

Using computer simulations to assess sampling effects on spatial genetic structure in forest tree species

Yessica Rico¹

Received: 14 July 2016 / Accepted: 31 January 2017 / Published online: 4 February 2017
© Springer Science+Business Media Dordrecht 2017

Abstract Quantifying spatial genetic structure is key to inform forest management and restoration strategies. Reliable evaluations of genetic structure require sound sampling schemes because inappropriate sampling may over- and under-estimate spatial patterns of genetic structure. Sampling bias has been investigated through computer simulations mostly for animal species with continuous distributions. For tree species that have different life history traits, results from such studies may not apply. Here, I used spatially explicit landscape genetic simulations to assess the effects of spatial sampling scheme (random, systematic, and cluster), sampling intensity (35, 50, 65, and 80%), and the number of microsatellite loci (8, 14, and 20) on inferences of genetic structure under isolation by distance (IBD) in two forest tree species with varying dispersal distances and patchy distributions. Results showed that random sampling with 20 loci was the best performing sampling scheme, irrespective of sampling intensity and the strength of IBD. In contrast, the cluster and systematic sampling were sensitive to sample size. For the three sampling schemes, the number of loci had a large effect because with 8 loci there was an increasing chance of underestimating IBD. Increasing the number of samples over the number of loci, did not improve the performance of sampling schemes. Hence, researchers should put more effort on increasing the number of loci over increasing sample size. Results also showed that sampling error rates varied between species, and sampling bias appeared stronger for the species with a more aggregated spatial distribution.

Keywords Microsatellites · CDPOP · Isolation by distance · Random sampling · Systematic sampling · Cluster sampling · Sample size

✉ Yessica Rico
yessica.ricom@gmail.com

¹ CONACYT, Instituto de Ecología, A.C., Centro Regional del Bajío, Avenida Lázaro Cárdenas, 61600 Pátzcuaro, Michoacán, Mexico

Introduction

Forest management strategies accounting for spatial patterns of genetic structure are critical for delimiting conservation management units, germ plasm collection, sources and target sites for reforestation, and assisted migrations (Broadhurst et al. 2008; Frankham 2010). Moreover, spatial genetic structure provides insights for understanding relevant ecological and evolutionary processes, such as demography, progeny fitness, colonization history (Kalisz et al. 2001; Zhao et al. 2009), as well as potential for local adaptation in the light of environmental change (Alberto et al. 2013; Savolainen et al. 2013). Efforts to protect biodiversity thus should include the quantification of spatial patterns of genetic structure across species' ranges.

Reliable evaluations of genetic structure require sound sampling schemes that involve matching the scale of the spatial process driving genetic structure, which is determined by the life history traits of the study species (Anderson et al. 2010; Keller et al. 2013). Although conducting an extensive sampling is recommended to avoid biased results from genetic analyses, in real situations this condition is often difficult to achieve (Hall and Beissinger 2014). Sampling can be constrained by varying factors, such as limited economic resources for processing a small number of individuals for genetic analysis, permitted only in few sites or within geographically accessible areas.

Plants due to their sessile life form are expected to develop strong genetic structure over distance (mostly due to seed dispersal), where neighboring individuals are more closely related than distant individuals, resulting in a genetic pattern of isolation by geographical distance (IBD) (Ennos 2001; Hampe and Petit 2005; Epperson 2007). IBD is a prevalent driver of genetic structure in tree populations (Vekemans and Hardy 2004), and is likely the most common spatial process investigated in landscape genetic studies (Jenkins et al. 2010). Inappropriate sampling is recognized to mislead inferences of genetic structure (Latch and Rhodes 2006; Koen et al. 2013; Naujokaitis-Lewis et al. 2013; Prunier et al. 2013; Hoban and Schlarbaum 2014; Tucker et al. 2014). For instance, typical analyses in population genetics, such as estimates of genetic differentiation F_{ST} (Schwartz and McKelvey 2009; Landguth and Schwartz 2014), genetic clustering algorithms (Schwartz and McKelvey 2009), and correlative distance methods (Landguth and Schwartz 2014; Oyler-McCance et al. 2013) are sensitive to sampling bias under IBD, which may over- and under-estimate genetic structure (Meirmans 2012).

Sampling effects on inferences of spatial genetic structure have been explored using computer simulations (e.g., Schwartz and McKelvey 2009; Landguth et al. 2012; Oyler-McCance et al. 2013; Landguth and Schwartz 2014). From simulation studies, it has been suggested that random or systematic sampling (e.g., uniform sampling within transects) perform well for identifying the underlying spatial process (Schwartz and McKelvey 2009; Oyler-McCance et al. 2013). Other factors, such as the number of loci and alleles per locus are also important (Landguth et al. 2012). Sampling issues in simulation studies have been developed on scenarios that emulate animals with high dispersal abilities and mainly with continuous distributions (but see Prunier et al. 2013). For plants that have different life history traits, results from such studies may not apply. Moreover, recent evidence from landscape genetic simulations suggest that IBD is not only determined by dispersal distances *per se*, but also by habitat configuration and the spatial locations of populations (van Strien et al. 2014). This implies that sampling bias on inferences of spatial genetic structure under IBD, may be far more complex than anticipated for species with non-continuous distributions, such as in many tree species.

The aim of this study was to assess the effects of sampling design on inferences of spatial genetic structure in forest tree species with non-continuous distributions. I used spatially explicit, individual-based computer simulations to evaluate the sensitivity of sampling design through varying three factors: the spatial sampling scheme (i.e. spatial arrangement of samples), the sampling intensity (i.e. number of sites), and the number of microsatellite loci. I compared the performance of each sampling design combination to quantify spatial genetic differentiation under varying degrees of isolation by distance. Specifically, I asked: (1) which is the best performing spatial sampling scheme? (2) Does performance of spatial sampling largely depend on the number of sites sampled and the number of loci? (3) How does performance of sampling design vary with the strength of spatial genetic structure (i.e., IBD)? Through simulating and contrasting two species with different characteristics, this study also illustrates how bias in sampling design may depend on the study species.

Methods

Model system

To illustrate the effects of sampling design on inferences of spatial genetic structure under approximate real-like scenarios, I selected two tropical dry forest trees varying in spatial distributions, individual abundances, and mating systems. The selected species are characteristic of the tropical dry forest of the Bajío region in Central Mexico, utilized for timber and valuable for reforestation: *Cedrela dugesii* and *Bursera palmeri*.

Cedrela dugesii is a monoicous deciduous tree endemic to the Bajío region, where the species is commonly associated and abundant on rocky outcrops (Calderón de Rzedowski and Germán 1993). It has seeds adapted to dispersal by wind and is pollinated by insects (Calderón de Rzedowski and Germán 1993). *B. palmeri* is a dioicous deciduous tree (sometimes shrub) of the tropical dry forest and scrublands of the Mexican Plateau. It has a larger distribution in the Bajío region relative to *C. dugesii* (Fig. 1). *B. palmeri* has seeds dispersed by animals and is pollinated by insects (Rzedowski and Guevara-Féfer 1992).

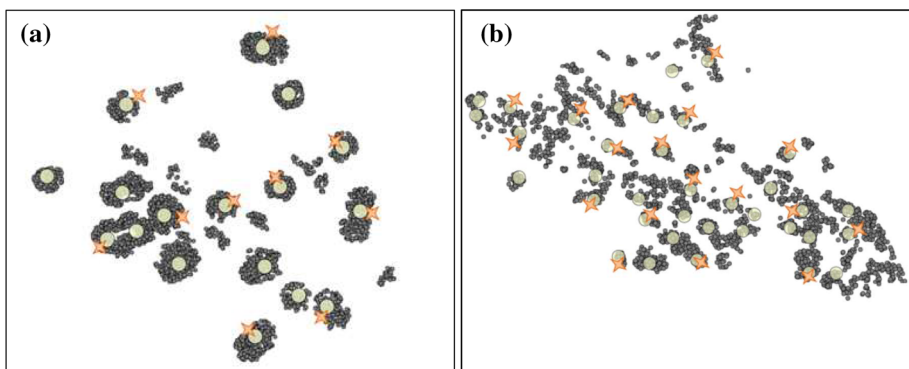


Fig. 1 Sample design under isolation by distance in two simulated forest tree species with non-continuous distributions: **a** *Cedrela dugesii* and **b** *Bursera palmeri*. The circles denote the 10 km² radius sampling locations for the systematic sampling. The stars denote the sampling sites included for the cluster sampling scheme. For the random sampling, individuals were collected randomly across the landscape

The tropical dry forest in the Bajío region covered large areas, but currently it is decreasing due to the intensification of agriculture, extensive livestock, and urbanization (Encino-Ruiz et al. 2014; Rzedowski et al. 2014). Both tree species are in decline and increasingly fragmented due to logging for timber and wood products (Rzedowski and Guevara-Féfer 1992; Calderón de Rzedowski and Germán 1993).

