

# A regression approach to estimate the relative roles of pollen- versus seed-mediated gene flow under an isolation by distance model

M. E. Barrandeguy<sup>1,2,3,\*</sup> and M. V. García<sup>1,2,3</sup>

<sup>1</sup>Departamento de Genética. Facultad de Ciencias Exactas, Químicas y Naturales. Universidad Nacional de Misiones; Posadas 3300 Misiones, Argentina

<sup>2</sup>Instituto de Biología Subtropical Nodo Posadas (UNaM, CONICET) Posadas, Argentina.

<sup>3</sup>Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET - Argentina)

The authors contributed equally to this work.

\*Corresponding author: María Eugenia Barrandeguy, E - mail address: ebarran@fceqyn.unam.edu.ar

## Abstract

In the present work, a new estimator to be known as  $r_{IBD}$  is proposed. The proposed estimator indirectly quantifies the relative role of gene flow mediated by pollen in relation to the gene flow mediated by seeds in hermaphrodite angiosperm species when an isolation by distance model is assumed. The proposed estimator complements the well-known estimator proposed by Ennos, which is appropriate for studies under the island model. In the present work, the proposed  $r_{IBD}$  index was used to analyze microsatellite data from uni- and biparentally inherited genomes generated by simulations, as well as for the analysis of an empirical data set obtained from public databases of forest tree species. The differences in median values of the proposed  $r_{IBD}$  index for simulated data using the Stepping Stone Model and Truncated Pareto Distribution Model coincided with the magnitude expected in terms of differences between levels of pollen and seed dispersion previously established in the simulations. In empirical data, the proposed  $r_{IBD}$  index shows lower levels of gene flow by seed versus gene flow by pollen, the ratio between them being three times lower than the ratio obtained by the Ennos index estimated under the island model. From the analyses carried out, it is feasible to consider the  $r_{IBD}$  index as a suitable estimator of the role of gene flow by seeds in relation to gene flow by pollen under isolation by distance.

**Keywords:** *Gene flow, nuclear microsatellites, chloroplast microsatellites, simulations,  $r_{IBD}$  index, isolation by distance*

## Introduction

The distribution of genetic variation over the space that a species occupies is an important part of population genetic processes (Epperson, 1993). Populations of species that share sufficiently low rates of gene flow become genetically isolated by distance (IBD) (Wright, 1943). In numerous species, dispersal is restricted by space, meaning that there is a higher probability that an individual mates with individuals born in close proximity to themselves than with individuals born further away (Leblois et al., 2003). Dispersal rates and population sizes or densities exert influence on the spatial dimension of a neighborhood, which has been defined as a function of the variance of the parent-offspring dispersal distribution (Crawford, 1984). Neighbourhood size ( $N_b$ ) is a concept that can be used to quantify spatial genetic structure by  $N_b = 4\pi\sigma^2D$ , in which  $\sigma^2$  is the parent-offspring dispersal variance measured around a zero mean and relative to a single reference axis passing through the population, and  $D$  is the density of the population (Wright 1946; 1969). Many studies have attempted to estimate gene dispersal rates and population sizes or densities, using either direct methods (e.g., mark-recapture methods) or indirect methods (genetic markers) (Leblois et al., 2004). Direct estimation of seed and pollen dispersal distances is not easy and requires considerable experimental effort, as the sample will inevitably include all reproductive adults in a well defined area. Indirect approaches have thus been developed to estimate gene flow because, although these methods are less precise, they are much simpler to implement (Ndiade-Bouroubou et al., 2010).

The importance of dispersal processes in shaping the genetic structure is difficult to evaluate in plants. Indeed, the

relationship between genetic structure and gene flow is notably complicated in seed plants by the existence of at least two components of gene movement: movement of pollen between mates and seed dispersal (Oddou-Muratorio et al., 2001). Also, seed and pollen flow may lead to asymmetrical migration of the biparentally inherited (nuclear) and maternally inherited (chloroplast and mitochondrial) genomes in angiosperm species and, in contrast, of the paternally and maternally inherited chloroplast and mitochondrial genomes, respectively, which occur in gymnosperms. These various manners of inheritance produce different levels of population differentiation for each genome (Hu and Ennos, 1997). To infer levels of pollen flow and seed flow indirectly from patterns of genetic structure, it is necessary to monitor genetic markers that permit differentiation between seed and pollen movements. Cytoplasmic markers, being generally uniparentally inherited, provide straightforward markers to track seed or pollen movement (Oddou-Muratorio et al., 2001).

Analyses of genetic differentiation for genomes with varying modes of inheritance contribute to an understanding of the relative importance of pollen and seed dispersal, and they may allow estimates of the relative rates of pollen and seed gene flow (Ennos, 1994; Hu and Ennos, 1997). However, in spite of having profound demographic and genetic consequences for plant populations, relative contribution of contemporary pollen and seed dispersal has not yet been thoroughly investigated (Barluenga et al., 2011).

The first estimator of relative rates of pollen and seed flow was developed by Ennos (1994) and relates expected values of population differentiation ( $F_{ST}$ ) for biparentally, paternally and maternally inherited markers to pollen and seed migration rates. This estimator was derived by assuming an island model of population structure, or the simplest model in which the total population is assumed to be divided into subgroups with the same effective population size, each breeding at random within itself except for a certain proportion of migrants drawn at random from the whole (Wright, 1943). The island model is not likely to be fully and completely carried out in nature. In most cases, the actual immigrants to a population come from immediately surrounding localities with excess populations and thus are not a random sample of the species (Wright, 1943). Whitlock and McCauley (1999) concluded that estimates of the number of migrants between subpopulations based on the  $F_{ST}$  statistic under assumption of an island model at equilibrium were only "likely to be correct within a few orders of magnitude," because assumptions regarding the genetic model (i.e. equal migration, no selection, and demographic stability) are often violated in biological systems. Thus, using an island model to describe pollen and seed movements, even when significant patterns of isolation by distance are detected, is inappropriate (Oddou-Muratorio et al., 2001). This limitation has led to considerations of alternative models with which to derive relative seed and pollen dispersal rates. There are several studies which estimate pollen and seed dispersal contributions under isolation by distance (Hu and Ennos, 1997; Oddou-Muratorio et al., 2001; Ndiade-Bourobou et al., 2010; Dainou et al., 2010; Barluenga et al., 2011). However, in these studies the

role of pollen versus seed gene flow under isolation by distance was not exhaustively estimated.

Ennos's estimator has become one of the most popular indices to show the relationship between pollen versus seed gene flow contribution, although it is not appropriate for species that show isolation by distance. Therefore, the main aim of the present study is to propose a new estimator called  $r_{IBD}$  to determine the relative role of pollen versus seeds gene flow in hermaphrodite plant species under an isolation by distance model. Gene flow by pollen and seeds were estimated by indirect methods using chloroplast and nuclear microsatellite data in order to separate pollen and seed contributions to gene flow. The proposed index arises from estimates of  $F_{ST}$  for pairs of populations rather than a single  $F$  statistic for the entire population. The proposed  $r_{IBD}$  index was tested by simulated and empirical data showing that this index works as more realistic estimator than others that do not consider isolation by distance.

