



ELSEVIER

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

A sampling formula for ecological communities with multiple dispersal syndromes



Thijs Janzen ^{a,*}, Bart Haegeman ^b, Rampal S. Etienne ^a

^a University of Groningen, Groningen Institute for Evolutionary Life Sciences, Box 11103, 9700 CC Groningen, The Netherlands

^b Centre for Biodiversity Theory and Modelling, Station d'Ecologie Expérimentale du CNRS, 2 route du CNRS, 09200 Moulis, France

HIGHLIGHTS

- We introduce a sampling formula that takes into account dispersal syndromes.
- Using simulated data we validate our sampling formula.
- We apply our sampling formula on tropical tree data from BCI, Panama.
- We show that including dispersal syndrome information improves the fit to the data.

ARTICLE INFO

Article history:

Received 6 September 2014

Received in revised form

10 March 2015

Accepted 13 March 2015

Available online 26 March 2015

Keywords:

Neutral theory

Guilds

Dispersal syndromes

BCI

ABSTRACT

Over the past decade, the neutral theory of biodiversity has stirred up community assembly theory considerably by suggesting that stochasticity in the form of ecological drift is an important factor determining community composition and community turnover. The neutral theory assumes that all species within a community are functionally equivalent (the neutrality assumption), and therefore applies best to communities of trophically similar species. Evidently, trophically similar species may still differ in dispersal ability, and therefore may not be completely functionally equivalent. Here we present a new sampling formula that takes into account the partitioning of a community into two guilds that differ in immigration rate. We show that, using this sampling formula, we can accurately detect a subdivision into guilds from species abundance distributions, given ecological data about dispersal ability. We apply our sampling formula to tropical tree data from Barro Colorado Island, Panama. Tropical trees are divided depending on their dispersal mode, where biotically dispersed trees are grouped as one guild, and abiotically dispersed trees represent another guild. We find that breaking neutrality by adding guild structure to the neutral model significantly improves the fit to data and provides a better understanding of community assembly on BCI. Our findings are thus an important step towards an integration of neutral and niche theory.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

The astonishing biodiversity around the globe, especially in the tropics, makes one wonder how this biodiversity has originated and how it can be maintained. Traditionally, species composition in an ecological community is explained by species-specific traits and species requirements. By contrast, the more recent neutral theory (Hubbell, 2001; Etienne and Olf, 2004; Rosindell et al., 2011) explains species composition in an ecological community by stochastic demography and dispersal. This theory deliberately

neglects species-specific differences (the neutrality assumption). It oversimplifies ecology in order to emphasize that ecological drift is an important factor in community assembly (Rosindell et al., 2011; Wennekes et al., 2012). Despite this simplification the model can convincingly explain various biodiversity patterns, suggesting that indeed ecological drift is an important factor in community assembly (Etienne and Olf, 2004; Alonso et al., 2006).

The neutrality assumption states that all the individuals within an ecological community have the same birth rates, death rates, dispersal rates and speciation rates, irrespective of the species the individuals belong to (Hubbell, 2001). The ecological community is assumed to consist of individuals of functionally equivalent species that compete with each other for space in the community. As a result, patterns in abundance predicted by the theory are purely

* Corresponding author.

E-mail address: thijsjanzen@gmail.com (T. Janzen).

the result of drift, speciation and immigration, and not the result of competitive asymmetries between the species in the local community. The neutrality assumption is the most debated assumption of the Neutral Theory of Biodiversity (McGill et al., 2006; Purves and Pacala, 2008; Turnbull et al., 2008; Gotelli et al., 2009). Most importantly, the neutrality assumption refutes the idea of the unique correspondence between a species and its niche (interpreted here as the set of conditions and requirements for a species to survive (Hutchinson, 1958), although the exact meaning of the niche concept is unclear (Chase and Leibold, 2003; McInerney and Etienne, 2012)). More specifically, the neutrality assumption ignores specific interactions between species and species-specific adaptations, such as habitat specialization; furthermore it ignores the effects of density dependence, ecological succession and the impact of trait differences (Purves and Turnbull, 2010).

Several models explore the continuum between niche and neutral models by looking at the effect of differences in birth and death rates, which might arise through differences in intraspecific and interspecific competition (Jabot and Chave, 2011). In the fully neutral case, intraspecific and interspecific competition are identical, whereas classic coexistence theory predicts that coexistence is promoted when intraspecific competition is stronger than interspecific competition (Adler et al., 2007). Combining community assembly with classic coexistence modelling, Noble and Fagan (2011) showed that when intraspecific competition exceeds interspecific competition, patterns similar to a fully neutral model emerge. Along similar lines, Haegeman and Loreau (2011) investigated how altering the difference between intraspecific and interspecific competition affects the species abundance distribution. They focused on the parameter space where intraspecific competition exceeds interspecific competition, i.e. where classical theory predicts coexistence. They found that with increasing interspecific competition, fluctuations in local community size increase, and the local community becomes more prone to extinction. More importantly they found that altering the difference between intraspecific and interspecific competition only influenced the species abundance distribution marginally, and concluded that from species abundance data alone it might be difficult to assess the degree of intraspecific versus interspecific competition. Proceeding even further, Pigolotti and Cencini (2013) found an analytical expression for the expected species abundance distribution where the degree of intraspecific and interspecific competition can be tuned by a single parameter. Their results suggest a profound impact of the degree of intraspecific versus interspecific competition not only on the species abundance distribution, but also on the average species lifetime and on the total variation in species lifetimes in the local community.

Competitive asymmetry could also result in differences in birth rate irrespective of competition. Du and colleagues found that introducing competitive asymmetry breaks down neutral patterns (Du et al., 2011), but also that these effects can be counteracted by negative density dependence: communities with intermediate competitive asymmetry and intermediate levels of negative density dependence show species abundance distributions that are indistinguishable from neutral distributions, suggesting that neutral patterns can emerge from non-neutral assumptions.

Breaking neutrality through the introduction of differences in dispersal rather than birth and death rates has been less well studied. Turnbull et al. (2008) investigated the effect of an equalizing trade-off between seed mass and seed number on neutrality. They found that after including such a trade-off, neutral patterns break down as soon as seed arrival becomes stochastic. Liu and Zhou (2011) relaxed the neutrality assumption by introducing stochastic differences in dispersal ability between species. As the standard deviation of the Gaussian distribution governing these differences increases, the neutral patterns break down and

community assembly becomes deterministic, where species with a high dispersal ability tend to dominate the local community. Liu and colleagues compared the effect of differences in dispersal ability to data generated with the neutral model without these differences, but did not confront their model with empirical data.

Trophically similar species may come close to fitting the neutrality assumption, but differences in dispersal may prevent them from being functionally equivalent. Differences in dispersal might arise through differences in seed size (Muller-Landau and Hardesty, 2005), differences in fruit size (Seidler and Plotkin, 2006) but might also manifest themselves as differences in flight prowess (Valtonen et al., 2013) or differences in pelagic larval duration in coral reef fish (Victor and Wellington, 2000; Almany et al., 2007). In this paper we will study such differences in dispersal, focusing on tropical trees. The majority of tropical tree species (73%) disperse through animal means (Muller-Landau and Hardesty, 2005), such as bats, birds, mammals, ants and sometimes even fish. The other 27% of tree species relies on abiotic factors to disperse their seeds, such as wind, water or ballistics.

By definition, neutral models fail to include differences in dispersal between species that share the same local community and metacommunity. Here we present a model where we classify species according to their dispersal syndrome. We will call the resulting classes guilds. This is a simple, but important step towards incorporating differences between species without needing to explicitly quantify these differences for every species in the community. Instead we only need to quantify the differences between guilds, and assess the importance of these differences for community assembly. Our model breaks the neutrality assumption of the standard neutral model (Hubbell, 2001; Etienne and Alonso, 2005) by subdividing the community into two guilds, where each guild is a group of species that have the same dispersal rate. Between guilds, dispersal rates may differ, but the speciation rate, birth and death rates are identical. We show that our model can accurately distinguish between datasets including a guild structure, and datasets that do not have any guild structure. Our model is able to detect signatures of guild structure from the species abundance distribution when combined with ecological data regarding dispersal, refuting the idea that the species abundance distribution does not contain sufficient information to draw conclusions about underlying community assembly mechanisms. Secondly, we show that parameter estimates obtained with our model are accurate and differ considerably from estimates obtained using the standard neutral model without guild structure. Lastly we illustrate the model by applying it to the tropical tree dataset of Barro Colorado Island (BCI).

2. Model

We assume that there are two guilds X and Y that differ in their immigration parameter m_i ($i=X, Y$); all species *within* each guild share the same migration parameter m_i . All species, regardless of the guild they belong to, have the same fundamental biodiversity number θ , as in the standard neutral model. In the metacommunity, every time step one individual dies and is replaced by an individual from either guild X or guild Y . With probability ν_X a speciation event occurs resulting in a new species that belongs to guild X and to guild Y with probability ν_Y . With probability $1-\nu_X-\nu_Y$ no speciation event occurs; then the new individual belongs to guild X or Y depending on the relative abundance of guilds X and Y in the metacommunity. Over time the relative abundances of both guilds reach a dynamical equilibrium. The equations we derive in Appendix A are applicable to the general case where $\nu_X \neq \nu_Y$. However, we found that the statistical power in such cases is much reduced. Furthermore, we focus here on

breaking neutrality in dispersal rather than speciation. Hence, we treat speciation to be equally likely between guilds, such that half of the speciation events result in a new species of guild X and half of the speciation events result in a new species of guild Y. This implies that $\nu_X = \nu_Y = \nu/2$.

In the local community, a deceased individual can be replaced by either an individual from the local community with probability $1 - m_X - m_Y$ or by an immigrant from the metacommunity; this is an individual from guild X with probability m_X or an individual of guild Y with probability m_Y . The migration probability m_i of guild i depends on the dispersal ability of guild i , here called α_i , and the relative abundance of the guild in the metacommunity, p_i . The migration probability of guild i is then given by $m_i = \alpha_i p_i$. The dispersal ability α_i is bounded between [0,1] where values close to zero indicate low dispersal ability and values close to one indicate good dispersal ability.

3. Sampling formula

In the case of a single guild, fitting the neutral community model to data makes use of the Etienne sampling formula (Etienne, 2005), which is a dispersal-limited extension of the Ewens sampling formula (Ewens, 1972). This formula gives the probability of a data set of species abundances in a sample as a function of the model parameters. Here we briefly describe the extension of the neutral sampling formula to the case of two guilds and we provide a more detailed derivation in Appendix A.

