Using molecular markers to investigate genetic diversity, mating system and gene flow of Neotropical trees

Authors: Vinson C. C.[1*],Mangaravite, E.[2], Sebbenn, A. M.[3], Lander, T. A.[4]

Addresses:

[1] Departamento de Botânica, Universidade de Brasília, 70910-900, Brasília, DF, Brazil

[2]Departamento de Microbiologia, Universidade Federal de Viçosa, Viçosa. MG, Brazil

[3] Instituto Florestal de São Paulo, C.P. 1322, CEP 01059-970 São Paulo, SP, Brazil

[4] Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, UK

[*]Corresponding author: Christina Vinson, Departamento de Botânica, Universidade de Brasília, Brasília, DF, Brazil. E-mail: ccvison@gmail.com

Word count for main text: words: 8670

Keywords: genetic structure; pollen dispersal; seed dispersal; microsatellites; SNPs; DNA sequencing

Abstract

Genetic analysis provides a robust method for understanding the ways in which human interventions in ecosystems may affect genetic diversity and species survival, providing an empirical basis for evidence-based conservation and management planning. This review aims to (1) describe the most informative and widely-used molecular markers for genotype analysis, and describe the methods for using that genotype data to understand (2) genetic diversity and structure; and (3) mating system, pollen and seed dispersal. For each area of investigation we discuss the types of analyses that can be performed, the software available for the analyses, what the analyses indicate, and provide examples from the literature of studies using these approaches. The review focuses on Neotropical tree species because Neotropical forests are highly diverse, are under particular threat from anthropogenic land-use change, and trees are useful model organisms to investigate patterns over evolutionary time and across geographical space.

Introduction

Species diversity underpins ecosystem function (Balvanera et al. 2006), and, in turn, individual species depend upon ecosystems for their survival (Chapin et al. 2000). The Neotropical forests are among the most biologically diverse ecosystems on Earth and are under particular threat from anthropogenic land-use change. Between 2000 and 2012, 2.3 million km^2 of forest were lost globally, and 0.8 million km^2 of new forest was established. Of this, tropical rainforest ecozones totalled 32% of global forest cover loss, nearly half of which was in South America (Hansen et al. 2013). These losses and changes to forest ecosystems are contributing to global species declines (see Chapin et al. 2000; Sala et al. 2000), and to loss of genetic diversity among the species that remain (Lowe et al. 2005. but see Lander et al. 2010). Loss of alleles and erosion of heterozygosity can reduce resilience and adaptive potential of individual species in the face of abiotic and biotic environment change (Reed and Frankham 2003), which, as noted above, threatens continued ecosystem function. Many human interventions in the environment, such as logging, forest fragmentation and climate change, may affect genetic diversity and survival of forest organisms (e.g. Leal et al. 2016), and therefore genetic analyses are valuable in conservation and management planning for both individual species and whole ecosystems (Allendorf 2017).

The aims of this review are to (1) describe the most informative and widely-used molecular markers for genotype analysis, and describe the methods for using that genotype data to understand (2) genetic diversity and structure; and (3) mating system, pollen and seed dispersal; processes which can be directly affected by human activity. We discuss the main challenges in the use of these molecular markers, sampling strategies to maximize the value of the genetic data that can be produced, and key directions for future work in this area. This review focuses on Neotropical tree species because Neotropical forests are highly diverse and are under

particular threat from anthropogenic land-use change, and trees are useful model organisms to investigate patterns over evolutionary time and across geographical space. This is because trees can maintain high levels of diversity and accumulate new mutations slowly, but can also display rapid local adaptation (Petit and Hampe 2006). Thus Neotropical tree species preserve an imprint of Neotropical evolutionary history (Cavers and Dick 2013).

Molecular markers

Molecular markers allow us to determine the genotypes of individuals, and are a powerful means of investigating and quantifying genetic diversity, spatial genetic structure, tree mating systems, gene flow, and breeding patterns of tree species. These markers can be divided into two groups based on the pattern of inheritance: dominant and codominant markers. Dominant markers (Randomly Amplified Polymorphic DNAs, Inter-Simple-Sequence-Repeats, Amplified Fragment Length Polymorphisms, minisatellites) can be used to identify two genotype classes: the homozygous recessive genotype (aa), and genotypes containing a dominant allele (AA or Aa). Dominant markers do not allow the user to distinguish the homozygous dominant genotype from the heterozygous genotype. In contrast, co-dominant markers (Allozymes, Restriction Fragment Length Polymorphism, Microsatellites, Single Nucleotide Polymorphisms) can be used to identify all three genotype classes (AA, Aa and aa).

Allozymes were the first molecular markers used to distinguish individuals (Lewontin and Hubby 1966). Genotypes are based on enzyme size, which can be identified by charge separation of different sized products in an electrophoresis gel. In tree species, allozyme variation has been widely used to estimate genetic and taxonomic relationships among different populations, and also has been applied in conservation genetic analysis (e.g. Hamrick et al. 1992).

The ability to study DNA variation directly began with the discovery of restriction enzymes by Arber, Smith and Nathans during the 1960s (Schlötterer 2004). Restriction enzymes have been used for Restriction Fragment Length Polymorphism (RFLP; Kerem et al. 1989) and minisatellite (Gill et al. 1985; Jeffreys et al. 1990) analysis. RFLPs are variations of one or more bases within 4 to 8 base pairs (bp) at a given locus, which the enzyme will recognize, and have been used for genetic diversity studies (Wünsch and Hormaza 2002). Minisatellites are length variations at loci with at least 6 bp length tandem repeats. These length variations can occur due to unequal crossing over between sister chromatids or replication slippage (Jeffreys et al. 1990). Minisatellite data provide such high variability that they were the first data to be called 'DNA fingerprints' (Gill et al. 1985), and they have also been used to distinguish tree species (e.g. Besse et al. 1993). Nevertheless, this marker is a dominant multi locus, thus, not well suited for mapping or association studies.

The development of the polymerase chain reaction (PCR, Saiki et al. 1985) facilitated the proliferation of new molecular marker methods during the 1990s. Markers such as PCR-RFLP (Rasmussen 2012), Randomly Amplified Polymorphic DNAs (RAPDs; Williams et al. 1990), Inter-Simple-Sequence-Repeats (ISSRs; Zietkiewicz et al. 1994); and Amplified Fragment Length Polymorphisms (AFLPs; Vos et al. 1995) were developed and have allowed analysis of species with no or little prior genetic information available. For tree species, which are generally non-model organisms, these new markers allowed estimations of gene dispersal (e.g. Hardy et al. 2006), and genetic structure and diversity. These data were often generated with a view to genetic conservation (e.g. Feyissa et al. 2007; Jump and Peñuelas 2007), or with a focus on neutral and adaptive drivers of genetic divergence (Brousseau et al. 2015). However, these markers can give low reproducibility and, apart from PCR-RFLP, they are dominant multi locus markers (Newton et al. 1999).

Currently, the three approaches most frequently used to analyze genetic variation of populations and individuals are microsatellites or Simple Sequence Repeats (SSRs), Single Nucleotide Polymorphisms (SNPs), and Restriction Site-Associated DNA Sequencing (RADseq). Similar to minisatellites, SSRs are sequences of a repeat motif, ranging from 1 to 6 bp, and varying in length between individuals (Tautz 1984). In addition to the facility of this PCR-based method, several characteristics of SSRs, such as high mutation rates and high levels of polymorphism, codominance, and random distribution across the genome (Zane et al. 2002), make them a more sensitive tool for revealing population genetic structure and gene flow than the markers described previously. SSR analysis requires the development of a specific primer pair flanking the individual SSR site. The primers are developed through a number of methods (see reviews Zane et al. 2002; Kalia et al. 2011), and may be unique to the species or transferable between closely related species (Barbará et al. 2007). The ability of at least some SSRs to be transferred across species has facilitated their use in Neotropical tree studies (e.g. Lander et al. 2007; Sebbenn et al. 2011; Caetano et al. 2012; Scotti-Saintagne et al. 2013; Vinson et al. 2015; Mangaravite et al. 2016; Tambarussi et al. 2017).

It is unclear when SNPs were first mentioned as markers for identification of polymorphism, but they rose to prominence with their use in the study of the human genome (e.g. Wang et al. 1998). These markers consist of a substitution in a single nucleotide in either coding or non-coding sequences, and exist genome-wide. They are a useful tool for mapping disease loci and for candidate gene association studies (Wang et al. 1998; Chen and Sullivan 2003). SNPs have also been used in studies of genetic and population demographics with a conservation aim (Allendorf 2017). However, this type of marker is of limited utility for non-model species because it requires prior genome information, which is expensive and time-consuming to obtain. For this reason SNPs have generally been applied in studies of animals

(Vignal et al. 2002) and crop plant species (Ganal et al. 2009; Inghelandt et al. 2010) rather than Neotropical tree species (Olsson et al. 2017).

Genetic studies based on RADseq (Davey and Blaxter 2011; Davey et al. 2011) arose from the development of Next Generation Sequencing (NGS) tools. The RADseq technique uses restriction endonucleases to randomly shear sample DNA at restriction sites. DNA fragments from regions next to the restriction sites, and within a chosen length range, are sequenced, and variation in these sequences is the basis for distinguishing individuals (Davey et al. 2011). RADseq targets a random subset of the genome instead of sequencing the whole genome. This approach reduces the cost of having high depth of coverage per locus, potentially allowing the researcher to process a larger number of samples (Andrews et al. 2016). Because RADseq analysis does not require prior genomic information for the species, this approach has been widely used for SNP discovery and in ecological and evolutionary studies of non-model organisms (Andrews et al. 2016). To date, this approach has not been widely used for plant species in general (Barnard-Kubow and Galloway 2015; Fernández-Mazuecos et al. 2017; Maguilla et al. 2017), or tree species in particular (Hipp et al. 2018).

