

## LETTER

## Frequency-dependent selection for rare genotypes promotes genetic diversity of a tropical palm

Luke Browne<sup>1,2\*</sup> and Jordan Karubian<sup>1,2</sup>

<sup>1</sup>Department of Ecology and Evolutionary Biology, Tulane University, New Orleans, Louisiana

<sup>2</sup>Foundation for the Conservation of the Tropical Andes, Quito, Ecuador

\*Correspondence: E-mail: lbrowne2@tulane.edu

**Abstract**

Negative frequency-dependent selection among species is a key driver of community diversity in natural systems, but the degree to which negative frequency-dependent selection shapes patterns of survival and genetic diversity within species is poorly understood. In a 5-year field experiment, we show that seedlings of a tropical palm with rare genotypes had a pronounced survival advantage over seedlings with common genotypes, with effect sizes comparable to that of light availability. This ‘rare genotype advantage’ led to an increase in population-wide genetic diversity among seedlings compared to null expectations, as predicted by negative frequency-dependent selection, and increased reproductive success in adult trees with rare genotypes. These results suggest that within-species negative frequency-dependent selection of genotypes can shape genetic variation on ecologically relevant timescales in natural systems and may be a key, overlooked source of non-random mortality for tropical plants.

**Keywords**

Chocó rainforest, demography, genetic diversity, intraspecific variation, *Oenocarpus bataua*, rapid evolution, reproductive success, seedling recruitment, tropics.

*Ecology Letters* (2016)

**INTRODUCTION**

Understanding processes that maintain diversity is a key challenge in ecology and evolutionary biology. Negative frequency-dependent selection among species within a community, where the relative fitness of a species decreases as its frequency increases, is a fundamental mechanism that promotes species coexistence and maintains community diversity (Chesson 2000; Wright 2002; Vellend 2010; HilleRisLambers *et al.* 2012). Negative frequency-dependent selection among genotypes within a species is predicted to maintain genetic diversity (Ayala & Campbell 1974; Asmussen & Basnayake 1990), but the extent to which this happens on ecologically relevant timescales in natural populations remains unclear (Antonovics & Ellstrand 1984; Schoener 2011; Hughes *et al.* 2013; Takahashi & Kawata 2013). As genetic diversity within a species can have substantial impacts on population dynamics, communities and ecosystem function (Bohannan & Lenski 2000; Crutsinger *et al.* 2006; Johnson & Stinchcombe 2007; Lankau & Strauss 2007; Hughes *et al.* 2008; Pruitt & Ferrari 2011; Ehlers *et al.* 2016), there is a growing appreciation that considering intraspecific genetic variation and microevolutionary processes improves our understanding of the ecology of a system. Therefore, determining how negative frequency-dependent selection interacts with intraspecific genetic variation to drive non-random mortality and resulting patterns of diversity is likely to improve the ability of ecologists to understand natural systems.

In plants, frequency-dependent survival of genotypes within species arises through competition or cooperation of genotypes, or through interactions with predators and pathogens (Antonovics & Ellstrand 1984; File *et al.* 2012). A rare genotype advantage may come from reduced pressure from predators and pathogens that preferentially attack common

genotypes (Haldane 1949; Barrett 1988; Chaboudez & Burdon 1995) or the ability of rare genotypes to avoid competition with neighbours and exploit a wider range of resources (‘niche partitioning hypothesis’, File *et al.* 2012). Alternatively, related individuals with similar genotypes may outperform unrelated individuals if competitive growth is reduced when surrounded by related individuals (‘kin selection hypothesis’, Dudley *et al.* 2013; Crepy & Casal 2015). The majority of experiments assessing the role of genetic variation in plant population dynamics manipulate clonal genotypes of short-lived, herbaceous species (Hughes *et al.* 2008), which limits our ability to understand the impacts of genetic variation on survival in long-lived, outcrossed tree species (but see Hufford & Hamrick 2003; Jones & Comita 2008; Fine *et al.* 2013; Liu *et al.* 2015). More broadly, the relative influence of intraspecific genetic variation on survival compared to other commonly studied ecological factors like light availability and habitat quality is unclear, as few studies on genetic variation simultaneously compare these effects (Hughes *et al.* 2008; Crutsinger *et al.* 2013).

Both the genetic characteristics and mating system of adult trees play important roles in determining the genetic composition of offspring. Rare alleles within the adult population may also be rare within offspring if the frequencies of alleles among seedlings are similar to those of adults. Consequently, a rare genotype survival advantage among offspring will provide a relative reproductive advantage for adults with rare genotypes that pass rare alleles to their offspring. Alternatively, mating system – whether trees are outcrossing or self-fertilising – may influence the rarity of offspring genotypes, especially among groups of siblings (Antonovics & Ellstrand 1984). Outcrossing and biparental inbreeding rates (i.e. the degree of mating between relatives), along with diversity of pollen donors, can impact reproductive success in the absence

of negative frequency-dependent selection (Burd 1994; Eckert *et al.* 2010; Breed *et al.* 2012). Determining how adult genotype and mating system parameters influence the composition of seedling genotypes in populations undergoing negative frequency-dependent selection may allow us to predict variation in reproductive success among adults, which is useful for understanding the dynamics of populations.

In northwestern Ecuador, male *Cephalopterus penduliger* (the long-wattled umbrellabird) create high density, genetically heterogeneous groups of dispersed seeds, which may promote recruitment of the palm *Oenocarpus bataua* Mart. (Karubian *et al.* 2010, 2012; Scofield *et al.* 2012). *Oenocarpus bataua* is the 7th most abundant tree species in Amazonian forests (ter Steege *et al.* 2013) and is distributed widely across the Neotropics, from Panama to Bolivia (Henderson *et al.* 1995). With a 5-year field experiment, we showed that *O. bataua* seedlings planted with siblings had similar survival probabilities as those planted among unrelated seedlings, suggesting that kin cooperation was not a prominent aspect of this system (Karubian *et al.* 2016). However, we were unable to assess the relative rarity of seedling genotypes within the population and the role of negative frequency-dependent selection in shaping survival, genetic diversity or reproductive success. In this study, we build upon this previous work to address the following questions: (1) Is there a rare genotype advantage in seedling survival, and how does this compare to the effects of habitat characteristics? (2) If there is a rare genotype advantage, does this lead to increased population-level genetic diversity as predicted under negative frequency-dependent selection? and finally, (3) Does the rarity of the maternal genotype or mating system parameters (e.g. outcrossing and biparental inbreeding rates) better predict the rarity of an offspring's genotype and variation in adult reproductive success?