The stationary abundance distribution of the neutral community depends on the fundamental biodiversity number θ and on the fundamental dispersal number I , which are defined as

$$\theta = \frac{\nu(J_M - 1)}{1 - \nu} \text{ and } I = \frac{m(J - 1)}{1 - m}$$

where ν is the speciation rate, J_M is the metacommunity size, J is the local community size, and m is the migration probability. Likewise, the stationary abundance distribution of the two-guild neutral community can be expressed in terms of guild-specific biodiversity numbers θ_X and θ_Y and guild-specific dispersal numbers I_X and I_Y (see Appendix (S4A) and (S4B)). Our assumption on neutrality with respect to speciation implies $\theta_X = \theta_Y = \theta/2$. We define migration from the metacommunity to the local community as the product of dispersal ability and the relative frequency of the guild in the metacommunity, such that for guild i : $m_i = \alpha_i p_i$, where α_i is the dispersal ability of guild i and p_i is the relative frequency of guild i in the metacommunity. For the dispersal numbers, using the guild-specific immigration probabilities $m_i = \alpha_i p_i$, we obtain I_X and I_Y :

$$I_X = \frac{\alpha_X p_X (J - 1)}{1 - \alpha_X p_X - \alpha_Y p_Y} \text{ and } I_Y = \frac{\alpha_Y p_Y (J - 1)}{1 - \alpha_X p_X - \alpha_Y p_Y} \tag{1}$$

Note that $\theta = \theta_X + \theta_Y$ and $I = I_X + I_Y$, that is, the speciation and immigration processes are split out over the two guilds. Using the guild-specific biodiversity and dispersal numbers, the two-guilds abundance distribution is (see Eq. (S8)):

$$P(\vec{D}_X, \vec{D}_Y | \theta, I_X, I_Y, J) = P(J_X, J_Y | I_X, I_Y, J) P(\vec{D}_X | \theta_X, I_X, J_X) P(\vec{D}_Y | \theta_Y, I_Y, J_Y) \tag{2}$$

where vector \vec{D}_X contains the species abundances in guild X and vector \vec{D}_Y contains the species abundances in guild Y. The second factor in the right-hand side, $P(\vec{D}_X | \theta_X, I_X, J_X)$, is the one-guild Etienne sampling formula of guild X, as if it was isolated from guild Y (but with the appropriate biodiversity and dispersal number). The third factor in the right-hand side, $P(\vec{D}_Y | \theta_Y, I_Y, J_Y)$, is the Etienne sampling formula of guild Y, as if it was isolated from guild X. The two isolated guild abundance distributions are combined through the

probability distribution of the guild sizes (see Eqs. (S1) and (S6)),

$$P(J_X, J_Y | I_X, I_Y, J) = \frac{J!}{(I_X + I_Y)^J} \frac{(I_X)_{J_X} (I_Y)_{J_Y}}{J_X! J_Y!} \tag{3}$$

Hence, as far as the stationary abundance distributions are concerned, the dependence between guilds is concentrated in the guild sizes. In other words, after conditioning the abundance distribution on the guild sizes, and for given values of I_X and I_Y , the abundance distributions of guild X and guild Y are independent. This shows that the one-guild abundance distributions are the fundamental building blocks of the abundance distribution of a community consisting of two (or more) neutral guilds.

Eq. (2) is not yet the full sampling formula, because I_X and I_Y depend on p_X and p_Y (see Eq. (1)), which are variables, not parameters. The distribution of the metacommunity guild sizes p_X and $p_Y = 1 - p_X$ is a beta distribution (set $\theta_X = \theta_Y = \theta/2$ in Eq. (S10)),

$$\rho(p_X | \theta) = \frac{\Gamma(\theta)}{\Gamma(\theta/2)^2} p_X^{(\theta/2)-1} (1 - p_X)^{(\theta/2)-1} \tag{4}$$

Hence, we find the full two-guild sampling formula by integrating over all possible values of p_X ,

$$P(\vec{D}_X, \vec{D}_Y | \theta, \alpha_X, \alpha_Y) = \int_0^1 P(J_X, J_Y) P(\vec{D}_X | \theta, I_X, J_X) P(\vec{D}_Y | \theta, I_Y, J_Y) \rho(p_X | \theta) dp_X \tag{5}$$

Code to calculate this sampling formula for a dataset is available in the GUILDS package for R.

4. Conditioning on guild size

Using Eq. (3), we can calculate the expected guild sizes, given α_X , α_Y and J :

$$E(J_X | \alpha_X, \alpha_Y, J) = \frac{J I_X}{(I_X + I_Y)} = \frac{J \alpha_X p_X}{\alpha_X p_X + \alpha_Y p_Y} \tag{6a}$$

$$E(J_Y | \alpha_X, \alpha_Y, J) = \frac{J I_Y}{(I_X + I_Y)} = \frac{J \alpha_Y p_Y}{\alpha_X p_X + \alpha_Y p_Y} \tag{6b}$$

From the expected guild sizes it follows that the ratio of guild sizes is equal to the ratio of dispersal rates: $E[J_X] / E[J_Y] = \alpha_X p_X / \alpha_Y p_Y$. This is close to α_X / α_Y for typical values of θ such as $\theta = 50$, because then the beta distribution of Eq. (4) is sharply peaked around 0.5. Explorations of the sampling formula confirmed that our estimated values for the dispersal parameter closely mimic the ratio of guild sizes (Fig. A1). This is also intuitively understandable, consider two guilds, with one guild being twice as large as the other guild (i.e. $J_X = 2J_Y$). In order to reach such a skewed distribution of individuals, either this distribution is already present in the metacommunity, or there is a large skew in dispersal ability. The beta distribution we assume in the metacommunity (Eq. (4)) does allow for some divergence from a 50/50 ratio of guild sizes, but on average we do not expect the metacommunity to be highly skewed towards one particular guild. As a result, we expect that given a dataset with differently sized guilds, our sampling formula will estimate differences in dispersal ability and hence assume some form of guild structure, even when these differences in guild size are not caused by guild effects. We have circumvented this problem by conditioning our sampling formula on guild sizes. This yields the probability of our data given the parameter values and the guild sizes. As a result, differences in parameters, and any detected guild effects, are independent of guild size and solely dependent on differences in dispersal ability.

The joint probability of a combination of guild sizes is given by (see Appendix equation (S12)):

$$P(J_X, J_Y | \theta, \alpha_X, \alpha_Y, J) = \int_0^1 P(J_X, J_Y | I_X, I_Y, J) \rho(p_X | \theta) dp_X \quad (7)$$

We condition by dividing our sampling formula by this likelihood, and thus obtain:

$$P(\bar{D}_X, \bar{D}_Y | \theta, \alpha_X, \alpha_Y, J_X, J_Y) = \frac{\int_0^1 P(J_X, J_Y) P(\bar{D}_X | \theta, I_X, J_X) P(\bar{D}_Y | \theta, I_Y, J_Y) \rho(p_X | \theta) dp_X}{\int_0^1 P(J_X, J_Y | I_X, I_Y, J) \rho(p_X | \theta) dp_X} \quad (8)$$

The conditioned sampling formula no longer results in a relation between guild size and estimated dispersal ability (Fig. A1). Code to numerically integrate the conditioned sampling formula, given a dataset, is partly based on the TeTaMe programme (Jabot et al., 2008) and is available in the GUILDS package for R (Janzen, 2014).

5. Testing on artificial data

The two-guilds sampling formula can be reduced to the Etienne sampling formula of the standard, single-guild, neutral model by setting the dispersal ability of both guilds to the same value ($\alpha_X = \alpha_Y$). Throughout the text we will refer to this model as D0. As an alternative model, we allow the dispersal rates to differ between guilds ($\alpha_X \neq \alpha_Y$). We will refer to this model as D1. To assess how well we can distinguish the two models from each other, we generated 100 replicate datasets for all unique combinations of $\theta = [30, 100, 300]$ and $\alpha = [0.001, 0.01, 0.1]$. There are 9 different combinations for D0 (all combinations of θ and α , 3×3), and 9 different combinations for D1 (3 different θ values with one of three α combinations: $[0.001, 0.01]$, $[0.001, 0.1]$ or $[0.01, 0.1]$, which again yields 3×3 combinations). Community size was set at 20,000 individuals. We generated artificial datasets using a three-step procedure: first, the sizes of guilds X and Y in the metacommunity (assuming that the metacommunity size is infinite), were drawn from a beta distribution with parameter θ (Eq. (4)). Secondly, the total number of individuals (J_i) of each guild i in the local community was drawn from Eq. (2) with parameters J , I_i . The species abundance distribution of each guild was then generated using the urn scheme as described in Etienne (2005) with parameters J and I . Code to generate a local community according to the aforementioned procedure is available in the GUILDS package for R.

For every artificial dataset we performed maximum likelihood estimation for the two models (D0, D1), where the likelihood maximization was started at the parameter values used to generate the data with. The obtained likelihood values for the maximum likelihood optimum were used to calculate the Akaike Information Criterion (AIC) (Akaike, 1974):

$$AIC = 2k - 2 \ln(L)$$

where k is the number of free parameters of the model, and L is the Maximum Likelihood of the model. The number of free parameters in the model is 2 for the D0 model, (θ and α) and 3 for the D1 model (θ , α_X , and α_Y). After calculation of our AIC values, we compared the AIC scores with AIC weights (Wagenmakers and Farrell, 2004):

$$w_i(AIC) = \frac{\exp(-(1/2)\Delta_i AIC)}{\sum_{k=1}^K \exp(-(1/2)\Delta_k AIC)}$$

where $\Delta_i AIC = AIC_i - \min(AIC)$, and K is the total number of models compared (in this case, 2). AIC weight w_i can be interpreted as the probability of model i being the best model among the models considered.

To assess the accuracy of our parameter estimates we performed Maximum Likelihood estimation for the same simulated communities, but now starting at a grid of 2^d initial parameter combinations (with d being the number of free parameters in the model, 2 for D0, 3 for D1), not necessarily including the values used to generate the data. The initial values contained all possible combinations for θ of $[30, 300]$ and α of $[0.001, 0.1]$. Using the 100 obtained Maximum Likelihood estimates we calculated the 25th, 50th and 75th percentiles.

6. Empirical data

To illustrate the application of our sampling formula, we performed both model selection and parameter estimation techniques on a well-studied dataset of tropical forest trees: the Neotropical community dataset of Barro Colorado Island (BCI), Panama (Condit et al., 1996, 2002; Hubbell, 2001; Volkov et al., 2003; Etienne, 2005). The dataset consists of the abundance of all free-standing woody plants with > 10 cm diameter at breast height in 50 ha of forest. We analysed censuses from 1982, 1985, 1990, 1995, 2000 and 2005. The resulting dataset consists of recorded abundances of 6 different years, 252 woody plant species, with a summed total over 20,000 individuals per census. Tree species in this data set are grouped according to their dispersal syndrome, where all biotically dispersed (i.e. via birds (171 species), bats (37 species) and mammals (194 species)) trees are grouped together in one guild, and all abiotically dispersed trees (i.e. wind (33 species), water (1 species), ballistic means (10 species)) are grouped together in another guild (Muller-Landau and Hardesty, 2005).

