We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Microsatellites as a Tool for the Study of Microevolutionary Process in Native Forest Trees

Maria Eugenia Barrandeguy and
Maria Victoria Garcia

Additional information is available at the end of the chapter

Abstract

The main aim of this work is an attempt to help researchers that use microsatellite markers to analyze microevolutionary forces in natural populations of native forest species. This kind of studies drives the researchers to make decisions regarding management or conservation of such species. This chapter pays attention to the entire process—from development of microsatellite markers, going through data analysis and ending with interpretation of these results. This work helps to researchers that are not familiarizing with methods and population genetics theories to analyze nuclear and chloroplast microsatellite data. These methods allow quantification of genetic variation and genetic structure in native forest species, and theoretical content allows knowledge about the past and the present genetic states of populations for making inferences about the future of these populations.

Keywords: cpSSRs, nuSSRs, native forest trees, population genetics

1. Introduction

Patterns of distribution of genetic variation in the landscape reflect the responses of species to evolutionary forces operating within current and past environments, and it can tell us much about how species have evolved and may continue to evolve in the future [1]. Most studies on genetic variation patterns within tree species were primarily motivated by attempts to improve our understanding of biodiversity at the intraspecific level or the evolutionary dynamics within plant species in an early stage of domestication [2]. However, forest tree species have many valuable subjects to be explored, and problems could be solved using microsatellite markers in combination with appropriate statistical analyses to make recommendations for

conservation of forest genetic resources [3], infer the origin of forest plants and woods [2], and conduct molecular tree improvements [4].

The main aim of this work is an attempt to help researchers that use microsatellite markers to analyze microevolutionary forces in natural populations of native forest species. This kind of studies drives the researchers to make decisions regarding management or conservation of such species.

2. The challenge to work with microsatellite markers in species without economical interest

Native forest species are interesting models of biodiversity study because they give valuable information about current and past conditions that could have influence on the amount and distribution of genetic variation in natural populations. Hence, long-lived tree species have witnessed climatic, demographic, and/or ecological changes, and all these changes left genetic traces that can be studied using microsatellite markers (Simple Sequence Repeats - SSRs). However, every interesting point about working with native species has its unfavorable counterpart because of low economic value of native forest species. One of the limitations is the lack of DNA sequence information needed to develop and use simple sequence repeats (SSRs). Unfortunately, SSRs are not universal markers, and species specificity of SSR loci in plants is a major constraint to their ubiquitous adoption [5], although limited cross-species transferability of SSR loci of closely related taxa is possible.

The starting point of a genetic study using SSRs in native forest trees is getting species-specific SSRs, e.g., searching in the nucleotide section of GeneBank public database. In the case of unavailability of species-specific SSRs, an alternative is to search SSR primers developed for a phylogenetically closely related species because as mentioned above, empirical studies have demonstrated that cross-species transfer of nuclear microsatellite markers is possible [6]. Using latter methodology, SSRs developed for one species can be used to detect polymorphism at homologous sites in related species. However, the repeat sequence and the flanking regions-containing primer binding sites must be conserved across taxa to detect polymorphism at homologous sites in related species [5].

The success of heterologous PCR amplification will depend upon evolutionary distance between the source and the target species because empirical studies have shown an inverse relationship between primer site conservation and evolutionary relationship between tested taxa [5]. Cross-species transferability of polymorphic markers in plants is mainly successful within genera (success rate close to 60% in eudicots and close to 40% in the reviewed monocots), whereas between genera, cross-species transfer rates are approximately 10% for eudicots [6]. There are studies with native forest tree species in which cross-species transfer were successful, e.g., *Quercus* [7–9], *Prosopis* [10], *Eucalyptus* [11], *Enterolobium* [12], *Pithecellobium* [13], *Araucaria* [14], and *Taxus* [15].

In the worst of cases, cross-species transfer of SSRs may not work. Hence, we propose two nonmutually excluding alternatives: the development of species-specific microsatellites for

nuclear genome (nuclear Simple Sequence Repeats – nuSSRs) and the use of chloroplast microsatellite markers (chloroplast Simple Sequence Repeats - cpSSRs). These alternatives are very different regarding to genetic information they provide and its cost in terms of time and money. Many laboratories have enough resources and expertise for conducting SSR-based research but not for characterizing new loci [5].

Microsatellite markers are present in chloroplast genome but particular traits of this genome provide different population genetic information than nuSSRs. Organelle genomes are typically nonrecombinant, uniparentally inherited, and effectively haploid [16]. Unlike the conventional approach for obtaining nuclear microsatellites, when cpSSRs primers designed for one species can regularly cross-amplify in related species, giving an opportunity to develop efficient "universal" SSR primers that show widespread intraspecific polymorphism. Chloroplast SSR primers developed by Weising and Gardner [17] are the most popular in Angiosperms. However, the low mutation rates associated with the chloroplast genome meant that detection of enough variation represents a major technical barrier for the widespread application of a particular marker [16].

3. Development of species-specific nuclear microsatellite markers

Population geneticists, forestry breeders, and ecologists starting a new research that must contend with a dichotomous decision: the isolation of species-specific microsatellite markers or application of multilocus fingerprinting approaches. The advantages of hypervariable, codominant markers as SSRs are well documented [18], but in many cases, the perceived difficulties of SSRs isolation act as a deterrent for the utilization of this class of markers [19].

In recent years, publications of new species-specific nuSSRs in forest tree species and in other plant taxa are frequent in most of journals. However, development of species-specific nuSSRs is time and cost consuming. Also, specific laboratory and technical conditions are needed, costly and laborious cloning and screening procedures limit the number of species that can be studied.

As a consequence of the diverse publications and techniques for nuSSRs development, in this section, we will exclusively focus on the SSR development procedure in plants addressing to describe the typical situation that the researchers must consider when working with nonmodel organisms. Our own experience comes from the development of specific nuSSRs for *Anadenanthera colubrina* var. *cebil* (Mimosoideae, Leguminosoideae), a native forest tree species from South America [20]. Laboratory work was started to cross-species transfer of nuSSRs developed for other legume tree species because SSR primers from species of the same genera were not available. Eighteen primer pairs from six different species were tested including *Koompasia malascensis*, *Acacia nilotica*, *Geoffroea spinosa*, *Prosopis sp.*, *Dinizia excelsa*, and *Parkia panurensis*. Results of cross-species transfer were unsatisfactory, and development of species-specific nuSSRs was necessary.

The successful isolation of SSRs involves several steps: (1) preparation of a microsatellite-enriched genomic library, (2) cloning and sequencing of fragments containing microsatellites,

(3) primer design, and (4) testing the functionality of SSR primers and polymorphisms in tested genotypes. There is a potential loss of loci at each stage. A number of loci that will finally constitute the working primer set are a fraction of the original number of sequenced clones, which is called attrition rate [19]. Microsatellite markers were developed from two microsatellite-enriched genomic libraries screening for an increase in the variability of microsatellite motifs. The results were notoriously different as only one library gave positive results. The libraries were developed using the enrichment procedure proposed by Fischer and Bachmann [21] and modified by Prinz et al. [22]. **Table 1** shows the attrition rates for this work. The development of specific nuSSRs for *A. colubrina* var. *cebil* demanded 3 months of work in a fully equipped laboratory. Human resources involved in the development of SSRs included a technical assistant, a doctoral student, and an experienced researcher.

Library	Oligo pool	Efficiency of enrichment			Efficiency of primer design					Functionality	
		N° of inserts sequenced	N° of inserts with SSRs	High quality sequences with SSRs	Sequences with flanking region suitable to design primers	Sequences with a unique SSRs locus	Sequences without restriction site inside the SSRs region	Primers designed	Tested primers	Clearly amplified loci	Polymorphic loci
A	(GA)$_{10}$	106	77 (73%)	56 (73%)	84%	70%	98.2%	30 (52%)	30 (52%)	16 (51.7%)	8(50%)
B	(CA)$_{10}$ (GAA)$_8$ (CAA)$_8$	98	20 (20.5%)	6 (30%)	100%	100%	100%	6 (100%)	1 (17%)	0(0%)	–
Total		204	97 (47.6%)	62 (64%)	85.5%	72.6%	98.4%	36 (56.5%)	30 (48.4%)	16(50%)	8(50%)

Table 1. Attrition rates for *A. colubrina* var. *cebil* specific nuSSRs development.