We hypothesised that seedlings with a rare genotype would have a survival advantage over common genotypes, potentially due to decreased competition or attack from natural enemies (Antonovics & Ellstrand 1984; File *et al.* 2012). We predicted this difference in survival would maintain genetic diversity over time as common genotypes are removed from the population (Ayala & Campbell 1974). We also hypothesised that the rarity of the maternal genotype would be a better predictor of the rarity of a seedling's genotype than would mating parameters because of extensive pollen dispersal and high rates of outcrossing in this system (Ottewell *et al.* 2012), and as a result, maternal trees with rare genotypes would have higher reproductive success than maternal trees with common genotypes.

We found that seedlings with rare genotypes had a higher probability of survival than seedlings with common genotypes and that, as predicted under negative frequency-dependent selection, population-level genetic diversity among surviving seedlings was higher than null expectations. Furthermore, the rarity of a seedling's genotype was largely a function of the rarity of the maternal tree genotype and this led to increased reproductive success for maternal trees with rare genotypes. This work shows that negative frequency-dependent selection of genotypes can maintain genetic variation on ecologically relevant timescales under natural conditions in a tropical forest. As well, this study suggests that intraspecific negative

frequency-dependent selection on genotypes may be a key, overlooked driver of non-random mortality in tropical forests.

## MATERIAL AND METHODS

### Study site and organism

The study was conducted at Bilsa Biological Station (BBS;  $0^{\circ}20'35.6''$  N  $79^{\circ}42'31.1''$  W; Fig. 1) in Esmeraldas Province, Ecuador. The study area receives between 2000 and 3500 mm of rainfall each year and ranges between 400 and 650 m in elevation (Clark *et al.* 2006; Carrasco *et al.* 2013). *Oenocarpus bataua* is a common, generalist palm tree (Family: Arecaceae), with adults growing to 20–40 m in height (Henderson *et al.* 1995). Within BBS, we have established a 130 ha study parcel where all mature *O. bataua* ( $n = 188$ ) were identified and mapped (Ottewell *et al.* 2012). Progenies of *O. bataua* in BBS are highly outcrossed, with little inbreeding, and *c.* 5.4 effective pollen donors per family group (Ottewell *et al.* 2012). Pollen is dispersed by various insect species in Coleoptera and Hymenoptera (Núñez-Avellaneda & Rojas-Robles 2008), with a mean pollen dispersal distance of 300 m in BBS (Ottewell *et al.* 2012). At BBS, *O. bataua* seeds are dispersed by a variety of terrestrial and arboreal mammals and birds (Browne and Karubian *unpublished*), with long-wattled umbrellabirds moving seeds on average 200–300 m (Karubian *et al.* 2012).

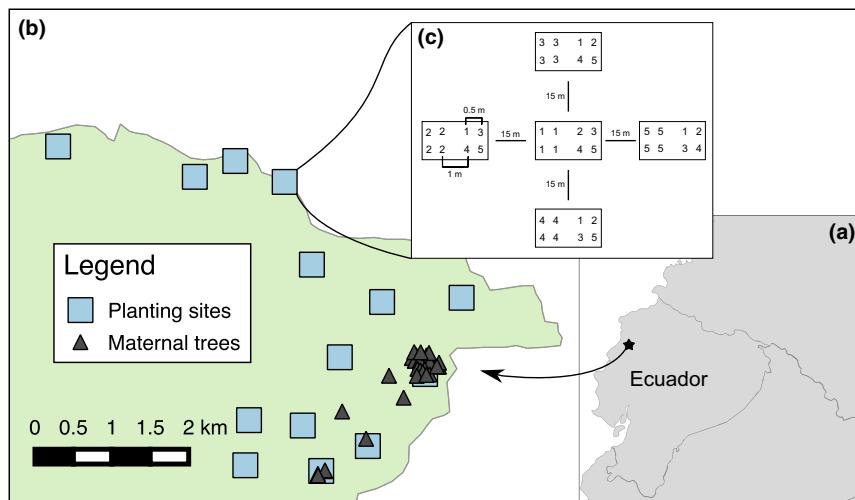
### Field experiment

In September 2008, we planted a cohort of 560 *O. bataua* seedlings at 14 sites in BBS ( $n = 40$  seedlings per site, Fig. 1). Within each site, we established five plots where eight seedlings were planted (Fig. 1). Within each plot, we planted four seedlings from a single maternal tree (i.e. half-to-full siblings), and four seedlings from four different maternal trees (i.e. likely unrelated, but half-sibs via shared pollen source possible). Prior to planting, we collected seeds from 30 adult *O. bataua* in BBS ( $n = 16$ –32 per adult; Fig. 1). These seeds were germinated in a nursery, individually marked with aluminium tags, and transplanted into the field at 3–4 months of age. Each seedling was between 30 and 60 cm in height at time of planting. For more details on the collection, germination and experimental design see Karubian *et al.* (2016).

We censused seedlings once every 9–12 months between September 2008 and July 2014 and recorded whether a plant was alive or dead during each census. Seven sites were located in breeding areas of long-wattled umbrellabirds (i.e. leks), whereas seven sites outside leks were considered control areas. In a previous study, survival and growth of *O. bataua* did not differ between lek and non-lek sites (Karubian *et al.* 2016), so we consider each site as an independent sampling location.

### Habitat characteristics

To determine the relative effects of habitat characteristics on survival compared with genetic rarity, we measured percent canopy cover using a spherical densiometer (Lemmon 1956), and the presence of trees  $> 50$  cm diameter at breast height (DBH) within 10 m of each plot ('large trees') at the time of



**Figure 1** Map of study area, maternal tree locations, planting sites and planting schematic. Inset (a) shows the location of study area near the western coast of Ecuador. Inset (b) shows the location of maternal trees (triangles), from which seeds were collected and germinated and the location of planting sites (squares,  $n = 14$ ) and outline of Bilsa Biological Station (BBS). Inset (c) shows representative planting arrangement for one site, where each number shows the ID of a maternal tree. Seedlings were planted in groups either surrounded by other seedlings from the same maternal tree, or surrounded by seedlings from other maternal trees.

planting in each plot, which were both previously shown to be strong predictors of *O. bataua* seedling survival (Karubian *et al.* 2016) and are likely related to light availability and habitat quality, particularly degree of past disturbance (Durães *et al.* 2013).

## Genotyping

Prior to transplanting each seedling in the field, we cut a 2 cm portion of leaf tissue from each seedling for genetic analysis, which was dried and stored in silica gel. We extracted genomic DNA with a Qiagen DNeasy kit and amplified an array of 10 microsatellite loci (Table S1), following the protocol and polymerase chain reaction (PCR) methods described in Ottewell *et al.* (2012). In this study, we use multiple microsatellite loci to sample the genome-wide genetic rarity of individuals, rather than identifying specific loci under selection. Microsatellites are generally assumed to be selectively neutral and appropriate for analysis of mating systems, population structure and genetic rarity; however, we acknowledge that microsatellites may be under selection in some situations if they play a functional role in gene activity or are located near loci under selection (Koshi & King 2006; Selkoe & Toonen 2006; Putman & Carbone 2014).

