Density dependence generates spatial overdispersion in adult tropical trees

Authors: Michael Kalyuzhny[1*], Jeffrey K. Lake[2], S. Joseph Wright[3], Annette M. Ostling[1].

1. Department of Integrative Biology, The University of Texas at Austin, 2415 Speedway #C0930, Austin, TX 78712
2. Department of Biology, Adrian College, Adrian, MI
3. Smithsonian Tropical Research Institute, Apartado 0843-03092, Balboa, Republic of Panama

*Correspondence should be addressed to: michael.kalyuzhny@utexas.edu

Keywords: Janzen-Connell, enemy pressure, Neutral Theory, spatial statistics, clumped, even-spaced, random distribution, Poisson model, Complete Spatial Randomness

Abstract

For multiple species to stably coexist, it is necessary that, as species increase in abundance, they suffer from Conspecific Negative Density Dependence (CNDD)[1]. For trees, there is ample evidence for such negative effects of abundance on the survival and frequency of juveniles, typically on small spatial scales[2-11]. Despite this short distance "repulsion" of juveniles, adult trees are generally spatially aggregated[12-16], and there is little evidence that CNDD acting on juveniles propagates to the adult stage[1,2,17], casting doubt on its importance for coexistence. However, evidence of aggregation was previously obtained by comparing the spatial distribution of adults to a random placement across the landscape[12,13], ignoring dispersal limitation. Here we show that, when compared to the predictions of a calibrated null model of stochastic birth, death and limited dispersal, the adults of dozens of tropical forest species are almost ubiquitously and strongly overdispersed, often up to distances of ~100 meters. We also show that such strong overdispersion can only occur if CNDD considerably exceeds Heterospecific Negative Density Dependence, creating stabilizing niche differences, which are an important requirement for stable coexistence. Together these findings suggest that CNDD propagates from juveniles to adults, strongly influencing adult spatial distribution. Furthermore, these results refine our view of the fundamental nature of species spatial distributions, a key interest in ecology[12,18-20], and demonstrate the power of using limited dispersal null models to detect biotic interactions and habitat associations from spatial distributions.

Introduction

How can hundreds of species coexist on small spatial scales? A necessary condition for species coexistence is that the performance of species decreases as they become more common, preventing any one species from taking over, a phenomenon known as Conspecific Negative Density Dependence (CNDD)[1]. CNDD can be caused by competition for specialized resources[2,11], but for tropical trees, it is believed to be caused primarily by species-specific natural enemies - herbivores and pathogens, attacking seeds and juveniles[2,3,11,21,22]. Importantly, both resource competition and natural enemies are hypothesized to operate on relatively small spatial scales[3-5]. This distance dependence of CNDD may lead to the emergence of spatial patterns, such as spacing between parent and successful offspring and overall spatial overdispersion of individuals[21,23]. Conversely, such spatial patterns can be used to identify the signature of CNDD[19,24,25].

CNDD has been suggested to play a major role in maintaining species diversity and its variation across the globe[2,6,7,21,26,27], especially for trees. Considerable evidence was found for negative effects of certain life stages on the frequency and survival of early life stages. For example, the abundance of seedlings, seeds and adults has been shown to reduce the number and survival of seedlings and saplings[3,8,9,23]. Several works have also found evidence that the strength of CNDD varies between species, with common species suffering weaker CNDD than rare species[3,7,8,10], and that CNDD is stronger than Heterospecific Negative Density Dependence (HNDD)[3,7,8]. These characteristics have important implications for the potential of CNDD to enable coexistence[28]. In particular, the situation of CNDD > HNDD implies the existence of niche differences between species, which have the potential to stabilize coexistence[1,2]. Nevertheless, recent works have cast doubt over the validity of many of these findings on methodological grounds[1,9].

Crucially, however, classical works that focused on the spatial distributions of *adult* trees found them to be almost ubiquitously *aggregated* in multiple tropical forests. Hence, individuals are typically closer to each other than expected, especially at small scales where CNDD is expected to be the strongest[12,13,24]. Furthermore, many spatial patterns in tropical forests resemble the expectations of neutral models without CNDD[14,29,30]. Since CNDD leads to "repulsion" between individuals and is expected to generate spatial *overdispersion*, the spatial patterns of adults are in contradiction to the findings on the wide prevalence and importance of CNDD at early life stages. These patterns could suggest that sensitivity to CNDD at early life stages is either severely diluted before reaching adulthood, traded-off against other effects, or completely overwhelmed by processes that create aggregation such as habitat specificity and dispersal limitation[1,2]. If CNDD at early life stages does not propagate to the adult stage (as evidence for this are lacking[17]) – how can it maintain species diversity?

Classical studies on spatial patterns of adult trees compared their distributions to the expectation of the Complete Spatial Randomness (CSR) null model[12,13,19]. This model distributes individuals with equal probability across the (often arbitrary) survey area, inevitably assigning

some to sites that they would not be able to reach due to dispersal limitation. This issue is particularly severe for rare species and species that disperse short distances, which will be artificially "spread" across the landscape by the CSR null model (Fig. 1a,b). We believe that a more ecologically sound null model is Dispersal Limitation (DL, spatially explicit neutral theory)[29,31], which assumes that abundance and distribution are determined by stochastic births and deaths and random, yet limited, dispersal. Using DL as a null model should incorporate the aggregation created by dispersal limitation into the null expectations (Fig 1c, Extended Data Figs. 1-2), potentially revealing signals of biotic interactions and habitat specificity.

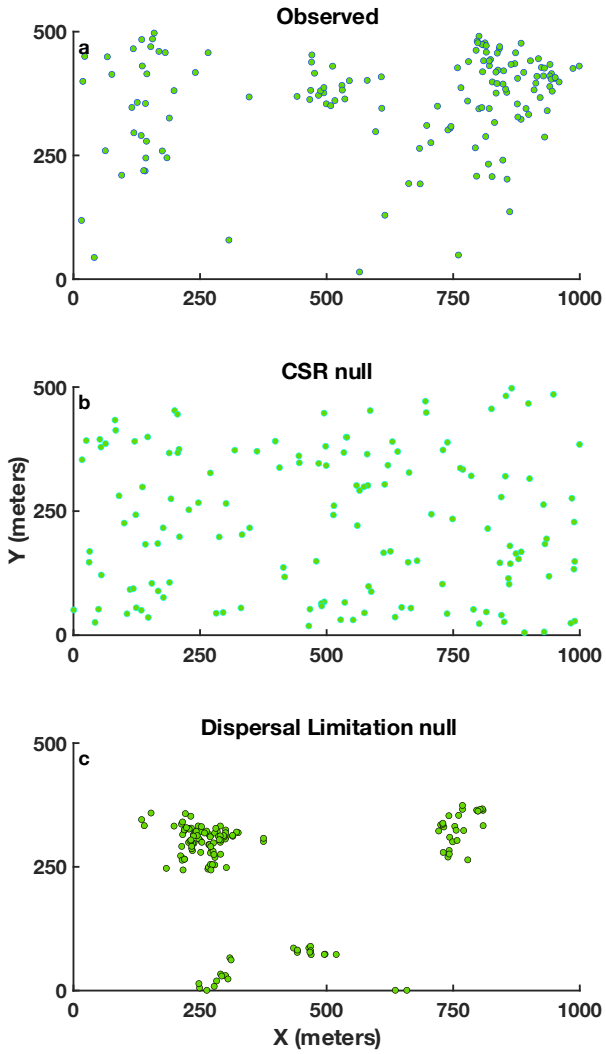


Fig. 1 | Spatial point patterns of adult *Beilschmiedia pendula* trees; The (**a**) observed appears aggregated compared to (**b**) realizations of the Complete Spatial Randomness (CSR) null model, but overdispersed compared to (**c**) the standard variant of the Dispersal Limitation (DL) null model. Both nulls use the same abundance as the observed. The mean dispersal distance of *B. pendula* is estimated at 10.7 meters[32]. Extended Data Fig. 1 presents realizations of the other variants of DL, while Extended Data Fig. 2 zooms in on the clump of trees at the top-left of (**c**).

Here we compare the spatial patterns of tropical tree species in a Panamanian forest to their expectation under DL, using previously estimated species-specific dispersal kernels[32,33]. Our goal is to examine the prevalence, magnitude and scale of overdispersion and aggregation with respect to DL. We hypothesize that some species will show aggregation and some overdispersion, the latter on small spatial scales typical of CNDD (~ 20 meters[3-5]). Finally, we examine the implications of our findings for tropical tree diversity.

Methods summary

We perform the analysis with 41 species that have reliable dispersal kernel estimates using known locations of seed-bearing trees and seed arrivals[32] in the Barro Colorado Island 50 ha forest plot (BCI)[34-36]. DL simulations, parameterized with the kernels, are run for every species, and the observed spatial distribution of adults in 2015 is compared to the expectation under DL. For that comparison we develop analogs of the two most commonly used point-pattern statistics: the Clark-Evans nearest neighbor statistic[13,37] and the relative neighborhood density (a.k.a. pair correlation function)[12,19,29] statistic. First, we compare the observed mean (or median) distance of trees to their nearest conspecific neighbor, NND_{obs}, to its value under the null (NND_{null}) using the Excess Distance (ED) statistic:

(eq. 1) $$ED = \log(NND_{obs}) - \overline{\log(NND_{null})},$$

where the averaging (denoted by an overbar in eqs. 1 and 2) is over samples of the DL model. We also compute the Excess neighborhood Abundance ($EA(r)$), namely the observed number of conspecific neighbors of an average tree at distances r to $r + \Delta r$, $N_{obs}(r)$, relative to its value under the null N_{null}(r):

(eq. 2) $$EA(r) = \log(N_{obs}(r)) - \overline{\log(N_{null}(r))}.$$

We examine this statistic at distances up to 20 meters, as well as at longer distances. However, while classical statistics compared the observed pattern to the CSR null model, we compare to a DL null model. To examine the robustness of our results, we repeated this analysis for the standard DL null hypothesis and five additional variants of the DL null model. The five variants 1) use earlier estimates of dispersal kernels for a broader set of 81 species ('H2001'[33]); 2) incorporate a lag of one generation from dispersal to recruitment ('Lag'); 3) add 10% of global dispersal, representing widespread long-distance dispersal ('LDD')[38]; 4) start with the distribution of conspecific trees in 1985 and examine only the recruits to the adult class appearing by 2015 ('Recruits'); 5) Fix the observed locations of all adult trees ('Fixed'), while allowing their species identity to vary dynamically.

