Age structure in neutral theory resolves inconsistencies related to reproductive-size threshold

James Rosindell[1,2,]*, Patrick A. Jansen[3,4,5] and Rampal S. Etienne[5]

[1] Institute of Integrative and Comparative Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, West Yorkshire, UK
[2] Institute of Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID 83844, USA
[3] Center for Tropical Forest Science, Smithsonian Tropical Research Institute, PO Box 0843-03092, Balboa, Ancon, Republic of Panama
[4] Centre of Ecosystem Studies, Wageningen University, PO Box 47, 6700 AA Wageningen, the Netherlands
[5] Community and Conservation Ecology, Centre for Ecological and Evolutionary Studies, University of Groningen, PO Box 11103, 9700 CC Groningen, the Netherlands
*Correspondence address. Institute of Integrative and Comparative Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, West Yorkshire, UK. E-mail: james@rosindell.org

Abstract

Aims

Neutral theory consists of a suite of models that assume ecological equivalence among individual organisms. They have been most commonly applied to tropical forest tree communities either as null models or as approximations. Neutral models typically only include reproductive adults; therefore, fitting to empirical tree community data requires defining a reproductive-size threshold, which for trees is usually set arbitrarily to a diameter at breast height (DBH) of 100 mm. The inevitable exclusion of some reproductive adults and inclusion of some saplings cause a non-random sampling bias in neutral model fits. Here, we investigate this problem and resolve it by introducing simple age structure into a neutral model.

Methods

We compared the performance and sensitivity of DBH threshold of three variants of a spatially explicit neutral model: the traditional model, a model incorporating random sampling and a model with two distinct age classes—reproductive adults and saplings. In the age-structured model, saplings are offspring from adults that disperse according to a Gaussian dispersal kernel around the adults. The only extra parameter is the ratio of adults to saplings, which is not a free parameter but directly measurable. We used species–area relationships (SARs) to explore the predicted effect of saplings on the species richness at different scales in our model. We then evaluated the three model variations to find the parameters required to maintain the observed level of species richness in the 50-ha plot on Barro Colorado Island (BCI). We repeated our analysis filtering the data at different minimum tree-size thresholds in order to find the effect this threshold has on our results. Lastly, we used empirical species–individual relationships (SIRs) to test the pre-existing hypothesis that environmental filtering is the primary cause of differences between the assemblage of saplings and that of adults on BCI.

Important Findings

Our age-structured neutral model was characterized by SARs that were insensitive to the presence of saplings at large scales and highly sensitive to them at small scales. Both models without age structure were highly sensitive to the DBH threshold chosen in a way that could not be explained based on random samplings alone. The age-structured neutral model, which allowed for non-random sampling based on life stage, was consistent with species richness observations. Our analysis of empirical SIRs did not support environmental filtering as a dominant force, but it did show evidence for other differences between age classes. Age can now be easily incorporated into future studies of neutral models whenever there is a concern that a sample is not entirely composed of reproductive adult individuals. More generally, we suggest that modeling studies using tree data subject to a minimum size threshold should consider the sensitivity of their results to that threshold.

Keywords: neutral theory • age structure • sampling • DBH

INTRODUCTION

Neutral theory refers to a collection of neutral models each assuming ecological equivalence between individuals (Bell 2000; Caswell 1976; Hubbell 1997, 2001). Hubbell (2001) introduced the best-known neutral models and put forward a strong case for their general use in ecology. Since then neutral models have received much interest and criticism (Clark

2009; McGill 2003; Ricklefs 2003; for reviews, see Alonso *et al.* 2006; Etienne and Alonso 2007; Leigh 2007; Leigh *et al.* 2010; Rosindell *et al.* 2011). The objective of neutral theory is to investigate the consequences of making this neutrality assumption, not to claim that the neutrality assumption is generally true (Rosindell *et al.* 2011). Remarkably, neutral theory can accurately predict a number of macroecological patterns including species–area relationships (SARs) and species abundance distributions. Furthermore, it provides a possible mechanism to explain the co-occurrence of many species without requiring niche differences. Neutral models do fail to explain some patterns (Dornelas *et al.* 2006; McGill *et al.* 2006). However, model failure reveals something interesting about the system and can often be rectified with more advanced (possibly non-neutral) models built on the foundations laid down by neutral theory (Rosindell and Cornell 2009; Rosindell *et al.* 2010).