7. Posterior analysis

To elucidate the differences between the models, we calculated the expected species abundance distribution for every dataset, using a hybrid approach of simulation and exact calculation. The expected species abundance distribution was approximated as follows: Given J_X and J_Y we obtained I_X and I_Y by first drawing p_X from:

$$P(p_X | \theta, \alpha_X, \alpha_Y, J_X, J_Y) = \frac{P(J_X, J_Y | I_X, I_Y, J) \rho(p_X | \theta)}{\int_0^1 P(J_X, J_Y | I_X, I_Y, J) \rho(p_X | \theta) dp_X} \quad (9)$$

and then calculated I_X and I_Y using Eq. (1). We then calculated the expected number of species in guild i with n individuals using Eq. (6) from Etienne and Alonso (2005):

$$E[S_n | \theta, I, J] = \frac{\theta}{(I)_J} \binom{J}{n} \int_0^1 (Ix)_n [I(1-x)]_{J-n} \frac{(1-x)^{\theta-1}}{x} dx \quad (10)$$

Because drawing from the distribution in Eq. (9) is inherently stochastic, we averaged over 100 replicates to obtain the final expected abundance distribution.

Furthermore we studied the power of the imposed guild structure on the data, i.e. we determined whether adding guild structure to the data adds information. We did this by randomizing the datasets 100 times, by randomly assigning species to a guild (with equal probability), thus removing any guild structure. For every randomized dataset we then used Maximum Likelihood Estimation with both models, initialized at a grid of initial values (all possible combinations for θ of $[30, 300]$ and α of $[0.001, 0.1]$). Using AIC weights we estimated which of the two models best explained the data. If the imposed guild structure provides additional information on the dataset, we would expect that after randomization any signals of guild structure are lost, and the D0 model is favored. Conversely, if any random subdivision into two

groups would also cause detection of guild structure and the D1 model would be favored after randomization the detection of guild structure in the original data set is ecologically meaningless.

Lastly we used our sampling formula to evaluate the goodness-of-fit of the model with guild structure to empirical data, by performing an ‘Exact’ test of neutrality (Etienne 2007). Using the parameter estimations obtained with Maximum Likelihood Estimation, we generated 100 different data sets (using the maximum likelihood estimates for θ and α). Datasets were generated by first drawing p_x from Eq. (9), then using J_x and J_y from the data and our obtained maximum likelihood estimates for θ and α we generated the species abundance distribution for each guild using the urn scheme as described in Etienne (2005). For these 100 replicate datasets we then again calculated the model parameters and likelihood using Maximum Likelihood Estimation as described in the previous section (this is different from Etienne (2007) who did not maximize the likelihood – the procedure used here is less conservative (Efron and Tibshirani, 1993)). If the likelihood of the empirical data is smaller than the large majority of likelihoods from the replicate datasets, the observed community differs strongly from a neutral community. If however the likelihood of the empirical data is not different from the obtained frequency distribution, the observed abundance distribution does not contain a detectable signal of non-neutrality.

8. Results

The ability to accurately select the correct model is essential for the implementation of our sampling formula. Not only should our sampling formula favour the more complex model when the data warrants it, it should also reject the more complex model if the data shows no sign of guild structure. We tested the ability of the sampling formula to detect guild structure by confronting it with artificially generated data. The artificial data contained either no guild structure at all, or was generated including different degrees of difference in dispersal limitation between guilds. Data generated using the D0 model (no guild structure), was correctly identified as having no guild structure in 88% of all simulated datasets (794 out of 900 datasets were correctly identified as D0). Data generated using the D1 model was correctly identified as having guild structure in 85% of all simulated datasets (765 out of 900 datasets were correctly identified as D1). Hence, type I ($116/900=12.88\%$) and Type II ($135/900=15\%$) errors are very similar, and the model adequately detects guild structure in the majority of the simulated datasets we analysed.

Using artificial data generated with either the D0 or the D1 model, we tested the precision and bias of the new guilds sampling formula. We report the 25th, 50th and 75th percentiles of 100 replicates (Table 1). For the D0 model, the parameter value used to generate the data fell between the 25th and 75th percentiles for 8 out of 9 parameter combinations. For the D1 model, the parameter values used to generate the data all fell between the 25th and 75th percentiles. The bias of the D0 model was small: the 50th percentiles of maximum likelihood estimates for datasets simulated with high dispersal values ($\alpha > 0.001$) were close to those used to simulate the datasets. Combinations with low dispersal ($\alpha=0.001$) tended to have a median slightly underestimating θ , but an accurate estimate of α_x , except for the combination [300,0.001], for which none of the percentiles included the correct α_x value. Precision of the D0 model was high, with the overall spread of estimated parameter values closely clustered around the median value, with a notable exception for the combination [300,0.001], where estimates for α_x have a large spread. The D1 model had a similarly low bias as the D0 model and median estimates were close to parameters used to generate the

data. Precision of the D0 model was high, with the 25th and 75th percentile generally close to each other, except for one combination: [30, 0.1, 0.01], where the 75th percentile of the estimate for α_x was 1.

For the empirical dataset, the D1 model had a much higher likelihood than the D0 model, for all censuses. After penalizing the likelihood for added complexity and calculating the corresponding AIC score and AIC weights, the D1 model was convincingly selected.

For the six BCI censuses, we found parameter estimates using both the D0 and the D1 model (Table 2, Table A1). The D0 model has been shown to have two competing optima (Etienne et al., 2006), one with a high value for θ and low value for α and another with a low value for θ and a high value for α . For the D1 model we found two competing optima as well. One of the two optima combines high diversity with high dispersal limitation and typically has a high θ value (~ 200), combined with low α values (~ 0.005 and ~ 0.0008), whilst the other optimum combines low diversity with low dispersal limitation and has a lower θ value (~ 53) combined with one extreme α value (of 1.0) and one much lower α value (~ 0.0006). Table 2 only shows the most likely optima for each model, parameter estimates for all optima can be found in Table A1.

For three out of six BCI censuses, a low diversity and low dispersal limitation (for one guild) optimum is favored over the high diversity, high dispersal limitation optimum. In all censuses, however, the dispersal ability of the guild that relies on biotic dispersal is much higher than the dispersal ability of the guild that relies on abiotic dispersal.

Considering both the ability of our sampling formula to detect or reject guild structure in artificial datasets and the large differences in dispersal limitation parameters between guilds, we conclude that using our guild sampling formula we have convincingly detected guild structure based on dispersal syndromes for all six BCI censuses.

We performed the ‘exact’ test of neutrality as described in the methods section, to estimate whether the observed data is the result of a guild structured process, or whether perhaps the observed data is the result of a different process. All optima for the D1 model have non-significant p -values for the ‘exact’ test of neutrality (Table 2), and hence we cannot distinguish patterns in these communities from those generated with our model with two guilds. All optima for the D0 model have significant p -values, except for the census of 1990, which has p -values of 0.05 and 0.06, which are barely non-significant. This is in contrast with previous comparisons between the BCI data and the neutral model, where using summary statistics Jabot and Chave (2011) were unable to reject neutrality. Hence, our direct use of the likelihood provides more statistical power. The combined results for the D0 and D1 models therefore strongly suggest that guild structure is an important aspect of the empirical data.

Randomization tests revealed that for all six datasets, randomization removed any signal of guild structure (Fig. 1). AIC weight was higher for the D0 model than for the D1 model for all 100 replicates for censuses 1985, 1990, 1995, 2000 and 2005. The 1982 census had 90 out of 100 replicates for which the AIC weight of the D0 model was larger than the AIC weight of the D1 model, retaining a guild signal after randomization in 10% of the replicates. Together these results indicate that the subdivision based on ecological data conveys more information than a random subdivision in guilds does.

For all datasets we plotted the empirical species abundance distribution versus the expected abundance distribution under the Maximum Likelihood Estimates (Fig. 2). For all datasets we observe that the D0 model tends to underestimate abundances for guild X, whilst overestimating abundances for guild Y. It appears that, in an attempt to fit best to both guilds, neither of

Table 1

Bias and precision of the maximum-likelihood estimates as shown by the median and the 25th and 75th percentiles of the estimated parameter values of 100 simulated data sets per parameter combination.

Parameters used to generate data				Estimated parameter values (25th, 50 and 75th percentiles)								
Model	θ	α_X	α_Y	θ			α_X			α_Y		
D0	30	0.100		28.49	30.07	35.24	0.06219	0.09033	0.13310			
	30	0.010		24.66	29.94	36.17	0.00679	0.00960	0.01674			
	30	0.001		10.81	21.70	31.83	0.00097	0.00148	0.00691			
	100	0.100		93.89	101.93	111.64	0.06711	0.09420	0.12661			
	100	0.010		80.80	104.43	126.31	0.00784	0.00955	0.01369			
	100	0.001		40.39	61.69	106.27	0.00102	0.00127	0.00163			
	300	0.100		277.82	313.17	351.06	0.07192	0.09331	0.12084			
	300	0.010		230.08	283.26	355.46	0.00909	0.01029	0.01224			
	300	0.001		90.20	173.06	255.91	0.25069	0.50046	0.75023			
	D1	30	0.100	0.010	28.64	33.29	39.75	0.02107	0.07899	1.00000	0.0046	0.0079
30		0.100	0.001	27.57	31.15	36.15	0.04524	0.08278	0.15556	0.0007	0.0009	0.0012
30		0.010	0.001	23.86	29.23	39.47	0.00565	0.01010	0.03525	0.0007	0.0009	0.0012
100		0.100	0.010	87.01	98.99	113.95	0.05214	0.11171	0.25117	0.0066	0.0076	0.0109
100		0.100	0.001	95.00	104.36	128.20	0.04673	0.08621	0.11942	0.0008	0.0010	0.0012
100		0.010	0.001	73.29	90.36	121.90	0.00760	0.01087	0.01855	0.0008	0.0010	0.0011
300		0.100	0.010	270.00	308.83	365.12	0.06318	0.09173	0.14091	0.0084	0.0093	0.0104
300		0.100	0.001	273.45	304.43	370.33	0.07156	0.09956	0.13416	0.0008	0.0010	0.0011
300		0.010	0.001	242.87	287.32	340.56	0.00890	0.01012	0.01261	0.0009	0.0010	0.0011

Table 2

Parameter estimates for six different censuses of Barro Colorado Island. The D0 model does not take into account differences in dispersal between the guilds, the D1 model does take these differences into account. Guild X represents tree species with biotic dispersal, and guild Y represents tree species with abiotic dispersal. The *p*-value of the 'exact' test of neutrality is reported in the last column.