DNA sequencing technology was first developed in 1977 by Frederick Sanger and colleagues using a dideoxyribonucleotide (ddNTP) termination method. Sequencing of nuclear DNA, mitochondrial DNA (mtDNA) and chloroplast DNA (cpDNA) has been used extensively in population genetic studies, giving rise to the area of research now known as molecular phylogeography. Molecular phylogeography deals with the principles and processes that affect the geographical distributions of genealogical lineages, especially those within and among closely related species, providing information on genetic diversity not only in the spatial dimension, but also in the temporal dimension (Avise 2000). With the emergence of the coalescing theory, population genetics analyses have shifted focus, from the allelic frequency

of future generations to using sample alleles to model the genealogies of genes under almost any complex demographic history, in order to estimate phylogenetic parameters such as population sizes, divergence times, and migration rates (Wakeley 2008). In the case of neotropical trees, several DNA regions have been identified that show variation over a wide range of taxa, for cpDNA such as psbA-trnH, trnS-trnG, (Shaw et al.2005), matK (Dunning & Savolainen 2010) and the íntron trnL-trnF (Taberlet et al. 1991) and the nuclear DNA region ITS (Internal Transcribed Spacer, White et al. 1990). In early phylogeography studies, chloroplast markers were chosen because of their low recombination rates, putative neutrality, and smaller effective population size (Hickerson et al. 2010). However, phylogeography studies have also made use of markers such as SSRs and SNPs because they are more useful for detecting gene flow or genetic structure between intra-specific populations. In addition, in the last eight years the technique of NGS (Next Generation Sequencing) has been applied to tree species (see Mccormack et al. 2013), and recently the first assembly of a neotropical tree species genome was carried out (Pink Ipe, Silva-Junior et al. 2018).

Genetic diversity and structure

Genetic diversity

Among its many important roles in species function, genetic diversity underpins a species' ability to respond to disease threats, population bottlenecks, and environmental change. Moreover, analysis of genetic diversity can be used to understand the effects of population size, mating patterns, spatial distribution of individuals, mutation, migration and natural selection on individual populations and a species as a whole. Thus, knowledge of the amount of genetic diversity is considered critical for analysing the status of, threats to, and genetic viability of whole species and individual populations (Escudero et al. 2003).

Genetic diversity, at its most basic, can be equated with genetic polymorphism, which is the idea that at a given locus there is a greater than 99% probability of observing more than one allele. Genetic polymorphism can be quantified in terms of allele frequencies. The convention for expressing the frequencies of alleles A and a at a given locus, represented by p and q, respectively, where the numbers of AA, Aa and aa genotypes are expressed as n_{AA}, n_{Aa} and n_{aa} is: $p = (2n_{AA} + n_{Aa})/2n$ and $q = 1 - p$. Allele frequencies and allele frequency-based analyses, including F statistics, population assignment, relatedness, and probabilities of identity (the probability that two randomly chosen individuals will have the same genotype when using a given set of makers (see reviews Luikart and England 1999; Waits et al. 2001)) can be analysed in Arlequin (Excoffier et al. 2005), GenAlEx (Peakall and Smouse 2006), and FSTAT (Goudet 2001).

In addition to the measures of allelic diversity described above, genetic diversity can be measured at the scale of single base pairs, nucleotide diversity, or groups of alleles, haplotype diversity. Nucleotide diversity (π) is the average number of nucleotide differences per site in pairwise comparisons among DNA sequences (Nei 1987). A haplotype is defined as a group of alleles in an individual organism. This group of alleles is further defined as being contained on a single chromosome and/or tightly linked such that there is a high probability of the alleles being inherited together. Haplotype diversity (H) is usually measured as the probability that two randomly chosen haplotypes are different (Nei 1987). However, haplotype diversity can also be measured as the total number of haplotypes, or the number of unique or 'private' haplotypes in a population. Haplotype diversity is frequently used in phylogeographical and demographic analyses of Neotropical trees (e.g. Scotti-Saintagne et al. 2013, Collevatti et al. 2015). Nucleotide diversity and several measures of haplotype diversity are standard statistics in programs such as DnaSP (Librado and Rozas 2009).

For selectively neutral loci, the number and frequency distribution of alleles are the result of an equilibrium between mutations and genetic drift, which is a function of mutation rate and effective population size (Cornuet and Luikart 1996). When effective population size is reduced, it can cause a reduction in allele number and heterozygosity. Importantly, allele number is expected to decrease faster than heterozygosity following a bottleneck, creating a potential window of time during which the observed number of alleles is less than expected at mutation drift equilibrium. In this case, a significant heterozygosity excess would indicate a recent genetic bottleneck. According to Cornuet and Luikart (1996), the allele deficiency is a function of the time since the beginning of the bottleneck, effective population size, and the mutation rate. A distinction must be made between 'heterozygosity excess' compared to expectations under mutation-drift equilibrium and 'excess of heterozygotes' compared to expectations under Hardy-Weinberg Equilibrium. Cornuet and Luikart (1996) suggest that a test of the difference between observed heterozygosity and expected heterozygosity provides a robust means of testing for population bottlenecks. Tests of observed and expected heterozygosity can be implemented in GenAlEx (Peakall and Smouse 2006), Genepop (Raymond and Rousset 1995), and FSTAT (Goudet 2001).

Given knowledge of current allele frequencies in a population (p and q above), expected allele and genotype frequencies in successive generations is generally modelled via estimation of expected genotype frequencies under Hardy-Weinberg Equilibrium (HWE) (Hartl and Clark 2007). There are a number of assumptions inherent in the calculation of HWE genotype frequencies; specifically HWE assumes the study organism is diploid, reproduces sexually, generations are non-overlapping, and there is no migration, mutation, natural selection, or assortative mating (Hartl and Clark 2007). These assumptions may appear strict and over-simplistic, but the HWE has been shown to provide a useful first

approximation, valuable in suggesting key experiments or observations and as a guide for interpreting data.

Deviations from HWE are thought to indicate inbreeding or population stratification, but may also be used to check for errors in genotyping (Hosking et al. 2004). HWE is commonly tested using the χ^2 goodness-of-fit test, however Wigginton et al. (2005) suggests that the χ^2 test can have inflated type I error rates, even in large samples. Alternative exact tests, and computational methods for their implementation, are provided in their review (Wigginton et al. 2005). Deviations from HWE can be calculated in FSTAT (Goudet 2001) and Arlequin (Excoffier et al. 2005), followed by Bonferroni correction of the p-values (Rice 1989).

Phylogeography and population expansion and contraction can also be estimated using tests of neutrality. Specifically, Tajima's D (Tajima 1989), a test for non-random evolution, Fu and Li's F^* and D^* (Fu and Li 1993), tests for gene selection or demographic changes in population based on the coalescent, and Fu's F_s, based on haplotype distribution (e.g. Turchetto-Zolet et al. 2012), are suitable statistics. Tests of Tajima's D, Fu and Li's F^* and D^*, and Fu's F_s can be implemented in DnaSP (Librado and Rozas 2009). In addition to the analyses above, demographic history can be studied using a suite of nonparametric and semi-parametric approaches based on coalescent theory. Coalescent theory quantifies the relationship between the genealogy of the sequences and the demographic history of the population, and allows the use of more complex models of population history beyond traditional exponential or logistic growth or decline. This group of approaches includes skyline-plot methods (reviewed in Ho and Shapiro 2011), and other Bayesian demographic methods (reviewed in Bijak and Bryant 2016). Skyline analyses can be implemented using BEAST (Drummond and Rambaut 2007) and GENIE (Pybus and Rambaut 2002), and other

Bayesian approaches can be implemented in a range of packages, including DIYABC (Cornuet et al. 2008).

Genetic structure

Biological organisms generally exhibit some degree of spatial/geographical structure. Population subdivision in particular generally occurs in response to either environmental heterogeneity or social behavior, in the case of mobile organisms (Manel et al. 2003). Analysis of population genetic structure is generally used to estimate differences or relationships between populations, and then infer evolutionary process such as gene flow, migration, mutation or drift. A number of statistics have been developed to describe spatial genetic structure, the most common being the F-statistics (Wright 1976) and R-statistics (Slatkin 1995). The F-statistic is described as the genetic correlation or inbreeding coefficient or the coefficient of differentiation. Because allelic differentiation can occur at different hierarchical levels, three coefficients, or F-statistics have been defined: (1) F_{IT}, mean reduction in heterozygosity of an individual relative to the total population, also described as deviation from Hardy-Weinberg expectations in the total population; (2) F_{IS}, the mean reduction in heterozygosity of an individual due to non-random mating in a subpopulation, also described as deviation from Hardy-Weinberg expectations in the subpopulation (often due to inbreeding); (3) F_{ST}, the mean reduction in heterozygosity of a subpopulation due to genetic drift, also described as genetic differentiation between subpopulations of the total sampled population. Large populations with high migration between them are expected to show little genetic differentiation and a small F_{ST} value, whereas populations with little migration between them are expected to be more genetically differentiated and have a high F_{ST} value (Holsinger and Weir 2009). Similarly, if there are a large number of allelic differences between populations

this will produce a high R_{ST} index (Slatkin 1995). Holsinger and Weir (2009) suggest that when using SSR markers, R_{ST} provides greater insight into the relationships among populations because it is based on differences in the actual number of repeats between alleles at each SSR locus. If SSR mutation is assumed to conform to the Stepwise Mutation Model (SMM, Valdes et al. 1993; but see Jarne and Lagoda 1996), differences in the number of SSR motif repeats indicate something about the genetic distance between samples, rather than simply indicating that one population is different from the other. In contrast, SNP markers are expected to conform to the Infinite Sites Model of mutation (Vignal et al. 2002; Morin et al. 2004), in which there is no measure of genetic distance between samples and F_{ST} is a more suitable statistic. F and R statistic analyses may be conducted in a number of software packages, including FSTAT (Goudet 2001), GDA (Lewis and Zaykin 2001), and Genepop (Raymond and Rousset 1995).