	Excess Abundance (EA) at 20 meters			Excess nearest neighbor Distance (ED)		
Null	Values	Overdispersed	Aggregated	Values	Overdispersed	Aggregated
Standard	-1.5 ± 1.2 [-3.9, 1.3]	38 (34)	3 (2)	0.9 ± 0.61 [-0.65, 2.1]	39 (36)	2 (1)
H2001	-1.1 ± 1.3 [-4.7, 1.3]	69 (47)	12 (3)	0.69 ± 0.67 [-0.45, 2.8]	75 (60)	6 (3)
Lag	-0.99 ± 1.1 [-3.3, 1.5]	33 (28)	8 (2)	0.57 ± 0.52 [-0.72, 1.7]	36 (31)	5 (2)
LDD	-0.97 ± 1.1 [-3.4, 1.4]	35 (27)	6 (3)	0.77 ± 0.71 [-1.1, 2.1]	37 (32)	4 (2)
Recruits	-0.65 ± 1.1 [-3.1, 1.9]	34 (22)	7 (1)	0.69 ± 0.66 [-1.2, 2.6]	38 (33)	3 (1)
Fixed	-1.3 ± 1.1 [-3.7, 1.3]	37 (33)	4 (2)	0.86 ± 0.68 [-1.1, 2.1]	38 (36)	3 (2)

Table 1 | **Distributions and significance of the spatial statistics of Barro Colorado tree species in** 2015 using the six variants of the DL null model as references. We compare the abundance of conspecific neighbors within 20 meters (left, eq. 2) and the nearest neighbor distance (right, eq. 1) to their expectations. For abundances (left), negative values represent overdispersion and vice versa. For distances (right), positive values represent overdispersion. Exponentiating the statistics would give the factor by which the observed density or distance exceeds expectations. The mean ± SD and the range (in square brackets) of the statistics over all species are presented, as well as the number of overdispersed and aggregated species, along with the number of statistically significant (with $\alpha = 0.05$) species in parentheses. For the LDD and Recruits nulls, median nearest neighbor distance is used, while the mean is calculated in the other cases.

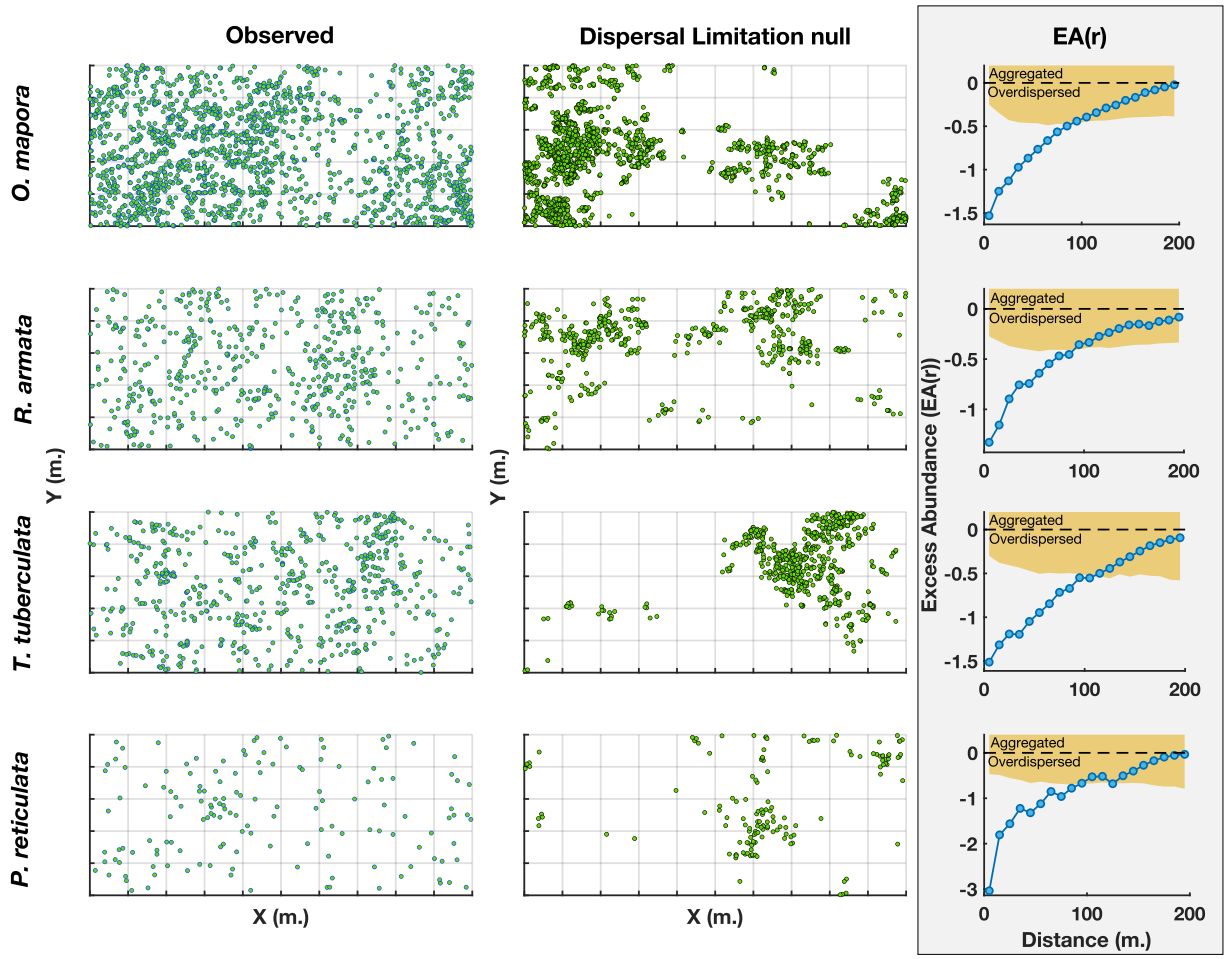


Fig 2. | **Spatial distributions of four typical species and their Excess neighborhor Aundance at distance *r***, *EA*(*r*) (eq. 2). Each row presents a different species, while the columns present for each species the observed distribution (left) versus a single realization of the Standart DL null (center). *EA*(*r*) is presented in the rightmost, shaded column, with the orange region indicating the 95% simulation envelope of the DL null. *EA*(*r*) of 0 indicates a random distribution with respect to DL while values < 0 indicate overdispersion. All distribution maps are 1000 by 500 meters. Compared to the DL null, trees have considerably less neighbors than expected up to long distances.

Results and discussion

Contrary to previous findings based on CSR[12,13], BCI tree species are almost ubiquitously and strongly overdispersed when compared to all six variants of the Dispersal Limitation null model (Table 1, Extended Data Table 1). On average, trees have 50% - 80% fewer conspecific neighbors than expected within 20 meters, and their nearest conspecific neighbor is 1.8 - 2.5 times more distant than expected (these numbers are the values of the mean statistics in Table 1, exponentiated). Even the most common shrubs, treelets and trees (e.g. *Trichilia tuberculata*, Fig. 2), are overdispersed, and these patterns are qualitatively similar under the different DL null

model variants (Table 1). Fig. 2 shows typical, monotonically increasing *EA*(*r*)s for four species, along with their distribution maps in comparison with the expectations of the DL null, while Extended Data Fig. 3-4 show *EA*(r) for additional species. Not surprisingly, some light demanding and habitat specialist species deviate from this typical pattern, exhibiting instead aggregation reflective of the limited spatial extent of the conditions they require (Extended Data Fig. 5).

What underlies this widespread, strong and robust (under different dispersal kernels and alternative assumptions of the null models) overdispersion? Theoretical studies show that while aggregation can be caused by dispersal limitation[14,25,39], habitat specificity[40] and facilitation[41], overdispersion is driven by competition, causing density dependence[19,39,42]. Importantly, we show here using a simulation model that the observed overdispersion *cannot* be created by competition if CNDD and HNDD are similar in magnitude. Fig. 3 shows the degree of overdispersion generated in spatially explicit simulations of community dynamics with different magnitudes of competition (*Q*) and with identical CNDD and HNDD (i.e., HNDD = CNDD = *Q*). For a dispersal distance of 20 meters (roughly the median across species), *ED* asymptotes with increasing competition at ~ 0.11 and *EA*(20) at ~ -0.18 when HNDD = CNDD. These values are much closer to the null expectation than are observed values (Table 1). Even for a very short dispersal distance of 7 meters, overdispersion simulated with HNDD = CNDD remains closer to the null expectation than to observed values. Intuitively, when HNDD = CNDD the competitive effect of neighbors is equal for all neighbors regardless of their species, and so competition cannot generate a strong repulsion from conspecifics. Moreover, in a theoretical study of simulations across a wide range of parameter space of the strengths of CNDD and HNDD[42], we found that large values of HNDD "erode" the species-specific overdispersion created by CNDD even when CNDD > HNDD. Hence, only the scenario CNDD >> HNDD is compatible with patterns of overdispersion observed here. Finally, our results using the 'Fixed' null variant (Table 1), in which the locations of all adults are preserved, are similar to the standard analysis. This null variant preserves the spacing between all trees of all species, resulting from competition for light, soil resources and other non-species-specific factors, hence incorporating HNDD and some CNDD into the null expectation. The strong overdispersion compared with this variant further confirms the weakness of HNDD compared to CNDD.

While we expect overdispersion to occur primarily at the short distances at which CNDD operates[3,5,8], species are typically overdispersed at distances up to ~100 meters, having ~ 45% fewer neighbors 75 – 125 meters away from a typical tree (under the Standard null, but this is not the case under all null variants, Figure 2, Extended Data Table 1). Our recent simulation study shows that such long-distance overdispersion can be observed in a model with short-distance CNDD[42]. Intuitively, while each tree "repels" only its neighbors, they in turn repel their neighbors and so forth, creating large-scale patterns from short-distance interactions[43]. Since the total number of individuals is matched to the observed under our null, the paucity of neighbors at distances < 100 meters must be compensated by an excess of individuals that are "repelled" to

larger distances, although it is difficult to study the patterns at larger scales due to the spatial extend of the data.