Landscape genetic simulations

Simulations were implemented in CDPOP v1.2.31 (Landguth and Cushman 2010), which is an individual-based landscape genetics program. CDPOP simulates gene flow under Mendelian inheritance among spatially referenced genotypes based on probabilistic functions of individual movement and mating across multiple generations (for more information on the software see Landguth and Cushman 2010 and the github site <https://github.com/ComputationalEcologyLab/CDPOP>). To map the spatial distribution for each tree species in the Bajío region, I accessed all collection records from specialized flora web-databases congregated by the National Commission for Knowledge and Use of Biodiversity of Mexico (REMIB-CONABIO 2016; <http://www.conabio.gob.mx/remib/cgi-bin/>). Records were also obtained from the herbarium IEB (Instituto de Ecología, A. C., Centro Regional del Bajío), which is specialized in the flora of the Bajío region.

Within an area of approximately 3.5×2.9 km that corresponded to the main area of species distribution in the Bajío region, 5000 individuals of *C. dugesii* (Fig. 1a) and 3000 individuals of *B. palmeri* (Fig. 1b) were randomly positioned within a 10 km² radius of each recorded location. Estimates of population size do not exist, but population sizes were established according to field observations. Although floristic surveys have been carried in the Bajío region in the last decades to record all flora species (which is the main objective of the herbarium IEB, <https://plants.jstor.org/partner/IEB>), there is still a possibility that some sites of species occurrence are unknown. To include the effect of missing sites on sampling schemes (see sampling design), I assigned individuals within few random locations across the landscape. The distribution of *C. dugesii* was more spatially aggregated relative to the more scattered distribution of *B. palmeri*. Mating parameters were set in CDPOP to represent a dioicous (equal number of females and males in *B. palmeri*) and a monoicous (*C. dugesii*) species. Pollen and seed dispersal were modeled from a probabilistic distribution proportional to an inverse linear function, which is determined by the maximal dispersal distance travelled by seeds (female- D_{max}) and pollen (male- D_{max}) (Landguth and Cushman 2010). Pollen distances were specified three times larger than seed dispersal as pollen is assumed to travel longer distances relative to the distances travelled by seeds (McCauley 1997; Petit et al. 2005). For *C. dugesii*, pollen flow was set to reach $D_{max} = 25\%$ (60, 952.8 m) of the maximal Euclidean distance between individuals (243,811.3 m), while seed dispersal distance was set to $D_{max} = 5\%$ (36, 571.6 m). For *B. palmeri*, pollen flow was set to $D_{max} = 15\%$ (51, 816.9 m) and seed dispersal distances were set to $D_{max} = 3\%$ (10, 363.4 m) of the maximal Euclidean pairwise individual distance (345, 445.8 m). For both species, there is no knowledge of pollen and seed dispersal, but the parameterized seed dispersal distance corresponded to dispersal distances in related *Bursera* species reported in the literature (Cantarello et al. 2011). For *C. dugesii*, there is no dispersal distances information in related species, but dispersal parameters selected allow building a genetic pattern of isolation by distance without crashing the population to extinction as IBD become stronger over generations.

Population growth followed a logistic function with a maximum size of 5000 (*C. dugesii*) or 3000 (*B. palmeri*) individuals (population size was constant at every

generation), while allowing seeds fecund from multiple fathers. Based on information from the literature (Bonfil-Sanders et al. 2008; Sánchez-Martínez et al. 2011), age structure and percentage of mortality per class were defined as: (1) seeds (40% mortality), (2) seedlings (80% mortality), (3) juveniles (50% mortality), (4) reproductive individuals (20% mortality), and (5) mature individuals (100% mortality for non-overlapping generations). Simulations were set from generation 1 to 300, which generated genotypes of 20 microsatellites and ten alleles per loci. Because CDPOP simulates stochastic processes, I ran 25 replicates to account for the mean and variability of the spatial genetic structure for each species.

Sampling design combinations

I was interested to analyze the effects of sampling given three common factors: (1) spatial sampling scheme, (2) sampling intensity (number of sites), and (3) number of microsatellite loci. I applied three common types of spatial sampling schemes (random, geographically clustered, and systematic sampling within documented localities). For the random sampling, I randomly collected individuals across the whole landscape, without distinguishing sampling sites. For the geographically clustered sampling, I sampled individuals from small portions of the landscape within a 10 km² radius. This type of sampling is common when logistic and economic resources are limited to collect few samples, but the researcher categorize sampling sites based on their geographic or environmental differences (e.g., Abeysinghe et al. 2000; Rico et al. 2008; Carrillo-Ángeles and Mandujano 2011). For the systematic sampling, I sampled individuals within a 10 km² radius of documented locations. This sampling type is more extensive than the cluster sampling since a larger spatial representation of the species distribution is achieved, but some areas are missed because of their geographically inaccessibility or because the researcher have incomplete knowledge of species occurrence (e.g., He et al. 2009; Dubreuil et al. 2010). For each sampling strategy, four sampling intensities were applied: 80, 65, 50 and 35%. For the cluster and systematic sampling, sampling intensity was the number of documented sites of occurrence, while for the random sampling was the number of individuals across the whole landscape. The number of individuals collected within each of the four intensity levels in the random sampling, was set equal to the number of individuals sampled in each level of the systematic sampling. I implemented three sets of microsatellites: 20, 14, and 8 microsatellite loci, which reflect similar numbers of markers commonly employed in genetic studies for trees (e.g., *Astronium urundeuva* n = 7 loci, Caetano et al. 2005; *Dipteryx alata* n = 8 loci, de Campos Telles et al. 2014; *Cedrela odorata* n = 9 loci, Hernández et al. 2008; *Tetragastris panamensis* n = 15 loci, Kenfack and Dick 2009; *Protorhus deflexa* n = 19 loci, Sato et al. 2014; *Laurus* n = 20 loci, Arroyo et al. 2010).

Table 1 Simulated sampling factors and associated levels of analysis investigated in two forest tree species

Factor	Levels	Number of levels
Sampling strategy	Random, cluster, systematic	3
Sampling intensity	35%, 50%, 60%, 80%	4
Number of microsatellite loci	8, 14, 20	3

Levels within the three main sampling factors resulted in 36 sampling design combinations (Table 1).

Individual-based sampling

The objective of this study was not to assess the performance of individual-based and population-based analysis, but to examine how sampling scheme and effort can bias genetic differentiation estimates. Population-based analysis assumes that each sampling site constitute a distinct population, which in many tree species that are recently fragmented such assumption is likely inadequate. Empirical evidence suggests that individual-based analysis is adequate for quantifying spatial genetic structure in species with continuous and non-continuous distributions (Bankenhol and Fortin 2016). Prunier et al. (2013) and Luximon et al. (2014) using spatially-explicit computer simulations showed that individual-based analysis provides more statistical power for detecting isolation by distance relative to population-based analysis. Specifically, Prunier et al. (2013) for species with patchy distributions showed that as few as three to four sampling individuals per aggregate was sufficient for estimating spatial genetic structure using individual-based analysis. To avoid artificial delimitations of putative populations and to make analyses comparable among sampling schemes, I performed individual-based analysis.

