

Research

Adding stage-structure to a spatial neutral model: implications for explaining local and regional patterns of biodiversity

Ryan A. Chisholm and Tak Fung

R. A. Chisholm (<https://orcid.org/0000-0002-9847-1710>) ✉ (ryan.chis@gmail.com) and T. Fung (<https://orcid.org/0000-0003-1039-4157>), Dept of Biological Sciences, National Univ. of Singapore, Singapore, Singapore.

Oikos

130: 1976–1987, 2021

doi: 10.1111/oik.08489

Subject Editor:

Isabel M. Smallegange

Editor-in-Chief: Dries Bonte

Accepted 12 August 2021

Theoretical ecologists have analysed a range of neutral models but few including stage structure. Here we introduce a stage-structured neutral model, by extending the standard spatial neutral model to have two-stage classes: a juvenile stage and a reproductive stage. We find that formulas for biodiversity patterns (e.g. species–area relationships and species abundance distributions) of reproductives in the stage-structured model can be obtained from the corresponding standard formulas via a parameter rescaling, which involves calculating an effective speciation rate parameter and an effective dispersal parameter. This is useful because it means existing knowledge about the non-stage-structured model can be transferred to the stage-structured model, providing that applications focus exclusively on the reproductive stage. One surprising implication is that the presence of a juvenile stage can substantially increase the species richness of reproductive individuals: a juvenile stage with a length fraction k that of the reproductive stage increases reproductive species richness by roughly the same factor. We apply our new formulas to a case study of tropical forest trees in Panama and find that while the stage-structured model makes different predictions than the standard model, it does not fix known problems with cross-scale predictions. We speculate that some of our results, in particular the result that the presence of a juvenile stage increases community diversity, likely apply to non-neutral systems as well.

Keywords: neutral model, spatial coalescence, species–area relationship, stage structure, tropical forest trees

Introduction

Initial enthusiasm for ecological neutral theory in the early 2000s was driven largely by its ability to accurately reproduce classic empirical biodiversity patterns, including the triphasic species–area relationship (SAR) and the log-series species abundance distribution (SAD) (Hubbell 2001, Volkov et al. 2003, Rosindell and Cornell 2007). This enthusiasm waned with the gradual revelation that the theory performed substantially less well when predicting patterns of temporal change (Chisholm and O'Dwyer 2014, Chisholm et al. 2014, Kalyuzhny et al. 2014) or when attempting to simultaneously reproduce patterns at local and regional scales (O'Dwyer and Cornell 2018).



www.oikosjournal.org

© 2021 Nordic Society Oikos. Published by John Wiley & Sons Ltd

Evidently, important processes that operate in nature are missing from neutral theory – or at least from most neutral models explored to date.

Mechanisms missing from most published neutral models can be classified into four broad categories: realism in the dispersal process (Rosindell and Cornell 2009), realism in the speciation process (Rosindell et al. 2010, Etienne and Haegeman 2011), niche-based mechanisms (Kalyuzhny et al. 2015, Cazelles et al. 2016, Chisholm et al. 2016, Fung et al. 2016), and mechanisms associated with demographic heterogeneity (Melbourne and Hastings 2008). The latter refers to variation in vital rates of individuals within a species by size, age or stage. Some types of demographic heterogeneity can be introduced without breaking neutrality, e.g. if the same size- or age-structured variation is present for all species. For example, O’Dwyer et al. (2009) added size structure to a neutral model and found that it had little effect on the SAR but did influence other system properties, such as the species–biomass distribution. The focus of the present study will be on how stage structure – the presence of multiple life stages with different fecundity and mortality rates – affects neutral predictions.

Perhaps the most basic form of stage structure in nature is the presence of separate juvenile and reproductive stages. In applications to tree communities, the default approach in most applications of neutral models has been to make the assumption (usually unstated, but see Hubbell et al. 2008) that all individuals above a certain diameter-at-breast-height (DBH) threshold are reproductive, and to ignore the juvenile stage completely (Rosindell et al. 2012). The chosen threshold is typically 1 or 10 cm (each is used at various points in Hubbell 2001). This approach might be justifiable if results were robust to the choice of threshold, but in general they are not (Jabot et al. 2008, Rosindell et al. 2012). For example, some quick calculations show that changing the DBH threshold has major effects on the SAR. The spatial neutral SAR formula of O’Dwyer and Cornell (2018) can be used to predict tree species richness in the 50-ha Barro Colorado Island (BCI) plot based on regional richness. If this formula is parameterised at the regional scale for trees above 10 cm DBH, with the implicit assumption that this is the reproductive threshold, then the formula predicts about 125 tree species with individuals above this threshold in the 50-ha plot, just over half the true value (Table 1) (O’Dwyer and Cornell 2018). If instead 1 cm DBH is chosen as the threshold, the formula predicts roughly 600 species above this size in the plot, or double the empirical value (Table 1; note that regional richness differs with the DBH threshold, but this accounts for only $\approx 20\%$ of the plot-scale difference). The seemingly innocuous choice between two thresholds causes a switch from severe underprediction to severe overprediction. The use of a threshold, in most cases without justification, is clearly problematic.

A more justifiable approach is to count the individuals that are actually reproductive and use these numbers when parameterising the model. For example, Preston’s (1960) classic bird SAR data set comes from counts of breeding

Table 1. Predicted local tree species richness in the BCI 50 ha plot from a regionally parameterised (non-stage-structured) spatial neutral model.

Input parameters ¹			
Reproductive DBH threshold ²	Regional richness ³	Predicted local richness ⁴	True local richness ⁵
1 cm	955	573	299–307
1 cm	1260	670	299–307
10 cm	716	124	222–238
10 cm	788	127	222–238

¹We also assume that the dispersal parameter is $\sigma=50$ m, and that the area of the regional scale is 2500 km² (corresponding roughly to the Panama Canal watershed).

²The 1 and 10 cm thresholds lead to respective density estimates of 400 000 and 40 000 reproductive individuals per km².

³The regional richness numbers are based on data from a regional network of small plots and inventories (Condit et al. 2001, 2005a, Pyke et al. 2001).

⁴The predicted local species richness values come from plugging the input parameters into Eq. 3 of O’Dwyer and Cornell (2018), with the speciation parameter ν fitted to match regional richness.

⁵The ranges of true local richness values come from the seven censuses of the plot between 1982 and 2010.

pairs, and the neutral SAR provides an excellent fit to these (Fig. 1) (Rosindell and Chisholm 2020) (albeit with one more free parameter than in the application to trees above, in the absence of an independent estimate of the dispersal

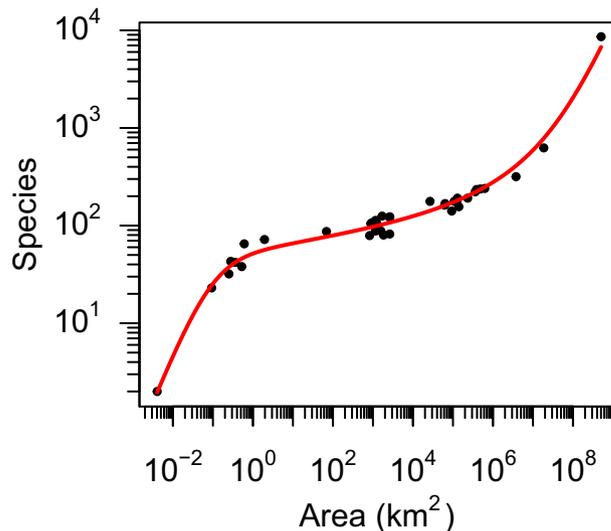


Figure 1. Species–area data for birds of North America from Preston (1960) (points), along with a fit of the standard spatial neutral model without stage structure (O’Dwyer and Cornell 2018) (red curve). The fitted parameter values are individual density $\rho=507$ km⁻², speciation rate $\nu=8.16 \times 10^{-10}$ and standard deviation of dispersal distance $\sigma=0.4$ km ($R^2=0.982$). The species counts in Preston’s data are for breeding pairs, i.e. individual density ρ is measured in units of breeding pairs. Note the graph differs slightly from the one in Rosindell and Chisholm (2020): here we more closely follow Preston’s (1960) approach where two small-area urban sites are omitted and the small-area point instead comes from the forested Neotoma Valley in Ohio (where one acre contains roughly two species).