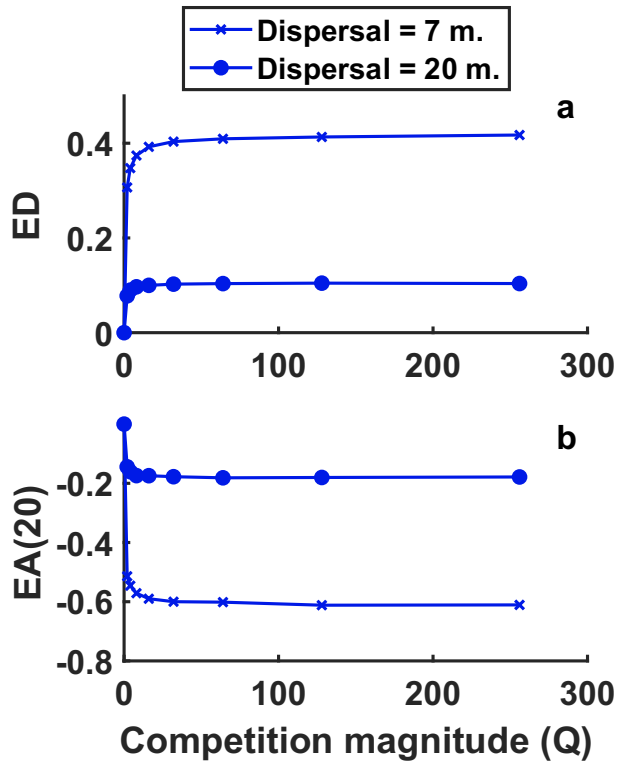


Fig. 3 | **Effect of competition on overdispersion when CNDD = HNDD = *Q*** in simulations of community dynamics. (**a**) presents the Excess nearest conspecific neighbor Distance (*ED*, eq. 1) while (**b**) presents the Excess nearest conspecific neighbor Abundance within 20 meters (*EA*(20), eq. 2). See methods for details of the simulations.

Conclusions

These findings have important implications for the maintenance of species diversity. First, our finding that CNDD must be substantially greater than HNDD for many species implies substantial niche differences among species, which may stabilize diversity because species that increase in abundance suffer reduced fitness. However, CNDD could be overwhelmed by fitness differences[44,45], which are also generated by interspecific differences in CNDD[28]. Moreover, given that the forest is overall saturated with individuals[31], the finding that individuals of most species have fewer *conspecific* neighbors up to long distances implies an increase in *heterospecific* neighbors and therefore spatial compositional turnover. This is in line with the reported increased spatial turnover (beta diversity) at scales of 20-200 meters in BCI compared with a neutral model[30]. Beta diversity has been extensively studied and is believed to be driven

mostly by spatial heterogeneity, dispersal limitation and interspecific competition[30,46,47]. Our results suggest that the role of CNDD in shaping turnover should also be carefully considered.

In his seminal work on CNDD, Janzen asserted that "most adults of a given tree species appear to be more regularly distributed than if the probability of a new adult appearing at a point in the forest were proportional to the number of seeds arriving at that point"[21]. Many have since claimed to reject this hypothesis based on the spatial aggregation of adults[13,16]. However, previous works have compared adult tree distributions to a random placement of adults in the entire observation window (the essence of CSR), which is, in our view, a misinterpretation of the original hypothesis. Our work is the first, to our knowledge, to test this classic hypothesis in multiple species and bring evidence that density dependence has a substantial effect on adult distribution and is much stronger among conspecifics than among heterospecifics. Since spatial overdispersion is the mark of CNDD[19,39,42], and given the substantial evidence for strong CNDD acting on early life stages of trees in BCI[3,8,9], we infer that the observed overdispersion is highly likely a result of CNDD acting on juveniles that propagates to the adult stage. A similar goal was attempted with a null model that considers the change in aggregation from the sapling to the adult life stages[12,24]. These studies found that only 15-22% of species are significantly less aggregated than expected at the adult stage, while we found that virtually all species were overdispersed at the adult stage. This difference is probably because CNDD in trees is accumulated through ontogeny with much of the effect happening at early life stages[48,49] before the sapling stage. This highlights the utility of comparison to the DL null model, which uses seed rain as the reference. Other works have fitted complex process-based models to all individuals (including subadults) of several species finding evidence for CNDD, but this approach could not compare the magnitudes of heterospecific and conspecific effects[25,50].

More broadly, patterns of species distributions at local scales have been investigated and compared to CSR for over one hundred years[18], often with the aim of detecting the signals of habitat specificity and biotic interactions[12,13,15,37,51]. However, theoretical models predict that dispersal limitation alone could override and obscure the signal of biotic interactions and dominate spatial patterns, creating ubiquitous aggregated distributions[39,42]. Indeed, spatial patterns have been found to be predominantly aggregated with respect to the CSR null [12,13,51]. In contrast, a richness of spatial patterns is revealed with respect to the DL null: most species are overdispersed, owing to CNDD, while some habitat and gap specialists are more aggregated. This demonstrates the power of the DL null to filter out much of the effect of dispersal limitation and to reveal the overall effect of biotic interactions and habitat specificity on adult individuals. Our work thus promotes a complementary definition of the classical terms "aggregation" and "overdispersion", rooted in ecological processes and theory. We believe that a more widespread use of this approach will have much capacity to shed light on the processes shaping the assembly of ecological communities.

Methods

Dispersal kernels

We use two sets of dispersal kernels that have been estimated previously using inverse-modeling with seed trap and adult census data. The main source we used ('H-2008') estimated the single parameter of a 2DT dispersal kernel with 3 degrees of freedom[52,53], which was chosen as the best overall, for 41 species[32]. This work also incorporated year to year variation in fecundity and attempted to incorporate dispersal from outside the plot, finding a mean dispersal distance of 27 meters across species. The second set, published earlier ('H-2001')[33], is comprised of 81 species (including 40 species from the first set), used a Weibull dispersal kernel and has a mean dispersal distance of 40 meters. Because it used a less sophisticated technique, we consider results obtained with this set as a "robustness test" of the main results.

Because of the evidence that some species may have surprisingly high levels of long-distance dispersal[38], we also considered a mixed kernel, with 90% of the seeds dispersed according to the H-2008 kernel[32], and 10% distributed completely at random.

The Dispersal Limitation null model

The observed spatial distributions of species are compared to the spatial point patterns generated by the Dispersal Limitation (DL) null model, a spatially-explicit, continuous time and space version of neutral theory[29,31]. Hence, the null model incorporates dispersal limitation and drift and eliminates all forms of niche processes. We use several versions of this null model, representing alternative assumptions.

The DL null is calibrated and run separately for each species with two empirical inputs:

a) the overall number of individuals in the model (J) – to obtain J we compute the density of all individuals in the BCI survey data[34] that have a Diameter at Breast Height (DBH) that equals or exceeds the reproductive threshold of the focal species[32,33], averaged over the surveys from 1985 – 2015. In line with previous works, we use 2/3 of the estimated reproductive threshold[32,33]. This density is transformed to J, to accommodate the larger area that is simulated.
b) The dispersal kernel.

Some additional empirical inputs that are used in variations of the null are mentioned in their descriptions below.

In the simulations, the community consists of a fixed number of J individuals belonging to multiple species on a continuous landscape of 1200 x 1200 meters. Each time step a new individual is recruited, replacing a random individual that is chosen to die. The new individual is an immigrant from a regional pool of 300 species (with equal frequencies) with probability $m = 10/J$ (or $2/J$ in the 'Fixed' version), landing at a random location, or an offspring of a randomly chosen local individual with probability $1 - m$. In the latter case, the offspring is dispersed from the parent tree a random distance drawn from the dispersal distance distribution (which is

derived from the species specific dispersal kernel using radial integration[53]) in a random direction, with torus boundary conditions (so that seeds dispersing "off the edge" of the landscape appear on the other side). The species-specific dispersal kernel (selected from the standard, H2001 and LDD options above) is used for all species in the simulation, making them neutral, and the only reason we use immigration is to avoid fixation due to drift and maintain some variability in abundances, which requires diversity. The simulations are initialized by drawing J individuals from the species pool and assigning them a random location, run for 5000 generations (each consisting of J time steps) to equilibrate, and then sampled every 10 generations, with 1000 samples overall. Each species-sample combination is treated as an independent sample of a spatial point pattern. All samples are truncated to an observation window of 1000 x 500 meters, which are the dimensions of the BCI forest plot.

Three additional versions of the null model are used – in the first ('Lagged'), each recruited individual "waits" for S time steps before it is recruited, representing the time it takes to grow from seed to adult. We used $S = J$ for all species since a generation consists of J time steps.

The second version ('Transient') considers recruitment to the adult class given the initial distribution observed in the data and the observed demographic rates. Hence, this version requires two additional inputs:

a) The initial spatial distribution of adults of the focal species – we used 1985 as the initial state.
b) The number of recruits (to the adult class) and mortality events that took place in the time frame of the data (calculated by comparing consecutive surveys).

We simulated only a single species each time on a landscape of 1000 x 500 meters, unlike in the other versions. The simulation is initiated from the observed distribution of adults in 1985 and the observed number of recruitment and mortality events over 30 years are reshuffled to occur in a random order. $2 \cdot 10^4$ realizations are run, and realizations where a species went extinct are re-run. Only distances from and densities around new recruits are analyzed. Other versions of the DL null model rely on spatial patterns that emerge over long periods of time in the simulation, while this version refrains from making any assumptions on long-term dynamics, relying on the observed initial distribution. This version also incorporates any demographic trends or large fluctuations that species undergo.

In the third version ('Fixed Trees'), we run the simulation using the observed coordinates of all trees of all species with DBH that equals or exceeds the reproductive threshold of the focal species in the 2015 survey. A tree is first chosen to die and is immediately replaced (at the same coordinates) by a recruit. The parent of this recruit is found by drawing a distance from the replaced tree using the dispersal distance distribution. The tree nearest to the drawn location (using torus boundary conditions on a 1000 x 500 m. landscape) is set to be the parent. This version is set to preserve the distancing of trees from each other that results from competition for space and light.

Overall, we use six different variants of the DL null model analysis: 1) the basic DL model with the standard H-2008 kernel ('Standard'); 2) The basic DL model with the H-2001 kernel ('H2001'); 3) the Lagged DL model with the standard H-2008 kernel ('Lag'); 4) the basic DL model with the kernel incorporating extra long-range dispersal ('LDD'); 5) the Transient null model, initialized with the observed spatial pattern in 1985, using the standard H-2008 kernel ('Recruits'); 6) the Fixed Trees null model, initialized with the locations of trees in 2015, using the standard H-2008 kernel ('Fixed'). Variants 2-6 of the null model analysis are used as robustness tests.

It is important to note that all our DL models assume that the community has a fixed size. Other DL models may be appropriate in different cases, such as if a species invades an empty habitat.

