



Reconciling neutral community models and environmental filtering: theory and an empirical test

Franck Jabot, Rampal S. Etienne and Jérôme Chave

F. Jabot and J. Chave (chave@cict.fr), Laboratoire Evolution et Diversité Biologique CNRS, Univ. Paul Sabatier Bâtiment 4R3, FR-31062 Toulouse cedex 4, France. FJ also at: AgroParisTech-Ecole Nationale du Génie Rural, des Eaux et des Forêts, 19 avenue du Maine, FR-75015 Paris, France. – R. S. Etienne, Community and Conservation Ecology Group, Univ. of Groningen, PO Box 14, NL-9750 AA Haren, the Netherlands.

It is widely believed that the neutral theory of biodiversity cannot be used for parameter inference if the assumption of neutrality is not met. The goal of this work is to extend this neutral framework to quantify the intensity of recruitment limitation (limited dispersal plus environmental filtering) in natural species assemblages. We model several local communities as part of a larger metacommunity, and we assume that neutrality holds in each local community, but not in the metacommunity. The immigration rate m does not only reflect dispersal limitation into a given local community, but also the intensity of environmental filtering. We develop a novel statistical method to infer the immigration parameter m in each local community. Using simulated datasets, we show that m indeed depends on both dispersal limitation and on the intensity of environmental filtering. We then apply this method to a network of tropical tree plots in central Panama. Inferred recruitment rates m were positively correlated with the fraction of trees dispersed by mammals, and with annual rainfall, possibly due to a weaker environmental filtering as rainfall increases. Finally, m , as estimated from trees greater than 1 cm trunk diameter, were significantly larger than an estimation based on trees greater than 10 cm trunk diameter. This suggests a cumulative effect of environmental filtering upon trees throughout their ontogeny.

In recent years, Hubbell's (2001) neutral theory of biodiversity has generated a remarkable amount of interest and controversy in ecology (Bell 2001, Chave 2004, Alonso et al. 2006, Holyoak and Loreau 2006, McGill et al. 2006). This theory assumes that all individuals in a community have the same prospects of birth and death, irrespective of the species they belong to, and that they also have the same limited dispersal ability, being more likely to disperse to neighbouring areas than to remote places. Dispersal limitation into a local community may be measured as an aggregated immigration rate m , the probability that a newly established individual is an immigrant from outside the local community. When m is small, the local community receive few propagules from outside, and it is therefore a 'dispersal-limited' sample of the regional species pool (Etienne 2005, Etienne and Alonso 2005), or 'metacommunity' (Chase et al. 2005). The second parameter of the neutral theory measures the species diversity of the metacommunity, a function of the metacommunity size J_M and of the rate of appearance of new species (speciation rate ν), both summarized by the parameter $\theta = \nu(J_M - 1)/(1 - \nu)$.

One of the great achievements of the neutral theory has been to relate community-wide processes of species coexistence and biogeographical processes, thus providing a simple tractable model of biodiversity across spatial scales (Hubbell 2001, Chave 2004, Leibold and McPeck 2006). If

dispersal limitation is an explanatory factor for the assembly of ecological communities, the similarity in species composition should decline across samples as inter-plot distance increases (Condit et al. 2002). An alternative explanation for spatial structure across communities is that the environment alone dictates where species may or may not establish (Wright 2002, Tuomisto et al. 2003, Condit et al. 2004). Many studies have sought to test the validity of the neutral assumption using species abundance data and ordination methods, and species were found not to occur randomly across samples, suggesting that environmental filtering is a dominant mechanism in the assembly of many ecological communities (Tuomisto et al. 2003, Cottenie 2005, Jones et al. 2006, John et al. 2007). However, these studies do not permit to construct dynamic community models that incorporate both dispersal limitation and environmental filtering. As a result, it is difficult to integrate the predictions of the neutral theory with those of the niche theory for spatial patterns of species abundance. This limitation has swamped recent discussions on the relative merit of the neutral theory as compared with alternative theories (Leibold and McPeck 2006, McGill et al. 2006).

As currently developed, the neutral theory of biodiversity overlooks the fact that post-dispersal environmental filtering is a crucial community assembly process (Ricklefs 1987, Hurtt and Pacala 1995, Nathan and Muller-Landau 2000).

Here, we propose that, under certain assumptions, the major features of the neutral theory hold when the immigration parameter m is not simply a measure of dispersal limitation, but also accounts for establishment limitation into the local community. As a result, we can no longer assume that species are equivalent at the regional scale, as sites may have different establishment constraints, hence species may be more likely to be absent in some sites than in others. However, we still assume a neutral dynamics at the local community scale. After the filtering stage, once an individual is established, it competes neutrally with its neighbours. If an environment filters out many species, the local species composition will represent poorly the meta-community (Fig. 1). Consequently, the parameter m , inferred from a local community, is expected to decrease as environmental filtering increases. Hence, m should be interpreted as a measure of recruitment limitation, that is, the compound influence of dispersal limitation and post-dispersal environmental filtering (Hurtt and Pacala 1995, Nathan and Muller-Landau 2000), and not of dispersal limitation alone.

One of the great advantages of the neutral theory is that it is associated with an exact sampling theory. Etienne (2005) found an exact mathematical expression for the likelihood function associated to Hubbell's dispersal-limited neutral theory, making it possible to estimate the parameters m and θ based on species abundance datasets. This approach also facilitates inter-model comparisons (Etienne and Olff 2005, Chave et al. 2006). An exact likelihood formula for the multi-samples equivalent of Hubbell's model was also recently obtained (Etienne 2007), and this allows in principle to estimate the immigration parameters m_i , i in $\{1, \dots, D\}$ from D samples belonging to the same metacommunity. However, Etienne's (2007) formula is only computationally tractable if the immigration parameters are the same among samples, thus reducing its usefulness. We here provide a tractable approximation of

Etienne's (2007) theory for jointly estimating the immigration parameters m_i , including cases where neutrality is not met at the regional scale. We use simulations to assess whether our new method of inference allows us to track the variation in both dispersal limitation and environmental filtering among communities.

Then, applying this method to a network of permanent tree plots in Panama, we investigate whether variations in our inferred recruitment rates may help assess the magnitude of non-neutral effects. In a purely neutral interpretation, variation in m may be explained by dispersal limitation created by geographical barriers, or by a lack of dispersers in part of the area. However, variation in m may also be due to non-neutral effects, for instance differential dispersal abilities of the tree species, or environmental filtering acting differentially upon species. We test whether the variation of m correlates with the dispersal syndromes of the trees in the community. If such a correlation is observed, then m effectively captures the intensity of dispersal limitation averaged over the coexisting species in the community. We then test whether environmental variables predict the variation in m across samples. If a significant correlation is observed, then the environment should contribute to determining the level of recruitment limitation into the local community. For instance, more productive environments may favour the presence of dispersers, hence lead to an increase in m . Environmental filtering may act upon individuals either independently of the species they belong to, or differentially across species (Fig. 1, Paoli et al. 2006, Comita et al. 2007). In the former case, the estimation of m should not change during the establishment phase of trees. Thus the minimal size of the individuals included in the census should not affect m , as long as only canopy tree species are considered. On the contrary, if environmental filtering differs among species, some species are expected to be filtered out of the community during their establishment phase (Fig. 1).

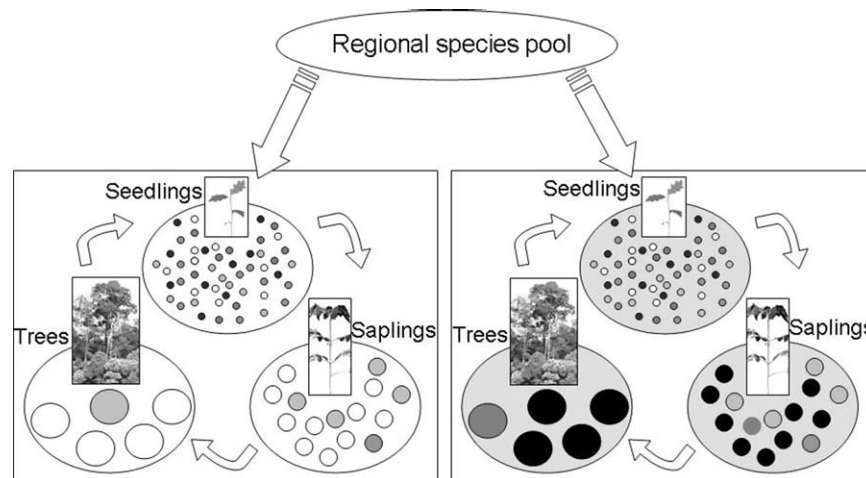


Figure 1. The influence of environmental filtering on tree species throughout the ontogeny. The two panels depict two communities with different environmental constraints, symbolized by a white tone on the left and a dark grey tone on the right, but with the same regional species pool. Each species is represented by a shade of grey, and it is assumed that light-grey species prefer light-grey environments. Newly emerged seedlings (top) have not been submitted to important environmental filters such as nutrient limitation, and the seedling communities therefore are almost the same in the two sites. As the plants grow up, white and light-grey species tend to be favoured in the white environment, while black and dark-grey species are favoured in the dark-grey environment. The neutral theory would have predicted that the community of adult tree is a random sample of the seedling community.

