

Niche modelling and landscape genetics of *Caryocar brasiliense* (“Pequi” tree: Caryocaraceae) in Brazilian Cerrado: an integrative approach for evaluating central–peripheral population patterns

José Alexandre F. Diniz-Filho · João Carlos Nabout · Luis Mauricio Bini ·
Thannya N. Soares · Mariana Pires de Campos Telles · Paulo de Marco Jr. ·
Rosane G. Collevatti

Received: 21 October 2008 / Revised: 4 March 2009 / Accepted: 13 April 2009 / Published online: 13 May 2009
© Springer-Verlag 2009

Abstract Complex and integrative approaches may be necessary to understand the abundant-centre model and the patterns in genetic diversity that may be explained by this model. Here we developed an integrated framework to study spatial patterns in genetic diversity within local populations, coupling genetic data, niche modelling and landscape genetics, and applied this framework to evaluate population structure of *Caryocar brasiliense*, an endemic tree from the Brazilian Cerrado. We showed different geographical patterns for genetic diversity, allelic richness and inbreeding levels, estimated using microsatellite data for ten local populations. Ecological suitability was estimated by combining five niche modelling techniques. Genetic diversity tend to follow a central-periphery model and is associated with ecological variables. On the other hand, inbreeding levels may be alternatively explained by isolation processes and habitat fragmentation more related to intense recent human occupation in the southern border of the biome, or by deeper historical patterns in the origin of the populations. Although still suffering from some of the problems of central-periphery analysis (small number of local populations), our analyses show how these patterns

can be better investigated and offering a better understanding of the processes structuring genetic diversity within species’ geographic ranges.

Keywords Genetic diversity · Human impacts · Inbreeding · Niche modelling · Landscape genetics · Population structure

Introduction

In a macroecological context, the analysis of range edges has been usually developed in a comparative way, trying to understand how ecological and microevolutionary mechanisms drive populations situated inside the extent of occurrence of a species and how they vary towards their edges (Brown 1984; Gaston 2003). Different empirical data on demographic and genetic parameters (e.g., population density, growth rates, survival and genetic diversity), obtained in a geographical framework, are usually compared with expectations from theoretical models to infer which ecological and microevolutionary processes explain why species have limited ranges (Holt and Keit 2005).

There are many hypotheses to explain the dynamics of range edges; many of them involving some loss of adaptive or neutral genetic variation in peripheral populations (see Gaston 2003 for a review). The starting point of most of these mechanistic hypotheses is the abundant-centre model, a pattern described since early 1920s and that predicts a peak of abundance in the geographic centre of populations (Soulé 1973; Hoffman and Blows 1994; Sagarin and Gaines 2002). Under this hypothesis, it is expected that peripheral populations, because of their reduced effective population size, suffer loss of genetic variation due to drift and inbreeding, which can in turn reduce individual fitness

Communicated by A. Kremer.

J. A. F. Diniz-Filho (✉) · L. M. Bini · T. N. Soares ·
M. P. de Campos Telles · P. de Marco Jr. · R. G. Collevatti
Departamento de Biologia Geral, Instituto de Ciências
Biológicas—ICB, Universidade Federal de Goiás,
Campus Samambaia (Campus II), caixa postal: 131,
CEP: 74001-970 Goiânia, GO, Brazil
e-mail: diniz@icb.ufg.br

J. C. Nabout
Programa de Pós-Graduação em Ciências Ambientais (CIAMB),
PRPPG-UFG,
Goiânia, Brazil

and population adaptability, causing, then, lower reproduction and higher mortality rates and, in the end, even smaller populations and extinction (the extinction vortex explanation) (Fig. 1a). If intensity of these processes is geographically structured towards range edges, there will be some equilibrium level which will limit species' extent of occurrence. At the same time, adaptive processes that could increase population sizes in the range edges would be hindered by asymmetric gene flow from centre to edges (Kirkpatrick and Barton 1997; Lenormand 2002; Alleaume-Benharira et al. 2006).