In the simplest neutral models, individuals die at a constant rate, each leaving a gap. These gaps are filled by the immediately maturing offspring of other individuals in the system. The identity of reproducing individuals is chosen at random based on a dispersal model that may assume complete mixing or may be more complex involving a spatially explicit dispersal kernel. Under these assumptions, species abundances follow a random walk to eventual extinction. Speciation replenishes the lost species so that species richness is maintained in a dynamic equilibrium akin to that proposed by MacArthur and Wilson (1967). Many varieties of neutral models exist, differentiated by the way they model dispersal and speciation. The best-known, and most frequently tested, neutral model assumes a two-tier (spatially implicit) spatial structure consisting of a small local community and a much larger metacommunity or species pool where speciation occurs at a constant per-individual rate (Etienne 2005; Hubbell 2001).

There is generally no age structure in neutral theory; all individuals are assumed to be reproducing adults (but see O'Dwyer *et al.* 2009). Consequently, a rigorous comparison with empirical data (Etienne 2005) requires that data contain only reproductive individuals, but in practice we most often do not know which sampled individuals are reproductive. In the case of tropical forest tree data, which remains the most common form of data used to test neutral models, all individuals typically are assumed to attain adulthood at a diameter at breast height (DBH) of 100 mm. This arbitrary cut-off may approximately correspond with the actual reproductive-size threshold of many tree species (Thomas 1996; Wright *et al.* 2005) but will inevitably lead to the inclusion of many juveniles and the exclusion of many adults—a potentially serious problem that has received surprisingly little attention from either the proponents or the critics of neutral theory (but see Jabot *et al.* 2008).

Jabot *et al.* (2008) began to address these issues by testing the effect of DBH threshold on the fit of a spatially implicit neutral model to empirical tropical forest data. They repeated the entire fitting process for a number of different DBH thresholds

ranging from 10 to 100 mm and found that the fitted values for the dispersal limitation parameter *m* consistently decreased with larger DBH thresholds. This meant that according to the model, communities including larger trees had decreased dispersal from outside the local community. Jabot *et al.* (2008) interpreted their results as a signature of environmental filtering, arguing that a sample of older trees is more dissimilar to the species pool than a sample including younger trees because unsuitable species are environmentally filtered out before reaching adulthood. However, there are two alternative explanations for the inconsistency. The first is that inconsistencies are an artifact of the strongly negative relationship between sample size and DBH threshold. The second explanation is that the outcomes depend not only on the number of individuals above the DBH threshold but also on the number below it.

O'Dwyer *et al.* (2009) developed an elegant mathematical framework that allows size structure to be integrated into the best-known non-spatial version of neutral theory. Interestingly, they found that the species abundance distributions are independent of the underlying size structure. Our work complements this with a different approach based on coalescence that enables us to study a spatially explicit model and take into account sampling effects that are size structure dependent. We build on the work of Jabot *et al.* (2008) by re-evaluating the biases caused by DBH threshold choice and presenting a solution to the problem of these biases. Using a spatially explicit model, we show that the fit of simple neutral models is highly sensitive to the arbitrary choice of DBH threshold, as in Jabot *et al.* (2008), which runs these models unstable. To explore the possibility that inconsistencies are due to changes in sample size with DBH threshold, we extended the model with random sampling as to account for these differences but find that the instability persists albeit at a reduced level. Then, to explore the possibility that explicit consideration of juveniles yields more stable outcomes, we add simple age structure (saplings and adults) to the model and find that this does indeed solve the problem. Finally, we analyze species–individual relationships (SIRs) for the empirical data and find that there are distinctions between sapling and adult life stages that are detectable from species richness alone but that contrary to Jabot *et al.* (2008), environmental filtering cannot have a key role in shaping this difference for trees with DBH ⩾10 mm.

METHODS

We first developed a spatially explicit neutral model that allows to take into account random sampling and age structure. We analyzed this alongside two other models using simulations that are based on coalescence methods, a powerful and efficient technique imported from population genetics. We first explored the behavior of the new age-structured model by plotting SARs that showed the effect of sapling density on species richness at different scales. Next, for three

different neutral models: a simple neutral model, a neutral model with sampling to correct for sample size bias and an age structured neutral model, we calculated which combinations of model parameters would be consistent with observed empirical species richnesses at different DBH thresholds. We compared the performance and sensitivity to DBH threshold choice for the alternative models. Finally, we analyzed SIRs drawn using individuals from differently sized subplots and with different minimum DBH thresholds to test for the presence of processes influencing age structure, such as environmental filtering (cf. Jabot *et al.* 2008).