Spatial statistics

We compared the observed spatial distributions of adults to the null spatial distributions using two analogs of classical spatial statistics: Excess neighborhood Abundance ($EA(r)$) and Excess Distance to conspecific neighbors (ED), analogs of the pair correlation function and Clark-Evans nearest neighbor statistics[12,13,19,37], respectively. All the distances we compute incorporate torus boundary conditions.

Unlike the classical statistics, we design our statistics to be symmetric on the logarithmic scale and have a mean of zero under the null. To compute $EA(r)$ (eq. 2), the number of conspecific neighbors within an annulus at distances r to $r + \Delta r$ is computed for every tree and averaged over all the observed ($N_{obs}(r)$) or simulated (($N_{null}(r)$) trees in the spatial distribution, and the logarithm of both is computed. To obtain the expectation under the DL null model, $\overline{\log(N_{null}(r))}$, we perform the aforementioned calculation for every point pattern sampled with an abundance similar to the observed (see more about abundance binning below), obtaining a distribution of log abundances within each annulus and using its mean over realizations of DL. Moreover, to evaluate the variability of $EA(r)$ under the null, we construct 95% simulation envelopes by examining the 2.5 and 97.5 percentiles of the distribution of simulated log abundances[19]. This is presented as an orange area in Fig. 2 and Extended Data Figures 3 - 4. $EA(r)$ is computed at distance bins of 0 - 10, 10 - 20, 20 - 30, …, 190 – 200 meters. Moreover, to obtain a single overdispersion statistic for each species, we compute $EA(r)$ for a single circle with radius 20 meters (equivalent to a single distance bin of 0 – 20 m, $EA(20)$), which is often believed to be roughly the radius of interactions[3-5]. If no individuals are observed within some distance bin, and to avoid taking the logarithm of zero, we assume that a single tree has a single neighbor at this distance. This equals half of the minimal possible strictly positive number of neighbors, since in real data at least two trees would have each other in this neighborhood (or no trees at all). Values of $EA(r) > 0$ indicate a higher abundance than expected at some radius, or aggregation, while $EA(r) < 0$ indicates overdispersion.

Notice that if $EA(r)$ is negative at small r, the trees 'absent' at a short distance must appear at larger distances, creating positive $EA(r)$ at these larger distances. The analogous issue appears with the relative neighborhood density statistic (pair-correlation function)[12,19]. Still, these patterns are typically interpreted at short distances[19].

We also study the Excess Distance (ED) of trees to their nearest neighbors. For every tree, the distance to the nearest conspecific neighbor is calculated, and the mean ($\overline{NND}_{obs}$, sometimes median, $\widehat{\mathrm{NND}}_{\mathrm{obs}}$, see below) of this distribution over the trees is compared to the expectation under the null, in one of two ways:

(eq. 3) $$ED_{mean} = \log(\overline{\mathrm{NND}}_{\mathrm{obs}}) - \overline{\log(\overline{\mathrm{NND}}_{\mathrm{null}})},$$

where $\overline{\mathrm{NND}}_{\mathrm{obs}}$ is the observed mean distance of a tree to its nearest neighbor and $\overline{\mathrm{NND}}_{\mathrm{null}}$ is this mean for a sample of the null, whose log is averaged over the samples. Alternatively, we use medians instead of means, computing the logarithm of the median distance of a tree to its nearest neighbor, and taking the median over samples of the null:

(eq. 4) $$ED_{median} = \log\left(\widehat{\mathrm{NND}}_{\mathrm{obs}}\right) - \log\left(\widehat{\widehat{\mathrm{NND}}}_{\mathrm{null}}\right),$$

where "hat" indicates median. Values of $ED > 0$ mean that the distances to neighbors are larger than expected and the distribution is overdispersed, while $ED < 0$ indicates aggregation. To preserve the analogy with the Clark-Evans statistic, we typically used ED_{mean} for analyses, accept when using the LDD and Recruits variants of DL. We used medians with the LDD variant because there is an arbitrary proportion of seeds that are dispersed very long distances, and the median is less sensitive to this. For the Recruits variant, we computed only nearest neighbor distances of recruits to all trees, because other trees are preserved under the null. The distribution of nearest neighbor distances in this case is very skewed, and our simulations showed that using the mean (but not the median) would lead to deviations from neutrality when the method is applied to neutral simulations (results not shown). Hence, we used ED_{median} here as well. Intuitively, if ED or $EA(r)$ have a value of x, it indicates that the nearest neighbor distance or number of conspecific neighbors at distance r, respectively, are larger by e^x then their expectation, which is the geometric mean (or median) over realizations of the null.

The significance of both $EA(20)$ and ED are evaluated by comparing the logarithm of the observed nearest neighbor distance or abundance with the distribution of their expectations under the nulls with a two-sided test (with $\alpha = 0.05$). Importantly, both nearest neighbor distances and neighborhood density depend strongly on abundance. For this reason, the observed patterns must be compared with samples of similar abundance. Therefore, for all null models except for Recruits (which is initialized with the observed abundance), a species of abundance N is compared to null samples with abundances in the range (round($N - B$), round($N + B$)), where B, the bin width, is $0.3*N^{0.75}$. This non-standard binning meant that a species with abundance < 9 will be compared only to samples of its exact abundance, species with an abundance of 20 will

be compared to species with abundance in the range (17,23) and a species with abundance 500 will be compared with the range (468, 532).

Simulation model

To examine the mechanistic causes of overdispersion and specifically to assess the potential degree of overdispersion that can result from a situation when CNDD = HNDD, we run simulations of such a case resembling BCI in other manners. We simulate a community of sessile organisms on a landscape with edge length L = 600 m and toroidal boundary conditions. The landscape is inhabited by a fixed number of J = 5,500 adults. This density corresponds to the density of trees with Diameter at Brest Height (DBH) of 20 cm (the mode and median of the reproductive thresholds of the species we analyzed) and above in the 50-ha forest plot on Barro Colorado Island (BCI). Each time step, with probability $m = 10^{-3}$, a seed arrives from a uniform species pool of S_{reg} = 300 species at a random location, or, with probability $1 - m$, a random local individual is chosen to reproduce. The offspring is dispersed with a 2DT (3df) dispersal kernel[32]. Two mean distances were used: 20 meters, corresponding to the median of the dispersal distance among the species we analyzed, and seven meters, which is close to the minimal distance among species. The probability an offspring would establish equals $\frac{1}{1+NCI}$, where NCI is the Neighborhood Crowding Index, or the summed contribution of the competitive effects of neighbors, C_i. In line with previous analyses[4,54,55], we assume that only neighbors within 20 meters have any effect on the focal individual, and for these neighbors the competitive effect of individual i at distance D_i on the focal individual is:

(eq. 5) $$C_i(D_i) = \frac{Q}{1+\left(\frac{D_i}{b_1}\right)^{b_2}},$$

where Q is the magnitude of competition, b_1 is the distance at which the competition is reduced in half, and b_2 governs how fast competitive effects decline with distance . This functional form enables setting competition to be relatively fixed up to some distance, followed by a sharp decline at longer distances (when b_2 is large, see Extended Data Fig. 5). We believe this to be a good model for shading, whereas under the crown the effect is quite fixed, followed by a rapid decline beyond the edge of the crown. We set b_1 = 4.5 meters, representing roughly the crown diameter of adult trees, and b_2 = 5, so that at a distance of 20 meters the effect drops to ~ 0.0005Q (see Extended Data Fig. 6). If an offspring establishes, it immediately becomes an adult, preserving the species identity of its parent, and a random adult is chosen to die.

Simulations were run with values of Q of 0, 2, 4, 8, …, 256. All simulations were initiated as random samples from the pool and run for 1000 generations (J timesteps each) to equilibrate, and then sampled 400 - 1500 times, every ten generations. The spatial distribution of all species with more than five individuals was compared to a Dispersal Assembly null model with similar parameters, but where Q = 0, using a procedure similar to the empirical analyses.

Note that the *m* parameter we chose is lower than values estimated for BCI[56,57], because we wanted to avoid the spatially homogenizing effect of random immigration from outside the plot. We expect that incorporating higher levels of *m* would reduce overdispersion in simulations even further.

Data Availability

All data used in this analysis is publicly available and referenced to.

Code availability

All the code used to generate the results is included in the supplement.

References

1 Broekman, M. J. E. *et al.* Signs of stabilisation and stable coexistence. *Ecology Letters* **22**, 1957-1975, doi:10.1111/ele.13349 (2019).

2 Hülsmann, L., Chisholm, R. A. & Hartig, F. Is Variation in Conspecific Negative Density Dependence Driving Tree Diversity Patterns at Large Scales? *Trends in Ecology & Evolution* (2020).

3 Comita, L. S., Muller-Landau, H. C., Aguilar, S. & Hubbell, S. P. Asymmetric density dependence shapes species abundances in a tropical tree community. *Science* **329**, 330-332, doi:10.1126/science.1190772 (2010).

4 Uriarte, M., Canham, C. D., Thompson, J., Zimmerman, J. K. & Brokaw, N. Seedling recruitment in a hurricane-driven tropical forest: light limitation, density-dependence and the spatial distribution of parent trees. *Journal of Ecology* **93**, 291-304, doi:10.1111/j.0022-0477.2005.00984.x (2005).

5 Hubbell, S. P., Ahumada, J. A., Condit, R. & Foster, R. B. Local neighborhood effects on long-term survival of individual trees in a neotropical forest. *Ecological Research* **16**, 859-875, doi:10.1046/j.1440-1703.2001.00445.x (2001).

6 Bagchi, R. *et al.* Pathogens and insect herbivores drive rainforest plant diversity and composition. *Nature* **506**, 85-+, doi:10.1038/nature12911 (2014).

7 Johnson, D. J., Beaulieu, W. T., Bever, J. D. & Clay, K. Conspecific Negative Density Dependence and Forest Diversity. *Science* **336**, 904-907, doi:10.1126/science.1220269 (2012).

8 Lebrija-Trejos, E., Reich, P. B., Hernandez, A. & Wright, S. J. Species with greater seed mass are more tolerant of conspecific neighbours: a key driver of early survival and future abundances in a tropical forest. *Ecology Letters* **19**, 1071-1080, doi:10.1111/ele.12643 (2016).

9 Detto, M., Visser, M. D., Wright, S. J. & Pacala, S. W. Bias in the detection of negative density dependence in plant communities. *Ecology Letters* **22**, 1923-1939, doi:10.1111/ele.13372 (2019).

10 Mangan, S. A. *et al.* Negative plant-soil feedback predicts tree-species relative abundance in a tropical forest. *Nature* **466**, 752-U710, doi:10.1038/nature09273 (2010).

