Genetica (2013) 141:479–489
DOI 10.1007/s10709-013-9747-0

A new eigenfunction spatial analysis describing population genetic structure

José Alexandre Felizola Diniz-Filho · João Vitor Barnez P. L. Diniz · Thiago Fernando Rangel · Thannya Nascimento Soares · Mariana Pires de Campos Telles · Rosane Garcia Collevatti · Luis Mauricio Bini

Received: 12 September 2012 / Accepted: 18 October 2013 / Published online: 27 October 2013

Abstract Several methods of spatial analyses have been proposed to infer the relative importance of evolutionary processes on genetic population structure. Here we show how a new eigenfunction spatial analysis can be used to model spatial patterns in genetic data. Considering a sample of n local populations, the method starts by modeling the response variable (allele frequencies or phenotypic variation) against the eigenvectors sequentially extracted from a geographic distance matrix ($n \times n$). The relationship between the coefficient of determination (R^2) of the models and the cumulative eigenvalues, which we named the spatial signal-representation (SSR) curve, can be more efficient than Moran's I correlograms in describing different patterns. The SSR curve was also applied to simulated data (under distinct scenarios of population differentiation) and to analyze spatial patterns in alleles from microsatellite data for 25 local populations of *Dipteryx alata*, a tree species endemic to the Brazilian Cerrado. The SSR curves are consistent with previous phylogeographical patterns of the species, revealing combined effects of isolation-by-distance and range expansion. Our analyses demonstrate that the SSR curve is a useful exploratory tool for describing spatial patterns of genetic variability and for selecting spatial eigenvectors for models aiming to explain spatial responses to environmental variables and landscape features.

Keywords Cerrado · *Dipteryx alata* · Eigenfunction analyses · Microsatellites · Spatial autocorrelation · Spatial genetic structure

J. A. F. Diniz-Filho (✉) · T. F. Rangel · L. M. Bini
Laboratório de Ecologia Teórica e Síntese, Departamento de Ecologia, Instituto de Ciências Biológica (ICB), Universidade Federal de Goiás (UFG), C.P. 131, Goiânia, GO 74001-970, Brazil
e-mail: diniz@ufg.br; jafdinizfilho@gmail.com

J. V. B. P. L. Diniz
Graduação em Engenharia da Computação, Universidade Federal de Goiás (UFG), Goiânia, Brazil

T. N. Soares · M. P. C. Telles · R. G. Collevatti
Laboratório de Genética e Biodiversidade, Departamento de Biologia Geral, Instituto de Ciências Biológica (ICB), Universidade Federal de Goiás (UFG), Goiânia, Brazil

Introduction

Natural selection, adaptation, and the balance between gene flow and genetic drift create complex spatial patterns in genetic data (Epperson 2003; Rousset 2004). Thus, several methods incorporating the spatial arrangement of local populations under study have been developed to detect and infer the evolutionary processes underlying these patterns (Sokal and Oden 1978a, b; Slatkin and Arter 1991; Sokal et al. 1997; see Guillot et al. 2009; Balkenhol et al. 2009; Diniz-Filho and Bini 2012; Wagner and Fortin 2013 for recent methodological reviews).

Recently, methods such as spatial autocorrelation, Bayesian clustering, Wombling, Monmonier and multivariate analyses have been used within the field of landscape genetics to describe patterns of genetic variation, to detect spatial boundaries, to infer gene flow limitation (Manel et al. 2003; Holderegger and Wagner 2006, 2008; Kelly et al. 2010; Wagner and Fortin 2013) and which dispersal routes between local populations better explain genetic similarity variation (Spear et al. 2005; Storfer et al. 2007; Meister et al. 2010; Croucher et al. 2011). Also, these methods allow associating spatial patterns of genetic variation with landscape characteristics, which have

usually been modified by human activities (Meister et al. 2010; Manel et al. 2010a; Segelbacher et al. 2010; Holderegger et al. 2010 for recent reviews).

Eigenfunction spatial analyses can also be used to explore spatial patterns of genetic variation and to take spatial autocorrelation into account in linear modeling, evaluating the relative explanatory power of environmental and spatial components (and their shared importance) at distinct spatial scales (see Borcard and Legendre 2002; Griffith 2003; Desdevises et al. 2003; Diniz-Filho and Bini 2005; Dray et al. 2006; Griffith and Peres-Neto 2006; Dormann et al. 2007; Peres-Neto and Legendre 2010; Manel et al. 2010b; Logue et al. 2011; Bertin et al. 2012; Diniz-Filho et al. 2012a, b). However, despite its usefulness and flexibility, there have been only a few attempts to use these approaches to infer evolutionary processes from spatial patterns in molecular genetic data (see Diniz-Filho et al. 2009; Manel et al. 2010b; Bertin et al. 2012).

In standard applications of eigenfunction spatial analyses, response variables are regressed against a selected set of eigenvectors extracted from pairwise matrices containing different transformations of geographic distances or spatial connectivity among spatial units (see Griffith and Peres-Neto 2006; Dray et al. 2006). The coefficient of determination (R^2) of this regression model indicates the amount of spatial patterns in the data. Also, the effects of environmental or landscape features on response variables may be assessed after taking the selected eigenvectors (i.e., spatial variables) into account in more complex models (Peres-Neto and Legendre 2010). However, results from eigenfunction spatial analyses to estimate spatial dependence or to control for this dependence in analyses of environmental effects depend on how the eigenvectors are selected. There is no single consensual criterion to select eigenvectors and a few attempts for discussing this important issue were performed (see Blanchet et al. 2008; Bini et al. 2009; Diniz-Filho et al. 2012b). A strategy entails, for instance, trying a range of models allowing for an increasingly number of eigenvectors as explanatory variables (see Diniz-Filho et al. 2012c).

Our goals are threefold. First, we propose a new application of eigenfunction spatial analyses to describe the geographic patterns of genetic variation. We call this method spatial signal-representation (SSR) curve, which was originally developed in the context of phylogenetic comparative analyses (see Diniz-Filho et al. 1998, 2012c). Second, using simulated data, we compare SSR curves with Moran's *I* spatial correlograms, which have been more commonly used to describe genetic variation. Third, we apply the SSR curves to an empirical dataset in which the variation in microsatellite molecular markers was used to describe the genetic population structure in *Dipteryx alata* Vogel (Fabaceae), a widely distributed tree species endemic to Brazilian Cerrado.

Materials and methods

The spatial signal-representation (SSR) curve

The first step to build a SSR curve is to extract a total of $n - 1$ eigenvectors (and associated eigenvalues) from a double-centered matrix **G**, given as $-0.5D_{ij}$ (where D_{ij} are the geographical distances between populations) (Fig. 1). The eigenvectors are extracted from the double-centered **G** matrix without truncating or squaring distances, as in other spatial eigenfunction methods, such as Principal Coordinate of Neighbour Matrices (Borcard and Legendre 2002) or Moran's Eigenvector Mapping (Dray et al. 2006; Griffith and Peres-Neto 2006; Peres-Neto and Legendre 2010) (see spatial filtering by Griffith 2003).