Thus, increasing the diameter threshold for censused trees should lead to increasingly filtered communities. Consequently the community of large trees should have a smaller m than a random and equal-sized sample of communities including smaller trees.

Material and methods

Metacommunity model

We define a metacommunity by a set of D neutral communities, each of size C_i , i in $\{1, \dots, D\}$, and connected through migration. The metacommunity has a total size of $J_M = \sum_{i=1}^D C_i$. We consider a field sampling protocol consisting in collecting a single sample in each local community. Thus, the dataset has D samples of size J_i , i in $\{1, \dots, D\}$, with, for all i , $J_i \ll C_i$. The relative species abundances in the regional pool are summarized by the vector $X \equiv x_j$, j in $\{1, \dots, S\}$, where x_j is the regional relative abundance of species j . In sample i , the abundance of species j is denoted n_{ij} , and the community matrix may be denoted by $N \equiv \{n_{ij}\}$, i in $\{1, \dots, D\}$, j in $\{1, \dots, S\}$. Our definition of a local community depends on the sampling scheme, rather than on the environmental characteristics of the ecological community, as two different local communities may be experiencing the same environmental conditions. With our definition, each of the D communities may be defined as a sample of the metacommunity (Etienne and Alonso 2005), limited by both dispersal and by the habitat characteristics of the site. Importantly, by using the full regional species abundance distribution X instead of θ to describe the metacommunity, we do not assume that the metacommunity is itself neutral, as it is a sum of local communities with potentially different environmental filters.

The probability that an individual establishes in community i having been produced outside of this community is defined as m_i , and we shall write $M \equiv \{m_i\}$, i in $\{1, \dots, D\}$. A size-unbiased version of m_i is $I_i = m_i (J_i - 1) / (1 - m_i)$ (Etienne 2005). As shown in Fig. 2, I is however slightly biased for small sample sizes (Tavaré and Ewens 1997). In the empirical datasets studied, we had samples with as few as $J_{\min} = 249$ individuals. We therefore used a rarefaction method to estimate a comparable value of m across samples. We randomly drew $J_{\min} = 249$ individuals 100 times in all the samples and computed an average of m_i across these draws.

Statistical inference

The inference of the parameters m_i from several samples embedded in the same metacommunity is based on the maximization of a likelihood function of the inferred parameters given empirical data. A simple method for computing the m_i in a similar situation was proposed by Munoz et al. (2007). It consists in computing first θ from the dispersal-unlimited neutral sampling formula in the pooled samples (Ewens 1972), and then using the Etienne (2005) sampling formula independently for each local sample, constraining the empirical θ to be equal to the value estimated in the first step of the procedure. This

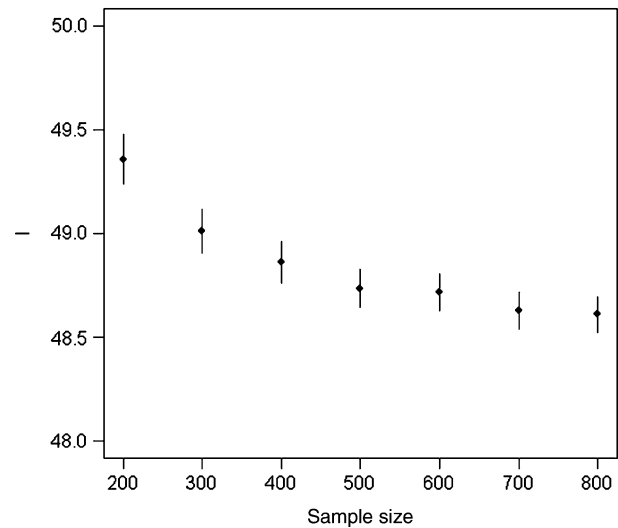


Figure 2. Sampling bias in the estimation of the immigration rate: mean of the estimated I over 10 000 simulated samples for samples varying from $J = 200$ to $J = 800$, using Etienne's (2005) formula. Samples were simulated assuming $\theta = 92$ and $I = 48$. The normalized immigration rate I was then estimated by fixing θ to its value of 92. Vertical bars stand for SE of the means.

assumes that the metacommunity dynamics is neutral and panmictic, since θ is obtained from the neutral sampling distribution of Ewens (1972). We here describe a more general method for jointly estimating the parameters m_i . Our inference method uses regional species abundances as parameters instead of summarizing them by the neutral parameter θ . Based on simulations of a structured metacommunity, we show that our strategy outperforms Munoz et al.'s method.

The sampling formula $L(M, X|N)$ for the model parameters M , and the relative species abundances in the regional pool X , given the data N may be expressed as the product over the D plots of Eq. 25 in Etienne and Alonso (2005), erratum in 2006; Appendix 1):

$$L(M, X|N) = \prod_{i=1}^D \frac{J_i!}{\prod_{j=1}^S n_{ij}!} \frac{\prod_{j=1}^S (I_i x_j)^{n_{ij}}}{(I_i)_{J_i}} \quad (1)$$

In this formula, $(x)_y$ denotes the Pochhammer symbol:

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

In contrast with the Etienne sampling formula, this equation depends on the regional relative species abundances x_j . From this formula, the regional species abundances and the dispersal limitation parameters may be jointly estimated from the local species abundances. Maximizing such a function at once is a serious challenge because the parameter space is high-dimensional (535 dimensions for the empirical dataset presented below). Therefore, we choose to adopt a different strategy. We fixed the regional relative abundances to their observed relative abundances obtained by pooling the D local samples, hence neglecting spatial species clumping. That is, we assumed that, for each species j , $x_j = \sum_{i=1}^D n_{ij} / \sum_{i=1}^D J_i$. A direct

Markov chain Monte Carlo (MCMC) search of the full maximum likelihood written in Eq. 1 suggests that this assumption does not seriously bias the final results (unpubl.). For the number of plots we used in our application ($D = 42$), the regional abundances converge to a constant value (Appendix 2). Would the regional pool be unequally sampled, the inferred m_i would be biased towards higher values for oversampled communities. Our approach assumes that the Panama sampling scheme is not biased towards a particular habitat and/or a geographical area. Once the regional abundances are fixed, the maximization of Eq. 1 is simply made by an optimization method (a simplex method in our implementation).

Computations were performed using the freeware tetame (<<http://www.edb.ups-tlse.fr/equipe1/chave/tetame>>). All additional programs are available in the Supplementary material.

Test with simulated data

We assessed the efficiency of our inference method on neutral simulated datasets. We simulated two spatially explicit neutral communities placed one besides the other, connected by migration, and experiencing different intensities of dispersal limitation. We constructed the datasets using the simulation approach of Chave et al. (2002). In each community, we assumed that a single tree occupies a cell in a grid of size $J = 512 \times 256$, and the two communities occupy $J_M = 512 \times 512$ cells in total, with toroidal boundary conditions (Chave et al. 2002). At each timestep, a random site is vacated, and it is replaced by a propagule chosen at random from the offspring of the neighbouring sites according to a Gaussian dispersal kernel. The variance

σ^2 of the dispersal kernel is fixed at a different value in the two communities (Fig. 3). The left community has a lower σ^2 , and it is therefore more dispersal-limited. Diversity is maintained due to the random appearance of individuals of new species in the metacommunity. We then subdivided the simulated system into quadrats of size 16×16 each, and we estimated m in each quadrat using the inference method developed above. We thus obtained 512 estimates of m for each of the two communities. We computed the means and standard deviations of m along vertical stripes (Fig. 3). In a given vertical stripe, the quadrats are experiencing similar conditions of dispersal limitation, and they are equidistant from the other community.