## Materials and Methods

### *Development of the proposed $r_{IBD}$ index from Wright's neighbourhood model*

Wright's neighbourhood model shows that the spatial dimension of a neighbourhood is a function of the variance of the parent-offspring dispersal distribution. If parent-offspring dispersal distances follow a zero mean circular normal distribution with variance  $\sigma^2$  along any axis, the neighbourhood is a circle of radius  $2\sigma$  and has an area  $A = 4\pi\sigma^2$  (Wright, 1946). The effective number of individuals within the neighbourhood area or neighbourhood size ( $N_b$ ) is a critical parameter that generally controls the degree of isolation by distance. This parameter is defined as an area from which the parents of central individuals may be treated as if drawn at random, being  $N_b = AD_e$  where  $D_e$  is the density of breeding individuals (Wright, 1943; 1946), thereby,  $N_b = 4\pi\sigma^2 D_e$ ; used thus,  $N_b$  has been demonstrated as a useful and popular measure of dispersal distance standardized for population density (Epperson, 2007).

Estimates of pollen ( $\sigma_p^2$ ) and seed ( $\sigma_s^2$ ) dispersal variances must be combined to estimate the axial parent-offspring dispersal variance for a biparentally inherited marker, such as nuclear microsatellites. Paternal dispersal is through pollen and seed, whereas maternal dispersal is exclusively through seed in angiosperm species (Figure 1). Male gamete dispersal variance is  $\sigma_p^2$  and female gamete dispersal is zero in higher plants. Therefore, the average gamete dispersal variance is  $1/2(\sigma_p^2 + 1/2\sigma_p^2)$ . Seed dispersal is a post-fertilization event and as such should be regarded as progeny dispersal, so the total parent-offspring dispersal variance is,

$$\sigma_n^2 = \frac{1}{2}\sigma_p^2 + \sigma_s^2 \quad (1a) \quad (\text{Crawford, 1984})$$

whereas for a uniparentally inherited marker the dispersal variance is,

$$\sigma_{cp}^2 = \frac{1}{2}[(1-\alpha)(\sigma_p^2 + \sigma_s^2) + \alpha\sigma_s^2] \quad (2a) \quad (\text{Oddou-Muratorio et al., 2001})$$

being  $\alpha$  the female transmission rate. With strict maternal inheritance of the cytoplasmic genome as chloroplast microsatellites in angiosperms ( $\alpha = 1$ ),

$$\sigma_{cp}^2 = \frac{1}{2}\sigma_s^2 \quad (2b) \text{ (Figure 2)}$$

from (1a)

$$\sigma_p^2 = 2(\sigma_n^2 - \sigma_s^2) \quad (1b)$$

from (2b)

$$\sigma_s^2 = 2\sigma_{cp}^2 \quad (2c)$$

hence, from (1b) and (2c),

$$\sigma_p^2 = 2(\sigma_n^2 - 2\sigma_{cp}^2) \quad (1c)$$

In this way, we propose the index  $r_{IBD}$  in order to estimate the relative role of pollen- versus seed-mediated gene flow in angiosperm species under isolation by distance ( $\sigma_p^2/\sigma_s^2$ ) from the expressions (1c) and (2c) as,

$$r_{IBD} = \frac{\sigma_p^2}{\sigma_s^2} = \frac{(\sigma_n^2 - 2\sigma_{cp}^2)}{\sigma_{cp}^2} \quad (3a)$$

This index is estimated from pollen- versus seed-dispersal variances and allows inference of the relative role of pollen- versus seed-mediated gene flow under isolation by distance.

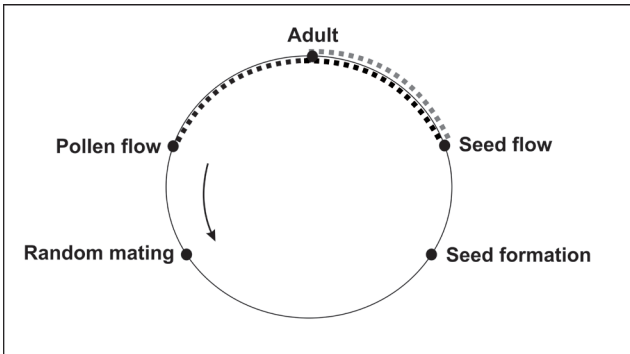


Figure 1

Basic processes of pollen and seed flow occurring among populations within a life cycle of a hermaphrodite plant population. Black dotted line shows dispersal events that can be studied by nuclear markers and gray dotted line shows dispersal event that can be studied by chloroplast markers in angiosperm species. Figure adapted from Hu and Ennos (1999)

### Development of the proposed $r_{IBD}$ index using molecular markers

Rousset (1997) defined a method based on  $F_{ST}$  values for pairs of populations showing that the expected relationship between  $F_{ST}/(1-F_{ST})$  and the logarithm of distance between pairs of populations in a two-dimensional isolation by distance model is approximately linear:  $y=a+bx$  with slope

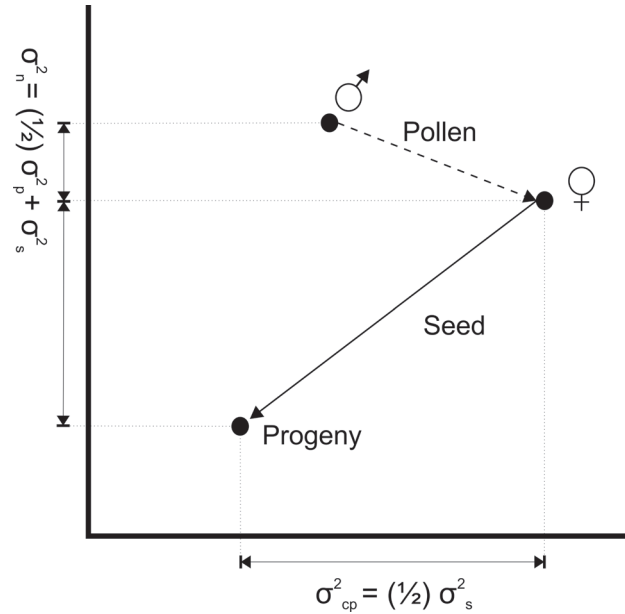


Figure 2

Estimates of pollen and seed dispersal variances,  $\sigma_p^2$  and  $\sigma_s^2$ , from biparental and uniparental inherited microsatellite markers,  $\sigma_n^2$  and  $\sigma_{cp}^2$ , in hermaphrodite plant species