Population genetic structure may also be quantified using a hierarchical Analysis of Molecular Variance (AMOVA; Excoffier et al. 1992). AMOVA is based on conventional analysis of variance (Excoffier and Heckel 2006), and provides a statistical framework for testing patterns of genetic variance within and among populations and subpopulations. Because genetic variation is partitioned into hierarchical levels, AMOVA, like F-statistics, provides a means of detecting inbreeding, genetic drift and gene flow. Whereas Wright's F-statistics are based on allelic variation, AMOVA can be used with a variety of molecular data (e.g. genetic sequences, SSRs, or SNPs). An AMOVA analysis can be conducted in Arlequin (Excoffier et al. 2005).

Analysis of the genetic structure of naturally occurring populations frequently involves the parallel analysis of spatial data. Possibly the most common and direct method of evaluating the relationship between genetic data and spatial data is the Mantel Test. The normalised

Mantel statistic is interpreted as a Pearson correlation coefficient between a matrix of genetic distance and a matrix of spatial distance. The statistic is tested by iterative randomization in one of the two matrices. For this analysis spatial distance can be measured as Euclidean distance, non-Euclidean metric distance, or ranked distance (Escudero et al. 2003). Mantel tests can be implemented using the ade4, *vegan* or *ecodist* libraries in R (R Core Team 2017). There has been discussion in the literature about the statistical performance of the Mantel test, however Diniz-Filho (2013) found that alternative approaches, including Spatial Eigenfunction Analysis, resulted in similar estimates of the magnitude of spatial structure in their study of genetic structure in *Dipteryx alata* Vogel in the Brazilian Cerrado.

There are a number of traditional statistical alternatives to the Mantel Test for analyzing spatial genetic structure. Spatial Autocorrelation Analysis (SAA) is useful for analysis of multiallelic, codominant, multilocus arrays, and can incorporate differential weighting of low-frequency alleles (Smouse and Peakall 1999). An Allelic Aggregation Index Analysis (AAIA) tests for non-random spatial patterns of genetic diversity. Both SAA and AAIA can be conducted in the Alleles In Space software (Miller 2005).

In addition to traditional statistical approaches, there are a number of methods for investigating spatial genetic structure based on Bayesian approaches. STRUCTURE (Pritchard et al. 2000) is possibly the most widely used Bayesian approach to population genetic structure analysis. The popularity of this approach is indicated by the fact that the original STRUCTURE paper has been cited more than 20,000 times (Google Scholar accessed in October 2017). Pritchard et al. (2000) implement a model-based clustering method for multilocus genotype data to infer population structure in which each individual in each population is characterized by a set of alleles at each locus. An important advance of this approach compared to the F and R statistics or the ANOVA approach is that a single individual may be probabilistically

assigned to two or more populations in a. ‘admixture inference’ model (Pritchard et al. 2000). This model does not assume a particular mutation process, and it can be applied to most of the commonly used genetic markers including SSR, SNPs and RFLPs (Pritchard et al. 2000).

Null alleles present a challenge for genotyping, and therefore estimation of genetic structure. SSR loci in particular suffer from the presence of null alleles and variable mutation patterns (see review Dakin and Avise 2004), which can introduce ambiguity into analyses (Morin et al. 2004). The STRUCTURE software can be run on a dataset which includes null alleles (Falush et al. 2007) although the software does not produce an estimate of their frequency. However, there is software available, such as MICROCHECKER (Van Oosterhout et al. 2004), CERVUS (Kalinowski et al. 2007) and INEST 2.0 (Chybicki and Burczyk, 2009), which can estimate and correct null allele frequencies.

The Bayesian Analysis of Population Structure (BAPS) software (Corander et al. 2008) can be used to infer spatial genetic structure while treating both the allele or nucleotide frequencies and the number of genetically diverged groups as random variables or by using a fixed number of groups. Geneland (Guillot et al. 2012) uses a Bayesian approach plus geographic information to estimate null allele frequencies, and also estimates the population genetic structure of each individual. Finally, Genetic Landscape Shape interpolations are a visualization technique implemented in The Alleles In Space software (Miller 2005) which allows researchers to visually explore spatial genetic structure.

An interesting point of discussion is that frequently individuals are assigned to ‘populations’ by the researcher when in fact highly mobile organisms, organisms that exist as large groups occupying large continuous habitats, and species in marine or aquatic systems may not be objectively assigned to a particular population. The Alleles In Space software (Miller 2005) allows researchers to avoid arbitrary groupings of individuals, and instead

performs joint analyses of inter-individual spatial and genetic information. Similarly, SAMOVA (Spatial Analysis of Molecular Variance; http://cmpg.unibe.ch/software/samova2/) partitions samples into groups by maximizing the genetic differentiation among the groups while also taking into account the geographical coordinates of the samples.

Conclusions drawn from analyses of spatial genetic structure can be used to provide valuable guidelines for conservation strategies and management of the genetic diversity of tree species (Hamrick et al. 1992, see Table 1 in Ratnam et al. 2014), and a number of studies with an emphasis on conservation have employed molecular marker-based approaches. One such study was the evaluation of the genetic structure of ten populations of the endangered tropical tree species *Caryocar brasiliense* Cambess. (Caryocaraceae), located in the Brazilian Savanna, using ten SSR loci (Collevatti et al. 2001). This study showed that although habitat fragmentation can potentially reduce genetic variability, its effects on *C. brasiliense* populations were difficult to detect because disturbed populations were probably already present prior to fragmentation. Sebbenn et al. (2011), in a study evaluating eight SSR markers in populations of the tropical tree *Copaifera langsdorffii* Desf. (Caesalpinioideae) found that habitat fragmentation leads to the spatial isolation of populations and can increase genetic structure, even when adult plants exhibit high genetic diversity. In contrast, eight SSR loci revealed low genetic diversity in the threatened tropical tree species *Dipteryx alata* Vogel (Fabaceae) from the Brazilian Neotropical savannas, even though structured and divergent populations were also detected (Collevatti et al. 2013). Finally, seven SSR markers indicated that populations of a critically endangered tree of the Brazilian Savanna, *Dimorphandra wilsonii* Rizzini (Caesalpinioideae), display low levels of genetic diversity. The authors suggested that monitoring dispersion in order to maintain the genetic diversity and structure are strategies that may ensure the continued survival of this species (Vinson et al. 2015).

In addition to concerns regarding conservation, our understanding of the evolutionary forces shaping the origins and maintenance of Neotropical tree diversity is still limited (Hipp et al. 2018). Thus, molecular markers developed for Neotropical tree species have found significant use in studies of phylogeography. One example is in the use of six SSRs to better understand the recent colonization of the Galápagos islands from the Peruvian mainland by the tree *Geoffroea spinosa* Jacq. (Leguminosae). This study revealed two possibilities: recent natural dispersal or human introduction (Caetano et al. 2012). Another dispersal pattern was revealed using chloroplast SSR variation in the widespread species *Cordia alliodora* (Ruiz & Pav.) Cham. (Boraginaceae), to investigate phylogeographic structure in Central and South America, and the West Indies (Rymer et al. 2013). SSR markers were also effective in indicating the phylogeography of a species complex from lowland Neotropical rain forests (*Carapa guianensis* Aubl., Meliaceae), which pointed to the Amazon as a center of origin and diversification, with subsequent migration to the Pacific coast of South America and Central America (Scotti-Saintagne et al. 2013). Here, two genetic clusters were detected and show apparent gene flow among them. Similarly, Mangaravite et al. (2016) identified two genetic groups by using ten SSR loci to genotype populations of *Cedrela fissilis* Vell. (Meliaceae). The genetic diversity detected using this approach indicated structure in accordance with geography: the Atlantic range and the Chiquitano range, in South America, and an admixture of recent origin, resulting from population expansion. For more details on phylogeography studies of neotropical trees see Orlando et al in this special issue.

The study of evolution, diversification, phylogenetics and phyogeography in non-model organisms is often restricted by a lack of genetic data, which limits access to the newest and most powerful analytical techniques. Neotropical trees are not used as model organisms for a number of reasons, including size and generation time, and so, to date, studies of Neotropical trees have largely relied on SSR markers because they are highly variable,

codominant and relatively inexpensive to develop for non-model species. However, low-coverage draft genome sequencing, ddRAD, and RADseq data is increasingly within reach, even for small projects, and this flood of new data will facilitate exciting new avenues of study for non-model organisms (Ekblom and Wolf 2014). Apart from facilitating the development of molecular markers, including SSRs, and facilitating the identification of candidate genes (Weitemier et al. 2014), genome sequence data itself can be used in studies of evolutionary history and phylogenetics. Genome sequence data can be used in programs such as SIMCOAL and SIMCOAL2 (Excoffier et al. 2000), which use a coalescent approach to investigate population demography, and DIYABC (Cornuet et al. 2008) and PopABC (Lopes et al. 2009) which use inference-based approximate Bayesian computation to model population history (see review Lopes and Beaumont 2010). A few studies have used these approaches for Neotropical trees; for example, Olsson et al. (2017), where RADseq data are used to investigate the evolutionary and phylogenetic history of a widespread tropical tree, *Symphonia globulifera* L.f. (Clusiaceae) and Hipp et al (2018) where RADseq data is used to investigate the diversification of Mexican oaks. These powerful new tools are making it increasingly possible to use RADseq to identify the complex and interacting forces that shape the evolution of tree biodiversity.

