MR. MARTIN F BREED (Orcid ID : 0000-0001-7810-9696)

Article type : Biodiversity Research

Biodiversity Research

Standardised genetic diversity-life history correlates for improved genetic resource management of Neotropical trees

Running title: Standardised tree life history-population genetic correlates

Andrew J. Lowe[1]*§, Martin F. Breed[1]§, Henri Caron[2], Nathalie Colpaert[3], Christopher Dick[4], Bryan Finegan[5], Mike Gardner[1,6], Godelieve Gheysen[3], Rogério Gribel[7], J. Berton C. Harris[1,8], Antoine Kremer[2], Maristerra R. Lemes[9], Rogerio Margis[10], Carlos M. Navarro[5,11], Fabiano Salgueiro[12], Heidy M. Villalobos-Barrantes[13], Stephen Cavers[14]*

[1]Environment Institute and School of Biological Sciences, University of Adelaide, North Terrace, SA 5005, Australia. Email: andrew.lowe@adelaide.edu.au, martin.breed@adelaide.edu.au

[2]INRA, UMR 1202, BIOGECO, Cestas, France. Email: henri.caron@ecofog.gf, antoine.kremer@pierroton.inra.fr

[3]Department of Molecular Biotechnology, Faculty of Bioscience Engineering, Ghent University, B-9000 Ghent, Belgium. Email: nacol701@yahoo.com, godelieve.gheysen@ugent.be

[4]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, U.S.A. and Smithsonian Tropical Research Institute, Republic of Panama. Email cwdick@umich.edu

[5]Tropical Agricultural Research and Higher Education Center, Turrialba, Cartago, 30501, Costa Rica. Email: bfinegan@catie.ac.cr

[6]Biological Sciences, Flinders University, Adelaide, SA, Australia. Email: michael.gardner@flinders.edu.au

[7]Coordenação de Biodiversidade, Instituto Nacional de Pesquisas da Amazônia - INPA, Manaus, AM, Brazil. Email: rgribel@jbrj.gov.br
[8]Rainforest Trust, Warrenton, VA, U.S.A. Email: aramidopsis@gmail.com
[9]Laboratório de Genética e Biologia Reprodutiva de Plantas (LabGen), Instituto Nacional de Pesquisas da Amazônia, Manaus, Amazonas, Brazil. Email: mlemes@inpa.gov.br
[10]Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil. Email: rogerio.margis@ufrgs.br
[11]100 metros este UNED Betania, Condominium Sabanilla. Apartment 13. Turrialba, Costa Rica. Email: carlosmanuelnavarro@gmail.com
[12]Programa de Pós-graduação em Ciências Biológicas (Biodiversidade Neotropical), Universidade Federal do Estado do Rio de Janeiro (UNIRIO). Email: fabiano.salgueiro@gmail.com
[13]Centro de Investigación en Biología Celular y Molecular and Escuela de Química, Universidad de Costa Rica, 11501-2060 San José, Costa Rica. Email: heidy.villalobos@ucr.ac.cr
[14]Centre for Ecology and Hydrology, CEH Edinburgh, Bush Estate, Penicuik, Midlothian, Scotland, UK. Email: scav@ceh.ac.uk
§These authors contributed equally

Authors for correspondence: Andrew J. Lowe, telephone +61 8 8313 1149, fax: + 61 8 8313 4364, e-mail: andrew.lowe@adelaide.edu.au; Stephen Cavers, telephone +44 131 445 8552, e-mail: scav@ceh.au.uk

Words in the Abstract = 269
Number of words in main body of the paper = 4835

(A) ABSTRACT

(B) Aim. Life history traits and range size are key correlates of genetic diversity in trees. We used a standardized sampling protocol to explore how life history traits and range size relate to the magnitude, variance and structuring (both between and within population) of genetic diversity in Neotropical tree species.

(B) Location. The Neotropics

(B) Methods. We present a meta-analysis of new population genetic data generated for 23 Neotropical tree species (= 2966 trees, 86 populations) across a shared and broad geographic area. We compared established population genetic metrics across these species (e.g. genetic diversity, population structure, fine-scale genetic structure), plus we estimated the rarely used variance in genetic diversity among populations. We used a multivariate, maximum likelihood,

multi-model inference approach to explore the relative influence of life history traits and range size on patterns of neutral genetic diversity.

(B) Results. We found that pioneer and narrow range species had lower levels but greater variance in genetic diversity – signs of founder effects and stronger genetic drift. Animal dispersed species had lower population differentiation, indicating extensive gene flow. Abiotically dispersed and pioneer species had stronger fine-scale genetic structure, suggesting restricted seed dispersal and family cohort establishment.

(B) Main conclusions. Our multi-variable and multi-species approach allows ecologically relevant conclusions, since knowing whether one parameter has an effect, or one species shows a response in isolation, is dependent on the combination of traits expressed by a species. Our study demonstrates the influence of ecological processes on the distribution of genetic variation in tropical trees, and will help guide genetic resource management, and contribute to predicting the impacts of land-use change.

Keywords: effective population size, founder effects, gene flow, genetic resource management, seed dispersal

(A) INTRODUCTION

The life history traits and range size of tree species play critical roles in defining the magnitude and spatial arrangement of their genetic diversity (Duminil *et al.*, 2007; Meirmans *et al.*, 2011; Breed *et al.*, 2015; Broadhurst *et al.*, 2017). Consequently, traits and geographic ranges have become key considerations for planning genetic resource management (Montoya *et al.*, 2008; Breed *et al.*, 2013), the next generation of species distribution models (Swab *et al.*, 2012; Fordham *et al.*, 2014), and for underpinning studies of ecosystem function, conservation and restoration strategies (FAO, 2014; IPBES, 2014; Suding *et al.*, 2015).

For over 30 years, researchers have debated the relative influence of a range of life history traits and geographic patterns on population genetic variation in tree species (Loveless & Hamrick, 1984; Hamrick *et al.*, 1992; Hamrick *et al.*, 1993; Hamrick & Godt, 1996; Nybom & Bartish, 2000; Degen *et al.*, 2001; Hardy *et al.*, 2006; Duminil *et al.*, 2007; Montoya *et al.*, 2008; Meirmans *et al.*, 2011; Harata *et al.*, 2012; Broadhurst *et al.*, 2017). Previous meta-analyses have shown that range size, growth form and mating system can be important predictors of the magnitude of genetic diversity, and that growth form, seed dispersal vector and mating system are associated with species-wide genetic structure. While these previous meta-analyses have advanced our understanding of patterns of population genetic variation, most have explored single life history traits or geographic patterns in isolation (but see Hamrick & Godt, 1990;

Hamrick & Godt, 1996; Broadhurst *et al.*, 2017). Multivariate approaches are superior to single variable approaches when attempting to rank the importance of several competing predictor variables. Additional work is warranted to explore predictors of population genetic structure within populations, and whether patterns of population genetic variation within populations scale up to species-level patterns.