Census	General statistics			S_X	S_Y	Model	Parameter estimates			Model fit		ΔAIC	AIC W	P-value
	J	J_X	J_Y				θ	α_X	α_Y	LL	AIC			
1982	20914	18321	2593	196	46	D0	33.66	0.0515	0.0515	-398.24	870.55	93.35	0	0.02
						D1	255.64	0.0047	0.0008	-369.79	777.2	0	1	0.58
1985	20742	18203	2539	197	44	D0	32.83	0.0569	0.0569	-400.20	874.77	97.76	0	0
						D1	285.24	0.0045	0.0008	-369.71	777.01	0	1	0.28
1990	21253	18641	2612	189	42	D0	31.10	0.0541	0.0541	-392.60	858.1	86.65	0	0.05
						D1	53.45	1.0000	0.0006	-366.14	771.45	0	1	0.87
1995	21460	18812	2648	188	41	D0	30.90	0.0521	0.0521	-403.60	879.48	87.99	0	0
						D1	53.19	1.0000	0.0005	-376.52	791.49	0	1	0.66
2000	21205	18607	2598	187	42	D0	30.44	0.0567	0.0567	-392.25	856.69	84.23	0	0.04
						D1	52.86	1.0000	0.0006	-366.63	772.46	0	1	0.93
2005	20860	18321	2539	188	43	D0	30.25	0.0630	0.0630	-390.61	853.23	88.97	0	0.04
						D1	232.02	0.0046	0.0008	-363.42	764.26	0	1	0.56

them is fitted well, which explains the poor performance of the D0 model on our empirical data. For the D1 model we observe that for both the high diversity, high dispersal limitation optima (1982, 1985, 200) and the low diversity, low dispersal limitation optima (1990, 1995, 2000) expected abundance distributions closely match the empirical data (Fig. 2). For guild Y both optima show similar patterns, whereas for the larger guild X, the high diversity, high dispersal limitation optima tend to expect a higher number of rare species than the low diversity, low dispersal limitation optimum.

9. Discussion

In this paper we have presented a novel sampling formula that extends the neutral model to a non-neutral setting of two guilds with different dispersal modes. The purpose of our sampling

formula is two-fold: (1) to assess whether a subdivision into two guilds, based on ecological information regarding dispersal, amounts to a statistically significant difference in community structure and (2) to illustrate how to determine, for empirical data sets, to what extent the two guilds differ in their dispersal ability.

Using simulated data we have shown that our sampling formula can detect guild structure from data generated including guild structure and reject guild structure when guild structure was not imposed on the simulated data. Furthermore, the simulation results showed parameter estimates obtained using our sampling formula to be unbiased (i.e. close to the parameters used to generate the data), and precision of our parameter estimates was generally high (but see below) (i.e. spread in parameter estimates was low). Our new guilds sampling formula allowed us to conclude that for all six censuses of tropical tree communities in BCI, Panama, inclusion of guild structure was favoured and tree

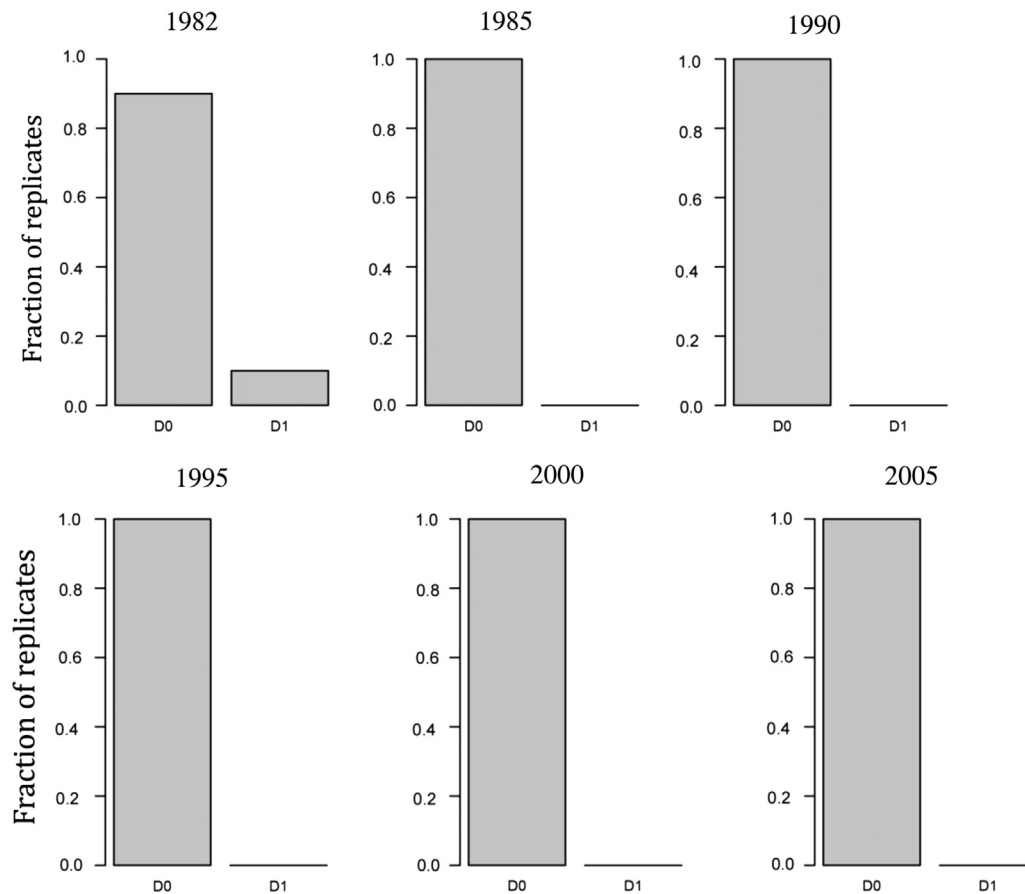


Fig. 1. Fraction of replicates assigned to each model (D0 or D1) after randomizing the BCI datasets by randomly assigning species to a guild. The number of replicates is 100 for every census.

species relying on biotic dispersal tend to be statistically and biologically significantly less dispersal limited than tree species relying on abiotic dispersal.

It has been suggested that the species abundance distribution contains insufficient information to distinguish between competing models (Cohen, 1968; McGill, 2003; McGill et al., 2006, 2007; Ricklefs, 2006), and that additional data are needed to test the validity of community assembly models, for instance in the form of phylogenetic diversity or spatial abundance patterns (Jabot and Chave, 2009; McGill et al., 2006). Here we show that we can distinguish between competing models, using the species abundance distribution combined with ecological information about dispersal mode. It should be noted however that we can only distinguish between competing models when we keep speciation rates between guilds equal in the metacommunity, i.e., $\theta_x = \theta_y$. When allowing for differences in speciation between guilds by leaving both parameters free to be optimized independently, preliminary analyses suggested that we can no longer distinguish between competing models, due to a lack of information in the data. Furthermore, including information on dispersal and guild structure does not resolve the multiple optima problem of the Etienne Sampling Formula. The Etienne Sampling Formula can potentially yield multiple optima with similar likelihood values. Situated at opposite ends of the parameter continuum, one optimum is typically associated with a high value for θ and a low value for I (or m) and the other optimum with a low value for θ and a high value for I (or m). Additional information about the local community tends to favour one of these two optima. Jabot and Chave (2009) combined abundance data and phylogenetic data within an approximate Bayesian framework and recovered only one optimum, with high θ and low I . In another approach,

Etienne (2007) combined information on multiple local communities to obtain estimates for the neutral model and also found only a single optimum. Here we have included information on guild structure, based on ecological information about dispersal, and recover two competing optima. However, one of these optima seems to be a mathematical abnormality, which is situated at the very limit of parameter space. Parameter estimates for this optimum reflect limited diversity, but extremely low dispersal limitation for one guild ($\alpha = 1$) and high dispersal limitation for the other guild ($\alpha < 0.001$). Expected abundance distributions for these extreme dispersal optima seem to reflect the empirical abundance distributions well, although the ecological interpretation of unlimited dispersal remains problematic.

Parameter estimates for the Tropical Tree datasets from BCI suggest high values for θ (average value of 215.42) and low values for α (average values of 0.0050 and 0.00077 for the biotic and abiotic dispersing guilds respectively, ignoring the optima with extreme α values), implying that the tropical tree ecosystem in BCI is highly diverse and fairly dispersal limited. This agrees with previously obtained estimates (Jabot and Chave, 2009). The guild that relies on biotic dispersal (e.g. through birds, bats and mammals) consistently has a 5.75 ($sd = 0.125$) times higher estimated dispersal ability (again ignoring the optima with extreme α values), than the guild relying on abiotic dispersal, such as dispersal through ballistics, gravity, wind and water. This is a much smaller difference than previously estimated (Thomson et al., 2011), where a more than 100-fold difference was measured in dispersal distance between abiotic and biotic dispersers. By contrasting, focusing more on tropical trees, Muller-Landau et al. (2008) were unable to detect significant differences in dispersal distance between abiotic and biotic dispersers. Dispersal ability does however not only capture dispersal