The spatially explicit neutral model

We followed the established approach for individual-based neutral models. A random individual is chosen to die, leaving a gap in the habitat. This is then filled with offspring from another individual, with probability $(1 - v)$, or with an individual that founds a new species by 'point mutation speciation', with probability v. In the most frequently studied spatially implicit neutral model (Hubbell 2001), individuals are spatially arranged into a metacommunity and a local community that operate at different timescales from one another. The local community receives potentially new species by migration from the metacommunity. Both the local community and the metacommunity are well mixed with no internal spatial structure. In the spatially explicit neutral model that we use here, individuals are spatially structured at a density of ρ on an infinitely large grid of squares of side $\rho^{-0.5}$. There is exactly one (adult) individual in each grid cell. When an empty cell is to be replaced with offspring of another individual, that individual is chosen using a 'dispersal kernel' that describes the probability of dispersal as a function of position relative to the empty cell. We used a dispersal kernel given by a symmetric normal distribution with variance ρL^2 centered about the habitat gap, where L is a variable representing dispersal distance (Rosindell and Cornell 2007). The output of this model is the community structure found within a survey area of size $A\rho^{-1}$, which supports A individuals. The survey area is usually assumed to be square, but we used a rectangle so that it exactly matched the shape of the true survey area of the empirical data we used.

To tractably simulate a spatially explicit infinite area, we used a special technique of simulation known as coalescence. Based on techniques used in population genetics (Kingman 1982), coalescence has very significant advantages in simulating neutral community models (Rosindell *et al.* 2008) and can be adapted to allow for age structure. The natural way to simulate an individual-based neutral model is to start with some initial condition and apply the rules of the model systematically forward in time until an equilibrium is reached. Coalescence does the opposite: it begins at the present and works backward in time, tracing the ancestry of just the sampled individuals. Once common ancestors and speciation events have been identified for each sampled individual, we know their species identity and the simulation is complete. For more details, refer to Rosindell *et al.* (2008).

There are different ways to sample from a spatially explicit environment, even when ignoring the issues relating to age and size structure. For example, one could perform a census of a bounded survey area or a random sample of the same number of individuals from a larger survey area. Many neutral models are non-spatial in which case the distinction becomes irrelevant or they have focused on a census of a bounded survey area that covers part (or all) of the larger simulated landscape. However, it is straightforward to account for different types of spatially structured and/or random sampling in neutral model simulations and we expect the sampling choice to influence the results.

Introducing three model variations

The simplest model that we investigated sampled all individuals within the survey area set within an infinite area—this meant assuming all those individuals were adults and that no adults were missed, the standard assumption of neutral models. We refer to this as the infinite area neutral (IAN) model. The second model that we investigated incorporated random sampling to account for differences in sample size related to DBH threshold. This model randomly sampled a subset of those individuals and thus all those sampled were assumed to be adults but not all adults need be sampled. We refer to the randomly sampled version of IAN as an RIAN.

The third model that we investigated included simple age structure, with two distinct age classes: saplings and adults. This was achieved by introducing parameter D so that the ratio of adults to saplings was 1:D. As always there was exactly one adult per grid cell, but now additional saplings were scattered into grid cells according to a uniform distribution until the total number of samplings achieved the required level. It is easiest to explain and justify the way we included saplings in terms of forward (not coalescent) simulations. An explicit consideration of the sapling life stage requires dead adults to be replaced by randomly maturing saplings and dead (or maturing) saplings to be replaced by random offspring from adults selected according to a dispersal kernel. The species of a maturing adult is then selected at random from a set of saplings, the species identities of which were in turn selected according to a dispersal kernel. Consequently, the species identity of an adult was essentially selected according to the same dispersal kernel as the model without saplings but based on what was present in the surroundings half a generation ago instead of what was present now. We do not expect the half generation time lag to be significant. The primary influence of the sapling life stage in our model is thus the presence of the saplings that may be sampled, not the influence of the sapling life stage on the adult assemblage. Simulating sampling from the model according to a DBH threshold involves sampling all the adults and a proportion of saplings so that the total sample is of the required size. To achieve this, we effectively scattered a carpet of saplings onto the completed simulation with the species identity of each sapling chosen according to the dispersal kernel as offspring of a nearby adult. Saplings may of course be offspring

from adults located outside the boundaries of the sampling area. Our model reflects this exactly: the coalescence approach allows a simulation landscape that extends indefinitely outside of the survey area and therefore enabled us to make saplings the offspring of parents that need not be in the survey area themselves. We refer to the IAN model that includes an explicit sapling life stage as an SIAN.