parameter for birds). If, as in this birds example, the data do not contain any juveniles, there still remains the problem of how the juvenile stage might affect the overall dynamics of the neutral model and thus the interpretation of fitted parameters. For organisms with a short juvenile stage, it can perhaps be safely ignored, as it has been in most neutral models to date. But many species, including birds (Hickey 1952) and trees (Hubbell 1998, Metcalf et al. 2009), can spend a good fraction of their lives as juveniles.

A spatial version of a neutral model with two stage classes – juvenile and reproductive – was discussed by Rosindell et al. (2012). They conjectured that the presence of a juvenile stage would have little effect on the population dynamics of the reproductives, relative to the case without stage structure. They therefore ran a standard spatial neutral model without stage structure, i.e. with only reproductives, and created a static distribution of juveniles around each reproductive. They used this to explore how an imperfect sampling process that includes not just reproductives but also some juveniles can affect the SAR. A fully dynamic stage-structured model was outside the scope of their study, but the implicit assumption was that such a model would produce similar results. By contrast, D’Andrea and O’Dwyer (2017) did study a fully dynamic stage-structured neutral model in a non-spatial context and found that the SAD could exhibit marked deviations from the non-stage-structured expectation if adult fecundity and mortality were low, but that the progeny distribution was more robust.

Here we build on this previous work by quantitatively analysing a spatial neutral model with two-stage classes. Our stage-structured model is similar to the one Rosindell et al. (2012) discussed verbally without exploring quantitatively the juvenile dynamics. Our specific goals are to explore the effects of stage structure on the neutral model’s SAR and on its ability to make accurate cross-scale predictions. Pleasingly, in some cases, one can derive formulas for translating results from a non-stage-structured model to a stage-structured model, but these must be used carefully because the parameters gain different interpretations and the values of individual density must be chosen appropriately. We show with an application to Panama forest trees that this improved approach to neutral modelling of stage-structured communities changes conclusions on when and how neutral theory fails, in particular, whether the neutral theory can simultaneously predict species richness at local and regional scales via the SAR.

Methods

The model

The mathematical model concerns a non-zero-sum community on an infinite landscape, with a focal disc-shaped region of area A . The average densities of juvenile and reproductive individuals per unit area are ρ_{juv} and ρ_{repr} , respectively, and thus the corresponding expected numbers of individuals in the

focal area are $J_{\text{juv}} = \rho_{\text{juv}}A$ and $J_{\text{repr}} = \rho_{\text{repr}}A$. Individuals can exit the juvenile stage either by being promoted to reproductive status or dying, processes which occur at per-capita rates c and d_{juv} (per unit time), respectively. Individuals exit the reproductive stage only via death, which occurs at a per-capita rate d_{repr} . Additionally, new juveniles appear in the landscape as a result of two types of event: birth events among the reproductives, which occur at a per-capita rate $b(1 - \nu)$, with $\nu \in [0, 1]$; and speciation events which occur at a per-unit area rate $\rho_{\text{repr}}b\nu$. To maintain constant average population densities of both stage classes, we enforce the constraints $b\rho_{\text{repr}} = (c + d_{\text{juv}})\rho_{\text{juv}}$ and $c\rho_{\text{juv}} = d_{\text{repr}}\rho_{\text{repr}}$. (Technical note: Because we are dealing with a non-zero-sum model, the speciation rate is defined per-unit area rather than per-individual to guarantee long-term persistence of the community, which otherwise has an absorbing state with zero individuals. Most of the time, though, the system will be close to the dynamic equilibrium, and the parameter ν can be thought of as the probability of speciation during a birth event, as in the zero-sum model.)

Observe that the average time spent in the juvenile stage is $\tau_{\text{juv}} = 1/(c + d_{\text{juv}})$ and the average time spent in the reproductive stage is $\tau_{\text{repr}} = 1/d_{\text{repr}}$. Putting all this together, we get

$$\begin{aligned} c &= \frac{1}{\tau_{\text{repr}}} \frac{\rho_{\text{repr}}}{\rho_{\text{juv}}} \\ b &= \frac{1}{\tau_{\text{juv}}} \frac{\rho_{\text{juv}}}{\rho_{\text{repr}}} \\ d_{\text{juv}} &= \frac{1}{\tau_{\text{juv}}} - \frac{1}{\tau_{\text{repr}}} \frac{\rho_{\text{repr}}}{\rho_{\text{juv}}} \\ d_{\text{repr}} &= \frac{1}{\tau_{\text{repr}}} \end{aligned} \tag{1}$$

These formulas can be useful for parameterising the model with empirical data because they involve only estimates of individual density and time spent in each stage.

When a reproductive individual gives birth, the new individual disperses away from the parent with a distance and direction determined by drawing from a bivariate normal dispersal kernel with zero mean, zero covariance terms and variances equal to σ_d^2 . When a juvenile is promoted to the reproductive stage, it can change spatial position, and the distance and direction it moves is again drawn from a bivariate normal kernel with zero mean and zero covariances, and with variances equal to σ_g^2 . In some communities this distance will be small or negligible, e.g. in tree communities, a reproductive tree will be at essentially the same location that it occupied as a juvenile. In other communities, though, the distance could be appreciably large, e.g. for birds, it represents the distance a fledgling travels before establishing its own nest.

We analysed this model mathematically and also via numerical simulations. The definition of the model given

above is non-zero-sum, which facilitates the application of related published analytical solutions (O'Dwyer and Cornell 2018). For the simulations, we created a zero-sum version of the model because this is easier to implement numerically. The individuals in the zero-sum simulation exist on two spatially coincident grids that generally have different resolutions (equal only if $\rho_{\text{juv}} = \rho_{\text{repr}}$), with the juveniles on one grid and reproductive individuals on another. Each reproductive death is immediately replaced by the promotion of a juvenile into the vacated grid cell, and each juvenile death or promotion is immediately replaced by a birth event from a reproductive, with speciation occurring with probability ν per new individual produced. The zero-sum assumption introduces a weak form of density dependence into the model, which can be important if diversity is very low (just a few species), but the zero-sum and non-zero-sum models produce almost identical results otherwise (Chisholm and O'Dwyer 2014).

We coded two versions of the zero-sum simulation in R using different algorithms: one that runs forwards in time, and one that runs backwards in time using coalescence methods (Rosindell et al. 2008, Thompson et al. 2020). The backwards-in-time algorithm has two major advantages: it runs much faster than the forwards-in-time version; and it can run on effectively infinite landscapes (limited only by numerical precision) consistent with our original model definition above. Our backwards-in-time algorithm differs from the forwards-in-time algorithm described above only in that we allow two or more lineages in the juvenile stage to simultaneously occupy the same grid cell because there was no way to prevent this in general without keeping track of all lineages in the infinite landscape. This assumption does not matter in the limit of high juvenile density, and fortunately, in practice, it appears not to matter providing that juvenile density is equal to or greater than reproductive density ($\rho_{\text{juv}} \geq \rho_{\text{repr}}$; Fig. 2a). With this justification, we used the backwards-in-time algorithm for the case study below, where empirically we have $\rho_{\text{juv}} \gg \rho_{\text{repr}}$.