$b=1/(4D\pi\sigma^2)$ . The slope of the regression can be used to estimate  $D\sigma^2$  (Rousset, 1997). This slope does not depend on the spatial scale due to the logarithmic effect of distance; also, it is inversely proportional to the product of population density ( $D$ ) by the second moment of dispersal distance ( $\sigma^2$ ) (Rousset, 1997). Oddou-Muratorio et al. (2001) showed that this method can be used to provide estimates of the relative importance of pollen and seed dispersal if it is applied to markers that show contrasting modes of inheritance. In the present work, patterns of isolation by distance were studied for both nuclear (biparentally inherited) and chloroplast (maternally inherited) DNA markers by the regression of  $F_{STn}/(1-F_{STn})$  and  $F_{STcp}/(1-F_{STcp})$  for pairs of populations against the logarithm of geographic distance between them, respectively. The slope values from these regressions (i.e.  $b_n=1/(4D\pi\sigma_n^2)$  and  $b_{cp}=1/(4D\pi\sigma_{cp}^2)$ ) allow the estimation of the relative role of pollen versus seed gene flow under isolation by distance from equation (3a) [ $r_{IBD}=\sigma_p^2/\sigma_s^2=(\sigma_n^2-2\sigma_{cp}^2)/(\sigma_{cp}^2)$ ]. This derivation is based on the substitution of  $\sigma_n^2$  and  $\sigma_{cp}^2$  with Wright's (1946) neighbourhood size  $N_b$ , used after Epperson (2007) as a measure of dispersal distance standardized for population density. By analogy, with equation (3a) using  $N_b$  we can write  $r_{IBD}=\sigma_p^2/\sigma_s^2=(4\pi D\sigma_n^2-8\pi D\sigma_{cp}^2)/4\pi D\sigma_{cp}^2$  (4a). Simplifying the algebraic expression we obtain  $r_{IBD}=\sigma_p^2/\sigma_s^2=4\pi D(\sigma_n^2-2\sigma_{cp}^2)/4\pi D\sigma_{cp}^2$  (4a') and  $r_{IBD}=\sigma_p^2/\sigma_s^2=(\sigma_n^2-2\sigma_{cp}^2)/\sigma_{cp}^2$  (4a''). Hence we can rewrite the equation (3a) from the regression slopes in the following way:  $r_{IBD}=\sigma_p^2/\sigma_s^2=[1/b_n-2(1/b_{cp})]/(1/b_{cp})$  (4b).

In order to summarize:

$$r_{IBD} = \frac{\sigma_p^2}{\sigma_s^2} = \frac{(\sigma_n^2 - 2\sigma_{cp}^2)}{\sigma_{cp}^2} \quad (3a)$$

$$r_{IBD} = \frac{\sigma_p^2}{\sigma_s^2} = \frac{(4\pi D\sigma_n^2 - 8\pi D\sigma_{cp}^2)}{4\pi D\sigma_{cp}^2} \quad (4a) \quad r_{IBD} = \frac{\sigma_p^2}{\sigma_s^2} = \frac{4\pi D(\sigma_n^2 - 2\sigma_{cp}^2)}{4\pi D\sigma_{cp}^2} \quad (4a')$$

$$r_{IBD} = \frac{\sigma_p^2}{\sigma_s^2} = \frac{\left[\left(\frac{1}{bn}\right) - 2\left(\frac{1}{bcp}\right)\right]}{\left(\frac{1}{bcp}\right)} \quad (4b)$$

Finally, the proposed  $r_{IBD}$  index is represented by equation (4b). This index is estimated from slopes of regression between  $F_{ST}/(1-F_{ST})$  and the logarithm of distance between pairs of populations using nuclear and chloroplast markers, and it allows the inference of the relative role of pollen- versus seed-mediated gene flow under isolation by distance. The proposed  $r_{IBD}$  index assumes absence of heteroplasmy and strict maternal inheritance of the cytoplasmic genome. In addition, a prerequisite for using the proposed  $r_{IBD}$  index is a significant pattern of isolation by distance for both genomes.

### Implementation of the proposed $r_{IBD}$ index in simulated data

Using IBDSim (Leblois et al., 2008) we simulated a population composed by a square grid of 20 x 20 subpopulations with 30 diploid individuals per node. Each individual was characterized by a multilocus genotype defined by ten nuclear and ten chloroplast microsatellite loci. This analysis was based on a model which assumes that the plant species under study is both diploid and hermaphrodite. IBDSim simulates dispersal backwards in time before adding in mutations. The life cycle is as follows: i) gamete production and death of adults; ii) gamete mutations following a generalized stepwise model with rate ( $\mu$ ) of  $10^{-3}$  and  $10^{-4}$  for nuclear and chloroplast microsatellites, respectively, with a maximum number of alleles set to ten and five per nuclear and chloroplast locus, respectively; repeated motif size was set to two and one nucleotide per nuclear and chloroplast microsatellite locus, respectively; iii) gamete dispersal according to a predefined distribution of dispersal distances; iv) constitution of diploid individuals or haplotypes for nuclear and chloroplast genetic data, respectively; v) regulation of the population size ( $n$ ) to ten individuals per node.

Dispersal distances followed two models: a) the Stepping Stone Model (SSM) which assumes discrete populations, discrete number of generations, genetic drift within each population, and migration between adjacent or spatially proximal population (Kimura, 1953; Kimura and Weiss, 1964, Weiss and Kimura, 1965) and b) a Truncated Pareto Distribution (TPD), where the probability of dispersing  $k$  steps in each dimension is given by  $f_k = M/K^n$  for  $k \leq k_{max}$ , being  $M$  the total dispersal rate in one dimension,  $n$  a parameter that controls the shape of the distribution and  $k_{max}$  the maximum dispersal distance (Leblois et al., 2008).

Different values for the parameter  $M$  were considered to obtain a range of dispersal situations that represented nuclear and chloroplast dispersal distribution for plants with high or low levels of gene flow by pollen and seeds under SSM (Figure S1). Different values for the parameters:  $M$ ,  $n$  and  $K_{max}$  were

considered to obtain a range of dispersal situations that represent nuclear and chloroplast dispersal distributions for plants with high or low levels of gene flow by pollen and seeds under TPD model (Figure S1). Thus, simulations differ in the values taken for  $\sigma^2$  (giving the strength of IBD) and also in the nature of the dispersal kernels characterized by  $M$ ,  $n$  and  $K_{max}$ .

Four simulation scenarios were defined which differ in their dispersal conditions (Table 1). Scenarios 1 and 4 evaluate the consequences on the proposed  $r_{IBD}$  index in the case of strong differences between biparental and uniparental dispersion under SSM and TPD models, respectively. In contrast, scenarios 2 and 3 evaluate the consequences on the proposed  $r_{IBD}$  index in the case of weak differences between biparental and uniparental dispersion under SSM and TPD models, respectively. To avoid edge effects, a two-dimensional lattice is represented in a torus (Leblois et al., 2008). At grid edges, we used 'absorbing' boundaries in IBDSim whereby 'the probability mass of going outside the lattice is equally shared on all movements inside the lattice' (Leblois et al., 2012).

The total simulated population was kept constant, but samples were taken from within a smaller area of 6 x 6 subpopulations with ten individuals per node. This sampling strategy was implemented because it is important to restrict the sampling design to a relatively small geographical area in order to work at a local geographical scale (Leblois et al., 2003). Each simulation was replicated 100 times for each genome (i.e. the greatest number to capture almost all of the variance across replicates), giving  $4 \times 200 = 800$  simulations overall to test for IBD and to infer the proposed  $r_{IBD}$  index from simulated genetic data. The statistical significance of the regression slope values was evaluated by a 95 % confidence interval calculated by ABC bootstrap method using Genepop V4.5 software (Rousset, 2008). The proposed  $r_{IBD}$  index was calculated for pairwise nuclear and chloroplast simulations by scenario (100 estimations for each scenario). The spread and skew of the proposed  $r_{IBD}$  index estimated in all simulations by each scenario was shown using box and whisker plots that display a five-number summary: minimum, maximum, median, upper and lower quartiles. The central rectangle spans the first quartile to the third quartile, or the interquartile range (IQR). A segment inside the rectangle shows the median while "whiskers" to the left and to the right show the locations of the minimum and maximum. This simple box plot displays the full range of variation (from min to max), the likely range of variation (the IQR), and a typical value (the median). If the data happens to be normally distributed,  $IQR=1.35\sigma$ , where  $\sigma$  is the standard deviation of data. These estimations were calculated using Microsoft Excel software.