Second, we assessed whether m was able to track variation in environmental filtering. We simulated a metacommunity where environmental filtering varied across local communities. We defined a one-dimensional 'niche axis' of total length 1, with periodic boundaries (Schwilk and Ackerly 2005). In each local community j , the environmental filter is modelled by a function equal to 1 in the interval $[h_j - f_j, h_j + f_j]$ and to zero elsewhere, where f_j measures the amount of local filtering at a site (when it equals 0.5, there is no more environmental filtering), and h_j measures the position of the habitat of community j on the 'niche axis'. The niche of species i is modelled by a function equal to 1 in the interval $[u_i - w_i, u_i + w_i]$ and to zero elsewhere, where w_i models the niche specialization of species i , and u_i represents the position of species i on the 'niche axis'. Species i can establish in community j only if the two respective segments overlap. In this model, all species able to establish locally have the same local competitive advantage. We simulated a species pool of 1000 species, and drew at random the u_i from $[0,1]$ and

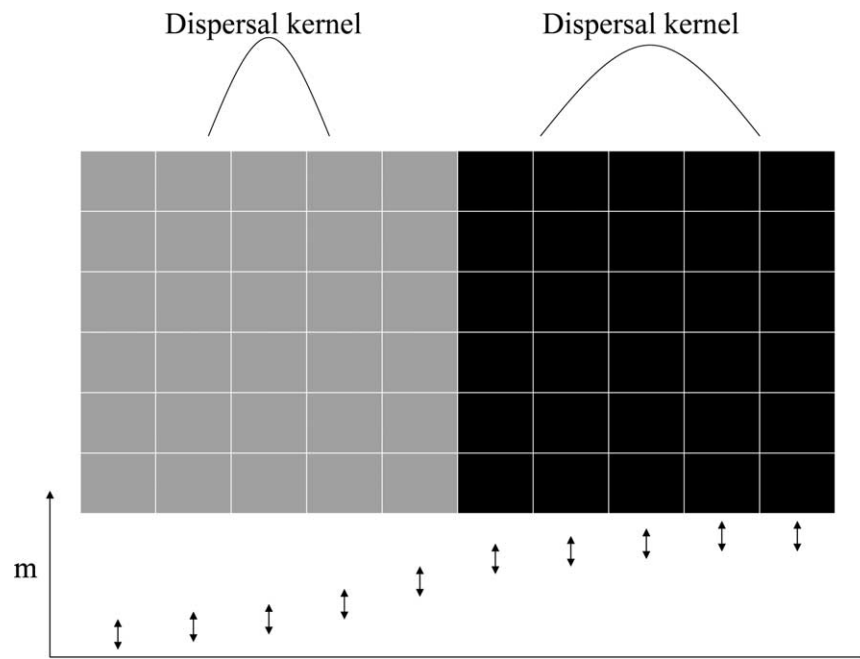


Figure 3. Simulated neutral community. The individuals located in the left community (in grey) are poor dispersers, whereas individuals in the right community (in black) are good dispersers. The simulated value of m was computed in each square subplot of 16×16 individuals, and they were averaged along vertical stripes. As expected (bottom panel), the value of m increases from the left community (low dispersal) towards the right community (high dispersal).

the w_i from $[0, 1/16]$. We then simulated 400 local communities, with h_j randomly drawn from $[0, 1]$, and a log-transformed environmental filtering intensity $\ln(f_j)$ randomly drawn from $[\ln(0.001), \ln(0.5)]$. We drew a sample of 256 individuals from each of the 400 local communities, following Etienne's (2005) algorithm with a dispersal limitation parameter fixed to 0.1, and constraining the immigrants to belong to the species able to establish in this community. Our sampling formula was then used to estimate m in the resulting dataset with 400 samples of 256 individuals. We replicated this analysis five times.

Test with empirical data

We applied our inference method to a published census of tropical forest trees of central Panama (Condit et al. 2002). It comprises one 50 ha plot (Barro Colorado Island, henceforth BCI), and 40 1-ha plots scattered across the Panama Canal Watershed (PCW). All trees over 10 cm in diameter at breast height (dbh) were tagged, mapped, and identified to species, except for 154 trees out of the 22 955 (0.7%) that remained unidentified. We used only two 1-ha subplots out of the 50 ha plot to avoid over-representing the BCI forest. Since the plots are samples of the Panama Canal Watershed forest, with no particular geographical or habitat bias, it is appropriate to apply our method.

Annual rainfall was monitored from weather stations nearby each of the plots, and geological information was collected based on a United States Geological Survey map. The age of the forest was inferred from the size of the largest trees and recorded as young, secondary, or old growth (Pyke et al. 2001). Plot-averaged percentages of trees dispersed ballistically, by wind, by birds or bats, and by mammals was quantified in the same fashion as in Seidler and Plotkin (2006), using the Phenology Database of the Smithsonian Tropical Research Institute (http://striweb.si.edu/esp/tesp/plant_species.htm). We also explored whether plot-averaged log-transformed seed weight, computed from the Seed Information Database of the Royal Botanic Gardens, Kew (Flynn et al. 2006) was a good predictor of m . When seed weight was unavailable at the species level, we used the mean value of the species in the same genus. Overall, 88% of the trees had seed weight information.

Our model assume that the D plots are sufficiently distant to be considered as independent samples of the metacommunity (between-plot migration may be neglected). Some of the study plots are relatively close, thus the dataset may slightly violate this independence assumption. To assess this potential bias, we repeated our analysis on a subset of 16 plots separated by large distances, but found no detectable difference. We tested whether m varied significantly across the metacommunity by comparing the model with variable immigration parameter m_i to the model with constant m , using a likelihood ratio test (Burnham and Anderson 2002). The likelihood of the model with equal immigration is given by Eq. 1 simply by replacing all the I_i by a single parameter I .

We then assessed whether m was related with the environmental conditions and the seed dispersal syndromes of the tree species found in the samples. We categorized the species into four possible dispersal syndromes (ballistic,

wind, birds or bats, and mammals). We performed a multiple linear regression of m_i against the fraction of trees dispersed according to all four dispersal syndromes. We also performed a linear regression of m_i against the plot-averaged seed weight, defined as $\langle \ln(w_j) \rangle$, where w_j is the seed weight of tree j in the sample, and the brackets denote an average over all trees in the sample. Then, we regressed m_i against the environmental data: annual rainfall, soil type, and forest age. The linear regressions above assume that m is known with no measurement error. We then also repeated this analysis by propagating the error in m to the final regression, and found similar results.

To test the importance of environmental filtering, we used a different test and a different dataset. Specifically, we used three plots in the Panama Canal Watershed, for which all saplings and trees over 1 cm in dbh have been tagged, mapped and identified. We used a 4-ha subplot in the BCI plot, the full 4.96-ha Sherman plot on the Atlantic side of the Panama Canal Zone, and the 4-ha Cocoli plot on the Pacific side of the Panama Canal Zone (<http://www.ctfs.si.edu/doc/datasets.html>). These represented a total of 45 325 trees ≥ 1 cm dbh, and 5402 trees ≥ 10 cm dbh. To ensure that, in a given plot, the absence of certain species in the high dbh classes cannot be due to the fact that they never grow up to this size, we excluded all the species that had no tree greater than 10 cm dbh in the total PCW dataset. We computed m_i for each of the three plots, both based on trees ≥ 10 cm dbh, and based on saplings ≥ 1 cm dbh. The sapling dataset was rarefied 100 times to get the same sample sizes for trees and for saplings. For the BCI plot, where the dbh of each tree was available in the publicly available database, we also computed m_i based on trees greater than a threshold dbh, from 2 to 9 cm dbh, in increments of 1 cm.

Results

We first analyzed the data obtained from the neutral simulation with variable dispersal limitation intensities. Unsurprisingly, using the estimation of both parameters θ and m based on Etienne (2005) on each sample independently, we failed to detect a variation in m (Fig. 4A). Using Eq. 1, we found that the inferred variations in m were consistent with the dispersal parameters used in the simulation (Fig. 4C). We also compared our method with Munoz et al.'s two-step inference method (Fig. 4B). The estimation of m based on Eq. 1 led to a variance in m 29% lower than with the two-step method i.e. $\text{variance}(\text{intra-stripe}) / \text{variance}(\text{inter-stripe})$. Therefore, in the remainder of this work, we only report the results obtained with our method.

Then we turned to the simulated community with environmental filtering. We found that m , as inferred from simulated samples, was negatively correlated with the intensity of environmental filtering (Fig. 5, mean $R^2 = 0.66$ across five replicates). As environmental filtering decreased, the inferred m tended towards the value of dispersal limitation used in the simulations (equal to 0.1, dotted line in Fig. 5). Together these results show that our inference method permits to infer variation both in environmental filtering and in dispersal limitation.