Case study: tropical forest trees in Panama

We parameterised the model for tropical forest trees in central Panama. The key parameters of the model are the densities ρ_{juv} and ρ_{repr} , the mean times spent in each stage τ_{juv} and τ_{repr} , the variances of the dispersal and recruitment distances σ_d^2 and σ_g^2 and the speciation probability ν , which controls landscape-scale diversity.

We estimated reproductive tree density at the 50 ha permanent plot at BCI, where each stem greater than 1 cm DBH has been censused (Condit et al. 2017), using census data together with unpublished species-specific average DBHs of reproduction onset for 278 species estimated by Robin Foster (Wright et al. 2005, Muller-Landau et al. 2008). By using different DBH reproductive thresholds for each species and a fixed juvenile-to-reproductive transition rate for all species, we effectively assume that species with larger reproductive DBH thresholds grow faster. We acknowledge this as a limitation of our model, necessary to maintain parsimony and tractability, and encourage future work that models DBH dynamics explicitly.

Juvenile density is harder to estimate because it should include all pre-reproductive stages including seeds, seedlings, and saplings. Simulation results suggested, however, that the dynamics of reproductives are largely insensitive to the value of juvenile density, provided it is not lower than reproductive density, in which case juvenile density can be approximated as $\rho_{\text{juv}} = \infty$. The assumption that juvenile density is greater than reproductive density is easily met at BCI: for most species, the density of individuals below the reproductive threshold is several times higher than that above.

To parameterise the reproductive stage dynamics, we estimated the annual mortality rate of reproductives $d_{\text{repr},i}$ for each species i from the census data. Recall that the mean reproductive stage length is $\tau_{\text{repr},i} = 1/d_{\text{repr},i}$ (Eq. 1). We then took the median of these values across species (the distribution was positively skewed) as the value of τ_{repr} for the model.

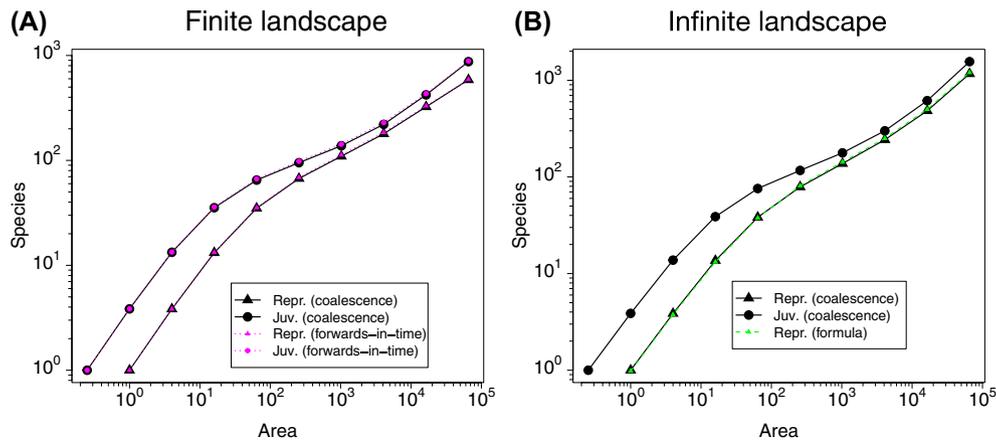


Figure 2. Theoretical SARs for reproductive (triangles) and juvenile (circles) individuals in the stage-structured neutral model on a square focal landscape with $J_{\text{repr}} = 2^8 \times 2^8$, $\rho_{\text{repr}} = 1$ (so that there is one reproductive individual per unit area) and $\rho_{\text{juv}} = 4$. Other parameters are $\sigma_d = 5$, $\sigma_g = 3$, $\nu = 10^{-3}$, $k = \tau_{\text{juv}}/\tau_{\text{repr}} = 1/3$. Each data point shows the average across 10 independent simulations. (a) Simulations on a finite landscape verify that the two simulation algorithms – coalescence (black) and forwards-in-time (magenta) – give consistent results. (b) On an infinite landscape, Eq. 5 gives a very accurate approximation to the SAR for reproductive individuals (green), as verified here by comparison to results from the coalescence algorithm (black).

To parameterise the juvenile stage dynamics, we estimated the mean juvenile stage length τ_{juv} from census data. Note that this value of τ_{juv} was estimated assuming that the mortality rate is the same for all juveniles. In reality, this mortality tends to decrease with age. We acknowledge this as a limitation of our model, which could potentially be addressed by splitting the juvenile stage class (e.g. into seedling and sapling classes), but in the Discussion we explain that this would not change our results qualitatively.

Our definition of the juvenile stage includes the seedling and the sapling stages, so we can write

$$\tau_{\text{juv}} = \tau_{\text{seedling}} + p_{\text{seedling} \rightarrow \text{sap}} \tau_{\text{sap}} \quad (2)$$

with the threshold for the seedling-to-sapling transition being 1 cm DBH to coincide with the size at which individuals enter the census data. We are ignoring the seed stage here because we assumed the seed bank is negligible in a tropical forest (Dalling and Brown 2009). To estimate τ_{seedling} , we took the estimated $m=0.13$ annual mortality rate of seedlings in the plot (Comita et al. 2010) and Hubbell's (1998) estimate of $T=16.6$ years for the average time taken for a seedling to reach 1 cm DBH in the BCI forest, and thus estimated τ_{seedling} and $p_{\text{seedling} \rightarrow \text{sap}}$ from the following formulas:

$$\tau_{\text{seedling}} = T(1-m)^T + \sum_{t=0}^{T-1} tm(1-m)^t = \frac{(1-m)(1-(1-m)^T)}{m}$$

$$p_{\text{seedling} \rightarrow \text{sap}} = (1-m)^T$$

The first term in the first expression for τ_{seedling} accounts for seedlings that successfully make it to the sapling stage; the second term accounts for seedlings that die before they reach the sapling stage. More sophisticated estimates of these quantities would account for variation in τ_{seedling} across individuals. We then estimated time spent in the sapling stage, $\tau_{\text{sap},i}$ for each species i by sampling randomly from the observed species-specific DBH growth and mortality rate data (from the BCI censuses) to estimate how long it takes a sapling, starting from 1 cm DBH, to either die or reach its species-specific reproductive DBH. This then allowed us to calculate $\tau_{\text{juv},i}$ for each species i from Eq. 2 and hence $k_i = \tau_{\text{juv},i} / \tau_{\text{repr},i}$ and ultimately a median value of k across species. We also explored the hypothetical limiting scenario in which $\tau_{\text{juv}} = 0$ and thus $k = 0$.

We estimated the dispersal variance σ_d^2 from seed trap data at BCI (Muller-Landau 2001). We set the juvenile-to-reproductive movement variance parameter to the arbitrarily low value of $\sigma_g^2 \approx 1 \text{ m}$, in which case it has a negligible effect on the results (this could represent trunk plasticity; Strigul et al. 2008).