Exploring age structure in a neutral model with SARs

We used SARs to explore the predictions of the SIAN model before considering empirical data. The SARs show the species richness as a function of area and in this case assume that a census is taken within the sampling area. We chose SARs because they capture the scale dependence of species richness that is relevant for our comparisons with empirical data and because SARs are extremely well-studied macroecological patterns that are frequently used as summary statistics. SARs enable us to demonstrate the key features of adding saplings. For the age-structured SIAN model, SARs depend on three parameters: dispersal distance L, speciation probability v and adult-to-sapling ratio D. The behavior of the SARs for the SIAN with $D = 0$ is already well explored (Rosindell and Cornell 2007) and yields a constant shape of SAR (on logarithmic axes) for all values of L that are not very small. The SAR can be moved diagonally in logarithmic space by changing L but its shape remains constant. In order to show this, Rosindell and Cornell (2007) plotted $\log(S/g(L))$ against $\log(A/g(L))$, where $g(L)$ represents the function of L that causes the SARs to overlap exactly showing that they have the same shape. $g(L)$ is well approximated by L^2 in most cases that makes $A/g(L)$ dimensionless. The same pattern emerges when $D > 0$, and thus plotting $S/g(L)$ against $A/g(L)$ reduces the effective number of parameters by one. The effects of speciation rate on SARs (Pigolotti and Cencini 2009; Rosindell and Cornell 2007, 2009) are less relevant here, so we concentrate on the sapling-to-adult ratio D. Because D affects the density of sampled individuals (albeit not the density of sampled adults), we analyzed two sets of SARs, one where $\rho = 1$ (constant density of adults) and one where $\rho = 1 + D$ (constant density of individuals whether adults or saplings).

Testing for model consistency with empirical data

Our comparison with empirical data uses the species richness of empirical data at different DBH thresholds and tests for consistency of the three alternative models with these data. We used species richness because using the full species abundance distribution is not feasible. Also, good fits of species abundance distribution always coincide with matching the correct species richness (Etienne *et al.* 2007), so failure to match the correct species richness does seem enough to rule out some possible models. We used the techniques described in Rosindell *et al.* (2008) to draw a 'manifold' that is a graph of speciation rate v against dispersal distance L to show every possible combination of parameters that yields an expected species richness that is

the same as the known empirical species richness. Because there is some limited uncertainty in the manifold, we plotted the mean ± standard deviation (as a triplet of lines in the same style).

We used data from the 2005 census of the 50-ha (1,000 × 500 m) forest dynamics plot on Barro Colorado Island (BCI), Panama (Condit 1998; Hubbell *et al.* 1999, 2005). All trees with DBH >10 mm within the plot have been censused and their species, location and DBH recorded, making it a perfect data set for this project. We analyzed the data for different DBH thresholds. For each, we recorded the number of individuals and the species richness.

If the manifolds drawn for different DBH thresholds are non-overlapping for realistic parameter values, then there exist no parameter values consistent with the species richness from all these thresholds and the model can be rejected as an explanation for richnesses of trees at different sizes. If, in contrast, the manifolds cross at realistic parameter values, then these parameter values are consistent with the species richness for all DBH thresholds. We applied this test to all three spatially explicit neutral models. The three DBH thresholds that we used for our analysis were 10, 31 and 100 mm that are approximately evenly distributed in logarithmic space. In the RIAN model that allowed for random sampling within the survey area, all individuals sampled had to be adults, but not all adults needed to be sampled. Nevertheless, this still meant that assuming all individuals with a DBH ⩾10 mm were adults. In the most advanced model SIAN, which included saplings, we assumed that all adults were sampled but that some saplings could also be sampled to make up the total sample size to the correct number of individuals.

Our simulations (and thus our manifolds) used the parameter L for dispersal distance measured in grid cell widths (not meters). In order to make this meaningful, we converted values of L to a true dispersal distance measured in meters as follows. We repeatedly drew dispersal distances from the dispersal kernel, we then discretized them into unit grid spaces and disallowed dispersal distances of zero, just as in the model simulations. For each dispersal draw, we calculated the Euclidean dispersal distance using Pythagoras' theorem. We then found the median dispersal distance by ordering all the independent readings. Finally, the distance was converted to meters from grid spaces using the empirically measured density of individuals per square meters (given by ρ).

Testing for environmental filtering with SARs

The simplest test for environmental filtering from empirical data on BCI would be to count how many species are present in the empirical data for 10 mm ⩽ DBH < 100 mm and DBH ⩾ 100 mm. However, this test is biased by the number of individuals present in each DBH range and there would be many more individuals expected in the smaller size category. To correct for this, we used SIRs in a more accurate test. SIRs are a non-spatial alternative to SARs; SIRs represent the accumulation of species richness with increasing sample size. SIRs are preferable to SARs for our tests because SARs are influenced by

the density of individuals that is strongly correlated with DBH threshold. The species richness in any sample will be strongly influenced by the number of individuals in the sample; SIRs control for this, whereas SARs do not. We plotted SIRs for the smaller and larger DBH ranges on the same axes. We used a uniform distribution to randomly choose individuals for the SIR from those within the specified DBH range. This was repeated to obtain estimates for the mean and standard deviation of the SIR for each size class. A noticeable difference in SIRs between age classes implies a difference in diversity. If environmental filtering is present, the SIR for smaller DBH should have significantly higher diversity than the SIR for larger DBH. We repeated the test for different plot sizes <50 Ha to check for possible effects of scale, thus incorporating the important qualities of the SAR across a number of SIR plots.