A subset of the 560 seedlings in this study was used to estimate mating system parameters and pollen dispersal for *O. bataua* (Ottewell *et al.* 2012). We genotyped those seedlings not used in Ottewell *et al.* (2012) and re-scored previously published genotypes from Ottewell *et al.* (2012) to ensure consistent allele calls. We obtained high quality genotypes from 511/560 (91%) of seedlings; samples were either not available or did not amplify for the remaining 49 seedlings. Details on marker quality control (e.g. null alleles, linkage disequilibrium) are available in Table S1.

We used previously published genotypes for 24/30 maternal trees in this study from Ottewell *et al.* (2012). For the six

remaining maternal trees, which we did not have genetic data for, we used the most-likely genotype generated by MLTR v3.2 (Ritland 2002), which estimates maternal genotypes based on progeny arrays. We excluded the genotype for one maternal tree (No. 64) because not enough genotyped progeny ( $n = 2$ ) were available to infer a reliable maternal genotype. To assess the robustness of our results using inferred rather than directly observed maternal genotypes, we analysed data with and without the most-likely maternal genotypes in our dataset.

## Seedling and maternal genetic rarity

To assess genetic rarity of individuals, we first estimated the relative frequency of alleles for separate reference populations of seedlings and maternal trees. For seedlings, not all potential parent trees were genotyped because some pollen sources were outside the 130 ha study parcel (Ottewell *et al.* 2012). We therefore estimated pollen and ovule allele frequencies, representing allele frequencies of the parental population, using MLTR v3.2 (Ritland 2002). Using either pollen, ovule or the diploid genotypes of the seedlings themselves produced highly correlated values of the genetic rarity index (see below) for seedlings (Pearson's  $R > 0.94$ ); we present results using pollen as reference allele frequencies. To estimate reference allele frequencies of maternal trees, we calculated allele frequencies using genotypes of 189 total adults in BBS from Ottewell *et al.* (2012) and additional maternal trees sampled in this study. The relative frequencies of alleles found in both seedling and adult populations (58/63 alleles) were strongly correlated (Pearson's  $R = 0.98$ ), suggesting that alleles rare among adults were also rare among seedlings.

To estimate individual level genetic rarity, we calculated a genetic rarity index, where for each allele in each individual, we determined the corresponding relative frequency of that allele in the reference population. We then averaged across

the relative frequencies of alleles within each locus with diploid genotype  $a, b$  to estimate the value of the genetic rarity index for individual  $i$  (eqn 1):

$$\text{Genetic rarity index}_i = \frac{\sum_{l=1}^L (p_a + p_b)_l}{2L} \quad (1)$$

where  $L$  is the total number of loci, and  $p_a$  and  $p_b$  are the frequencies of alleles  $a$  and  $b$  at locus  $l$  in the reference population. Note that  $p_a$  and  $p_b$  are equal in the case of homozygotes.

Lower values of the genetic rarity index correspond to individuals with genotypes containing rare, low frequency alleles compared to individuals harbouring relatively common, higher frequency alleles. The genetic rarity index of seedlings ranged from 0.231 to 0.449, with an average of 0.350 (Table S2). Maternal trees sampled in this study represented the range of genetic rarity found in the overall adult population at BBS (Fig. S1).

### Mating system parameters

We estimated mating system parameters for each family group of seedlings ( $n = 27$  with  $> 14$  progeny genotyped) using the program MLTR v3.2 and the Expectation-Maximisation (EM) algorithm with default settings (Ritland 2002). Specifically, we estimated multilocus outcrossing rate ( $t_m$ , the probability of outcrossing rather than self-fertilisation), biparental inbreeding ( $t_m - t_s$ , the degree of mating between relatives) and paternity correlation ( $r_p$ ), which estimates the proportion of full-sibs within a progeny array and provides information on the number of pollen sources. We estimated the variance of each family level mating parameter by bootstrapping individuals within families ( $n = 1000$ ).

### Statistical analyses

#### *Seedling survival model*

We modelled cumulative survival of each genotyped seedling ( $n = 511$ ) from planting to each census point with a hierarchical logistic regression. Survival was coded as a binary response (1 = survival, 0 = death) and modelled with a logit link function and a linear combination of the following predictor variables: genetic rarity index of each seedling, canopy cover (%), and whether there was a tree  $> 50$  cm DBH within 10 m of the plot (binary). Site ( $n = 14$ ) and plot ( $n = 70$ ) were treated as random effects to account for potential correlations between individuals within a plot and plots within a site, along with unmeasured environmental conditions common to each site and plot. The model took the following form:

$$\text{Survival}_{i,y} \sim \text{Bernoulli}(\hat{y}_{i,y})$$

$$\text{logit}(\hat{y}_{i,y}) = \alpha_y + \beta_{1y} \times \text{Canopy}_p + \beta_{2y} \times \text{Trees}_p + \beta_{3y} \times \text{RarityIndex}_i + \alpha_{\text{sites},y} + \alpha_{\text{plot}_p,y}$$

where  $\text{Survival}_{i,y}$  is the observed survival of the  $i$ -th seedling in the  $y$ -th census,  $\alpha_y$  is the intercept (i.e. global mean) for all seedlings for the  $y$ -th census,  $\beta_{1y}$  is the estimated effect of canopy cover for the  $y$ -th census,  $\beta_{2y}$  is the effect of large trees

for the  $y$ -th census,  $\beta_{3y}$  is the estimated effect of the genetic rarity index for the  $y$ -th census,  $\alpha_{\text{sites},y}$  is the random effect term for the  $s$ -th site in the  $y$ -th census and  $\alpha_{\text{plot}_p,y}$  is the random effect term for the  $p$ -th plot in the  $y$ -th census. In all cases, we used covariate values at the time of planting (e.g. % canopy cover in 2008).