The speciation parameter ν is then the only unknown in the model and can be determined by any single further piece of information. To fix ν , we used estimated values of

species richness at the scale of the Panama Canal watershed, of area $A=2500 \text{ km}^2$ (Pyke et al. 2001, Condit et al. 2005b). We then explored whether the resulting fully parameterised SAR accurately predicted species richness of the BCI 50 ha plot and whether the corresponding SAD at the 50 ha scale accurately matched the BCI SAD. This largely follows the procedure of O'Dwyer and Cornell (2018), who conducted a similar exercise but for a non-stage-structured version of our model. In addition, we examined whether the model accurately captures the distance decay of similarity at the 50-ha scale and below, as measured by the probability that two reproductive individuals separated by a distance r come from the same species, $F(r)$ (Chave and Leigh 2002). This largely follows the procedure of Condit et al. (2002), who looked at diversity decay in a non-stage-structured version of our model.

Our results revealed that the regionally parameterised model could not accurately predict local richness. To calculate what regional richness would hypothetically need to be if the model were to match local scale richness, we also explored the SAR when ν is instead fixed using observed species richness at the 50 ha scale.

Results

The model

Initial investigations indicated that the SAR for all individuals (juveniles plus reproductives) in the stage-structured model is not a simple rescaling of the known formulas for the non-stage-structured version of the model (Fig. 2a shows the SAR decomposed into its juvenile-stage and reproductive-stage components). Interestingly, however, a formula for just the reproductive-stage SAR can be obtained by thinking about the model's dynamics in a coalescence framework. Suppose we start with a set of lineages corresponding to all reproductive individuals in the focal area at the present day and trace these lineages backwards through time, which involves first tracing the lineage for each focal reproductive individual back to its juvenile stage, then tracing this back to a previous reproductive individual, and so on. The dynamics are similar to the coalescence process in the non-stage-structured model except that lineages now can coalesce only when stepping backwards from the juvenile stage to the reproductive stage, not when stepping backwards from the reproductive to the juvenile stage. The biological reason is simply that two reproductive individuals cannot come from the same juvenile, whereas two juveniles can come from the same reproductive individual.

Following the above logic, we can infer that the rate of coalescence will be a factor $\tau_{\text{repr}} / (\tau_{\text{repr}} + \tau_{\text{juv}}) = 1 / (1 + k)$ slower in the stage-structured model, where $k \equiv \tau_{\text{juv}} / \tau_{\text{repr}}$. However, after accounting for this effective stretching of time, the coalescence tree should be statistically identical to one from the standard non-stage-structured model. We can account for this time-stretching mathematically and convert the stage-structured model (for reproductives only, recall) to

an equivalent non-stage-structured model by calculating an effective speciation rate and an effective dispersal distance:

$$\nu_{\text{eff}} = 1 - (1 - \nu)^{1+k} \approx (1+k)\nu \quad (3)$$

$$\sigma_{\text{eff}}^2 = (1+k)(\sigma_d^2 + \sigma_g^2) \quad (4)$$

where k is as defined above. The logic here is that the effective stretching of time by a factor of $1+k$ increases both the effective speciation rate and the effective dispersal distance by approximately a factor of $1+k$ (the average distance travelled over time in a Gaussian random walk is linearly related to the variance of each step). Note that for the effective speciation rate, the exact value is slightly less than $(1+k)\nu$ to account for the fact that tracing of a lineage in the coalescence algorithm is terminated when the first speciation event occurs (so there cannot be more than one speciation event on a lineage). Note also that the formula for σ_{eff}^2 involves σ_g^2 as well as σ_d^2 accounting for the possibility that an individual's position moves from the juvenile to the reproductive stage (which is essentially irrelevant to trees but potentially relevant to other applications).

We emphasise that Eq. 3 and 4 should not be taken to mean that speciation and dispersal are actually somehow greater in the stage-structured model than the specified values ν and σ_d^2 , but rather that the resulting model behaves like a non-stage-structured model in which ν_{eff} and σ_{eff}^2 are the speciation rate and dispersal variance.

Armed with the rescaling formulas, we can write the SAR for reproductive individuals in the stage-structured model as

$$S = (1+k)(\sigma_d^2 + \sigma_g^2) \Psi \left(\frac{\rho_{\text{repr}} A}{(1+k)(\sigma_d^2 + \sigma_g^2)}, 1 - (1-\nu)^{1+k} \right) \quad (5)$$

where $\Psi(A, \nu)$ is the SAR in a spatial non-stage-structured neutral model with $\rho=1$ and $\sigma=1$. The function Ψ has elsewhere been named the Preston function, for its broad utility in spatial neutral ecology (Chisholm et al. 2018, Thompson et al. 2019, Rosindell and Chisholm 2020). An excellent approximation to Ψ has been derived by O'Dwyer and Cornell (2018). The good match between Eq. 5, using O'Dwyer and Cornell's formula for Ψ , and simulation data are shown in Fig. 2b.

Linearising Eq. 5 around $k=0$ shows that species richness is roughly proportional to $1+k$ for small k . This result can in fact be derived from the first principles as follows. Consider a two-stage neutral community (juveniles and reproductives) of constant size at dynamic speciation–extinction equilibrium. For ease of exposition let the units of time be years, and let the average number of new species arising from speciation per year be N (uppercase ν). Over a long period t of geological time, the total number of species seen will be roughly equal

to the total number of speciation events, Nt (this ignores only species that arose before the start of the interval but persisted into it, a number that becomes negligible relative to Nt for large t). If T is the average species lifetime, then the total number of species-years is NtT . An alternative formula for the total number of species-years is St , where S is the average number of species alive at any one time. Thus we have $St=NtT$, showing that $S=NT$, i.e. average species richness is linearly related to the average species lifetime.

Using the same rescaling (Eq. 3–4) used to derive the SAR (Eq. 5) in our stage-structured model, we can write a formula for the SAD for reproductive individuals in our model by substituting ν_{eff} and σ_{eff}^2 for ν and σ^2 into a corresponding formula for the SAD in the spatial non-stage-structured model (O'Dwyer and Cornell 2018). And we can also write formulas specifying $F(r)$ for reproductive individuals in our model by applying the same substitution to the corresponding formulas for $F(r)$ in the spatial non-stage-structured model (Eq. 4.7 and 4.8 in Chave and Leigh 2002, or equivalently Eq. 1 and 2 in Condit et al. 2002).

Case study: tropical forest trees in Panama

From the average of seven censuses at BCI, we estimated $\rho_{\text{repr}} = 152\,932 \text{ km}^{-2}$. As noted in the Methods, juvenile density was set at $\rho_{\text{juv}} = \infty$. The median time spent in the reproductive stage across all species was estimated as $\tau_{\text{repr}} = 55.9$ years. We estimated $\tau_{\text{seedling}} = 6.1$ years and $p_{\text{seedling} \rightarrow \text{sap}} = 0.12$. The median time spent in the sapling stage (given that an individual reaches this stage) across species was estimated as $\tau_{\text{sap}} = 26.8$ years, and the corresponding median time spent in the whole juvenile stage was $\tau_{\text{juv}} = 9.3$ years (this is lower than τ_{sap} because most juveniles die as seedlings, before they reach the sapling stage). The median ratio of time spent in the juvenile stage to time spent in the reproductive stage across species was $k = \tau_{\text{juv}}/\tau_{\text{repr}} = 0.18$. Similar values for k were obtained when we instead divided the median juvenile stage length by the median reproductive stage length ($k=0.17$) or calculated an abundance weighted value ($k=0.16$). We performed analyses for $k=0.1$ and $k=0.2$ to cover an envelope that includes the empirically estimated values. The dispersal parameter was estimated as $\sigma_d = 50$ m, based on dispersal data for 81 tree species at BCI (Muller-Landau 2001, Chisholm and Lichstein 2009). With the assumed value of $\sigma_g = 1$ m, this gave $\sigma_{\text{eff}}^2 = (1+k) \times (50^2 + 1^2)$ and thence $\sigma_{\text{eff}} = 52.5$ or 54.8 m, depending on the value of k chosen.