In this study, we present a meta-analysis of new data generated by a collaboration of researchers from ten institutions. Our study used standardized sampling of 23 tree species across a shared and broad geographic area – the Neotropics – to explore how key life history traits (seed dispersal vector and successional stage) and range size associated with the magnitude and structure of genetic diversity. We also estimated the standard deviation (σ) and coefficient of variation ($CV = \sigma/\bar{x}$) of genetic diversity among populations, which have rarely been used to compare differences among species since they were first proposed by Brown and Weir (1983) and further developed by Schoen and Brown (1991). We expect that variation in genetic diversity among populations will be higher in species that have traits that increase the risk of episodic but dramatic losses in genetic diversity, such as pioneer species that undergo strong founder effects (Davies *et al.*, 2010).

We used a multi-variable statistical approach that explores the relative influence of life history traits and range size on patterns of neutral genetic diversity, while accounting for potential correlations among characters. Our multi-variable and multi-species approach allows more ecologically relevant conclusions, since knowing whether one parameter has an effect, or one species shows a response in isolation, is dependent on the combination of traits expressed by a species. We investigated the following questions: (1) how do life history traits and range size relate to the magnitude, variance and structuring (both between and within population) of genetic diversity in 23 Neotropical tree species? (2) are these patterns consistent with findings from previous meta-analyses? Finally, we interpret our results in terms of relevance to the management of Neotropical tree genetic resources.

(A) METHODS

(B) Study species

Our 23 study species are all trees that largely occur in tropical and sub-tropical forest, with some extending into seasonally dry forests, are taxonomically resolved, and either dioecious or mixed to strongly outcrossing Neotropical trees (between 60-100% outcrossing Ward *et al.*, 2005), which limited variation in mating system and plant habit. Mating system and life form are characters that have been identified as confounding variables in previous studies, as both have been shown to have strong effects on patterns of neutral genetic diversity (Hamrick & Godt,

1996; Duminil *et al.*, 2007). To further minimize confounding effects, we used a consistent approach to study each species (see Fig. S1 in Supporting Information). Where possible, we standardized population sampling (mean ± SD populations per species = 3.7 ± 1.7, range = 2 to 9), focusing our efforts on populations of individually mapped trees (one population per species; mean ± SD n = 67 ± 18, range = 32 to 89), together with one or more populations close to (50-100 km) and distant from (>500 km) the mapped population, and focusing on a single geographic area (i.e. the Neotropics) which incorporated a significant proportion of the species' range in each case (Fig. 1; Table 1). We used standardized laboratory protocols and genetic markers (AFLPs Vos *et al.*, 1995) (details of laboratory protocols in Methods S1) to achieve consistency and comparability of the estimates of population genetic parameters (Vekemans & Hardy, 2004; Cavers *et al.*, 2005; Kremer *et al.*, 2005; Petit *et al.*, 2005; Hardy *et al.*, 2006; Jump & Peñuelas, 2007; Dick *et al.*, 2008).

Species were stratified by three variables central to standing hypotheses, based on data available at the time of our analysis (Loveless & Hamrick, 1984; Hamrick *et al.*, 1992; Hamrick *et al.*, 1993; Hamrick & Godt, 1996; Duminil *et al.*, 2007): range size, seed dispersal vector and successional stage (Table 2). Pollination syndrome has been an important factor to consider in studying genetic diversity, however we had insufficient variation in this parameter to include it in our study (18 of 23 were insect pollinated). These categories were used as predictor variables of patterns of variation in population genetic parameters. The 23 study species were from 22 different genera and 15 families, indicating that our species do not share patterns of population genetic variation due to recent ancestry, as might conceivably be the case for recently diverged sister species. For all study species, the magnitude and spatial distribution of genetic variation is independently acquired.

Species were defined as having wide (>50,000 km^2; n = 15) or narrow (<50,000 km^2; n = 8) ranges (local endemics, sensu Gentry, 1986). In theory, range size should have a positive effect on genetic diversity because larger ranges should correlate with larger effective population sizes (assuming effective density is constant) and reduce the influence of random genetic drift (Loveless & Hamrick, 1984). This hypothesis has been generally supported by empirical data (Hamrick *et al.*, 1992; Hamrick & Godt, 1996; Broadhurst *et al.*, 2017). Range size has also been hypothesized to have a negative effect on population differentiation because larger range size should correlate with greater dispersal ability and hence greater levels of gene flow (Loveless & Hamrick, 1984; Hamrick *et al.*, 1992). However, several studies found conflicting patterns in empirical data (Loveless & Hamrick, 1984; Hamrick *et al.*, 1992; Hamrick & Godt, 1996; Duminil *et al.*, 2007), a pattern that may be explained by sampling over geographic barriers

within wider ranging species, or a greater age of some widespread species (Dick & Heuertz, 2008; Dick *et al.*, 2013), allowing time for genetic differentiation to accrue.

Species were grouped as either late successional (n = 11) or pioneer (n = 12) based on functional trait data (traits included wood density, seed size and specific leaf area; see Table S1), plus field observations reported in primary literature (Forget, 1992; Huc *et al.*, 1994; Jones *et al.*, 2005; Flores *et al.*, 2006; Silva & Pinheiro, 2009). Pioneer species have been hypothesized to have lower genetic diversity (Loveless & Hamrick, 1984) and stronger spatial genetic structure (Davies *et al.*, 2010; Harata *et al.*, 2012), reflecting the habit of copious reproductive output and recruitment following disturbance, with few overlapping generations, which results in elevated genetic drift and founding of family groups plus a narrower window of opportunity for incoming gene flow (for exception, see Born *et al.*, 2008). Expectations of successional stage effects on population differentiation are mixed (Loveless & Hamrick, 1984), but generally, pioneer species are expected to exhibit higher levels of population differentiation because founder effects and few overlapping generations increase genetic drift, leading to rapid divergence among populations, and reduce opportunities for incoming gene flow.

We classified species according to their primary seed dispersal vector and sampled 13 animal-dispersed (*e.g.* bird, bat, monkey, rodent) and 10 abiotically dispersed species (*e.g.* gravity, explosive capsules, water, wind). Two species are known to undergo both abiotic and biotic seed dispersal (*Araucaria angustifolia*, *Calophyllum brasiliense*) but were grouped into the abiotically dispersed group in our analysis. Species with abiotically dispersed seeds are generally expected to have more limited seed dispersal than species with animal dispersed seeds (Howe & Smallwood, 1982), hence the former have been found to exhibit stronger population differentiation (Loveless & Hamrick, 1984; Hamrick *et al.*, 1992; Hamrick & Godt, 1996; Duminil *et al.*, 2007) and stronger spatial genetic structure (Loveless & Hamrick, 1984; Hamrick *et al.*, 1993; Harata *et al.*, 2012). The same reasoning suggests that population differentiation should correlate with spatial genetic structure due to the similar influence of seed dispersal (Dick *et al.*, 2008), but this remains largely untested.

(B) Genetic analysis

We performed a genome scan of an average of 228 AFLP loci (± 30 SE, range = 61 to 673) across our uniform sampling design of 23 Neotropical tree species from 96 populations, 2966 trees in total (Table 1; for details of AFLP laboratory methods see Methods S1). We estimated the percentage of polymorphic loci (P; n = 23 species), mean expected heterozygosity across populations (H_E; n = 23 species), and total expected heterozygosity within species (H_T; n = 23 species), and differentiation among populations (F_{ST}; n = 21 species) in AFLPsurv (Vekemans,

2002). Mean and total expected heterozygosity were tightly correlated ($r^2 = 0.85$), and to minimize redundancy in our results, our analysis will focus on mean expected heterozygosity.