We fit the model to our data with a Bayesian framework and Markov Chain Monte Carlo (MCMC) algorithms. Covariates were mean-centred and standardised to  $SD = 1$  prior to analyses to aid in model convergence and to allow comparison of coefficient estimates between parameters (Schielzeth 2010; Bolker *et al.* 2013). The mean, SD and range of covariates are provided in Table S2. We considered effects of covariates on survival to be significant if their 95% posterior credible intervals did not overlap with zero. We did not find high levels of correlation between covariates (Pearson's  $|R| < 0.10$ , Fig. S2). We set vaguely non-informative priors for all parameters:  $\text{Uniform}(-10, 10)$  for fixed regression parameters and  $\text{Uniform}(0, 10)$  for the standard deviation of random effect parameters. We estimated posterior distributions for each parameter using MCMC methods implemented in JAGS v4.1.0 (Plummer 2003) using the 'rjags' R package (R Core Development Team 2015; Plummer 2016). We ran three MCMC chains for each parameter for 50 000 iterations, thinning chains every 25 iterations, with the first 10 000 samples discarded as burn in. We assessed chain convergence by visually inspecting traceplots and ensuring that the potential scale reduction factor statistic ('rhat')  $< 1.10$  (Kéry 2010), which assesses the degree of within-chain to between-chain variance. We assessed model goodness-of-fit with a posterior predictive check (Fig. S3). Model code is available in Appendix S1, and genotype and survival data of seedlings is publicly available online (<https://dx.doi.org/10.6084/m9.figshare.3817023.v1>).

#### *Seedling genetic rarity model*

To assess whether mating parameters or maternal genetic rarity better predicted the rarity of a seedling's genotype, we used a multiple linear regression with the genetic rarity index of each seedling as the response variable and genetic rarity index of the maternal tree, biparental inbreeding ( $t_m - t_s$ ), paternity correlation ( $r_p$ ) and outcrossing rate ( $t_m$ ) as explanatory variables. The model was fit to the data using the 'lm' function in R (R Core Development Team 2015). Correlations between covariates were small ( $|R| \leq 0.40$ ) for all pairwise combinations (Fig. S4). Our sample sizes for this model included 497 seedlings and 27 maternal trees.

#### *Reproductive success model*

We defined the reproductive success of maternal trees as the proportion of seedlings surviving to the last census (5.5 years after planting). To determine whether the genetic rarity index of maternal trees or mating parameters (e.g. biparental inbreeding, paternity correlation and outcrossing rate) better predicted variation in reproductive success, we modelled reproductive success with a logistic regression and the logit link function. Because five maternal trees had 32 seedlings planted in the experiment rather than 16 seedlings for the other 25 maternal trees, we weighted the

logistic regression by the number of seedlings initially planted. The model was fit to the data using the ‘glm’ function in R (R Core Development Team 2015).

### Change in genetic diversity and allele frequencies over time

To determine if genetic diversity increased over the study period, we calculated population-level gene diversity ( $H_s$ ) (Nei 1987) and allelic richness (rarefied to  $n = 100$  samples) and compared these estimates to the expected changes in genetic diversity if survival was random with respect to genotype. Gene diversity ( $H_s$ ) provides information on the relative frequency of alleles in a population and may be more sensitive to short-term changes than allelic richness (Lowe *et al.* 2005). To generate the null distribution for each genetic diversity metric, we randomised seedling survival among individuals 999 times and calculated the metrics of genetic diversity at each census point. If the observed genetic diversity metric did not overlap with the 95% confidence interval (CI) of the null distribution, we rejected the null hypothesis that seedling survival was random with respect to genotype. Allelic richness, which is especially sensitive to differences in sample size, was rarefied to 100 samples to account for changes in population size across census sampling events, though qualitative results did not differ using un-rarefied allelic richness. Gene diversity and allelic richness were calculated in the R package ‘hierfstat’ (Goudet 2005; R Core Development Team 2015). To estimate levels of Type 1 error in our randomisation procedure, we simulated genotypes for 511 seedlings, with equivalent numbers of alleles at each of 9 loci as our real dataset. We then performed the exact randomisation procedure described above for 500 simulated data sets and calculated the percentage of false positives. Overall, rates of false positives were low (< 0.03%, Table S4).

To determine whether common, high-frequency alleles declined between parental and seedling generations, we calculated the change in allele frequency for each allele between the parental population (using pollen pool frequencies) and seedlings alive at the time of the last census. We estimated the correlation between allele frequency in the parental population

**Table 1** Standardised regression coefficients (posterior means) and 95% credible intervals (CI) for the effects of genetic rarity index of seedlings, canopy cover (%) and presence of large trees within a 10 m radius on cumulative survival of 511 *Oenocarpus bataua* seedlings from planting to 5.5 years after planting

Parameter	Estimate	2.5%	97.5%
Intercept	-0.84	-1.62	-0.12
Genetic rarity index (seedling)	<b>-0.34</b>	<b>-0.58</b>	<b>-0.11</b>
Canopy cover	<b>-0.76</b>	<b>-1.29</b>	<b>-0.28</b>
Large trees	<b>-0.69</b>	<b>-1.14</b>	<b>-0.30</b>

We modelled survival as a hierarchical logistic regression (see Methods for model details). We considered model parameters statistically significant if their 95% CI did not cross 0 (shown in bold). Predicted effects of each covariate on survival to 5.5 years are shown in Fig. 2, and mean, SD and range of each covariate are available in Table S2. Coefficient estimates for intermediate censuses prior to 5.5 year are shown in Table S3.

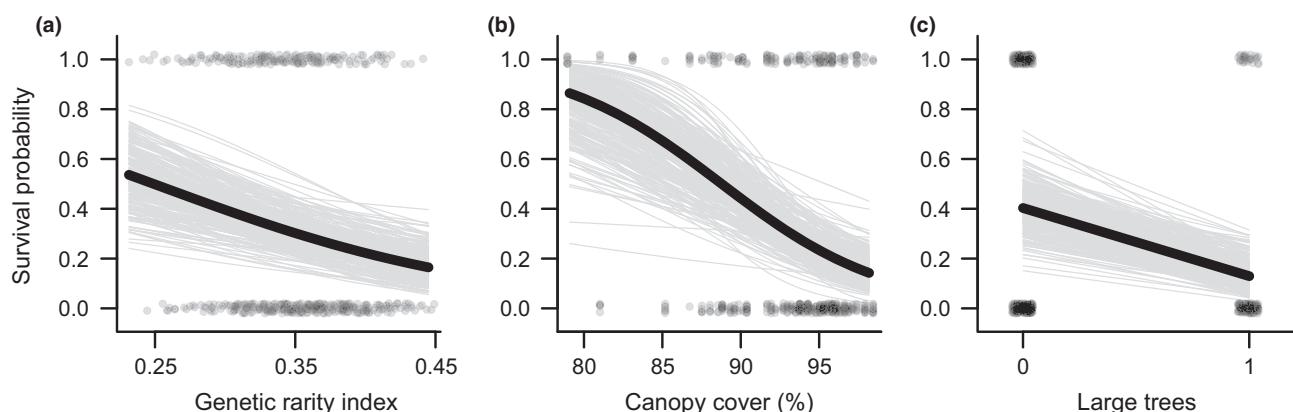
and the change in allele frequency between generations with a linear regression, controlling for locus identity, with the ‘lm’ function in R (R Core Development Team 2015).