11 Forrister, D. L., Endara, M. J., Younkin, G. C., Coley, P. D. & Kursar, T. A. Herbivores as drivers of negative density dependence in tropical forest saplings. *Science* **363**, 1213-+, doi:10.1126/science.aau9460 (2019).
12 Condit, R. *et al.* Spatial patterns in the distribution of tropical tree species. *Science* **288**, 1414-1418, doi:10.1126/science.288.5470.1414 (2000).
13 Hubbell, S. P. Tree dispersion, abundance, and diversity in a tropical dry forest. *Science* **203**, 1299-1309, doi:10.1126/science.203.4387.1299 (1979).
14 Seri, E., Maruvka, Y. E. & Shnerb, N. M. Neutral Dynamics and Cluster Statistics in a Tropical Forest. *American Naturalist* **180**, E161-E173, doi:10.1086/668125 (2012).
15 Plotkin, J. B. *et al.* Species-area curves, spatial aggregation, and habitat specialization in tropical forests. *Journal of Theoretical Biology* **207**, 81-99, doi:10.1006/jtbi.2000.2158 (2000).
16 Leigh, E. G. *Tropical Forest Ecology : A View from Barro Colorado Island*. (Oxford University Press, USA, 1999).
17 Mittelbach, G. G. & McGill, B. J. *Community ecology*. (Oxford University Press, 2019).
18 Gleason, H. A. Some applications of the quadrat method. *Bulletin of the Torrey Botanical Club* **47**, 21-33 (1920).
19 Wiegand, T. & Moloney, K. A. *Handbook of spatial point-pattern analysis in ecology*. (CRC press, 2013).
20 McIntire, E. J. B. & Fajardo, A. Beyond description: the active and effective way to infer processes from spatial patterns. *Ecology* **90**, 46-56, doi:10.1890/07-2096.1 (2009).
21 Janzen, D. H. Herbivores and the number of tree species in tropical forests. *American Naturalist* **104**, 501-528, doi:10.1086/282687 (1970).
22 Connell, J. H. On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees. *Dynamics of populations* **298**, 312 (1971).
23 Marchand, P. *et al.* Seed-to-seedling transitions exhibit distance-dependent mortality but no strong spacing effects in a Neotropical forest. *Ecology* **101**, doi:10.1002/ecy.2926 (2020).
24 Bagchi, R. *et al.* Spatial patterns reveal negative density dependence and habitat associations in tropical trees. *Ecology* **92**, 1723-1729, doi:10.1890/11-0335.1 (2011).
25 Detto, M. & Muller-Landau, H. C. Fitting Ecological Process Models to Spatial Patterns Using Scalewise Variances and Moment Equations. *American Naturalist* **181**, E68-E82, doi:10.1086/669678 (2013).
26 LaManna, J. A. *et al.* Plant diversity increases with the strength of negative density dependence at the global scale. *Science* **356**, 1389-1392, doi:10.1126/science.aam5678 (2017).
27 Levi, T. *et al.* Tropical forests can maintain hyperdiversity because of enemies. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 581-586, doi:10.1073/pnas.1813211116 (2019).
28 Stump, S. M. & Comita, L. S. Interspecific variation in conspecific negative density dependence can make species less likely to coexist. *Ecology Letters* **21**, 1541-1551, doi:10.1111/ele.13135 (2018).
29 May, F., Huth, A. & Wiegand, T. Moving beyond abundance distributions: neutral theory and spatial patterns in a tropical forest. *Proceedings of the Royal Society B-Biological Sciences* **282**, doi:10.1098/rspb.2014.1657 (2015).

30 Condit, R. *et al.* Beta-diversity in tropical forest trees. *Science* **295**, 666-669, doi:10.1126/science.1066854 (2002).
31 Hubbell, S. P. *The unified neutral theory of biodiversity and biogeography*. Vol. 32 (Princeton University Press, 2001).
32 Muller-Landau, H. C., Wright, S. J., Calderon, O., Condit, R. & Hubbell, S. P. Interspecific variation in primary seed dispersal in a tropical forest. *Journal of Ecology* **96**, 653-667, doi:10.1111/j.1365-2745.2008.01399.x (2008).
33 Muller-Landau, H. C. Seed dispersal in a tropical forest: empirical patterns, their origins, and their consequences for community dynamics. (2001).
34 Condit, R. *et al.* Complete data from the Barro Colorado 50-ha plot: 423617 trees, 35 years. *URL https://doi. org/10.15146/5xcp-0d46* (2019).
35 Hubbell, S. P. *et al.* Light-gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. *Science* **283**, 554-557, doi:10.1126/science.283.5401.554 (1999).
36 Condit, R. *Tropical Forest Census Plots: Methods and Results from Barro Colorado Island, Panama and a Comparison with Other Plots*. (Springer, 1998).
37 Clark, P. J. & Evans, F. C. Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology* **35**, 445-453 (1954).
38 Wright, S. J. *et al.* Understanding strategies for seed dispersal by wind under contrasting atmospheric conditions. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 19084-19089, doi:10.1073/pnas.0802697105 (2008).
39 Chave, J., Muller-Landau, H. C. & Levin, S. A. Comparing classical community models: Theoretical consequences for patterns of diversity. *American Naturalist* **159**, 1-23, doi:10.1086/324112 (2002).
40 Brown, C., Illian, J. B. & Burslem, D. Success of spatial statistics in determining underlying process in simulated plant communities. *Journal of Ecology* **104**, 160-172, doi:10.1111/1365-2745.12493 (2016).
41 Molofsky, J., Bever, J. D. & Antonovics, J. Coexistence under positive frequency dependence. *Proceedings of the Royal Society B-Biological Sciences* **268**, 273-277, doi:10.1098/rspb.2000.1355 (2001).
42 Kalyuzhny, M., Lake, J. K. & Ostling, A. M. Spatial patterns from a dispersal limitation perspective: revealing biotic interactions. *bioRxiv*, 2022.2002.2010.480007, doi:10.1101/2022.02.10.480007 (2022).
43 Levin, S. A. The Problem of Pattern and Scale in Ecology. *Ecology* **73**, 1943-1967, doi:10.2307/1941447 (1992).
44 Chisholm, R. A. & Fung, T. Janzen-Connell Effects Are a Weak Impediment to Competitive Exclusion. *American Naturalist* **196**, 649-661, doi:10.1086/711042 (2020).
45 Chisholm, R. A. & Muller-Landau, H. C. A theoretical model linking interspecific variation in density dependence to species abundances. *Theoretical Ecology* **4**, 241-253, doi:10.1007/s12080-011-0119-z (2011).
46 Chase, J. M. & Myers, J. A. Disentangling the importance of ecological niches from stochastic processes across scales. *Philosophical Transactions of the Royal Society B-Biological Sciences* **366**, 2351-2363, doi:10.1098/rstb.2011.0063 (2011).
47 Segre, H. *et al.* Competitive exclusion, beta diversity, and deterministic vs. stochastic drivers of community assembly. *Ecology Letters* **17**, 1400-1408, doi:10.1111/ele.12343 (2014).

48 Zhu, Y., Comita, L. S., Hubbell, S. P. & Ma, K. P. Conspecific and phylogenetic density-dependent survival differs across life stages in a tropical forest. *Journal of Ecology* **103**, 957-966, doi:10.1111/1365-2745.12414 (2015).
49 Zhu, Y. *et al.* Density-dependent survival varies with species life-history strategy in a tropical forest. *Ecology Letters* **21**, 506-515, doi:10.1111/ele.12915 (2018).
50 Detto, M. & Muller-Landau, H. C. Rates of formation and dissipation of clumping reveal lagged responses in tropical tree populations. *Ecology* **97**, 1170-1181 (2016).
51 Levine, J. M. & Murrell, D. J. The community-level consequences of seed dispersal patterns. *Annual Review of Ecology Evolution and Systematics* **34**, 549-574, doi:10.1146/annurev.ecolsys.34.011802.132400 (2003).
52 Clark, J. S., Silman, M., Kern, R., Macklin, E. & HilleRisLambers, J. Seed dispersal near and far: Patterns across temperate and tropical forests. *Ecology* **80**, 1475-1494, doi:10.2307/176541 (1999).
53 Nathan, R., Klein, E., Robledo-Arnuncio, J. J. & Revilla, E. Dispersal kernels. *Dispersal ecology and evolution*, 187-210 (2012).
54 Uriarte, M., Canham, C. D., Thompson, J. & Zimmerman, J. K. A neighborhood analysis of tree growth and survival in a hurricane-driven tropical forest. *Ecological Monographs* **74**, 591-614, doi:10.1890/03-4031 (2004).
55 Grote, S., Condit, R., Hubbell, S., Wirth, C. & Ruger, N. Response of Demographic Rates of Tropical Trees to Light Availability: Can Position-Based Competition Indices Replace Information from Canopy Census Data? *Plos One* **8**, doi:10.1371/journal.pone.0081787 (2013).
56 Chisholm, R. A. & Lichstein, J. W. Linking dispersal, immigration and scale in the neutral theory of biodiversity. *Ecology Letters* **12**, 1385-1393, doi:10.1111/j.1461-0248.2009.01389.x (2009).
57 Volkov, I., Banavar, J. R., Hubbell, S. P. & Maritan, A. Neutral theory and relative species abundance in ecology. *Nature* **424**, 1035-1037, doi:10.1038/nature01883 (2003).
58 Harms, K. E., Condit, R., Hubbell, S. P. & Foster, R. B. Habitat associations of trees and shrubs in a 50-ha neotropical forest plot. *Journal of Ecology* **89**, 947-959, doi:10.1111/j.1365-2745.2001.00615.x (2001).
59 Wright, S. J. *et al.* Functional traits and the growth-mortality trade-off in tropical trees. *Ecology* **91**, 3664-3674, doi:10.1890/09-2335.1 (2010).