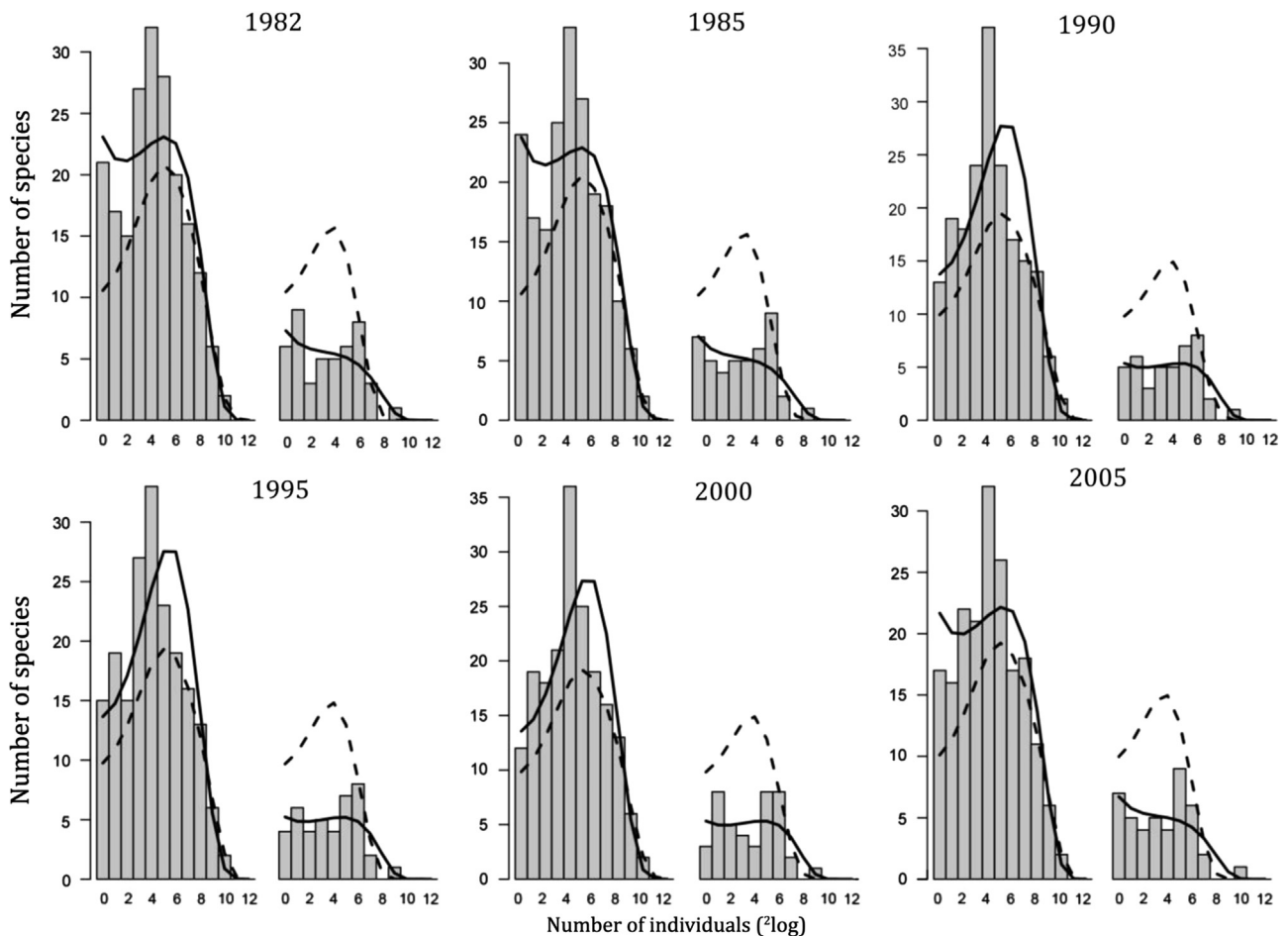


Fig. 2. Empirical (grey bars) and expected (solid curve: D1 model, dashed curve: D0 model) species abundance plots for all BCI censuses. Histograms on the left hand side represent the biotically dispersing guild, histograms on the right hand side the abiotically dispersing guild. Abundances are binned in $^2\log$ bins.

distance, but also establishment and recruitment, which might account for these discrepancies. Furthermore our estimates are in line with clustering patterns (Seidler and Plotkin, 2006), and previous estimates of migration of plots of tropical trees where plots containing a higher relative proportion of mammal-dispersed trees were found to have higher migration (Jabot et al., 2008). Although wind dispersed trees could disperse over long distances, the tight canopy of tropical forests restricts air movement and generally abiotically dispersed trees tend to disperse over shorter distances than animal dispersed trees (Seidler and Plotkin, 2006; Beaudrot et al., 2013). The difference in dispersal ability between the two guilds as inferred by our model thus stresses that although empirical differences in dispersal distance might be negligible for tropical tree species (Muller-Landau et al., 2008), the combined effect of dispersal distance, recruitment and establishment is not, and should be taken into account in future studies, empirical or theoretical.

Our current subdivision in guilds has lumped together trees with fairly different modes of dispersal; we have for instance lumped tree species dispersed by birds as well as tree species dispersed by small mammals in the same guild (biotically dispersed). We expect however that these differences will be less important than the differences between guilds, also because previous estimates of spatial aggregation have shown that within guild differences are smaller than between guild differences (Seidler and Plotkin, 2006). Extending the sampling formula towards more than two guilds is fairly straightforward, but it remains questionable whether this will yield additional understanding of the system. We expect that a larger total sample size is needed to reveal differences in dispersal ability with an increased number of guilds.

An important question that automatically arises when looking at guild structured data is whether the suggested dichotomy introduces more information and structure to the data than a random subdivision into two guilds would. This would quantify the importance of including guild structure in the analysis of community assembly. In our analysis we have tried to approach this question by randomly assigning species to a guild, and assessing which model best explains the (now randomized) data. We found that after randomization the signal of guild structure was almost always lost. This implies that the differences in dispersal limitation we found are not a coincidence and that our method is robust, that is, it is able to detect guild structure even in the presence of other factors that always influence real communities, but not predict guild structure if such structure is absent. Furthermore, randomizing the data requires making an a priori choice about how to divide species over guilds (either 50/50 or some other distribution). An alternative to randomly assigning species to different guilds would be to randomly assign individuals to different guilds (whilst keeping the total number of individuals per guild constant). The number of individuals and number of species are tightly linked however, and it appears non-trivial how to correctly assign species to the randomized individuals without assigning the same species label to individuals in both guilds. Ultimately, validating the guild structure lies not so much in finding a randomization that can test the added value of the imposed guild structure, but in validating the ecological causes that determine why species belong to different guilds.

In a recent paper, Humphreys and Barraclough (2014) also considered a metacommunity divided into multiple “guilds”, and

studied the effect of differences in dispersal. Dispersal in their model is not defined as dispersal between a local community and a metacommunity, but rather defined as the connectivity between the two guilds – dispersal represents here the probability of a species from one guild to disperse towards the other guild. This would be analogous to a speciation event of a species from one guild speciating into a species from another guild in our model. Humpreys and Barraclough focus on the emergence of higher Evolutionary Significant Units (hESU's) as the result of a lack of dispersal between guilds and show that when the exchange of species between guilds is low, this leads to a clear phylogenetic pattern, where both guilds cluster into two distinct clades separated by long external branches. In our sampling formula we have chosen not to focus on speciation dynamics in the metacommunity in favour of unravelling the effects of differences in dispersal. It would be very interesting to look into a guild structured model where both within-guild speciation (e.g. an individual of guild X speciates into a new species belonging to guild X) and between-guild speciation (e.g. an individual of guild X speciates into a new species belonging to guild Y) is modelled. However, because this would introduce at least three new parameters to estimate, we doubt whether such a large number of parameters can be accurately estimated using species abundance data and information on guild structure; perhaps this requires the inclusion of additional information about phylogeny (Jabot and Chave, 2009).

Our sampling formula resembles the multiple samples sampling formula presented by Etienne (2007). That sampling formula considers multiple local communities with independent migration, which all share the same metacommunity (with one single estimate for θ). If we interpret these different local communities as different guilds, the multiple samples model closely resembles our multiple guilds model. An important difference, however, is hidden in the metacommunity structure. The multiple samples metacommunity consists of one single metacommunity, without any structure. Our multiple guilds metacommunity is explicitly structured such that there are two separate guilds in the metacommunity that have independent dispersal towards the local community. Due to their independent dispersal, guild sizes and number of species in the local community can differ from each other, whereas the linked local communities from the multiple samples model all sample from the same species pool.

Our guilds sampling formula disentangles migration, dispersal ability, and metacommunity abundance. In classical neutral theory, dispersal limitation between the local and metacommunity is governed by one single parameter, m (migration) (Hubbell, 2001). This can be interpreted as the combined effects of dispersal, recruitment and establishment. In our sampling formula we defined migration analogously, but here the dependence on the relative size of the guild in the metacommunity becomes explicit: $m_i = \alpha_i p_i$ (in the neutral case there was only one guild with $p=1$) Our newly defined dispersal ability α still includes dispersal, recruitment and establishment. Because we have redefined the migration parameter, and have focused on estimating α , estimates of our model cannot be directly compared with previously obtained estimates of immigration (Etienne, 2007; Jabot and Chave, 2009). Inferences with our D0 model however, provide a good reference point, as this model assumes no guild structure and reduces to the Etienne Sampling Formula (Etienne, 2005) with the migration parameter substituted by our new dispersal parameter and the relative metacommunity abundance.

In our model we assumed independence of migration and speciation ability. We assumed that in the metacommunity there are no differences between guilds with respect to speciation and have only focused on differences of migration between the metacommunity and local community. If there are profound differences in dispersal ability between guilds, however, we would expect this to

also influence the probability of speciation. This is a general problem of two-scale neutral models (Leigh, 2007). From empirical data it becomes clear that a lack of dispersal tends to lead to more patchy distributions (Seidler and Plotkin, 2006) and can thus facilitate geographical isolation of populations. As a result we expect an interaction between speciation and dispersal ability, such that either low dispersal is associated with high speciation rates due to the patchy distribution of the population, or high dispersal is associated with high speciation, as populations come in contact with novel environments more often. The exact relationship between dispersal and speciation will depend on the life-history of species. Furthermore, correctly implementing this interaction would require extending our current model towards a spatially explicit form. Rosindell and Phillimore (2011) took a first step towards a further integration of dispersal and speciation by identifying the difference between in situ speciation on an island (cladogenesis) and speciation through drift over time, where an immigrant on an island diverges from its ancestor on the mainland (anagenesis). Future work could focus on a more direct connection between dispersal and speciation and could provide a more explicit link between spatially explicit processes driving both dispersal limitation and speciation.

Our new sampling formula incorporates ecological reality into a neutral approach of community assembly. It enhances our understanding and appreciation of the interplay between stochasticity, dispersal and species-specific requirements that govern the patterns we observe in ecological communities and the underlying processes of community assembly.

Acknowledgements

We thank Joe Wright, Helene Muller-Landau and Denise Hardisty for help with categorizing BCI tree species according to dispersal mode. Financial support for B.H. was provided by the TULIP Laboratory of Excellence (ANR-10-LABX-41). T.J. and R.S.E. thank the Netherlands Organization for Scientific Research (NWO) (number: VIDI project 864.07.007) for financial support through VIDI and VICI grants awarded to R.S.E. We thank the Van Gogh programme (Project number: 23195PD) for providing financial support. Computer code for this work has been made available as the R package "GUILDS".