Mating system, pollen and seed dispersal

The basic processes of reproductive biology are the sexual system, incompatibility mechanisms, flowering patterns and pollination processes (Boshier 2000). The integration of these reproductive processes with the spatial structure of a species influences the levels and dynamics of genetic diversity (Boshier 2000).

Mating system

A key step in understanding a species' mating system is defining its sexual system. The range of sexual systems found in trees can be broadly classified into three major groups: (1) dioecious individuals that are male or female; (2) monoecious individuals that are hermaphrodites, but with individual flowers that are female or male; and (3) monoecious individuals that present individual hermaphrodite flowers with both male and female organs (Boshier 2000). The sexual system is one of the main determinants of a species' current mating patterns. In addition, as current mating patterns determine how genes are transferred from one generation to the next, and how genes are recombined in the next generation, the sexual system also determines the relatedness and, to some extent, the levels of inbreeding in following generations.

In addition to being influenced by the sexual system, levels of inbreeding are also determined by levels of self-pollination and degree of self-incompatibility. Flowers can be self-pollinated and/or cross-pollinated (Sedgley and Griffin 1989). Self-pollination (selfing) involves the transfer of pollen from the anther to the stigma of either the same flower or another flower of the same tree, whereas, cross-pollination involves the transfer of pollen from one plant to the stigma or micropyle of another plant by a pollination agent (Sedgley and Griffin 1989). Self-incompatibility, which can be pre or post-zygotic is the inability of fertile plants to reproduce upon selfing, and is a genetically-controlled mechanism which reduces the prevalence of inbreeding in a population (Boshier 2000). In self-compatible trees, inbreeding (the inbreeding coefficient, F, defined as the probability that two alleles at a locus are identical by descent) can be caused by self-fertilization ($s = 1 - t_m$) and mating among related individuals ($t_m - t_s$). Selfing produces an inbreeding value of at least 50% ($F_s = 0.5(1 + F_m)$, where F_m is the inbreeding value of the mother) and mating among related individuals ($t_m - t_s$) produces inbreeding values (F_r) at the same level as coancestry (θ_p) between the parents

($F_r = \theta_p$). For example, mating among two half-sibs (θ_p= 0.125) will produce 12.5% inbreeding and mating among two full-sibs (θ_p= 0.25) will produce 25% inbreeding. Mating among related individuals occurs if populations present intra-population genetic structure (SGS) and pairwise near-neighbours are more likely to be related than pairwise trees separated by longer distances (See next section, Cavers and Dick 2013).

By definition, in dioecious species self-fertilization is prevented (Boshier 2000) and seeds are only produced by outcrossing. Thus in an open-pollinated progeny array from a dioecious species there may be two kinds of relatedness: half-siblings and full-siblings. Half-siblings are the result of two ovules from the same mother being fertilized by pollen from different pollen donors. Full-siblings are the result of two ovules from the same mother being fertilized by pollen from the same pollen donor. Inbreeding only occurs in a dioecious species through mating between related individuals. The breeding success of dioecious species depends upon the frequency of sexual forms in the population, the spatial distribution of individuals, and the abundance and behaviour of pollinators (Barrett and Thomson 1982). For example, Silva et al. (2008) found biparental inbreeding (6.7%) in dioecious *Bagassa guianensis* Aubl. (Moraceae), which the authors suggested was the result of the behaviour of pollinators that visit few trees together paired with short range seed dispersal and significant genetic structure. Studies using molecular markers for dioecious species show high estimates of outcrossing rate (t_m), ranging from 0.97-1.00 for populations of *Araucaria angustifolia* Bertol. (Bittencourt and Sebbenn 2007; Medina-Macedo et al. 2015), *Bagassa guianensis* (Silva et al. 2008; Arruda et al. 2015), and *Myracrodruon urundeuva* Allemão (Gaino et al. 2010).

Monoecious species are generally classified as outcrossing, mixed mating system or selfing. According to Goodwillie et al. (2005), an outcrossed species is one that has an outcrossing rate (t_m) higher than 0.8, a species with a mixed mating system has an t_m between

0.2 and 0.8, and a species is selfing when t_m is lower than 0.2. In self-compatible monoecious species, mating may involve self-fertilization and outcrossing, thus within an open-pollinated progeny array there may be four different kinds of relatedness: self-siblings, half-siblings, full-siblings, and self-half-siblings. Self-siblings are the result of self-fertilization by the mother tree. Self-half-siblings are the description of the relatedness between two seeds from the same mother tree, one of which is the result of self-fertilization and one of which is the result of outcrossing. Inbreeding occurs in monoecious species through self-fertilization or mating between related individuals. Factors such as asynchronous flowering of male and female flowers and an unequal sex ratio in the population may contribute to inbreeding. 60-70% of Neotropical tree species are hermaphroditic, and variable values for t_m have been reported in the literature (i.e. t_m= 0.99 for *Dipteryx odorata* Aublet but as low as 0.45 to 0.90 for *Dipteryx alata* Vogel, Tarazi et al. 2010; Vinson et al. 2014b; Tambarussi et al. 2017). Furthermore, the amount of selfing occurring in a hermaphroditic species depends on factors such as the density of neighbouring flowering trees, density and efficiency of pollinators, flowering time and frequency, and the impact of the area surrounding the study plot.

Patterns of breeding are influenced not only by the sexual and mating systems of the species, but also by the species' flowering pattern. Flowering pattern analysis involves the study of the variation in timing, duration, synchrony and frequency of flowering at the population level over a number of flowering seasons (Bawa 1983). The factors that promote flowering are: (1) competition for pollinators and interspecific gene flow, where habitat segregation appears to be at least as important as temporal segregation; (2) pollinator availability and nature of floral rewards; and (3) selection for the optimization of life history traits other than flowering. Most Neotropical trees flower annually, however, some do not, such as the endemic *Manilkara huberi* Ducke which only flowers every 3 to 5 years (Azevedo et al. 2007). A species' flowering can range from an extended flowering period where individuals

produce a few flowers per day for a long period, to mass flowering where individuals produce a large number of flowers per day for a few days. Most neotropical trees have an extended flowering period, which has the advantages of adjusting the number of flowers to the resources available, as well as increasing the chance of reproduction with a large number of mates and thereby increasing diversity of mating and decreasing the risk of reproductive failure. Whilst mass flowering reduces the chances of herbivory there is a risk of no pollinators being present during the flowering event. For example *Tabebuia roseoalba* (Ridl.) Sandwith (Bignoniaceae), known as "Ipé branco", flowers on average three days a year and has a late-acting self-incompatibility system (LSI) (Gandolphi and Bittencourt Jr 2010), therefore if pollinators are not efficient during this short flowering period there will be no viable fruits for that year. In addition, asynchronous flowering among individuals can reduce the effective population size and limit the diversity of mating, though under other circumstances asynchronous flowering may actually increase the diversity of mating and avoid intraspecific competition for pollinators (Bawa 1983). In *Dipteryx odorata* strong asynchrony in flowering is likely to reduce effective population size, with a maximum of 34% of the trees flowering per event in this population (Maués 2006). This asynchrony, in extreme, may result in trees in small fragments having no available breeding partners (Vinson et al. 2014a, b). Finally, the variation in timing, duration, synchrony and frequency of flowering of a population, together with the interaction of pollinators, can affect the outcrossing rate and, consequently, population-level genetic diversity.

To undertake molecular analysis of mating systems, open-pollinated seeds, mother trees and, optionally, other potential pollen-donor trees must be sampled and genotyped using a set of molecular markers. The mating system of populations can be investigated using molecular marker data and models such as the mixed mating model (Ritland and Jain 1981), correlated mating model (Ritland 1989), neighbourhood mating model (see next section, Burczyk et al.

2006), and paternity analysis (see next section, Meagher 1986). The mixed mating model assumes that seeds within families may originate from a mixture of selfing and outcrossing, whereas the correlated mating model allows for the possibility that outcrossed seeds may originate from successive matings between the mother and the same father and the seeds may therefore be full-siblings. These two models are implemented in the MLTR software (Ritland 2002), which provides an estimate of the pollen and ovule gene frequencies, fixation index of mother trees (F_m), multilocus outcrossing rate (t_m), single-locus outcrossing rate (t_s), mating among related individuals ($t_m - t_s$), selfing correlation (r_s), single-locus correlation of paternity ($r_{p(s)}$), multilocus correlation of paternity ($r_{p(m)}$) and rate of crossing between pollen donor trees that are related ($r_{p(m)} - r_{p(s)}$). From these indices, many other indices may be estimated. For example, the selfing rate ($s = 1 - t_m$), effective number of pollen donors ($N_{\text{ep}} = 1/r_{p(m)}$, Ritland 1989), the co-ancestry coefficient within families ($\Theta = 0.125(1 + F_m)\left[4s + (t_m^2 + \text{st}_m r_s)\left(1 + r_{p(m)}\right)\right]$, Sebbenn 2006), the variance effective size ($N_e = 0.5/\Theta$, Sebbenn 2006), and the number of seed trees for seed collection ($m = N_{e(r)}/N_e$, where $N_{e(r)}$ is the effective population size the collector aims to retain in the progeny array sampled, Sebbenn 2006). The estimate of m is based on the following assumptions: (1) seed trees are not related; (2) seed trees do not receive an overlapping pollen pool (each seed tree mates with a different set of pollen donors); and (3) seed trees do not mate with each other. If the assumptions are met, in the whole progeny sample related seeds should only occur within family groups.