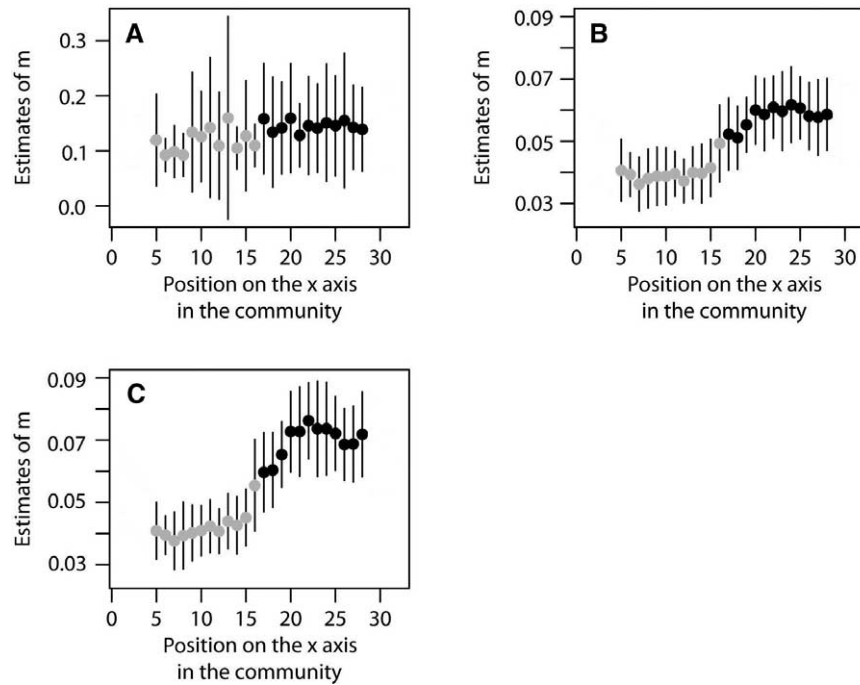


Figure 4. Detection of the variation in recruitment limitation in simulated neutral communities (Fig. 2). Panel (A): inference based on the Etienne (2005) sampling formula. Panel (B): inference based on a two-step method where θ is computed first, and then the m parameters (Munoz et al. 2007). Panel (C): inference based on the present work (Eq. 1). The vertical bars stand for standard deviations of m among subplots in the same vertical stripe (Fig. 2). The last two methods detect the variation in dispersal limitation in the simulations.

In the Panama tree dataset, m varied between 0.05 and 0.3 across samples. This variation in the intensity of recruitment limitation was significant. A test based on the difference in log-likelihoods between the variable- m model and the constant- m model showed that the variable- m model clearly outperforms the constant- m model ($\chi^2_{41} = 248.9$, $p < 0.001$).

The value of m was positively correlated with the percentage of mammal-dispersed trees ($n = 42$, $R^2 = 0.18$, $p = 0.009$, Fig. 6). However, m did not correlate significantly

with the community-wide log-transformed seed weight, or with the fraction of trees dispersed ballistically, by wind, and by birds or bats ($p > 0.05$). The parameter m was found to be significantly smaller in young forests than in old-growth forests ($R^2 = 0.33$, $p = 0.003$), and it was also significantly higher in high-rainfall forests ($R^2 = 0.2$, $p = 0.003$). We did not detect any influence of soil type on the value of m . To account for the potential effect of correlations between precipitation and of the percentage of mammal-dispersed trees in the plots, we also performed a multiple linear regression including both independent variables. Both variables turned out to be significant ($p = 0.03$ for both variables). The model including an interaction term between rainfall and the percentage of mammal-dispersed species did not perform significantly better ($p = 0.053$).

For all three plots for which sapling data were available (BCI, Sherman and Cocoli), m as inferred from trees over 10 cm dbh was significantly lower than the value inferred from all trees and saplings over 1 cm in dbh (Fig. 7B–D). Likewise, we observed a decrease in m with increasing lower cut-off dbh for the BCI plot (Fig. 7A). The slight increase of m for trees of dbh greater than 2 and 3 cm in BCI was only marginally significant.

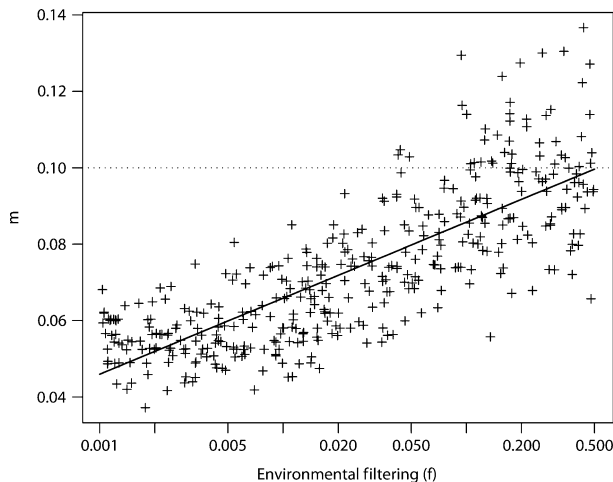


Figure 5. Influence of environmental filtering on the value of m in a simulated non-neutral metacommunity. The dotted line represents the value of the dispersal limitation used in the simulation. By definition (Methods), environmental filtering is more intense for lower values of f , and reaches zero as $f = 0.5$.

Discussion

In sessile organisms, dispersal and establishment limitation both play a central role in explaining patterns of species distribution and abundance (Hurt and Pacala 1995, Wang and Smith 2002, Levin et al. 2003). Communities limited by dispersal tend to have more spatially clumped species, and spatial turnover in diversity would be predicted to

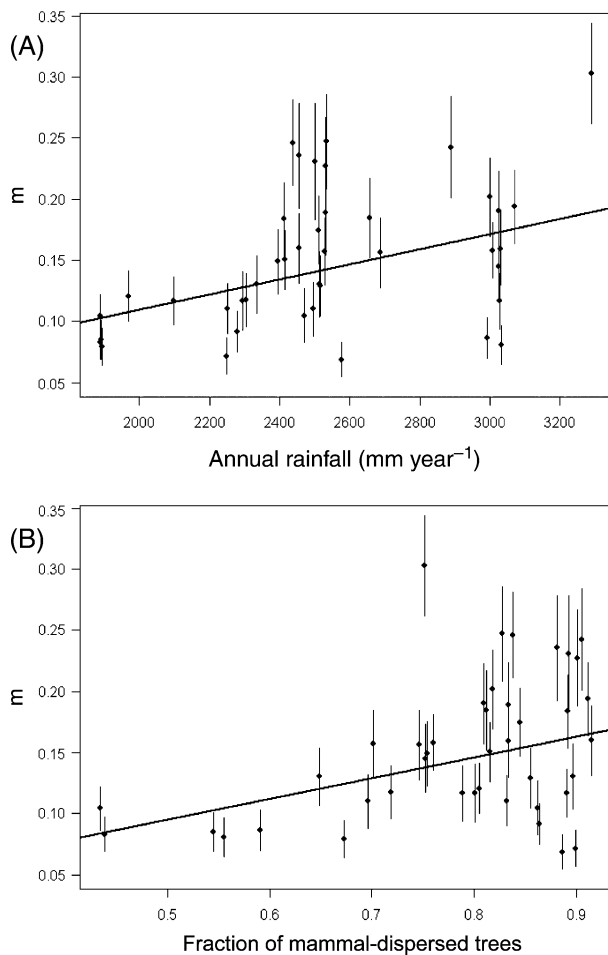


Figure 6. (A): regression of inferred values of m against annual rainfall (mm year^{-1}) across 42 plots of the Panama Canal zone ($R^2 = 0.20$, $p = 0.003$). (B): regression of m against the fraction of mammal-dispersed trees in the same 42 plots ($R^2 = 0.18$, $p = 0.009$). Vertical bars stand for standard deviations among the 100 rarefaction draws.

depend mostly on geographical distance (Condit et al. 2002). In contrast, communities limited by establishment are predicted to have ecologically more similar species, and spatial turnover in diversity would depend mostly on environmental factors (Tuomisto et al. 2003). Both processes are critical in a context of global change, since the inability of plants to spread into new suitable habitat may result in a rapid extinction if their habitats shift at a rapid pace (Higgins et al. 2003, Thuiller et al. 2005). Estimating community-wide migration rates from biodiversity surveys is therefore an important challenge.