We also calculated the standard deviation of P and H_E (σP and σH_E) and the coefficient of variation of P and H_E ($_{CV}P$ and $_{CV}H_E$) among populations, which are underutilized metrics to explore the variance in diversity across populations (and derived from a parameter first proposed by Brown and Weir in 1983, and further developed by Schoen and Brown 1991). The variance of population genetic diversity is rarely estimated in tree species because they usually exhibit very low differentiation for allelic frequencies and correspondingly low differentiation for diversity across populations. However, the variance in genetic diversity may be an important metric to observe in trees because it could, for example, be impacted by the strength of founder effects. Older, better-connected populations would be expected to have higher diversity than recently founded populations, as the latter may suffer from genetic bottlenecks (Davies *et al.*, 2010).

Spatial genetic structure was analysed in SPAGeDi (Hardy & Vekemans, 2002), following the procedure described in (Vekemans & Hardy, 2004), and using the Loiselle pairwise kinship coefficients between individuals, F_{ij} (Loiselle *et al.*, 1995). To define the slope of the relationship between average F_{ij} and geographic distance, we defined distance classes following the authors' recommendations, where, for each distance class, 50% of all individuals were represented at least once and the coefficient of variation of the number of times each individual represented was <1. Mean F_{ij} was plotted over the logarithm of the distance class. Pairwise kinship coefficients were regressed on the logarithm of pairwise distance to estimate the regression slope, *b*, and the significance of this slope was tested with 10,000 permutations. The strength of spatial genetic structure was then quantified by calculating Sp (Vekemans & Hardy, 2004). $Sp = -b/(F_1-1)$, where F_1 was the average kinship coefficient between individuals within the first distance class (all species: mean ± SE = 316 ± 137 m, $n = 19$; pioneer: mean ± SE = 232 ± 130 m, $n = 7$; late successional: mean ± SE = 364 ± 206 m, $n = 13$) and *b* was the regression slope of F_{ij} regressed on the logarithm of pairwise distance. Sp is a reciprocal of neighbourhood size, where low Sp indicates that the neighbourhood size is large and therefore weaker spatial genetic structure is observed.

(B) Statistics

We used general linear models in a maximum likelihood, multi-model inference framework (Burnham & Andersen, 2002) in R v. 3.4.1 (2017) to test for hypothesized relationships between the three life history and geographic predictor variables (range size, seed vector, successional stage) and the eight genetic response variables (P, σP, $_{CV}P$, H_E, σH_E, $_{CV}H_E$, F_{ST}, Sp) at the species level. We estimated Akaike's Information Criterion corrected for small sample sizes

(AICc; calculated in the MuMIn package – https://cran.r-project.org/web/packages/MuMIn/index.html) and Akaike weights ($_w$AIC) for each model (Burnham & Andersen, 2002). To select predictor variables of greatest importance to each response variable, we derived the index of the relative importance of predictor variable *i* ($AICc_i$), the sum of Akaike weights for all models that included parameter *i* (Burnham & Andersen, 2002; Giam & Olden, 2016). We also calculated ratios of the absolute value of the *t* statistic for each variable to judge variable importance, as suggested by Cade (2015).

We used a square root transformation for F_{ST} and $_{CV}H_E$, cube root transformation for Sp, and log base 10 transformation for σP and $_{CV}$P to meet the assumption of normality of residuals. We verified that the models met the statistical assumptions of general linear models by (1) testing the normality of residuals of fitted models by examining quantile-quantile plots (Crawley, 2007) and running Shapiro-Wilk tests (Shapiro & Wilk, 1965), and (2) checking for heteroscedasticity by examining plots of the residuals versus fitted values and scale-location (Crawley, 2007) as well as running Breusch–Pagan tests in the lmtest library (https://cran.r-project.org/web/packages/lmtest/index.html) (Breusch & Pagan, 1979). None of the top-ranked models had $P > 0.05$ for Shapiro-Wilk or Breusch–Pagan tests, but the multivariate F_{ST} and Sp models showed signs of heteroscedasticity in the residuals vs. fitted values plots. For P, we also used binomial generalized linear models with polymorphic loci as the successes and non-polymorphic loci as failures. The response variable for P was created by taking the sum of the loci that were polymorphic and not polymorphic for each species across all populations.

We ran our main analyses with the species that are known to undergo both abiotic and biotic seed dispersal (*Araucaria angustifolia* and *Calophyllum brasiliense*) classified as biotic rather than abiotic seed dispersers. In addition to species-level analysis, we also analysed the effects of the same predictor variables on population-level H_E and P data. For P, we used binomial generalized linear mixed-effect models with the lme4 package (https://cran.r-project.org/web/packages/lme4/citation.html) with species as the random effect. For H_E, we used Gaussian mixed-effect models with species as the random effect.

(B) Data accessibility

The genetic summary statistics supporting the findings of this study are available within the Supporting Information. The raw AFLP data will be uploaded to a data repository (e.g. Dryad) if our paper is accepted for publication.

(A) RESULTS

We found genetic diversity differences that correlated with range size (large vs. small range: mean P = 88.66 vs. 80.09, mean H_E = 0.31 vs. 0.25; $AICc_i$ P = 1.00; |t| ratio P = 0.97; $AICc_i$ H_E = 0.67; |t| ratio H_E = 1.00) as well as successional stage (late successional vs. pioneer: mean P = 90.98 vs. 80.82, mean H_E = 0.30 vs. 0.28; $AICc_i$ P = 1.00; |t| ratio P = 1.00; $AICc_i$ H_E = 0.67; |t| ratio H_E = 0.36), where pioneer and range restricted species had lower genetic diversity (Fig. 2; Table 3; Table S2, S3). These trends were largely consistent when comparisons were run individually within our three main study regions (south-east Brazil, Costa Rica, and French Guyana – inset maps in Fig. 1; Table S4), when binomial generalized linear models were used for P (Table S5), when mixed-effects models at the population-level were run (for P but not H_E; Table S6), and when univariate models were run (for both P and H_E; Table S7, S8). The percentage of polymorphic loci was positively correlated with expected heterozygosity (Fig. S2, S3; coefficient of determination r^2 = 0.51).

The standard deviation in the percentage of polymorphic loci (σP) and the coefficient of variation for both percentage of polymorphic loci ($_{CV}$P) and expected heterozygosity ($_{CV}H_E$) were each affected by successional stage (late successional vs. pioneer: mean σP = 4.35 vs. 10.70; $AICc_i$ σP = 0.87; |t| ratio σP = 1.00; σH_E did not differ; mean $_{CV}$P = 15.30 vs. 41.24; $AICc_i$ $_{CV}$P = 0.88; |t| ratio $_{CV}$P = 1.00; mean $_{CV}H_E$ = 0.04 vs. 0.01; $AICc_i$ $_{CV}H_E$ = 0.98; |t| ratio $_{CV}H_E$ = 1.00), and pioneer species generally exhibited greater variation of genetic diversity across populations within species than late successional species (Fig. 2; Table 3; Table S2, S3). These trends were consistent when we ran univariate models (Table S7). Variation in the percentage of polymorphic loci was correlated with the variance in expected heterozygosity (coefficient of determination r^2 = 0.58), but neither standard deviation metric was correlated with the corresponding mean estimate (σP ~ P: coefficient of determination r^2 = 0.07; σH_E ~ H_E: coefficient of determination r^2 = 0.07) or population differentiation (σP ~ F_{ST}: coefficient of determination r^2 = 0.03; σH_E ~ F_{ST}: coefficient of determination r^2 < 0.01).