RESULTS

Exploring age structure in a neutral model with SARs

The SARs emerging from the model as a function of sapling-to-adult ratio are shown in Fig. 1. At the largest scales, the saplings had no effect on species richness because almost all saplings in the large survey area also had conspecific adults in the survey area. At small scales, saplings had the same effect as adults on the species richness. This is because at these scales most individuals (whether matured or not) were offspring from others outside of the survey area, so the community represents a sample of offspring from a broader area defined by the dispersal kernel and the age of those offspring are almost irrelevant (Fig. 1).
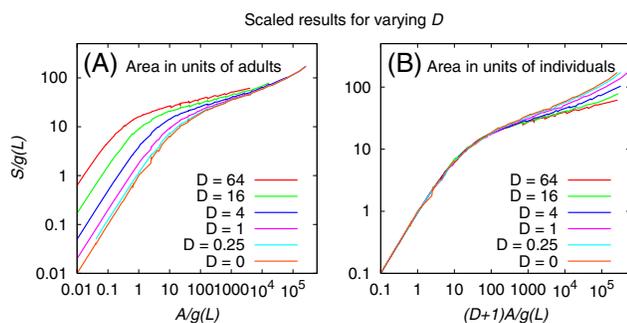
Testing for model consistency with empirical data

The basic model without saplings or random sampling (IAN) produced parameter manifolds at different DBH thresholds of 10, 31 and 100 mm that were completely inconsistent with one another (Fig. 2A). This showed that the observed species richnesses at different DBH thresholds cannot be explained by simple neutral models, in agreement with Jabot *et al.* (2008). Our calculations using this model had to assume that every adult in the survey area has been sampled and no other
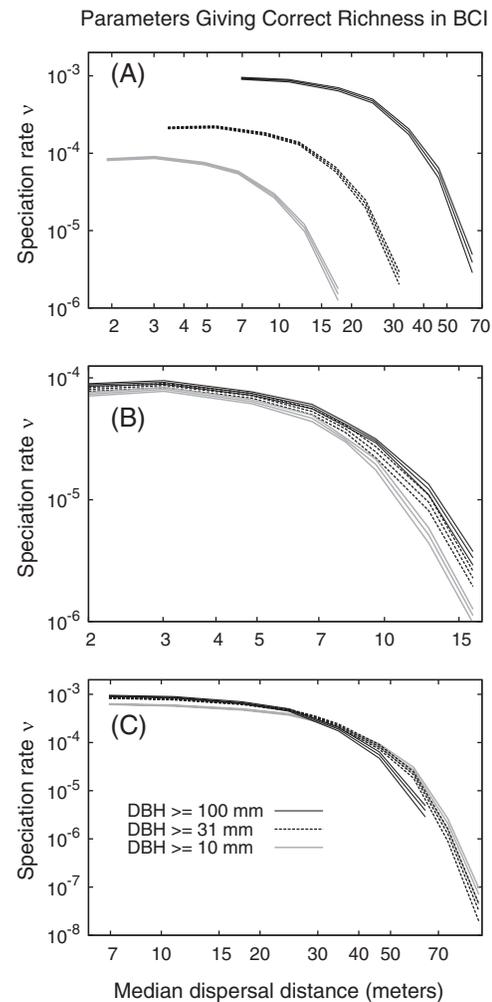

Figure 2: the effects of DBH threshold choice on the parameter manifolds of three alternative spatially explicit model applied to the 50-ha plot on BCI, Panama. The manifolds give the combinations of speciation rate n and dispersal distance L that are consistent with the observed tree species richness of the 50-ha plot on BCI, Panama. The three manifolds correspond to different reproductive-size thresholds (DBH). Each manifold is plotted as three lines of the same style: the mean and the mean ± standard deviation. Shown are manifolds for the basic spatially explicit neutral model (**A**), the spatially explicit neutral model with random sampling (**B**) and the more advanced age-structured model that accounts for random sampling within the survey area (**C**). Only for age-structured model do the manifolds cross, yielding possible parameters that are consistent with the species richness at all three DBH thresholds.