## RESULTS

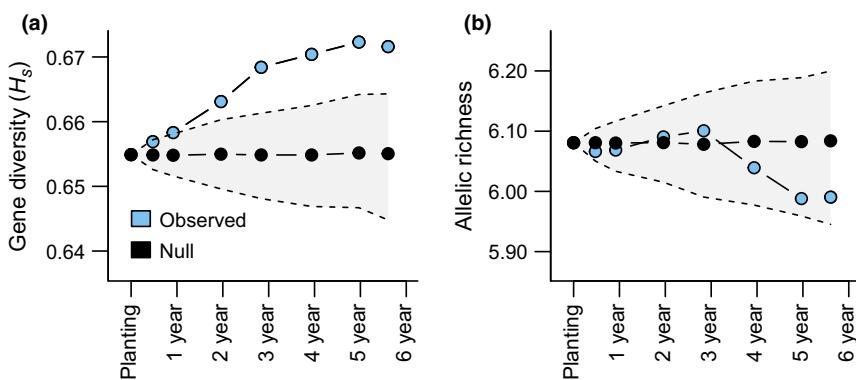
### Is there a rare genotype advantage in seedling survival, and how does this compare to the effects of habitat characteristics?

We found a consistent effect of seedlings with rare genotypes (i.e. those with lower values of the genetic rarity index) having an increased probability of survival starting 1 year post-planting and continuing until the final census 5.5 years after planting (Table 1, Table S3). A seedling with the rarest genotype had on average a 54% (37–70%, 95% credible interval) probability of surviving to 5.5 years compared to a 16% (9–25%) probability for a seedling with the most common genotype (Fig. 2a).

Comparing the effect sizes of habitat characteristics and genetic rarity, canopy cover and presence of large trees generally had a higher mean effect size on survival probability than did the genetic rarity index of seedlings, though the posterior



**Figure 2** Predicted survival probability of an *Oenocarpus bataua* seedling to 5.5 years after planting in relation to (a) genetic rarity index of seedlings, with lower values corresponding to rarer genotypes, (b) canopy cover (%) and (c) presence of large trees (> 50 cm diameter at breast height DBH) within a 10 m radius. Black line shows mean predicted relationship, and light grey lines show uncertainty in predicted relationship from 250 random samples from posterior distributions. Black dots show observed survival status for each seedling at 5.5 years after planting and are jittered along the y axis to reduce point overlap. Regression coefficients of the full model are shown in Table 1.



**Figure 3** Changes in population-level genetic diversity over time compared to the null expectation that survival is not related to genotype for (a) gene diversity ( $H_s$ ) and (b) allelic richness rarified to 100 samples. The blue circles show observed values at each census point. Black circles and dashed lines show the mean and 95% quantile range of null expectation values after randomising survival with respect to genotype of each seedling 999 times.

**Table 2** Standardised regression coefficients of multiple linear regression of the effects of genetic rarity index of maternal trees and mating parameters on the genetic rarity of seedlings

Parameter	Estimate	SE	T	P
Intercept	0.350	0.002	215.395	< 0.001
Genetic rarity index (maternal tree)	<b>0.018</b>	<b>0.002</b>	<b>11.053</b>	<b>&lt; 0.001</b>
Paternity correlation ( $r_p$ )	-0.002	0.002	-1.250	0.212
Biparental inbreeding ( $t_m - t_s$ )	-0.001	0.002	-0.492	0.623
Outcrossing rate ( $t_m$ )	<b>-0.004</b>	<b>0.002</b>	<b>-2.173</b>	<b>0.030</b>

Shown are the estimated coefficients, standard error (SE), *t*-statistic and *P* value. *P* values < 0.05 are shown in bold. Mean, SD and range of each covariate are available in Table S2.

95% credible interval of estimated effects overlapped (Table 1, Table S3, Fig. 2). The timing of the onset of effects was similar between large trees and the genetic rarity index, with effects on survival apparent at the 1 year post-planting and continuing until the final census, and the effect of canopy cover apparent after the 2 year census (Table S3).

#### Does a rare genotype advantage increase population-level genetic diversity?

Starting 2 years post-planting, population-level genetic diversity in terms of gene diversity ( $H_s$ ) was higher than would be expected if survival were random with respect to genotype; this pattern continued until the final census (Fig. 3a). In contrast, changes in allelic richness did not differ from null expectations at any census point (Fig. 3b). Comparing the change in allele frequencies between parental and seedling generations, the frequency of common alleles declined ( $t = -3.44$ , d.f. = 51,  $P = 0.001$ ), with the four most common alleles declining 3–9% (Fig. S5).

#### Does the rarity of the maternal genotype or mating system parameters better predict the rarity of an offspring's genotype and variation in adult reproductive success?

The genetic rarity (i.e. genetic rarity index) of a seedling was best predicted by the genetic rarity of its maternal tree (Table 2), which alone explained 20% of variation in seedling

genetic rarity (adjusted  $R^2$ ). Genetic rarity of seedlings was weakly related to outcrossing rate, such that higher levels of outcrossing led to rarer genotypes in seedlings, but not significantly correlated with other mating system parameters – paternity correlation or biparental inbreeding (Table 2).

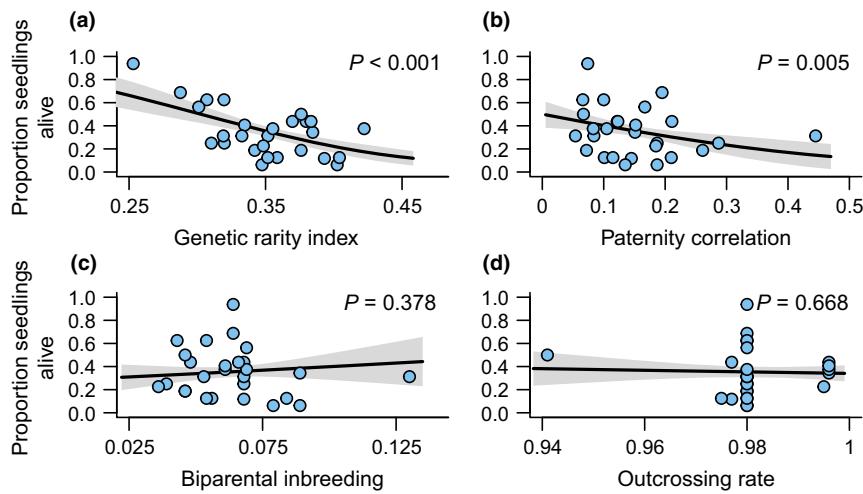
Reproductive success (proportion of seedlings alive 5.5 years post-planting) was highest among maternal trees with rare genotypes and low paternity correlation ( $r_p$ , the proportion of half-sibs among progeny; Fig. 4, Table S5). Reproductive success was not significantly correlated with biparental inbreeding or outcrossing rate (Fig. 4, Table S5). Excluding inferred maternal genotypes from MLTR produced similar results (Table S5).