Effects of sampling design on estimates of IBD

I used Mantel test (Mantel 1967) to evaluate the correlation of genetic and geographical distances for each of the 36 scenarios using the library *ecodist* (Goslee and Urban 2007) in the R statistical software (R Development Core Team 2016). Inter-individual genetic distances were calculated using the proportion of shared alleles (D_{PS} , Bowcock et al. 1994), and Euclidean distances were calculated from the X and Y coordinates between all pairs of individuals. This was applied for each of the 25 replicate runs and for seven generation periods: 25, 50, 100, 150, 200, 250, and 300. To facilitate comparisons among sampling combinations and their deviation from the true r coefficient, I calculated a standardized measure of relative error (RE; Kossinets 2006): $RE = (r \text{ coefficient}_{\text{True-value}} - r \text{ coefficient}_{\text{Replicate-value}} / r \text{ coefficient}_{\text{True-value}}) \times 100\%$. The relative error represents the proportional difference between the mean true r coefficient and the estimated mean r coefficient among all simulated runs and for each sampling combination, which means that higher absolute values represent a greater divergence between the true and estimated value.

Significant deviations of IBD r coefficients from the true values and effect size

Three-way ANOVAs were applied to identify main factors and their interactions using as response variable the r coefficients. A logarithmic transformation to the response variable was applied to meet ANOVA assumptions. For all ANOVAs, I calculated the effect size of each resultant significant factor and their interactions to determine their relative strength on the response variable. The effect size (η^2) was calculated as the ratio of the effect variance (sum of squares SS) to the total variance (total SS; Tabachnick and Fidell 2007). To assess whether the 25 replicate Mantel r coefficients significantly deviated from the observed distribution of “true” r coefficients (using the total genotype dataset for each of the 25 runs), I applied a Kolmogorov-Smirnoff test (KS) for each of the 36 sampling combinations.

Results

Effects of sampling design on estimates of IBD

For both tree species, the strength of the association between genetic and Euclidean distances increased with the number of generations, but *B. palmeri* resulted in a stronger IBD pattern (generation 25: $r = 0.09$ and generation 300: $r = 0.35$, $p < 0.05$) relative to the strength of IBD shown by *C. dugesii* (generation 25: $r = 0.05$ and generation 300: $r = 0.29$, $p < 0.05$). Figure 2 shows an example of the variation of RE values within sampling combinations and from generation 25 to 300. To facilitate visualization, Fig. 2 shows only comparisons of the three sampling strategies at sampling intensities of 80 and 35%, and for 20 and 8 loci, which were the scenarios with marked differences. For both species, RE values were larger when the r coefficient was low (generation 25) and tended to decrease as the number of generations progressed and the strength of the correlation

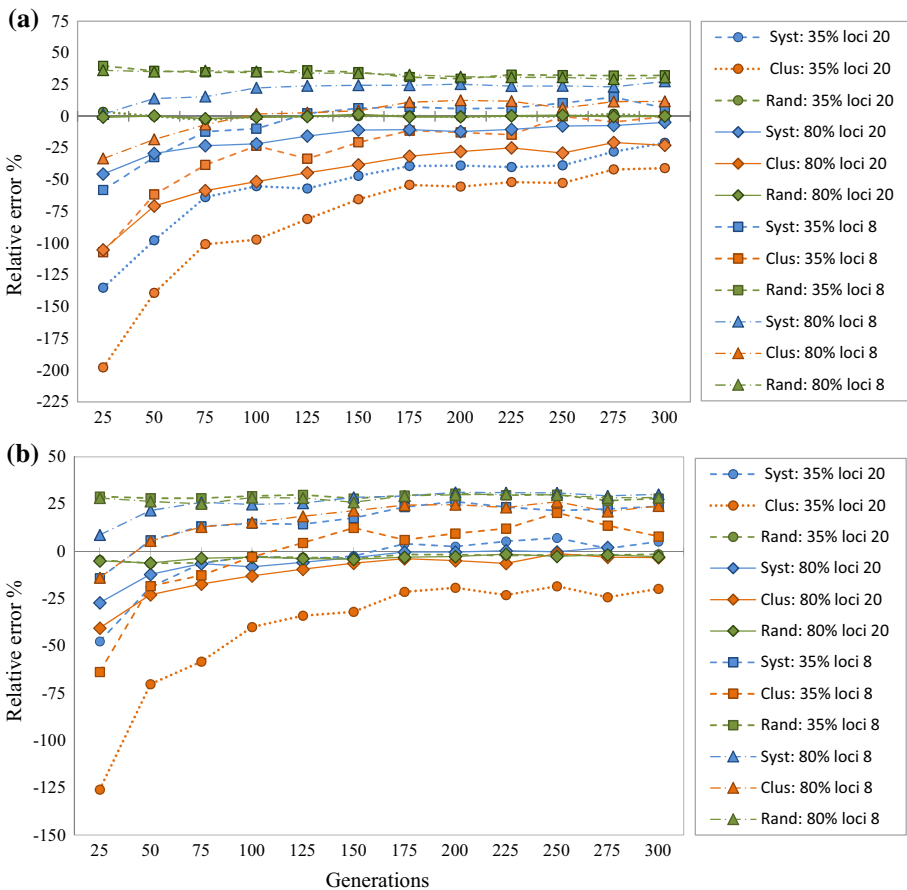


Fig. 2 Relative error plots of the three spatial sampling schemes at 80 and 35% of sampling intensity, and 20 and 8 microsatellite loci combinations for **a** *Cedrela dugesii* and **b** *Bursera palmeri*. RE values were calculated every 25 generations from generation 25 to generation 300. Sampling abbreviations: *Syst* Systematic, *Clus* Cluster and *Rand* Random

coefficient increased. This trend was more evident in *C. dugesii* (Fig. 2a). An exception to this pattern was for the random sampling with 20 loci, which regardless of sampling intensity and generation time, was the sampling combination that produced the lowest error (<10%). Instead, the number of loci was more important in the random sampling, with an increase of >30% in RE values over scenarios with 8 loci.

For both species, the cluster sampling with 20 loci and 35% of sampling intensity was the sampling combination that had the largest RE values independently of generation time. Overall, the cluster and systematic sampling were more sensitive to sampling intensity and number of loci, but for *B. palmeri* RE values within each sampling combination varied more and trends were less predictable (Fig. 2b). Moreover, the cluster and systematic sampling at 35% of sampling intensity resulted in higher RE variation, which changed from negative (above the true mean r coefficient) to positive (below the true mean r coefficient) as IBD become stronger over generations. Cluster sampling with 8 loci resulted in lower RE values relative to 20 loci regardless of sampling intensity, which was more evident in *C. dugesii* (Fig. 2a).

Significant deviations of IBD r coefficients from the true values

Figures 3 and 4 show boxplots of the distribution of estimated r coefficients around the “true” mean r coefficient (dashed line) for the 36 sampling combinations and for three

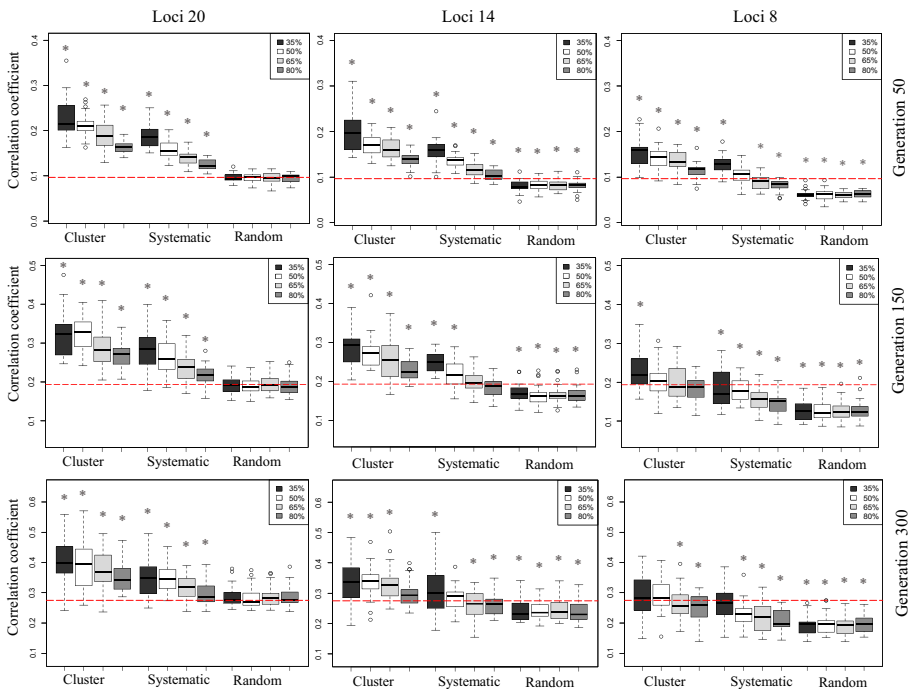


Fig. 3 Boxplots of the distribution of estimated r coefficients for each of the 36 sampling combinations in three generations (50, 150, and 300) in *Cedrela dugesii*. The dashed line represents the “true” mean Mantel r coefficient, calculated across 25 replicates for the complete data set. The stars above each box denotes a significant deviation ($p < 0.05$) of the distribution of estimated r coefficients from the true r values using the Kolmogorov-Smirnoff test