From our experience, we suggest to pay attention on the following: starting the process with DNA of good quality and enough quantity; ensuring good conditions of sterility during enrichment procedure and in the whole process; making two simultaneous libraries using different sets of repeated motifs for enrichment; avoiding repetitive bases in primer sequences; analyzing the primer sequences directly in the electropherograms to ensure that primers were designed on sequences of good quality with high peaks; resequencing amplified products after functionality tests; and aligning the original fragment obtained from the enrichment procedure.

New and revolutionary sequencing methods, referred to as next-generation sequencing (NGS), are extremely high-throughput technologies that produce thousands or millions of sequences at once at a fraction of the cost of traditional Sanger methods [23]. A specific application of this new technology in plants is the possibility of rapid and cost-effective discovery of microsatellite loci [23]. Despite this modern technology is more cost effective than traditional enrichment

procedures, currently it is not yet widely used for nuSSRs development in plant species. A commonly cited weakness of microsatellites is their high development cost and relatively low-throughput when compared to SNPs but the same technologies that have widened the use of SNPs have also benefited microsatellite development processes [24].

Once a set of nuSSRs primers was developed, the final step was the statistical analyses of data to confirm the utility of these markers to population genetic studies. These analyses consist of: (1) estimation of observed and expected heterozygosity, (2) test of Hardy-Weinberg equilibrium, (3) test of genotypic linkage disequilibrium, (4) test of null alleles and genotyping errors, and (5) perform neutrality test. There are free software available for these analyses, e.g., Genalex [25], Genepop [26], and/or Microchecker [27]. Expected good results for these analyses include high heterozygosity, high number of loci in Hardy-Weinberg equilibrium and linkage disequilibrium, low number of loci with null alleles and absence of genotyping errors, and lack of traces of selection.

4. What do microsatellite markers say us about natural populations of forest tree species?

The development of molecular genetic markers has had a great impact on our understanding of the processes that determine structure and variation within and among natural populations [16]. Microsatellites, as other molecular markers, are particular characteristics of DNA molecule that enable the identification of individuals at DNA level. However, a molecular marker must be considered as genetic marker when its particular genetic features are known. The knowledge on precise molecular basis and a mode of inheritance of a genetic polymorphism are crucial for the appropriate interpretation of molecular marker data in a population context [28].

Plants show a remarkable variety of inheritance modes, and further, some of their reproductive patterns permit genetic study with means not available in other types of organisms [29]. The mitochondrial genome in plants shows a large size, slow nucleotide substitution rates and extensive levels of intramolecular recombination, and has been of limited use in genetic diversity studies. The chloroplast genome shows conserved gene order and a general lack of heteroplasmy and recombination, and it is an attractive tool for demographic and phylogenetic studies [16]. There is considerable potential for hypervariable chloroplast microsatellites to provide markers with uniparental inheritance for indirect measures of seed or pollen gene flow. Studies of angiosperms, where chloroplast DNA (cpDNA) is predominantly maternally inherited, might offer further insights and provide information on the patterns and extent of localized seed dispersal [16]. Furthermore, its uniparental mode of inheritance makes it possible to elucidate the relative contributions of seed and pollen gene flow to the genetic structure of natural populations by comparing nuclear and chloroplast markers [16].

The use of genetic markers with uni- and biparental inheritance (i.e., cpSSRs and nuSSRs) differentiates the historical contributions of the movement of seed and pollen on the levels of gene flow. This information is relevant to distinguish between genetic consequences of

colonization by seed and the exchange of genes through pollen between established populations [30, 31]. Given haploid genome of organelles, effective population size in hermaphrodite outcrossing plants is half that of diploid nuclear genome, and as a result, chloroplast-specific markers should be good indicators of historical bottlenecks, founder effects, and genetic drift [16].

Differences in mutation rates, ploidy levels, and recombination presence or absence between nuclear and chloroplast genomes make cpSSRs and nuSSRs valuable tools for the study of the effects of historical and recent fragmentation on the contemporary genetic variation and current population genetic structure. This allows contrasting the relative role of genetic drift and gene flow as microevolutionary process that shapes population genetic structure [32].

In addition, due to their high rate of polymorphism, nuclear microsatellites are often cited as being very useful for studying recent evolutionary events among subpopulations within an individual species [24].

Microevolutionary process	cpSSRs	Both	nuSSRs
Gene flow	Gene flow by seeds	–	Gene flow by pollen and seed
Genetic drift	Historical fragmentation	–	Recent fragmentation
Genetic drift vs. gene flow	–	Relative contributions of seed and pollen flow to the genetic structure of natural populations	–
Factor			
Demographic events	Colonization/ expansion Historical bottlenecks Founder effects	–	–

Table 2. Microevolutionary processes and demographic events that can be studied by microsatellite markers.

The movement of alleles within and between natural populations and their interaction with genetic drift, mutation, and natural selection determine the genetic composition of a population, including its genetic diversity and genetic structure [28]. Microsatellites allow taking a high-resolution snapshot of a given allelic composition at a given time for certain loci, and the studying of mechanisms that generate and maintain genetic variability is possible by means of population genetics theories and methods.

Great potential exists for the application of coalescent-based models to cpSSRs [16]. Coalescent approaches can be extremely useful in assessing a range of demographic histories but their application to intraspecific studies in plants has been hampered by the slow mutation rate of

the nonrecombinant genomes, such as chloroplast DNA. Although limitations exist, cpSSRs represent a potentially informative data source with which coalescent-based approaches can be explored [16]. Microevolutionary processes and demographic events that can be studied by microsatellite markers are showed in **Table 2**.

5. Population genetic data analysis

This section attempts to guide the researchers to make decisions regarding the statistical analysis of nuclear and chloroplast microsatellite data. Nevertheless, those attempting to use these analyses for the first time will need to read the cited bibliography here for each particular analysis. Advances in computing technology have inspired the use of intensive statistical approaches such as maximum likelihood, Bayesian probability theory, and Markov chain Monte Carlo simulation contributing to the recent technical advancements of molecular ecology [33].
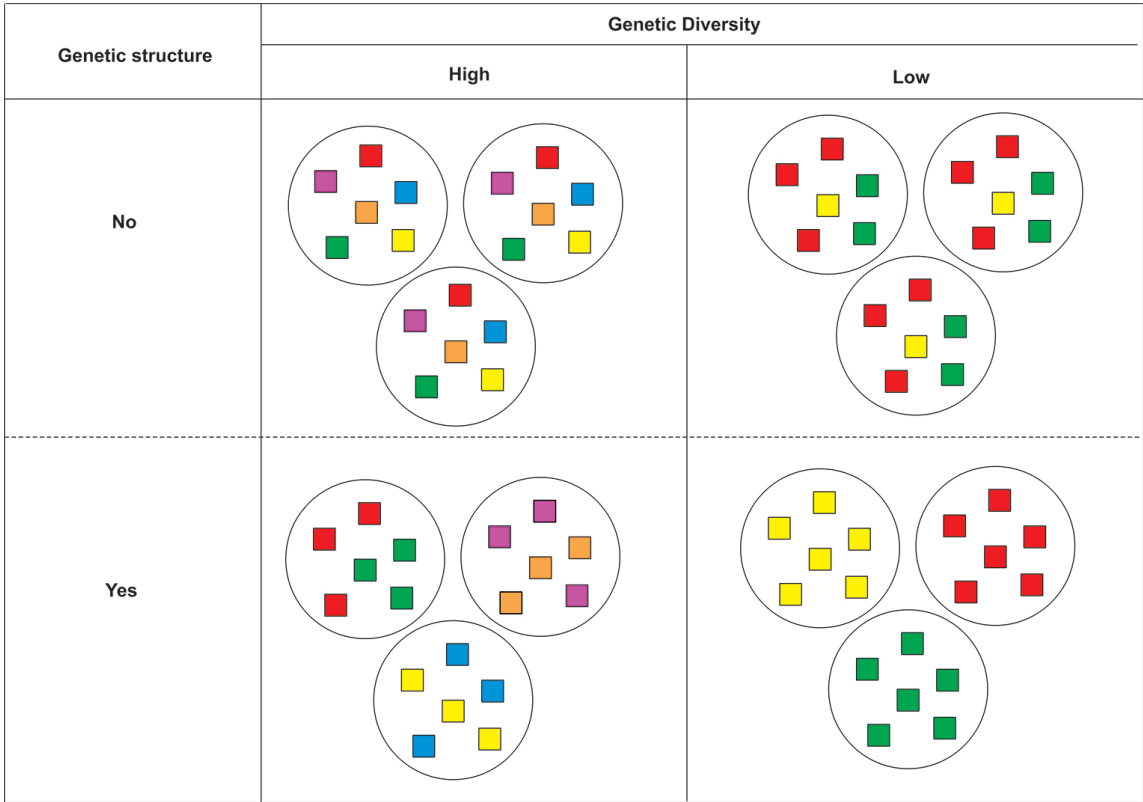