Appendix A. Deriving the sampling formula

A.1. One-guild sampling formula

First, we sketch a derivation of the one-guild sampling formula. We start by deriving the abundance distribution of the local community in Hubbell's neutral community model. Community size J is fixed. Individuals die at a constant rate and are replaced with probability $1-m$ by offspring from within the community, or with probability m by an immigrant from outside the community. For the time being, we assume the composition of the pool of immigrants (that is, the metacommunity) to be fixed. We denote the relative abundances in the metacommunity by $p_1, p_2 \dots p_s$ and the absolute abundances in the community by $N_1, N_2 \dots N_s$. Note that

$$\sum_{i=1}^s p_i = 1 \quad \text{and} \quad \sum_{i=1}^s N_i = J$$

Then, using the fundamental dispersal number $I = m(J-1)/(1-m)$, the stationary distribution of the community abundances is

$$P(\vec{N} | \vec{p}, I, J) = \frac{J!}{(I)_J} \frac{(I p_1)_{N_1} \dots (I p_s)_{N_s}}{N_1! \dots N_s!} \quad (S1)$$

where the Pochhammer symbol $(x)_y$ is defined as

$$(x)_y = \prod_{i=1}^y (x+i-1)$$

Eq. (S1) describes the abundance distribution in the local community. It can also be used to obtain the abundance distribution of the metacommunity. To do so, the migration probability m must be interpreted as the speciation probability ν , the metacommunity size is denoted by J_M and the fundamental dispersal number l is replaced by the fundamental biodiversity number $\theta = \nu(J_M - 1) / 1 - \nu$. The abundance distribution of the metacommunity is used to integrate out the relative abundances p_1, \dots, p_S in the abundance distribution of the local community.

Etienne (2005) used Eq. (S1) to derive the sampling formula of Hubbell's neutral model, that is, the probability that a sample taken from Hubbell's neutral community has abundance vector D . The sample is taken from the local community, described by parameters I and J , while the metacommunity abundance distribution is described by parameter θ . The sampling formula is given by

$$P(\vec{D} | \theta, I, J) = \frac{J!}{\prod_{i=1}^S J_i! \prod_{j=1}^J (S_j)!} \frac{(\theta)^S}{(I)_J} \sum_{A=S}^J \left(K(\vec{D}, A) \frac{(I^A)}{(\theta)^A} \right) \quad (S2)$$

where \vec{D} is a vector of the number of individuals per species, S is the number of species in vector \vec{D} and J is the total number of individuals in vector \vec{D} . $K(\vec{D}, A)$ is defined as follows:

$$K(\vec{D}, A) := \sum_{\{a_1, \dots, a_S | \sum_{i=1}^S a_i = A\}} \prod_{i=1}^S \frac{\bar{s}(n_i, a_i) \bar{s}(a_i, 1)}{\bar{s}(n_i, 1)}$$

where n_i is the number of individuals of species i and $\bar{s}(n_i, a_i)$ is the unsigned Stirling number of the first kind.

A.2. Two-guild sampling formula

Next, we derive the two-guilds sampling formula. We start by deriving the abundance distribution of the local community. Species belong to one of two guilds X and Y with different dispersal ability α_X and α_Y . Total community size J is a fixed parameter, but guild sizes J_X and J_Y are dynamic variables. The species relative abundances in the metacommunity are $p_{X,i}$ ($i=1, \dots, S_X$) for guild X and $p_{Y,i}$ ($i=1, \dots, S_Y$) for guild Y . For the time being, we assume the relative abundances of the metacommunity to be fixed. The guild relative abundances are

$$p_X = \sum_{i=1}^{S_X} p_{X,i} \text{ and } p_Y = \sum_{i=1}^{S_Y} p_{Y,i} \text{ with } p_X + p_Y = 1$$

We denote the local community abundances by $N_{X,i}$ ($i=1, \dots, S_X$) for guild X and by $N_{Y,i}$ ($i=1, \dots, S_Y$) for guild Y so that

$$J_X = \sum_{i=1}^{S_X} N_{X,i} \text{ and } J_Y = \sum_{i=1}^{S_Y} N_{Y,i} \text{ with } J_X + J_Y = J$$

As in the case of a single guild, dead individuals are replaced with probability $1 - m$ by local offspring and with probability m by immigration. In contrast to the case of a single guild, the immigration probability of a specific species is not only determined by its metacommunity abundance, but also by the guild it belongs to. In particular,

Immigration by species i of guild X has probability $\alpha_X p_{X,i}$

Immigration by species i of guild Y has probability $\alpha_Y p_{Y,i}$ so that

$$m = \sum_{i=1}^{S_X} \alpha_X p_{X,i} + \sum_{i=1}^{S_Y} \alpha_Y p_{Y,i} = \alpha_X p_X + \alpha_Y p_Y$$

We use Eq. (S1) to compute the stationary community

composition. To do so we construct a virtual metacommunity with relative abundances.

Species i of guild X has relative abundance $\alpha_X p_{X,i} / (\alpha_X p_X + \alpha_Y p_Y)$

Species i of guild Y has relative abundance $\alpha_Y p_{Y,i} / (\alpha_X p_X + \alpha_Y p_Y)$

We then consider neutral immigration from this virtual metacommunity. The total immigration probability is equal to $\alpha_X p_X + \alpha_Y p_Y$. Explicitly,

Immigration by species i of guild X has probability

$$m \frac{\alpha_X p_{X,i}}{\alpha_X p_X + \alpha_Y p_Y} = (\alpha_X p_X + \alpha_Y p_Y) \frac{\alpha_X p_{X,i}}{\alpha_X p_X + \alpha_Y p_Y} = \alpha_X p_{X,i}$$

Immigration by species i of guild Y has probability

$$m \frac{\alpha_Y p_{Y,i}}{\alpha_X p_X + \alpha_Y p_Y} = (\alpha_X p_X + \alpha_Y p_Y) \frac{\alpha_Y p_{Y,i}}{\alpha_X p_X + \alpha_Y p_Y} = \alpha_Y p_{Y,i}$$

Comparing these immigration probabilities with the previous ones, we see that neutral immigration from this virtual metacommunity is equivalent with the original immigration process. Therefore, we can apply Eq. (S1) to the virtual metacommunity to obtain the abundance distribution of the two-guild local community,

$$P(\vec{N}_X, \vec{N}_Y | \vec{p}_X, \vec{p}_Y, \alpha_X, \alpha_Y, I, J) = \frac{J!}{(I)_J} \frac{(I(\alpha_X p_{X,1} / (\alpha_X p_X + \alpha_Y p_Y)))_{N_{X,1}} \dots (I(\alpha_X p_{X,S_X} / (\alpha_X p_X + \alpha_Y p_Y)))_{N_{X,S_X}}}{N_{X,1}! \dots N_{X,S_X}!} \times \frac{(I(\alpha_Y p_{Y,1} / (\alpha_X p_X + \alpha_Y p_Y)))_{N_{Y,1}} \dots (I(\alpha_Y p_{Y,S_Y} / (\alpha_X p_X + \alpha_Y p_Y)))_{N_{Y,S_Y}}}{N_{Y,1}! \dots N_{Y,S_Y}!} \quad (S3)$$

Introducing the guild fundamental dispersal numbers

$$I_X = \frac{\alpha_X p_X}{\alpha_X p_X + \alpha_Y p_Y} J = \frac{\alpha_X p_X (J-1)}{1 - \alpha_X p_X - \alpha_Y p_Y} \quad (S4A)$$

$$I_Y = \frac{\alpha_Y p_Y}{\alpha_X p_X + \alpha_Y p_Y} J = \frac{\alpha_Y p_Y (J-1)}{1 - \alpha_X p_X - \alpha_Y p_Y} \quad (S4B)$$

we get

$$P(\vec{N}_X, \vec{N}_Y | \vec{p}_X, \vec{p}_Y, \alpha_X, \alpha_Y, I, J) = \frac{J!}{(I)_J} \frac{(I_X (p_{X,1} / p_X))_{N_{X,1}} \dots (I_X (p_{X,S_X} / p_X))_{N_{X,S_X}}}{N_{X,1}! \dots N_{X,S_X}!} \times \frac{(I_Y (p_{Y,1} / p_Y))_{N_{Y,1}} \dots (I_Y (p_{Y,S_Y} / p_Y))_{N_{Y,S_Y}}}{N_{Y,1}! \dots N_{Y,S_Y}!} = \frac{J!}{(I)_J} \frac{(I_X)_{J_X} (I_Y)_{J_Y}}{J_X! J_Y!} \times \frac{J_X!}{(I_X)_{J_X}} \frac{(I_X (p_{X,1} / p_X))_{N_{X,1}} \dots (I_X (p_{X,S_X} / p_X))_{N_{X,S_X}}}{N_{X,1}! \dots N_{X,S_X}!} \times \frac{J_Y!}{(I_Y)_{J_Y}} \frac{(I_Y (p_{Y,1} / p_Y))_{N_{Y,1}} \dots (I_Y (p_{Y,S_Y} / p_Y))_{N_{Y,S_Y}}}{N_{Y,1}! \dots N_{Y,S_Y}!}$$

In the last equality, the two last lines give the abundance distribution of the species belonging to guild X and guild Y , respectively. Both are instances of the one-guild formula (S1). Hence, the remaining factors (first line of the last equality) give the probability distribution of the guild sizes J_X and J_Y ,

$$P(J_X, J_Y | I_X, I_Y, J) \stackrel{\text{As a result,}}{=} \frac{J!}{(I_X + I_Y) J} \frac{(I_X)_{J_X} (I_Y)_{J_Y}}{J_X! J_Y!} \quad (S6)$$

$$P(\vec{N}_X, \vec{N}_Y | \vec{p}_X, \vec{p}_Y, \alpha_X, \alpha_Y, I, J) = P(J_X, J_Y | I_X, I_Y, J) P\left(\vec{N}_X | \frac{\vec{p}_X}{p_X}, I_X, J_X\right) P\left(\vec{N}_Y | \frac{\vec{p}_Y}{p_Y}, I_Y, J_Y\right) \quad (S7)$$

The product structure shows that, given guild sizes J_X and J_Y and

parameters I_X and I_Y , the abundance distributions of guilds X and Y are independent.

Eq. (S7) describes the abundance distribution in the local community. In particular, the product structure of Eq. (S7) carries over to the sampling formula for dispersal-limited sample from the local community for given values of the metacommunity relative abundances of the two guilds,

$$P(\bar{D}_X, \bar{D}_Y | \theta, I_X, I_Y, J) = P(J_X, J_Y | I_X, I_Y, J) P(\bar{D}_X | \theta_X, I_X, J_X) P(\bar{D}_Y | \theta_Y, I_Y, J_Y) \quad (S8)$$

which is Eq. (2) in the main text.