Pollen and seed dispersal

Understanding pollen and seed dispersal is key to understanding the impacts of anthropogenic land-use change on population and species viability (see Table 1 in Ratnam et al. 2014).

Genetic markers can be used to study the pollen-flow and seed dispersal patterns using (1) direct paternity analysis, (2) neighbourhood mating model or (3) the TWO-GENER method. Direct paternity analysis methods implemented in CERVUS (Marshall et al. 1998; Kalinovski et al. 2007) and Famoz (Gerber et al. 2003) use the exclusion and likelihood method approach. For paternity analysis open-pollinated seeds, mother trees and all of the other adult trees with a study area must be sampled and genotyped. Direct paternity analysis aims to identify the pollen donor for each sampled seed. This data, paired with the location of the pollen donor, allows the researcher to investigate the minimum, average and maximum distance of pollen dispersal, the number of pollen donors for each seed tree, and percentage of different pollen donors for each seed tree. If there is a pollen donor that is contributing disproportionately to the pollen pool, these data allow the researcher to investigate which factors may be important for pollination success, for example diameter of the pollen donor, wind, and distance from the pollen recipients. Pollen dispersal can be investigated using the Neighbourhood mating model (Burczyk et al. 2006), which is a paternity analysis approach that assumes that seeds of a seed-tree (mother-tree) may results from selfing (s), outcrossing with pollen donors located within a pre-defined circular neighbour area (t_w), and outcrossing with pollen donors located outside of the pre-defined circular neighbour area (t_o). This model is implemented in the NM+ software (Chybicki and Burczyk 2010) and provides an estimate of pollen flow (m_p) from outside of the sample area, the effective number of pollen donors (N_{ep}), and the mean distance of pollen dispersal (δ). For this analysis, open-pollinated seeds, mother trees, and other adult trees must be sampled and genotyped within a defined area (Neighbor area), and other trees may be sampled outside of the area. Pollen dispersal can also be indirectly investigated, using the TWO-GENER method (Smouse et al. 2001), implemented in the POLDIST software (Robledo-Arnuncio et al. 2007). The principle of the TWO-GENER method is to estimate the differentiation of allele frequencies among the pollen pools sampled by several mother trees in

a population. The parameter is calculated by an analysis of molecular variance (AMOVA, Excoffier et al. 1992). For this analysis, open-pollinated seeds and mother trees and, optionally, other potential pollen-donor trees, are sampled and genotyped.

Parentage analysis based on nuclear microsatellite loci may permit us to identify both parents of a progeny, but does not distinguish which is the female and which is the male parent. Pollen dispersal distance is the distance between the both assigned parents, and the identification of which parent is the mother and which is the father is not relevant. However, in order to determine seed dispersal distance, we must know which parent is the mother, since seed dispersal distance is calculated as the linear distance between the seedling and the mother tree. Many studies have assumed that the closest assigned parent for progeny is the mother and the most distant assigned parent is the father based on the maximum locomotion distance capacity of pollen and seed dispersal vectors (Sebbenn et al. 2011; Baldauf et al. 2014). For dioecious species, putative parents can be sexed visually by observation of female and male flowers, as well as by the presence and absence of fruits. Cytoplasmatic gene markers can be especially helpful in this case if their pattern of inheritance is known, as is the case for coniferous trees, where mtDNA is inherited maternally whilst cpDNA is inherited paternally. For angiosperm trees the situation is more complicated as mtDNA and cpDNA may not be uniquely heritable from the mother and father, respectively.

Biotic pollinator-mediated pollen dispersal depends largely on pollinator foraging behaviour. Trees use odour and visual attraction and offer compensation for the visit of the pollinator, including pollen, nectar and oils, protection and brood locations (Sedgley and Griffin 1989). Different animal pollinators, such as bees, beetles, lepidopterans and syrphid flies, birds and bats, have different minimum and maximum flying distances and different visiting behaviours. This interaction between plants and pollinators has been modelled based

on food-gathering and predator-prey models (Waddington 1983). The ideal pollinator is rewarded for making regular visits at the time of pollen shed, visits many flowers of many trees, carries significant loads of viable pollen and makes frequent contact with stigmas to produce effective pollination. An ideal tree has high effective population size, a high proportion of trees flowering in synchrony, a large quantity of pollen produced, and is attractive to pollinators (Maués 2006).

Here we present two case studies, both of which employed SSR markers for direct paternity assignment, which produced contrasting results. *Jacaranda copaia* (Aubl.) D. Don (Bignoniaceae) pollen disperses over long distances, and individual trees generally mate with a large number of other trees. The species does not have a specific pollinator, and can be pollinated by small and medium sized bees and occasionally butterflies and humming birds. In contrast, within the same forest, the pollinators (Thysanoptera) of *Bagassa guianensis* form 'thrip clouds' and visit the nearest trees, therefore studies of pollen dispersal found only three pollen donors per mother tree among narrowly distributed trees (Silva et al. 2008). In general, in canopy and emergent species, outbreeding and self-incompatibility is common, some pollinators apparently travel long distances, gene flow may be extensive and seeds are largely or completely outcrossed. Therefore, tropical tree populations have extensive gene flow and seem to exist as large and genetically connected populations (Loveless 2002).

Intra-population spatial genetic structure (SGS) depends largely on seed dispersal (Vekemans and Hardy 2004; Azevedo et al. 2007; Bittencourt and Sebbenn 2007; Carneiro et al. 2007; Cloutier et al. 2007; Eduardo et al. 2008; Silva et al. 2008; Degen and Sebbenn 2014; Vinson et al. 2014b). Species whose seeds are dispersed near the maternal plant (e.g. gravity dispersal) or species whose seeds are deposited in clumps or patches should have more fine-scale genetic structure than species whose seeds are dispersed singly by mobile animals.

Furthermore, due to the overlap of seed shadows, species with high adult densities should have less genetic structure than species with lower densities (Vekemans and Hardy 2004). Mating among related individuals occurs if populations present intra-population genetic structure (SGS) and pairwise near-neighbours are more likely to be related than pairwise trees separated by longer distances (Cavers et al. 2005). SGS can be determined using an estimation of the coancestry coefficient (θ_{xy}) between pairwise trees within different distance classes, an analysis which can be implemented in the SPAGEDI software (Hardy and Vekemans 2002). It can also be compared between different generations, populations and species by the statistic Sp-statistic (Vekemans and Hardy 2004): $Sp = -b_k/(1-\theta_1)$ where θ_1 is the average coancestry coefficient calculated between all pairwise individuals within the first distance class and b_k is the slope of the regression of the coancestry coefficient against the logarithm of spatial distance. Populations of tropical trees generally exhibit SGS (Degen and Sebbenn 2014), but small populations that have been isolated for a long time are expected to exhibit the highest SGS (Wright 1976). There are four main SGS patterns generally observed: (1) Short seed dispersal distance leads to a population with genetic structure, trees surrounded by related trees (Table 1, species A,B, E, F, I and J). If pollen dispersal is short, pollinators only visit neighbouring trees, related trees will cross, and there is a high probability of selfing (Table 1, species E and I) and inbreeding (Table 1, species A, E and I). (2) Short seed dispersal but long median pollen dispersal, that is, pollen dispersal is longer than seed dispersal: In this case the probability of selfing and inbreeding is moderate, as some pollinators will visit close trees (related) and some pollinators will travel further and visit unrelated trees (Table 1, species B, F, J). (3) Long-distance seed dispersal results in little or no genetic structure in the population and in association to short pollen dispersal there is a low probability of mating among related trees and, consequently of seeds present inbreeding (Table 1, species C, G, K). (4) Long-distance seed dispersal and long-distance pollen dispersal results in very low probability of selfing and

inbreeding (Table 1, species D, H, L). (5) If species have a self-incompatibility system, such as species A, B, E, F, I and J in Table 1, this will probably lead to abortion of seeds resulting from selfing and inbreeding and will produce a reduced number of seeds capable of germinating.

Conclusions and sampling strategies for genetic studies

Sampling seeds is an important step in a genetic study. Depending on the question, and in order to avoid bias, trees from the centre and corners of the study area should be selected. If clumps are present in the study area, one tree per clump may be selected. Ideally one seed per fruit should be collected, together with fruits from different positions on the tree (different heights and different cardinal directions) since there may be pollinators that pollinate only flowers at the top of the tree, whilst other pollinators visit the flowers below, or on the north or south side of the tree. These measures aim to guarantee representative sampling of each tree. DNA can be extracted from the embryo, or seeds can be germinated and when seedlings have at least two leaflets, DNA can be extracted. However, the second approach can cause early selection against inbred individuals leading to an underestimation of inbreeding.