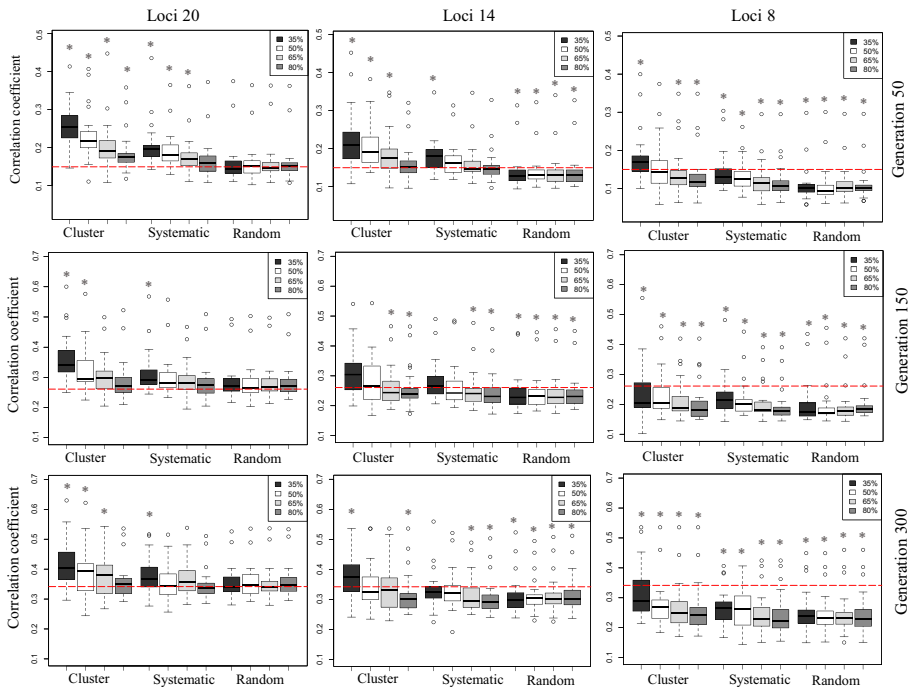


Fig. 4 Boxplots of the distribution of estimated r coefficients for each of the 36 sampling combinations in three generations (50, 150, and 300) in *Bursera palmeri*. The dashed line represents the “true” mean Mantel r coefficient, calculated across 25 replicates for the complete data set. The stars above each box denotes a significant deviation ($p < 0.05$) of the distribution of estimated r coefficients from the true r values using the Kolmogorov-Smirnoff test

generations (50, 150, and 300). For *C. dugesii*, random sampling with 20 loci resulted in estimated r coefficients closest to the true mean r , irrespective of sampling intensity and the strength of IBD. Results from the KS test indicated non-significant differences between distributions of the true r and estimated r coefficients for this sampling combination (Fig. 3). Cluster and systematic sampling at 20 loci across the three generations, showed significant differences from the true correlation for all sampling intensities. For both sampling strategies, estimated r coefficients were above the true mean r , and hence associations of IBD were overestimated. In scenarios with 14 and 8 loci, random sampling across all levels of sampling intensity resulted in estimated r coefficients below (underestimating) the true mean r , and with marked differences at 8 loci. This trend was observed across the three generations, and these deviations from the true r value were statistically significant. For the cluster and systematic sampling, estimated r coefficients were higher at loci 20, which monotonically decreased as the number of loci went from 14 to 8, which was more evident at generation 25 relative to generation 300.

Similarly in *B. palmeri*, the random sampling resulted in estimated r coefficients that were not significantly different from the true r across the three generations analyzed. Although estimated r values for the systematic sampling were overall higher relative to the mean true r in scenarios with 20 loci, those differences were not statistically significant. Moreover, estimated r values for the systematic sampling were closer to the true r value with increasing the number of sites sampled, which was more evident as the strength of

IBD increased. This trend was similar for the cluster sampling with 85% intensity at generation 150 and 300 (Fig. 4). Unlike *C. dugesii*, sampling intensity produced a lower variation within the three sampling strategies and irrespective of the number of loci (Fig. 4). Moreover, all sampling combinations with 8 loci (except cluster sampling at 35% intensity) had similar distributions of estimated r coefficients at all sampling intensity levels, which distributions were significantly lower than the true r coefficient.

Effect size of sampling factors

Results from ANOVA showed that effects of sampling strategy, sampling intensity, number of loci, and the interaction between sampling strategy and number of loci were statistically significant across the seven analyzed generations in both species (Tables 2 and 3). However, there were important differences in the relative effect sizes among factors, which values varied in relation to generation time and thus the strength of the r coefficient. Specifically, for *C. dugesii*, sampling strategy had the largest effect size, which comprised 74% of the variation at generation 25 ($r = 0.05$), but that decreased to 50% at generation 150 ($r = 0.2$). After generation 200 ($r = 0.23$), the number of loci reversed the trend and had a largest importance (>50% of the variation). Sampling intensity followed by the interaction between sampling strategy and number of loci accounted for less than 10% of the total variation (Table 2). For *B. palmeri*, sampling strategy had the largest effect size only at generation 25 ($r = 0.1$), which comprised 56% of the variation, while number of loci was the most important factor after generation 50 ($r = 0.15$), explaining 54%. Effect size values for the number of loci continued to increase as associations of IBD become stronger, and for a maximum effect size of 86% at generation 300 ($r = 0.35$). Also, sampling intensity followed by the interaction between sampling strategy and number of loci accounted for less than 15% of the variation (Table 3).

Discussion

Performance of sampling design on the correct quantification of IBD

As expected, the dioicous species with shorter dispersal range (*B. palmeri*) developed stronger genetic differentiation over distance across all generations relative to the monoicous species with larger dispersal (*C. dugesii*). Results showed that the spatial sampling scheme and the number of loci were important factors influencing the correct estimation of IBD. Specifically, the random sampling with 20 microsatellite loci was the sampling combination that performed best across generations and for both species, and thus performance was independent of the strength of IBD (generation time) and the spatial distribution of individuals.