In a second step, a response variable of interest, measured in *n* local populations (e.g., phenotypic traits or allele frequencies), is modeled as a function of an increasingly number of eigenvectors (Fig. 1). Thus, the first regression model describes the relationship between the response variable and the first eigenvector (the one associated with the highest eigenvalue). The second eigenvector is then added to estimate a second regression model. Regression models are then built sequentially until the addition of all eigenvectors is completed. Third, both coefficients of determination (R^2) and the cumulative eigenvalues ($\lambda\%$) of the eigenvectors used in each model are recorded (Fig. 1). The SSR curve is obtained by plotting the unadjusted R^2 of each model against $\lambda\%$, providing a description of how the response variable is accounted for by the spatial arrangement of the local populations at different scales (Fig. 1).

An R-package for calculating SSR curves and testing it against the null expectation of absence of spatial patterns is available from CRAN (see http://cran.r-project.org/web/packages/PVR/index.html). A friendly-user program written in Delphi for PSR/SSR curves is also available from the authors upon request.

Simulations of population genetic structure and the interpretation of the SSR curve

We simulated well-know population genetic processes (see Hardy and Vekemans 1999; Epperson et al. 2010) to show the usefulness of SSR curves in describing the spatial structure of genetic data generated under four evolutionary scenarios (i.e., isolation by distance, the effect of barriers on gene flow, range expansion, and panmixia, which generate a random pattern) (Fig. 2). All simulations were performed in a software developed in JAVA for this work and available from the authors upon request.

First, we simulated an isolation-by-distance (IBD) process (Wright 1943; see also Sokal and Wartenberg 1983;

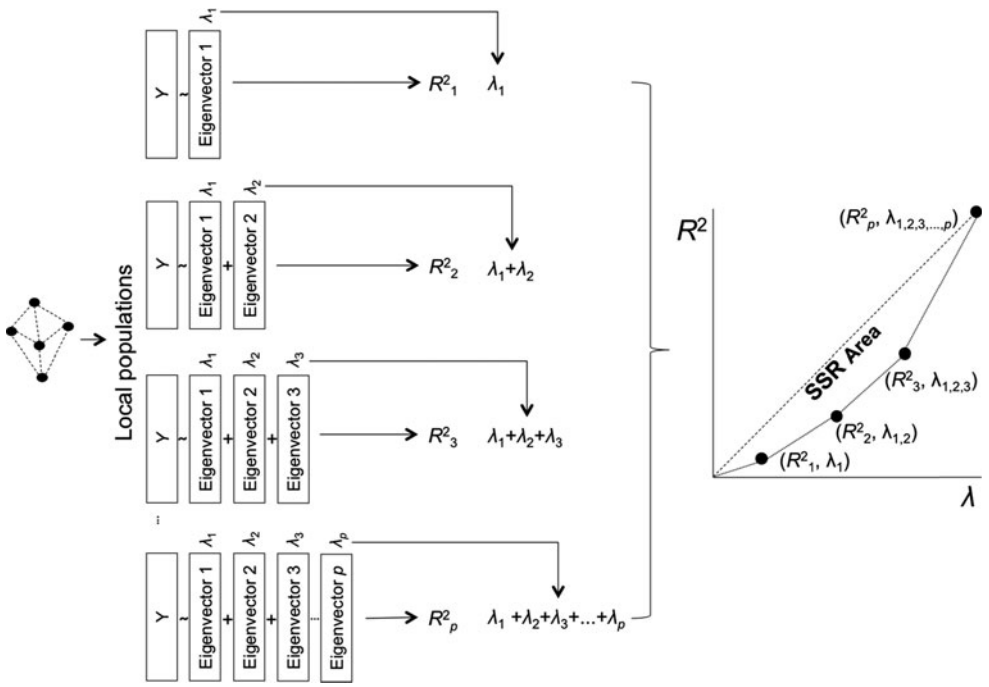

Fig. 1 Schematic representation of the calculations needed to construct a SSR curve. Eigenvectors (from 1 to p, where $p = n - 1$) extracted from a distance (or connectivity) matrix containing the spatial relationships among the local populations (represented by the *inset* map) are successively used to model the response variable (Y; e.g. allele frequency). Each model produces an unadjusted R^2. Eigenvalues (λ) associated to the eigenvectors used in the models are summed up and plotted against R^2. The region occupying the space between the curve (*dotted line*) and the 45° line (*dashed*) is the SSR area. It can be used to measure the deviation from a neutral model of spatial variation. The same procedure can be applied to study phylogenetic autocorrelation

Epperson 1995, 1996; Hardy and Vekemans 1999). All individual-based simulations were performed in an area with 100 × 100 cells, in which 2,000 diploid individuals (1,000 males and 1,000 females) were randomly allocated (the individuals thus occupy 1/5 of the available positions in the area). A single locus with two alleles was assigned to each individual, so that initial allele frequencies were equal to 0.5 and 0.5. The probability of mating [$P(D)$] between the pairs of individuals is given by the following inverse exponential function:

$$P(D) = ce^{-\alpha D},$$

where α is a parameter regulating the strength of the spatial patterns, D is the geographic distance between pairs of individuals, and c is a scaling factor. Thus, a large α indicates that individuals tend to mate primarily with their neighbors, meaning that allele frequencies will be strongly autocorrelated within short geographic distances. In our simulations we used α values from 1 to 5 (results shown for $\alpha = 5$), although the shape of the SSR curve is not strongly affected by variations in this range of the parameter.

After reproducing, the genotypes of the offspring from each pair were stochastically defined by standard Mendelian inheritance rules. There is no overlap between generations and, in the basic IBD simulation scenario, the two offspring resulting from each couple replace their parents in the same geographical coordinates. The simulations were run for 1,000 generations, and the entire process was replicated 100 times.

Starting from the simple IBD simulations, other evolutionary scenarios were defined by the way in which individuals mate and by how they disperse across space after reproducing. Thus, for our second scenario, we added a barrier in our grid, limiting mating between individuals on the opposite east–west sides of this barrier (the "barrier scenario"). Thus, genetic differentiation between both sides of the barrier will occur when (1) α is small, (2) the effects of IBD do not structure the genetic variation of nearby neighbors, and (3) the simulation time is long (as the "effective" population size is relatively high on each side of the barrier). For this barrier scenario (see below), the α was reduced to 1.0, allowing for the emergence of broad-scale geographic structures, and 2,000 generations were used. It is important to emphasize that the SSR curve does not incorporate any information about the physical barrier, as our goal was to explore how a SSR curve would behave under this effect and the other scenarios.