A 'tree' for the purposes of this analysis is defined as any free-standing woody plant that can reach 1 cm DBH (coinciding with the census data at BCI). We used a range of values for tree species richness at the scale of the Panama Canal watershed: $S_{\text{PCW}} \in [955, 1260]$. The lower value in this range is the lower bound on species richness of trees above 1 cm DBH obtained from a network of small plots and inventories throughout the watershed (Table 1); the upper value is the upper bound from this same data source (Table 1) plus 20% to account for uncertainties including undersampling (e.g. singletons or other locally rare species that may exist in

Table 2. Predicted local species richness of reproductive trees (using species-specific reproductive DBH thresholds) in the BCI 50-ha plot from a regionally parameterised stage-structured spatial neutral model.

Input parameters ¹			
Ratio of juvenile to reproductive stage duration ($k = \tau_{\text{juv}}/\tau_{\text{repr}}$)	Regional richness (S_{PCW})	Predicted local richness (\hat{S}_{BCI})	True local richness ²
0.0	955	349	291–297
0.1	955	369	291–297
0.2	955	389	291–297
0.0	1260	384	291–297
0.1	1260	409	291–297
0.2	1260	432	291–297

¹In addition, we used the parameter values $\sigma_d = 50$ m and $\sigma_g = 1$ m, as described in the main text.

²This is the range of values observed over the seven censuses between 1982 and 2010.

the watershed without having been observed). The resulting downscaled estimates of tree species richness at the 50-ha plot scale, \hat{S}_{BCI} , are given in Table 2. For the empirically estimated values of k , the values of \hat{S}_{BCI} range from about 26–47% more than the observed values of $S_{\text{BCI}} \in [291, 297]$ over seven censuses (Table 2, Fig. 3). Thus the model overestimates local species richness. Even for the lowest value of regional richness considered and the lowest possible value of $k=0$, corresponding to the case of no juvenile stage and thus

a population without stage structure, local species richness is still overestimated by about 19% (Table 2). For the downscaling to give accurate results at the 50-ha scale, regional richness would have to be less than half its current estimated value ($S_{\text{PCW}} = 436, 465, \text{ and } 502$ for $k=0.0, 0.1, \text{ and } 0.2$, respectively; Fig. 3, grey curves).

The model also fails to predict the form of the SAD at the 50-ha plot scale: predicted interspecific variation in abundance is much lower than observed, with a deficiency of both very rare and very abundant species (Fig. 4). The results for the similarity index $F(r)$ reflect both the model's overprediction of species richness and underprediction of the proportion of very abundant species at local scales: in the regionally parameterised model, the predicted similarity at all distances within the 50-ha plot is a factor of 8–43 lower than the observed similarity (Fig. 5).

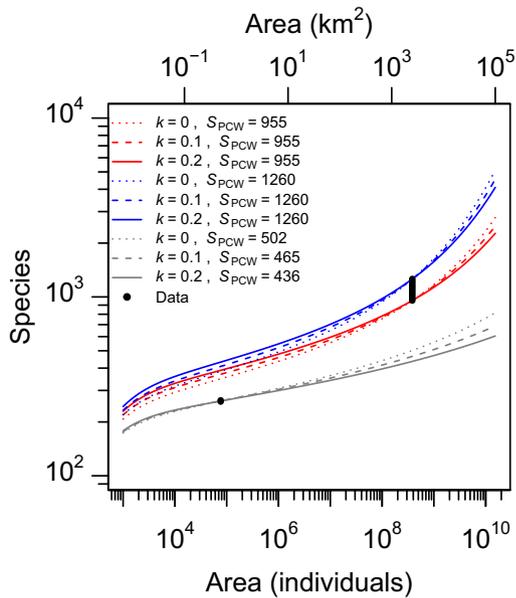


Figure 3. Theoretical SARs for reproductive individuals in a spatial stage-structured neutral model for forest trees in central Panama. The SARs show various combinations of the parameter k (ratio of juvenile to reproductive stage duration) and the regional species richness S_{PCW} (graphical legend and Table 2). The blue and red SARs are tuned to fit the regional scale (for which the empirically estimated range of species richness is indicated by the cluster of black dots at right); the grey SARs are tuned to fit the local scale of the 50-ha BCI plot (black dot at left). The regionally parameterised SAR overpredicts local richness, and the locally parameterised SAR underpredicts regional richness. The area on the lower axis is in units such that individual reproductive tree density $\rho = 1$; the area on the upper axis is in units of km^2 for ease of interpretation.

Discussion

Many ecological communities exhibit pronounced stage structure. When modelling such communities with neutral theory, one sensible approach is to build a stage-structured model that properly accounts for the different demographic rates of each stage. Here we have shown that a mathematically equivalent, and often easier, alternative is to use a reparameterised version of a corresponding non-stage-structured model in which individual density (ρ) is defined as the density of reproductives, and the speciation and dispersal parameters are reinterpreted as effective parameters. The rescaled model reveals surprising insights about the effects of stage structure on biodiversity, some of which go beyond neutral theory.

Perhaps our most surprising result is that the presence of a juvenile stage can have a substantial effect on the species richness of the reproductive stage. Assuming all other model parameters are held constant, the presence of a juvenile stage having length a fraction k that of the reproductive stage increases reproductive species richness by roughly the same factor (note the increase of richness with k in our case study in Table 2 is less than this because we constrain the model to fit regional richness and thus ν also varies with k). This effect can be large. For example, consider a hypothetical herbaceous plant community with two stages, a seed bank

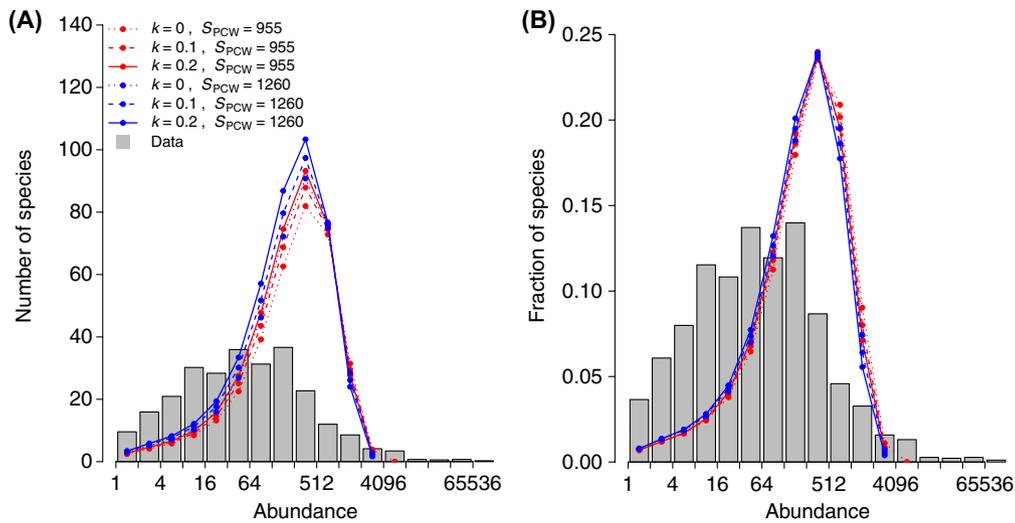


Figure 4. (a) Theoretical SADs (coloured curves with points) for reproductive individuals in the BCI 50-ha plot from the spatial stage-structured neutral model parameterised at the scale of the Panama Canal watershed as in Fig. 3. The six different theoretical SADs show different combinations of the input parameters k (ratio of juvenile to reproductive stage duration) and S_{PCW} (species richness at the watershed scale) (graphical legend and Table 2). The empirical SAD (averaged over seven censuses) is shown for comparison (grey bars). Total species richness at the 50-ha plot scale is overpredicted, as was evident from the SAR in Fig. 3, but, in addition, from the SADs we see that variation in species abundance across species is underpredicted. (b) The latter discrepancy is not due solely to the overprediction of total richness, as is made clear by plotting the proportional SADs instead.