The good performance of the random sampling has been observed in previous computer simulation studies, which also found that increasing sample size increased the accuracy for identifying the underlying spatial process (Landguth et al. 2012; Oyler-McCance et al. 2013). I found that the random sampling with 20 loci was equally sufficient for quantifying the “true” strength of IBD at the lowest and highest proportions of sampled individuals. Indeed, the random sampling scheme was unaffected by sampling intensity across all scenarios. The random sampling in this study, was a good unbiased distribution of samples across all sites of species occurrence, which allowed to capture the existing pattern of

Table 2 Summary statistics of three-way ANOVAs to test differences among sampling factors and their interactions on mantel correlation coefficients generated for simulated data of seven generations in *Cedrela dugesii*

Generation	Factor	d.f.	F	Prob.	Effect size	r_{true}
25	ST	2	1541.09	***	0.738	0.05
	SI	3	90.60	***	0.065	
	Loc	2	314.62	***	0.151	
	ST:SI	6	29.18	***	0.042	
	ST:Loc	4	2.64	*	0.003	
	SI:Loc	6	0.43	NS	–	
	ST:SI:Loc	12	0.45	NS	–	
50	ST	2	1673.34	***	0.683	0.01
	SI	3	100.06	***	0.061	
	Loc	2	536.97	***	0.219	
	ST:SI	6	28.28	***	0.035	
	ST:Loc	4	1.95	NS	–	
	SI:Loc	6	0.046	NS	–	
	ST:SI:Loc	12	0.37	NS	–	
100	ST	2	884.87	***	0.558	0.15
	SI	3	53.63	***	0.051	
	Loc	2	575.20	***	0.362	
	ST:SI	6	13.12	***	0.025	
	ST:Loc	4	0.58	NS	–	
	SI:Loc	6	0.33	NS	–	
	ST:SI:Loc	12	0.81	NS	–	
150	ST	2	528.90	***	0.496	0.2
	SI	3	36.65	***	0.052	
	Loc	2	446.23	***	0.418	
	ST:SI	6	9.894	***	0.028	
	ST:Loc	4	0.63	NS	–	
	SI:Loc	6	0.31	NS	–	
	ST:SI:Loc	12	0.72	NS	–	
200	ST	2	288.42	***	0.415	0.23
	SI	3	24.27	***	0.052	
	Loc	2	348.44	***	0.501	
	ST:SI	6	6.48	***	0.028	
	ST:Loc	4	0.33	NS	–	
	SI:Loc	6	0.205	NS	–	
	ST:SI:Loc	12	0.266	NS	–	
250	ST	2	343.37	***	0.408	0.26
	SI	3	21.201	***	0.038	
	Loc	2	439.13	***	0.521	
	ST:SI	6	6.97	***	0.025	
	ST:Loc	4	0.13	NS	–	
	SI:Loc	6	1.45	NS	–	
	ST:SI:Loc	12	0.45	NS	–	

Table 2 continued

Generation	Factor	d.f.	F	Prob.	Effect size	r_{true}
300	ST	2	203.65	***	0.363	0.29
	SI	3	13.81	***	0.037	
	Loc	2	315.24	***	0.562	
	ST:SI	6	4.49	***	0.024	
	ST:Loc	4	0.74	NS	–	
	SI:Loc	6	0.44	NS	–	
	ST:SI:Loc	12	0.85	NS	–	

Largest effect size denoted in bold. Mantel r is the mean true value of the complete data at each simulated run

ST Sampling type, SI Sampling intensity, Loc Number of loci, d.f. Degree of freedom, NS non-significant values, ST:SI Interaction

P values are indicated as follow: $p < 0.0001$ ***; $p < 0.001$ **; $p < 0.01$ *

genetic differentiation with acceptable degree on the condition that the number of loci was high (20 loci).

Simulation studies have found that the systematic sampling, where the sampling is distributed at regular intervals across the landscape, perform with similar acceptable rates to the random sampling (Naujokaitis-Lewis et al. 2013; Oyler-McCance et al. 2013). In contrast to previous studies, the systematic sampling did not show similar trends to the random sampling. Unlike the random sampling, performance of the systematic sampling was related to sampling intensity and the strength of IBD. Specifically, in *B. palmeri*, the effectiveness of the systematic sampling for successfully describing the “true” pattern of IBD, increased with the proportion of sampled sites and the number of loci. For *C. dugesii*, the above was observed only for the combination with 14 loci and when the pattern of IBD was moderate. The difference observed on the performance of the random versus the systematic sampling scheme is related to the influence of missing sites and the spatial configuration of the species distribution. In simulation studies for species with continuous distributions, the systematic sampling spaced at regular intervals and covering all areas of the species occurrence is more likely to capture the prevailing pattern of genetic structure, thus performing equally well to the random sampling. For the systematic sampling in this study, samples were collected only in recorded locations, while missing “undocumented” sites. Hence, as fewer sites were sampled, there was a larger chance of not accurately capturing the genetic variation found across the landscape (e.g., Koen et al. 2013; Naujokaitis-Lewis et al. 2013). The random and systematic sampling differ only in the spatial allocation of samples because the number of individuals within each intensity level were set equal, and hence differences are not due to sample size. The simulation study of van Strien et al. (2014), showed that habitat configuration has a large influence on spatial patterns of genetic differentiation under IBD, where the highest differentiation between pairs of demes (as measure with genetic distance metrics) may not necessarily corresponds to the largest pair-wise Euclidean distance. This implies that the strength of IBD may not appear as a uniform spatial gradient of genetic differentiation across the landscape. The different performance of sampling schemes in the two simulated species is likely related to the spatial configuration of individuals as suggested from the work of van Strien et al. (2014).

Table 3 Summary statistics of three-way ANOVAs to test differences among sampling factors and their interactions on mantel correlation coefficients generated for simulated data of seven generations in *Bursera palmeri*

Generation	Factor	d.f.	F	Prob.	Effect size	r_{true}
25	ST	2	233.78	***	0.561	0.09
	SI	3	23.37	***	0.084	
	Loc	2	122.95	***	0.295	
	ST:SI	6	7.58	***	0.055	
	ST:Loc	4	0.37	NS	–	
	SI:Loc	6	0.08	NS	–	
	ST:SI:Loc	12	0.22	NS	–	
50	ST	2	94.44	***	0.332	0.15
	SI	3	13.79	***	0.073	
	Loc	2	153.54	***	0.541	
	ST:SI	6	4.79	***	0.051	
	ST:Loc	4	0.039	NS	–	
	SI:Loc	6	0.14	NS	–	
	ST:SI:Loc	12	0.09	NS	–	
100	ST	2	60.09	***	0.214	0.23
	SI	3	10.45	***	0.056	
	Loc	2	192.52	***	0.685	
	ST:SI	6	3.78	**	0.04	
	ST:Loc	4	0.29	NS	–	
	SI:Loc	6	0.06	NS	–	
	ST:SI:Loc	12	0.08	NS	–	
150	ST	2	18.10	***	0.078	0.28
	SI	3	7.34	***	0.048	
	Loc	2	191.79	***	0.831	
	ST:SI	6	2.76	*	0.036	
	ST:Loc	4	0.15	NS	–	
	SI:Loc	6	0.12	NS	–	
	ST:SI:Loc	12	0.18	NS	–	
200	ST	2	20.28	***	0.075	0.31
	SI	3	7.21	***	0.038	
	Loc	2	229.24	***	0.845	
	ST:SI	6	3.023	**	0.033	
	ST:Loc	4	0.35	NS	–	
	SI:Loc	6	0.23	NS	–	
	ST:SI:Loc	12	0.18	NS	–	
250	ST	2	10.89	***	0.034	0.33
	SI	3	7.45	***	0.035	
	Loc	2	286.47	***	0.898	
	ST:SI	6	2.54	*	0.024	
	ST:Loc	4	0.24	NS	–	
	SI:Loc	6	0.11	NS	–	
	ST:SI:Loc	12	0.35	NS	–	

Table 3 continued

Generation	Factor	d.f.	F	Prob.	Effect size	r_{true}
300	ST	2	15.94	***	0.057	0.35
	SI	3	7.24	***	0.041	
	Loc	2	240.45	***	0.860	
	ST:SI	6	3.09	**	0.033	
	ST:Loc	4	0.25	NS	–	
	SI:Loc	6	0.29	NS	–	
	ST:SI:Loc	12	0.19	NS	–	

Largest effect size denoted in bold. Mantel r is the mean true value of the complete data at each simulated run

ST Sampling type, SI Sampling intensity, Loc Number of loci, d.f. Degree of freedom, NS non-significant values, ST:SI Interaction

P values are indicated as follow: $p < 0.0001$ ***; $p < 0.001$ **; $p < 0.01$ *

Oyler-McCance et al. (2013) through an individual-based simulation study for an animal species with continuous distributions found that the cluster sampling had a low probability of identifying the underlying spatial process irrespective of sample size. Overall, the results here also suggest that the cluster sampling does not perform well for species with non-continuous distributions. Remarkably, at the lowest proportion of sampled sites and combined with 20 loci, the cluster sampling performed the worst, overestimating the strength of IBD. This sampling scheme, was more sensitive to sample size and the number of loci, resulting in large variance of estimated r values across all scenarios. This pattern was more evident in *C. dugesii*, which spatial distribution was more aggregated.