## DISCUSSION

In this study, we provide evidence for a rare genotype advantage in survival for seedlings of a hyper-dominant tropical palm, *Oenocarpus bataua*. We found that population-level genetic diversity of seedlings increased over time compared to null expectations, consistent with the prediction that negative frequency-dependent selection maintains genetic diversity. We also provide evidence that the rarity of a seedling's genotype was primarily a function of the rarity of the maternal tree's genotype, rather than mating system parameters such as biparental inbreeding and outcrossing rates. Consequently, maternal trees with rare genotypes had higher reproductive success than did trees with common genotypes. Together, these results show that negative frequency-dependent selection can increase genetic diversity on ecologically relevant time-scales in natural systems and suggest that the influence of genotype on survival may be an under-appreciated driver of non-random seedling recruitment in tropical forests, with consequences for the maintenance of population-level genetic diversity and reproductive success of adult trees.

#### Rare genotype survival advantage

The observed rare genotype survival advantage is likely caused by either of two ecological mechanisms: frequency-dependent enemy attack or niche-partitioning. Pathogens and predators have been shown to respond in a frequency-dependent manner,



**Figure 4** Reproductive success of maternal trees (proportion of seedlings alive 5.5 years after planting) in relation to (a) the genetic rarity index of maternal trees, with lower values corresponding to rarer genotypes, (b) paternity correlation ( $r_p$ ), (c) biparental inbreeding ( $t_m - t_s$ ) and (d) outcrossing rate ( $t_m$ ). The proportion of seedlings alive was modelled with a logistic regression. Solid line shows the model predictions and shaded grey area shows uncertainty in model predictions ( $\pm 2 \text{ SE}$ ). The  $P$ -value for each parameter estimate is shown within the graph, and full parameter estimates are available in Table S5.

where common genotypes of plants are preferentially attacked (Chaboudez & Burdon 1995; Brunet & Mundt 2000), though the majority of relevant studies are from agricultural systems or clonally reproducing plants. Despite the importance of predators and pathogens in regulating populations and communities in the tropics (Swamy & Terborgh 2010; Terborgh 2012; Bagchi *et al.* 2014), the specificity of natural enemies to within-species genotypes in *O. bataua* and other tropical trees is currently not well understood, and represents a priority for future work. The niche-partitioning hypothesis, where rare genotypes experience less competition (File *et al.* 2012), provides an alternative mechanism that also predicts higher survival for rare genotypes. Given that tropical seedlings are thought to not be strong direct competitors (Paine *et al.* 2008; Swamy & Terborgh 2010), we question the degree to which niche-partitioning among seedlings explains the observed rare genotype advantage. However, experiments that manipulate and measure both resource use or attack from natural enemies among seedlings with rare and common genotypes would be necessary to tease apart the relative importance of these two mechanisms.

The strength and timing of the effect of genetic rarity on seedling survival was lower on average but comparable (95% CI overlapping) to the effects of habitat characteristics like canopy cover. Most studies of seedling recruitment include some measure of habitat like canopy cover to explain survival (see Karubian *et al.* 2016 for a detailed discussion of the influence of habitat on *O. bataua* seedling survival), but the genetic composition of seedlings is rarely considered. Likewise, most studies of the influence of genetic variation on survival fail to compare the effects of other commonly studied ecological factors (Hughes *et al.* 2008), which would help determine the relative importance of these effects on seedling recruitment. Our results suggest that quantifying within-species genetic variation is a worthwhile step towards fully explaining patterns of non-random survival in tropical plants.

### Consequences for genetic diversity

As predicted under negative frequency-dependent selection, we found that the frequency of common alleles declined between the parental and seedling generations and that population-level genetic diversity among seedlings increased over time, beginning 2 years post-planting. Observed increases in gene diversity ( $H_s$ ) were relatively small, but the fact that we detected change over a short timescale relative to the lifespan of the plant is remarkable. To our knowledge, this is among the first studies that show increased genetic diversity to result as a consequence of a rare genotype advantage in a tropical plant, and reinforces the idea that rapid evolutionary changes in allele frequencies can occur on ecologically relevant timescales as a result of micro-evolutionary processes (Hairston *et al.* 2005).

### Variation in adult reproductive success

We found higher reproductive success among maternal trees with rare genotypes. Due to the nature of frequency dependence, this reproductive advantage at the individual level may be temporary, as the frequency of once rare alleles increases in the population and no longer confer a survival advantage. In contrast, we observed an independent effect of increased pollen diversity (lower paternity correlation,  $r_p$ ) leading to increased reproductive success, which is consistent with other studies (Breed *et al.* 2012) and may arise from increased pollen competition (Charlesworth 1988). Although we did not measure spatial genetic structure or dispersal in this study, our results suggest that long-distance dispersal, which may bring rare genotypes into a population if there is spatial structuring of genotypes, may confer a survival advantage to dispersed individuals due to the rarity of their genotypes. Given that both seed dispersal (McConkey *et al.* 2012) and pollination processes (Eckert *et al.* 2010) are threatened by forces like climate change, hunting and habitat loss, understanding how these

neutral dispersal processes interact with negative frequency-dependent survival of offspring may aid in predicting how ongoing anthropogenic changes will impact plant populations.

While this study provides evidence that negative frequency-dependent selection can be a powerful driver of seedling mortality within a population of a dominant tropical tree species, additional studies on a wider range of taxa and sites are needed to assess the extent of this effect across species and across systems. It also remains to be investigated whether negative frequency-dependent selection acts across life stages (e.g. seed to seedling transition) and if observed increases in genetic diversity among seedlings persist to later life stages. Although intraspecific genetic variation, especially within dominant species, can enhance species coexistence in some systems (Vellend & Geber 2005; Ehlers *et al.* 2016), the degree to which negative frequency-dependent selection among genotypes in maintains community diversity in tropical forests remains unknown. Finally, expanding beyond microsatellite makers to genetic markers that target potentially adaptive loci, such as single-nucleotide polymorphisms, will likely help elucidate the mechanisms underlying negative frequency-dependent selection (Ouborg & Vriezen 2007).

## CONCLUSION

In conclusion, we show through a long-term planting experiment with a tropical palm that micro-evolutionary processes like negative frequency-dependent selection can impact seedling recruitment, genetic diversity and adult reproductive success on ecologically relevant timescales. Incorporating the interactions between genetic variation and micro-evolutionary processes like negative frequency-dependent selection in studies of tropical forests may therefore improve our ability to explain patterns of non-random mortality and variation in reproductive success, which in turn should aid ecologists in understanding and predicting the dynamics of tropical tree populations and communities.