Our third scenario consists of a process of range expansion in which each simulation starts with a small number of individuals (10 instead of 2,000) located in the

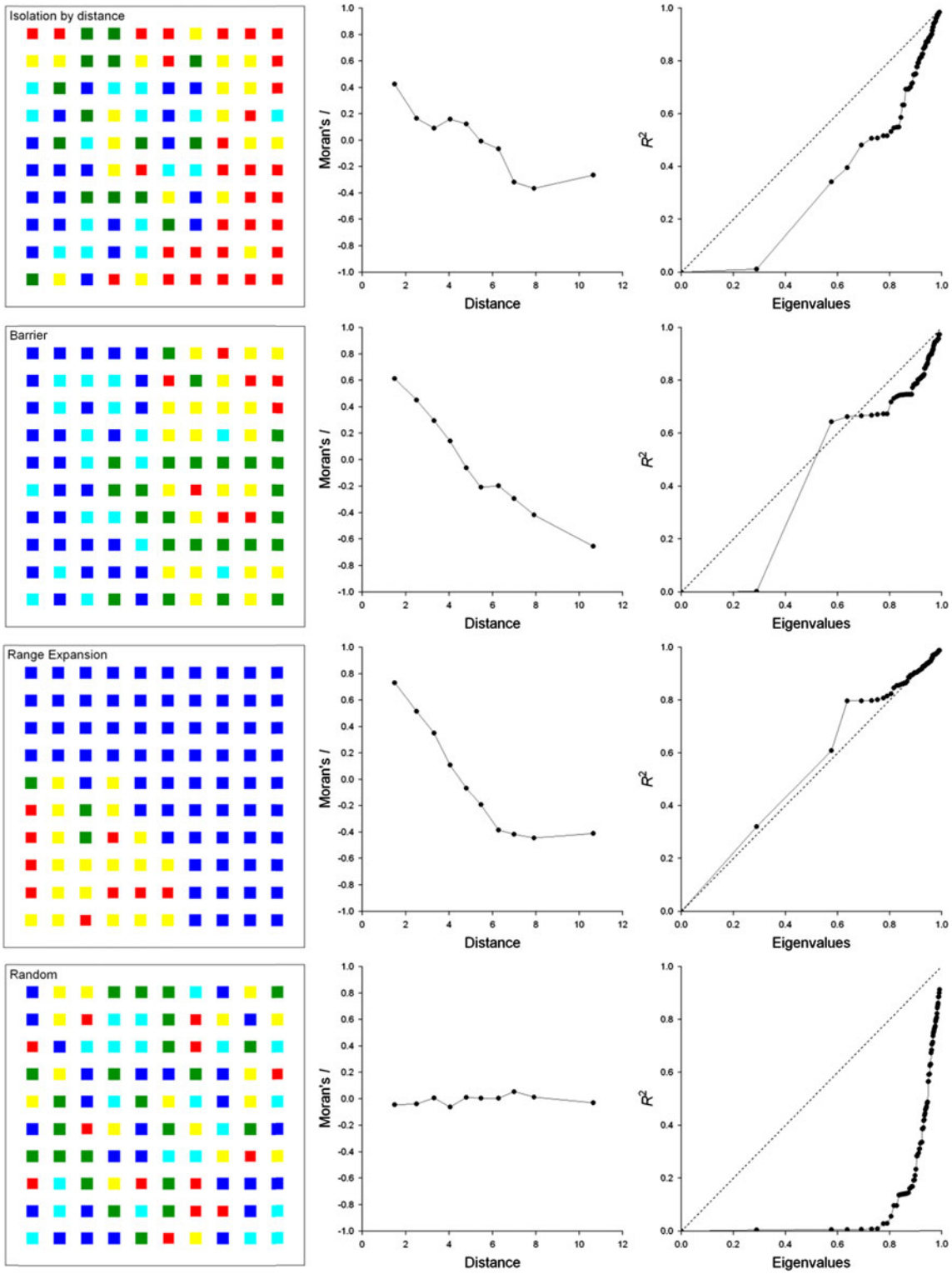

Fig. 2 Examples of allele surfaces simulated under different scenarios (isolation-by-distance, barrier, range expansion and random variation), with respective Moran's *I* correlograms and SSR curves

lower left corner of the grid (the "expansion" scenario). We allowed the population to expand (pairs will generate, on average, 1.2 offspring) and gradually disperse throughout the grid until a population of 2,000 individuals was reached, following the same exponential relationship defined above. Under this process of range expansion, it is

expected that alleles will display a clinal pattern of variation, with frequencies changing gradually when departing from the initial location (see Excoffier and Ray 2008 for more complex effects due to allele surfing).

Finally, a fourth scenario was generated by randomly assigning values of allele frequencies in geographic space, as a null model for comparing the SSR curves (the "null model" scenario), biologically reflecting panmixia. This is equivalent to setting $\alpha = 0$ in the IBD scenario defined above, so mating between individuals will occur randomly at any geographic distance (random scenario).

After the simulations, the geographical space with 10,000 cells in which the 2,000 individuals were allocated was subdivided in a regular lattice of 10 × 10 cells, forming "populations". The genotypes of the individuals in each cell were recorded and the allele frequency in each cell was calculated and used to produce the SSR curve. The matrix **G** used to extract eigenvectors (after a double-center transformation) was also used to represent the geographic relationship between the centroids of the 100 populations. In all scenarios, eigenvectors were extracted from the same **G** matrix, so that variations in the shape of the curves are expected to track the processes underlying these different scenarios.

Using the simulated datasets, we compared the SSR curves with Moran's *I* correlograms, whose use for describing spatial patterns in genetic variation was pioneered by Sokal and Oden (1978a, b; see Diniz-Filho and Bini 2012 for a review). For the different alleles, Moran's *I* correlograms were estimated with 10 distance classes, using a batch version of the SAM (Spatial Analysis in Macroecology) v. 4.0 software (Rangel et al. 2006, 2010).

Sokal and Wartenberg (1983) and Sokal et al. (1997; see also Sokal and Jacquez 1991) showed that, under IBD, allele frequencies of different loci will be stochastically independent even though their Moran's *I* spatial correlograms will be similar; this is because their spatial patterns will be driven by a common dispersion parameter (see also Diniz-Filho et al. 2012a). They tested this prediction by correlating a Manhattan distance matrix (**M**) (Legendre and Legendre 2012) containing the dissimilarities between the correlograms for the different allele frequencies (using Moran's *I* as "variables") with a Pearson correlation matrix (**R**) containing the spatial correlations between allele frequencies (correlation between "maps"). Thus, it is possible to evaluate how similarity between maps of allele frequencies is related to the similarity among the patterns described by their correlograms (Manhattan distances among correlograms). Here, we performed a similar test with the SSR curves, by using a Pearson correlation between the Manhattan distances between pairs of SSR curves (using R^2 as "variables") obtained for different alleles and the correlation matrix between the maps of allele frequencies.

Finally, we also calculated the area of the SSR curve (SSR area) as the region occupying the space between the curve and the proportional variation line (the 45° line) on the scatter plot of R^2 against the cumulative $\lambda\%$ of the eigenvalues (Fig. 1) (see Diniz-Filho et al. 2012c). As a convention, the area below the curve was coded as "negative", and areas above the curve were assigned as "positive". Null distributions of the SSR areas can be obtained by randomizing the allele frequencies in the populations.