Although many papers showed a decline in genetic diversity towards range edges, as expected under the abundant-centre hypothesis (see Eckert et al. 2008 for a recent review), the generality of this model has been doubted based on empirical data and theoretical concepts

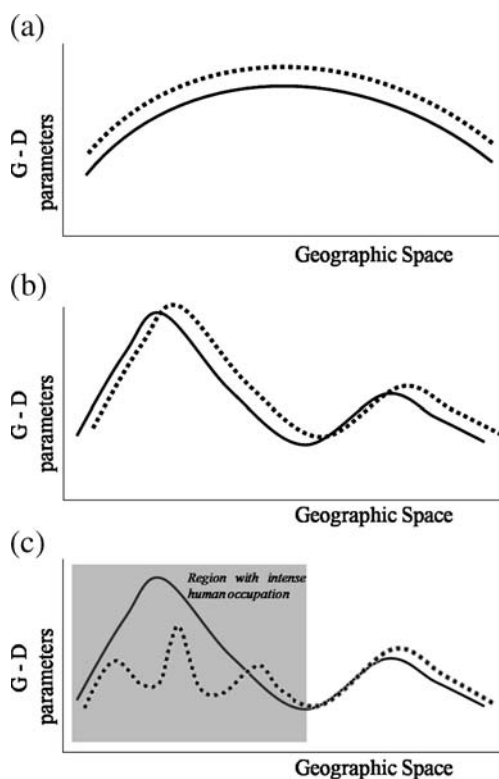


Fig. 1 Theoretical relationship between genetic and demographic parameters, such as ecological suitability (*solid line*) and genetic diversity (*dashed line*), in geographical space. In a classical central-periphery model (**a**), the low ecological suitability in the extremes of geographic distribution, expressed here across a single dimension, generates a reduction in genetic diversity (and, inversely, an increase in inbreeding). However, if ecological suitability has a more complex geographical pattern (**b**), the genetic variation can also track this pattern but does not generate central–peripheral pattern, even though the mechanistic basis would be similar. Finally, these patterns can be disrupted by human occupation increasing habitat loss and fragmentation (**c**), so even in a region with high ecological suitability, genetic diversity could be low due to these effects

(Maurer 1994; Sagarin and Gaines 2002; Sagarin et al. 2006). If a species does not possess a simple abundance peak in the geographical centre of its range, but instead has multiple abundance peaks determined by environmental conditions (Maurer 1994), then the basic assumption of genetic explanations does not hold and other mechanisms must be involved to explain the loss of genetic diversity. However, the overall reasoning of the abundant-centre hypothesis would still hold if there is a correlation between genetic diversity and relative position of local populations in niche space. The idea is that, although the largest populations (with higher genetic diversity, lower inbreeding rates and more connected to adjacent populations by gene flow) are not in the centre of the geographic distribution, they may exist in regions of high environmental suitability, and thus genetic parameters would track these more complex demographic patterns (Fig. 1b). In this case, the non-centrality of highly dense populations would generate a more complex pattern of genetic variation due to a different balance between processes driven by abundance (e.g., inbreeding and genetic drift) and asymmetric dispersal. Finally, in an explicit landscape genetics context (see Manel et al. 2003; Holderegger and Wagner, 2006), even central populations, both in geographical centre and in niche centre, can be affected by habitat loss and fragmentation and thus their genetic diversity or inbreeding levels are not within the expected values based on simple range-wide demographic processes (Fig. 1c).

Empirical analyses of these patterns (and consequently the choice among these alternative explanatory mechanisms) suffer from several problems. For instance, as an implicit shortfall of the macroecological research programme (e.g., Gaston and Blackburn 2000), it is operationally difficult to obtain detailed population parameters at broad geographic scales, so that comparisons of central and peripheral populations are usually based on small data sets, and worst of all, they usually do not encompass the entire extent of occurrence of a species (Eckert et al. 2008). Moreover, despite the currently wide use of molecular markers to estimate genetic variation at different geographical scales and originated from different evolutionary processes (see Diniz-Filho et al. 2008a for a recent review), it is still difficult to match these data with ecological parameters at broad scales. Also, despite many papers testing the simple relationship between genetic diversity and the geographical position of the populations, where we could expect that peripheral populations exhibit lower genetic diversity and higher genetic differentiation than central populations, only a few explicitly tested alternative mechanisms leading to such patterns (such as colonisation history and human effects) by integrating genetic and demographic in both ecological and historical (phylogeographical) frameworks (Eckert et al. 2008).