To obtain the full sampling formula, we proceed as in the derivation of the one-guild sampling formula. We integrate over all possible values of the relative abundances of the two guilds, weighted by the probability distribution of the guilds' relative abundances in the metacommunity. This probability distribution is obtained by lifting Eq. (S7) from the local community to the metacommunity. To do so we make the substitutions $J_X \rightarrow J_{M,X}$, $J_Y \rightarrow J_{M,Y}$, $J \rightarrow J_M$, $I_X \rightarrow \theta_X$ and $I_Y \rightarrow \theta_Y$, analogously to the case of a single guild. As a result, given the relative abundances p_X and p_Y of guilds X and Y , the metacommunity abundance distributions of the guilds are independent. To compute the distribution of p_X and p_Y we have

$$P(J_{M,X}, J_{M,Y} | \theta_X, \theta_Y, J_M) = \frac{J_M!}{(\theta_X + \theta_Y)_{J_M}} \frac{(\theta_X)_{J_{M,X}} (\theta_Y)_{J_{M,Y}}}{J_{M,X}! J_{M,Y}!}$$

or

$$P(J_{M,X} | \theta_X, \theta_Y, J_M) = \frac{J_M!}{(\theta_X + \theta_Y)_{J_M}} \frac{(\theta_X)_{J_{M,X}} (\theta_Y)_{J_M - J_{M,X}}}{J_{M,X}! (J_M - J_{M,X})!} \quad (S9)$$

Then we take the coupled limit $J_{M,X} \rightarrow \infty$ and $J_M \rightarrow \infty$ with $J_{M,X} = p_X J_M$. That is, we transform the discrete probability distribution $P(J_{M,X})$ of absolute abundances to a continuous distribution $\rho(p_X)$ of relative abundances. We have

$$P(J_{M,X} = \alpha J_M) = P\left(\frac{\alpha J_M}{J_M} \leq p_X < \frac{\alpha J_M + 1}{J_M}\right) = \int_{\alpha}^{\alpha + (1/J_M)} \rho(p_X) dp_X \approx \rho(\alpha) \frac{1}{J_M}$$

such that

$$\rho(p_X | \theta_X, \theta_Y, J_M) = \lim_{J_M \rightarrow \infty} J_M P(p_X J_M | \theta_X, \theta_Y, J_M)$$

$$\begin{aligned} &= \lim_{J_M \rightarrow \infty} J_M \frac{J_M!}{(\theta_X + \theta_Y)_{J_M}} \frac{(\theta_X)_{p_X J_M} (\theta_Y)_{(1-p_X)J_M}}{(p_X J_M)! ((1-p_X)J_M)!} \\ &= \frac{\Gamma(\theta_X + \theta_Y)}{\Gamma(\theta_X)\Gamma(\theta_Y)} \lim_{J_M \rightarrow \infty} J_M \frac{J_M!}{\Gamma(J_M + \theta_X + \theta_Y)} \frac{\Gamma(p_X J_M + \theta_X)}{(p_X J_M)!} \frac{\Gamma((1-p_X)J_M + \theta_Y)}{(1-p_X J_M)!} \end{aligned}$$

Now we apply three times the formula,

$$\lim_{L \rightarrow \infty} \frac{\Gamma(L + \alpha)}{L^{\alpha-1} L!} = 1$$

which can be proved using Stirling's approximation. Hence,

$$\begin{aligned} \rho(p_X | \theta_X, \theta_Y, J_M) &= \frac{\Gamma(\theta_X + \theta_Y)}{\Gamma(\theta_X)\Gamma(\theta_Y)} \lim_{J_M \rightarrow \infty} J_M \frac{1}{J_M^{\theta_X + \theta_Y - 1}} (p_X J_M)^{\theta_X - 1} ((1-p_X)J_M)^{\theta_Y - 1} \\ &= \frac{\Gamma(\theta_X + \theta_Y)}{\Gamma(\theta_X)\Gamma(\theta_Y)} p_X^{\theta_X - 1} (1-p_X)^{\theta_Y - 1} \lim_{J_M \rightarrow \infty} J_M \frac{J_M^{\theta_X - 1} J_M^{\theta_Y - 1}}{J_M^{\theta_X + \theta_Y - 1}} \\ &= \frac{\Gamma(\theta_X + \theta_Y)}{\Gamma(\theta_X)\Gamma(\theta_Y)} p_X^{\theta_X - 1} (1-p_X)^{\theta_Y - 1} \quad (S10) \end{aligned}$$

Eq. (S10) holds for a speciation process in which a fraction $\theta_X/(\theta_X + \theta_Y)$ of speciation events gives rise to species of guild X and a fraction $\theta_Y/(\theta_X + \theta_Y)$ of speciation events gives rise to species of guild Y . In the main text we make the additional assumption that these fractions are equal, that is, $\theta_X = \theta_Y = \theta/2$. With this assumption, Eq. (S10) reduces to Eq. (4) in the main text.

Combining Eqs. (S8) and (S10), and integrating over all possible values of p_X gives us the full sampling formula (Eq. (5) in the main text):

$$P(\bar{D}_X, \bar{D}_Y | \theta, \alpha_X, \alpha_Y) = \int_0^1 P(J_X, J_Y | I_X, I_Y, J) P(\bar{D}_X | \theta_X, I_X, J_X) P(\bar{D}_Y | \theta_Y, I_Y, J_Y) \rho(p_X | \theta) dp_X \quad (S11)$$

A.3. Conditioning on guild size

Guild sizes J_X and J_Y are central to Eq. (S11) and we therefore expect that differences in guild size might disproportionately affect parameter estimates. To see to which extent parameter estimates are influenced by differences in guild size, rather than differences in dispersal ability between guilds, we plotted the ratio of guild sizes versus the ratio of dispersal ability estimates obtained using Eq. (S8) (we used parameter estimates obtained using the procedure described under the ‘‘model selection’’ part of the methods

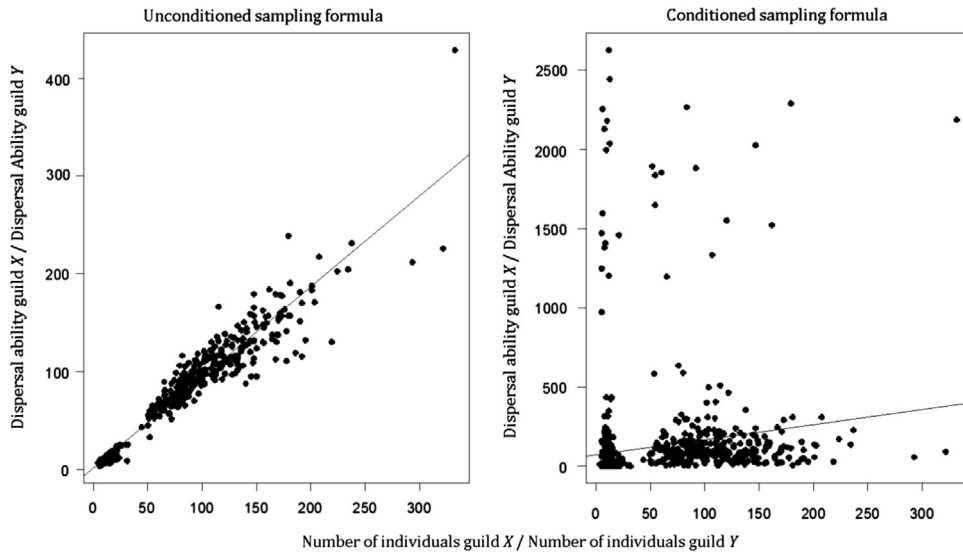


Fig. A1. Ratio of dispersal abilities of both guilds, versus the ratio of individuals in these guilds. Left hand plot shows ratios obtained using the unconditioned sampling formula, right hand plot shows ratios obtained using the conditioned sampling formula. For the unconditioned sampling formula, a clear relation is found between the ratio of dispersal abilities and ratio of guild sizes ($R^2=0.95$, slope=0.927, $p < 2e-16$). Using the conditioned sampling formula, this relationship vanishes however ($R^2=0.00276$, $p=0.115$).

Table A1

Parameter estimates for the six different censuses of BCI, including both found optima for the D0 model. Guild X represents tree species with biotic dispersal, and guild Y represents tree species with abiotic dispersal. The *p*-value of the Neutrality test is reported in the last column.

Census	General statistics					Model	Parameter estimates			Model fit				<i>p</i> -value
	<i>J</i>	<i>J_X</i>	<i>J_Y</i>	<i>S_X</i>	<i>S_Y</i>		θ	α_X	α_Y	<i>LL</i>	<i>AIC</i>	ΔAIC	<i>AIC W</i>	
1982	20,914	18,321	2593	196	46	D0	111.04	0.0041	0.0041	-398.65	801.31	55.72	0	0.00
1982						D0	33.663	0.0515	0.0515	-398.24	800.48	54.89	0	0.02
1982						D1	255.64	0.0047	0.0008	-369.79	745.58	0	0.87	0.58
1982						D1	55.77	1.0000	0.0006	-371.70	749.39	3.81	0.13	0.91
1985	20,742	18,203	2539	197	44	D0	32.83	0.0569	0.0569	-400.20	804.41	59.00	0	0.00
1985						D0	115.51	0.0040	0.0040	-400.70	805.40	59.98	0	0.00
1985						D1	56.00	1.0000	0.0006	-372.36	750.72	5.31	0.07	0.90
1985						D1	285.24	0.0045	0.0008	-369.71	745.41	0	0.93	0.28
1990	21,253	18,641	2612	189	42	D0	31.10	0.0541	0.0541	-392.60	789.21	50.93	0	0.05
1990						D0	82.52	0.0049	0.0049	-395.84	795.67	57.39	0	0.06
1990						D1	53.45	1.0000	0.0006	-366.14	738.28	0	0.72	0.87
1990						D1	198.29	0.0050	0.0008	-367.06	740.12	1.84	0.28	0.59
1995	21,460	18,812	2648	188	41	D0	30.90	0.0521	0.0521	-403.60	811.19	52.16	0	0.00
1995						D0	81.68	0.0048	0.0048	-406.48	816.97	57.94	0	0.00
1995						D1	53.19	1.0000	0.0005	-376.51	759.03	0	0.64	0.66
1995						D1	197.21	0.0049	0.0007	-377.08	760.15	1.12	0.36	0.17
2000	21,205	18,607	2598	187	42	D0	30.44	0.0567	0.0567	-392.25	788.50	49.25	0	0.04
2000						D0	81.04	0.0050	0.0050	-395.68	795.35	56.11	0	0.03
2000						D1	52.86	1.0000	0.0006	-366.60	739.25	0	0.72	0.93
2000						D1	189.58	0.0051	0.0008	-367.59	741.18	1.92	0.28	0.53
2005	20,860	18,321	2539	188	43	D0	30.25	0.0630	0.0630	-390.61	785.22	52.38	0	0.04
2005						D0	104.84	0.0040	0.0040	-392.17	788.34	55.50	0	0.01
2005						D1	53.32	1.0000	0.0006	-365.05	736.11	3.26	0.16	0.97
2005						D1	232.01	0.0046	0.0008	-363.42	732.85	0	0.84	0.48

section). Only the parameter estimate of the model with the highest AIC weight was taken into consideration. We found a positive relation between the ratio of dispersal abilities of both guilds and the ratio between guild sizes ($R^2=0.95$, slope=0.927, $p < 2e-16$, Fig. A1). It seems thus, that Eq. (S8) overly emphasizes the impact of differences in guild size and negates any differences in the abundance distributions of the two guilds. A next step would be to condition our sampling formula on guild size.