Acknowledgements

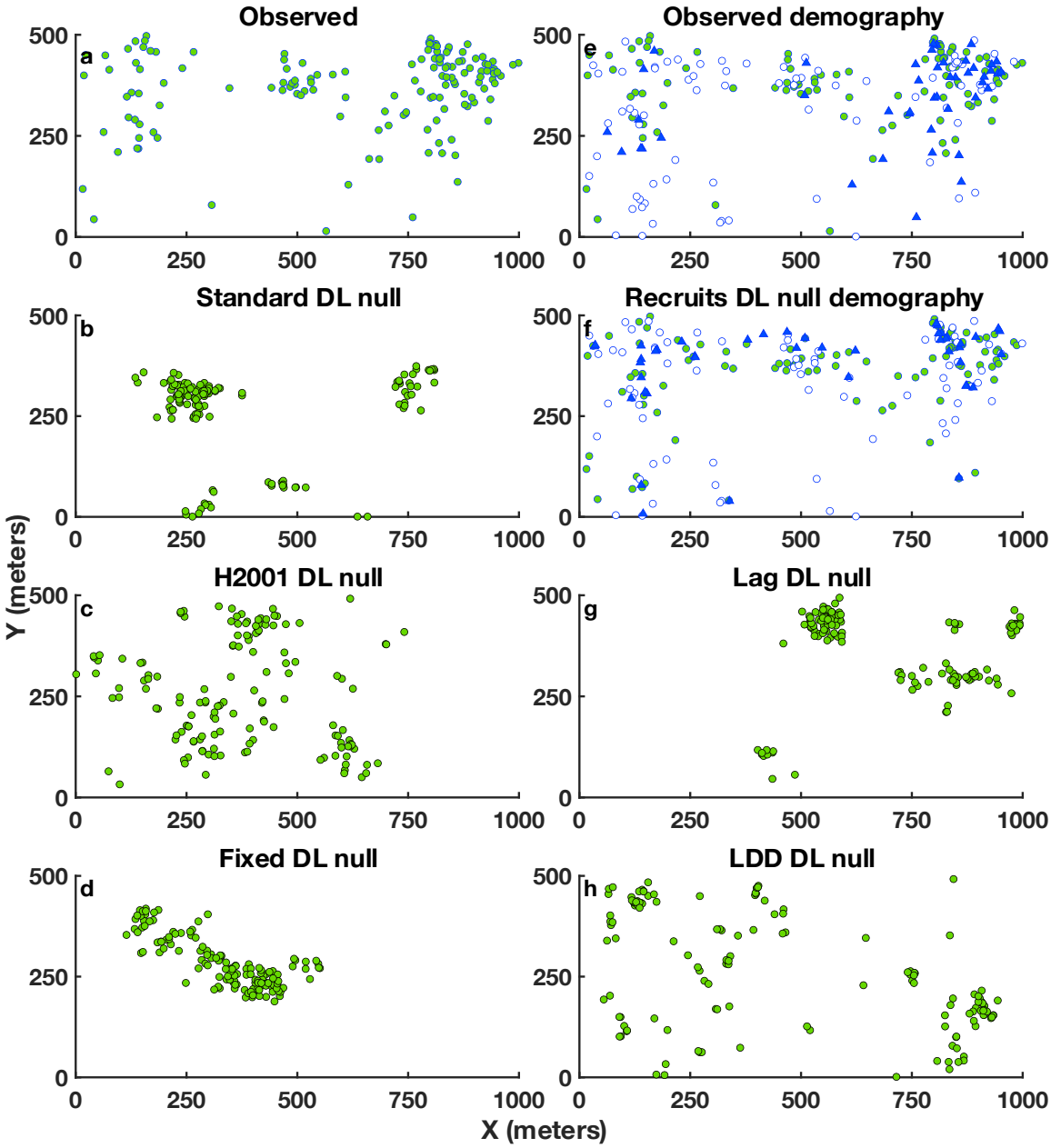
This work was supported by a Michigan Life Sciences Fellowship for MK, the Zuckerman STEm Leadership Program, funds from MCubed and the Associate Professor Support Fund at the University of Michigan. JKL acknowledges Adrian College and the University of Michigan for sabbatical support. We thank R. Condit and R. Kadmon for comments. BCI forest dynamics research project was made possible by National Science Foundation grants to Stephen P. Hubbell: DEB-0640386, DEB-0425651, DEB-0346488, DEB-0129874, DEB-00753102, DEB-9909347, DEB-9615226, DEB-9615226, DEB-9405933, DEB-9221033, DEB-9100058, DEB-8906869, DEB-8605042, DEB-8206992, DEB-7922197, support from the Forest Global Earth Observatory, the Smithsonian Tropical Research Institute, the John D. and Catherine T.

MacArthur Foundation, the Mellon Foundation, the Small World Institute Fund, and numerous private individuals, and through the hard work of over 100 people from 10 countries over the past three decades. The plot project is part the Forest Global Earth Observatory (ForestGEO), a global network of large-scale demographic tree plots.

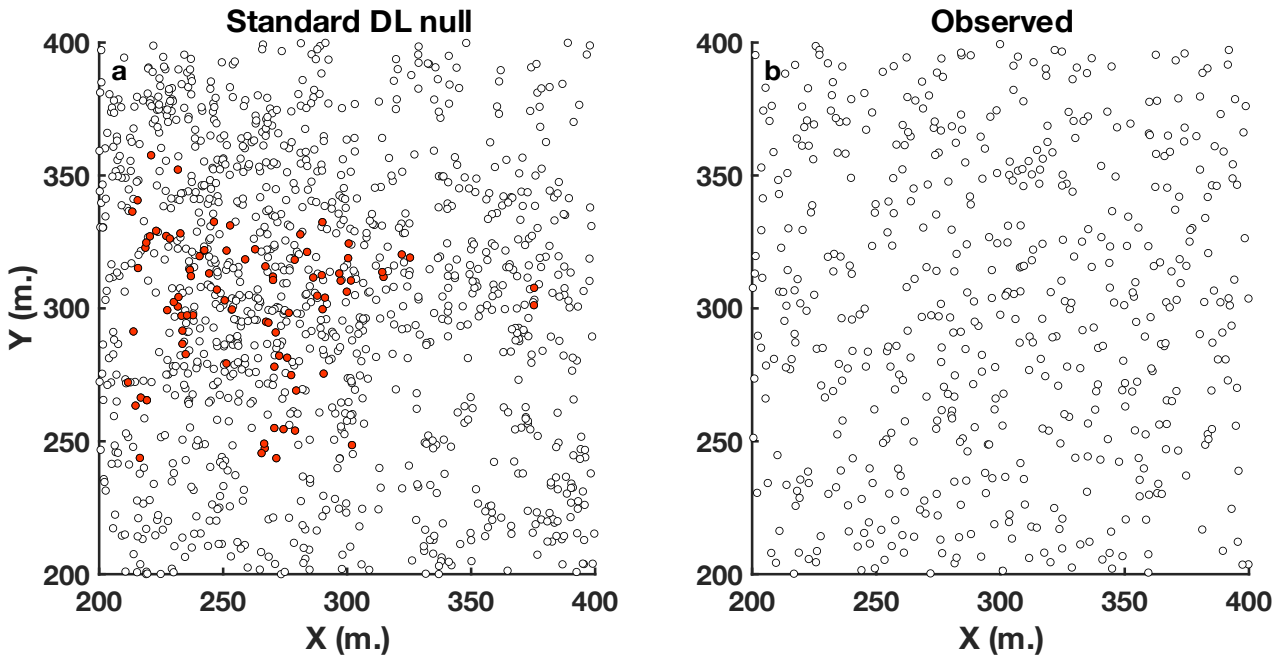
Author contributions

MK conceptualized the study, MK, JKL and AMO planned the study, SJW generated the data, MK performed the analysis and wrote the first draft of the manuscript, and all authors contributed substantially to interpreting