Figure 1. Extreme states of allelic configuration of one theoretical population integrated by three subpopulations. Circles: subpopulations; Squares: nuSSRs genotypes or cpSSRs haplotypes. Colors show differences in the allelic compositions.

Fourth extreme and simple states of allelic configuration of a theoretical population integrated by three subpopulations are showed in **Figure 1**. Each state was defined from the relationship between genetic diversity levels and genetic structure of the theoretical population. In the

nature, these extremes are exceptions while complex allelic configurations are the usual situations. Statistical analyses are the most appropriate tools to infer levels and distribution patterns of genetic variation while population genetic theory gives the knowledge for interpreting statistical data analysis results.

5.1. Genetic diversity

A prerequisite for starting population studies is to detect the genetic diversity underlying phenotypic variation, and understand the genetic diversity as the total genetic variation among individuals within a population. Several measures of genetic diversity have been developed over the years. The simplest measure of genetic diversity from molecular data is the number of alleles at a given locus (N_A), which is also known as gene multiplicity [34]. At the same time, these alleles have their own frequencies in each population, which represent their abundance. As multiplicity and abundance vary independently, genetic diversity can be expressed as the effective number of alleles (N_E) [35]. N_E will be equal to N_A if alleles show the same frequency. In case that allele frequency distribution is not uniform, N_E will be lower than N_A. The number of alleles that can be found in only one population is defined as private alleles (N_P) [36]. In this way, N_P is a simple measure of genetic distinctiveness. Private alleles can also have low frequencies being able to call them rare alleles. These kinds of alleles are very informative because their presence and frequency allow quantification of gene flow levels.

Since the number of detected alleles in a population depends on its size, it is not advisable to compare genetic diversity parameters among subpopulations with different sizes. An useful parameter to compare the number of alleles between samples that differ in size is the allelic richness (R). This parameter predicts the expected number of alleles if samples have the same size using the rarefaction method [37]. However, the original and most important measure of genetic diversity is Nei's gene diversity index (h) estimated as $h = \sum_1^n \left(1 - x_i^2\right)$, where x_i indicates the allele frequency [38]. This parameter represents the probability that two alleles randomly and independently selected from a gene pool will represent different alleles. This index analyzes allele frequency variation directly in the terms of heterozygosity without consideration of the number of alleles at a given locus or the pattern of evolutionary forces [38]. In this way, the treatment of this index is biologically most appropriate because it has been formulated entirely in terms of allele and genotype frequencies [39].

Since particular genetics features of chloroplast genome, combination of cpSSRs alleles from different loci allows determining the chloroplast haplotypes. Hence, haplotype genetic multiplicity could also be characterized from haplotype number, genetic abundance from haplotype frequencies, genetic distinctiveness from the number of private haplotypes, and genetic diversity from Nei's haplotypic diversity index (H) estimated as $H = \left\{\left[\frac{n}{n-1}\right]\left[1 - \sum_1^n p_i^2\right]\right\}$, where n is the number of analyzed individuals and p_i is the frequency of haplotypes in the population [40, 41] (**Table 3**).

Population genetic measure	Objective of analysis	Analysis		Statistical estimation	Suggested software[a]
Genetic diversity	Characterization	Genetic multiplicity		Number of alleles (N_A) Number of chloroplast haplotypes (N_H)	Genalex [25]
		Genetic abundance		Allele frequencies Haplotype frequencies	Genalex [25] Arlequin [59]
		Multiplicity vs. abundancy		Effective number of alleles (N_E)	Genalex [25]
		Genetic distinctiveness		Number of private alleles (N_P) Number of private haplotypes (N_{PH})	Genalex [25]
		Heterozygosity		Observed heterozygosity (H_O) Expected heterozygoity (H_E)	Genalex [25]
	Quantification	Genetic diversity		Nei´s gene diversity index (h) Nei's haplotypic diversity index (H)	Genalex [25]
				Allelic richness (R)	ADZE [60] FSTAT [61]
Genetic structure	Determination	Individual-based methods	Methods based on distance	Median joining trees	Darwin [62] PopArt [63]
				Haplotype network	Network [64]
			Methods based on models	Bayesian admixture analysis for nuclear data	Structure [44] Structure Harvester [65]
				Bayesian mixture analysis for linked cpSSR data	BAPS [45]
		Subpopulation-based methods	Hierarchical structure using nongenetic criteria	AMOVA using geographic definition of groups	Arlequin [59]
			Hierarchical structure using a genetic criteria	AMOVA using clusters defined by Bayesian analysis	Arlequin [59]
	Quantification	Wright's F_{ST}		Nuclear genetic structure (F_{STnu})	Arlequin [59]

Population genetic measure	Objective of analysis		Analysis	Statistical estimation	Suggested software[a]
				Chloroplast genetic structure (F_{STcp})	
Gene flow	Quantification	Historical	Indirectly estimated from F_{ST}	Gene flow by pollen and seed (N_em_{nu})	Arlequin [59]
				Gene flow by seeds (N_em_{cp})[b]	
			From rare alleles	Gene flow by pollen and seed (N_em_{nu})	Genepop [26]
		Recent	Probability to be a migrant or a recent migrant descendent	Bayesian genetic structure analysis with *a priori* individual´s geographical position information	Structure [44]
	Gene flow by pollen versus gene flow by seeds		Indirectly estimated from levels of nuclear and chloroplast genetic structure	Ennos´s equation (r)	–
Inbreeding	Quantification		Hierarchical subpopulation-based analyses including within individual level (F_{IS})	AMOVA including within individual levels	Arlequin [59]
			Bayesian inbreeding inference (*f*)	Bayesian analysis	Hickory [52]
			Inbreeding coefficient inference considering null alleles F_{ISnull}	Computer simulations to simultaneous estimation of null alleles by locus and inbreeding coefficient as a multilocus parameter	INEST [53]
Demographic events	Population expansion determination		Neutrality tests	F_S neutrality test (F_S) D_{Tajima} (D_{Tajima})	Arlequin [59]
	Phylogeography		Approximate Bayesian computation (ABC)	Parameter estimation (effective sizes of current and ancestral populations, immigration rates, splitting times, and tree topology) and ancestral model comparison	DiYABC [66]

[a]Free available software.
[b]For angiosperm species.

Table 3. Classification of methods for analysis of microevolutionary processes and demographic events and suggested software.

The differences in the genetic diversity parameters among populations must be statistically significant to arrive at the conclusion of which population is the most diverse. These differences could be tested by permutation, a nonparametric procedure.