stage and an adult stage, and assume that the average lengths of the two stages are equal, and imagine we want to predict species richness with a neutral model. If we have $J = 10\,000$

individuals, a dispersal distance of $\sigma = 4$ (in units of distance defined such that individual density per unit area equals one), and a speciation rate of $\nu = 10^{-8}$, then naively ignoring the

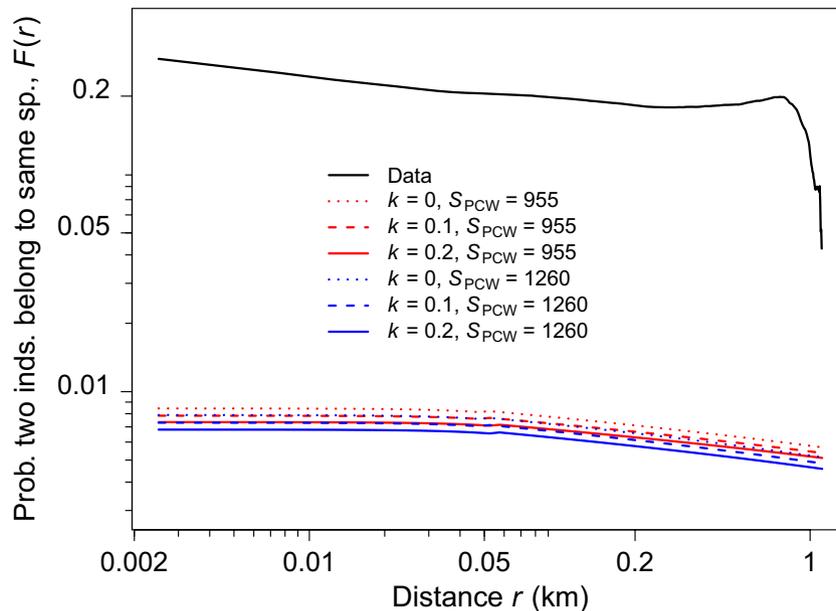


Figure 5. Theoretical similarity (coloured curves with points) for reproductive individuals in the BCI 50-ha plot from the spatial stage-structured neutral model parameterised at the scale of the Panama Canal watershed as in Fig. 3. The similarity is measured as the probability of two reproductive individuals separated by a distance of r belonging to the same species, $F(r)$. For ease of visual representation of the empirical curve, we binned the r values for all pairs of individuals into bins of width 5 cm, calculated the mean $F(r)$ for each bin, and then plotted these against the mid-points of the bins. The six different theoretical curves of $F(r)$ against r show different combinations of the input parameters k (ratio of juvenile to reproductive stage duration) and S_{PCW} (species richness at the watershed scale) (graphical legend and Table 2). The empirical curve of $F(r)$ against r (averaged over seven censuses) is shown for comparison (black curve).

juvenile stage by setting $k = 0$ gives estimated species richness $S = 19.0$, but accounting for it by setting $k = 1$ gives $S = 37.6$, a nearly two-fold difference.

These potentially large effects of the juvenile stage on reproductive community diversity run counter to our prior expectations and to those of a previous study on stage-structured neutral models (Rosindell et al. 2012). Rosindell et al. (2012) correctly pointed out that a juvenile stage could affect richness in two ways: 1) via direct inclusion of juveniles in the sample and 2) via the indirect effect of juveniles on reproductive community dynamics, which amounts to a time lag. But they argued that (2) would be negligible, which we now see is not the case in general. One way to understand how the presence of a juvenile stage increases community diversity is to note that it leads to longer species lifetimes but does not affect the total number of speciation events over a fixed period of evolutionary time, and the only way to reconcile these two facts is to have more species concurrently alive at any one time (see the mathematical arguments in the paragraph following Eq. 5).

To illustrate this more vividly, consider an extreme hypothetical example of a forest plot where the average juvenile stage length is 10 million years, implying that the ancestors of the current crop of reproductive trees are drawn from individuals that have lived in the area over tens of millions of years, rather than over just the last few centuries as would be typical for a real forest. Obviously, the current diversity will be higher in our hypothetical scenario than in reality (because we will sample trees representing diverse species from different geological ages), even though we have not changed the overall speciation rate. Our hypothetical example is extreme but serves to provide a general intuition for how the presence of a juvenile stage can increase the diversity of reproductives by stretching the fingers of the coalescence process deeper into the past. Importantly, this general insight is not dependent on the specific neutral assumptions of our model, and thus likely applies beyond neutral theory.

In the application to BCI, we found that parameterising the model at the scale of the Panama Canal watershed ($\approx 2500 \text{ km}^2$) results in overestimation of tree species richness at the local 50-ha scale by 26–47%. This is opposite to the conclusion reached by O'Dwyer and Cornell (2018), and the different result is attributable to two features of the stage-structured model. The first is that we estimate the density of reproductive individuals using published species-specific DBH thresholds from BCI, whereas O'Dwyer and Cornell (2018) implicitly assumed that the reproductive trees are all trees greater than 10 cm DBH (as have, it should be noted, most other applications of neutral theory to these data). This change alone is enough to switch the regionally parameterised model from under-predicting to slightly over-predicting species richness in the 50-ha plot (rows with $k = 0$ in Table 2). The second distinguishing feature of our model is the presence of a juvenile stage. The key parameter is the aforementioned ratio k . The higher this ratio, the higher the species richness, and thus in the BCI application, the worse the over-prediction problem (Table 2). For BCI, future empirical

research can refine the estimate of k , for example by incorporating the seed stage (Dalling and Brown 2009) and using the maximum likelihood models of Condit et al. (2017) to estimate mortality and transition rates for each stage and thus the length of each stage. But even if k turns out to be lower than estimated here, overprediction of species richness will remain an issue, as evidenced by the $k = 0$ case in Table 2. In our BCI case study, the effect of more accurately counting reproductives on estimated local richness was substantially larger than the effect of explicitly modelling juvenile stage dynamics. However, the latter effect is a more novel and surprising result and can be expected to be more important in communities with longer juvenile stages, as we noted above with reference to a hypothetical herbaceous plant community.

Another critical insight from the stage-structured model is that if the dispersal parameter σ is estimated by fitting the non-stage-structured SAR formula to empirical SAR data (or by fitting to related spatial patterns), the resulting value cannot be interpreted literally as the standard deviation of dispersal distance but instead must be interpreted as an effective parameter that also incorporates the parameter k , i.e. the ratio of juvenile to reproductive stage duration. In our case study, the estimated standard deviation of dispersal distance for forest trees in Panama was 50 m, but the effective dispersal distance σ_{eff} was closer to 54 m. This is not to suggest, of course, that trees actually disperse further than field measurements suggest. Rather, the stage-structured community behaves like a non-stage-structured community with $\sigma \approx 54 \text{ m}$. Following intuitions about the coalescence process outlined earlier, we can see that this is because in a stage-structured community the new recruits are drawn not just from the current crop of reproductives but also indirectly from reproductive individuals that may have died some time ago, via the 'buffer' of the juvenile stage class. The diversity of this pool of potential parents is thus greater than in a non-stage-structured model, and the mathematics must somehow take account of this. Instead of doing it explicitly by directly reaching back in time, the topological symmetries of coalescent trees from the neutral model allow us to incorporate this extra diversity by inflating the dispersal parameter and thus drawing recruits from a larger spatial region. This is the fundamental reason that $\sigma_{\text{eff}} > \sigma$.