### Implementation of the proposed $r_{IBD}$ index in empirical data

A set of molecular data of *Quercus lobata* was obtained from Dryad digital repository (Gugger et al., 2013a) in order to estimate the proposed  $r_{IBD}$  index in empirical data. *Quercus lobata* is a dominant or co-dominant species in oak savannas, oak woodlands and riparian forests across California. This species

has an extended latitudinal distribution (34–40°) and altitude range from sea level to 1700 m. Gene movement occurs through wind-pollination and the mating system is predominantly outcrossing. In estimates of contemporary gene flow through pollen, average pollen dispersal distance computed with indirect and direct approaches yielded an estimate of 65–114 m, with a high propensity for short distance dispersal (Sork et al., 2010). However, these populations have a fat-tailed dispersal kernel which indicates long distance dispersal as well. Dispersal by birds and rodents typically results in movement of seeds less than 150 m (Sork et al., 2010). Molecular data set includes 267 nuclear microsatellite genotypes across 65 sites (Grivet et al., 2008; Sork et al., 2010) and 190 chloroplast microsatellite haplotypes across 64 sites throughout the species' range (Grivet et al., 2006; Grivet et al., 2008; Sork et al., 2010). These data sets share 44 sites, and they cover the full species' range. Both data sets were obtained using six microsatellite primers (Gugger et al., 2013b). Regression analyses and statistical significance of the regression slope values, evaluated by a 95 % confidence interval calculated with the ABC bootstrap method, were estimated using Genepop V4.5 software (Rousset, 2008). The proposed  $r_{IBD}$  index was calculated from nuclear and chloroplast empirical data sets (Figure 3).

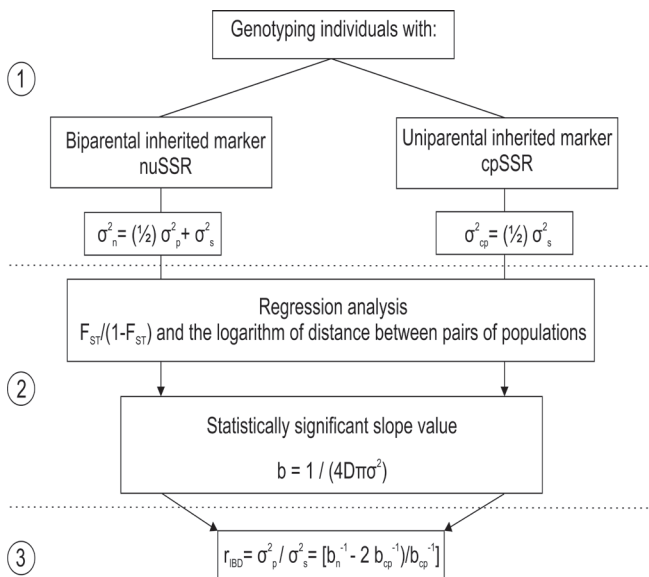


Figure 3

Steps for estimate the proposed  $r_{IBD}$  index: 1) Genotyping with nuclear and chloroplast microsatellite markers, 2) Regression analysis, and 3)  $r_{IBD}$  index estimation

## Results

Implementation of the proposed  $r_{IBD}$  index in simulated data Each simulation gave the genotypes or haplotypes for 360 individuals defined by 10 polymorphic nuclear or chloroplast loci denoted by their coordinates in the lattice, respectively. The slope value of the regression line between  $F_{ST}/(1-F_{ST})$  and

the logarithm of geographical distance was computed for each simulation (Table S1). Average slope values per genome and scenario are shown in Table 1. Slope values were higher for chloroplast simulations than nuclear simulations and standard deviations showed low values indicating little variation around the mean.

Under SSM all estimated slope values were positive and two slope values for chloroplast simulated data were statistically non-significant in scenario 2, consequently they were excluded from the estimation of the proposed  $r_{IBD}$  index. A negative slope value could be understood as an infinite estimate of  $D\sigma^2$  (Leblois et al., 2003) while the positive slope values of simulated data could be considered as evidence of isolation by distance in both genomes (i.e. a prerequisite for the estimation of the proposed  $r_{IBD}$  index). From nuclear and chloroplast slope values the proposed  $r_{IBD}$  index was calculated using equation (4b). Seven values of the proposed  $r_{IBD}$  index were zero in scenario 2. The median value of the proposed  $r_{IBD}$  index in scenario 1 was 30.5, while it was two in scenario 2 (Figure 4).

Under the TPD model all estimated slope values were positive and only three values for chloroplast simulated data were statistically non-significant. For nuclear simulated data under the TPD model, three values were negative and 31 values were statistically non-significant. Consequently, they were excluded from the calculation of the proposed  $r_{IBD}$  index. From nuclear and chloroplast slope values, the proposed  $r_{IBD}$  index was calculated using equation (4b). The median value of the proposed  $r_{IBD}$  index in scenario 3 was eight while it was 23 in scenario 4 (Figure 4).

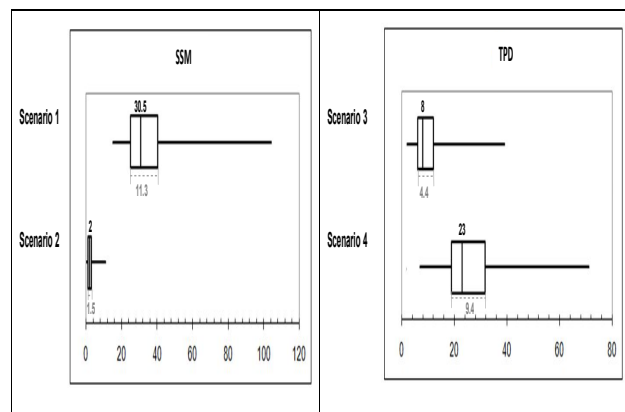


Figure 4

Box-whisker plot of  $r_{IBD}$  proposed index for simulated scenarios. Median values are showed above boxes and standard deviations are showed below boxes. SSM= Stepping Stone Model; TPD= Truncated Pareto Distribution model

The box and whisker plots for each simulated scenario allowed estimation of the median, the most probable values and the dispersion of the estimates of the proposed  $r_{IBD}$  index by scenario under the different dispersion conditions together with the theoretical simulated IBD models (SSM and TPD). The

differences in median values of the proposed  $r_{IBD}$  index using both models coincide with the magnitude expected in terms of differences between levels of pollen and seed dispersion previously established in the simulations. That is, when the differences in the pollen and seed dispersion distances are greater (Scenarios 1 and Scenario 4) the proposed  $r_{IBD}$  index took values of 30.5 and 23 in the SSM and TPD models, respectively. In the same way, when the differences in pollen and seed dispersion distances are lower (Scenarios 2 and Scenario 3), the proposed  $r_{IBD}$  index took values of 2 and 8 under the SSM and TPD models, respectively.

