunity level. The performance of MNND was most robust also as a multi-trait metric of divergence.

As noted above, the detection of significant patterns would be more difficult in actual studies because the uncertainties that affect the statistical power are not exclusive. Therefore, the careful selection of metrics that are robust to these uncertainties is strongly encouraged. It is also important to assume that signs of nonrandom assembly are sometimes difficult to detect even when they actually exist, especially at a single local-community level.

Recommended metrics

Based on these results, we tentatively recommend the following metrics for use in studies on trait distribution patterns as a sign of community assembly processes (Table 3). For a single-trait analysis of convergence, we recommend use of the variance and range simultaneously. Although the range outperformed variance in some atypical assembly scenarios, variance was slightly more robust to the problem of scaling. For a multi-trait analysis of convergence, FRic and FDis slightly outperformed FD in their robustness to alternative assembly scenarios. Of the single-trait metrics for divergence, MNND, SDNDR, and variance were relatively tolerant to atypical scenarios. Additionally, the responses of

TABLE 3. Recommended metrics for the detection of trait distribution patterns.

Pattern	Main metric	Complementary metric
Trait convergence		
Single trait	variance, (range)	range, (variance)
Multiple trait	One of FRic or FDis	
Trait divergence		
Single trait	MNND	MinNND
Multiple trait	one of MNND, FDis or FD	one of MinNND or Min/Max MST

Notes: FRic, FDis, and FD indicate functional richness, functional dispersion, and functional diversity based on a dendrogram, respectively. MNND, MinNND, and Min/Max MST represents mean nearest-neighbor distance, minimum nearest-neighbor distance, and ratio of the shortest distance to the longest distance in a minimum spanning tree, respectively.

MinNND to atypical scenarios were unique and complementary to the other metrics. Because MNND and MinNND are relatively robust to the problem of scaling and low levels of local species richness, we recommend MNND as the main metric and MinNND as a complement. Of the multi-trait metrics of divergence, performance of MNND, FDis, and FD was complementary to that of MinNND and Min/Max MST. Responses to the scaling problem also varied between the two groups, with the former being relatively robust. Therefore, we recommend use of MNND, FDis, or FD in combination with either MinNND or Min/Max MST.

Conclusion

We showed that the choice of metrics can potentially have a substantial impact on the results of community-assembly studies. The high frequency of nonsignificant results in past studies could be partly due to the use of metrics sensitive to uncertainties in assembly processes. Unstable performances of the metrics may lead to detection of false differences among traits and studies. Additionally, past studies have shown that the choice of a null model can also substantially affect the results of studies of community assembly (e.g., Gotelli 2000, Kembel 2009). Therefore, future studies of community-assembly processes should be conducted based on standardized methods that have been validated by recent studies (Götzenberger et al. 2011). Reanalyses of past studies using such methods and the comparison of the results with the original study would also assist both our understanding of assembly processes and further methodological refinements.

ACKNOWLEDGMENTS

We thank the Forest Department of Sarawak and Sarawak Forest Research Corporation for their kind permission to conduct research in Lambir Hills National Park. We also thank the many people who contributed to the 52-ha Long-Term Ecological Research Project. The 52-ha Long-Term Ecological Research Project is a collaborative project of the Forest Department of Sarawak, Malaysia, the Center for Tropical

Forest Science of the Smithsonian Institution, USA (under NSF awards DEB-9107247 and DEB-9629601), and Osaka City University, Japan (under Monbusho grants 06041094 to T. Yamakura, 08NP0901 to S. Tamura, and 09NP0901 to S. Sasaki). Two anonymous referees provided helpful comments on a previous version of this manuscript. This study was partly supported by a Research Fellowship for Young Scientists (to M. Aiba) and grants from the Ministry of Environment (No. S-9-3 to M. Aiba, M. Katabuchi, and T. Hiura and No. S-9-4 to S. Matuzaki). M. Katabuchi was supported by a Chinese Academy of Sciences postdoctoral research award and a Grant-in-Aid for JSPS Fellows.