In addition, for studies of genetic diversity and structure we suggest the sampling of at least 50 trees per site due to the fact that these analyses are dependent on robust estimates of gene frequencies (Sebbenn 2002). When sampling 50 trees per site we are likely to obtain a robust representation of gene frequencies within populations (Sebbenn 2002). However, in cases where population size is lower than 50 trees, we suggest sampling all trees available at the study sites. For mating system analysis, we suggest the collection of at least 200 open-pollinated seeds from at least 20 trees (10 seeds per tree) in order to investigate the individual

variation in outcrossing rate, mating among related trees and correlated mating. If the species under study produces more than one seed per fruit, we suggest collecting more than one seed per fruit and increasing the sampling of seeds to at least 20 seeds per tree (total of 400 seeds) in order to investigate the hierarchical paternity correlation within and among fruits (Sebbenn 2006). For parentage analysis, when aiming to investigate pollen and seed dispersal, we suggest sampling all reproductive trees in a forest fragment or in the case of continuous forest or species with a high population density (>10 trees per hectare), establish experimental plots where all reproductive trees must be sampled and open pollinated seeds must be collected mainly from trees located in the centre of the plot to try to maximize pollen donor assignment. Moreover, if possible, regenerants (saplings, seedlings or juveniles) must also be sampled, mainly in the centre of the plots, to try to maximise the assignment of mothers and fathers and investigate the pollen flow and seed dispersal distances.

Finally, a key goal for future studies of population genetics of tropical trees is to investigate the effects of inbreeding originating from both selfing and mating among relatives. At present there are a small number of studies of this subject in tropical trees (Ismail et al. 2012; Rymer et al. 2015; Duminil et al. 2016; Tambarussi et al. 2017), even though tree diversity is very high in tropical biomes. To do this, in studies based on paternity analysis, we suggest the parallel investigation of seed germination rate, seedling growth rate, plant height and diameter. Paternity analysis based on samples of open-pollinated seeds permits the determination of which individuals (germinated seeds) originated from selfing and mating among related trees, and the germination rate and growth characteristics of these seeds can then be compared with seeds originating from mating among unrelated trees (Ismail et al. 2012; Rymer et al. 2015; Tambarussi et al. 2017). This will provide important information regarding the effects of inbreeding originating from selfing or mating among related trees, as well as different seed and seedling traits.

References

Allendorf FW (2017) Genetics and the conservation of natural populations: allozymes to genomes. Mol Ecol 26:420–430. doi: 10.1111/mec.13948

Andrews KR, Good JM, Miller MR, et al (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet 17:81–92. doi: 10.1038/nrg.2015.28

Arruda CCB, Silva MB, Sebbenn AM, et al (2015) Mating system and genetic diversity of progenies before and after logging: a case study of *Bagassa guianensis* (Moraceae), a low-density dioecious tree of the Amazonian forest. Tree Genet Genomes 11:1-9

Avise JC (2000) Phylogeography: the history and formation of species. Harvard University Press.

Azevedo VCR, Kanashiro M, Ciampi AY, Grattapaglia D (2007) Genetic structure and mating system of *Manilkara huberi* (Ducke) A. Chev., a heavily logged Amazonian timber species. J Hered 98:646–654

Baldauf C, Ciampi-Guillardi M, Aguirra TJ, Correa CE, Santos FAM, Souza AP, Sebbenn AM (2014) Genetic diversity, spatial genetic structure and realised seed and pollen dispersal of *Himatanthus drasticus* (Apocynaceae) in the Brazilian savanna. Conserv Genet 15:1073–1083

Balvanera P, Pfisterer AB, Buchmann N, et al (2006) Quantifying the evidence for biodiversity effects on ecosystem functioning and services. Ecol Lett 9:1146–1156.

Barbará T, Palma-Silva C, Paggi GM, et al (2007) Cross-species transfer of nuclear microsatellite markers: potential and limitations. Mol Ecol 16:3759–67. doi: 10.1111/j.1365-294X.2007.03439.x

Barnard-Kubow KB, Debban CL, Galloway LF (2015) Multiple glacial refugia lead to genetic structuring and the potential for reproductive isolation in a herbaceous plant. Am J Bot 102:1842–1853. doi: 10.3732/ajb.1500267

Barrett SCH, Thomson JD (1982) Spatial pattern, floral sex ratios, and fecundity in dioecious *Aralia nudicaulis* (Araliaceae). Can J Bot 60:1662–1670

Bawa KS (1983) Patterns of flowering in tropical plants. In: C.E. Jones and R.J. Little (ed.) Handbook of Experimental Pollination Biology, pp 394–410

Besse P, Lebrun P, Seguin M, Lanaud C (1993) DNA Fingerprints in *Hevea-Brasiliensis* (Rubber Tree) Using Human Minisatellite Probes. Heredity (Edinb) 70:237–244. doi: 10.1038/hdy.1993.35

Bijak, J., and J. Bryant. 2016. Bayesian demography 250 years after Bayes. Population Studies **70**:1–19.

Bittencourt JVM, Sebbenn AM (2007) Patterns of pollen and seed dispersal in a small, fragmented population of the wind-pollinated tree *Araucaria angustifolia* in southern Brazil. Heredity (Edinb) 99:580–591

Boshier DH (2000) Mating System. In: Young A, Boshier DH, Boyle T (eds) Conservation Genetics: Principles and Practice. CSIRO, Collingwood, Australia, pp 350

Brousseau L, Foll M, Scotti-Saintagne C, Scotti I (2015) Neutral and adaptive drivers of microgeographic genetic divergence within continuous populations: The case of the

neotropical tree *Eperua falcata* (Aubl.). PLoS One 10:1–23. doi: 10.1371/journal.pone.0121394

Burczyk J, Adams WT, Birkes DS, Chybicki IJ (2006) Using genetic markers to directly estimate gene flow and reproductive success parameters in plants on the basis of naturally regenerated seedlings. Genetics 173:363–372

Caetano S, Currat M, Pennington RT, et al (2012) Recent colonization of the Galápagos by the tree *Geoffroea spinosa* Jacq. (Leguminosae). Mol Ecol 21:2743–60. doi: 10.1111/j.1365-294X.2012.05562.x

Carneiro F da S, Magno Sebbenn A, Kanashiro M, Degen B (2007) Low interannual variation of mating system and gene flow of *Symphonia globulifera* in the Brazilian Amazon. Biotropica 39:628–636

Cavers S, Dick CW (2013) Phylogeography of Neotropical trees. J Biogeogr 40:615–617. doi: 10.1111/jbi.12097

Chapin Iii FS, Zavaleta ES, Eviner VT, et al (2000) Consequences of changing biodiversity. Nature 405:234–242

Chen X, Sullivan PF (2003) Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. Pharmacogenetics J 3:77–96

Chybicki IJ, Burczyk J (2009) Simultaneous estimation of null alleles and inbreeding coefficients. J Hered 100:106–113.

Chybicki IJ, Burczyk J (2010) NM+: software implementing parentage-based models for estimating gene dispersal and mating patterns in plants. Mol Ecol Resour 10:1071–1075

Cloutier D, Kanashiro M, Ciampi AY, Schoen DJ (2007) Impact of selective logging on inbreeding and gene dispersal in an Amazonian tree population of Carapa guianensis Aubl. Mol Ecol 16:797–809

Collevatti RG, Grattapaglia D, Hay JD (2001) Population genetic structure of the endangered tropical tree species *Caryocar brasiliense*, based on variability at microsatellite loci. Mol Ecol 10:349–356

Collevatti RG, Telles MPC, Nabout JC, et al (2013) Demographic history and the low genetic diversity in *Dipteryx alata* (Fabaceae) from Brazilian Neotropical savannas. Heredity (Edinb) 111:97–105. doi: 10.1038/hdy.2013.23

Collevatti, R. G., L. C. Terribile, S. G. Rabelo, and M. S. Lima-Ribeiro. 2015. Relaxed random walk model coupled with ecological niche modeling unravel the dispersal dynamics of a Neotropical savanna tree species in the deeper Quaternary. Frontiers in Plant Science **6**:653.

Corander, J., J. Sirén, and E. Arjas. 2008. Bayesian Spatial Modelling of Genetic Population Structure. Computational Statistics **23**:111-129.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001–2014.

Cornuet JM, Santos F, Beaumont MA, Robert CP, Marin JM, Balding DJ, Guillemaud T, Estoup A (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. Bioinformatics **24**:2713-2719

Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. Heredity (Edinb) 93:504–509

Davey JL, Blaxter MW (2011) RADseq: Next-generation population genetics. Brief Funct Genomics 9:416–423. doi: 10.1093/bfgp/elq031

Davey JW, Hohenlohe PA, Etter PD, et al (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499–510. doi: 10.1038/nrg3012

Degen B, Sebbenn AM (2014) Genetics and tropical forests. In: Tropical Forestry Handbook. Pancel L, and Köhl M (eds). Berlin Heidelberg: Springer, pp 885–920

Diniz-Filho JAF, Diniz JVBPL, Rangel TF, et al (2013) A new eigenfunction spatial analysis describing population genetic structure. Genetica 141:479–489

Drummond, A., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. Bmc Evolutionary Biology. doi: 10.1186/1471-2148-7-214.

Duminil J, Mendene Abessolo DT, Ndiade Bourobou D, Doucet J-L, Loo J, Hardy OJ (2016) High selfing rate, limited pollen dispersal and inbreeding depression in the emblematic African rain forest tree *Baillonella toxisperma* – Management implications. For Ecol Manag 379:20–29

Dunning LT, Savolainen V (2010) Broad-scale amplification of matK for DNA barcoding plants, a technical note. Botanical Journal of the Linnean Society 164: 1–9.