Empirical data analysis

We analyzed 28 allele frequencies derived from eight microsatellite loci of *D. alata*, an endemic tree of the Brazilian Cerrado. We obtained the genotypes of 644 individuals collected from 25 locations throughout Central Brazil (see Diniz-Filho et al. 2012d), with sample sizes ranging from 12 to 32 individuals per locality (32 individuals were analyzed in 15 of the 25 populations). Although more alleles were found for these eight loci (a total of 52), many of them were rare and occurred in a single population; therefore, we restricted our spatial analyses to those alleles present in at least 8 local populations. Details of genotyping, sample size, private alleles and the genetic characteristics of each local population are given in Soares et al. (2012) and Diniz-Filho et al. (2012d). The same analytical protocols described above (SSR curves and Moran's *I* correlograms) were applied to analyze each of the alleles in this dataset.

Results

Simulations

The maps obtained from the simulations are in line with the spatial patterns expected under each process. For the simulations performed under the IBD scenario (Figs. 2, 3), a short-distance spatial structure is observed, and Moran's *I* correlograms show the expected exponential-like distance decay. The SSR curve also shows an exponential shape with points lying below the proportional variation line. Thus, the variation in genetic data explained by the first eigenvectors, representing broad spatial scales, is smaller than the relative eigenvalues associated with these eigenvectors (Fig. 3a).

Under IBD, the pairwise Manhattan distances between SSR curves and correlograms are correlated ($r = 0.337$; $P > 0.01$; Fig. 4a), although there is wide scatter. Even so, the absolute values of Manhattan distances are relatively small, indicating that the curves and correlograms are similar. Most importantly, the relationships between the

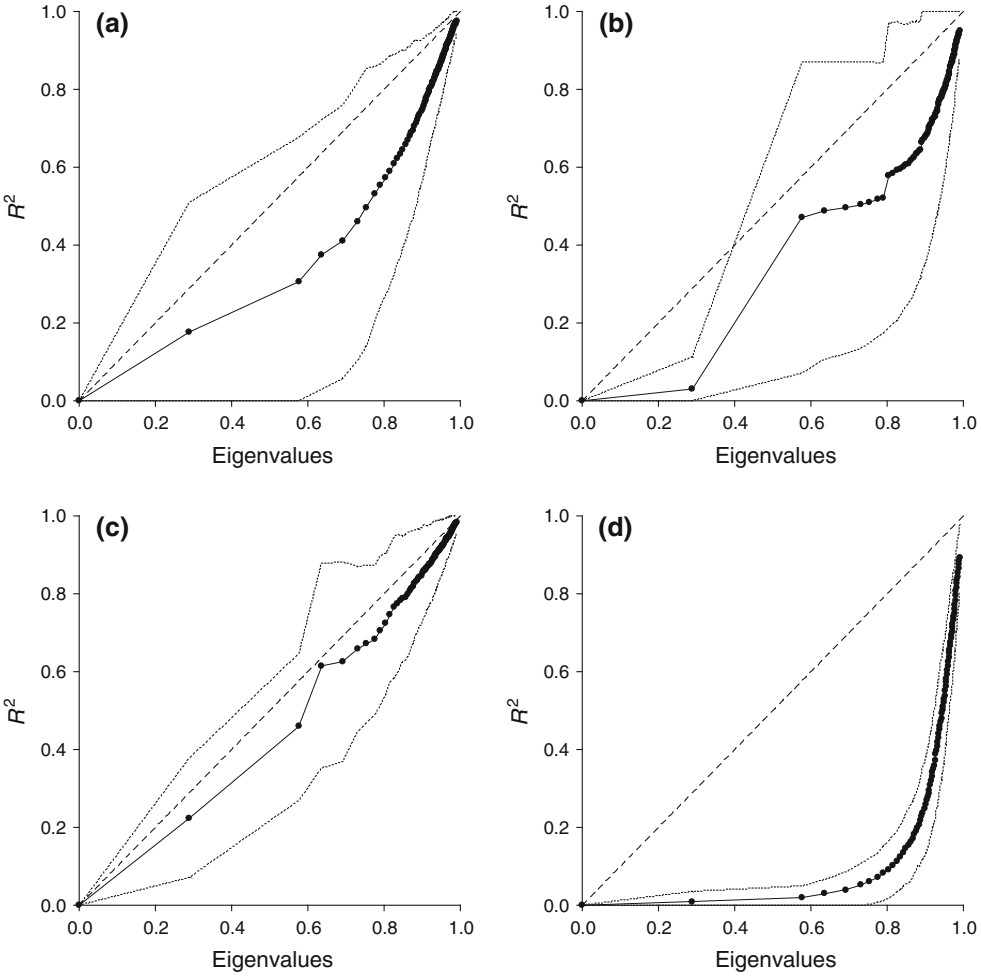

Fig. 3 Average SSR curves (*filled circles* and *continuous lines*) derived from the different scenarios: **a** isolation by distance, **b** barriers, **c** range expansion and **d** random variation. The *dotted lines* represent the 95 % confidence intervals. The 45° line (*dashed*) is also shown for comparison

matrix containing the Pearson's correlations between allele frequencies and the Manhattan distances between correlograms (Fig. 4b) or SSR curves (Fig. 4c) were not significant ($r = 0.04$; $P \gg 0.05$ and $r = -0.003$; $P \gg 0.05$, respectively).

With the addition of a dispersal barrier, the allele frequencies tend to differ between the two sides of the barrier (east and west sides—Figs. 2, 3b). The correlogram shows a more continuous decline of Moran's I from small to large geographic distance classes. On the other hand, the SSR curve shows a conspicuous increase in R^2 when approximately half of the variation in the geographic distance matrix was represented by the eigenvectors used in the model (i.e., $\lambda\% = 50$ % in the abscissa). Thus, the increase in the variation of allele frequencies accounted for by spatial eigenvectors is higher than detected for the SSR curve estimated with the data from IBD simulations.

Under the range expansion scenario (Figs. 2, 3c), the correlograms show a distance decay in Moran's I coefficients up to the sixth distance class. On the other hand, the SSR curve tends to be linear and close to the proportional variation line.

Finally, in contrast to the patterns described above, the lack of spatial patterns (i.e., randomness of allele frequencies in space or panmixia) is easily recognized by the Moran's I correlograms. The SSR curve is also very different from the ones previously described, depicting a line that starts flat, but moves upward only after the addition of a large number of eigenvectors (Figs. 2, 3d). Thus, even after allowing for a large number of eigenvectors (which account for most of the variation in matrix **G**; for instance, $\lambda\% = 80$ %), the coefficients of determination of models for allelic frequencies are always lower than 10 %.

The areas of the SSR curves differed significantly among the scenarios (ANOVA's $R^2 = 0.795$; $F = 442$; $P \ll 0.01$) as well as between each scenario and the null expectation. The areas estimated for barrier and expansion scenarios did not differ significantly (Tukey's post hoc test), but both differed from IBD. All three simulated scenarios differed from the null model (i.e. random or panmixia) scenario (Fig. 5).