Here we developed an integrated framework to evaluate geographic patterns of genetic variability within local population and test the central–peripheral models. We applied this framework to evaluate patterns and processes involved in population structure of *Caryocar brasiliense* Camb. (Caryocaraceae), an abundant and widely distributed endemic tree from Brazilian Cerrado (Ratter et al. 2003; Walter and Guarino 2006). First, we showed that geographical patterns in different population genetic parameters (expected and observed heterozygosity, allelic richness and inbreeding levels) can furnish insights about processes involved in central–peripheral patterns. We then showed how these genetic data can be directly linked to estimates of ecological suitability based on niche modelling approaches, to demographic population structure (age-structure) and to surrogate variables of isolation and habitat fragmentation processes caused by human occupation in the biome. Previous links between species' distribution models and molecular data have been performed in the context of phylogeographical inference (e.g., Kidd and Ritchie 2006; Knowles et al. 2007; Jakob et al. 2007), but to our knowledge this is the first time

that niche models are used as macroecological surrogates of abundance and related to geographical genetic patterns. Although still suffering from some of the problems mentioned above (mainly the small number of local populations), our analyses show how central–peripheral population patterns can be investigated, offering a better understanding of the processes limiting species' geographic ranges.

Material and methods

Genetic data and analyses

Our basic data consist of molecular genetic data for ten local populations of *C. brasiliense*, an abundant and widely distributed tree species which is endemic to the savannah region (the “Cerrado”) in Central Brazil (Ratter et al. 2003). Although there is a relatively small number of local populations for which genetic data is available, these populations are widely distributed in the biome (Fig. 2; see Collevatti et al. 2001 for sampling details).

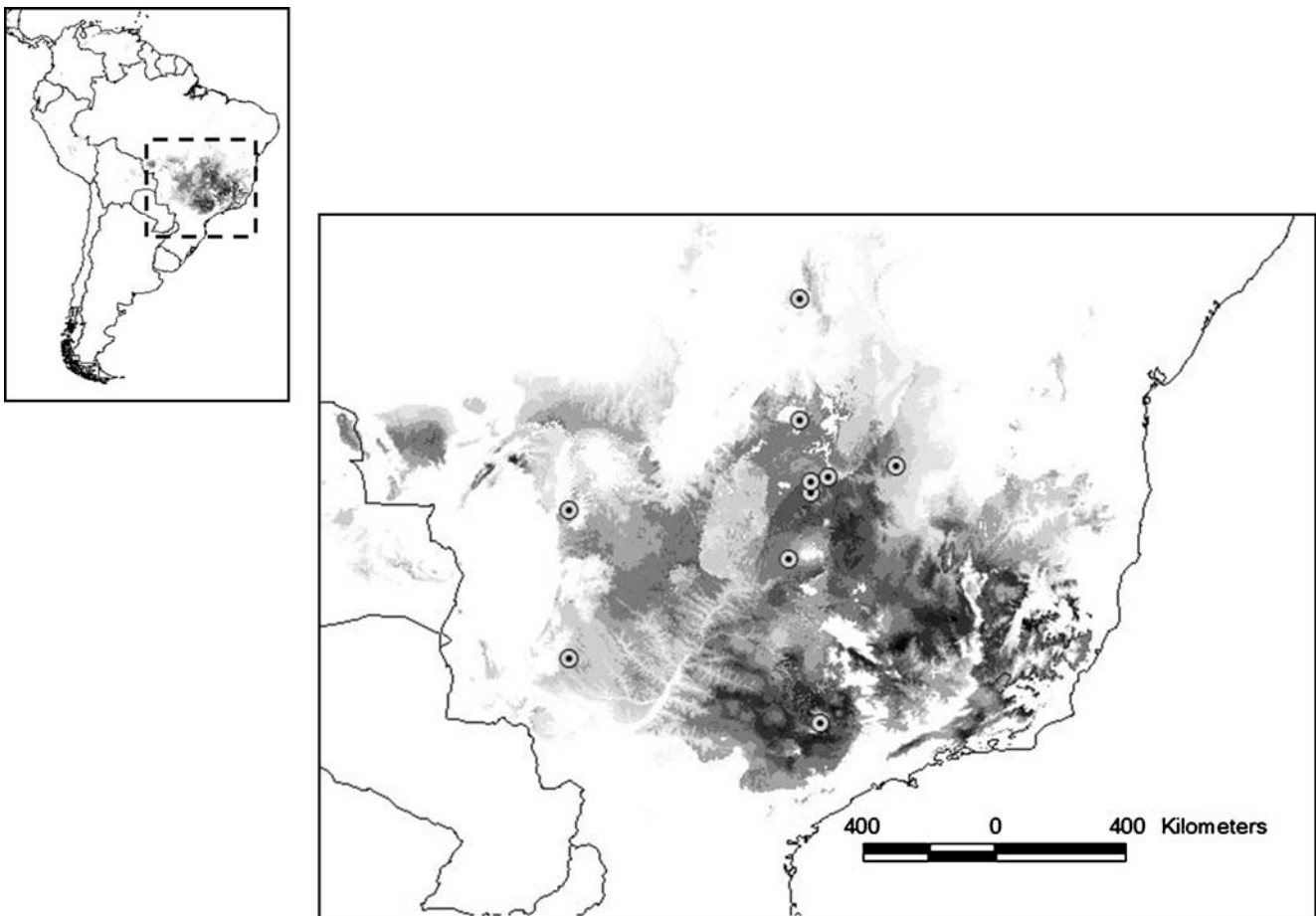


Fig. 2 Modelled geographic distribution of *Caryocar brasiliense*, shown as the average probability of occurrence from five distinct modelling techniques (BIOCLIM, Euclidian and Mahalanobis Distan-