5.2. Genetic structure

Population genetic structure is the amount of genetic variability and its distribution within and among local populations and individuals within a species [42]. Given the central role of population genetic structure to microevolutionary processes, additional tools for its measurement and quantification are necessary. In this way, we perform a classification of the several statistical methods to study population genetic structure. However, the researchers must keep in mind the scope and aims of the study to define the analyses of their data.

5.2.1. Individual-based methods

The starting points for these analyses are individual microsatellite's genotypes. The simplest methods are those based on distances, e.g., median joining trees (MJ) and networks (NWK) [43]. These methods are graphical representation of genetic distances among multilocus genotypes (nuSSRs data) or among haplotypes (cpSSRs data). Distance-based methods are usually easy to apply and are often visually appealing. However, the clusters identified may be dependent on both the distance measure and graphical representation chosen, being difficult to assess confidence of clusters obtained [44].

Nowadays, the most popular individual-based methods are those based on models as Bayesian admixture analysis for nuSSRs data [44] and Bayesian mixture analysis for linked loci for cpSSRs data [45]. Methods based on Bayesian theory are extensively used because they give information about the genetic origin of individuals making clusters and assigning individuals to these clusters to infer population structure based on a probabilistic criterion. In addition, these methods include *a priori* information of the geographic origin of individuals to help in the population genetic structure determination [44, 45] and the identification of migrants or descendants of recent immigrants [44]. The results of Bayesian admixture analyses must be analyzed by the Evanno method [46] to determine the most likely level of population subdivision.

5.2.2. Subpopulation-based methods

The starting points of these analyses are groups of individuals. Here these groups are called subpopulations. Different criterion could be considered to grouping individuals: a nongenetic criterion (e.g., geographical groups of individuals, cohort, etc.) or a genetic criterion (e.g., previously identified Bayesian clusters). Subpopulation-based methods are based on the analysis of molecular variance (AMOVA) [47]. This method consists in the analysis of distribution of molecular genetic variation in the previously established different hierarchical levels. Once genetic structure was determined, the strength of genetic structure must to be quantified. The most appropriate way is the estimation of Wright's fixation index (F_{ST}). Sewall Wright [48] devised the fixation index to describe correlations among alleles sampled at hierarchically

organized levels of a population. Hence, this index could be estimated from the AMOVA as $F_{ST} = \left(\sigma_a^2 - \sigma_T^2\right)/\sigma_{T'}^2$, where σ_a^2 is the variance among subpopulations and σ_T^2 is the total population variance [49]. Statistical significance estimation of this index could be performed using permutations. F_{ST} index could be estimated among pairwise subpopulations to determine patterns of genetic differentiation (**Table 3**).

5.3. Gene flow

Genetic exchange between local populations is called gene flow, and it is an evolutionary force that occurs between populations with distinct gene pools [42]. We are going to introduce two ways to estimate levels of gene flow among subpopulations from microsatellite data. The first way is the indirect method based on genetic differentiation among populations [50]. After Wright [48] developed fixation index, he went on to demonstrate that there is a simple relationship between the genetic divergence of two populations, measured as F_{ST}, and the amount of gene flow between them, which is given as $F_{ST} = 1/4 \ \left(N_e m + 1\right)$, where Ne is the effective size of each population and m is the migration rate between populations, and therefore $N_e m$ is the number of breeding adults that are migrants. Hence, for nuSSRs data, the number of migrants could be estimated as $N_e m = \left(1/F_{ST} - 1\right)/4$ and it quantifies the historical gene flow by pollen and seed, while for cpSSRs data, the number of migrants could be estimated as $N_e m = \left(1/F_{ST} - 1\right)/2$ and it quantifies the historical gene flow by seeds in angiosperm species [42]. The second way to estimate gene flow is the method of rare alleles described by Slatkin [51]. He proposed the estimation of Nm from the spatial distribution of rare alleles. He demonstrated that $\log_{10}\left[\bar{p}_{(1)}\right]$, where $\bar{p}_{(1)}$ indicates the average frequency of private alleles, is approximately lineal with $\log_{10}Nm$.

The estimation of relative rates of pollen and seed gene flow could be estimated from an estimator proposed by Ennos [31], which is based on the conception that the effectiveness of pollen and seed in bringing about gene flow depends upon the mode of inheritance of the genetic marker. In most of the angiosperms, gene flow occurs by pollen and seed for nuclear and paternally inherited markers; however, gene flow occurs only by seeds for maternally inherited markers. Consequently, different levels of population differentiation for markers with contrasting modes of inheritance are expected. The relative levels of pollen versus seed gene flow among populations could be estimated as $r = \left[\left(1/F_{STb} - 1\right) - 2 \times \left(1/F_{STm} - 1\right)\right]/\left(1/F_{STm} - 1\right)$, where F_{STb} and F_{STm} are fixation index for nuclear and chloroplast markers for an angiosperm species, respectively.

Finally, Pritchard et al. [43] extended their approach to infer genetic structure including in the algorithm, the geographic position of individuals. In essence, it assumes that each individual originated, with high probability, in the geographical region in which it was sampled, but to allow some small probability that it is an immigrant (or has immigrant ancestry). Immigrants

would be individuals whose genetic makeup suggests they were misclassified, and in this way it is possible to quantify recent gene flow (**Table 3**).

5.4. Inbreeding

In its most basic sense, inbreeding is mating between biological relatives [42]. It is not a microevolutionary process because its effect does not change allele frequencies of populations. However, it is important because genotypic composition of populations could be determined by its influence. The presence of inbreeding informs us about reproductive dynamics of the species. Inbreeding coefficient (F_{IS}) could be estimated from AMOVA if the hierarchical level "within individuals" is included in it. Inbreeding could also be estimated using a Bayesian approach (*f*) [52].

Although microsatellites are a very efficient tool for many population genetics applications, they may occasionally produce null alleles, which, when present in high proportion at a particular locus, the observed heterozygosity would be underestimated. As a consequence, the population parameter estimates based on the proportion of heterozygotes could be affected by null alleles. Estimates of Wright's inbreeding coefficient F_{IS} based on microsatellite data could be unclear regarding to the extension of actual level of inbreeding in the studied population and in the degree affected by the presence of null alleles. Population inbreeding model can be applied for simultaneous estimation of null allele frequencies and of the inbreeding coefficient as a multilocus parameter [53] (**Table 3**).

5.5. Demographic events

There has been little focus on the potential of chloroplast microsatellites for demographic inference. Navascués et al. [54] investigated the utility of cpSSRs data for the detection of demographic expansions. The study of historical demography by means of genetic information is based on coalescent theory [55]. One alternative is the development of the F_S neutrality test for determination of population expansion events [56]. This test is based on different expectations for the number of haplotypes when comparing a stationary with expansion demography [56]. Another alternative is the estimation of D_{Tajima} index as $D_{Tajima} = \pi - (\theta)/\sqrt{V[\pi-(\theta)]}$ where, π is the number of different sites between sequences, V is the numerator variance, and θ is estimated as $\theta = S/a$, where S is the number of polymorphic sites and a is calculated by $a = \sum_{i=1}^{n-1} 1/i$, where i takes values of [1 – (*n*-1)] and where *n* is the number of analyzed sequences [57]. Both parameters should be estimated from the distribution of the differences between individuals within a population, and these differences are considered as allelic differences between cpSSRs haplotypes, being this data considered as binary [54].

A new and robust method to examine a species' phylogeography using microsatellite markers is the approximate Bayesian computation (ABC). This model-based method is useful to infer parameters and compare models in population genetics [58] (**Table 3**).

6. Understanding population genetic data analysis results

The challenge in the study of population genetic events based on microsatellite markers is the interpretation of the statistical analysis results from the biological point of view. The aim of this section is to serve as a guide about how to interpret these results in a forest tree species in order to infer which are the forces that determine the distribution of current genetic variation of nuSSRs and cpSSRs?