Table 1  
Description of considered scenarios

Scenario	Distribution model	Genome	Dispersal distribution conditions	Difference in $M$ between nuclear vs chloroplast
1	Stepping Stone model	Nuclear	$M = 0.400$	0.375
		Chloroplast	$M = 0.025$	
2	Stepping Stone model	Nuclear	$M = 0.400$	0.200
		Chloroplast	$M = 0.200$	
3	Truncated pareto distribution	Nuclear	$M = 0.200$ $n = 2$ $k_{max} = 5$	0.100
		Chloroplast	$M = 0.100$ $n = 3.300$ $k_{max} = 1$	
4	Truncated pareto distribution	Nuclear	$M = 0.200$ $n = 2$ $k_{max} = 5$	0.160
		Chloroplast	$M = 0.040$ $n = 3.3$ $k_{max} = 1$	

$M$ = total dispersal rate in one dimension,  $N$ = parameter that controls the shape of the distribution,  $k_{max}$ = maximum dispersal distance

### Implementation of the proposed $r_{IBD}$ index in empirical data

The analysis of regression performed from *Q. lobata* chloroplast and nuclear microsatellite data result in statistically significant and positive slope values (Table S1 and Table 2). These results confirm isolation by distance in both genomes enabling the estimation of the proposed  $r_{IBD}$  index. The proposed  $r_{IBD}$  index reached a value of 19.222, demonstrating that pollen flow is approximately nineteen times higher than seed flow.

## Discussion

Under the island model, differences in dispersal distances do not produce differences in the relative importance of gene flow by pollen vs gene flow by seeds; however under the isolation by distance model, differences in dispersal distances impact the spatial patterns of distribution of alleles of chloroplast and nuclear genomes that can be detected via molecular

Table 2  
Average regression slope for simulated genetic data by scenario and regression slope for empirical system

System	Scenario	Genome	Average regression slope (SD)
Simulated	1	Nuclear	0.010 (0.002)
		Chloroplast	0.334 (0.079)
	2	Nuclear	0.010 (0.002)
		Chloroplast	0.041 (0.011)
	3	Nuclear	0.008 (0.002)
		Chloroplast	0.162 (0.023)
	4	Nuclear	0.008 (0.002)
		Chloroplast	0.202 (0.056)
Empirical	<i>Q. lobata</i>		<b>Regression slope</b>
		Nuclear	0.031
		Chloroplast	0.658

markers. Some authors estimate the ratio of pollen vs seed gene flow taking into consideration the IBD model. Hu and Ennos (1997) develop a theory for indirectly estimating the ratio of pollen to seed flow among plant populations by a variety of methods and conclude, "In the isolation-by-distance case it is possible to obtain analytical expressions for estimating this ratio but in practice this formula will be very difficult to apply. In the first place it requires estimates of neighbourhood size for the three different genomes, these are difficult to measure in the field. The model also assumes a random mating population, reaching an infinite number of generations back to its ancestors." Oddou-Muratorio et al. (2001) derive an estimator of pollen/seed migration ratio under isolation by distance based on Rousset's theoretical frame. However, the species under study was androdieious, and consequently its proposed estimator takes into account the unequal contribution of the two sexes to the next generation. Dainou et al. (2010) proposed a method based on the regression of pairwise kinship coefficients between individuals on the logarithm of the distance to estimate the  $Sp$  statistic. In dioecious plants with a balanced sex ratio,  $Sp$  statistics are expected to approach  $1/(4\pi D\sigma_g^2)$  for nuclear and  $1/(\pi D\sigma_s^2)$  for chloroplast markers under IBD. However, for the computation of the  $Sp$  statistics, the slope must be assessed for pre-establish spatial scale. Following a similar method based on fine scale genetic structure, Ndiade-Bouroubo et al. (2010) proposed the  $Sp$  statistic for the estimation of pollen and seed gene flow which also requires the knowledge of effective population density. Finally, Barluenga et al. (2011) estimated the relative contributions of pollen-mediated and seed-mediated gene flow by comparing the patterns revealed by autosomal vs. X-linked microsatellite markers. Despite its vital importance for an understanding of species biology, as well as for the interpretation of population genetic structure, we have not found recent studies that attempt to resolve the estimation of the proportion of gene flow by pollen vs seeds; therefore, in this study we propose a new analytical expression with its respective practical solution using molecular markers.

The most commonly used measures of population genetic structure is Wright's  $F_{ST}$  (Holsinger and Weir, 2009; Nielsen and Slatkin, 2013). This statistics provides a measure of genetic structure by describing the genetic diversity partitioned within and among populations (Wright, 1951). As noted by Holsinger and Weir (2009),  $F_{ST}$  is much more than a simple measure of genetic differentiation because it directly infers the variance in allele frequencies among populations. Several related statistics have been described since Wright's  $F_{ST}$ . For example,  $R_{ST}$  is commonly used for microsatellite data,  $\Phi_{ST}$  for sequence data and  $G_{ST}$  a multi-allele generalization of  $F_{ST}$  (Holsinger and Weir, 2009, Jost et al., 2018). These measures and their relatives collectively share many properties and measure a kind of demographic differentiation (Jost et al., 2018). When defined as a ratio of genetic variances (Cockerham, 1973),  $F_{ST}$  and its related statistics work by relating the amount of genetic variation among populations to the total genetic variation over all populations. For biallelic markers, this makes sure that  $F_{ST}$  is bounded between zero and one, with zero representing no differentiation and one representing fixation of different alleles within populations. For multiallelic markers, however, the maximum possible value is determined by the amount of within-population diversity (Charlesworth, 1998; Hedrick, 1999). Furthermore, Jakobsson et al. (2013) shown that not only does diversity constrain the range of  $F_{ST}$ , the frequency of the most frequent allele has a strong influence on the values that  $F_{ST}$  can take. Nowadays, a large number of different marker types are available for population genetic studies, with a large range of allelic diversities, from SNPs that are essentially biallelic to microsatellites that can have several alleles. The dependency of many  $F_{ST}$  estimators on the level of diversity could cause difficulties in their interpretation especially the case when markers are compared that have different mutation rates or when species are compared with different effective population sizes (Meirmans y Hedrick, 2011).

At present, there are two broad groups of measures that quantify population structure—fixation measures like  $F_{ST}$ ,  $G_{ST}$  and  $\Phi_{ST}$ , and allelic differentiation measures, like Jost's  $D$  and entropy differentiation (Jost et al., 2018). In the literature, there has been much discussion of the purpose, application, and performance of these two groups of metrics (e.g. Gerlach et al., 2010; Leng and Zhang, 2011; Heller and Siegismund, 2009; Jost, 2008; Meirmans and Hedrick, 2011; Whitlock, 2011). These two groups have sometimes been treated as rivals, or as if members of the second family are “correcting” or estimating the members of the first however the two groups are designed to quantify complementary aspects of population structure. Thus, attempts to compare them as if they are measuring the same feature are improper (Jost et al., 2018).