Assessing the model assumptions

The main goal of this study was to extend the inference framework of the neutral theory of biodiversity to cases where multiple samples of the same metacommunity are available, potentially with different environmental filtering intensities across these samples. Using simulations, we showed that the immigration parameter m , as estimated

by our new statistical inference method, is influenced by both the intensity of environmental filtering into each local community, and by dispersal limitation. Importantly, our generalization of the neutral theory does not relax the hypothesis of functional equivalence of species at the local scale, once these species are established. This assumption allows us to build a quantitative statistical framework, and to analyze community diversity data more powerfully than previous neutral-based methods. Although this is a simplification of the real dynamics, this assumption is a good working hypothesis and a clear improvement over the neutral models. We assume that the establishment phase of a tree lasts until the plant reaches a trunk diameter of 10 cm, and that the subsequent dynamics of adult trees is neutral. Since the most important environmental filters, both biotic and abiotic, are likely to have already acted upon a tree when it reaches 10 cm in diameter (Poorter 2007), assuming a neutral dynamics for trees is not as crude as it would be for seedlings or saplings (Norden et al. 2007).

At the regional scale, our model relaxes the assumption of neutrality. Species may differ in their establishment constraints, and may thus be filtered differentially across the plots. In this case, the regional abundance of species i (X_i) should be seen as an 'effective' measure of regional abundance (Alonso et al. 2006), which depends on the real regional abundance of this species, but also on its niche width.

Robustness of the model to violations of neutrality

In what cases should departures from neutrality alter our results? In our model, species differences in competitive ability, in dispersal or in reproductive rate, are possible so long as all species have the same establishment rates (Chesson 2000, Purves and Pacala 2005). If several species guilds associated to different micro-habitats coexist, then strict neutrality cannot hold. For instance, Sheil et al. (2006) analyzed the distribution of tropical tree species with respect to light exposure, and they showed that the species better exposed to sunlight as saplings also tended to be better exposed as adults (but see Poorter et al. 2005 for a different viewpoint). However, as long as many species share similar strategies, the dynamics resemble a purely neutral one (Purves and Pacala 2005, Schwillk and Ackerly 2005). In our model, the presence of more than one species guild implies a larger number of available micro-sites for establishment, hence a less intense establishment limitation (but see Kadmon and Allouche 2007). Finally, if species are non-neutral even after they have grown past the 10-cm diameter limit, then some species will be superior to others in a given plot, and will increase in abundance. This will lead to a decrease of m in that site, since the plot will be even more different than the rest of the metacommunity. In this case, our interpretation of m as a measure of recruitment limitation would keep being coherent.

Determinants of recruitment limitation in the Panama Canal Watershed

We used our inference method to explore the determinants of community assembly of tropical trees in the Panama

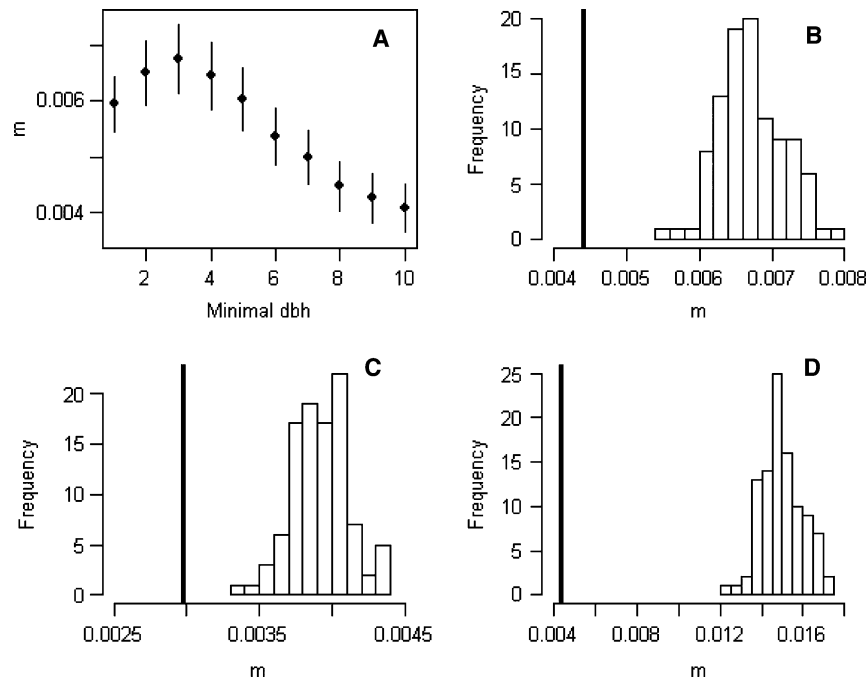


Figure 7. Evidence that environmental filtering is more intense on a community of trees ≥ 10 cm dbh, than on a community of trees ≥ 1 cm dbh. (A): estimates of m on BCI using a minimal threshold dbh ranging from 1 cm to 10 cm. Vertical bars stand for the standard deviations associated to the estimation. (B) – (D): distributions of m estimated for random subsamples of trees ≥ 1 cm dbh ((B): BCI, (C): Sherman, (D): Cocoli). In all three (B) to (D), the vertical lines correspond to the m values estimated based on trees over 10 cm dbh.

Canal watershed, central Panama. We found that our samples were recruitment-limited (m was significantly less than 1), consistent with a previous analysis based on single plots (Etienne 2005, Chave et al. 2006). We also attempted to test whether the interpretation of m as a pure measure of dispersal limitation was supported by our data, and found that it was not. By comparing two different life stages (trees over 1 cm in dbh and trees over 10 cm in dbh), our analysis revealed that the community of large trees was more dissimilar to the rest of the metacommunity than the community of trees also including saplings. We interpret this result as a signature of environmental filtering acting throughout plant ontogeny, which is more easily detected when large trees are singled out than when both saplings and trees are included in the analysis (Webb and Peart 2000, Paoli et al. 2006, Comita et al. 2007). The fact that m is significantly larger in old growth forests than in young forests may indicate that filtering is less intense in old forests that have more microhabitats due to vertical stratification.

We tested alternative explanations for the variation in m among communities. We explored whether this variation may be caused by a variation in seed dispersal syndromes of the species making up the local communities. Animals play a key role in tropical forest dynamics during the phases of plant dispersal and establishment (Janzen 1970, Stevenson 2000, Cordeiro and Howe 2001, Engelbrecht et al. 2007). Recently, Seidler and Plotkin (2006) provided a community-wide analysis of trees in the Pasoh tropical forest, Malaysia, and also at the BCI plot. They found a relationship between seed dispersal syndromes and tree spatial clustering patterns: the least clustered species tended to be dispersed by mammals. Hence, immigration rates should

depend both on the proportion of species belonging to dispersal syndromes types, and on the community of seed dispersers available to actually disperse the seeds. We found that m was positively correlated with the percentage of mammal-dispersed trees in the plots, a result consistent with that of Seidler and Plotkin. The observed decline of seed dispersing animals in the tropics may thus have a massive, though delayed, impact upon plant communities (Muller-Landau 2007). It would be critical to assess this hypothesis directly in the field.

We also observed a significant correlation between rainfall and m . This correlation might be explained either by a variation in the intensity of environmental filtering, or of dispersal limitation across the gradient. If drier areas have a lower primary productivity (Condit et al. 2004), they would support a lower density of seed-dispersing animals by providing them with less food. This may explain the positive and significant correlation between m and annual rainfall. In this case, drier areas would tend to be more dispersal-limited. An alternative hypothesis is that differential environmental filtering explains the correlation between m and annual rainfall. It has often been emphasized that the neutral theory is inconsistent with habitat specialization of the plants making up the communities (Tuomisto et al. 2003). Using both ordination analyses and regressions on distance matrices of floristic similarity, Chust et al. (2006) previously found a significant correlation between among-plot similarity in environmental variables (elevation, annual rainfall, dry season length) and the similarity of tree species composition in the same permanent tree plots we used here. They interpreted this result as an indication that environmental effects could not be

ignored in these communities. The positive correlation we found between m and annual rainfall is also consistent with a signature of a stronger environmental filtering in drier habitats (Engelbrecht et al. 2007). To further test the respective roles of dispersal and environment on recruitment limitation, one would need to quantify the amount and the origin of the seeds of all species migrating into a quadrat, and the corresponding amount of seedlings established into this quadrat. Empirical attempts to quantify the intensity of plant recruitment limitation have so far used seed trap monitoring programs (Harms et al. 2000, Wright et al. 2005), and seedling censuses combined with a quantification of environmental conditions (Webb and Peart 2000, Russo et al. 2005, Comita et al. 2007, Norden et al. 2007). In the future, it will be possible to refine these tests by a direct quantification of gene flow across a broader range of species than the ones being investigated today (Jones et al. 2005, Hardesty et al. 2006, Hardy et al. 2006).