Population differentiation was associated with range size (large vs. small range: mean F_{ST} = 0.126 vs. 0.049; $AICc_i$ F_{ST} = 0.86; |t| ratio F_{ST} = 1.00) and seed dispersal vector (animal vs. abiotic dispersal: mean F_{ST} = 0.072 vs.0.131; $AICc_i$ F_{ST} = 0.65; |t| ratio F_{ST} = 0.83), and animal dispersed and narrow range species had lower population differentiation (Fig. 2; Table 3; Table S2, S3). When we ran univariate models, range size remained as a strong predictor whereas seed dispersal vector was not (Table S7). Population differentiation did not correlate with mean geographic distance between populations (coefficient of determination r^2 = 0.04).

We observed marked differences in fine-scale spatial genetic structure associated with seed dispersal vector (animal vs. abiotic dispersal: mean Sp = 0.011 vs. 0.028; $AICc_i$ Sp = 0.71; |t| ratio Sp = 1.00) as well as successional stage (late successional vs. pioneer: mean Sp = 0.010

vs. 0.030; $AICc_i$ Sp = 0.62; |t| ratio Sp = 0.75), where abiotically dispersed and pioneer species had stronger fine-scale spatial genetic structure than biotically dispersed and late successional species (Fig. 2; Table 3; Table S2, S3). These trends were largely consistent when univariate models were run (Table S7). We also observed that population differentiation and spatial genetic structure were positively correlated, potentially driven by two species (*Pinus oocarpa* and *Vochysia ferruginea*), although our results were robust to bootstrapping (Fig. S3, S4; coefficient of determination r^2 = 0.40, ß = 0.133; n = 17; 2.5 and 97.5 percentiles of slope distribution of 10,000 bootstrap iterations = 0.003 and 0.232).

Our results were generally robust, but were less clear, when the two species that are known to undergo both abiotic and biotic seed dispersal were switched from abiotic to biotic seed dispersal classification (*Araucaria angustifolia*, *Calophyllum brasiliense*) (Table S9, S10).

(A) DISCUSSION

We show that with consistent sampling and analysis, range size, successional stage and seed dispersal vector are useful predictors of the magnitude, variance and structuring of genetic diversity. Our standardized approach included using the same genetic marker type, focusing our sampling to the same geographic region – the Neotropics – and sampling across a significant proportion of the species' range, which are factors that have not been controlled in previous studies (Duminil *et al.*, 2007). Our results should be interpreted with some caution as our study region does cross known biogeographic areas (Cavers & Dick, 2013), but our results appear robust to this sampling design. Further, since we analysed all characters together in a multi-variable, maximum likelihood, multi-model inference framework, which allowed more robust, ecologically relevant conclusions to be made by decoupling potential correlations among characters. We used a rarely used population genetic metric – the population genetic diversity standard deviation (σP, σH_E) – that proved sensitive to the successional stage of our study species. Together, our study provides the first consistently designed, multi-species study to explore whether species characteristics can predict the magnitude and structuring of genetic diversity.

Among our 23 study species, pioneer species had lower genetic diversity than late successional species. These findings support the hypothesis that pioneer species colonize gaps in sibling cohorts, leading to bottlenecks and the loss of genetic diversity (Nybom & Bartish, 2000; Davies *et al.*, 2010; Harata *et al.*, 2012). These findings indicate that pioneer species either risk losing adaptive variation during colonization due to genetic drift, which could impact their adaptive potential, or that these species are intrinsically well equipped to cope with reduced genetic diversity. Our findings are consistent with the review by Nybom and Bartish (2000), but

several other reviews did not observe an effect of successional stage on genetic diversity, potentially due to the limitations or level of variance of previous studies (Loveless & Hamrick, 1984; Hamrick *et al.*, 1992; Meirmans *et al.*, 2011).

Pioneer species also had higher variation in genetic diversity (for σP, but not σH_E). There has been little discussion in the literature on the drivers of variation in genetic diversity, but our findings provide justification for further investigation of this parameter, and indicate that succession and founder effects during gap-colonization are potentially important characters influencing this variable. This was most likely due to stronger population sampling effects during gap-colonization and scaling-up of genetic turnover from within-population to inter-population levels (Dick *et al.*, 2008), as supported by the positive association we observed between F_{ST} and Sp. It is perhaps expected that F_{ST} and Sp associate as both are measurements of isolation by distance processes, and as such, both are likely to be impacted by the same factors (e.g. limited seed dispersal). However, the strength of our conclusions is limited by the variable number of populations per species, which could adversely affect variance estimates, and we were unable to test alternative factors that could potentially influence variation in genetic diversity (e.g. historical demography, asymmetrical gene flow). As such, we suggest that simulation studies should be undertaken to develop testable hypotheses to better understand the causes and consequences of variation in genetic diversity, and the associations between fine-scale and population genetic structure.

We observed that range restricted species had lower genetic diversity than wide range species, which is consistent with the theory that large range sizes buffer genetic diversity (Loveless & Hamrick, 1984). Species with larger range sizes should also, at least in part, have greater dispersal capacity or maintain larger effective population sizes, and both would result in reduced effects of random genetic drift on genetic diversity. Our findings were consistent with some previous reviews (Hamrick *et al.*, 1992; Hamrick & Godt, 1996; Broadhurst *et al.*, 2017), but not others (Nybom & Bartish, 2000). As previously reported, we also found redundancy in the different measures of genetic diversity (Hamrick & Godt, 1990; Meirmans *et al.*, 2011; Broadhurst *et al.*, 2017), where the percentage of polymorphic loci was highly correlated with H_E.

Population genetic differentiation was strongly associated with seed dispersal vector, supporting previous theoretical expectations that animals have the capacity to disperse seeds further, on average, than abiotic means (e.g. wind, water; Loveless & Hamrick, 1984; Hamrick *et al.*, 1992; Hamrick & Godt, 1996; Duminil *et al.*, 2007) (for exceptions, see Nybom & Bartish, 2000; Meirmans *et al.*, 2011). Furthermore, population genetic differentiation was strongly associated with species range size. Species with wider ranges had stronger population genetic

differentiation than species with smaller ranges, which is contrary to the expectation that species with larger ranges have greater capacity to disperse and thus have lower population genetic differentiation (Loveless & Hamrick, 1984; Duminil *et al.*, 2007). We suggest that this result reflects our species-wide sampling efforts, where, despite the absence of an F_{ST}-geographic distance correlation, species with wider ranges are likely to also span biogeographic barriers (e.g. mountains, rivers), increasing isolation by distance. Future studies should explore this result in more detail by, for example, conducting multi-species studies within areas that do not contain major dispersal barriers and sampling many populations per species.

The strength of spatial genetic structure within populations appeared to be most influenced by seed dispersal vector and successional stage. Abiotically dispersed plants and pioneer species had stronger fine-scale spatial genetic structure than biotically dispersed and late successional species, most likely due to restricted seed dispersal and family cohorts establishing together. These findings are largely consistent with previous findings (Loveless & Hamrick, 1984; Hamrick *et al.*, 1993; Davies *et al.*, 2010; Harata *et al.*, 2012), and support the use of these categorical traits to predict levels of gene flow at local scales (Dick *et al.*, 2008).