ces, GARP and MAXENT), and the ten local populations for which genetic data are available from Collevatti et al. (2001, 2003). Darker regions indicate more overlap of models and more ecological suitability

A detailed description of genetic data can be obtained from Collevatti et al. (1999, 2001). In short, ten loci obtained from nuclear microsatellite markers, with number of alleles per locus ranging from 20 to 27, were used to genetically characterise 314 individuals of *C. brasiliense*. Sample sizes within local populations ranged from 30 to 41, so that estimates of genetic parameters for each local population are probably not biased by small samples sizes. Data from chloroplast (cpDNA) sequences and microsatellites were used to describe phylogeographical history of the populations (Collevatti et al. 2003; see below). Our analyses are based mainly on three different genetic parameters estimated from nuclear genetic data within each local population. First, the expected heterozygosity (H_e) is given by

$$H_e = 1 - \sum p_i^2$$

where p_i is the frequency of alleles across all loci. Then, H_e is a measure of genetic diversity that takes into account both frequency and richness of alleles within local populations, whereas the observed heterozygosity (H_o) is given by the frequency of heterozygotes in the populations. By comparing the H_e and H_o , it is possible to define the inbreeding level of a local population, usually referred as F_{IS} , in a hierarchical context of population structure, on simply f , which is given by

$$f = (H_e - H_o)/H_o$$

so that high f -values indicate an excess of homozygous genotypes in populations, higher than expected by random combination of alleles (Hardy–Weinberg equilibrium). High f -values are expected in small populations due to non-random mating even in large populations (Ginzburg and Golenberg 1985; Avise 2004).

Analyses of genetic structure of central–peripheral populations also emphasise that, in range edges, isolation will create more divergence among populations because drift within each population (due to their low local abundances) will drive them into different “directions” in genetic space (Eckert et al. 2008). So, we also measured genetic divergence among populations and expressed them in a pairwise matrix of F_{ST} values, estimated by φ_P from an analysis of molecular variance (Excoffier et al. 1992), taking into account allele size (Slatkin’s R_{ST} statistics) and evaluated geographical patterns using simple techniques of matrix correspondence (Manly 1985; Legendre and Legendre 1998; see below).

A pairwise matrix of F_{ST} values was also obtained using haplotypes from microsatellites of the cpDNA (see Collevatti et al. 2003). Because cpDNA is uni-parentally inherited, these markers can provide a useful estimate of historical, or genealogical, relationships among local populations and help to disentangle historical and recent processes driving genetic

divergence. The first eigenvector of this matrix (Diniz-Filho et al. 1998, 1999) was used as a predictor of “historical” components driving geographical and central–peripheral patterns within populations.

Ecological data

For each of the ten populations, Collevatti et al. (2001) also provided some ecological data that can help the understanding of processes underlying genetic variation within populations. An estimate of the percentage of juveniles (J) in the areas sampled was obtained by counting the proportion of individuals with less than 10 cm circumference at 30 cm above the ground. A low value of J may suggest then recruitment failures or high juvenile mortality, both of them leading to low intrinsic growth rates and local abundances. At the same time, this variable may also indicate a time-lag in microevolutionary and ecological processes inferred, since in local populations with low J the genetic parameters are being estimated for old individual trees, reflecting past events not necessarily still affecting the current populations.

We also searched in the literature for occurrence records of *C. brasiliense* (compiled from many papers and from the online database available in the “Centro de Referência em Informação Ambiental”—CRIA; <http://www.cria.org.br/>) and used five different niche models (or species distribution models—SDM) based on presence-only data (Tsoar et al. 2007). Species’ potential geographic distribution was modelled using BIOCLIM in DIVA-GIS (Hijmans et al. 2005), Euclidian and Mahalanobis Distances (Farber and Kadmon 2003), GARP (Stockwell and Peterson 2002) and MAXENT (Phillips et al. 2006, Phillips and Dudik 2008) (see also Segurado and Araújo 2004; Elith et al. 2006; Tsoar et al. 2007 for comparative evaluations of the performance of SDM techniques). A total of 130 records (occurrence points) were used in the modelling process, which was based on seven environmental variables (altitude, annual mean temperature, maximum temperature of warmest month, mean temperature of wettest quarter, annual precipitation, precipitation of warmest quarter and precipitation of coldest quarter) derived from the WORLDCLIM (<http://www.worldclim.org/>) and converted to a grid resolution of 0.0417 degrees.