Dynamic models and multivariate statistics

Previous approaches based on canonical analysis and regressions on distance matrices have sought to explore the respective roles of dispersal limitation, and of the environment on the floristic dissimilarity of communities by variation partitioning techniques (Tuomisto et al. 2003, Legendre et al. 2005, Chust et al. 2006, Jones et al. 2006, Pélissier and Coutron 2007). These approaches have led to the general conclusion that both mechanisms play a role, with varying intensities depending on the group of organisms under study (Cottenie 2005, Leibold and McPeck 2006). Such methods are however not based on a genuine sampling theory, and they do not pave the road for a dynamic modelling of ecological communities, potentially in relation to dynamic global vegetation models. Consequently, it is unclear how to translate the results of ordination analyses into predictions of community dynamics, and notably the role that dispersal limitation may have in the ability of ecological communities to track environmental changes (Thuiller et al. 2005).

Here we propose a simple dynamic model of ecological communities, associated with a proper stochastic sampling theory (Alonso et al. 2006). This model provides a simple interpretation of the causes of variation in recruitment limitation. Species abundance distributions, when analysed with the appropriate methods, are useful to understand how biodiversity is structured (McGill et al. 2007, Alonso et al. 2008). Although our method is based on simplifying assumptions, it does capture the cross-sample variation in recruitment limitation. Compared with ordination methods, our model does not yield a partition of the variation into geographic and environmental components. To do this, we would need to model explicitly an environmental filter, acting differentially among species, instead of using a neutral-based framework to infer this filtering. Although variation among species may be modelled, attributing to each species different parameters is a non-trivial task. In ordination methods, this problem is bypassed because the species dynamics is not modelled explicitly. Instead, the correlations between diversity and environmental axis summarize among-plots diversity /environmental structure.

In our approach, one possible solution to the over-parameterization problem could be to use hierarchical modelling, in which species-specific parameters are not determined, but instead belong to a certain distribution, whose parameters are subject to estimation (Clark 2007). Alternatively a set of relevant species functional traits may be used to constrain the species model parameters. If such a general model were available, it would permit not only to detect the intensity of dispersal limitation, but also the intensity of environmental filtering upon each species, and would therefore be much more informative than ordination approaches.

Perspectives

The neutral theory of biodiversity has been criticized for the crudeness of its speciation model, whereby each individual has a fixed probability of mutating into an altogether new species (Hubbell 2001, Nee 2005, Ricklefs 2006). In previous approaches, the species abundances were constrained by the mode of speciation in the regional species pool (Etienne et al. 2007). Our extension of the theory does not assume that the regional abundance distribution is predicted under the assumption of neutrality. Rather, this distribution is reflected by the total available sampling in the metacommunity. Therefore, parameter θ disappears from our formulation of the theory, and it is replaced by the vector of regional species abundances, X . As more data are being collected in species-rich environments such as tropical forests, and worldwide herbarium data are being made publicly available, the issue of estimating the structure of the regional species pool should become more easily tractable (Graham et al. 2004). Therefore, the framework we defined here is ideally suited to relate local-scale data (typically plot-based for tropical trees) to regional-scale data. Finally, our inference approach may be useful for constraining community models aimed at predicting the impacts of global change on biodiversity (Thuiller et al. 2005). Such models would take into account both dispersal limitation and habitat specialization, within a framework that would permit a direct prediction of sampling uncertainties (Higgins et al. 2003).

Acknowledgements – We thank François Goreaud, Simon Levin and Christophe Thébaud for constructive comments. Diego Vázquez, David Alonso and Marc Cadotte provided useful insights that improved the manuscript. This work was partly funded by the ANR through the BRIDGE project. The Panama Canal Watershed dataset was coordinated through the Center for Tropical Forest Science (CTFS).

References

- Alonso, D. et al. 2006. The merits of neutral theory. – *Trends Ecol. Evol.* 21: 451–457.
- Alonso D. et al. 2008. The implicit assumption of symmetry and the species abundance distribution. – *Ecol. Lett.* 11: 93–105.
- Bell, G. 2001. Ecology – neutral macroecology. – *Science* 293: 2413–2418.

- Burnham, K. P. and Anderson, D. R. 2002. Model selection and multimodel inference, a practical information-theoretic approach. – Springer.
- Chase, J. M. et al. 2005. Competing theories for competitive metacommunities. – In: Holyoak, M. et al. (eds), *Metacommunities, spatial dynamics and ecological communities*. Chicago Univ. Press.
- Chave, J. 2004. Neutral theory and community ecology. – *Ecol. Lett.* 7: 241–253.
- Chave, J. et al. 2002. Comparing classical community models: theoretical consequences for patterns of diversity. – *Am. Nat.* 159: 1–23.
- Chave, J. et al. 2006. Theoretical biology – comparing models of species abundance. – *Nature* 441: E1–E1.
- Chesson, P. 2000. Mechanisms of maintenance of species diversity. – *Annu. Rev. Ecol. Syst.* 31: 343–366.
- Chust, G. et al. 2006. Determinants and spatial modeling of tree beta-diversity in a tropical forest landscape in Panama. – *J. Veg. Sci.* 17: 83–92.
- Clark, J. S. 2007. *Models for ecological data: an introduction*. – Princeton Univ. Press.
- Comita, L. S. et al. 2007. Developmental changes in habitat associations of tropical trees. – *J. Ecol.* 95: 482–492.
- Condit, R. et al. 2002. Beta-diversity in tropical forest trees. – *Science* 295: 666–669.
- Condit, R. et al. 2004. Tropical forest dynamics across a rainfall gradient and the impact of an El Niño dry season. – *J. Trop. Ecol.* 20: 51–72.
- Cordeiro, N. J. and Howe, H. F. 2001. Low recruitment of trees dispersed by animals in African forest fragments. – *Conserv. Biol.* 15: 1733–1741.
- Cottenie, K. 2005. Integrating environmental and spatial processes in ecological community dynamics. – *Ecol. Lett.* 8: 1175–1182.
- Engelbrecht, B. M. J. et al. 2007. Drought sensitivity shapes species distribution patterns in tropical forests. – *Nature* 447: 80–82.
- Etienne, R. S. 2005. A new sampling formula for neutral biodiversity. – *Ecol. Lett.* 8: 253–260.
- 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. and Alonso, D. 2005. A dispersal-limited sampling theory for species and alleles. – *Ecol. Lett.* 8: 1147–1156.
- Etienne, R. S. and Olff, H. 2005. Confronting different models of community structure to species-abundance data: a Bayesian model comparison. – *Ecol. Lett.* 8: 493–504.
- Etienne, R. S. and Alonso, D. 2006. A dispersal-limited sampling theory for species and alleles. Erratum – *Ecol. Lett.* 9: 500.
- Etienne, R. S. et al. 2007. Modes of speciation and the neutral theory of biodiversity. – *Oikos* 116: 241–258.
- Ewens, W. J. 1972. The sampling theory of selectively neutral alleles. – *Theor. Popul. Biol.* 3: 87–112.
- Flynn, S. et al. 2006. Seed Information Database (rel. 7.0, Oct. 2006) <<http://www.kew.org/data/sid>>.
- Graham, C. H. et al. 2004. New developments in museum-based informatics and applications in biodiversity analysis. – *Trends Ecol. Evol.* 19: 497–503.
- Hardesty, B. D. et al. 2006. Genetic evidence of frequent long-distance recruitment in a vertebrate-dispersed tree. – *Ecol. Lett.* 9: 516–525.
- Hardy, O. J. et al. 2006. Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. – *Mol. Ecol.* 15: 559–571.
- Harms, K. E. et al. 2000. Pervasive density-dependent recruitment enhances seedling diversity in a tropical forest. – *Nature* 404: 493–495.
- Higgins, S. I. et al. 2003. Forecasting plant migration rates: managing uncertainty for risk assessment. – *J. Ecol.* 91: 341–347.
- Holyoak, M. and Loreau, M. 2006. Reconciling empirical ecology with neutral community models. – *Ecology* 87: 1370–1377.
- Hubbell, S. P. 2001. *The unified neutral theory of biodiversity and biogeography*. – Princeton Univ. Press.
- Hurttt, G. C. and Pacala, S. W. 1995. The consequences of recruitment limitation: reconciling chance, history and competitive differences between plants. – *J. Theor. Biol.* 176: 1–12.
- Janzen, D. H. 1970. Herbivores and the number of tree species in tropical forests. – *Am. Nat.* 104: 501–528.
- John, R. et al. 2007. Soil nutrients influence spatial distributions of tropical tree species. – *Proc. Natl Acad. Sci. USA* 104: 864–869.
- Jones, F. A. et al. 2005. A genetic evaluation of seed dispersal in the neotropical tree *Jacaranda copaia* (Bignoniaceae). – *Am. Nat.* 166: 543–555.
- Jones, M. M. et al. 2006. Effects of mesoscale environmental heterogeneity and dispersal limitation on floristic variation in rain forest ferns. – *J. Ecol.* 94: 181–195.
- Kadmon, R. and Allouche, O. 2007. Integrating the effects of area, isolation, and habitat heterogeneity on species diversity: a unification of island biogeography and niche theory. – *Am. Nat.* 170: 443–454.
- Legendre, P. et al. 2005. Analyzing beta diversity: partitioning the spatial variation of community composition data. – *Ecol. Monogr.* 75: 435–450.
- Leibold, M. A. and McPeck, M. A. 2006. Coexistence of the niche and neutral perspectives in community ecology. – *Ecology* 87: 1399–1410.
- Levin, S. A. et al. 2003. The ecology and evolution of seed dispersal: a theoretical perspective. – *Annu. Rev. Ecol. Syst.* 34: 575–604.
- McGill, B. J. et al. 2006. Empirical evaluation of neutral theory. – *Ecology* 87: 1411–1423.
- McGill, B. J. et al. 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. – *Ecol. Lett.* 10: 995–1015.
- Muller-Landau, H. C. 2007. Predicting the long-term effects of hunting on plant species composition and diversity in tropical forests. – *Biotropica* 39: 372–384.
- Munoz, F. et al. 2007. Estimating parameters of neutral communities: from one single large to several small samples. – *Ecology* 88: 2482–2488.
- Nathan, R. and Muller-Landau, H. C. 2000. Spatial patterns of seed dispersal, their determinants and consequences for recruitment. – *Trends Ecol. Evol.* 15: 278–285.
- Nee, S. 2005. The neutral theory of biodiversity: do the numbers add up? – *Funct. Ecol.* 19: 173–176.
- Norden, N. et al. 2007. Is temporal variation of seedling communities determined by environment or by seed arrival? A test in a neotropical forest. – *J. Ecol.* 95: 507–516.
- Paoli, G. D. et al. 2006. Soil nutrients and beta diversity in the Bornean Dipterocarpaceae: evidence for niche partitioning by tropical rain forest trees. – *J. Ecol.* 94: 157–170.
- Pélissier, R. and Couteron, P. 2007. An operational, additive framework for species diversity partitioning and beta-diversity analysis. – *J. Ecol.* 95: 294–300.
- Poorter, L. 2007. Are species adapted to their regeneration niche, adult niche, or both? – *Am. Nat.* 169: 433–442.
- Poorter, L. et al. 2005. Beyond the regeneration phase: differentiation of height – light trajectories among tropical tree species. – *J. Ecol.* 93: 256–267.
- Purves, D. W. and Pacala, S. W. 2005. Ecological drift in niche-structured communities: neutral pattern does not imply