To condition on guild size, we can use the likelihood of having guilds of size J_X and J_Y (Eq. (S6)):

$$P(J_X, J_Y | I_X, I_Y, J) = \frac{J! (I_X)_{J_X} (I_Y)_{J_Y}}{(I_X + I_Y)_J J_X! J_Y!}$$

Taking into account all possible combinations of I_X and I_Y , and remembering from that I_X and I_Y depend on p_X we obtain:

$$P(J_X, J_Y | \theta, I_X, I_Y) = \int_0^1 \frac{J! (I_X)_{J_X} (I_Y)_{J_Y}}{(I_X + I_Y)_J J_X! J_Y!} \rho(p_X | \theta) dp_X \tag{S12}$$

Using Eq. (S12), we condition Eq. (S11) and obtain:

$$P(\bar{D}_X, \bar{D}_Y | \theta, I_X, I_Y, J_X, J_Y) = \frac{\int_0^1 P(J_X, J_Y) P(\bar{D}_X | \theta, I_X, J_X) P(\bar{D}_Y | \theta, I_Y, J_Y) \rho(p_X | \theta) dp_X}{\int_0^1 (J! / (I_X + I_Y)_J) ((I_X)_{J_X} (I_Y)_{J_Y} / J_X! J_Y!) \rho(p_X | \theta) dp_X} \tag{S13}$$

Using Eq. (S13), we repeated the procedure to obtain the ratio between guild sizes and the ratio between dispersal abilities. This time, no significant correlation between the ratio of dispersal abilities and the ratio of guild sizes was detected anymore ($R^2=0.00276$, $p=0.115$, Fig. A1).

References

Adler, P.B., Hillerislambers, J., Levine, J.M., 2007. A niche for neutrality. *Ecol. Lett.* 10, 95–104.

Akaike, H., 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Control* 19 (6).

Almany, G.R., Berumen, M.L., Thorrold, S.R., Planes, S., Jones, G.P., 2007. Local replenishment of coral reef fish populations in a marine reserve. *Science* 316, 742–744.

Alonso, D., Etienne, R.S., McKane, A.J., 2006. The merits of neutral theory. *Trends Ecol. Evol.* 21, 451–457.

Beaudrot, L., Rejmánek, M., Marshall, A.J., 2013. Dispersal modes affect tropical forest assembly across trophic levels. *Ecography* 36, 984–993.

Chase, J., Leibold, M., 2003. *Ecological Niches: Linking Classical and Contemporary Approaches*. The University Press, Chicago, USA.

Cohen, J., 1968. Alternate derivations of a species-abundance relation. *Am. Nat.* 102, 165.

Condit, R., Hubbell, S., LaFrankie, J., 1996. Species-area and species-individual relationships for tropical trees: a comparison of three 50-ha plots. *J. Ecol.* 84, 549–562.

Condit, R., Pitman, N., Leigh, E.G., Chave, J., Terborgh, J., Foster, R.B., Núñez, P., Aguilar, S., Valencia, R., Villa, G., Muller-Landau, H.C., Losos, E., Hubbell, S.P., 2002. Beta-diversity in tropical forest trees. *Science* 295, 666–669.

Du, X., Zhou, S., Etienne, R.S., 2011. Negative density dependence can offset the effect of species competitive asymmetry: a niche-based mechanism for neutral-like patterns. *J. Theor. Biol.* 278, 127–134.

Efron, B., Tibshirani, R., 1993. *An Introduction to the Bootstrap*. Chapman & Hall, New York.

Etienne, R.S., 2007. A neutral sampling formula for multiple samples and an “exact” test of neutrality. *Ecol. Lett.* 10, 608–618.

Etienne, R.S., 2005. A new sampling formula for neutral biodiversity. *Ecol. Lett.* 8, 253–260.

Etienne, R.S., Alonso, D., 2005. A dispersal-limited sampling theory for species and alleles. *Ecol. Lett.* 8, 1147–1156.

Etienne, R.S., Latimer, A.M., Silander, J.A., Cowling, R.M., 2006. Comment on “Neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot”. *Science* 311, 610.

Etienne, R.S., Olf, H., 2004. A novel genealogical approach to neutral biodiversity theory. *Ecol. Lett.* 7, 170–175.

Ewens, W., 1972. The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* 112, 87–112.

Gotelli, N.J., Anderson, M.J., Arita, H.T., Chao, A., Colwell, R.K., Connolly, S.R., Currie, D.J., Dunn, R.R., Graves, G.R., Green, J.L., Grytnes, J.-A., Jian, Y.-H., Jetz, W., Lyons,

- S.K., McCain, C.M., Magurran, A.E., Rahbek, C., Rangel, T.F.L.V.B., Soberón, J., Webb, C.O., Willig, M.R., 2009. Patterns and causes of species richness: a general simulation model for macroecology. *Ecol. Lett.* 12, 873–886.
- Haegeman, B., Loreau, M., 2011. A mathematical synthesis of niche and neutral theories in community ecology. *J. Theor. Biol.* 269, 150–165.
- Hubbell, S.P., 2001. The unified neutral theory of biodiversity and biogeography. (MPB-32) Vol. 32, Princeton University Press.
- Humphreys, A.M., Barraclough, T.G., 2014. The evolutionary reality of higher taxa in mammals. *Proc. R. Soc. B: Biol. Sci.* 281, 20132750.
- Hutchinson, G.E., 1957. Concluding remark. Cold spring harbor symposia on Quant. Bio. 22 (2), 415–427.
- Jabot, F., Chave, J., 2011. Analyzing tropical forest tree species abundance distributions using a nonneutral model and through approximate Bayesian inference. *Am. Nat.* 178, E37–E47.
- Jabot, F., Chave, J., 2009. Inferring the parameters of the neutral theory of biodiversity using phylogenetic information and implications for tropical forests. *Ecol. Lett.* 12, 239–248.
- Jabot, F., Etienne, R., Chave, J., 2008. Reconciling neutral community models and environmental filtering: theory and an empirical test. *Oikos* 117, 1308–1320.
- Janzen, T., 2014. GUILDS in R: Implementation of sampling formulas for the unified neutral model of biodiversity and biogeography, with or without guild structure. Available from: (<http://cran.r-project.org/web/packages/GUILDS/index.html>).
- Leigh, E.G., 2007. Neutral theory: a historical perspective. *J. Evol. Biol.* 20, 2075–2091.
- Liu, J., Zhou, S., 2011. Asymmetry in species regional dispersal ability and the neutral theory. *PLoS One* 6, e24128.
- McGill, B.J., 2003. A test of the unified neutral theory of biodiversity. *Nature* 422, 881–885.
- McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., Dornelas, M., Enquist, B.J., Green, J.L., He, F., Hurlbert, A.H., Magurran, A.E., Marquet, P.A., Maurer, B. a, Ostling, A., Soykan, C.U., Ugland, K.I., White, E.P., 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.* 10, 995–1015.
- McGill, B.J., Maurer, B.A., Weiser, M.D., 2006. Empirical evaluation of neutral theory. *Ecology* 87, 1411–1423.
- McInerney, G.J., Etienne, R.S., 2012. Ditch the niche – is the niche a useful concept in ecology or species distribution modelling? (S. Higgins, Ed.). *J. Biogeogr.* 39, 2096–2102.
- Muller-Landau, H.C., Hardesty, B.D., 2005. Seed dispersal of woody plants in tropical forests: Concepts, examples, and future directions. In: Burslem, D.F.R. P., Pinard, M.A., Hartley, S. (Eds.), *Biotic interactions in the tropics*. Cambridge University Press, Cambridge, p. 580.
- Muller-Landau, H.C., Wright, S.J., Calderón, O., Condit, R., Hubbell, S.P., 2008. Interspecific variation in primary seed dispersal in a tropical forest. *J. Ecol.* 96, 653–667.
- Noble, A., Fagan, W., 2011. A unification of niche and neutral theories quantifies the impact of competition on extinction. arXiv preprint arXiv 1102.0052.
- Pigolotti, S., Cencini, M., 2013. Species abundances and lifetimes: from neutral to niche-stabilized communities. *J. Theor. Biol.* 338C, 1–8.
- Purves, D., Pacala, S., 2008. Predictive models of forest dynamics. *Science* 320, 1452–1453.
- Purves, D.W., Turnbull, L.A., 2010. Different but equal: the implausible assumption at the heart of neutral theory. *J. Anim. Ecol.* 79, 1215–1225.
- Ricklefs, R.E., 2006. The unified neutral theory of biodiversity: do the numbers add up? *Ecology* 87, 1424–1431.
- Rosindell, J., Hubbell, S., Etienne, R., 2011. The unified neutral theory of biodiversity and biogeography at age ten. *Trends Ecol. Evol.* 26, 340–348.
- Rosindell, J., Phillimore, A.B., 2011. A unified model of island biogeography sheds light on the zone of radiation. *Ecol. Lett.* 14, 552–560.
- Seidler, T.G., Plotkin, J.B., 2006. Seed dispersal and spatial pattern in tropical trees. *PLoS Biol.* 4, e344.
- Thomson, F.J., Moles, A.T., Auld, T.D., Kingsford, R.T., 2011. Seed dispersal distance is more strongly correlated with plant height than with seed mass. *J. Ecol.* 99, 1299–1307.
- Turnbull, L.A., Rees, M., Purves, D.W., 2008. Why equalising trade-offs aren't always neutral. *Ecol. Lett.* 11, 1037–1046.
- Valtonen, A., Molleman, F., Chapman, C., 2013. Tropical phenology: bi-annual rhythms and interannual variation in an Afrotropical butterfly assemblage. *Ecosphere* 4, 1–28.
- Victor, B., Wellington, G., 2000. Endemism and the pelagic larval duration of reef fishes in the eastern Pacific Ocean. *Mar. Ecol. Prog. Ser.* 205, 241–248.
- Volkov, I., Banavar, J.R., Hubbell, S.P., Maritan, A., 2003. Neutral theory and relative species abundance in ecology. *Nature* 424, 1035–1037.
- Wagenmakers, E.-J., Farrell, S., 2004. AIC model selection using Akaike weights. *Psychon. Bull. Rev.* 11, 192–196.
- Wennekes, P.L., Rosindell, J., Etienne, R.S., 2012. The neutral-niche debate: a philosophical perspective. *Acta Biotheor.* 60, 257–271.