In the field of population genetics, it is becoming increasingly necessary to focus more attention on understanding the practical limitations of various analyses and applying increased caution when interpreting results generated by molecular markers [24]. Regardless of the question, a molecular marker must fundamentally be selectively neutral and follow Mendelian inheritance in order to be used as a tool for detecting demographic patterns and microevolutionary forces as genetic drift and gene flow [33].

Recombination, selection, and genetic drift affect different genes and regions of the genome in a different way. Consequently, multiple samples of the genome by combining the results from many loci provide a precise and statistically powerful way of comparing populations and individuals [33]. Microsatellites have high mutation rates that generate the high levels of allelic diversity necessary for genetic studies of processes acting on ecological time scales [67, 68]. The new approaches use more of the information in a data set than the summary statistics of traditional approaches (e.g., F_{ST}).

Nowadays, demography and history of populations and relationships of individuals can be described in a detailed manner because the typical data set contains high number of individuals sampled at many loci. Hence, genetic tools allow to address many basic ecological questions for the first time or in new ways [33].

Genetic diversity is essential for the long-term survival of species; without it, species cannot adapt to environmental changes and are more susceptible to extinction. Measuring levels of genetic variation within and among populations is an important first step in evaluating the evolutionary biology and tree improvement potential of a species [1]. Most forest tree species possess considerable genetic variation, much of which can be found within populations, and the expected heterozygosity is approximately 50% higher in a population of forest trees than average heterozygosity expected in populations of annuals and perennials cycle short life species [1].

A number of factors that contribute to the high levels of genetic diversity typically found in forest tree populations are large population size, longevity, high levels of outcrossing, strong gene flow by pollen and seed between populations, and balancing selection [68]. Nuclear DNA is often highly variable, and it is biparentally inherited. Efficient gene flow, in particular, via pollen is the main factor contributing to the high diversity within populations of trees but low differentiation among spatially separated populations [69].

Gene diversity index is the expected heterozygosity averaged over all loci sampled and is the most widely used measure of genetic variation employing genetic markers. Because low-

frequency alleles contribute very little to h, it is relatively insensitive to sample size. However, when sampled populations show markedly differences in size, additionally allelic richness (R) could be informed. The observation of variation at nonrecombining chloroplast DNA (cpDNA) is of particular importance for plants. Low mutation rates of cpDNA are responsible for, in general, low variation within species [69]. As a consequence of previous statements, higher levels of genetic diversity with nuSSRs than with cpSSRs are expected.

Populations of forest trees often differ in allele frequencies (especially when they are separated geographically), and it is often of interest to determine the degree to which genetic variation in a region is distributed within and among populations. This information is useful for understanding the degree to which gene flow by pollen and seed counters population subdivision due to selection or genetic drift [70]. It also has practical value when planning seed collections for breeding or gene conservation purposes. Therefore, knowledge of natural patterns of genetic variation and their evolutionary bases also are of great practical significance [1]. The pattern of genotypic variation (heterozygosity vs. homozygosity) among individuals within a subpopulation is highly dependent upon the mating system, whereas the distribution of allelic variation within and among subpopulations is influenced by both gene flow and genetic drift. Because of the opposite effects of gene flow and genetic drift, the balance between them is a primary determinant of the genetic population structure of a species [42]. Total diversity in forest trees is also generally higher than that found in other plants. However, only a small proportion of the total gene diversity in trees is due to differences among populations [1].

Woody species contain more variation within populations but have less variation among them than species with other life forms. Woody species that present large geographic ranges, outcrossing breeding systems, and wind or animal-ingested seed dispersal are more genetically diverse than woody species with other combinations of traits [68]. Hence, from AMOVA results in forest tree species, higher levels of genetic variation is expected in the hierarchical level within populations than the hierarchical level among populations. Genetic differentiation among populations estimated by F_{ST} also varies widely among tree species, ranging from low values in species that have more or less continuous distributions to high values in species with disjunct population distributions [1]. For the interpretation of F_{ST}, the value scale suggested by Wright [71] is a useful tool. The four values are (1) 0–0.05 indicate little genetic differentiation, (2) 0.05–0.15 indicate moderate genetic differentiation, (3) 0.15–0.25 indicate great genetic differentiation and, (4) values above 0.25 indicate very great genetic differentiation. Nuclear F_{ST} in trees is frequently 10% or lower, which is only one-half to one-quarter of the F_{ST} estimates typically found in annuals or other herbaceous species. The lower F_{ST} in trees is most likely because most tree species are outcrossing while a large proportion of annuals and herbaceous plants are either self-pollinated or self-pollination features prominently in their mating system. High levels of self-pollination not only promote inbreeding but also limit pollen gene flow between populations. Both pollen and seed gene flow between populations of forest trees can be extensive [1].

Patterns of seed dispersal shape the composition and genetic structure of plant populations. Species with low levels of gene flow by seeds have high probability to show genetic hetero-

geneity among subpopulations, whereas species with high levels of gene flow by seeds have low levels of genetic structure [72].

Compared to biparentally inherited and paternally inherited markers, maternally inherited markers detected strong genetic differentiation between populations [69] because normally seeds are distributed to shorter distances than pollen [73]. Moreover, being a haploid genome, effective population size for hermaphrodite outcrossing plants is half that of the corresponding diploid nuclear genome [16]. Hence, gene flow between populations of small size has a lower effective to counteract the effects of genetic drift in loci transmitted maternally [74]. When these occur, F_{ST} values for chloroplast DNA can be markedly higher than those for nuclear genes [75]. Genetic structure of chloroplast genetic variation is also affected by the interaction of seed dispersal with other ecological and genetic processes. Deposition patterns of seeds, pollen dispersal, density of adults, microhabitat selection, and several aspects of the ecology of the species could have significant effects on patterns of genetic variation within species [72]. While both pollen and seed dispersal determines gene flow in plants, seed dispersal is most important because it allows species to colonize habitats and therefore influences the dynamics of populations [76].

Methods based on distance allow grouping individuals according to genetic distance while its graphical representation allows to relate these groups with other information, e.g., the geographical origin of individuals or phenotypic traits [44]. Even though these methods are statistically weak, they still represent a first approach to analyze population genetic structure. Conversely, grouping from methods based on models are statistically powerful and allow to determine the number of clusters using genetic information assigning individuals probabilistically to these clusters even when they require model assumptions (e.g., Hardy-Weinberg equilibrium within populations and complete linkage equilibrium between loci within populations) [44].

Species can become subdivided into genetically distinct subpopulations when gene flow is restricted, leading to variation in the frequency of a gene over space [42]. The number of migrants $N_e m$ represents an estimation of gene flow, and it is important to keep in mind that m is defined in the terms of gene pools, and therefore m represents the amount of exchange of gametes between subpopulations and not necessarily individuals [42]. Most trees species are wind pollinated, and the pollen can be blown from hundreds of miles by the wind. Hence, tree populations that are quite distant can still experience gene flow. Because gene flow requires both movement and reproduction, m is not just the amount of dispersal of individuals between subpopulations but instead m represents a complex interaction between the pattern of dispersal and the mating system [42]. As a consequence of this, in forest tree species, it is really important to know pollen and seed dispersal mechanisms and species matting system to interpret the estimated levels of gene flow from microsatellite data.

The effects of gene flow on genetic variation among and within subpopulations can be summarized as gene flow decreases genetic variation among subpopulations and increases genetic variation within subpopulations. Genetic drift causes an increase in genetic variation among subpopulations and decreases genetic variation within a subpopulation. Hence, the

effects of gene flow on genetic variation within and among subpopulations are the opposite of those of genetic drift [42].