- neutral process. – In: Burslem, D. et al. (eds), Biotic interactions in the tropics. Cambridge Univ. Press.
- Pyke, C. R. et al. 2001. Floristic composition across a climatic gradient in a neotropical lowland forest. – *J. Veg. Sci.* 12: 553–566.
- Ricklefs, R. E. 1987. Community diversity: relative roles of local and regional processes. – *Science* 235: 167–171.
- Ricklefs, R. E. 2006. The unified neutral theory of biodiversity: do the numbers add up? – *Ecology* 87: 1424–1431.
- Russo, S. E. et al. 2005. Soil-related performance variation and distributions of tree species in a Bornean rain forest. – *J. Ecol.* 93: 879–889.
- Schwilk, D. W. and Ackerly, D. D. 2005. Limiting similarity and functional diversity along environmental gradients. – *Ecol. Lett.* 8: 272–281.
- Seidler, T. G. and Plotkin, J. B. 2006. Seed dispersal and spatial pattern in tropical trees. – *Plos Biol.* 4: 2132–2137.
- Sheil, D. et al. 2006. Competition for light amongst tropical tree species. – *J. Ecol.* 94: 494–507.
- Stevenson, P. R. 2000. Seed dispersal by woolly monkeys (*Lagothrix lagothricha*) at Tinigua National Park, Colombia: dispersal distance, germination rates, and dispersal quantity. – *Am. J. Primat.* 50: 275–289.
- Tavaré, S. and Ewens, W. J. 1997. Multivariate Ewens distribution. – In: Johnson, N. et al. (eds), Multivariate discrete distributions. Wiley, pp. 232–246.
- Thuiller, W. et al. 2005. Climate change threats to plant diversity in Europe. – *Proc. Natl Acad. Sci. USA* 102: 8245–8250.
- Tuomisto, H. et al. 2003. Dispersal, environment, and floristic variation of western Amazonian forests. – *Science* 299: 241–244.
- Wang, B. C. and Smith, T. B. 2002. Closing the seed dispersal loop. – *Trends Ecol. Evol.* 17: 379–385.
- Webb, C. O. and Peart, D. R. 2000. Habitat associations of trees and seedlings in a Bornean rain forest. – *J. Ecol.* 88: 464–478.
- Wright, S. J. 2002. Plant diversity in tropical forests: a review of mechanisms of species coexistence. – *Oecologia* 130: 1–14.
- Wright, S. J. et al. 2005. Annual and spatial variation in seedfall and seedling recruitment in a neotropical forest. – *Ecology* 86: 848–860.

Appendix 1.

Alternative derivation of Eq. 1.

We develop the same strategy as the one in Etienne and Olff (2004). Looking backwards in time, one may trace each individual belonging to the local community back to its immigrating ancestor, that is, its last ascendant having immigrated in the local community. The derivation of the likelihood function consists in computing the probability distribution of a given (unknown) species-ancestry distribution, the probability that the immigrating ancestors have a given species abundance distribution (SAD), and that they each have a given number of descendants in the sample. The probability of the observed species–abundance distribution in the sample is the summed probabilities of all the species-ancestry distributions compatible with the observed SAD.

If the local community is small enough compared with the regional species pool, speciation may be neglected inside the local community. All the descendants of an immigrating ancestor thus belong to the same species. Consequently, the

species abundance distribution of the local community can be summarized by the vector of ancestry abundance $\bar{n}_A = (n_1, \dots, n_A)$, where n_i is the number of descendants of ancestor i , and A is the number of immigrating ancestors. Etienne and Olff (2004) established the probability distribution of \bar{n}_A . They showed that it is equivalent to the Ewens (1972) sampling distribution, replacing θ by $I = m(J - 1)/(1 - m)$, where m is the immigration rate into the local community and J is the size of this community:

$$P(\bar{n}_A/I) = \frac{J!}{\prod_{j=1}^J j^{\phi_j} \phi_j!} \frac{I^A}{(I)_J} \quad (\text{A.1})$$

where ϕ_j is the number of ancestors having j individuals in the sample, and $(x)_y$ is the Pochhammer symbol equal to $(x)_y = \prod_{i=1}^y (x + i - 1)$. See also Wakeley 1998, for the derivation of an equivalent formula in population genetics.

It is more useful to consider an ordered sample of the ancestry distribution, $\bar{n}_A^G = (n_1^G, \dots, n_A^G)$, where the individuals are labelled. For example, in a sample of four individuals descending from two ancestors, the set (ancestor 1, ancestor 1, ancestor 1, ancestor 2) is equivalent to (ancestor 1, ancestor 1, ancestor 2, ancestor 1) in the \bar{n}_A terminology, but not for the ordered ancestry distribution. The number of permutations for a given \bar{n}_A^G which give the same \bar{n}_A is equal to:

$$\frac{J!}{\prod_{j=1}^J j^{\phi_j} \phi_j!} \quad (\text{A.2})$$

It is the number of permutations of the J individuals ($J!$), that lead to the same \bar{n}_A , divided by the number of permutations which give rise to the same \bar{n}_A^G . This last number equals the product of permutations inside the groups of descendants of the same ancestor j ($j!$) by the number of permutations of labels of ancestors which have the same number of descendants ($\phi_j!$).