## ACKNOWLEDGEMENTS

We are grateful for the support of FCAT (Fundación para la Conservación de los Andes Tropicales), the Jatun Sacha Foundation, and the Ecuadorian Ministry of the Environment. We thank C. Aulestia, J. Bermingham, D. Cabrera, F. Castillo, M. Chambers, J.B. Cox, Z. Diaz-Martin, G. DiRenzo, J. Olivo, K. Ottewell, A. Ruohomäki, M. Gonzalez and three anonymous reviewers and the editors for their invaluable assistance and helpful comments. This project was supported by the Conservation, Food & Health Foundation; Disney Conservation Fund; National Science Foundation (EAGER #1548548, DDIG #1501514 and a Graduate Research Fellowship to LB); National Geographic Society; Tulane University; and the United States Fish & Wildlife Service (NMBCA # 5605). All research was conducted with approval of the Ecuadorian Ministry of the Environment (Permit 010-2015-IC-FLO-FAU-DPE-MA).

## AUTHORSHIP

LB performed laboratory and statistical analyses, helped conduct fieldwork and prepared the first draft of the manuscript.

JK designed the field experiment and contributed to manuscript preparation.

## REFERENCES

- Antonovics, J. & Ellstrand, N.C. (1984). Experimental studies of the evolutionary significance of sexual reproduction. I. a test of the frequency-dependent selection hypothesis. *Evolution*, 38, 103–115.
- Asmussen, M.A. & Basnayake, E. (1990). Frequency-dependent selection: the high potential for permanent genetic variation in the diallelic, pairwise interaction model. *Genetics*, 125, 215–230.
- Ayala, F.J. & Campbell, C.A. (1974). Frequency-dependent selection. *Annu. Rev. Ecol. Syst.*, 5, 115–138.
- Bagchi, R., Gallery, R.E., Gripenberg, S., Gurr, S.J., Narayan, L., Addis, C.E. *et al.* (2014). Pathogens and insect herbivores drive rainforest plant diversity and composition. *Nature*, 506, 85–88.
- Barrett, J.A. (1988). Frequency-dependent selection in plant-fungal interactions. *Philos. Trans. R. Soc. B Biol. Sci.*, 319, 473–483.
- Bohannan, B.J.M. & Lenski, R.E. (2000). Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. *Ecol. Lett.*, 3, 362–377.
- Bolker, B.M., Gardner, B., Maunder, M., Berg, C.W., Brooks, M., Comita, L. *et al.* (2013). Strategies for fitting nonlinear ecological models in R, AD Model Builder, and BUGS. *Methods Ecol. Evol.*, 4, 501–512.
- Breed, M.F., Gardner, M.G., Ottewell, K.M., Navarro, C.M. & Lowe, A.J. (2012). Shifts in reproductive assurance strategies and inbreeding costs associated with habitat fragmentation in Central American mahogany. *Ecol. Lett.*, 15, 444–452.
- Brunet, J. & Mundt, C.C. (2000). Disease, frequency-dependent selection, and genetic polymorphisms: experiments with stripe rust and wheat. *Evolution*, 54, 406–415.
- Burd, M. (1994). Bateman's principle and plant reproduction: the role of pollen limitation in fruit and seed set. *Bot. Rev.*, 60, 83–139.
- Carrasco, L., Berg, K., Litz, J., Cook, A. & Karubian, J. (2013). Avifauna of the Mache Chindul Ecological Reserve, northwest Ecuador. *Ornitol. Neotrop.*, 24, 321–334.
- Chaboudez, P. & Burdon, J.J. (1995). Frequency-dependent selection in a wild plant-pathogen system. *Oecologia*, 102, 490–493.
- Charlesworth, D. (1988). Evidence for pollen competition in plants and its relationship to progeny fitness: a comment. *Am. Nat.*, 132, 298–302.
- Chesson, P. (2000). Mechanisms of maintenance of species diversity. *Annu. Rev. Ecol. Syst.*, 31, 343–366.
- Clark, J.L., Neill, D.A. & Asanza, M. (2006). *Floristic checklist of the Mache-Chindul Mountains of Northwestern Ecuador*. Contrib. from United States Natl. Herb, Smithsonian Institution, Washington, D.C.
- Crepy, M.A. & Casal, J.J. (2015). Photoreceptor-mediated kin recognition in plants. *New Phytol.*, 205, 329–338.
- Crutsinger, G.M., Collins, M.D., Fordyce, J.A., Gompert, Z., Nice, C.C. & Sanders, N.J. (2006). Plant genotypic diversity predicts community structure and governs an ecosystem process. *Science*, 313, 966–968.
- Crutsinger, G.M., Carter, B.E. & Rudgers, J.A. (2013). Soil nutrients trump intraspecific effects on understory plant communities. *Oecologia*, 173, 1531–1538.
- Dudley, S.A., Murphy, G.P. & File, A.L. (2013). Kin recognition and competition in plants. *Funct. Ecol.*, 27, 898–906.
- Durães, R., Carrasco, L., Smith, T.B. & Karubian, J. (2013). Effects of forest disturbance and habitat loss on avian communities in a Neotropical biodiversity hotspot. *Biol. Cons.*, 166, 203–211.
- Eckert, C.G., Kalisz, S., Geber, M.A., Sargent, R., Elle, E., Cheptou, P.O. *et al.* (2010). Plant mating systems in a changing world. *Trends Ecol. Evol.*, 25, 35–43.
- Ehlers, B.K., Damgard, C.F. & Laroche, F. (2016). Intraspecific genetic variation and species co-existence in plant communities. *Biol. Lett.*, 12, 20150853.
- File, A.L., Murphy, G.P. & Dudley, S.A. (2012). Fitness consequences of plants growing with siblings: reconciling kin selection, niche partitioning and competitive ability. *Proc. R. Soc. B Biol. Sci.*, 279, 209–218.