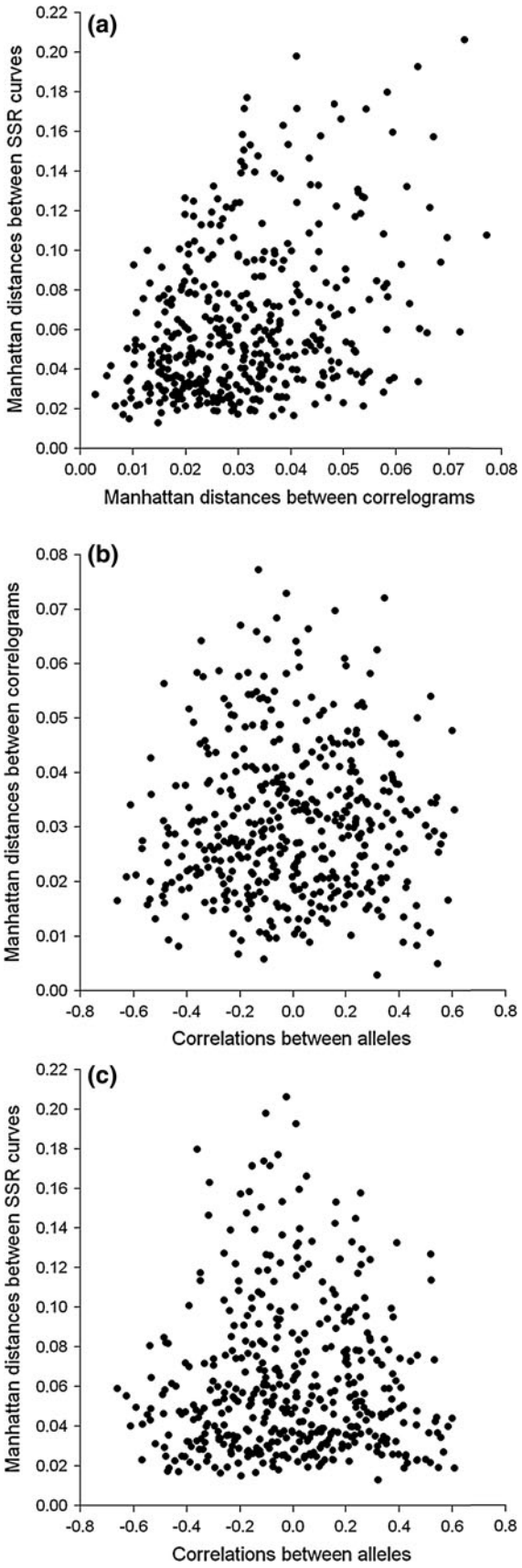

Fig. 4 Relationship between Manhattan distances between pairs of correlograms and SSR curves (**a**), relationship between correlations between alleles and Manhattan distances between correlograms (**b**) and relationship between correlations between alleles and Manhattan distances between SSR curves (**c**)

Geographic variation of allele frequencies in *Dipteryx alata* populations

Most SSR curves for the 28 allele frequencies of the *D. alata* populations are below or close to the proportional variation line and are thus similar to the ones obtained in our simulations for IBD and range expansion (Fig. 6a). The null distribution of the SSR area (Fig. 6b) shows that, for this sample size and spatial configuration, SSR areas lower than −0.266 below the proportional variation line can be considered random. SSR areas for fourteen alleles were below −0.266 and thus their variations across space were not different from those expected by chance alone (Fig. 5b). Indeed, of these 14 alleles, only three have Moran's *I* coefficients larger than 0.1 in the first distance class. The correlation between the SSR area and the Moran's *I* coefficient in the first distance class was positive ($r = 0.609$; $P < 0.01$), but gradually became negative as higher distance classes were considered. For instance, the correlation between SSR area and the Moran's *I* coefficient estimated for the last distance class was strongly negative ($r = -0.908$; $P \ll 0.01$), indicating that linear SSR curves are found for more linear correlograms too, both expressing clinal patterns in geographical space.

Discussion

Spatial autocorrelation and the SSR curve

Our results show that SSR curves can be a useful tool to explore spatial patterns of genetic variation and allow for the differentiation between evolutionary processes, as given by our simulated scenarios. In addition, SSR curves provide a better visual description of complex spatial surfaces than do Moran's *I* correlograms.

Previous studies showed that correlograms resulting from the IBD process tend to have positive autocorrelation (as estimated by Moran's *I* coefficients) at short distances and non-significant autocorrelation at long distances (Sokal and Oden 1978b; Sokal and Wartenberg 1983; Barbujani 1987; Epperson 1995, 1996, 2004, 2005; Hardy and Vekemans 1999). This pattern is found because, although nearby individuals or local populations tend to be genetically similar, similarity between individuals becomes unpredictable with the increase of geographic distance. The same pattern appears when using the SSR curve, which also follows an exponential-like shape under IBD.

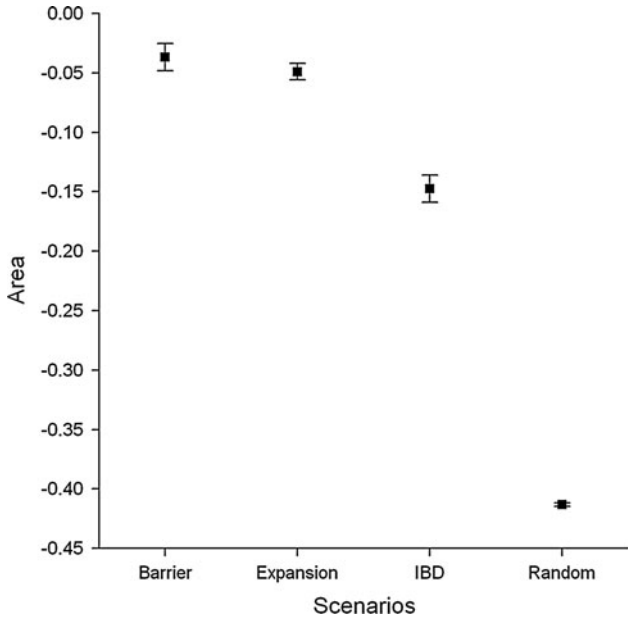

Fig. 5 SSR area for the different scenarios (isolation by-distance, barrier, range expansion and random variation)

Under the barrier scenario, the simulations show that Moran's *I* correlograms become more linear and similar to the ones that can emerge under long-distance migration waves or selection. In general, the SSR curve is a better description of the spatial pattern generated by the barrier scenario than Moran's *I* correlograms, with a peak of R^2 at intermediate geographic distances. Notice that the R^2 of the first model in the SSR curve (i.e., the one that includes the first eigenvector only) is very close to zero because of the omnidirectional behavior of our simulations. Because our simulations generate large differences in allele frequencies between the east and west sides of the grid, and the first eigenvectors depict broad-scale patterns, one could expect that these first eigenvectors would explain a large fraction of the variation in the allele frequencies. However, this expectation does not hold because of the strong similarity in the north–south direction; the R^2 tends toward zero and only after adding the second eigenvector is the R^2 near the proportional line. The ability of the SSR curve to detect a barrier is directly related to the difference between the two regions separated by the barrier. Despite the existence of other spatial methods to detect barriers, such as wombling (see Fortin and Dale 2005 for a detailed description, and Cercueil et al. 2007 for a recent development), one of the improvements of the SSR curve over these methods is that it is possible to identify which eigenvector is describing the pattern, which can be used later to model the variation in allele frequency in relation to environmental predictors.