In this context,  $F_{ST}$  remains of particular interest on the basis of its long history of use in population genetics and its connection to features of biological models (Whitlock, 2011; Jakobsson et al., 2013).

Despite of considerations about measures of genetic structure and the dependency of  $F_{ST}$  on genetic diversity the development of the proposed  $r_{IBD}$  index is based on Rousset method which consider that the expected relationship

between  $F_{ST}/(1-F_{ST})$  and the logarithm of distance between pairs of populations in a two-dimensional IBD model is approximately linear with slope  $b=1/(4D\pi\sigma^2)$  (Rousset, 1997). Actually, IBD model proposed by Wright (1943) is still one of the most commonly used and Jenkings et al. (2010) supported that IBD model continue as the simplest reference standard against which all other, more complex models should be compared in landscape genetics research.

Even the  $F_{ST}$  is the first and the most commonly used index to measure genetic differentiation, it can be estimated for chloroplast and nuclear markers with several free available software and it is easy to interpret, its dependency from the level of diversity is a serious subject to analyze in future studies in order to avoid possible pitfalls in the estimation of the proposed  $r_{IBD}$  index. Hence, the proposed  $r_{IBD}$  index is appropriate when its results are thoroughly analyzed and interpreted carefully.

In the present study we developed the  $r_{IBD}$  index, employed this index with simulated genetic data in order to understand the impact of different dispersal distances on its magnitude; and finally we calculate this index with empirical data in order to compare the results with the Ennos estimator based on the island model.

Simulations help to recreate the stochastic process that accompanies the transmission of genes from parents to offspring because they allow replication of the movement of alleles in space under the same dispersal condition several times. Any model (analytical, simulation, and otherwise) makes simplifying assumptions, barring that it be “an entire reconstruction of the actual system—whereupon it ceases to be a model” (Epperson et al., 2010). Simple models remain useful, however, for capturing some salient aspects of complex systems or problems, developing theories, testing hypotheses, and making predictions. More complex models tend to be more difficult to understand and derive general conclusions from. SSM are the simplest models of isolation by distance in which gene flow occurs between adjacent populations. The TPD model includes more parameters to define the dispersal distribution, thereby simulations under this model may be more similar to biological systems. Furthermore, many scenarios dealing with heterogeneities in dispersal rates between biparentally and uniparentally inherited loci may be performed. In the present study we chose to test two alternative scenarios for each IBD model in order to see how differences between nuclear and chloroplast dispersion affected the proposed  $r_{IBD}$  index.

Under SSM, similar emigration rates between nuclear and chloroplast genomes resulted in a high frequency of zero values for the proposed  $r_{IBD}$  index in scenario 2. Zero or negative values for the proposed  $r_{IBD}$  index mean that seed gene flow exceeds pollen gene flow. Highly skewed pollen/seed migration ratios may be detected only in the case that pollen gene flow is predominant, as in many of the species studied so far (Oddou-Muratorio et al., 2001). Therefore, a limitation of the method is that the proposed  $r_{IBD}$  index is useful for cases in which a low absolute level of seed migration is present. Since pollen migration rates are generally believed to be greater than seed migration rates, this assumption is likely to be

realistic (Ennos, 1994). The median value of the proposed  $r_{IBD}$  index was according to the differences between biparental and uniparental dispersion by scenario in the simulated system. Hence, in cases of high differences for M values between nuclear and chloroplast genomes, the proposed  $r_{IBD}$  index reached higher values than in cases of similar M values between nuclear and chloroplast genomes.

Under the TPD model, nuclear genome simulations resulted in a high frequency of statistically non-significant slope values. This result indicates absence of isolation by distance and thus could not be used for the estimation of proposed  $r_{IBD}$  index. The proposed  $r_{IBD}$  index was estimated using statistically significant results around 67 times in each scenario reaching values that showed that under TPD models, differences in uniparental and biparental dispersal conditions impact the proposed  $r_{IBD}$  index in the same way as under SSM, namely decreasing differences between pollen and seed gene flow as differences in dispersal distribution decrease, as well.

Repeated simulations of chloroplast and nuclear genetic data under each scenario allow consideration of stochastic differences in the movement of alleles under the same conditions; namely, allowing consideration of the expected pattern of dispersion of the proposed  $r_{IBD}$  index under each scenario and model. Genetic transmission, or the process by which genes are transmitted from parents to offspring, is also highly stochastic, even if it follows simple mating rules and Mendelian segregation (Epperson et al., 2010). From box and whisker plots it is possible to gather a quick understanding of the distribution of a dataset. From the results of the analysis of simulated data, a main role could be assigned to pollen as the agent responsible for gene flow in species with  $r_{IBD}$  values greater than 10, while  $r_{IBD}$  values less than 10 would allow the inference that both pollen and seeds act as agents responsible for the allelic exchange between individuals and populations, determining the distribution of genetic variability within and between populations.

In the empirical system analyzed, previous spatial autocorrelation analyses were consistent with a higher spatial genetic structure in the chloroplast than in the nucleus for *Q. lobata*. In this species, the spatial pattern most likely reflects restricted gene flow for both genomes (Grivet et al., 2006; 2008). The ratio of pollen to seed gene flow under the island model (Ennos, 1994) was around 60 for *Q. lobata*, indicating greater gene flow through pollen than gene flow by seeds (Grivet et al., 2008). However, Dutech et al. (2005) detected high kinship coefficients for short distance classes by means of spatial autocorrelation analysis, and it was compatible with very restricted pollen and seed dispersal. It is generally assumed that seed movement is more restricted than pollen movement in oaks (Ducousso et al., 1993). Despite the proposed  $r_{IBD}$  index's showing lower levels of gene flow by seed versus gene flow by pollen ( $r_{IBD} = 19$ ) the ratio between them is three times lower than the ratio obtained by the Ennos index estimated under the island model ( $r_{IM} = 60$ ). Since isolation by distance was detected in the analyzed populations of *Q. lobata*, the proposed  $r_{IBD}$  index seems appropriate. The result of the proposed  $r_{IBD}$  index for the empirical data is greater than 10, thus pollen could be

seen as the agent responsible for the gene flow of the species. In addition, the proposed  $r_{IBD}$  index seems compatible with the biological traits related to dispersal mechanisms and reproduction of *Q. lobata*. Also, it is important to take into account that genetic change in a population depends on both the amount of gene flow from other populations and the correlations of genetic variation between the source populations and the recipient population (Epperson, 1993). In a recent work, Van Strien et al. (2015) showed that IBD patterns should not be regarded as resulting only from distance-constrained migration, but also from the deme topology and habitat configuration. These authors found that the distance at which the highest  $F_{ST}$ -distance correlations were measured did not correspond to the maximum migration distance, and recommended that IBD and landscape effects on gene flow be assessed separately. From this study, a future challenge for the methodology proposed in the present study will be to abandon the regular lattice setup of demes and instead use irregular habitat configurations and demes topologies which better reflect natural systems. Conservation biologists have become concerned about the impact of human disturbance on gene flow because processes such as fragmentation and demographic erosion can increase genetic isolation and decrease local population size in a manner that may jeopardize the viability of remaining populations (Dutech et al., 2005). Population size, neighbourhood size, and gene flow between populations are all important factors to long term survival among endangered species (Tero et al., 2005).