Sampling in few locations of the whole species distribution, is common in population genetic studies due to logistics or financial constraints (e.g., Tucker et al. 2014; Dubreuil et al. 2010; Rico et al. 2008). The cluster sampling captures only a portion of the genetic variation found across the landscape, if by chance areas of high genetic differentiation are included, the strength of the association between genetic distances and Euclidean distances would appear stronger, thus overestimating IBD (Landguth and Schwartz 2014). On the other hand, if areas of the landscape with low genetic differentiation are included, IBD would be underestimated. This would explain the patterns shown for the systematic and the cluster sampling, but more marked in the cluster sampling at low levels of sampling intensity where the range of the r coefficient values was larger (i.e., interquartile size and whiskers of boxplots Figs. 3 and 4) relative to the variation in the other sampling schemes.

Relative importance of sampling factors

Although the three analyzed sampling factors had significant effects on the quantification of spatial genetic structure, the relative importance of each factor varied between species and across generations. For *C. dugesii*, the spatial sampling scheme was the most important from generation 25 to 150, which then was reversed for the number of loci, while for *B. palmeri* the number of loci was by far the most important factor from generation 50. In all cases, sampling intensity had a small effect size. Differences between species in the relative importance of sampling factors may be related to the strength of IBD. Specifically as IBD become stronger, the relative effect size values for the number of loci increased,

while decreased for the spatial sampling scheme. Landguth et al. (2012), in a simulation study found that increasing the number of loci (up to 25) was important for the correct identification of the underlying spatial process. This trend was more evident if IBD was strong (Mantel $r > 0.7$). However, the authors found that increasing the number of samples (from 100 to 500) if the number of loci was small ($n = 15$) improved the probability of correctly identifying the underlying spatial process (Landguth et al. 2012). In this study, increasing the number of samples (e.g., from 260 individuals at 35% to 700 individuals at 80%) did not improve the performance of the sampling scheme if loci was either 14 or 8. Specifically, using 8 loci underestimated the “true” pattern of IBD in most sampling combinations (except in *C. dugesii* for the cluster sampling at low sample sizes). Notably, underestimating IBD with 8 loci was marked at generation 300 when IBD was the strongest.

These results indicate the lower power to detect genetic structure in standard correlative matrix distance methods if few microsatellite loci are used (Landguth et al. 2012; Peterman et al. 2016). Mantel and partial Mantel test statistics have been criticized due to its poor performance for accurately identifying the main spatial process driving genetic structure if multiple and correlated competing landscape hypotheses occur (Guillot and Rousset 2013; Zeller et al. 2016). Because in this study isolation by distance is the only spatial process being modeled, I consider the simple Mantel test sufficient for illustrating the relative performance that sampling design have on the quantification of IBD. It would be interesting to explore under more complex simulation scenarios, such as landscape resistances to gene flow (e.g., Zeller et al. 2016) or adaptive genetic differentiation under gene flow (e.g., Landguth and Balkenhol 2012), the performance of alternative statistical methods, such as spatial regressions and multiple regression on distance matrices (MRM) (Wagner and Fortin 2016). Other important aspect to explore using computer simulations, would be the role of dispersal vectors on patterns of genetic structure. For instance, the influence of wind speed and turbulence is known to importantly influence the distance and direction of pollen flow (Wang et al. 2016) and seed deposition (Nathan et al. 2002). Moreover, the movement of pollinators can also be influenced by the speed and direction of wind (Ahmed et al. 2009). The effect of sampling schemes under more realistic scenarios beyond the species spatial configuration and landscape structure would be needed.

Recommendations and Conclusions

Results here highlight that covering the total spatial extent of the species distribution is critical because missing areas of occurrence is likely to under- or over-estimate patterns of genetic differentiation. By contrasting two species with varying spatial distributions, it was observed that a cluster sampling scheme is even more problematic for species with patchy distributions. Equally important is the number of microsatellite loci because as fewer loci are included, there is an increasing chance of underestimating IBD due to low power to detect spatial genetic structure. Increasing the number of samples over the number of loci, appears not to improve the performance of sampling schemes for correctly identifying the underlying process even if sampling is randomly distributed across the landscape. This implies that researchers should give more priority to obtain large number of markers over the effort of increasing sample size. Obtaining a large number of loci >20 and with good-level of polymorphism (i.e., 10 alleles) is still difficult for many research studies in non-model species with poor or non-existing genetic information. Fortunately, advances of next

generation sequencing technologies allow overcoming the limitation associated with marker discovery and polymorphism. Specifically, restriction site associated DNA sequencing and genotyping by sequencing methods that identify thousands of polymorphic markers across the genome (single nucleotide polymorphisms SNP) are becoming rapidly accessible to research studies in non-model species (Baird et al. 2008; Elshire et al. 2011).

Despite the increasing development of computer simulation software in ecological and landscape genetics (e.g., SPOTG of Hoban et al. 2013; SimAdapt of Rebaudo et al. 2013; CDMetaPOP of Landguth et al. 2016), computer simulations remain poorly used as a tool for sampling planning in research studies. The routine implementation of computer simulations with model parameters focused on the biology of the study species, the vectors, and the landscape structure would support more effective decisions on sampling design in population genetic studies (Hoban 2014).

Acknowledgements Thanks to Michelle DiLeo for explaining details of running the computer simulations in CDPOP. Also thanks to the organizing committee of the Gene Conservation of Tree Species workshop and to Kevin Potter and Richard Sniezko for the invitation to submit my work to this special issue. Special thanks to Erin Landguth and other anonymous reviewer, which comments improved the quality of this work. The author declare that no conflict of interest exists.

References

- Abeyinghe PD, Triest L, Greef BD et al (2000) Genetic and geographic variation of the mangrove tree *Bruguiera* in Sri Lanka. *Aquat Bot* 67:131–141. doi:[10.1016/S0304-3770\(99\)00096-0](https://doi.org/10.1016/S0304-3770(99)00096-0)
- Ahmed S, Compton SG, Butlin RK, Gilmartin PM (2009) Wind-borne insects mediate directional pollen transfer between desert fig trees 160 kilometers apart. *Proc Natl Acad Sci USA* 106:20342–20347. doi:[10.1073/pnas.0902213106](https://doi.org/10.1073/pnas.0902213106)
- Alberto FJ, Aitken SN, Alía R et al (2013) Potential for evolutionary responses to climate change—evidence from tree populations. *Glob Change Biol* 19:1645–1661. doi:[10.1111/gcb.12181](https://doi.org/10.1111/gcb.12181)
- Anderson CD, Epperson BK, Fortin MJ et al (2010) Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Mol Ecol* 19:3565–3575. doi:[10.1111/j.1365-294X.2010.04757.x](https://doi.org/10.1111/j.1365-294X.2010.04757.x)
- Arroyo JM, Rigueiro C, Rodríguez R et al (2010) Isolation and characterization of 20 microsatellite loci for laurel species (*Laurus*, Lauraceae). *Am J Bot* 97:e26–e30. doi:[10.3732/ajb.1000069](https://doi.org/10.3732/ajb.1000069)
- Baird NA, Eitter PD, Atwood TS et al (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3:e3376. doi:[10.1371/journal.pone.0003376](https://doi.org/10.1371/journal.pone.0003376)
- Bankenhol N, Fortin MJ (2016) Basics of study design: sampling landscape heterogeneity and genetic variation for landscape genetic studies. In: Bankenhol N, Cushman S, Storfer A, Waits L (eds) *Landscape genetics: concepts, methods, applications*. Wiley, Chichester, pp 58–76. doi:[10.1002/9781118525258.ch05](https://doi.org/10.1002/9781118525258.ch05)
- Bonfil-Sanders C, Cajero-Lázaro I, Evans RY (2008) Germinación de semillas de seis especies de *Bursera* del centro de México. *Agrociencia* 42:827–834
- Bowcock AM, Ruiz-Linares A, Tomfrohde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368:455–457
- Broadhurst LM, Lowe A, Coates DJ et al (2008) Seed supply for broadscale restoration: maximizing evolutionary potential. *Evol Appl* 1:587–597. doi:[10.1111/j.1752-4571.2008.00045.x](https://doi.org/10.1111/j.1752-4571.2008.00045.x)
- Caetano S, Silveira P, Spichiger R, Naciri-Graven Y (2005) Identification of microsatellite markers in a neotropical seasonally dry forest tree, *Astronium urundeuva* (Anacardiaceae). *Mol Ecol Notes* 5:21–23. doi:[10.1111/j.1471-8286.2004.00814.x](https://doi.org/10.1111/j.1471-8286.2004.00814.x)
- Calderón de Rzedowski G, Germán MT (1993) Meliaceae. *Flora del Bajío y de Regiones Adyacentes* 11:1–5
- Cantarello E, Newton AC, Hill RA et al (2011) Simulating the potential for ecological restoration of dryland forests in Mexico under different disturbance regimes. *Ecol Model* 222:1112–1128. doi:[10.1016/j.ecolmodel.2010.12.019](https://doi.org/10.1016/j.ecolmodel.2010.12.019)
- Carrillo-Ángeles IG, Mandujano MC (2011) Patrones de distribución espacial en plantas clonales. *Bol Soc Bot Mex* 89:1–18