LITERATURE CITED

- Aiba, M., H. Takafumi, and T. Hiura. 2012. Interspecific differences in determinants of plant species distribution and the relationships with functional traits. *Journal of Ecology* 100:950–957.
- Arita, H. T. 1997. Species composition and morphological structure of the bat fauna of Yucatan, Mexico. *Journal of Animal Ecology* 66:83–97.
- Cavender-Bares, J., K. H. Kozak, P. V. A. Fine, and S. W. Kembel. 2009. The merging of community ecology and phylogenetic biology. *Ecology Letters* 12:693–715.
- Cornwell, W. K., and D. D. Ackerly. 2009. Community assembly and shifts in plant trait distributions across an environmental gradient in coastal California. *Ecological Monographs* 79:109–126.
- Findley, J. 1976. The structure of bat communities. *American Naturalist* 110:129–139.
- Gotelli, N. J. 2000. Null model analysis of species co-occurrence patterns. *Ecology* 81:2606–2621.
- Gotelli, N. J., and D. J. McCabe. 2002. Species co-occurrence: a meta-analysis of J. M. Diamond's assembly rules model. *Ecology* 83:2091–2096.
- Götzenberger, L., et al. 2011. Ecological assembly rules in plant communities—approaches, patterns and prospects. *Biological Reviews* 87:111–127.
- HilleRisLambers, J., P. B. Adler, W. S. Harpole, J. M. Levine, and M. M. Mayfield. 2011. Rethinking community assembly through the lens of coexistence theory. *Annual Review of Ecology, Evolution, and Systematics* 43:227–248.
- Keddy, P. A. 1992. Assembly and response rules—2 goals for predictive community ecology. *Journal of Vegetation Science* 3:157–164.
- Kembel, S. W. 2009. Disentangling niche and neutral influences on community assembly: assessing the performance of community phylogenetic structure tests. *Ecology Letters* 12: 949–960.
- Kraft, N. J. B., and D. D. Ackerly. 2010. Functional trait and phylogenetic tests of community assembly across spatial scales in an Amazonian forest. *Ecological Monographs* 80: 401–422.
- Kraft, N. J. B., W. K. Cornwell, C. O. Webb, and D. D. Ackerly. 2007. Trait evolution, community assembly, and the phylogenetic structure of ecological communities. *American Naturalist* 170:271–283.
- Kraft, N. J. B., R. Valencia, and D. D. Ackerly. 2008. Functional traits and niche-based tree community assembly in an Amazonian forest. *Science* 322:580–582.
- Kursar, T. A., K. G. Dexter, J. Lokvam, R. T. Pennington, J. E. Richardson, M. G. Weber, E. T. Murakami, C. Drake, R. McGregor, and P. D. Coley. 2009. The evolution of antiherbivore defenses and their contribution to species coexistence in the tropical tree genus *Inga*. *Proceedings of the National Academy of Sciences USA* 106:18073–18078.
- Lake, J. K., and A. Ostling. 2009. Comment on “Functional traits and niche-based tree community assembly in an Amazonian forest”. *Science* 324:1015.
- Laliberte, E., and P. Legendre. 2010. A distance-based framework for measuring functional diversity from multiple traits. *Ecology* 91:299–305.
- MacArthur, R., and R. Levins. 1967. Limiting similarity convergence and divergence of coexisting species. *American Naturalist* 101:377–385.
- Mason, N. W. H., F. de Bello, D. Mouillot, S. Pavoine, and S. Dray. 2012. A guide for using functional diversity indices to reveal changes in assembly processes along ecological gradients. *Journal of Vegetation Science* 24:794–806.
- Mason, N. W. H., C. Lanoiselee, D. Mouillot, P. Irz, and C. Argillier. 2007. Functional characters combined with null models reveal inconsistency in mechanisms of species turnover in lacustrine fish communities. *Oecologia* 153:441–452.
- Matsuzaki, S., T. Sasaki, and M. Akasaka. 2013. Consequences of the introduction of exotic and translocated species and future extirpations on the functional diversity of freshwater fish assemblages. *Global Ecology and Biogeography* 22: 1071–1082.
- Mayfield, M. M., and J. M. Levine. 2010. Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecology Letters* 13:1085–1093.
- Mouchet, M. A., S. Villegger, N. W. H. Mason, and D. Mouillot. 2010. Functional diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules. *Functional Ecology* 24:867–876.
- Petchey, O. L., and K. J. Gaston. 2002. Functional diversity (FD), species richness and community composition. *Ecology Letters* 5:402–411.
- Petchey, O. L., and K. J. Gaston. 2006. Functional diversity: back to basics and looking forward. *Ecology Letters* 9:741–758.
- R Development Core Team. 2012. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.r-project.org
- Ricklefs, R. E., and J. Travis. 1980. A morphological approach to the study of avian community organization. *Auk* 97:321–338.
- Shmida, A., and M. V. Wilson. 1985. Biological determinants of species diversity. *Journal of Biogeography* 12:1–20.
- Spasojevic, M. J., and K. N. Suding. 2012. Inferring community assembly mechanisms from functional diversity patterns: the importance of multiple assembly processes. *Journal of Ecology* 100:652–661.
- Stubbs, W. J., and J. B. Wilson. 2004. Evidence for limiting similarity in a sand dune community. *Journal of Ecology* 92: 557–567.
- Villegger, S., N. W. H. Mason, and D. Mouillot. 2008. New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology* 89:2290–2301.
- Watanabe, K. 2012. Faunal structure of Japanese freshwater fishes and its artificial disturbance. *Environmental Biology of Fishes* 94:533–547.
- Webb, C. O., D. D. Ackerly, M. A. McPeck, and M. J. Donoghue. 2002. Phylogenies and community ecology. *Annual Review of Ecology, Evolution, and Systematics* 33: 475–505.
- Weihner, E., G. D. P. Clarke, and P. A. Keddy. 1998. Community assembly rules, morphological dispersion, and the coexistence of plant species. *Oikos* 81:309–322.
- Weihner, E., D. Freund, T. Bunton, A. Stefanski, T. Lee, and S. Bentivenga. 2011. Advances, challenges and a developing synthesis of ecological community assembly theory. *Philosophical Transactions of the Royal Society B* 366:2403–2413.
- Weihner, E., and P. A. Keddy. 1995. Assembly rules, null models, and trait dispersion—new questions from old patterns. *Oikos* 74:159–164.
- Weihner, E., and P. A. Keddy. 1999. *Ecological assembly rules: perspectives, advances, retreats*. Cambridge University Press, Cambridge, UK.