## Conclusion

We estimated the relative role of pollen and seed gene flow among populations under isolation by distance according to the equation  $r_{IBD} = \sigma_p^2 / \sigma_s^2 = [1/b_n - 2(1/b_{cp})] / (1/b_{cp})$ , which we developed ourselves, instead of using the island model and the corresponding equation  $r_{IM} = [(1/F_{STn} - 1) - 2(1/F_{STcp} - 1)] / (1/F_{STcp} - 1)$  developed by Ennos (1994). Our proposed method requires nuclear and chloroplast markers determined in the same populations under study, although both markers are not required for the same individual.

It has been suggested that the transformation of measurements of population structure into quantitative estimates of gene flow or dispersal are, at best, not necessary, and at worst, often ambiguous and open to multiple interpretations (Whitlock and McCauley, 1999; Oddou-Muratorio et al., 2001). Most, though not all, of this criticism applies to the island model. Other authors have argued that the availability of alternative methods to derive estimates of the parameters, especially when these methods are based on different assumptions, constitutes a clear progress (Sork et al., 1999; Neigel 1997, Oddou-Muratorio et al., 2001). In this way, the proposed  $r_{IBD}$  index represents an appropriate statistic to show the role of pollen versus seed gene flow in hermaphrodite species when isolation by distance is detected. The next step could be the



systematization of the  $r_{IBD}$  index by creating software that estimates this index from the molecular and geographical data.

The proposed  $r_{IBD}$  index should be analyzed as an estimator of the relative role of the realized pollen and seed gene flow, and as a consequence, the relation between them should not be seen to mark merely the simple differences arising from dispersal conditions, but instead should be expected to result from differences in the effective function of pollen and seed flow as micro-evolutive homogenizing forces influenced by species biology (pollen and seed morphological traits) and several other features, i.e. genetic (mating system, effective size, mode of inheritance, genetic drift, etc.), ecological (relationships with pollinators and seed disperser, relationships among species in the community, etc.), demographic (spatial fragmentation, population expansion or bottleneck) and landscape. This complexity leads to understanding the estimated  $r_{IBD}$  index value as a complex genetic estimator, rather than a simple ratio directly dependent on biparental and uniparental dispersal distribution conditions.

## Acknowledgement

This work was supported by Agencia Nacional de Promoción Científica y Tecnológica (AGENCIA) and Universidad Nacional de Misiones (UNaM) [PICTO UNaM 2011 N°0133]. The authors wish to thank the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina). We thank the referees for his helpful comments because his time and effort invariably improved our manuscript.

## Supplementary material

Figure S1: Dispersal distribution conditions for simulate data  
Table S1: Regression results and proposed  $r_{IBD}$  index for nuclear and chloroplast simulated genetic data and for empirical data

## References

- Barluenga M, Austerlitz F, Elzinga J A, Teixeira S, Goudet J, Bernasconi G (2011) Fine-scale spatial genetic structure and gene dispersal in *Silene latifolia*. *Heredity* 106: 13–24. <https://doi.org/10.1038/hdy.2010.38>
- Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Mol Biol Evol* 15: 538–543. <https://doi.org/10.1093/oxfordjournals.molbev.a025953>
- Cockerham CC (1973) Analysis of gene frequencies. *Genetics* 74: 679–700.
- Crawford T J (1984) The estimation of neighbourhood parameters for plant populations. *Heredity* 52(2): 273–283. <https://doi.org/10.1038/hdy.1984.29>
- Dainou K, Bizoux J, Doucet J, Mahy G, Hardy O J, Heuert M (2010) Forest refugia revisited: nSSRs and cpDNA sequences support historical isolation in a wide-spread African tree with high colonization capacity, *Milicia excelsa* (Moraceae). *Mol Ecol* 19: 4462–4477. <https://doi.org/10.1111/j.1365-294x.2010.04831.x>
- Ducoussou A, Michaud H, Lumaret R (1993) Reproduction and gene flow in the genus *Quercus* L. *Ann Sci Forest* 50(11): 91–106. <https://doi.org/10.1051/forest:19930708>
- Dutech C, Sork V L, Irwin A J, Smouse P E, Davis F W (2005) Gene flow and fine-scale genetic structure in a wind-pollinated tree species, *Quercus lobata* (Fagaceae). *Am J Bot* 92(2): 252–261. <https://doi.org/10.3732/ajb.92.2.252>
- Ennos, R A (1994) Estimating the relative rates of pollen and seed migration among plant populations. *Heredity* 72: 250–259. <https://doi.org/10.1038/hdy.1994.35>
- Epperson B K, Mcrae B H, Scribner K, Cushman S A, Rosenberg M S, Fortin M J, James P M, Murphy M, Manel S, Legendre P, Dale M R (2010) Utility of computer simulations in landscape genetics. *Mol Ecol* 19: 3549–3564. <https://doi.org/10.1111/j.1365-294x.2010.04678.x>
- Epperson, B K (1993) Spatial and spatial-time correlations in systems of subpopulations with genetic drift and migration. *Genetics* 133: 711–727.
- Epperson, B K (2007) Plant dispersal, neighbourhood size and isolation by distance. *Mol Ecol* 16: 3854–3865. <https://doi.org/10.1111/j.1365-294x.2007.03434.x>
- Gerlach G, Jueterbock A, Kraemer P, Deppermann J, Harmand P (2010) Calculations of population differentiation based on GST and D: Forget GST but not all of statistics! *Mol Ecol* 19(18): 3845–3852.
- Grivet D, Deguilloux M F, Petit R J, Sork V L (2006) Contrasting patterns of historical colonization in white oaks (*Quercus* spp.) in California and Europe. *Mol Ecol* 15: 4085–4093. <https://doi.org/10.1111/j.1365-294x.2006.03083.x>
- Grivet D, Sork V L, Westfall R D, Davis F W (2008) Conserving the evolutionary potential of California valley oak (*Quercus lobata* Née): a multivariate genetic approach to conservation planning. *Mol Ecol* 17: 139–156. <https://doi.org/10.1111/j.1365-294x.2007.03498.x>
- Gugger P F, Ikegami M, Sork V L (2013a) Data from: Influence of late Quaternary climate change on present patterns of genetic variation in valley oak, *Quercus lobata* Née. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.g645d>
- Gugger P F, Ikegami M, Sork V L (2013b) Influence of late Quaternary climate change on present patterns of genetic variation in valley oak, *Quercus lobata* Née. *Mol Ecol* 22(13): 3598–3612. <https://doi.org/10.1111/mec.12317>
- Hedrick P W (1999) Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution* 53: 313–318. <https://doi.org/10.1111/j.1558-5646.1999.tb03767.x>
- Heller R, Siegmund H R (2009) Relationship between three measures of genetic differentiation GST, DEST and G'ST: How wrong have we been? *Mol Ecol* 18(10): 2080–2083. <https://doi.org/10.1111/j.1365-294x.2009.04185.x>
- Holsinger K E, Weir B S (2009) Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nat Rev Genet* 10:639–650. <https://doi.org/10.1038/nrg2611>
- Hu X, Ennos R A (1997) On estimation of the ratio of pollen to seed flow among plant populations. *Heredity* 79: 541–552. <https://doi.org/10.1038/sj.hdy.6882080>
- Hu X, Ennos R A (1999) Impacts of seed and pollen flow on population genetic structure for plant genomes with three contrasting modes of inheritance. *Genetics* 152: 441–450.
- Jakobsson M, Edge M D, Rosenberg N A (2013) The relationship between FST and the frequency, <https://doi.org/10.1534/genetics.112.144758>
- Jenkins D G, Carey M, Czerniewska J, Fletcher J, Hether T, Jones A, Knight S, Knox J, Long T, Mannino M, McGuire M, Riffle A, Segelsky S, Shappell L, Sterner A, Strickler T, Tursi R (2010) A meta-analysis of isolation by distance: relic or reference standard for landscape genetics? *Ecography* 33: 315–320. <https://doi.org/10.1111/j.1600-0587.2010.06285.x>
- Jost L (2008) GST and its relatives do not measure differentiation. *Molecular Ecology* 17(18): 4015–4026. <https://doi.org/10.1111/j.1365-294x.2008.03887.x>
- Jost L, Archer F, Flanagan S, Gaggiotti O, Hoban S, Latch E (2018) Differentiation measures for conservation genetics. *Evolutionary Applications* 1–10. <https://doi.org/10.1111/eva.12590>
- Kimura M (1953) "Stepping stone" model of population. *Annu Rep Natio Inst Genet.* 3: 62–63.
- Kimura M, Weiss G H (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* 49: 561–576.
- Leblois R, Beeravolu C R, Rousset F (2012) IBDSim version 2.0 User manual.