- de Campos Telles MP, Dobrovolski R, da Silva e Souza K et al (2014) Disentangling landscape effects on population genetic structure of a neotropical savanna tree. *Nat Conserv* 12:65–70. doi:[10.4322/natcon.2014.012](https://doi.org/10.4322/natcon.2014.012)
- Dubreuil M, Riba M, González-Martínez SC et al (2010) Genetic effects of chronic habitat fragmentation revisited: strong genetic structure in a temperate tree, *Taxus baccata* (Taxaceae), with great dispersal capability. *Am J Bot* 97:303–310. doi:[10.3732/ajb.0900148](https://doi.org/10.3732/ajb.0900148)
- Elshire RJ, Glaubitz JC, Sun Q et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi:[10.1371/journal.pone.0019379](https://doi.org/10.1371/journal.pone.0019379)
- Encino-Ruiz L, Lindig-Cisneros R, Gómez-Romero M (2014) Desempeño de tres especies arbóreas del bosque tropical caducifolio en un ensayo de restauración ecológica. *Bot Sci* 91:107–114. doi:[10.17129/botsci.406](https://doi.org/10.17129/botsci.406)
- Ennos RA (2001) Inferences about spatial processes in plant populations from the analysis of molecular markers. In: Silvertown J, Antonovics J (eds) *Integrating ecology and evolution in a spatial context*. Blackwell, Oxford, pp 45–57
- Epperson BK (2007) Plant dispersal, neighbourhood size and isolation by distance. *Mol Ecol* 16:3854–3865. doi:[10.1111/j.1365-294X.2007.03434.x](https://doi.org/10.1111/j.1365-294X.2007.03434.x)
- Frankham R (2010) Challenges and opportunities of genetic approaches to biological conservation. *Biol Conserv* 143:1919–1927. doi:[10.1016/j.biocon.2010.05.011](https://doi.org/10.1016/j.biocon.2010.05.011)
- Goslee SC, Urban DL (2007) The ecodist package for dissimilaritybased analysis of ecological data. *J Stat Softw* 22:1–19
- Guillot G, Rousset F (2013) Dismantling the Mantel tests. *Methods Ecol Evol* 4:336–344. doi:[10.1111/2041-210x.12018](https://doi.org/10.1111/2041-210x.12018)
- Hall LA, Beissinger SR (2014) A practical toolbox for design and analysis of landscape genetics studies. *Landsc Ecol* 29:1487–1504. doi:[10.1007/s10980-014-0082-3](https://doi.org/10.1007/s10980-014-0082-3)
- Hampe A, Petit RJ (2005) Conserving biodiversity under climate change: the rear edge matters. *Ecol Lett* 8:461–467. doi:[10.1111/j.1461-0248.2005.00739.x](https://doi.org/10.1111/j.1461-0248.2005.00739.x)
- He J, Chen L, Si Y et al (2009) Population structure and genetic diversity distribution in wild and cultivated populations of the traditional Chinese medicinal plant *Magnolia officinalis* subsp. *biloba* (Magnoliaceae). *Genetica* 135:233–243. doi:[10.1007/s10709-008-9272-8](https://doi.org/10.1007/s10709-008-9272-8)
- Hernández G, Buonamici A, Walker K et al (2008) Isolation and characterization of microsatellite markers for *Cedrela odorata* L. (Meliaceae), a high value neotropical tree. *Conserv Genet* 9:457–459. doi:[10.1007/s10592-007-9334-y](https://doi.org/10.1007/s10592-007-9334-y)
- Hoban S (2014) An overview of the utility of population simulation software in molecular ecology. *Mol Ecol* 23:2383–2401. doi:[10.1111/mec.12741](https://doi.org/10.1111/mec.12741)
- Hoban S, Schlarbaum S (2014) Optimal sampling of seeds from plant populations for ex situ conservation of genetic biodiversity, considering realistic population structure. *Biol Conserv* 177:90–99. doi:[10.1016/j.biocon.2014.06.014](https://doi.org/10.1016/j.biocon.2014.06.014)
- Hoban S, Strand A (2015) Ex situ seed collections will benefit from considering spatial sampling design and species' reproductive biology. *Biol Conserv* 187:182–191. doi:[10.1016/j.biocon.2015.04.023](https://doi.org/10.1016/j.biocon.2015.04.023)
- Hoban S, Gaggiotti O, Bertorelle G (2013) Sample planning optimization tool for conservation and population genetics (SPOTG): a software for choosing the appropriate number of markers and samples. *Methods Ecol Evol* 4:299–303. doi:[10.1111/2041-210x.12025](https://doi.org/10.1111/2041-210x.12025)
- Jenkins DG, Carey M, Czerniewska J et al (2010) A meta-analysis of isolation by distance: relic or reference standard for landscape genetics? *Ecography*. doi:[10.1111/j.1600-0587.2010.06285.x](https://doi.org/10.1111/j.1600-0587.2010.06285.x)
- Kalisz S, Nason JD, Hanzawa FM, Tonsor SJ (2001) Spatial population genetic structure in *Trillium grandiflorum*: the roles of dispersal, mating, history, and selection. *Evolution* 55:1560–1568. doi:[10.1111/j.0014-3820.2001.tb00675.x](https://doi.org/10.1111/j.0014-3820.2001.tb00675.x)
- Keller D, Holderegger R, Van Strien MJ (2013) Spatial scale affects landscape genetic analysis of a wetland grasshopper. *Mol Ecol* 22:2467–2482. doi:[10.1111/mec.12265](https://doi.org/10.1111/mec.12265)
- Kenfack D, Dick CW (2009) Isolation and characterization of 15 polymorphic microsatellite loci in *Tetragastris panamensis* (Bursaceae), a widespread Neotropical forest tree. *Conserv Genet Resour* 1:385–387. doi:[10.1007/s12686-009-9089-5](https://doi.org/10.1007/s12686-009-9089-5)
- Koen EL, Bowman J, Garroway CJ, Wilson PJ (2013) The Sensitivity of genetic connectivity measures to unsampled and under-sampled sites. *PLoS One*. doi:[10.1371/journal.pone.0056204](https://doi.org/10.1371/journal.pone.0056204)
- Kossinets G (2006) Effects of missing data in social networks. *Soc Netw* 28(3):247–268
- Landguth EL, Balkenhol N (2012) Relative sensitivity of neutral versus adaptive genetic data for assessing population differentiation. *Conserv Genet* 13:1421–1426. doi:[10.1007/s10592-012-0354-x](https://doi.org/10.1007/s10592-012-0354-x)
- Landguth EL, Cushman SA (2010) Cdpop: a spatially explicit cost distance population genetics program. *Mol Ecol Resour* 10:156–161. doi:[10.1111/j.1755-0998.2009.02719.x](https://doi.org/10.1111/j.1755-0998.2009.02719.x)