For neutral alleles, when gene flow is interrupted, genetic drift is more effective than mutation to produce genetic differentiation among subpopulations [77]. Thereby, gene flow could be a force that maintains species integrated as well as influences the ecologic processes, e.g., determine the persistence and adaptation of local populations, determine pattern of distribution of species, etc. [78]. In this way, studies of gene flow become relevant for the interpretation of microevolutionary patterns and genetic structure of populations [80]. Even a small amount of gene flow can cause two populations to behave effectively as a single evolutionary lineage. One "effective" migrant per generation ($N_em = 1$) defines an inflection point from the relation between F_{ST} and N_em, with increasing effective number of migrants F_{ST} declines only very slowly when $N_em \geq 1$ but with decreasing effective number of migrants F_{ST} rises very rapidly when $N_em \leq 1$. As a consequence of this, $N_em = 1$ marks a transition in the relative evolutionary importance of gene flow to drift. It is impressive that only one or more effective migrants per generation on average are needed to cause gene flow to dominate over genetic drift, leading to great genetic homogeneity among subpopulations [42].

Ennos [31] demonstrated that estimation of the relative rates of pollen and seed migration among plant populations is possible from a simple comparison of F_{ST} values for nuclear and maternally inherited organelle genetic markers. Estimated rates of pollen migration are greater than rates of seed migration for all six species investigated by Ennos [31]; however, differences among species were substantial. The greatest contrast between pollen and seed migration rates was found for oak species, where interpopulation pollen flow is estimated to be 200 times greater than interpopulation seed flow. This result was interpreted by the species reproductive system. Oaks show high rates of interpopulation pollen dispersal because they are outbreeding, wind-pollinated, and disperse pollen from a substantial height. Also, dispersal of acorns by birds and rodents is likely to be restricted [31, 79]. In contrast, lower differences between pollen and seed migration rates were found for wild barley. Gene flow by pollen is estimated to be only four times greater than interpopulation gene flow by seeds. Opportunities for interpopulation pollen dispersal in such a highly self-pollinating species are expected to be rare, and it is not surprising that pollen and seed flow should be of the same order of magnitude for this species [31, 81]. Forest trees species are generally outbreeding, whereby levels of pollen flow versus seed flow (*r*) may vary according to the potential distances of dispersal related to mechanism of pollen and seed dispersal.

By itself, the mating system does not alter allele frequencies but does affect the relative proportions of different genotypes in populations, which under some circumstances deeply influences the viability and vigor of offspring [1]. Inbreeding coefficient (F_{IS}) measures the fractional reduction in heterozygosity relative to a random mating population with the same allele frequencies. Even though genotypic frequencies in natural populations of forest trees often approximate those expected under random mating, mating systems that depart from random mating do occur and have important implications. Individuals of most temperate forest trees are bisexual and have the capacity for self-fertilization. In addition, nearby trees may be related (e.g., siblings originating from seeds of the same mother tree), providing

opportunity for mating between relatives. Therefore, forest trees typically have mixed mating systems, whereby many and perhaps most mates are paired essentially at random.

There is also some mating between genetically related individuals, which occurs more often than expected from random pairings [1]. Inbreeding is of great significance to the genetic makeup of both natural populations and breeding populations of forest trees because it has two major consequences: (1) in comparison to random mating, inbreeding increases the frequency of homozygous offspring at the expense of heterozygotes and (2) mating between close relatives is usually detrimental to the survival and growth of offspring, called inbreeding depression. Therefore, the magnitude of inbreeding among parent trees used to produce seed for reforestation, such as in seed production areas or seed orchards, is of great practical concern [1].

In previous section, we presented three ways to estimate inbreeding coefficient that has different statistical power and assumptions: (1) F_{IS} estimated from AMOVA could be used as first measure of inbreeding for a determinate hierarchical structure, (2) F_{IS} estimated by a Bayesian approximation is a measure statistically more powerful to determine the level of inbreeding in a population, and (3) F_{IS} could be estimated considering null alleles when certain levels of null alleles were determined in the microsatellite loci considered in order to determine the proportion of homozygote genotypes consequence of inbreeding than homozygotes caused by null alleles.

The current distribution and population structure and potential fate in the future are better understood from the knowledge of historical distribution, postglacial phylogeography, and evolution of a species [82]. Regarding to the assessment of demographic history using the F_S neutrality test for population, F_S statistic takes a large negative value within a population affected by expansion due to an excess of rare haplotypes (recent mutations). Significance of the test must to be calculated with data bootstraps. A F_S statistic with $p(F_S) < 0.02$ ($\alpha = 0.05$, due to a particular behavior of this statistics, [56]) is considered as an evidence of population expansion.

Whereas using D_{Tajima} neutrality test, a D_{Tajima} statistic is expected to be close to zero in a population of constant size while statistically significant negative values indicate a sudden expansion of population size and positive values indicate population subdivision or recent bottlenecks. The statistical significance of D_{Tajima} is tested generating random samples using a coalescent simulation algorithm under the hypothesis of population balance. The p-value for D_{Tajima} is obtained by the ratio of random D_{Tajima} less than or equal to the observed D_{Tajima}. Computer-intensive statistical methods have been developed to extract as much information from the data as possible and to provide a flexible framework within which complex models of population history can be handled [83].

Approximate Bayesian computation is a computer-intensive method that has wide applicability, where populations diverge genetically through time, influenced by random genetic drift and migration, ABC uses summary statistics measured from microsatellite loci to make inferences about demographic parameters in different population models. The method can be

used to infer effective sizes of current and ancestral populations, immigration rates, splitting times, and tree topology [83].

As a final recommendation, researchers must define which is/are the problem/s and question/s that they would resolve with their study before starting a study with molecular marker in a native forest tree species. This is a founder requisite to determine sampling design, to decide molecular markers to use (keeping in mind the information required and laboratory work to obtain molecular data), and appropriate statistical analysis to obtain the required information. Of course, the researchers must pay attention to biological features of the studied species at the moment to design the study and back to these features at the moment to interpret the results of statistical analyses in a biological context.

7. Conclusion

This chapter helps to researchers that are not familiarizing with statistical methods and population genetics theories to analyze nuclear and chloroplast microsatellite data. Methods allow quantification of genetic variation and genetic structure in native forest species while theories allow knowledge about the past and the present genetic states of populations for making inferences about the future of these populations.

Acknowledgements

Author details

Maria Eugenia Barrandeguy[1,2,3*] and Maria Victoria Garcia[1,2,3]

*Address all correspondence to: ebarran@fceqyn.unam.edu.ar

1 Departamento de Genética, Facultad de Ciencias Exactas, Químicas y Naturales, Universidad Nacional de Misiones, Argentina

2 Instituto de Biología Subtropical Nodo Posadas (UNaM, CONICET) Posadas, Argentina

3 Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

References

[1] White TW, Adams WT, Neale DB. Forest Genetics. 1st ed. Cambridge, USA: CAB International Publishing; 2007;702 p. DOI: 10.1079/9781845932855.0000

[2] Finkeldey R, Leinemann L, Gailing O. Molecular genetic tools to infer the origin of forest plants and wood. Applied Microbiology Biotechnology. 2010;85:1251–1258. DOI: 10.1007/s00253-009-2328-6

[3] Finkeldey R, Ziehe M. Genetic implications of silvicultural regimes. Forest Ecology Management. 2004;197:231–244. DOI: 10.1016/j.foreco.2004.05.036

[4] Krutovsky KV, Neale DB. Nucleotide diversity and linkage disequilibrium in cold-hardiness- and wood quality-related candidate genes in Douglas fir. Genetics. 2005;171:2029–2041. DOI: 10.1534/genetics.105.044420

[5] Rossetto M. Sourcing SSR markers from related plant species. In: Henry EJ, editor. Plant Genotyping: The DNA Fingerprinting of Plants. 1st ed. Oxon UK: CAB Plant International; 2001;pp. 211–224. DOI: 10.1079/9780851995151.0000

[6] Barbará T, Palma-Silva C, Paggi GM, Bered F, Fay MF, Lexer C. Cross-species transfer of nuclear microsatellite markers: potential and limitations. Molecular Ecology. 2007;16:3759–3767. DOI: 10.1111/j.1365-294X.2007.03439.x