(A) CONCLUSIONS

Protecting and managing forest genetic resources is an urgent priority, particularly as the extent of forest continues to be reduced and fragmented in the face of ongoing land clearance and climate change. Forest genetic resources provide the raw material underpinning population genetic health, adaptive potential, restoration and breeding. A recent international initiative by the FAO developed the Global Plan of Action on forest genetic resources (http://www.fao.org/3/a-i3849e.pdf) designed to promote their protection and sustainable management, and regional consortia such as EUFORGEN (http://www.euforgen.org/) have made great strides in identifying and protecting temperate forest genetic resources. Yet a huge task remains, even in well-resourced regions such as Western Europe, in finding effective proxies for predicting the levels and distribution of genetic diversity in tree species as manual characterization of all forest genetic resources is not tractable. The task, and need, is greatest in the high-diversity forests of the tropics. Currently, proxy prediction is most commonly done using abiotic environmental predictors and little biotic knowledge is built in to forecasting where genetic diversity lies.

Understanding how ecology relates to genetic diversity can provide important predictive power for the management of tree species. For example, knowing the relationships between key characteristics and genetic parameters allows prediction of tree species' capacity to overcome gaps in distribution or to re-connect fragmented populations (Loveless & Hamrick, 1984), which could be used to inform the spatial arrangement of connecting corridors. Patterns of neutral

genetic diversity can also provide a baseline against which studies of adaptive potential and adaptation can be set, where populations with higher levels of neutral genetic diversity may also be those with higher levels of adaptive potential (Sgrò *et al.*, 2011; Broadhurst *et al.*, 2017), and for seed collections, where diversity sampling can be better targeted (e.g. for seed banking, seed-based restoration; Broadhurst *et al.*, 2016) should be adjusted based on species characteristics. While it would be preferable to assign species to continuous character states and to incorporate phenotypic trait variation for analytical purposes, and new evidence may allow this, using the categorical assignment and neutral genetic data proved a powerful standpoint on which to make informed genetic resource management decisions.

The relationships we established between species characters and the magnitude, variance and structure of genetic diversity can be directly used to make much-needed genetic resource management recommendations (FAO, 2014; IPBES, 2014). Our results on the magnitude of population genetic diversity indicate that pioneer and narrow range species have lower genetic diversity, suggesting that species with these characters may either be at risk of poor adaptability due to low genetic diversity or that they are intrinsically well suited to adapt with low genetic diversity. It may therefore be required to use multiple seed sources when undertaking seed-based restoration for these pioneer or narrow range species, to augment their genetic diversity (Breed *et al.*, 2013; Breed *et al.*, 2016). We also implement an infrequently used metric that describes the variance in genetic diversity across populations, and showed that pioneer species had higher variance than late successional species. Thus, more populations of pioneer species are likely to be required if representative species-wide sampling is desired (e.g. for seed banking, seed production areas; Broadhurst *et al.*, 2016).

Our findings for population genetic differentiation indicate that it is possible to predict species responses to biogeographic barriers based on seed dispersal vector, which can be integrated with other data to delineate seed zones (Breed *et al.*, 2013), or used to optimize sampling of database collections for tracking timber stocks (Dormontt *et al.*, 2015). Spatial genetic structure was most affected by successional stage and seed dispersal vector, and this knowledge can be used to inform seed collection strategies on how to avoid closely related individuals and to ensure representative sampling of population-level variation (Lowe *et al.*, 2015). Our findings can also help advance species distribution models by allowing the incorporation of these population genetic functional group classifications into existing simulation frameworks (Fordham *et al.*, 2014; McCallum *et al.*, 2014), which are now an important basis for improving predictions of how land-use changes alter biodiversity and ecosystem services for forest tree species more generally (IPBES, 2014).

(A) ACKNOWLEDGEMENTS. This research was supported by EU funding through the INCO-DEV funding program under projects GENEO-TROPECO (ICA4-CT-2001-10101) and SEEDSOURCE (contract 003708). The Australian Research Council supported AJL and MFB (DE150100542 awarded to MFB; DP150103414 awarded to AJL and MFB). We thank Xingli Giam for statistical advice.

(A) BIOSKETCH

The authors have an interest in the genetic management of Neotropical tree species for conservation and restoration. AJL, AK, BF, CD, RG, ML, RM, CN proposed the funded project; AJL, SC, AK designed the study; AJL and SC coordinated field and lab work; HC, CD, BF, RG, ML, RM, CN, FS, HMV-B undertook field work; SC, HC, NC, GG, MG, RG, ML, RM, CMN, FS, HMV-B generated data; MFB, CD, BF, JBCH did analyses; MFB, AFL wrote the first draft of the manuscript, all authors contributed substantially to revisions. The authors declare no conflicts of interest.

(A) REFERENCES

Born, C., Kjellberg, F., Chevallier, M.-H., Vignes, H., Dikangadissi, J.-T., Sanguié, J., Wickings, E.J. & Hossaert-McKey, M. (2008) Colonization processes and the maintenance of genetic diversity: insights from a pioneer rainforest tree, *Aucoumea klaineana*. *Proceedings of the Royal Society Biological Sciences Series B*, **275**, 2171-2179.

Breed, M.F., Mortimer, P.E. & Lowe, A.J. (2016) Restoration: 'Garden of Eden' unrealistic. *Nature*, **533**, 469.

Breed, M.F., Stead, M.G., Ottewell, K.M., Gardner, M.G. & Lowe, A.J. (2013) Which provenance and where? Seed sourcing strategies for revegetation in a changing environment. *Conservation Genetics*, **14**, 1-10.

Breed, M.F., Ottewell, K.M., Gardner, M.G., Marklund, M.H.K., Dormontt, E.D. & Lowe, A.J. (2015) Mating patterns and pollinator mobility are critical traits in forest fragmentation genetics. *Heredity*, **115**, 108-114.

Breusch, T.S. & Pagan, A.R. (1979) A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 1287-1294.

Broadhurst, L.M., Jones, T.A., Smith, F.S., North, T. & Guja, L. (2016) Maximizing seed resources for restoration in an uncertain future. *BioScience*, **66**, 73-79.

Broadhurst, L.M., Breed, M.F., Lowe, A.J., Bragg, J., Catullo, R., Coates, D., Encinas-Viso, F., Gellie, N., James, E., Krauss, S. & Byrne, M. (2017) Genetic diversity and structure of the Australian flora. *Diversity and Distributions*, **23**, 41-52.

Brown, A. & Weir, B. (1983) Measuring genetic variability in plant populations. *Isozymes in plant genetics and breeding, part A* (ed. by S.D. Tanksley and T.J. Orton), pp. 219-239. Elsevier Science Publishing, Amsterdam.

Burnham, K.P. & Andersen, D.R. (2002) *Model selection and multimodel inference*, 2nd edn. Springer, New York.

Cade, B.S. (2015) Model averaging and muddled multimodel inferences. *Ecology*, **96**, 2370-2382.

Cavers, S. & Dick, C.W. (2013) Phylogeography of Neotropical trees. *Journal of Biogeography*, **40**, 615-617.