Figure 1: SARs from an age-structured neutral model. In (**A**), area was measured in units of adult individuals, and in (**B**), it was measured in units of individuals (both saplings and adults combined). This diagram is distinct from a species accumulation curve because samples still represent a census of a sampling area of size A; however, the density of individuals is constant and thus number of individuals remains a convenient way to interpret area. Each SAR represents data from simulations for dispersal distance $L \in \{3, 4, 6, 8, 11, 16, 23, 32, 45, 64\}$ that fall on top of one another because of rescaling by $g(L) \approx L^2$ (see Rosindell and Cornell 2007). The speciation rate parameter v was held constant at $v = 0.00001$. The convergence of the SARs at large scales in (A) and at small scales in (B) shows that saplings have no effect at large scales and the same effect as adults at small scales.

individuals were sampled. Consequently, adult density decreases steeply with DBH threshold that appears inconsistent with reality where adult density is constant and DBH threshold is simply a sampling of those adults. Inclusion of random sampling effects to keep adult density constant in the RIAN model moved the manifolds for different DBH thresholds closer together, but they still did not coincide (Fig. 2B).

Incorporation of age structure did make the manifolds for DBH thresholds consistent. The parameter manifolds for three different DBH thresholds using the neutral model including both adults and saplings (SIAN) are shown in Fig. 2C. These manifolds assumed that the trees with DBH ⩾100 mm were all the adults and were always sampled, whilst individuals with DBH <100 mm were saplings, some of which were sampled depending on the DBH threshold. The manifolds produced were consistent with one another and crossed at a median dispersal distance of ∼30 m and a speciation rate of ∼0.0003. Although this rules out pure sampling as an explanation, it does not constitute sufficient evidence to accept or reject an age-structured neutral explanation for the age-structured tree community on BCI because we only looked at species richness.

Testing for environmental filtering with SIRs

If environmental filtering is a significant force, as Jabot *et al.* (2008) inferred, it will reduce species richness from saplings to adults. We found that 4.3% of species were absent in the 10 mm ⩽ DBH < 100 mm range, while 23.4% of species were absent in the DBH ⩾ 100 mm range, which seems to support the idea of environmental filtering. However, 90% of trees fell in the 10 mm ⩽ DBH < 100 mm range and we would expect a much larger sample to contain many more species regardless of environmental filtering. Therefore, we drew SIRs, which account for sample size. These showed that the species richness was, if anything, lower for the smaller size threshold. Our result here was robust to different spatial scales from which the SIRs were drawn Fig. 3. A sampling-based interpretation of these SIRs would require species to accumulate at the same rate with increasing sample size, regardless of DBH class. However, we do not really see this, especially in the top left panel of Fig. 3, which suggests that adults are more speciose than the same number of saplings. This could possibly be because some adults create a much larger number of saplings than others. However, we note from Fig. 1 that an entirely neutral model with age structure also expects a lower richness of saplings at larger scales, qualitatively in line with our observations.

DISCUSSION

Tests of neutral theory with empirical data have used the arbitrary DBH threshold of 100 mm to distinguish adults from juveniles. Jabot *et al.* (2008) showed that this choice can strongly affect the recruitment–limitation parameter for the model's best fit to observed data. Here, we used a spatially explicit model to assess how serious the problem is and to what degree it can be solved by incorporating sampling effects and

age structure. We found that species richness estimates are strongly affected by the DBH threshold chosen, as in Jabot *et al.* (2008). Even with both random and spatial sampling effects accounted for, the spatially explicit neutral model cannot reconcile the species richnesses at three different DBH thresholds. This discrepancy is perhaps not surprising given that the neutral model assumes that all individuals sampled are reproductive adults, while in reality the sample must contain some saplings and exclude some adults. For example, the reproductive-size threshold ranges from 61 to 467 mm DBH for just 16 species on BCI (Wright *et al.* 2005). More generally, small-statured tree species of the forest understorey may not ever reach a DBH of 100 mm, while large-statured tree species of the canopy may not reproduce before they attain a DBH of at least 500 mm. This difference in reproductive-size threshold across species is well known (Thomas 1996; Wright *et al.* 2005) and is non-neutral. Neutral models cannot deal with intraspecific variation of this nature and we cannot expect a neutral model to be entirely insensitive to the size threshold.

We resolved the problem this poses for fitting neutral models by including a basic age structure containing both saplings and reproductive adults as distinct life stages, which requires only one extra parameter: the ratio of adults to saplings. The age-structured model simultaneously explained the species richnesses observed at a range of different DBH thresholds. Although the dispersal distance parameter needed for this (a median distance of 30 m) seems reasonable, the speciation rate of 0.0003 *per capita* per generation seems unrealistic. However, this parameter should not be interpreted literally as the true *per capita* speciation rate because it also encompasses individuals arriving from long-distance dispersal events (Rosindell and Cornell 2009) as well as the arrival of barely recognizable 'incipient species' that go extinct before they become 'good species' (Rosindell *et al.* 2010). Our age-structured neutral model does a good job at fitting the species richness for different DBH thresholds, more data would thus be required to show observable effects of any intraspecific size structure.