[7] Streiff R, Labbe T, Bacilieri R, Steinkellner H, Glossl J, Kremer A. Within-population genetic structure in *Quercus robur* L. and *Quercus petrea* (Matt.) Liebl. assessed with isozymes and microsatellites. Molecular Ecology. 1998;7:317–328. DOI: 10.1046/j.1365-294X.1998.00360.x

[8] Degen B, Streiff R, Ziegenhagen B. Comparative study of genetic variation and differentiation of two pedunculate oak (*Quercus robur*) stands using microsatellite and allozyme loci. Heredity. 1999;83:597–603. DOI: 10.1038/sj.hdy.6886220

[9] Fernández JF, Sork VL, Gallego G, López J, Bohorques A, Tohme J. Cross-amplification of microsatellite loci in a neotropical *Quercus* species and standardization of DNA extraction from mature leaves dried in silica gel. Plant Molecular Biology Reporter. 2000;18:397a–397e. DOI: 10.1007/BF02825070

[10] Mottura MC, Finkeldey R, Verga AR, Gailing O. Development and characterization of microsatellite markers for *Prosopis chilensis* and *Prosopis flexuosa* and cross-species amplification. Molecular Ecology Notes. 2005;5:487–489. DOI: 10.1111/j.1471-8286.2005.00965.x

[11] Yasodha R, Ghosh M, Sumathi R, Gurumurthi K. Cross-species amplification of eucalyptus SSR markers in Casuarinaceae. Acta Botanica Croatia. 2005;64:115–120.

[12] Moreira PA, Souza SAS, Oliveira EA, Araújo NH, Fernandes GW, Oliviera DA. Charaterization of nine tranferred SSR markers in the tropical tree species *Enterolobium*

contortisiliquum (Fabaceae). Genetics and Molecular Research. 2012;11:2338–2342. DOI: 10.4238/2012

[13] Dayanandan S, Bawa KS, Kesseli R. Conservation of microsatellites among tropical trees (Leguminosae). American Journal of Botany. 1997;84:1658–1663.

[14] Moreno AC, Marchelli P, Vendramin GG, Gallo LA. Cross transferability of SSRs to five species of Araucariaceae: a useful tool for population genetic studies in *Araucaria araucana*. Forest Systems. 2011;20:303–314.

[15] Liu J, Gao LM, Li DZ, Zhang DQ, Möller M. Cross-species amplification and development of new microsatellite loci for *Taxus wallichiana* (Taxaceae). American Journal of Botany. 2011;98:e70–e73. DOI: 10.3732/ajb.1000445

[16] Provan J, Powell W, Hollingsworth PM. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends in Ecology and Evolution. 2001;16:142–147. DOI: 10.1016/S0169-5347(00)02097-8

[17] Weising K, Gardner R. A set of conserved PCR primers for the analysis of simple sequence repeat polymorphism in chloroplast genomes of dicotyledonous. Genome. 1999;42:9–19. DOI: 10.1139/gen-42-1-9

[18] Powell W, Machray GC, Provan J. Polymorphism revealed by simple sequence repeats. Trends in Plant Science. 1996;1:209–245. DOI: 10.1016/1360-1385(96)86898-1

[19] Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M, Powell W. How much effort is required to isolate nuclear microsatellites from plants? Molecular Ecology. 2003;12:1339–1348. DOI: 10.1046/j.1365-294X.2003.01825.x

[20] Barrandeguy ME, Prinz K, García MV, Finkeldey R. Development of microsatellite markers for *Anadenanthera colubrina* var. *cebil* (Fabaceae), a native tree from South America. American Journal of Botany. 2012;99:e372–e374. DOI: 10.3732/ajb.1200078

[21] Fisher D, Bachmann K. Microsatellite enrichment in organisms with large genomes (*Allium cepa* L.). Biotechniques. 1998;24:796–800.

[22] Prinz K, Schie S, Debener T, Hensen I, Weising K. Microsatellite markers for *Spergularia media* (L.) C. Presl. (Caryophyllaceae) and their cross-species transferability. Molecular Ecology Notes. 2009;9:1424–1426. DOI: 10.1111/j.1755-0998.2009.02680.x

[23] Zalapa JE, Cuevas H, Zhu H, Steffan S, Senalik D, Zeldin E, Mccown B, Harbut R, Simon P. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. American Journal of Botany. 2012;99:93–208. DOI: 10.3732/ajb.1100394

[24] Putman AI, Carbone I. Challenges in analysis and interpretation of microsatellite data for population genetic studies. Ecology and Evolution. 2014;4:4399–4428. DOI: 10.1002/ece3.1305

[25] Peakall R, Smouse PE. Genalex 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics. 2012;28:2537–2539. DOI: 10.1093/bioinformatics/bts460

[26] Raymond M, Rousset F. GENEPOP (version 1.2): population genetics software for exact test and ecumenicism. Journal of Heredity. 1995;86:248–249.

[27] Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes. 2004;4:535–538. DOI: 10.1111/j.1471-8286.2004.00684.x

[28] Avise JC. Molecular Markers, Natural History and Evolution. 2nd ed. Sunderland, Massachusetts: Sinauer Associates; 2004;684 p. DOI: 10.1007/978-1-4615-2381-9

[29] Linhart YB, Grant MC. Evolutionary significance of local genetic differentiation in plants. Annual Review of Ecology and Systematics. 1996;27:237–277. DOI: 10.1146/annurev.ecolsys.27.1.237

[30] Sork VL, Nason J, Campbell DR, Fernandez JF. Landscape approaches to historical and contemporary gene flow in plants. Trends in Ecology and Evolution. 1999;14:219–224. DOI: 10.1016/S0169-5347(98)01585-7

[31] Ennos RA. Estimating the relative rates of pollen and seed migration among plant populations. Heredity. 1994;72:250–259. DOI: 10.1038/hdy.1994.35

[32] Barrandeguy ME, García MV, Prinz K, Rivera Pomar R, Finkeldey R. Genetic structure of disjunct Argentinean populations of the subtropical tree *Anadenanthera colubrina* var. *cebil* (Fabaceae). Plant Systematic and Evolution. 2014;300:1693–1705. DOI: 10.1007/s00606-014-0995-y

[33] Selkoe KA, Toonen RJ. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology Letters. 2006;9:615–629. DOI: 10.1111/j.1461-0248.2006.00889.x

[34] Kimura M, Crow JF. The number of alleles that can be maintained in a finite population. Genetics. 1964;49:725–738.

[35] Barton NH, Slatkin M. A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. Heredity. 1986;56:409–415. DOI: 10.1038/hdy.1986.63

[36] Foulley J, Ollivier L. Estimating allelic richness and its diversity. Livestock Science. 2006;101:150–158. DOI: 10.1371/journal.pone.0115203

[37] Nei M. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences. 1973;70:3321–3323.

[38] Nagylaki T. Fixation indices in subdivided populations. Genetics. 1998;148:1325–1332.

[39] Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics. 1978;89:583–590.

[40] Nei M. Molecular Evolutionary Genetics. New York: Columbia University Press; 1987;512 p.

[41] Templeton AR. Population Genetics and Microevolutionary Theory. New Jersey: Wiley-Liss Publication; 2006;720 p. DOI: 10.1002/0470047356

[42] Bandelt HJ, Forster P, Rohl A. Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution. 1999;16:37–48. DOI: 10.1093/oxford-journals.molbev.a026036

[43] Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multi-locus genotype data. Genetics. 2000;155:945–959.

[44] Corander J, Sirén J, Arjas E. Bayesian spatial modelling of genetic population structure. Computational Statistics. 2008;23:111–129. DOI: 10.1007/s00180-007-0072-x

[45] Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology. 2005;14:2611–2620. DOI: 10.1111/j.1365-294X.2005.02553.x

[46] Excoffier L, Smouse P, Quattro J. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 1992;131:479–491.