Cavers, S., Degen, B., Caron, H., Lemes, M.R., Margis, R., Salgueiro, F. & Lowe, A.J. (2005) Optimal sampling strategy for estimation of spatial genetic structure in tree populations. *Heredity*, **95**, 281-289.

Crawley, M. (2007) *The R Book*. John Wiley & Sons, Ltd, Chichester, UK.

Davies, S., Cavers, S., Finegan, B., Navarro, C. & Lowe, A. (2010) Genetic consequences of multigenerational and landscape colonisation bottlenecks for a Neotropical forest pioneer tree, *Vochysia ferruginea*. *Tropical Plant Biology*, **3**, 14-27.

Degen, B., Caron, H., Bandou, E., Maggia, L., Chevallier, M.H., Leveau, A. & Kremer, A. (2001) Fine-scale spatial genetic structure of eight tropical tree species as analysed by RAPDs. *Heredity*, **87**, 497-507.

Dick, C., Hardy, O., Jones, F. & Petit, R. (2008) Spatial scales of pollen and seed-mediated gene flow in tropical rain forest trees. *Tropical Plant Biology*, **1**, 20-33.

Dick, C.W. & Heuertz, M. (2008) The complex biogeographic history of a widespread tropical tree species. *Evolution*, **62**, 2760-2774.

Dick, C.W., Lewis, S.L., Maslin, M. & Bermingham, E. (2013) Neogene origins and implied warmth tolerance of Amazon tree species. *Ecology and evolution*, **3**, 162-169.

Dormontt, E.E., Boner, M., Braun, B., Breulmann, G., Degen, B., Espinoza, E., Gardner, S., Guillery, P., Hermanson, J.C. & Koch, G. (2015) Forensic timber identification: It's time to integrate disciplines to combat illegal logging. *Biological Conservation*, **191**, 790-798.

Duminil, J., Fineschi, S., Hampe, A., Jordano, P., Salvini, D., Vendramin, G.G. & Petit, R.J. (2007) Can population genetic structure be predicted from life-history traits? *American Naturalist*, **169**, 662-672.

FAO (2014) The state of the world's forest genetic resources. In. Commission on Genetic Resources for Food and Agriculture Organization of the United Nations, Rome.

Flores, O., Gourlet-Fleury, S. & Picard, N. (2006) Local disturbance, forest structure and dispersal effects on sapling distribution of light-demanding and shade-tolerant species in a French Guianian forest. *Acta Oecologica*, **29**, 141-154.

Fordham, D.A., Brook, B.W., Moritz, C. & Nogués-Bravo, D. (2014) Better forecasts of range dynamics using genetic data. *Trends in ecology & evolution*, **29**, 436-443.

Forget, P.-M. (1992) Regeneration ecology of *Eperua grandiflora* (Caesalpiniaceae), a large-seeded tree in French Guiana. *Biotropica*, **24**, 146-156.

Gentry, A.H. (1986) Endemism in tropical versus temperate plant communities. *Conservation biology: The science of scarcity and diversity* (ed. by M. Soule), pp. 153-181. Sinaner, Sunderland, MA.

Giam, X. & Olden, J.D. (2016) Quantifying variable importance in a multimodel inference framework. *Methods in Ecology and Evolution*, **7**, 388-397.

Hamrick, J. & Godt, M. (1990) Allozyme diversity in plant species. *Plant population genetics, breeding, and genetic resources.* (ed. by A.H.D. Brown, M.T. Clegg, A.L. Kahler and B.S. Weir), pp. 43-63. Sinauer Associates Inc., Sunderland, MA.

Hamrick, J.L. & Godt, M.J.W. (1996) Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **351**, 1291-1298.

Hamrick, J.L., Godt, M.J.W. & Sherman-Broyles, S.L. (1992) Factors influencing levels of genetic diversity in woody plant species. *New Forests*, **6**, 95-124.

Hamrick, J.L., Murawski, D.A. & Nason, J.D. (1993) The influence of seed dispersal mechanisms on the genetic structure of tropical tree populations. *Vegetatio*, **107**, 281-297.

Harata, T., Nanami, S., Yamakura, T., Matsuyama, S., Chong, L., Diway, B.M., Tan, S. & Itoh, A. (2012) Fine-scale spatial genetic structure of ten dipterocarp tree species in a Bornean rain forest. *Biotropica*, **44**, 586-594.

Hardy, O.J. & Vekemans, X. (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618-620.

Hardy, O.J., Maggia, L., Bandou, E., Breyne, P., Caron, H., Chevallier, M.H., Doligez, A., Dutech, C., Kremer, A., Latouche-HallÉ, C., Troispoux, V., Veron, V. & Degen, B. (2006) Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. *Molecular Ecology*, **15**, 559-571.

Howe, H.F. & Smallwood, J. (1982) Ecology of seed dispersal. *Annual Review of Ecology and Systematics*, **13**, 201-228.

Huc, R., Ferhi, A. & Guehl, J.M. (1994) Pioneer and late stage tropical rainforest tree species (French Guiana) growing under common conditions differ in leaf gas exchange regulation, carbon isotope discrimination and leaf water potential. *Oecologia*, **99**, 297-305.

IPBES (2014) Plenary of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. *Antalya, Turkey, 9-14 December 2013*,

Jones, F., Chen, J., Weng, G. & Hubbell, S. (2005) A genetic evaluation of seed dispersal in the neotropical tree *Jacaranda copaia* (Bignoniaceae). *American Naturalist*, **166**, 543-555.

Jump, A. & Peñuelas, J. (2007) Extensive spatial genetic structure revealed by AFLP but not SSR molecular markers in the wind-pollinated tree, *Fagus sylvatica*. *Molecular Ecology*, **16**, 925-936.

Kremer, A., Caron, H., Cavers, S., Colpaert, N., Gheysen, G., Gribel, R., Lemes, M., Lowe, A.J., Margis, R., Navarro, C. & Salgueiro, F. (2005) Monitoring genetic diversity in tropical trees with multilocus dominant markers. *Heredity*, **95**, 274-280.

566 Loiselle, B.A., Sork, V.L., Nason, J. & Graham, C. (1995) Spatial genetic structure of a tropical
567 understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany*, **82**,
568 1420-1425.

569 Loveless, M.D. & Hamrick, J.L. (1984) Ecological determinants of genetic structure in plant
570 populations. *Annual Review of Ecology and Systematics*, **15**, 65-95.

571 Lowe, A.J., Cavers, S., Boshier, D., Breed, M.F. & Hollingsworth, P.M. (2015) The resilience of
572 forest fragmentation genetics - no longer a paradox - we were just looking in the wrong
573 place. *Heredity*, **115**, 97-99.

574 McCallum, K.P., Guerin, G.R., Breed, M.F. & Lowe, A.J. (2014) Combining population genetics,
575 species distribution modelling and field assessments to understand a species vulnerability
576 to climate change. *Austral Ecology*, **39**, 17-28.

577 Meirmans, P.G., Goudet, J., IntraBioDiv, C. & Gaggiotti, O.E. (2011) Ecology and life history
578 affect different aspects of the population structure of 27 high-alpine plants. *Molecular*
579 *Ecology*, **20**, 3144-3155.