- Landguth EL, Schwartz MK (2014) Evaluating sample allocation and effort in detecting population differentiation for discrete and continuously distributed individuals. *Conserv Genet* 15:981–992. doi:[10.1007/s10592-014-0593-0](https://doi.org/10.1007/s10592-014-0593-0)
- Landguth EL, Fedy BC, Oyler-McCance SJ et al (2012) Effects of sample size, number of markers, and allelic richness on the detection of spatial genetic pattern. *Mol Ecol Resour* 12:276–284. doi:[10.1111/j.1755-0998.2011.03077.x](https://doi.org/10.1111/j.1755-0998.2011.03077.x)
- Landguth EL, Bearlin A, Day CC, Dunham J (2016) CDMetaPOP: an individual-based, eco-evolutionary model for spatially-explicit simulation of landscape demogenetics. *Methods Ecol Evol*. doi:[10.1111/2041-210X.12608](https://doi.org/10.1111/2041-210X.12608)
- Latch EK, Rhodes OE (2006) Evidence for bias in estimates of local genetic structure due to sampling scheme. *Anim Conserv* 9:308–315. doi:[10.1111/j.1469-1795.2006.00037.x](https://doi.org/10.1111/j.1469-1795.2006.00037.x)
- Luximon N, Petit EJ, Broquet T (2014) Performance of individual vs. group sampling for inferring dispersal under isolation-by-distance. *Mol Ecol Resour* 14:745–752. doi:[10.1111/1755-0998.12224](https://doi.org/10.1111/1755-0998.12224)
- Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* 27:209–220
- McCauley DE (1997) The relative contributions of seed and pollen movement to the local genetic structure of *Silene alba*. *J Hered* 88:257–263. doi:[10.1093/oxfordjournals.jhered.a023103](https://doi.org/10.1093/oxfordjournals.jhered.a023103)
- Meirmans PG (2012) The trouble with isolation by distance. *Mol Ecol* 21:2839–2846. doi:[10.1111/j.1365-294X.2012.05578.x](https://doi.org/10.1111/j.1365-294X.2012.05578.x)
- Nathan R, Katul GG, Horn HS et al (2002) Mechanisms of long-distance dispersal of seeds by wind. *Nature* 418:409–413. doi:[10.1038/nature00844](https://doi.org/10.1038/nature00844)
- Naujokaitis-Lewis IR, Rico Y, Lovell J et al (2013) Implications of incomplete networks on estimation of landscape genetic connectivity. *Conserv Genet* 14:287–298
- Oyler-McCance SJ, Kahn NW, Burnham KP et al (1999) A population genetic comparison of large- and small-bodied sage grouse in Colorado using microsatellite and mitochondrial DNA markers. *Mol Ecol* 8:1457–1465. doi:[10.1046/j.1365-294X.1999.00716.x](https://doi.org/10.1046/j.1365-294X.1999.00716.x)
- Oyler-McCance SJ, Fedy BC, Landguth EL (2013) Sample design effects in landscape genetics. *Conserv Genet* 14:275–285. doi:[10.1007/s10592-012-0415-1](https://doi.org/10.1007/s10592-012-0415-1)
- Peterman W, Brocato ER, Semlitsch RD, Eggert LS (2016) Reducing bias in population and landscape genetic inferences: the effects of sampling related individuals and multiple life stages. *PeerJ* 4:e1813. doi:[10.7717/peerj.1813](https://doi.org/10.7717/peerj.1813)
- Petit RJ, Duminil J, Fineschi S et al (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Mol Ecol* 14:689–701. doi:[10.1111/j.1365-294X.2004.02410.x](https://doi.org/10.1111/j.1365-294X.2004.02410.x)
- Prunier JG, Kaufmann B, Fenet S et al (2013) Optimizing the trade-off between spatial and genetic sampling efforts in patchy populations: towards a better assessment of functional connectivity using an individual-based sampling scheme. *Mol Ecol* 22:5516–5530. doi:[10.1111/mec.12499](https://doi.org/10.1111/mec.12499)
- Rebaudo F, Le Rouzic A, Dupas S et al (2013) SimAdapt: an individual-based genetic model for simulating landscape management impacts on populations. *Methods Ecol Evol* 4:595–600. doi:[10.1111/2041-210X.12041](https://doi.org/10.1111/2041-210X.12041)
- Rico Y, Lorenzo C, González-Cózatl FX, Espinoza E (2008) Phylogeography and population structure of the endangered Tehuantepec jackrabbit *Lepus flavigularis*: implications for conservation. *Conserv Genet* 9:1467–1477
- Rzedowski J, Guevara-Féfer F (1992) Burseraceae. *Flora del Bajío y de regiones adyacentes* 3:46
- Rzedowski J, Zamudio S, Calderón G, Paizanni A (2014) El bosque tropical caducifolio en la cuenca lacustre de Pátzcuaro. Michoacán, México
- Sánchez-Martínez E, Hernández-Oria JG, Hernández-Martínez MM, Maruri-Aguilar B, Torres-Galeana LE, Chávez-Martínez R (2011) Consejo de Ciencia y Tecnología del Estado de Querétaro, Querétaro, México
- Sato H, Adeno C, Harata T et al (2014) Isolation and characterization of microsatellite loci for the large-seeded Tree *Protorhus deflexa* (Anacardiaceae). *Appl Plant Sci* 2:1300046. doi:[10.3732/apps.1300046](https://doi.org/10.3732/apps.1300046)
- Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nat Rev Genet* 14:807–820. doi:[10.1038/nrg3522](https://doi.org/10.1038/nrg3522)
- Schwartz MK, McKelvey KS (2009) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conserv Genet* 10:441–452. doi:[10.1007/s10592-008-9622-1](https://doi.org/10.1007/s10592-008-9622-1)
- Tabachnick BG, Fidell LS (2007) Using multivariate statistics, 5th edn. Allyn and Bacon Inc, Boston
- R Development Core Team (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. ISBN 3-900051-07-0. <http://www.R-project.org>
- Tucker JM, Schwartz MK, Truex RL et al (2014) Sampling affects the detection of genetic subdivision and conservation implications for fisher in the Sierra Nevada. *Conserv Genet* 15:123–136. doi:[10.1007/s10592-013-0525-4](https://doi.org/10.1007/s10592-013-0525-4)

- van Strien MJ, Holderegger R, Van Heck HJ (2014) Isolation-by-distance in landscapes: considerations for landscape genetics. *Heredity* 114:27–37. doi:[10.1038/hdy.2014.62](https://doi.org/10.1038/hdy.2014.62)
- Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Mol Ecol* 13:921–935. doi:[10.1046/j.1365-294X.2004.02076.x](https://doi.org/10.1046/j.1365-294X.2004.02076.x)
- Wagner HH, Fortin MJ (2016) Basics of spatial data analysis: linking landscape and genetic data for landscape genetic studies. In: Bankenhol N, Cushman S, Storfer A, Waits L (eds) *Landscape genetics: concepts, methods, applications*. Wiley, Chichester, pp 77–100. doi:[10.1002/9781118525258.ch05](https://doi.org/10.1002/9781118525258.ch05)
- Wang Z-F, Lian J-Y, Ye W-H et al (2016) Pollen and seed flow under different predominant winds in wind-pollinated and wind-dispersed species *Engelhardia roxburghiana*. *Tree Genet Genomes* 12:19. doi:[10.1007/s11295-016-0973-3](https://doi.org/10.1007/s11295-016-0973-3)
- Zeller KA, Creech TG, Millette KL et al (2016) Using simulations to evaluate Mantel-based methods for assessing landscape resistance to gene flow. *Ecol Evol*. doi:[10.1002/ece3.2154](https://doi.org/10.1002/ece3.2154)
- Zhao R, Xia H, Lu BR (2009) Fine-scale genetic structure enhances biparental inbreeding by promoting mating events between more related individuals in wild soybean (*Glycine soja*; Fabaceae) populations. *Am J Bot* 96:1138–1147. doi:[10.3732/ajb.0800173](https://doi.org/10.3732/ajb.0800173)