SUPPLEMENTAL MATERIAL

Appendix A

Summary of the metrics examined in the study ([Ecological Archives E094-265-A1](#)).

Appendix B

Summary of the configuration of the simulation regarding community assembly conditions and related data preparation ([Ecological Archives E094-265-A2](#)).

Appendix C

A schematic flow of the process of the analysis ([Ecological Archives E094-265-A3](#)).

Appendix D

The performance of the two single-trait metrics for convergence against 10 alternative assembly scenarios ([Ecological Archives E094-265-A4](#)).

Appendix E

The performance of the three multi-trait metrics for convergence against 10 alternative assembly scenarios ([Ecological Archives E094-265-A5](#)).

Appendix F

The performance of the eight single-trait metrics for divergence against 10 alternative assembly scenarios ([Ecological Archives E094-265-A6](#)).

Appendix G

The performance of the eight multi-trait metrics for divergence against 10 alternative assembly scenarios ([Ecological Archives E094-265-A7](#)).

Appendix H

The performance of the two single-trait metrics for convergence against the problems of scaling relationships and local species richness ([Ecological Archives E094-265-A8](#)).

Appendix I

The performance of the three multi-trait metrics for convergence against the problems of scaling relationships, the sufficiency of traits, and local species richness ([Ecological Archives E094-265-A9](#)).

Appendix J

The performance of the eight single-trait metrics for divergence against the problems of scaling relationships and local species richness ([Ecological Archives E094-265-A10](#)).

Appendix K

The performance of the eight multi-trait metrics for divergence against the problems of scaling relationships, the sufficiency of traits, and local species richness ([Ecological Archives E094-265-A11](#)).

Supplement

R code for community assembly simulation and calculation of metrics ([Ecological Archives E094-265-S1](#)).