580 Montoya, D., Zavala, M.A., Rodríguez, M.A. & Purves, D.W. (2008) Animal versus wind
581 dispersal and the robustness of tree species to deforestation. *Science*, **320**, 1502-1504.

582 Nybom, H. & Bartish, I.V. (2000) Effects of life history traits and sampling strategies on genetic
583 diversity estimates obtained with RAPD markers in plants. *Perspectives in Plant Ecology,*
584 *Evolution and Systematics*, **3**, 93-114.

585 Petit, R.J., Duminil, J., Fineschi, S., Hampe, A., Salvini, D. & Vendramin, G.G. (2005)
586 Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant
587 populations. *Molecular Ecology*, **14**, 689-701.

588 R Core Team (2017) *R: A language and environment for statistical computing*. R Foundation for
589 Statistical Computing.

590 Schoen, D.J. & Brown, A. (1991) Intraspecific variation in population gene diversity and
591 effective population size correlates with the mating system in plants. *Proceedings of the*
592 *National Academy of Sciences*, **88**, 4494-4497.

593 Sgrò, C.M., Lowe, A.J. & Hoffmann, A.A. (2011) Building evolutionary resilience for
594 conserving biodiversity under climate change. *Evolutionary Applications*, **4**, 326-337.

595 Shapiro, S.S. & Wilk, M.B. (1965) An analysis of variance test for normality (complete samples).
596 *Biometrika*, **52**, 591-611.

597 Silva, A.L.G.d. & Pinheiro, M.C.B. (2009) Reproductive success of four species of *Eugenia* L.
598 (Myrtaceae). *Acta Botanica Brasilica*, **23**, 526-534.

599 Suding, K., Higgs, E., Palmer, M., Callicott, J.B., Anderson, C.B., Baker, M., Gutrich, J.J.,
600 Hondula, K.L., LaFevor, M.C., Larson, B.M.H., Randall, A., Ruhl, J.B. & Schwartz,
601 K.Z.S. (2015) Committing to ecological restoration. *Science*, **348**, 638-640.

602 Swab, R.M., Regan, H.M., Keith, D.A., Regan, T.J. & Ooi, M.K.J. (2012) Niche models tell half
603 the story: spatial context and life-history traits influence species responses to global
604 change. *Journal of Biogeography*, **39**, 1266-1277.

605 Vekemans, X. (2002) AFLP-SURV version 1.0. Laboratoire de Genetique et Ecologie Vegetale.
606 *Universite Libre de Bruxelles, Belgium*,

607 Vekemans, X. & Hardy, O.J. (2004) New insights from fine-scale spatial genetic structure
608 analyses in plant populations. *Molecular Ecology*, **13**, 921-935.

609 Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J.,
610 Peleman, J., Kuiper, M. & Zabeau, M. (1995) AFLP: A new technique for DNA
611 fingerprinting. *Nucleic Acids Research*, **23**, 4407-4414.

612 Ward, M., Dick, C.W., Gribel, R. & Lowe, A.J. (2005) To self, or not to self ... A review of
613 outcrossing and pollen-mediated gene flow in neotropical trees. *Heredity*, **95**, 246-254.
614
615
616 **(A) SUPPORTING INFORMATION**
617 Additional Supporting Information may be found in the online version of this article
618
619 **Methods S1** AFLP methods
620 **Table S1** Functional trait data (sourced from TRY) by succession category
621 **Table S2** Genetic diversity, population genetic differentiation and fine-scale spatial genetic
622 structure data for the study species
623 **Table S3** Population genetic patterns investigated with general linear models
624 **Table S4** Mean population genetic diversity in the three main regions of our study
625 **Table S5** Binomial generalized linear model results for the effects of the species characters on P

Table S6 Population genetic patterns investigated at the population level with generalized mixed effects models

Table S7 Univariate population genetic patterns investigated with general linear models

Table S8 Univariate binomial generalized linear model results for the effects of species characters on P

Table S9 Population genetic patterns investigated with general linear models with the two species that are known to undergo both abiotic and biotic seed dispersal classified as biotic rather than abiotic

Table S10 Binomial generalized linear model results for the effects of the species characters on P with the two species that are known to undergo both abiotic and biotic seed dispersal classified as biotic rather than abiotic

Figure S1 We used a consistent study design, including species selection, population sampling and the genetic marker used

Figure S2 Plot of percentage of polymorphic loci against mean expected heterozygosity (H_E)

Figure S3 Plot of first two principal components of a PCA of the genetic response variables, showing the associations of the five main population genetic parameters

Figure S4 Plot of population differentiation (F_{ST}) estimates against fine-scale spatial genetic structure (Sp) for each species

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

650 **Table 1** Family, range size, seed dispersal vector, successional stage, number of AFLP loci scored, number of populations sampled and total number
651 of samples across all populations of the study species.

Species	Family	Range size	Seed dispersal vector	Successional stage	Loci	$n_{\text{populations}}$ ($n_{\text{total samples}}$)
Anacardium occidentalis	Anacardiaceae	Wide	Biotic (birds)	Pioneer	181	2 (89)
Araucaria angustifolia	Araucariaceae	Wide	Mixed (gravity, birds)	Shade tolerant	673	9 (190)*
Bocoa prouacensis	Fabaceae	Narrow	Biotic (monkeys, bats)	Shade tolerant	88	2 (123)*
Calophyllum brasiliense	Clusiaceae	Wide	Mixed (gravity, water, bats)	Shade tolerant	519	4 (159)*
Chrysophyllum sanguinolentum	Sapotaceae	Wide	Biotic (monkeys)	Shade tolerant	149	3 (121)*
Dicorynia guianensis	Fabaceae	Narrow	Abiotic (gravity)	Shade tolerant	134	3 (92)*
Eperua falcata	Fabaceae	Narrow	Abiotic (gravity)	Shade tolerant	107	4 (169)*
Eperua grandiflora	Fabaceae	Narrow	Abiotic (gravity)	Shade tolerant	173	3 (113)*
Eugenia uniflora	Myrtaceae	Wide	Biotic (birds)	Pioneer	205	5 (71)*
Hyeronima alchorneoides	Euphorbiaceae	Wide	Biotic (birds)	Shade tolerant	213	5 (244)*
Jacaranda copaia	Bignoniaceae	Wide	Abiotic (wind)	Pioneer	125	3 (92)
Lecythis ampla	Lecythidaceae	Wide	Biotic (rodents)	Shade tolerant	242	6 (157)*
Lonchocarpus costaricensis	Fabaceae	Narrow	Abiotic (wind)	Pioneer	487	6 (114)
Pinus oocarpa	Pinaceae	Wide	Abiotic (wind)	Pioneer	383	3 (132)*
Sideroxylon capiri	Sapotaceae	Narrow	Biotic (monkeys, bats)	Pioneer	254	4 (86)*
Simarouba amara	Simaroubaceae	Wide	Biotic (monkeys, birds)	Pioneer	157	5 (136)*
Swietenia macrophylla	Meliaceae	Wide	Abiotic (wind)	Pioneer	242	2 (106)*
Symphonia globulifera	Clusiaceae	Wide	Biotic (monkeys, bats)	Shade tolerant	184	3 (153)*
Tapirira guianensis	Anacardiaceae	Wide	Biotic (monkeys, birds)	Pioneer	198	4 (173)*
Tetragastris panamensis	Burseraceae	Wide	Biotic (monkeys, birds)	Shade tolerant	208	2 (115)*

Virola michelii	Myristicaceae	Narrow	Biotic (monkeys, birds)	Pioneer	240	2 (55)
Vochysia ferruginea	Vochysiaceae	Wide	Abiotic (wind)	Pioneer	61	4 (183)*
Vouacapoua americana	Fabaceae	Narrow	Biotic (rodents)	Shade tolerant	92	2 (93)*

*The larger population was spatially mapped for fine-scale spatial genetic structure analysis

Table 2 Predicted effects of three species characteristics (range size, seed dispersal, succession stage) on the levels, variance and structure of population genetic diversity. The process, support for and against these predictions from the literature are indicated, as are the findings from our study.