- Fine, P.V., Metz, M.R., Lokvam, J., Mesones, I., Zuniga, J.M.A., Lamarre, G.P. *et al.* (2013). Insect herbivores, chemical innovation, and the evolution of habitat specialization in Amazonian trees. *Ecology*, 94, 1764–1775.
- Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes*, 5, 184–186.
- Hairston, N.G., Ellner, S.P., Geber, M.A., Yoshida, T. & Fox, J.A. (2005). Rapid evolution and the convergence of ecological and evolutionary time. *Ecol. Lett.*, 8, 1114–1127.
- Haldane, J.B.S. (1949). Disease and evolution. *La Ric. Scientifica Suppl.*, 19, 1–11.
- Henderson, A., Galeano, G. & Bernal, R. (1995). *Field Guide to the Palms of America*. Princeton University Press, Princeton, New Jersey.
- HilleRisLambers, J., Adler, P.B., Harpole, W.S., Levine, J.M. & Mayfield, M.M. (2012). Rethinking community assembly through the lens of coexistence theory. *Annu. Rev. Ecol. Evol. Syst.*, 43, 227–248.
- Hufford, K.M. & Hamrick, J.L. (2003). Viability selection at three early life stages of the tropical tree, *Platypodium Elegans* (Fabaceae, Papilionoideae). *Evolution*, 57, 518–526.
- Hughes, A.R., Inouye, B.D., Johnson, M.T.J., Underwood, N. & Vellend, M. (2008). Ecological consequences of genetic diversity. *Ecol. Lett.*, 11, 609–623.
- Hughes, K.A., Houde, A.E., Price, A.C. & Rodd, F.H. (2013). Mating advantage for rare males in wild guppy populations. *Nature*, 503, 108–110.
- Johnson, M.T.J. & Stinchcombe, J.R. (2007). An emerging synthesis between community ecology and evolutionary biology. *Trends Ecol. Evol.*, 22, 250–257.
- Jones, F.A. & Comita, L.S. (2008). Neighbourhood density and genetic relatedness interact to determine fruit set and abortion rates in a continuous tropical tree population. *Proc. Biol. Sci.*, 275, 2759–2767.
- Karubian, J., Sork, V.L., Roorda, T., Durães, R. & Smith, T.B. (2010). Destination-based seed dispersal homogenizes genetic structure of a tropical palm. *Mol. Ecol.*, 19, 1745–1753.
- Karubian, J., Durães, R., Storey, J.L. & Smith, T.B. (2012). Mating behavior drives seed dispersal by the long-wattled umbrellabird *Cephalopterus penduliger*. *Biotropica*, 44, 689–698.
- Karubian, J., Browne, L., Cabrera, D., Chambers, M. & Olivo, J. (2016). Relative influence of relatedness, conspecific density, and microhabitat on seedling survival and growth of an animal-dispersed Neotropical palm. *Bot. J. Linn. Soc.*, 182, 425–438.
- Kéry, M. (2010). *Introduction to WinBUGS for Ecologists*. Academic Press, Burlington, MA.
- Koshi, Y. & King, D. (2006). Simple sequence repeats as advantageous molecular markers—mutations in evolution. *Trends Genet.*, 22, 252–259.
- Lankau, R.A. & Strauss, S.Y. (2007). Mutual feedbacks maintain both genetic and species diversity in a plant community. *Science*, 317, 1561–1563.
- Lemmon, P.E. (1956). A spherical densiometer for estimating forest overstory density. *For. Sci.*, 2, 314–320.
- Liu, X., Etienne, R.S., Liang, M., Wang, Y. & Yu, S. (2015). Experimental evidence for an intraspecific Janzen-Connell effect mediated by soil biota. *Ecology*, 96, 662–671.
- Lowe, A.J., Boshier, D., Ward, M., Bacles, C.F.E. & Navarro, C. (2005). Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees. *Heredity*, 95, 255–273.
- McConkey, K.R., Prasad, S., Corlett, R.T., Campos-Arceiz, A., Brodie, J.F., Rogers, H. *et al.* (2012). Seed dispersal in changing landscapes. *Biol. Conserv.*, 146, 1–13.
- Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press, New York, NY.
- Núñez-Avellaneda, L. & Rojas-Robles, R. (2008). Reproductive biology and pollination ecology of the milpesos palm *Oenocarpus bataua* in the Colombian Andes. *Caldasia*, 30, 101–125.
- Ottewell, K., Grey, E., Castillo, F. & Karubian, J. (2012). The pollen dispersal kernel and mating system of an insect-pollinated tropical palm, *Oenocarpus bataua*. *Heredity*, 109, 332–339.
- Ouborg, N.J. & Vriezen, W.H. (2007). An ecologist's guide to ecogenomics. *J. Ecol.*, 95, 8–16.
- Paine, C.E.T., Harms, K.E., Schnitzer, S.A. & Carson, W.P. (2008). Weak competition among tropical tree seedlings: implications for species coexistence. *Biotropica*, 40, 432–440.
- Plummer, M. (2003). JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. *Proc. 3rd Int. Work. Distrib. Stat. Comput.* ISSN 1609-395X.
- Plummer, M. (2016). rjags: Bayesian graphical models using MCMC. R Package version 4-6.
- Pruitt, J.N. & Ferrari, M.C.O. (2011). Intraspecific trait variants determine the nature of interspecific interactions in a habitat-forming species. *Ecology*, 92, 1902–1908.
- Putman, A.I. & Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecol. Evol.*, 4, 4399–4428.
- R Core Development Team. (2015). R: A language and environment for statistical computing.
- Ritland, K. (2002). Extensions of models for the estimation of mating systems using n independent loci. *Heredity*, 88, 221–228.
- Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods Ecol. Evol.*, 1, 103–113.
- Schoener, T.W. (2011). The newest synthesis: understanding the interplay of evolutionary and ecological dynamics. *Science*, 331, 426–429.
- Scofield, D.G., Smouse, P.E., Karubian, J. & Sork, V.L. (2012). Use of alpha, beta, and gamma diversity measures to characterize seed dispersal by animals. *Am. Nat.*, 180, 719–732.
- Selkoe, K.A. & Toonen, R.J. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecol. Lett.*, 9, 615–629.
- ter Steege, H., Pitman, N.C., Sabatier, D., Baraloto, C., Salomão, R.P., Guevara, J.E. *et al.* (2013). Hyperdominance in the Amazonian tree flora. *Science*, 342, 325–342.
- Swamy, V. & Terborgh, J.W. (2010). Distance-responsive natural enemies strongly influence seedling establishment patterns of multiple species in an Amazonian rain forest. *J. Ecol.*, 98, 1096–1107.
- Takahashi, Y. & Kawata, M. (2013). A comprehensive test for negative frequency-dependent selection. *Popul. Ecol.*, 55, 499–509.
- Terborgh, J. (2012). Enemies maintain hyperdiverse tropical forests. *Am. Nat.*, 179, 303–314.
- Vellend, M. (2010). Conceptual synthesis in community ecology. *Q. Rev. Biol.*, 85, 183–206.
- Vellend, M. & Geber, M.A. (2005). Connections between species diversity and genetic diversity. *Ecol. Lett.*, 8, 767–781.
- Wright, S.J. (2002). Plant diversity in tropical forests: a review of mechanisms of species coexistence. *Oecologia*, 130, 1–14.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

Editor, Liza Comita

Manuscript received 25 April 2016

First decision made 2 June 2016

Second decision made 8 August 2016

Manuscript accepted 8 September 2016