Thus, the probability of \bar{n}_A being equal to the sum of probabilities of \bar{n}_A^G , which lead to the same \bar{n}_A , follows the relation:

$$P(\bar{n}_A/I) = \frac{J!}{\prod_{j=1}^J j^{\phi_j} \phi_j!} P(\bar{n}_A^G/I) \quad (\text{A.3})$$

Combining Eq. A.1 and A.3 we obtain the formula for \bar{n}_A^G (Etienne and Olff 2004):

$$P(\bar{n}_A^G/I) = \prod_{j=1}^J (j - 1)!^{\phi_j} \frac{I^A}{(I)_J} \quad (\text{A.4})$$

The rest of the derivation of the likelihood function is also based on the assumption that the local community is much smaller than the regional pool, so that the immigrating ancestors can be considered as an instantaneous sample of this regional pool. In other words, the time between two immigration events inside the local community is small, so that regional changes in species abundances can be neglected. It follows that, assuming known the regional species abundances x_i i in $\{1, \dots, S\}$, the probability of

drawing a set of ordered ancestors with species abundance distribution (a_1^G, \dots, a_S^G) is given by the multinomial distribution where order matters:

$$P(a_1^G, \dots, a_S^G / A, x_1, \dots, x_S) = \prod_{i=1}^S x_i^{a_i} \quad (\text{A.5})$$

Looking now at the ordered species-ancestry distribution n_{SA}^G corresponding to the label for each individual of the local community of both its immigrating ancestor and species, we find that:

$$P(n_{SA}^G) = P(n_A^G) \times P(a_S^G) \quad (\text{A.6})$$

Moreover, as above, the corresponding unordered species-ancestry distribution n_{SA} is related to n_{SA}^G by:

$$P(n_{SA}) = \frac{J!}{\prod_{i=1}^S \prod_{j=1}^{n_i} j!^{\Phi_{ij}} \Phi_{ij}!} P(n_{SA}^G) \quad (\text{A.7})$$

where Φ_{ij} in $\{1, \dots, S\}$, j in $\{1, \dots, J\}$ is the number of ancestors of species i having j individuals in the sample.

Combining Eq. A.4, A.5, A.6 and A.7, we obtain the following relation:

$$P(n_{SA}) = \frac{J!}{\prod_{i=1}^S n_i!} \frac{I^A}{(I)_J} \prod_{i=1}^S \frac{n_i! x_i^{a_i}}{\prod_{j=1}^{n_i} j!^{\Phi_{ij}} \Phi_{ij}!} \quad (\text{A.8})$$

The probability of a given species abundance distribution (SAD) equals the sum of the probabilities of all the species-ancestry distributions, compatible with this SAD. We note these compatible species-ancestry distributions n_{SA}^+ . Pooling species-ancestries by their ancestors distributions \bar{a} , we obtain:

$$\begin{aligned} P(D/J, m, x_1, \dots, x_S) &= \sum_{\bar{a}} \sum_{n_{SA}^+/\bar{a}} P(n_{SA}/J, m, x_1, \dots, x_S) \\ &= \frac{J!}{\prod_{i=1}^S n_i!} \sum_{\bar{a}} \sum_{n_{SA}^+/\bar{a}} \frac{I^A}{(I)_J} \prod_{i=1}^S \frac{n_i! x_i^{a_i}}{\prod_{j=1}^{n_i} j!^{\Phi_{ij}} \Phi_{ij}!} \end{aligned} \quad (\text{A.9})$$

As stated in Etienne (2005, Eq. A9), the number $\sum_{n_{SA}^+/\bar{a}_i} \frac{n_i!}{\prod_{j=1}^{n_i} j!^{\Phi_{ij}} \Phi_{ij}!}$ equals $\bar{s}(n_i, a_i)$. Thus, we obtain:

$$P(D/J, m, x_1, \dots, x_S) = \frac{J!}{(I)_J \prod_{i=1}^S n_i!} \prod_{i=1}^S \sum_{a_i=1}^{n_i} \bar{s}(n_i, a_i) I^{a_i} x_i^{a_i} \quad (\text{A.10})$$

This expression is equivalent to that of Etienne and Alonso (2005, Eq. 25, erratum in Etienne and Alonso 2006). This correspondence is based on the equality:

$$(I x_i)_{n_i} = \sum_{a_i=1}^{n_i} \bar{s}(n_i, a_i) I^{a_i} x_i^{a_i} \quad (\text{A.11})$$

This equality is easily found by expanding the left member, and in recalling the recurrence relation $\bar{s}(n+1, m) = n \times \bar{s}(n, m) + \bar{s}(n, m-1)$ (Abramowitz and Stegun 1964).

Considering D samples of the same metacommunity, far enough so that migration between samples may be neglected, the likelihood function of the D samples is equal to the product of the likelihood function for each sample given by Eq. A.10. Hence (using Eq. A.11), we obtain our Eq. 1:

$$L(M, X/N) = \prod_{i=1}^D \frac{J_i!}{\prod_{j=1}^S n_{ij}!} \frac{\prod_{j=1}^S (I_i x_j)^{n_{ij}}}{(I_i)_{J_i}} \quad (\text{A.12})$$

where J_i , I_i , and n_{ij} refer to the sample size, immigration parameter, and abundance of species j for the sample i .

References

- Abramowitz, M. and Stegun, I. A. 1964. Handbook of mathematical functions with formulas, graphs and mathematical tables. – Dover Publications, INC.
- Etienne, R. S. 2005. A new sampling formula for neutral biodiversity. – Ecol. Lett. 8: 253–260.
- Etienne, R. S. and Olf, H. 2004. A novel genealogical approach to neutral biodiversity theory. – Ecol. Lett. 7: 170–175.
- Etienne, R. S. and Alonso, D. 2005. A dispersal-limited sampling theory for species and alleles. – Ecol. Lett. 8: 1147–1156.
- Ewens, W. J. 1972. The sampling theory of selectively neutral alleles. – Theor. Popul. Biol. 3: 87–112.
- Wakeley, J. 1998. Segregating sites in Wright's island model. – Theor. Popul. Biol. 53: 166–174.

Appendix 2.

The pooled 42 plots are approximating well species abundances in the Panama Canal region.

In our analysis, we make the assumption that the regional species abundances can be approximated by their relative abundances in the D pooled samples. As the number D of samples increases, this assumption is more and more reasonable. We want to know whether the dataset we are using is large enough so that this assumption is acceptable. If this is the case, then erasing one plot from the dataset should not change the relative species abundances a lot. We thus randomly erased one of the D plots, and recomputed the species abundances. We repeated this analysis 100 times, and computed the coefficient of variation of the total abundances of each species. Figure S.1 shows this coefficient of variation for each species. We see that except for seven species, this coefficient is always less than 5%. More problematic is the fact that species having the largest variation are also among the most abundant ones. However, this variation keeps reasonable, when reasoning in terms of relative species abundances. Indeed, the most abundant species (*Gustavia superba*)

represents 4% of the trees in the dataset, and thus, even with a coefficient of variation of 20%, the precision of its approximated relative abundance will be of one point of percentage. The dataset we are using thus seems to be large enough for using our approximation.

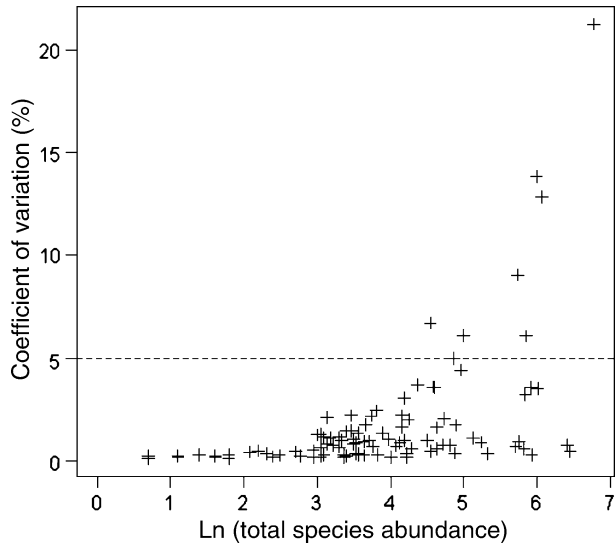


Figure S1. Coefficient of variation of total species abundances when erasing one of the 42 plots (based on 100 random draws).