Eduardo A, Lacerda B, Kanashiro M, Sebben AM (2008) Long-pollen Movement and Deviation of Random Mating in a Low density Continuous Population of a Tropical Tree *Hymenaea courbaril* in the Brazilian Amazon. Biotropica 40:462–470

Ekblom R, Wolf JBW (2014) A field guide to whole-genome sequencing, assembly and annotation. Evolutionary Applications 7:1026–1042.

Escudero A, Iriondo JM, Torres ME (2003) Spatial analysis of genetic diversity as a tool for plant conservation. Biol Conserv 113:351–365

Excoffier L, Heckel G (2006) Computer programs for population genetics data analysis: a survival guide. Nat Rev Genet 7:745–758. doi: 10.1038/nrg1904

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform online 1:47-50

Excoffier, L., J. Novembre, and S. Schneider. 2000. SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. Journal of Heredity **91**:506-509.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of Molecular Variance Inferred from Metric Distances among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. Genetics 131:479–491

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol Ecol Notes 7:574–578. doi: 10.1111/j.1471-8286.2007.01758.x

Fernández-Mazuecos M, Mellers G, Vigalondo B, et al (2017) Resolving Recent Plant Radiations: Power and Robustness of Genotyping-by-Sequencing. Syst Biol 67:250-268

Feyissa T, Nybom H, Bartish IV, Welander M (2007) Analysis of genetic diversity in the endangered tropical tree species *Hagenia abyssinica* using ISSR markers. Genet Resour Crop Evol 54:947–958. doi: 10.1007/s10722-006-9155-8

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709.

Gaino APSC, Silva AM, Moraes MA, et al (2010) Understanding the effects of isolation on seed and pollen flow, spatial genetic structure and effective population size of the dioecious tropical tree species *Myracrodruon urundeuva*. Conserv Genet 11:1631–1643

Ganal MW, Altmann T, Röder MS (2009) SNP identification in crop plants. Curr Opin Plant Biol 12:211–217. doi: 10.1016/j.pbi.2008.12.009

Gandolphi G, Bittencourt Jr NS (2010) Sistema reprodutivo do Ipê-Branco: *Tabebuia roseo-alba* (Ridley) Sandwith (Bignoniaceae). Acta Bot Brasilica 24:840–851

Gerber S, Chabrier P, Kremer A (2003) FAMOZ: a software for parentage analysis using dominant, codominant and uniparentally inherited markers. Mol Ecol Resour 3:479–481

Gill P, Jeffreys A, Werrett D (1985) Forensic application of DNA "fingerprints." Nature 318:577–579. doi: 10.1038/318577a0

Goodwillie C, Kalisz S, Eckert CG (2005) The evolutionary enigma of mixed mating systems in plants: occurrence, theoretical explanations, and empirical evidence. Annu Rev Ecol Evol Syst 36:47–79

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html. Updated from Goudet (1995).

Guillot G, Renaud S, Ledevin R, et al (2012) A Unifying Model for the Analysis of Phenotypic, Genetic, and Geographic Data. Syst Biol 61:897–911. doi: 10.1093/sysbio/sys038

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. New For 6:95–124. doi: 10.1007/978-94-011-2815-5

Hansen MC, Potapov P V, Moore R, et al (2013) High-resolution global maps of 21st-century forest cover change. Science 342:850–853

Hardy OJ, Maggia L, Bandou E, et al (2006) Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. Mol Ecol 15:559–571. doi: 10.1111/j.1365-294X.2005.02785.x

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Resour 2:618–620. doi: 10.1046/j.1471-8286.2002.00305.x

Hartl D, Clark A (2007) Principles of Population Genetics, Sinauer Associates, Sunderland. 1997. ISBN 0-87893-306-9

Hickerson MJ, Carstens BC, Cavender-Bares J, et al (2010) Phylogeography's past, present, and future: 10 years after Avise, 2000. Mol Phylogenet Evol 54:291–301. doi: 10.1016/j.ympev.2009.09.016

Hipp AL, Manos PS, González-Rodríguez A, et al (2018) Sympatric parallel diversification of major oak clades in the Americas and the origins of Mexican species diversity. New Phytol 217:439-452.

Ho SYW, Shapiro B (2011) Skyline-plot methods for estimating demographic historyfrom nucleotide sequences. Mol Ecol Resour **11**:423–434.

Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: defining, estimating and interpreting FST. Nat Rev Genet 10:639–50. doi: 10.1038/nrg2611

Hosking L, Lumsden S, Lewis K, et al (2004) Detection of genotyping errors by Hardy–Weinberg equilibrium testing. Eur J Hum Genet 12:395–399

Inghelandt D, Melchinger AE, Lebreton C, Stich B (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theor Appl Genet 120:1289–1299. doi: 10.1007/s00122-009-1256-2

Ismail SA, Ghazoul J, Ravikanth G, Uma Shaanker R, Kushalappa CG, Kettle CJ (2012) Does long-distance pollen dispersal preclude inbreeding in tropical trees? Fragmentation genetics of *Dysoxylum malabaricum* in an agro-forest landscape. Mol Ecol 21:5484–5496

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. Trends Ecol Evol 11:424–429

Jeffreys AJ, Neumann R, Wilson V (1990) Repeat unit sequence variation in minisatellites: A novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. Cell 60:473–485. doi: 10.1016/0092-8674(90)90598-9

Jump AS, Peñuelas J (2007) Extensive spatial genetic structure revealed by AFLP but not SSR molecular markers in the wind-pollinated tree, Fagus sylvatica. Mol Ecol 16:925–936. doi: 10.1111/j.1365-294X.2006.03203.x

Kalia RK, Rai MK, Kalia S, et al (2011) Microsatellite markers: an overview of the recent progress in plants. Euphytica 177:309–334. doi: 10.1007/s10681-010-0286-9

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106

Kerem B, Rommens JM, Buchanan JA, et al (1989) Identification of the cystic fibrosis gene: genetic analysis. Science 245:1073–80. doi: 10.1126/science.2570460

Lander TA, Boshier DH, Harris SA (2007) Isolation and characterization of eight polymorphic microsatellite loci for the endangered, endemic Chilean tree *Gomortega keule* (Gomortegaceae). Mol Ecol Resour 7:1332–1334.

Lander TA, Boshier DH, Harris SA (2010) Fragmented but not isolated: Contribution of single trees, small patches and long-distance pollen flow to genetic connectivity for *Gomortega keule*, an endangered Chilean tree. Biol Conserv 143:2583–2590. doi: 10.1016/j.biocon.2010.06.028

Leal BSS, Palma C, Pinheiro F, et al (2016) Phylogeographic Studies Depict the Role of Space and Time Scales of Plant Speciation in a Highly Diverse Neotropical Region. CRC Crit Rev Plant Sci 35:215–230. doi: 10.1080/07352689.2016.1254494

Lewis PO, Zaykin D (2001) Genetic data analysis: Computer Program for the Analysis of Allelic Data (version 1.1). Free program distributed by the authors over the internet from: http://lewis.eeb.uconn.edu/ lewishome/software.html.

Lewontin RC, Hubby JL (1966) A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics 54:595–609. doi: 10.1111/j.1601-5223.1973.tb01163.x

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25**:1451-1452.

Lopes, J. S., D. Balding, and M. A. Beaumont. 2009. PopABC: a program to infer historical demographic parameters. Bioinformatics **25**:2747-2749.

Loveless MD (2002) Genetic diversity and differentiation in tropical trees. In: Degen B, Loveless MD, Kremer A (eds) Modelling and experimental research on genetic processes in tropical and temperate forests, Kourou, French Guiana. Embrapa Amazônia Oriental, Belém, Brazil, pp 3–30

Lowe AJ, Boshier D, Ward M, et al (2005) Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees. Heredity (Edinb) 95:255–273

Luikart G, England PR (1999) Statistical analysis of microsatellite DNA data. Trends Ecol Evol 14:253–256

Maguilla E, Escudero M, Hipp AL, Luceño M (2017) Allopatric speciation despite historical gene flow: divergence and hybridization in *Carex furva* and *C. lucennoiberica* (Cyperaceae) inferred from plastid and nuclear RAD-seq data. Mol Ecol 26:5646–5662.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. Trends Ecol Evol 18:189–197

Mangaravite É, Vinson CC, Rody HVS, et al (2016) Contemporary patterns of genetic diversity of *Cedrela fissilis* offer insight into the shaping of seasonal forests in eastern South America. Am J Bot. 103:1-10 doi: 10.3732/ajb.1500370

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood based paternity inference in natural populations. Mol Ecol 7:639–655

Maués MM (2006) Estratégias reprodutivas de espécies arbóreas ea sua importância para o manejo e conservação florestal: Floresta Nacional do Tapajós (Belterra-PA). Instituto de Ciências Biológicas, Depatamento de Ecologia, Universidade de Brasilia. PhD Thesis.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. Mol Phylogenet Evol 66: 526-538

Meagher TR (1986) Analysis of paternity within a natural population of *Chamaelirium luteum*. 1. Identification of most-likely male parents. Am Nat 128:199–215

Medina-Macedo L, Sebbenn AM, Lacerda AEB, et al (2015) High levels of genetic diversity through pollen flow of the coniferous *Araucaria angustifolia*: a landscape level study in Southern Brazil. Tree Genet genomes 11:814 doi: 10.1007/s11295-014-0814-1

Miller, M. P. 2005. Alleles In Space (AIS): Computer Software for the Joint Analysis of Interindividual Spatial and Genetic Information. Journal of Heredity **96**:722–724.

Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. Trends Ecol Evol 19:208–216. doi: 10.1016/j.tree.2004.01.009

Nei, M. 1987. Molecular evolutionary genetics. Colombia University Press, New York.