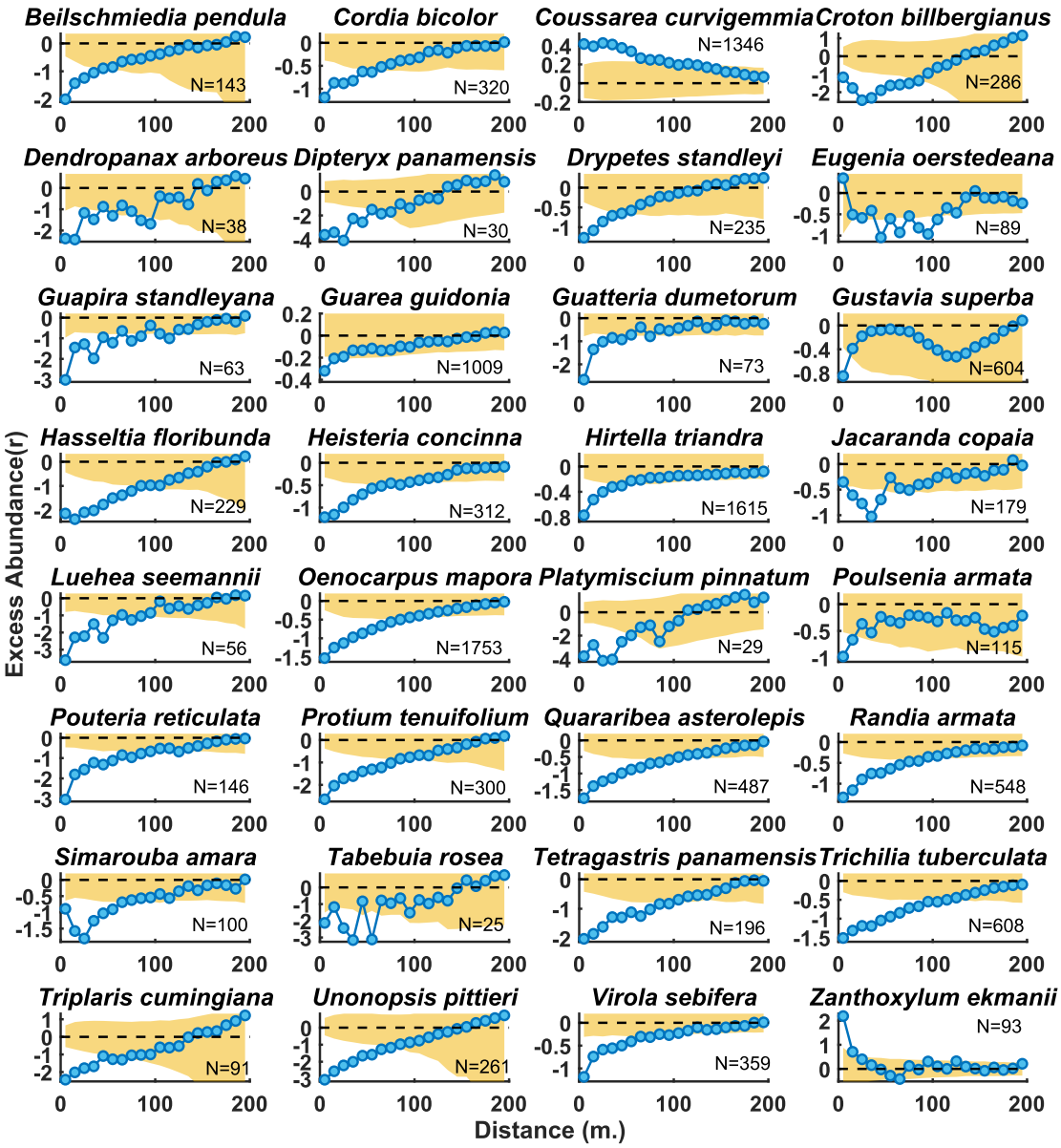
Extended data figures and tables



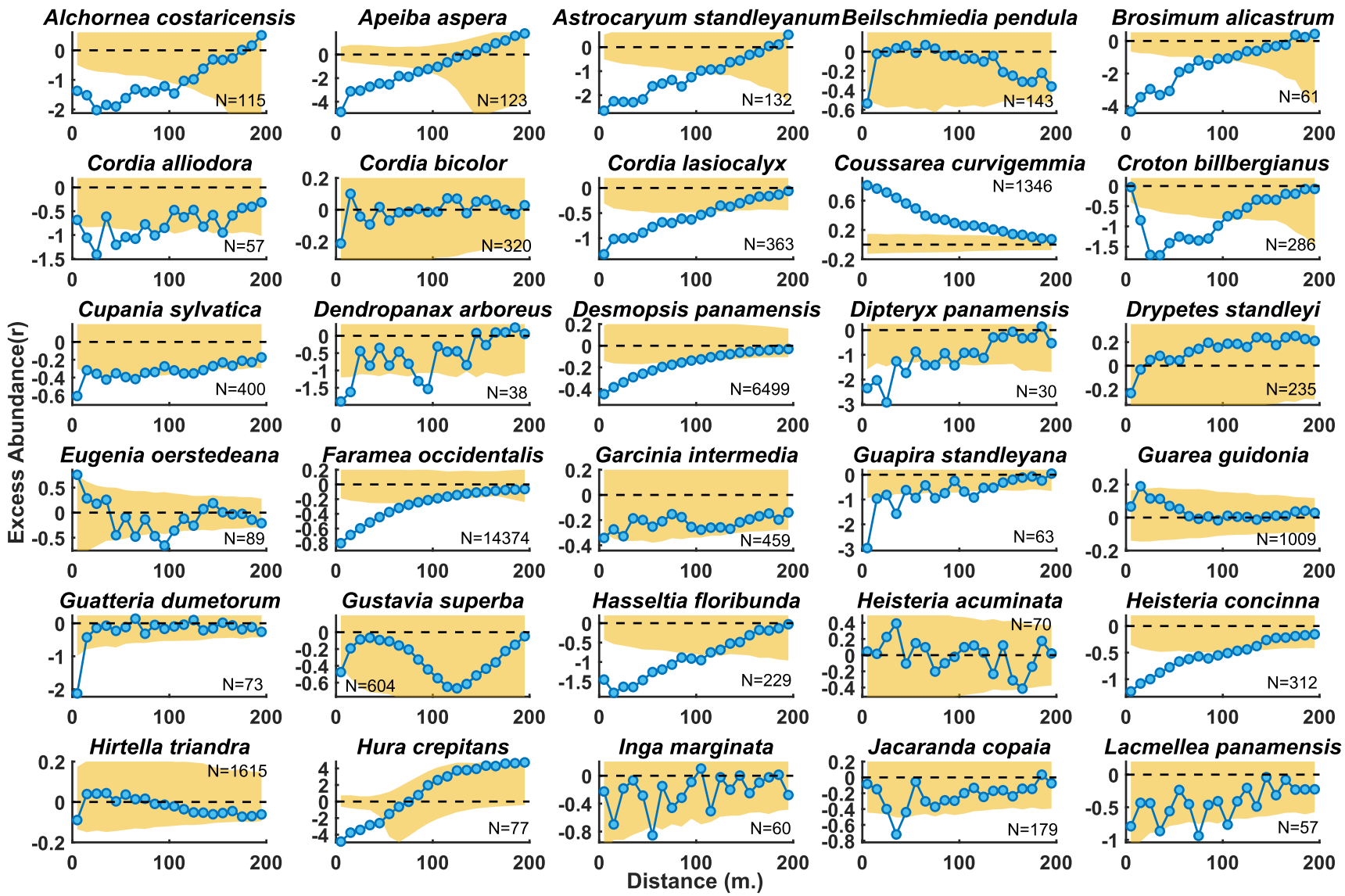
Extended Data Fig. 1 | **Spatial point patterns of adult *Beilschmiedia pendula* trees** – observed (**a,e**), compared to realizations of the null models (**b-d**, **f-h**). The nulls fix the observed abundance, while the Recruits null fixes the number of observed recruitment (blue triangle) and mortality (empty circle) events from 1985 to 2015.

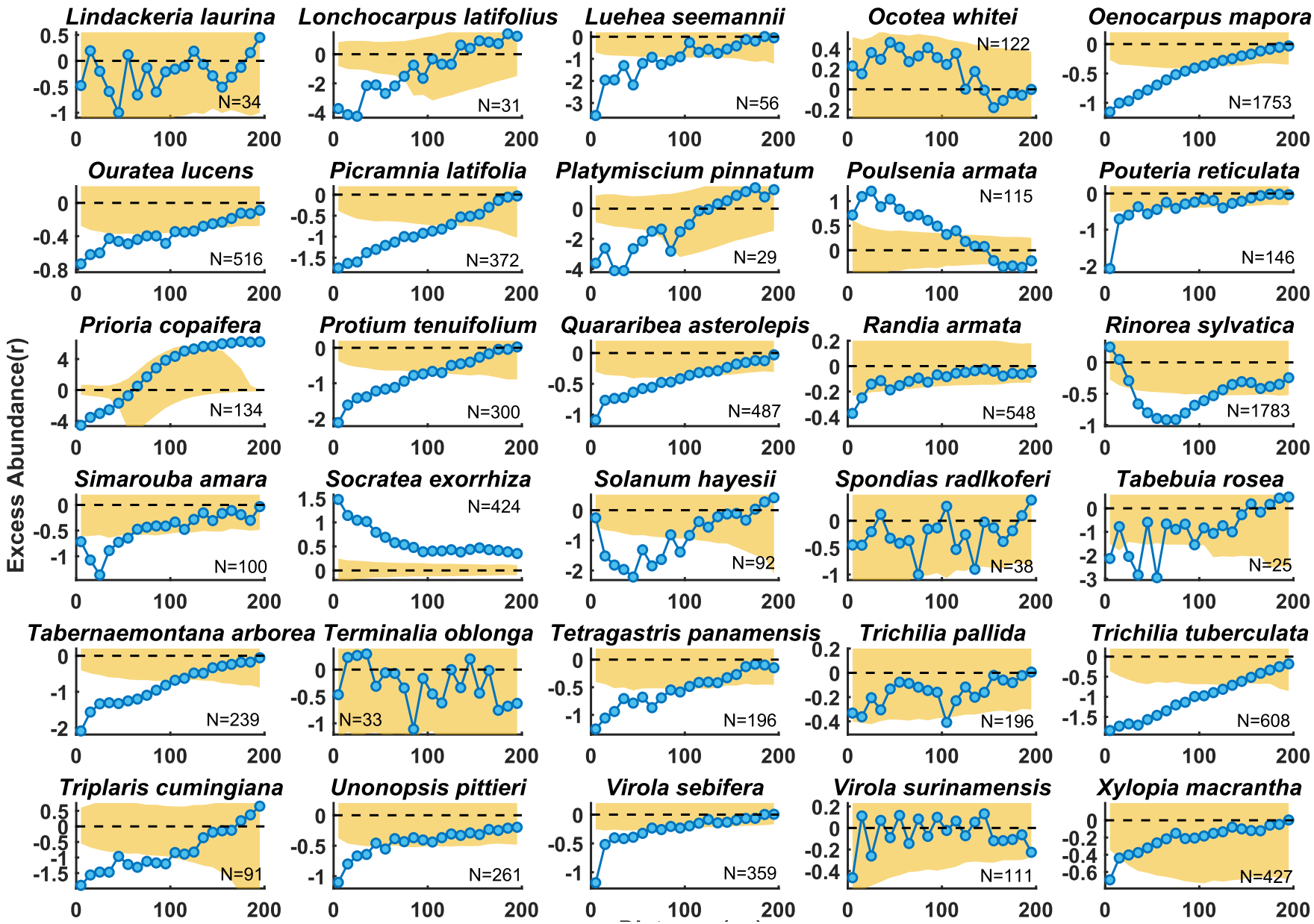


Extended Data Fig. 2 | **Observed vs. null spatial patterns, zoomed in**. (**a**) shows a piece of Fig. 1c, zooming on a clump of *Beilschmiedia pendula* trees (red) under the Standard DL null, shown along with heterospecifics (hollow black). This is compared to (**b**), showing the observed pattern of trees in the same area. The density of trees under the null in this area is somewhat higher than the observed, but it is lower in other parts of the landscape.