The range expansion scenario generates a simple spatial pattern that is described approximately by linear correlograms (at least considering the first distance classes) or SSR curves. Broad-scale clines in allele frequencies, which have sometimes been interpreted as resulting from selection, can indeed be generated by stochastic processes, as observed here (see Excoffier and Ray 2008). Further developments of the SSR curve and of the Sokal and Oden's (1978b) correlograms-based framework are necessary to allow better distinction between neutral and adaptive clines under range expansion scenarios. Nevertheless, this may be challenging because range expansion is also usually associated with climate changes (Excoffier and Ray

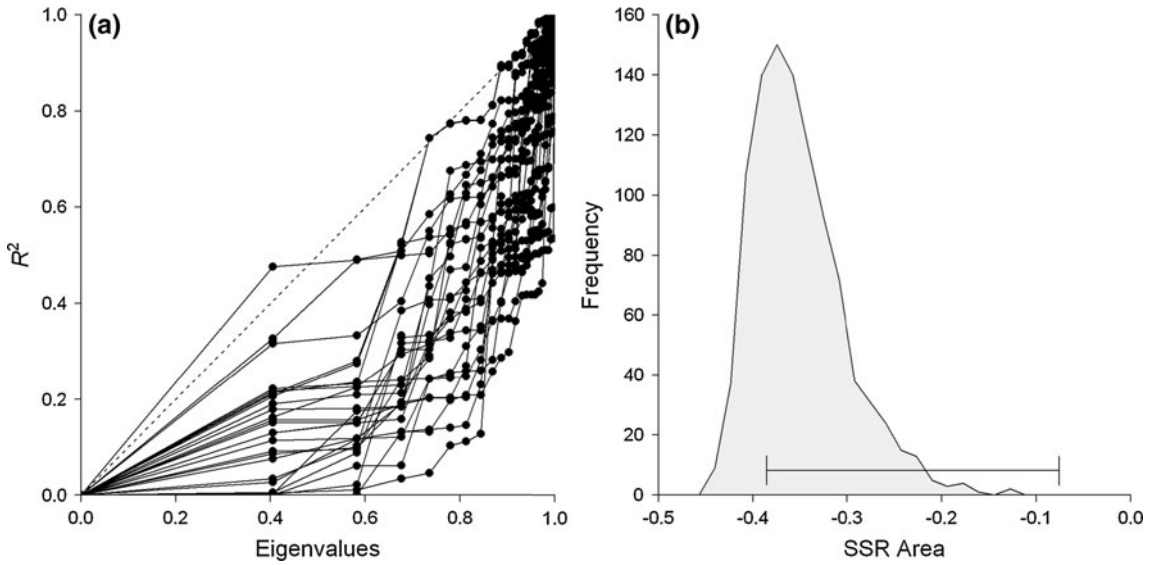

Fig. 6 SSR curves for 28 alleles derived from 8 microsatellite loci of *Dipteryx alata* in Brazilian Cerrado (**a**) and the null distribution of SSR area for the spatial configuration of 25 local populations (**b**). The *horizontal bar* delimits the range of SSR areas obtained for the 28 alleles

2008), so a correlation between the cline in allele frequencies and environmental drivers (triggering range expansion) will appears even under neutral processes.

Our results show that SSR curves can be used to infer microevolutionary processes, following the inferential framework proposed by Sokal and Oden (1978a, b; see also Sokal and Jacquez 1991; Sokal et al. 1997; Diniz-Filho and Bini 2012). For example, Sokal and Wartenberg (1983) showed that, under IBD, allele frequencies are uncorrelated (because they are independent realizations of the stochastic dispersal process). However, the correlograms of these allele frequencies are similar because the same dispersal distance controls the underlying structure of the spatial pattern. Thus, pairwise Manhattan distances between correlograms and pairwise correlations between allele frequencies are expected to be independent of each other (see also Diniz-Filho et al. 2012a for application of the same reasoning for ecological communities). Indeed, for our IBD simulations, this correlation (between Manhattan distances and pairwise correlations between allele frequencies) tended toward zero. Also, there was no correlation between the matrix of Manhattan distances between the curves and the correlation matrix between the allele frequencies. This result suggests that SSR curves can potentially replace Moran's *I* in Sokal and Oden's (1978b) inferential framework.

Epperson (1995, 1996) proposed that Moran's *I* coefficients and correlograms could be used to evaluate patterns within populations based on transformed genotypic data by assigning a value of 1 if the individual is homozygous for the allele, 0.5 if the individual is heterozygous, and 0 if the individual does not contain the allele (these numbers are the allele frequencies in a "region" so small that it contains only a single individual). In the same vein, although our analyses were based on allele frequencies in local populations, it is possible to use the SSR curve to analyze variation within local populations, highlighting the flexibility of eigenfunction spatial analyses (Griffith and Peres-Neto 2006).

However, it is important to emphasize that the flexibility of eigenfunctions (see Griffith and Peres-Neto 2006) has a downside, which is the large number of eigenvectors that can be extracted to represent the spatial relationships when many local populations are analyzed (Blanchet et al. 2008). Therefore, some form of variable selection is required and, in the realm of spatial eigenfunction analyses, it has usually been performed using several criteria with the main goal of controlling Type I error rates (see Diniz-Filho and Bini 2005; Griffith and Peres-Neto 2006; Blanchet et al. 2008; Bini et al. 2009; Diniz-Filho et al. 2012c). Our results suggest that SSR curves can also be useful in eigenvector selection by highlighting the geographic distance at which incorporation of the eigenvector in the model can track the spatial variation in the response variable (as performed by Manel et al. 2010b). For example, as noted above, the SSR curve can be used to identify which eigenvector is capturing the effect of a barrier or of other adaptive factors (see Manel et al. 2010a). One can then map the selected eigenvector as a way to verify whether it represents regions separated by an overlooked barrier in the studied landscape. Further studies using data of other species having similar dispersal abilities and a planned sampling design (to ensure representation of populations in both sides of the supposed barrier) could be performed to offer independent evidence of the barrier's efficiency in disrupting or minimizing gene flow.

The SSR curve, as presented here, is essentially univariate (i.e., one allele frequency is analyzed at time), similarly to Moran's *I* correlograms, allowing interesting evolutionary inferences. On the other hand, dealing simultaneously with many alleles, such as done with a Mantel test or Mantel correlogram based on genetic distances or similarity, does not allow a detailed description of multiple potential evolutionary processes driving patterns in each loci or allele frequency (except if these distances are defined with clearly defined expectations based on evolutionary models—see Diniz-Filho et al. 2012d). Even so, if the researcher is interested in a general description of several alleles and loci, the SSR curve can still be applied using a multivariate R^2 rather than the univariate R^2. Also, one can simply use the scores derived from any ordination analysis as response variable in the multiple regression models needed to generate the SSR curve.