- Leblois R, Estoup A, Rousset F (2003) Influence of mutational and sampling factors on the estimation of demographic parameters in a "continuous" population under isolation by distance. *Mol Biol Evol* 20(4): 491-502. <https://doi.org/10.1093/molbev/msg034>
- Leblois R, Estoup A, Rousset F (2008) IBDsim: a computer program to simulate genotypic data under isolation by distance. *Mol Ecol Res* 9(1): 107-109. <https://doi.org/10.1111/j.1755-0998.2008.02417.x>
- Leblois R, Rousset F, Estoup A (2004) Influence of spatial and temporal heterogeneities on the estimation of demographic parameters in a continuous population using individual microsatellite data. *Genetics* 166: 1081-1092. <https://doi.org/10.1534/genetics.166.2.1081>
- Leng L, Zhang D X (2013) Time matters: Some interesting properties of the population differentiation measures GST and D overlooked in the equilibrium perspective. *Journal of Systematics and Evolution* 51(1): 44-60. <https://doi.org/10.1111/j.1759-6831.2012.00231.x>
- Luximon N, Petit E J, Broquet T (2014) Performance of individual vs. group sampling for inferring dispersal under isolation-by-distance. *Mol Ecol Res* 14(4): 745-752. <https://doi.org/10.1111/1755-0998.12224>
- Meirmans P G, Hedrick P W (2011) Assessing population structure: FST and related measures. *Molecular Ecology Resources* 11(1): 5-18. <https://doi.org/10.1111/j.1755-0998.2010.02927.x>
- Ndiade-Bourboubou D, Hardy O J, Favreau B, Moussavou H, Nzengue E, Mignot A, Bouvet J M (2010) Long-distance seed and pollen dispersal inferred from spatial genetic structure in the very low-density rainforest tree, *Baillonella toxisperma* Pierre, in Central Africa. *Mol Ecol* 19: 4949-4962. <https://doi.org/10.1111/j.1365-294x.2010.04864.x>
- Neigel J E (1997) A comparison of alternative strategies for estimating gene flow from genetic markers. *Annu Rev Ecol Evol Syst* 28: 105-128. <https://doi.org/10.1146/annurev.ecolsys.28.1.105>
- Nielsen R, Slatkin M (2013) *An introduction to Population Genetics. Theory and Applications*. Sinauer Associates, Inc. Publishers. Sunderland, Massachusetts U.S.A. <https://doi.org/10.1086/673812>
- Oddou-Muratorio S, Petit R J, Le Guerroue B, Guesnet D, Demesure B (2001) Pollen- versus seed-mediated gene flow in a scattered forest tree species. *Evolution* 55(6): 1123-1135. [https://doi.org/10.1554/0014-3820\(2001\)055\[1123:pysmgf\]2.0.co;2](https://doi.org/10.1554/0014-3820(2001)055[1123:pysmgf]2.0.co;2)
- Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145: 1219-1228.
- Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol Ecol Res* 8: 103-106. <https://doi.org/10.1111/j.1471-8286.2007.01931.x>
- Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* 236: 787-792. <https://doi.org/10.1126/science.3576198>
- Sork V L, Davis F W, Westfall R, Flint A, Ikegami M, Wang H, Grivet D (2010) Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Née) in the face of climate change. *Mol Ecol* 19: 3806-3823. <https://doi.org/10.1111/j.1365-294x.2010.04726.x>
- Sork V L, Nason J, Campbell D R, Fernandez J F (1999) Landscape approaches to historical and contemporary gene flow in plants. *Trends Ecol Evol* 14(6): 219-224. [https://doi.org/10.1016/s0169-5347\(98\)01585-7](https://doi.org/10.1016/s0169-5347(98)01585-7)
- Tero N, Aspi J, Siikamäki P, Jäkäläniemi A (2005) Local genetic population structure in an endangered plant species, *Silene tatarica* (Caryophyllaceae). *Heredity* 94: 478-487. <https://doi.org/10.1038/sj.hdy.6800642>
- van Strien M J, Holderegger R, Van Heck H J (2015) Isolation-by-distance in landscapes: considerations for landscape genetics. *Heredity* 114: 27-37. <https://doi.org/10.1038/hdy.2014.62>
- Weiss G H, Kimura M (1965) A mathematical analysis of the stepping stone model of genetic correlation. *Appl Probab* 2: 129-149. <https://doi.org/10.2307/3211879>
- Whitlock M C (2011). G'ST and D do not replace FST. *Molecular Ecology* 20(6): 1083-1091. <https://doi.org/10.1111/j.1365-294x.2010.04996.x>
- Whitlock M C, McCauley D E (1999) Indirect measures of gene flow and migration:  $F_{ST} \neq 1/(4Nm+1)$ . *Heredity* 82: 117-125. <https://doi.org/10.1038/sj.hdy.6884960>
- Wright S (1943) Isolation by distance. *Genetics* 28: 114-138.
- Wright S (1946) Isolation by distance under diverse systems of mating. *Genetics* 31: 39-59.
- Wright S (1969) *Evolution and the genetics of population Vol. 2 The theory of gene frequencies*. University of Chicago Press, Chicago. <https://doi.org/10.1126/science.168.3932.722>
- Wright, S (1951) The genetical structure of populations. *Annals of Eugenics* 15: 323-354. <https://doi.org/10.1111/j.1469-1809.1949.tb02451.x>