Revisiting Preston's (1960) breeding bird data (Fig. 1), we can now appreciate that we should not take the fitted dispersal value of 0.4 km literally because it is an effective parameter (σ_{eff}). To compute the actual estimate of per-generation standard deviation of dispersal $\left((\sigma_d^2 + \sigma_g^2)^{1/2} \right)$ from Eq. 4 we need to estimate the length of juvenile and reproductive stages for birds. If we assume the mean length of the juvenile stage is 1–4 years while that of the reproductive stage is 3–5 years (Hickey 1952), we get estimates of per-generation standard deviation of dispersal of roughly 0.3 km, that is somewhat lower than the effective value. We emphasise that the discrepancies would be even larger for species with longer juvenile stages.

What to make of the result from our BCI case study that the neutral model overpredicts local species richness (Fig. 3)?

Of course, we must avoid overinterpreting the fit (or lack of such) to a single observation. Ideally, we would have multiple plots at the local scale from which to estimate local richness. But, absent such data, let us hypothesise that the overprediction is a general phenomenon. One implication is then that the most important omission from the neutral model is not niche stabilising mechanisms, but competitive exclusion that would reduce local species richness. The environment varies substantially across the Panama Canal Watershed, being much wetter on the Caribbean side than on the Pacific side of the isthmus (Condit et al. 2001). We can expect that at any given location, some species will be better adapted than others and hence have higher intrinsic population growth rates – in violation of the neutral assumption. Adding in these mechanisms to the model would make it more realistic and lead to lower local species richness.

It may be prudent to revisit other past applications of neutral theory to tropical forest trees, which, in light of our results, have not dealt with individual density and stage structure appropriately. Many of the results may not be robust, as we found here for the spatial downscaling to predict local species richness in O'Dwyer and Cornell (2018). Yet other results may be. For example, Condit et al. (2012) found that the empirically measured input rate of new species to the BCI plot over time was consistent with the prediction of neutral theory, using a threshold DBH of 1 cm. Repeating that analysis with a 10 cm threshold again gives results consistent with neutral theory (R. Condit pers. comm.). Thus, it would be reasonable to expect that repeating the analysis with reproductive trees only (which gives an intermediate density) would also give results consistent with neutrality. Also, another result of O'Dwyer and Cornell (2018), stating that spatial neutral theory predicts an overly narrow local SAD, does appear to be robust: whereas their model was parameterised with trees above 10 cm DBH, a very similar phenomenon arises in our model parameterised with reproductive trees (Fig. 4).

A broader lesson is that accounting for demographic heterogeneity can have surprisingly large effects on the biodiversity patterns predicted by neutral models. As noted earlier, the addition of a juvenile stage of length a proportion k that of the reproductive stage increases species richness by a similar proportion – a result we did not anticipate. Future studies could explore the effects of other kinds of demographic heterogeneity, such as the presence of more than two stages. One extension, relevant to tropical forest trees, would be to split the juvenile stage into separate seedling and sapling stages, which would be consistent with empirical data because these stages have different mortality rates (roughly 13% versus 4% per year for seedlings versus sapling, respectively). We predict that in this modified model the average juvenile stage length conditional on having reached reproductive age would be higher and, thus, following the intuition outlined earlier, community diversity would be even higher relative to the non-stage-structured case because the coalescence process would be able to reach back further in time. Though such an exercise would

certainly be worthwhile, increasing diversity would actually exacerbate the poor fit to the Panama tree data (Fig. 3), and thus would not qualitatively change our results for this case study. We suspect that rescaling formulas, similar to the ones presented here, could also be derived to handle such multi-stage scenarios and even perhaps a non-neutral versions of our models, thus relating them to known mathematical results. This provides an interesting avenue for future work.

Another extension involving demographic heterogeneity would be to allow reproductive output to increase with tree age. For trees, the first principles would predict that reproductive output increases linearly with canopy area, and canopy area is positively related to tree DBH and thus tree age (Cano et al. 2019). Demographic heterogeneity in reproductive output within species can also arise from genetic or site-specific factors, as suggested by data showing large variation in seed output among conspecific trees even after controlling for canopy area (De Steven and Wright 2002).

Accounting for stage structure is by no means a panacea for neutral theory's woes. We have shown that the problems with cross-scale predictions of spatial neutral models (O'Dwyer and Cornell 2018) are not fixed by the addition of stage structure, at least not by the addition of simple juvenile–reproductive stage structure (Fig. 3–5). Another, probably related, problem with neutral theory is its severe underprediction of temporal changes in species' abundances (Chisholm et al. 2014, Kalyuzhny et al. 2014, Fung et al. 2016), and stage structure will not fix this because the resulting neutral dynamics are still governed by drift – a very slow process. These problems we leave to other papers, past and future. In the meantime, a positive message emerging from this study is that the stage-structured problem is one that is relatively easy to account for via simple rescaling of existing formulas for non-stage-structured neutral models, allowing the growing knowledge base on non-stage-structured models to be leveraged for modelling stage-structured communities. Finally, our results lead to general insights that also apply beyond neutral theory, in particular the insight that a longer juvenile stage can increase community diversity by effectively broadening the period of time over which ancestors of present-day individuals are sampled.

Acknowledgements – The authors thank Richard Condit, Helene Muller-Landau, Deon Lum and James Rosindell for reading drafts of the manuscript and providing insightful comments. The BCI forest dynamics research project was founded by S. P. Hubbell and R. B. Foster and is now managed by R. Condit, S. Lao and R. Perez under the Center for Tropical Forest Science and the Smithsonian Tropical Research in Panama. Numerous organizations have provided funding, principally the US National Science Foundation, and hundreds of field workers have contributed. Acknowledgements for the Barro Colorado Island forest 50-ha plot: R. Foster as the plot founder and the first botanist able to identify so many trees in a diverse forest; R. Pérez and S. Aguilar for species identification; S. Lao for data management; S. Dolins for database design; the hundreds of field workers involved in the census work, which now

consists of over 2 million tree measurements; and the National Science Foundation, Smithsonian Tropical Research Institute and MacArthur Foundation for the bulk of the financial support. *Funding* – RAC and TF acknowledge support from a grant awarded to RAC by the James S. McDonnell Foundation (#220020470).

Author contributions

Ryan A. Chisholm: Conceptualization (lead); Formal analysis (lead); Software (lead); Writing – original draft (lead). **Tak Fung:** Formal analysis (supporting); Writing – review and editing (supporting).

Data availability statement

Data are available from the ForestGEO network website: <<https://forestgeo.si.edu/sites/neotropics/barro-colorado-island>>. (Chisholm and Fung 2021).