Geographical genetics of *D. alata*

SSR curves applied to the *D. alata* are consistent with the previous findings about the landscape and population genetics of this species. For instance, autocorrelation analyses detected a spatial genetic structure at local scales, within local populations (see Hardy et al. 2006, 2008), as expected by considering the ecological and life history traits of the species, especially reproductive mode (i.e. self-incompatible) and seed dispersal by mammals and pollination by large bees (see Collevatti et al. 2010). These characteristics tend to create local spatial structure in genetic variation, and the local patterns described by Collevatti et al. (2010) can be expanded to broader spatial scales and generate IBD-like decrease of genetic similarity with increasingly geographic distances. But at broad spatial scales these must be coupled with historical processes of range expansion to explain more clinal genetic structure among populations, such as studied here using the SSR curve (see also Soares et al. 2008; Collevatti et al. 2013).

Our current knowledge of genetic diversity and population structure in *D. alata* shows that the species possess a

strong spatial structure in genetic differentiation following an IBD-like pattern coupled with range expansion for distinct loci (Diniz-Filho et al. 2012d; Collevatti et al. 2013). This is also revealed by the SSR curves proposed here. However, inferring evolutionary processes from spatial patterns depicted by empirical data is a challenging task because these processes are complex, idiosyncratic and have a strong historical component (Slatkin and Arter 1991; Sokal et al. 1997). In this context, both SSR curves and Moran's *I* based correlograms have limitations. Even so, the SSR curves seem to be better than correlograms to detect the existence of barriers to gene flow at long distances.

Concluding remarks

Here, we show that our new approach based on SSR curves is a useful exploratory tool for describing spatial patterns of genetic variability. Our simulations show that different processes driving population divergence produce SSR curves with distinct shapes. Particularly, the SSR curve was more efficient in detecting the effect of barriers on gene flow than Moran's *I* correlogram. Although further developments in the application of the SSR curve to infer microevolutionary processes are still needed, our analyses suggest that the new approach can be applied akin to Moran's *I* correlograms. Moreover, the SSR curve can be useful for eigenvector selection to model phenotypic spatial variation. These different possibilities reveal the flexibility of eigenfunction spatial analyses when applied to geographical and landscape genetics.

Acknowledgments We thank Pedro Peres-Neto for several discussions about PSR/SSR curve and Guillaume Guénard for many suggestions that improved early versions of this manuscript. We also thank Thiago Santos for preparing the R-code that is available for analysis. This work was supported by several grants and fellowships to the research network GENPAC (Geographical Genetics and Regional Planning for natural resources in Brazilian Cerrado) from CNPq/MCT/CAPES (Projects # 564717/2010-0 and 563624/2010-8) and by the "Núcleo de Excelência em Genética e Conservação de Espécies do Cerrado"—GECER (PRONEX/FAPEG/CNPq CP 07-2009). Field work has been supported by Systema Naturae Consultoria Ambiental LTDA. Work by J.A.F.D.-F., L. M. B., T. F. R., M.P.C.T. and R.G.C. has been continuously supported by productivity fellowships from CNPq.

References

Balkenhol N, Waits LP, Dezzani RJ (2009) Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. Ecography 32:818–830

Barbujani G (1987) Autocorrelation of gene frequencies under isolation-by-distance. Genetics 177:772–782

Bertin A, Ruíz VH, Figueroa R, Gouin N (2012) The role of spatial processes and environmental determinants in microgeographic shell variation of the freshwater snail *Chilina dombeyana* (Bruguiére, 1789). Naturwissenschaften 99:225–232

Bini LM, Diniz-Filho JAF, Rangel TFLVB, Akre TSB, Albaladejo RG, Albuquerque FS, Aparicio A, Araújo MB, Baselga A, Beck J, Bellocq MI, Böhning-Gaese K, Borges PAV, Castro-Parga I, Chey VK, Chown SL, Marco P Jr, Dobkin DS, Ferrer-Castán D, Field R, Filloy J, Fleishman E, Gómez JF, Hortal J, Iverson JB, Kerr JT, Kissling WD, Kitching IJ, León-Cortés JL, Lobo JM, Montoya D, Morales-Castilla I, Moreno JC, Oberdorff T, Olalla-Tárraga MÁ, Pausas JG, Qian H, Rahbek C, Rodríguez MA, Rueda M, Ruggiero A, Sackmann P, Sanders NJ, Terribile LC, Vetaas OR, Hawkins BA (2009) Coefficients ships in geographical ecology: an empirical evaluation of spatial and non-spatial regression. Ecography 32:193–204

Blanchet FG, Legendre P, Borcard D (2008) Forward selection of explanatory variables. Ecology 89:2623–2632

Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecol Model 153:51–68

Cercueil A, Francois O, Manel S (2007) The genetical bandwidth mapping: a spatial and graphical representation of population genetic structure based on the Wombling method. Theor Popul Biol 71:332–341

Collevatti RG, Lima JS, Soares TN, Telles MPC (2010) Spatial genetic structure and life-history traits in Cerrado tree species: inferences for conservation. Nat Conserv 8:54–59

Collevatti RG, Telles MPC, Nabout JC, Chaves LJ, Soares TN (2013) Demographic history and the low genetic diversity in *Dipteryx alata* (Fabaceae) from Brazilian Neotropical savannas. Heredity. doi:10.1038/hdy.2013.23

Croucher PJP, Oxford GS, Gillespie RG (2011) Population structure and dispersal in a patchy landscape: nuclear and mitochondrial markers reveal area effects in the spider *Theridion californicum* (Araneae: Theridiidae). Biol J Linn Soc 104:600–620

Desdevises Y, Legendre P, Azouzi L, Morand S (2003) Quantifying phylogenetically structured environmental variation. Evolution 57:2647–2652

Diniz-Filho JAF, Bini LM (2005) Modelling geographical patterns in species richness using eigenvector-based spatial filters. Glob Ecol Biogeogr 14:177–185

Diniz-Filho JAF, Bini LM (2012) Thirty-five years of spatial autocorrelation analysis in population genetics: an essay in honour of Robert R. Sokal (1926–2012). Biol J Linn Soc 105:721–736

Diniz-Filho JAF, Sant'Ana CER, Bini LM (1998) An eigenvector method for estimating phylogenetic inertia. Evolution 52:1247–1262

Diniz-Filho JAF, Nabout JC, Telles MPC, Soares TN, Rangel TFLVB (2009) A review of techniques for spatial modeling in geographical, conservation and landscape genetics. Genet Mol Biol 32:203–211

Diniz-Filho JAF, Siqueira T, Padial AA, Rangel TF, Landeiro VL, Bini LM (2012a) Spatial autocorrelation allows disentangling the balance between neutral and niche processes in metacommunities. Oikos 121:201–210

Diniz-Filho JAF, Bini LM, Rangel TF, Morales-Castilla I, Olalla-Tárraga M, Rodriguez MA, Hawkins BA (2012b) On the selection of phylogenetic eigenvectors for ecological analysis. Ecography 35:239–249