[47] Wright S. The genetical structure of populations. Annals of Eugenics. 1951;15:323–354.

[48] Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984;38:1358–1370.

[49] Freeland JR. Molecular Ecology. 1st ed. England: Wiley-Liss; 2005;388 p.

[50] Slatkin M. Gene flow in natural populations. Annual Review of Ecology and Systematics. 1985;16:393–430.

[51] Holsinger KE, Lewis PO. University of Connecticut, Storrs, Connecticut USA. Hickory: a package for analysis of population genetic data version 1.1. Department of Ecology and Evolutionary Biology [Internet]. 2003. Available from: http://darwin.eeb.uconn.edu/hickory/hickory.html [Accessed: 2013-07-13].

[52] Chybicki IJ, Burczyk J. Simultaneous estimation of null alleles and inbreeding coefficients. Heredity. 2009;100:106–113. DOI: 10.1093/jhered/esn088

[53] Navascués M, Vaxevanidou Z, González-Martínez SC, Climent J, Gil L, Emerson BC. Chloroplast microsatellites reveal colonization and metapopulation dynamics in the

Canary Island pine. Molecular Ecology. 2006;15:2691–2698. DOI: 10.1111/j.1365-294X. 2006.02960.x

[54] Emerson BC, Paradis E, Thebaud C. Revealing the demographic histories of species using DNA sequences. Trends in Ecology and Evolution. 2001;16:707–716. DOI: 10.1016/S0169-5347(01)02305-9

[55] Fu YX. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics. 1997;147:915–925.

[56] Pybus OG, Shapiro B. Natural selection and adaptation of molecular sequences. In: Lemey P, Salemi M, Vandamme A, editors. The Phylogenetic Handbook. 2nd ed. Cambridge: Cambridge University Press; 2009;pp. 407–418. DOI: 10.1017/CBO9780511819049.015

[57] Beaumont MA, Nielsen R, Robert C, Hey J, Gaggiotti O, Knowles L, Estoup A, Panchal M, Corander J, Hickerson M, Sisson SA, Fagundes N, Chikhi L, Beerli P, Vitalis R, Cornuet JM, Huelsenbeck J, Foll M, Yang Z, Rousset F, Balding D, Excoffier L. In defence of model-based inference in phylogeography. Molecular Ecology. 2010;19:436–446. DOI: 10.1111/j.1365-294X.2009.04515.x

[58] Excoffier L, Lischer H. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology. 2010;10:564–567. DOI: 10.1111/j.1755-0998.2010.02847.x

[59] Szpiech ZA, Jakobsson M, Rosenberg NA. ADZE: a rarefaction approach for counting alleles private to combinations of populations. Bioinformatics. 2008;24:2498–2504. DOI: 10.1093/bioinformatics/btn478

[60] Goudet J. FSTAT (version 2.9.3.2): a computer program to calculate F statistics. Heredity. 1995;86:485–486.

[61] Perrier X, Jacquemoud-Collet JP. DARwin software [Internet]. 2006. Available from: http://darwin.cirad.fr/ [Accessed: 2016-04-28].

[62] Leigh J, Bryant D. Popart: full-feature software for haplotype network construction. Methods in Ecology and Evolution. 2015;6:1110–1116. DOI: 10.1111/2041-210X. 12410

[63] Forster JW, Jones ES, Kolliker R, Drayton CM, Dumsday JL, Dupal MP, Guthridge KM, Mahoney NL, Van Zijll de Jong E, Smith KF. Development and implementation of molecular markers for forage crop improvement. In: Spangenberg G, editor. Molecular Breeding of Forage Crops. Netherlands, Dordrecht: Springer; 2001;pp. 101–133. DOI: 10.1007/978-94-015-9700-5_6

[64] Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing structure output and implementing the Evanno method. Conservation Genetics Resources. 2011;4:359–361. DOI: 10.1007/s12686-011-9548-7

[65] Cornuet J, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R, Marin JM, Estoup A. DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. Bioinformatics. 2014;30:1187–1189. DOI: 10.1093/bioinformatics/btt763

[66] Schlötterer C. Evolutionary dynamics of microsatellite DNA. Chromosoma. 2000;109:365–371.

[67] Hamrick JL, Godt MJW, Sherman-Broyles SL. Factors influencing levels of genetic diversity in woody plant species. New Forests. 1992;6:95–124. DOI: 10.1007/BF00120641

[68] Finkeldey R, Hattemer HH. Tropical Forest Genetics. 1st ed. Heidelberg: Springer; 2007;316 p. DOI: 10.1007/978-3-540-37398-8

[69] Ouborg NJ, Piquot Y, Groenendael MV. Population genetics, molecular markers and the study of dispersal in plants. Journal of Ecology. 1999;87:551–668.

[70] Wright S. Evolution and the Genetics of Population, Variability Within and Among Natural Populations. Chicago: The University of Chicago Press; 1978;590 p.

[71] Hamrick JL, Murawski DA, Nason JD. The influence of seed dispersal mechanisms on the genetic structure of tropical tree populations. Vegetatio. 1993;107:281–297.

[72] Zhan QQ, Wang JF, Gong X, Peng H. Patterns of chloroplast DNA variation in *Cycas debaoensis* (Cycadaceae): conservation implications. Conservation Genetics. 2011;12:959–970. DOI: 10.1007/s10592-011-0198-9

[73] Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG. Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. Molecular Ecology. 2005;14:689–701. DOI: 10.1111/j.1365-294X.2004.02410.x

[74] McCauley DE. The use of chloroplast DNA polymorphism in studies of gene flow in plants. Trends in Ecology and Evolution. 1995;10:198–202.

[75] Ndiade-Bourobou D, Hardy OJ, Favreau B, Moussavou H, Nzengue E, Mignots A, Bouvet JM. Long distance seed and pollen dispersal inferred from spatial genetic structure in the very low-density rainforest tree, *Baillonella toxisperma* Pierre, in Central Africa. Molecular Ecology. 2010;19:4949–4962. DOI: 10.1111/j.1365-294X.2010.04864.x

[76] Slatkin M, Barton NH. A comparison of three indirect methods for estimating average levels of gene flow. Evolution. 1989;43:1349–1368.

[77] Planter EA. Flujo Génico: métodos para estimarlo y marcadores moleculares. In: Eguiarte L, Souza V, Aguirre X, editors. Ecología Molecular. 1st ed. México: INECC; 2007;pp. 49–61.

[78] Bossart DL, Prowell DP. Genetic estimates of population structure and gene flow: limitations, lessons and new directions. Trends in Ecology and Evolution. 1998;13:202–206.

[79] Sork VL. Examination of seed dispersal and survival in red oak, *Quercus Rubra* (Fagaceae), using metal tagged acorns. Ecology. 1984;65:1020–1022.

[80] Brown AHD, Zohary D, Nevo E. Outcrossing rates and heterozygosity in natural populations of *Hordeum spontaneum* Koch in Israel. Heredity. 1978;41:49–62.

[81] Zinck JWR, Rajora OP. Post-glacial phylogeography and evolution of a wide-ranging highly-exploited keystone forest tree, eastern white pine (*Pinus Strobus*) in North America: single refugium, multiple routes. BMC Evolutionary Biology. 2016;16:56. DOI: 10.1186/s12862-016-0624-1

[82] Zinck JWR, Rajora OP. Post-glacial phylogeography and evolution of a wide-ranging highly-exploited keystone forest tree, eastern white pine (Pinus strobus) in North America: single refugium, multiple routes. BMC Evolutionary Biology. 2016; 16 (1): 56. DOI: 10.1186/s12862-016-0624-1.

[83] Beaumont M. Joint determination of topology, divergence time and immigration in population trees. In: Matsumura S, Forster P, Renfrew C, editors. Simulations, Genetics and Human Prehistory. Cambridge: McDonald Institute for Archaeological Research; 2008. p. 134–154.