Our model differs from the model tested by Jabot *et al.* (2008) in many important respects. For example, the latter model was spatially implicit and the metacommunity abundance distribution is not based on a neutral model. Given these differences, it is not surprising that we obtained different results. However, our SIR results, which are not related to our choice of models, do not support the hypothesis that environmental filtering is prevalent: if it were, we would expect to see more diversity at earlier ages that is later filtered out. We have not demonstrated the absence of environmental filtering altogether, but our results do suggest a limited importance of filtering during the transition between 10 and 100 mm DBH size classes. Environmental filtering could still be prevalent in the transition from seed to seedling and from seedling to sapling that are not represented here. Indeed, earlier empirical work has highlighted fine-scale environmental filtering at very early life stages that becomes closer to an apparently neutral
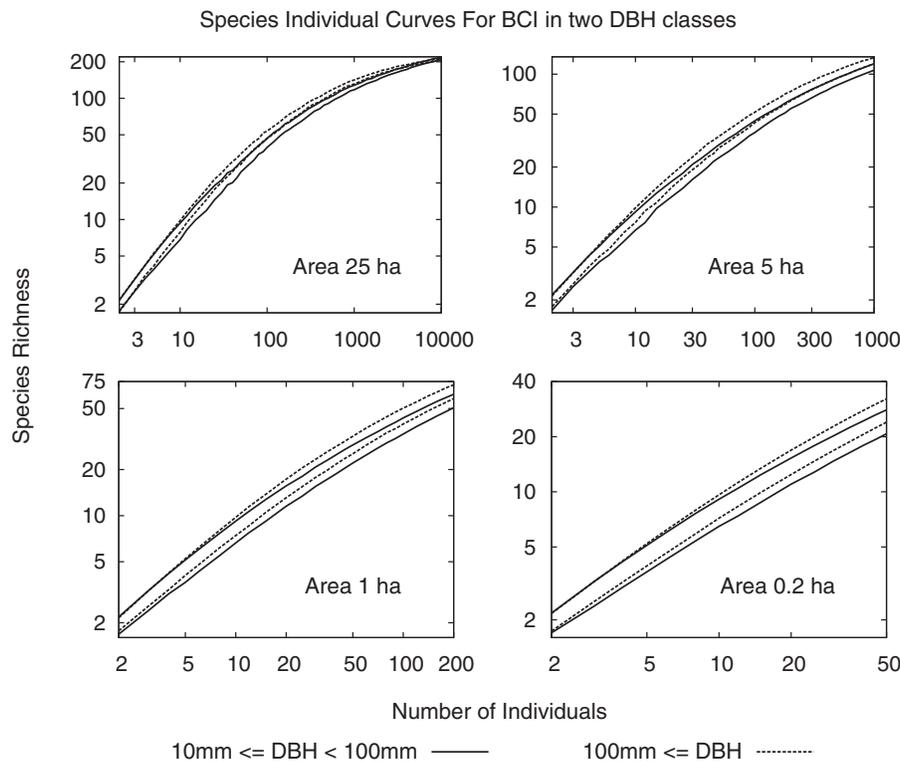
Figure 3: SIRs for two size classes of trees at BCI, Panama. Each SIR was constructed by taking 1000 random samples of the given number of individuals and plotting the mean ± standard deviation across those samples. For each area, the 50-ha plot was subdivided into as many non-overlapping squares of the given area as possible and SIRs were included from all of them, showing effects of spatial scale on the SIRs.

distribution of species at larger age classes (Gravel *et al.* 2008). Environmental filtering may still be present at larger size classes but is drowned out by other factors.

It would be easy to look at our initial results in Fig. 2 and dismiss the neutral model, but we have shown that the discrepancies can easily be accounted for. Our model's parameter manifolds can be made consistent with one another without dropping the assumption that adult individuals are those with 100 mm DBH or greater so perhaps the 100–200 mm DBH is justified as an approximation even though some adults will have DBH <100 mm. Future work can include additional age structure in the model that can still be approached using similar techniques.

We expect the seedling layer to have much more prevalent environmental filtering than the sapling layer because seedlings are far more susceptible to mortality and saplings have already survived in their environment for some time. An exciting direction for future research would be to use data from remote sensing to collect data on all trees reaching the canopy layer of the forest. This would enable a new definition of reproductive adulthood that is independent from DBH and most likely more reliable: trees generally begin reproduction when and only when they reach the canopy. We hope that our work will increase awareness of the consequences of choosing an arbitrary DBH cut-off for any analysis of reproductive adults (neutral or otherwise) and that we have provided a simple

way for age-structured sampling to be accounted for in future confrontations of neutral theory to data.