Diniz-Filho JAF, Rangel TF, Santos T, Bini LM (2012c) Exploring patterns of interspecific variation in quantitative traits using sequential phylogenetic eigenvector regression. Evolution 64:1079–1090

Diniz-Filho JAF, Collevatti RG, Soares TN, Telles MPC (2012d) Geographical patterns of turnover and nestedness-resultant components of allelic diversity among populations. Genetica 140:189–195

Dormann CF, McPherson J, Araújo MB, Bivand R, Bolliger J, Carl G, Davies RG, Hirzel A, Jetz W, Kissling WD, Kühn I, Ohlemüller R, Peres-Neto P, Reineking B, Schröder B, Schurr FM, Wilson R (2007) Methods to account for spatial autocorrelation in the analysis of distributional species data: a review. Ecography 30:609–628

Dray S, Legendre P, Peres-Neto PR (2006) Spatial modeling: a comprehensive framework for principal coordinate analysis of neighbor matrices (PCNM). Ecol Model 196:483–493

Epperson BK (1995) Spatial distribution of genotypes under isolation by distance. Genetics 140:1431–1440

Epperson BK (1996) Measurement of genetic structure within populations using Moran's I spatial autocorrelation statistics. Proc Natl Acad Sci USA 93:10528–10532

Epperson BK (2003) Geographical genetics. Princeton University press, Princeton

Epperson BK (2004) Multilocus estimation of genetic structure within populations. Theor Popul Biol 65:227–337

Epperson BK (2005) Estimating dispersal from short distance autocorrelation. Heredity 95:7–15

Epperson BK, McRae B, Scribner K, Cushman SA, Rosenberg MS, Fortin M-J, James PMA, Murphy M, Manel S, Legendre P, Dale MRT (2010) Utility of computer simulations in landscape genetics. Mol Ecol 19:3549–3564

Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. Trends Ecol Evol 23:347–351

Fortin M-J, Dale MRT (2005) Spatial analysis: a guide for ecologists. Cambridge University Press, Cambridge

Griffith DA (2003) Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization. Springer, New York

Griffith DA, Peres-Neto PR (2006) Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. Ecology 87:2603–2613

Guillot G, Leblois R, Coulon A, Frantz AC (2009) Statistical methods in spatial genetics. Mol Ecol 18:4734–4756

Hardy OJ, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. Genetics 83:145–154

Hardy OJ, Maggia L, Bandou E, Breyne P, Caron H, Chevallier MH, Doligez A, Dutech C, Kremer A, Latouche-Halle C, Troispoux V, Veron V, Degen B (2006) Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. Mol Ecol 15:559–571

Hardy OJ, Pearcy M, Aron S (2008) Small scale spatial genetic structure in an ant species with sex-biased dispersion. Biol J Linn Soc 93:465–473

Holderegger R, Wagner HH (2006) A brief guide to landscape genetics. Landscape Ecol 21:793–796

Holderegger R, Wagner HH (2008) Landscape genetics. Bioscience 58:199–208

Holderegger R, Buehler D, Gugerli F, Manel S (2010) Landscape genetics of plants. Trends Plant Sci 15:675–682

Kelly RP, Oliver TA, Sivasundar A, Palumbi SR (2010) A method for detecting population genetic structure in diverse, high gene flow species. J Hered 101:423–436

Legendre P, Legendre L (2012) Numerical ecology, 3rd edn. Elsevier, Amsterdam

Logue JB, Mouquet N, Peter H, Hillebrand H (2011) Empirical approaches to metacommunities: a review and comparison with theory. Trends Ecol Evol 26:482–491

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. Trends Ecol Evol 15:189–197

Manel S, Joost S, Epperson BK, Holderegger R, Storfer A, Rosenberg MS, Scribner K, Bonin A, Fortin MJ (2010a) Perspective on the use of landscape genetics to detect genetic adaptive variation in the field. Mol Ecol 19:3760–3772

Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010b) Common factors drive adaptive genetic variation at different scale in *Arabis alpina*. Mol Ecol 19:2896–2907

Meister B, Hofer U, Ursenbacher S, Baur B (2010) Spatial genetic analysis of the grass snake, *Natrix natrix* (Squamata: Colubridae) in an extensively used agricultural landscape. Biol J Linn Soc 101:51–58

Peres-Neto PR, Legendre P (2010) Estimating and controlling for spatial structure in the study of ecological communities. Glob Ecol Biogeogr 19:174–184

Rangel TFLVB, Diniz-Filho JAF, Bini LM (2006) Towards an integrated computational tool for spatial analysis in macroecology and biogeography. Glob Ecol Biogeogr 15:321–327

Rangel TFLVB, Diniz-Filho JAF, Bini LM (2010) SAM: a comprehensive application for spatial analysis in macroecology. Ecography 33:1–5

Rousset F (2004) Genetic structure and selection in subdivided population. Princeton University Press, Princeton

Segelbacher G, Cushman SA, Epperson BK, Fortin MJ, Francois O, Hard O, Holderegger R, Taberlet P, Waits LP, Manel S (2010) Landscape genetics: concepts and challenges in a conservation context. Conserv Genet 11:375–385

Slatkin M, Arter HE (1991) Spatial autocorrelation methods in population genetics. Am Nat 138:499–517

Soares TN, Chaves LJ, Telles MPC, Diniz-Filho JAF, Resende LV (2008) Landscape conservation genetics of *Dipteryx alata* (''baru'' tree: Fabaceae) from Cerrado region of central Brazil. Genetica 132:9–19

Soares TN, Melo DB, Resende LV, Vianello RP, Chaves LJ, Collevatti RG, Telles MPC (2012) Development of microsatellite markers for the Neotropical tree species *Dipteryx alata* (Fabacea). Am J Bot 99:e72–e73

Sokal RR, Jacquez GM (1991) Testing inferences about microevolutionary processes by means of spatial autocorrelation analysis. Evolution 45:152–168

Sokal RR, Oden NL (1978a) Spatial autocorrelation in biology. 1. Methodology. Biol J Linn Soc 10:199–228

Sokal RR, Oden NL (1978b) Spatial autocorrelation in biology. 2. Some biological implications and four applications of evolutionary and ecological interest. Biol J Linn Soc 10:229–249

Sokal RR, Wartenberg DE (1983) A test of spatial autocorrelation analysis using an isolation-by-distance model. Genetics 105:219–237

Sokal RR, Oden N, Thomson BA (1997) A simulation study of microevolutionary inferences by spatial autocorrelation analysis. Biol J Linn Soc 60:73–93

Spear SF, Peterson CR, Matocq MD, Storfer A (2005) Landscape genetics of the blotched tiger salamander (*Ambystoma tigrinum melanostictum*). Mol Ecol 14:2553–2564

Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF, Dezzani R, Delmelle E, Vierling L, Waits LP (2007) Putting the 'landscape' in landscape genetics. Heredity 98:128–142

Wagner H, Fortin MJ (2013) A conceptual framework for the spatial analysis of landscape genetic data. Conserv Genet 14:253–261

Wright S (1943) Isolation by distance. Genetics 28:114–138