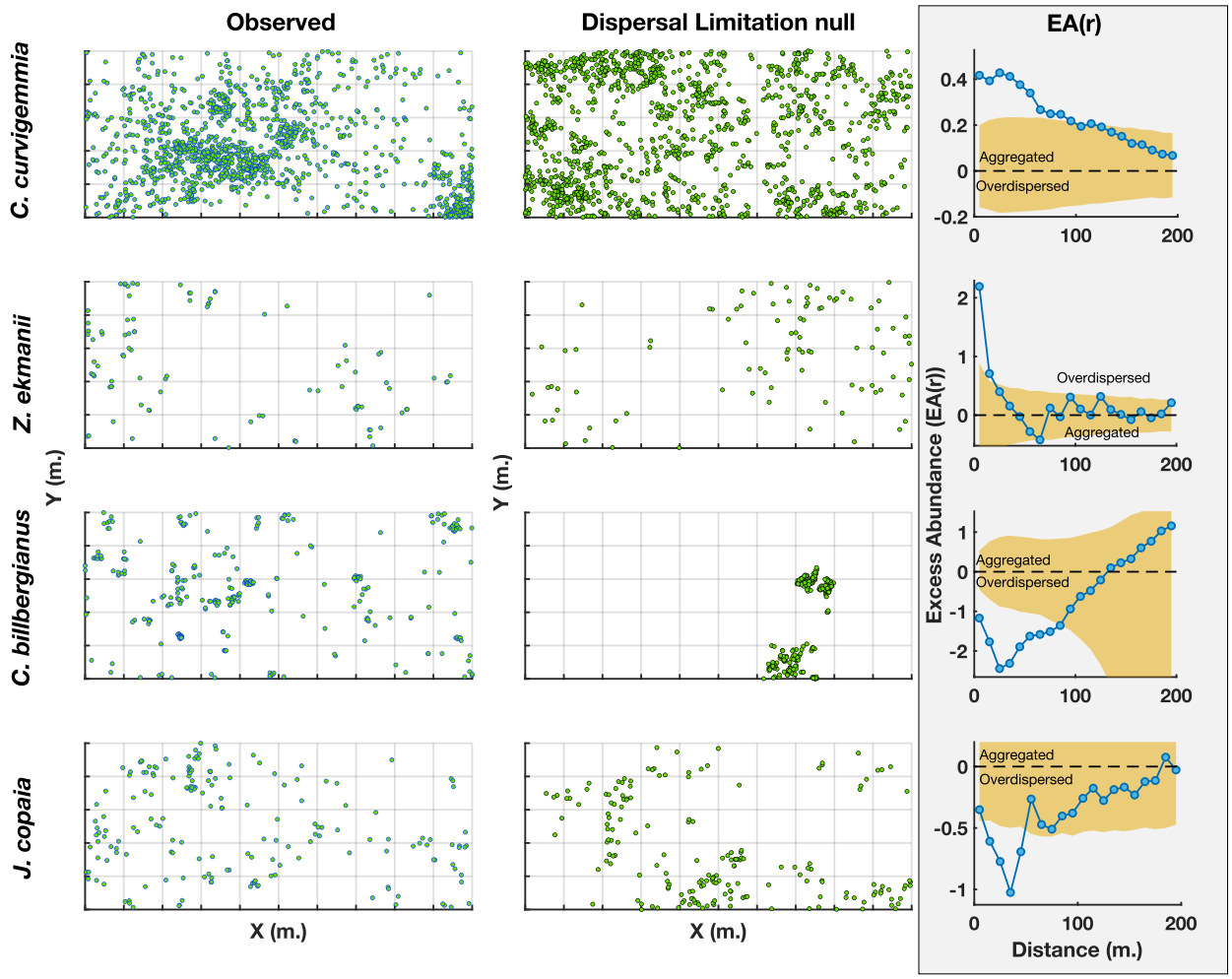


Extended Data Fig. 3 | Excess neighborhood Abundance at distance r (*EA*(r)), for all species with data on dispersal distance (in ref. 32) in 2015 with at least 25 individuals. (*N*, shown in each panel). The shaded region represents the 95% simulation envelope of the standard DL null. Note that for less common species, noisy patterns are to be expected. Negative values represent overdispersion w.r.t DL.

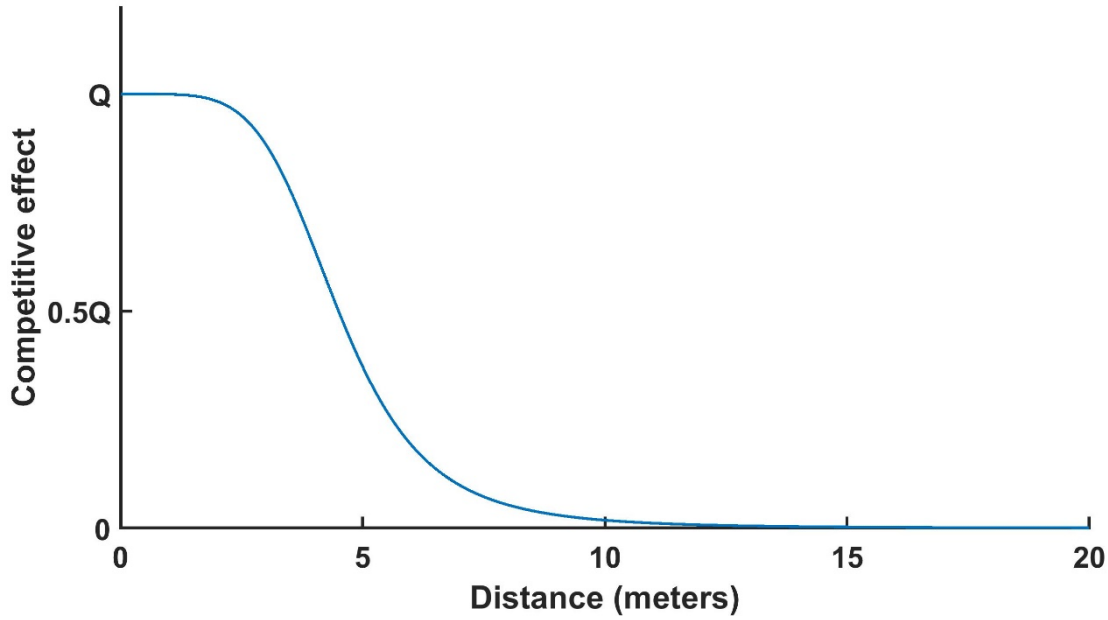

Alchornea costaricensis
N=115
Apeiba aspera
N=123
Astrocaryum standleyanum
N=132
Beilschmiedia pendula
N=143
Brosimum alicastrum
N=61
Cordia alliodora
N=57
Cordia bicolor
N=320
Cordia lasiocalyx
N=363
Coussarea curvigemmia
N=1346
Croton billbergianus
N=286
Cupania sylvatica
N=400
Dendropanax arboreus
N=38
Desmopsis panamensis
N=6499
Dipteryx panamensis
N=30
Drypetes standleyi
N=235
Eugenia oerstedeana
N=89
Faramea occidentalis
N=14374
Garcinia intermedia
N=459
Guapira standleyana
N=63
Guarea guidonia
N=1009
Guatteria dumetorum
N=73
Gustavia superba
N=604
Hasseltia floribunda
N=229
Heisteria acuminata
N=70
Heisteria concinna
N=312
Hirtella triandra
N=1615
Hura crepitans
N=77
Inga marginata
N=60
Jacaranda copaia
N=179
Lacmellea panamensis
N=57
Excess Abundance(r)
Distance (m.)


Excess Abundance(r)
Lindackeria laurina
N=34
Lonchocarpus latifolius
N=31
Luehea seemannii
N=56
Ocotea whitei
N=122
Oenocarpus mapora
N=1753
Ouratea lucens
N=516
Picramnia latifolia
N=372
Platymiscium pinnatum
N=29
Poulsenia armata
N=115
Pouteria reticulata
N=146
Prioria copaifera
N=134
Protium tenuifolium
N=300
Quararibea asterolepis
N=487
Randia armata
N=548
Rinorea sylvatica
N=1783
Simarouba amara
N=100
Socratea exorrhiza
N=424
Solanum hayesii
N=92
Spondias radlkoferi
N=38
Tabebuia rosea
N=25
Tabernaemontana arborea
N=239
Terminalia oblonga
N=33
Tetragastris panamensis
N=196
Trichilia pallida
N=196
Trichilia tuberculata
N=608
Triplaris cumingiana
N=91
Unonopsis pittieri
N=261
Virola sebifera
N=359
Virola surinamensis
N=111
Xylopia macrantha
N=427

Extended Data Fig. 4. | Excess neighborhood Abundance at distance r $\underline{EA}(r)$), for all species with data on dispersal distance (in ref. 33) in 2015 with at least 25 individuals (*N*, shown in each panel). The shaded region represents the 95% simulation envelope of the H2001 DL null. Note that for less common species, noisy patterns are to be expected. Negative values represent overdispersion w.r.t DL.



Extended Data Fig. 5 | **Spatial distributions of four atypical species and their Excess neighborhor Aundance at distance *r***, *EA*(*r*) (eq. 2). Each row presents a different species, while the columns present for each species the observed distribution (left) versus a single realization of the Standart DL null (center). *EA*(*r*) is presented in the rightmost, shaded column, with the orange region indicating the 95% simulation envelope of the DL null. *EA*(*r*) of 0 indicates a random distribution with respect to DL while values < 0 indicate overdispersion. All distribution maps are 1000 by 500 meters. *Coussarea curvigemmia* is a habitat specialist with a strong positive affiliation with the large low pletau habitat and negative asociations with the large high plateau and slope habitats[58], leading to aggreagation up to large distances. The three other species, *Zanthoxylum ekmanii*, *Croton billbergianus* and *Jacaranda copaia*, are gap specialists, as is evident from the very high relative growth rate of their fasters-growing juveniles[59]. They all have increased aggregation at short scales (0 – 30 meters), typical of gaps, compared with intermediate scales. *Z. ekmanii* is also estimated to have a very large dispersal distance (105 meters)[32] leading to wide expected dispersion under DL. Compared with this expected dispersion, the observed clumps of trees lead to the high positive values of *EA*(*r*) at small scales for this species.



Extended Data Fig. 6 | **Competition as a function of distance** in the simulation models. Eq. 5 is plotted with b_1 = 4.5 meters and b_2 = 5, which were used in all simulations.

	Excess Abundance (EA) at 75-125 meters		
Null	Values	Overdispersed	Aggregated
Standard	-0.59 ± 0.51 [-2.7, 0.57]	38 (8)	3 (1)
H2001	-0.43 ± 0.74 [-2.3, 3.6]	67 (23)	14 (3)
Lag	-0.49 ± 0.55 [-2.5, 0.46]	35 (5)	6 (1)
LDD	0.004 ± 0.5 [-2.3, 0.66]	12 (2)	29 (9)
Recruits	-0.021 ± 0.22 [-0.59, 0.36]	14 (6)	18 (3)
Fixed	-0.67 ± 0.57 [-2.9, 0.55]	38 (13)	3 (0)

Extended Data Table 1 | Distributions and significance of Excess Abundance within 75 – 125 meters for Barro Colorado tree species in 2015 using the six variants of the DL null model as references. We compare the abundance of conspecific neighbors within 75 - 125 meters to its expectation on a log scale, hence, negative values represent overdispersion and exponentiating the statistics would give the factor by which the observed density exceeds expectations. The mean ± SD and the range (in square brackets) of *EA*(75-125) are presented, as well as the number of overdispersed and aggregated species, along with the number of statistically significant (with $\alpha = 0.05$) results in brackets. Overdispersion is the norm accept in the Recruits and LDD models. The former indicates that recruitment at long distances is comparable to expectations. The latter could be a result of the (arguably) artificially high level of seeds dispersed with no dispersal limitation (10%).