FUNDING

ACKNOWLEDGEMENTS

REFERENCES

Alonso D, Etienne RS, McKane AJ (2006) The merits of neutral theory. *Trends Ecol Evol* **21**:451–7.

Bell G (2000) The distribution of abundance in neutral communities. *Am Nat* **155**:606–17.

Caswell H (1976) Community structure—neutral model analysis. *Ecol Monogr* **46**:327–54.

Clark JS (2009) Beyond neutral science. *Trends Ecol Evol* **24**:8–15.

Condit R (1998) *Tropical Forest Census Plots*. Berlin, Germany: Springer-Verlag.

Dornelas M, Connolly SR, Hughes TP (2006) Coral reef diversity refutes the neutral theory of biodiversity. *Nature* **440**:80–2.

Etienne RS (2005) A new sampling formula for neutral biodiversity. *Ecol Lett* **8**:253–60.

Etienne RS, Alonso D (2007) Neutral community theory: how stochasticity and dispersal-limitation can explain species coexistence. *J Stat Phys* **128**:485–510.

Etienne RS, Apol MEF, Olff H, Weissing FJ (2007) Modes of speciation and the neutral theory of biodiversity. *Oikos* **116**:241–58.

Gravel D, Beaudet M, Messier C (2008) Partitioning the factors of spatial variation in the regeneration density of shade tolerant tree species. *Ecology* **89**:2879–88.

Hubbell SP (1997) A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. *Coral Reefs* **16**:S9–21.

Hubbell SP (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton, NJ: Princeton University Press.

Hubbell SP, Foster RB, O'Brien ST, *et al.* (1999) Light-gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. *Science* **283**:554–7.

Hubbell SP, Condit R, Foster RB (2005) *Barro Colorado Forest Census Plot Data*. https://ctfs.arnarb.harvard.edu/webatlas/datasets/bci.

Jabot F, Etienne RS, Chave J (2008) Reconciling neutral community models and environmental filtering: theory and an empirical test. *Oikos* **117**:1308–20.

Kingman JFC (1982) On the genealogy of large populations. *J Appl Probab* **19A**:27–43.

Leigh EG (2007) Neutral theory: a historical perspective. *J Evol Biol* **20**:2075–91.

Leigh EG, Rosindell J, Etienne RS (2010) Unified neutral theory of biodiversity and biogeography. *Scholarpedia* **5**:8822.

MacArthur RH, Wilson EO (1967) *The Theory of Island Biogeography*. Princeton, NJ: Princeton University Press.

McGill BJ (2003) A test of the unified neutral theory of biodiversity. *Nature* **422**:881–5.

McGill BJ, Maurer BA, Weiser MD (2006) Empirical evaluation of neutral theory. *Ecology* **87**:1411–23.

O'Dwyer JP, Lake JK, Ostling A, *et al.* (2009) An integrative framework for stochastic, size-structured community assembly. *Proc Natl Acad Sci USA* **106**:6170–5.

Pigolotti S, Cencini M (2009) Speciation-rate dependence in species-area relationships. *J Theor Biol* **260**:83–9.

Ricklefs RE (2003) A comment on Hubbell's zero-sum ecological drift model. *Oikos* **100**:185–92.

Rosindell J, Cornell SJ (2007) Species-area relationships from a spatially explicit neutral model in an infinite landscape. *Ecol Lett* **10**:586–95.

Rosindell J, Cornell SJ (2009) Species-area curves, neutral models and long distance dispersal. *Ecology* **90**:1743–50.

Rosindell J, Wong Y, Etienne RS (2008) Coalescence methods for spatial neutral ecology. *Ecol Inform* **3**:259–71.

Rosindell J, Cornell SJ, Hubbell SP, Etienne RS (2010) Protracted speciation revitalizes the neutral theory of biodiversity. *Ecol Lett* **13**:716–27.

Rosindell J, Hubbell SP, Etienne RS (2011) The unified neutral theory of biodiversity and biogeography at age ten. *Trends Ecol Evol* **26**:340–8.

Thomas SC (1996) Relative size at onset of maturity in rain forest trees: a comparative analysis of 37 Malaysian species. *Oikos* **76**:145–54.

Wright SJ, Jaramillo MA, Pavon J, *et al.* (2005) Reproductive size thresholds in tropical trees: variation among individuals, species and forests. *J Trop Ecol* **21**:307–5.