Newton AC, Allnutt TR, Gillies ACM, Lowe AJ, Ennos RA (1999) Molecular phylogeography, intraspecific variation and the conservation of Tree Species. Trends Ecol Evol 14:140–145

Olsson S, Seoane-Zonjic P, Bautista R, et al (2017) Development of genomic tools in a widespread tropical tree, *Symphonia globulifera* L.f.: a new low-coverage draft genome, SNP and SSR markers. Mol Ecol Resour 17:614–630. doi: 10.1111/1755-0998.12605

Peakall ROD, Smouse PE (2006) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Resour 6:288–295

Petit RJ, Hampe A (2006) Some Evolutionary Consequences of Being a Tree. Annu Rev Ecol Evol Syst 37:187–214. doi: 10.1146/annurev.ecolsys.37.091305.110215

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–59

Pybus OG, Rambaut A (2002) GENIE: estimating demographic history from molecular phylogenies Bioinformatics **18**:1404–1405.

R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Rasmussen HB (2012) Restriction Fragment Length Polymorphism Analysis of PCR-Amplified Fragments (PCR-RFLP) and Gel Electrophoresis – Valuable Tool for Genotyping and Genetic Fingerprinting. In: Magdeldin S (2012) Gel Electrophoresis-Principles and Basics. InTech, pp 315–334

Ratnam W, Rajora OP, Finkeldey R, et al (2014) Genetic effects of forest management practices: global synthesis and perspectives. For Ecol Manag 333:52–65

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. Journal of Heredity 86:248–249

Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. Conserv Biol 17:230–237

Rice WR (1989) Analyzing tables of statistical tests. Evolution 43:223–225

Ritland K (1989) Correlated matings in the partial selfer *Mimulus guttatus*. Evolution 43:848–859

Ritland K (2002) Estimation of gene frequency and heterozygosity from pooled samples. Mol Ecol Resour 2:370–372

Ritland K, Jain S (1981) A model for the estimation of outcrossing rate and gene frequencies using n independent loci. Heredity (Edinb) 47:35–52

Robledo-Arnuncio JJ, Austerlitz F, Smouse PE (2007) POLDISP: a software package for indirect estimation of contemporary pollen dispersal. Mol Ecol Notes 7:763–766. doi: 10.1111/j.1471-8286.2007.01706.x

Rymer PD, Dick CW, Vendramin GG, et al (2013) Recent phylogeographic structure in a widespread “weedy” Neotropical tree species, *Cordia alliodora* (Boraginaceae). J Biogeogr 40:693–706. doi: 10.1111/j.1365-2699.2012.02727.x

Rymer PD, Sandiford M, Harris SA, Billingham MR, Boshier DH (2015) Remnant *Pachira quinata* pasture trees have greater opportunities to self and suffer reduced reproductive success due to inbreeding depression. Heredity 115:115–124

Saiki R, Scharf S, Faloona F, et al (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science. 230:1350–1354

Sala OE, Chapin FS, Armesto JJ, et al (2000) Global biodiversity scenarios for the year 2100. Science 287:1770–1774

Schlötterer C (2004) The evolution of molecular markers--just a matter of fashion? Nat Rev Genet 5:63–69. doi: 10.1038/nrg1249

Scotti-Saintagne C, Dick CW, Caron H, et al (2013) Phylogeography of a species complex of lowland Neotropical rain forest trees (*Carapa* , Meliaceae). J Biogeogr 40:676–692. doi: 10.1111/j.1365-2699.2011.02678.x

Scotti-Saintagne C, Dick CW, Caron H, Vendramin GG, Troispoux VR, Sire P, Casalis M, Buonamici A, Valencia R, Lemes MR, Gribel R, Scotti I (2013) Amazon diversification and cross-Andean dispersal of the widespread Neotropical tree species Jacaranda copaia (Bignoniaceae). J Biogeogr **40**:707–719

Sebbenn AM (2002) Amostragem de freqüências alélicas e índices de diversidade genética em espécies arbóreas. Rev Inst Flor 14:27–38

Sebbenn AM (2006) Sistema de reprodução em espécies arbóreas tropicais e suas implicações para a seleção de árvores matrizes para reflorestamentos ambientais. In: Higa AR, Silva LD (eds) Pomares de sementes de espécies florestais nativas. FUPEF, Curitiba, pp 193–198

Sebbenn AM, Carvalho ACM, Freitas MLM, et al (2011) Low levels of realized seed and pollen gene flow and strong spatial genetic structure in a small, isolated and fragmented population of the tropical tree *Copaifera langsdorffii* Desf. Heredity (Edinb) 106:134–45. doi: 10.1038/hdy.2010.33

Sedgley M, Griffin AR (1989) Sexual reproduction of tree crops. Academic Press, London

Shaw J , Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am J Bot 92: 142- 166

Silva MB, Kanashiro M, Ciampi AY, et al (2008) Genetic effects of selective logging and pollen gene flow in a low-density population of the dioecious tropical tree *Bagassa guianensis* in the Brazilian Amazon. For Ecol Manag 255:1548–1558

Silva-Junior OB, Grattapaglia D, Novaes E, Collevatti RG (2018) Genome assembly of the Pink Ipê (Handroanthus impetiginosus, Bignoniaceae), a highly valued, ecologically keystone Neotropical timber forest tree. GigaScience 7: 1–16 doi.org/10.1093/gigascience/gix125

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462

Smouse PE, Dyer RJ, Westfall RD, Sork VL (2001) Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females. Evolution 55:260–271

Smouse, P. E., and R. Peakall. 1999. Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. Heredity **82**:561–573

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 17: 1105-1109

Tajima F (1989) Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**:585-595

Tambarussi EV, Boshier D, Vencovsky R, et al (2017) Inbreeding depression from selfing and mating between relatives in the Neotropical tree *Cariniana legalis* Mart. Kuntze. Conserv Genet 18:225–234. doi: 10.1007/s10592-016-0896-4

Tarazi R, Moreno MA, Gandara FB, et al (2010) High levels of genetic differentiation and selfing in the Brazilian cerrado fruit tree *Dipteryx alata* Vog.(Fabaceae). Genet Mol Biol 33:78–85

Tautz D, Renz M, Molecular E (1984) Simple sequencecs are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res 12:4127–4138

Turchetto-Zolet, A. C., F. Cruz, G. G. Vendramin, M. F. Simon, F. Salgueiro, M. Margis-Pinheiro, and R. Margis. 2012. Large-scale phylogeography of the disjunct Neotropical tree species Schizolobium parahyba (Fabaceae-Caesalpinioideae). Molecular Phylogenetics and Evolution **65**:174-182

Valdes AM, Slatkin M, Freimer NB (1993) Allele frequencies at microsatellite loci: The stepwise mutation model revisited. Genetics 133:737–749

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-Checker: Software for Identifying and Correcting Genotyping Errors in Microsatellite Data. Mol Ecol Notes 4:535–538. doi: 10.1111/j.1471-8286.2004.00684.x

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. Mol Ecol 13:921–935

Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. Genet Sel Evol 34:275–305

Vinson CC, Dal'Sasso TCS, Sudré CP, et al (2015) Population genetics of the naturally rare tree *Dimorphandra wilsonii* (Caesalpinioideae) of the Brazilian Cerrado. Tree Genet Genomes 11:46. doi: 10.1007/s11295-015-0876-8

Vinson CC, Kanashiro M, Sebbenn AM, et al (2014a) Long-term impacts of selective logging on two Amazonian tree species with contrasting ecological and reproductive characteristics: inferences from Eco-gene model simulations. Heredity (Edinb) 115:130–139. doi: 10.1038/hdy.2013.146

Vinson CC, Kanashiro M, Harris SA, Boshier DH (2014b) Impacts of selective logging on inbreeding and gene flow in two Amazonian timber species with contrasting ecological and reproductive characteristics. Mol Ecol 24:38–53.

Vos P, Hogers R, M B, et al (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Waddington KD (1983) Foraging behavior of pollinators. In Real L (ed) Pollination Biology. Academic Press, NY, pp 213–239

Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Mol Ecol 10:249–256

Wakeley J (2008) Coalescent Theory: An Introduction. Roberts & Company Publishers, Greenwood Village, CO.

Wang DG (1998) Large-Scale Identification, Mapping, and Genotyping of Single-Nucleotide Polymorphisms in the Human Genome. Science 280:1077–1082. doi: 10.1126/science.280.5366.1077

Weitemier K, Straub S, Cronn R, Fishbein M, Schmickl R, McDonnell A, Liston A (2014) hybseq: Combining target enrichment and genome skimming for plant phylogenomics. Applications in Plant Sciences 2:1400042.

Wigginton JE, Cutler DJ, Abecasis GR (2005) A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet 76:887–893

Williams JG, Kubelik AR, Livak KJ, et al (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl Acids Res 18:6531–6535. doi: 10.1093/nar/18.22.6531

White, T.J., Bruns, T., Lee, S., and Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. PCR Protocols: a Guide to Methods and Applications, pp 315–322

Wright J (1976) Introduction to forest genetics. Academic Press, New York

Wünsch A, Hormaza J (2002) Cultivar identification and genetic fingerprinting of temperate fruit tree species using DNA markers. Euphytica 125:59–67. doi: 10.1023/a:1015723805293

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. Mol Ecol 11:1–16

Zietkiewicz E, Rafalski A, Labuda D (1994) Genome Fingerprinting by Simple Sequence Repeat (SSR)-Anchored Polymerase Chain Reaction Amplification. Genomics 20:176–183