References

- Cano, I. M. et al. 2019. Tropical tree height and crown allometries for the Barro Colorado Nature Monument, Panama: a comparison of alternative hierarchical models incorporating interspecific variation in relation to life history traits. – *Biogeosciences* 16: 847–862.
- Cazelles, K. et al. 2016. On the integration of biotic interaction and environmental constraints at the biogeographical scale. – *Ecography* 39: 921–931.
- Chave, J. and Leigh, E. G. J. 2002. A spatially explicit neutral model of β -diversity in tropical forests. – *Theor. Popul. Biol.* 62: 153–168.
- Chisholm, R. A. and Fung, T. 2021. Data from: Adding stage-structure to a spatial neutral model: implications for explaining local and regional patterns of biodiversity. – ForestGEO network website, <<https://forestgeo.si.edu/sites/neotropics/barro-colorado-island>>.
- Chisholm, R. A. and Lichstein, J. W. 2009. Linking dispersal, immigration and scale in the neutral theory of biodiversity. – *Ecol. Lett.* 12: 1385–1393.
- Chisholm, R. A. and O'Dwyer, J. P. 2014. Species ages in neutral biodiversity models. – *Theor. Popul. Biol.* 93: 85–94.
- Chisholm, R. A. et al. 2014. Temporal variability of forest communities: empirical estimates of population change in 4000 tree species. – *Ecol. Lett.* 17: 855–865.
- Chisholm, R. A. et al. 2016. Maintenance of biodiversity on islands. – *Proc. R. Soc. B* 283: 20160102.
- Chisholm, R. A. et al. 2018. Species–area relationships and biodiversity loss in fragmented landscapes. – *Ecol. Lett.* 21: 804–813.
- Comita, L. S. et al. 2010. Asymmetric density dependence shapes species abundances in a tropical tree community. – *Science* 329: 330–332.
- Condit, R. et al. 2001. The status of the Panama Canal watershed and Its biodiversity at the beginning of the 21st century. – *Bio-science* 51: 389–398.
- Condit, R. et al. 2002. Beta-diversity in tropical forest trees. – *Science* 295: 666–669.
- Condit, R. et al. 2005a. Geographic ranges and β -diversity: discovering how many tree species there are where. – *Biol. Skrifter* 55: 57–71.
- Condit, R. et al. 2012. Thirty years of forest census at Barro Colorado and the importance of immigration in maintaining diversity. – *PLoS One* 7: e49826.
- Condit, R. et al. 2017. Demographic trends and climate over 35 years in the Barro Colorado 50 ha plot. – *For. Ecosyst.* 4: 17.
- Condit, R. G. et al. 2005b. Spatial changes in tree composition of high-diversity forests: how much is predictable? – In: Bermingham, E. et al. (eds), *Tropical rainforests: past, present and future*. Univ. of Chicago Press, pp. 271–294.
- Dalling, J. W. and Brown, T. A. 2009. Long-term persistence of pioneer species in tropical rain forest soil seed banks. – *Am. Nat.* 173: 531–535.
- D'Andrea, R. and O'Dwyer, J. P. 2017. The impact of species-neutral stage structure on macroecological patterns. – *Theor. Ecol.* 10: 433–442.
- De Steven, D. and Wright, S. J. 2002. Consequences of variable reproduction for seedling recruitment in three neotropical tree species. – *Ecology* 83: 2315–2327.
- Etienne, R. S. and Haegeman, B. 2011. The neutral theory of biodiversity with random fission speciation. – *Theor. Ecol.* 4: 87–109.
- Fung, T. et al. 2016. Reproducing static and dynamic biodiversity patterns in tropical forests: the critical role of environmental variance. – *Ecology* 5: 1207–1217.
- Hickey, J. J. 1952. Survival studies of banded birds. – U.S. Fish and Wildlife Service, Washington, D.C.
- Hubbell, S. P. 1998. The maintenance of diversity in a neotropical tree community: conceptual issues, current evidence and challenges ahead. – In: Dallmeier, F. and Comiskey, J. A. (eds), *Forest biodiversity research, monitoring and modeling: conceptual background and old world case studies*. Parthenon Publishing, Pearl River, NY, pp. 17–44.
- Hubbell, S. P. 2001. *The unified neutral theory of biodiversity and biogeography*. – Princeton Univ. Press.
- Hubbell, S. P. et al. 2008. How many tree species are there in the Amazon and how many of them will go extinct? – *Proc. Natl Acad. Sci. USA* 105: 11498–11504.
- Jabot, F. et al. 2008. Reconciling neutral community models and environmental filtering: theory and an empirical test. – *Oikos* 117: 1308–1320.
- Kalyuzhny, M. et al. 2014. Niche versus neutrality: a dynamical analysis. – *Am. Nat.* 184: 439–446.
- Kalyuzhny, M. et al. 2015. A neutral theory with environmental stochasticity explains static and dynamic properties of ecological communities. – *Ecol. Lett.* 18: 572–580.
- Melbourne, B. A. and Hastings, A. 2008. Extinction risk depends strongly on factors contributing to stochasticity. – *Nature* 454: 100–103.
- Metcalf, C. J. E. et al. 2009. A time to grow and a time to die: a new way to analyze the dynamics of size, light, age and death of tropical trees. – *Ecology* 90: 2766–2778.
- Muller-Landau, H. C. 2001. Seed dispersal in a tropical forest: empirical patterns, their origins and their consequences for community dynamics. – Princeton Univ. Press.
- Muller-Landau, H. C. et al. 2008. Interspecific variation in primary seed dispersal in a tropical forest. – *J. Ecol.* 96: 653–667.
- O'Dwyer, J. P. and Cornell, S. 2018. Cross-scale ecological theory sheds light on the maintenance of biodiversity. – *Sci. Rep.* 8: 10200.
- O'Dwyer, J. P. et al. 2009. An integrative framework for stochastic, size-structured community assembly. – *Proc. Natl Acad. Sci. USA* 106: 6170–6175.
- Preston, F. W. 1960. Time and space and the variation of species. – *Ecology* 41: 611–627.

- Pyke, C. R. et al. 2001. Floristic composition across a climatic gradient in a neotropical lowland forest. – *J. Veg. Sci.* 12: 553–566.
- Rosindell, J. and Chisholm, R. A. 2020. The species–area relationships of ecological neutral theory. – In: Matthews, T. J. et al. (eds), *The species–area relationship: theory and application*. Cambridge Univ. Press, pp. 259–288.
- Rosindell, J. and Cornell, S. J. 2007. Species–area relationships from a spatially explicit neutral model in an infinite landscape. – *Ecol. Lett.* 10: 586–595.
- Rosindell, J. and Cornell, S. J. 2009. Species–area curves, neutral models and long-distance dispersal. – *Ecology* 90: 1743–1750.
- Rosindell, J. et al. 2008. A coalescence approach to spatial neutral ecology. – *Ecol. Inform.* 3: 259–271.
- Rosindell, J. et al. 2010. Protracted speciation revitalizes the neutral theory of biodiversity. – *Ecol. Lett.* 13: 716–727.
- Rosindell, J. et al. 2012. Age structure in neutral theory resolves inconsistencies related to reproductive-size threshold. – *J. Plant Ecol.* 5: 64–71.
- Strigul, N. et al. 2008. Scaling from trees to forests: tractable macroscopic equations for forest dynamics. – *Ecol. Monogr.* 78: 523–545.
- Thompson, S. E. D. et al. 2019. Characterizing extinction debt following habitat fragmentation using neutral theory. – *Ecol. Lett.* 22: 2087–2096.
- Thompson, S. E. D. et al. 2020. pycoalescence and rcoalescence: packages for simulating spatially explicit neutral models of biodiversity. – *Methods Ecol. Evol.* 11: 1237–1246.
- Volkov, I. et al. 2003. Neutral theory and relative species abundance in ecology. – *Nature* 424: 1035–1037.
- Wright, S. J. et al. 2005. Reproductive size thresholds in tropical trees: variation among individuals, species and forests. – *J. Trop. Ecol.* 21: 307–315.