Characteristic	Prediction	Process	Support for	Support against	This study
Range size	Species with larger ranges have higher genetic diversity	Weaker genetic drift	(Hamrick & Godt, 1990; Hamrick *et al.*, 1992; Hamrick & Godt, 1996)	(Nybom & Bartish, 2000)	Species with larger ranges had higher genetic diversity
	No predicted effect on genetic diversity standard deviation				No effect detected
	Species with larger ranges have weaker population genetic differentiation	Greater colonizing ability connects populations	(Hamrick & Godt, 1990; Hamrick *et al.*, 1992; Hamrick & Godt, 1996)	(Loveless & Hamrick, 1984; Duminil *et al.*, 2007)	Species with larger ranges had stronger population genetic differentiation
	No predicted effect on spatial genetic structure				No effect detected
Seed dispersal	No predicted effect on genetic diversity				No effect detected
	No predicted effect on genetic diversity standard deviation				No effect detected
	Species with biotically dispersed seeds have weaker population genetic differentiation	Wider seed dispersal	(Loveless & Hamrick, 1984; Hamrick *et al.*, 1992; Hamrick & Godt, 1996; Duminil *et al.*, 2007)	(Nybom & Bartish, 2000; Meirmans *et al.*, 2011)	Species with biotically dispersed seeds had weaker population genetic differentiation
	Species with biotically dispersed seeds	Wider seed dispersal	(Loveless & Hamrick,		Species with biotically

	have weaker spatial genetic structure		1984; Hamrick *et al.*, 1993; Harata *et al.*, 2012)		dispersed seeds had weaker spatial genetic structure
Successional stage	Pioneer species have lower genetic diversity	Founder effects leading to genetic bottlenecks	(Nybom & Bartish, 2000; Davies *et al.*, 2010; Harata *et al.*, 2012)	(Loveless & Hamrick, 1984; Hamrick *et al.*, 1992; Meirmans *et al.*, 2011)	Pioneer species had lower genetic diversity
	Pioneer species have larger genetic diversity standard deviations	Stronger population sampling effects during colonization	(Dick *et al.*, 2008)		Pioneer species had larger variance in genetic diversity
	Pioneer species have stronger population genetic differentiation	Founder effects increase genetic drift, leading to rapid differentiation			No effect detected
	Pioneer species have stronger spatial genetic structure	Founder effects leading to family group establishment	(Davies *et al.*, 2010; Harata *et al.*, 2012)	(Born *et al.*, 2008)	Pioneer species had stronger spatial genetic structure

655 **Table 3** Population genetic patterns investigated with general linear models. % DE, percentage
656 deviance explained by the model; ΔAICc, indicator of difference between model Akaike's
657 Information Criterion corrected for small samples sizes (AICc) and the minimum AICc in the
658 model set; *w*AICc, weight that show the relative likelihood of model *j*; *k*, the number of parameters;
659 only models with a ΔAICc less than the null model (~ 1) are shown.

Model	**% DE**	**ΔAICc**	***w*AICc**	***k***
Population expected heterozygosity (H_E)				
H_E ~ range	29.53	0.00	0.39	2
H_E ~ range + succession	38.02	0.01	0.39	3
H_E ~ range + seed	29.74	2.89	0.09	3
H_E ~ range + seed + succession	38.19	3.25	0.08	4
H_E ~ 1	0.00	5.39	0.03	1
Expected heterozygosity variance (σH_E)				
σH_E ~ 1	0.00	0.00	0.32	1
Expected heterozygosity coefficient of variation ($_{CV}H_E$)				
$_{CV}H_E$ ~ succession	37.48	0.00	0.63	2
$_{CV}H_E$ ~ seed + succession	38.61	2.54	0.18	3
$_{CV}H_E$ ~ range + succession	37.48	2.96	0.14	3
$_{CV}H_E$ ~ range + seed + succession	38.63	5.84	0.03	4
$_{CV}H_E$ ~ 1	0.00	8.14	0.01	1
Percentage of polymorphic loci variance (σP)				
σP ~ succession	24.56	0.00	0.43	2
σP ~ seed + succession	30.81	0.97	0.27	3
σP ~ range + succession	25.04	2.81	0.11	3
σP ~ 1	0.00	3.82	0.06	1
Percentage of polymorphic loci coefficient of variation ($_{CV}P$)				
$_{CV}P$ ~ succession	24.37	0	0.47	2
$_{CV}P$ ~ seed + succession	29.79	1.25	0.25	3
$_{CV}P$ ~ range + succession	24.45	2.94	0.11	3
$_{CV}P$ ~ 1	0	3.76	0.07	1

Population differentiation (F_{ST})				
F_{ST} ~ range + seed	38.52	0.00	0.48	3
F_{ST} ~ range	23.35	1.54	0.22	2
F_{ST} ~ range + seed + succession	39.97	3.00	0.11	4
F_{ST} ~ 1	0.00	4.38	0.05	1
Fine-scale spatial genetic structure (Sp)				
Sp ~ succession + seed	38.30	0.00	0.29	3
Sp ~ range + seed + succession	46.62	1.01	0.17	4
Sp ~ range + seed	34.77	1.06	0.17	3
Sp ~ succession	19.29	1.84	0.11	2
Sp ~ seed	15.97	2.61	0.08	2
Sp ~ range	15.02	2.82	0.07	2
Sp ~ 1	0.00	3.07	0.06	1

NB: Model results for effects of the species characters on P are in Table S8 since we ran binomial generalized linear models.

Figure Legends

Fig. 1 Maps showing the location of sampled populations for all species. Inset maps show greater detail of Costa Rica (CR), French Guyana (FG) and southeast Brazil (SEB). Populations of each species are represented by unique symbols, and the population in which trees are individually mapped is underlined.

Fig. 2 Partitioning of population genetic metrics for Neotropical trees across life history traits and geographic distribution. In plots A-C and D-F, two parameters per plot are shown for each column: A-C - percentage of polymorphic loci (P, filled squares, on left) and expected heterozygosity (H_E, open squares, on right); D-F - standard deviation of polymorphic loci (σP, filled squares, on left) and expected heterozygosity (σH_E, open squares, on right). In plots G-I and J-L a single parameter per plot is shown for each column: G-I = population differentiation (F_{ST}); J-L = spatial genetic structure (Sp). Range size shown in columns A, D, G, J: seed dispersal vector in columns B, E, H, K: and successional stage in C, F, I, L. The index of the relative importance of each predictor variable ($AICc_i$) is shown. All samples sizes are in Table 1.