The quantitative variables produced as output of the different SDMs (e.g., the output of Maxent, the number of models that predict occurrence in a given cell using GARP or the inverse of the Euclidean/Mahalanobis distances) was standardised to vary between 0 and 1 throughout the domain and can be considered as at least monotonically related to ecological “suitability” of local populations under the niche model that underlies all these methods (Anderson and Martinez-Meyer 2004; Peterson 2001; Phillips et al. 2006; Araújo and Guisan 2006). Thus, we defined the

ecological suitability S of each local population from Collevatti et al. (2001) as the average value of occurrence provided by each model, weighted and un-weighted by the performance of each model estimated using the area under the curve (AUC) from a receiver operator curve (ROC) analysis (Luoto et al. 2005; but see Lobo et al. 2008). Under the idea of ensemble forecasting (Araújo and New 2007), the distribution of *C. brasiliense* can be conservatively defined as encompassing all cells where all models suggested probabilities of occurrence higher than the threshold established by the ROC curve of each model.

Human occupation data at landscape level

Human occupation can be also important in altering isolation patterns among local populations (Manel et al. 2003; Holderegger and Wagner 2006). Thus, we used two metrics of habitat loss and fragmentation (a growing problem in Brazilian Cerrado, see Klink and Machado 2005) to evaluate the potential effects of human pressure on the observed patterns of genetic variation within local populations. First, we recorded the value of the “Human Footprint” (HF) in each locality, obtained from the map generated by Sanderson et al. (2002), which combined global records of population density, land use, access to humans (roads, rivers, etc) and electrical power infrastructure to create this map, at a resolution of 1 km² (data available at: http://www.ciesin.columbia.edu/wild_areas/) (accessed May 2007).

We also used, as a surrogate of human impacts, a regional estimate the amount of habitat remnants (R) surrounding each local population. Data on habitat loss were derived from the Cerrado Remnant Vegetation Map, recently released within the scope of the Brazilian Ministry of the Environment Biodiversity Programme (Brazilian Ministry of Environment (MMA), Conservation and Sustainable Use of Brazilian Biodiversity Project (PROBIO), which is freely available at <<http://mapas.mma.gov.br/mapas/aplic/probio/datadownload>>). This map, comprising both a wide range of remnant physiognomies and converted areas, was produced at the 1:250,000 scale, relying on unsupervised classification (image segmentation) and visual interpretation of 114 Landsat scenes from 2002 (mostly from August, September and October), analysis of MODIS vegetation index temporal datasets, ancillary landscape information from the former RadamBrasil project and ground truth (see Sano et al. 2008 for details). As continued validation initiatives have demonstrated, this map represents the most detailed knowledge of the biome’s remaining vegetation and human-induced changes currently available. For this particular study, only the remnant physiognomies, grouped into a single remnant class, were considered. As continued validation initiatives have demonstrated, this map represents the most detailed knowledge of the biome’s

remnant vegetation and human-induced changes currently available. The value of R was the calculated proportion of natural remnants around each point, using a grid of 1° of latitude and longitude (about 100,000 km²) covering the Brazilian Cerrado (Diniz-Filho et al. 2006).

Data analysis

Central–peripheral patterns were tested, for all variables analysed here, by comparing the improvement of the coefficient of determination (R^2) of a quadratic trend surface analysis (TSA; a parabolic pattern) in respect to a simpler linear trend surface analysis (modelling a north–south or east–west linear trend), based on the statistical significance of partial regression slopes of the quadratic polynomial expansion. The Akaike information criterion (AIC) was not used here because its penalty term, which is a function of the number of estimated parameters, is very stringent when small sample sizes are used and makes the evaluation of improvement very sensitive to model definition (i.e., choice of which quadratic terms are added to the model). Although the better fit of a quadratic TSA does not necessarily ensure complete symmetry of central–peripheral variation, it shows that at least there is a non-linear trend towards the borders. Because of low statistical power and over-parameterization in the quadratic trend surface models, we also performed another test of central–peripheral patterns in genetic and ecological data by computing the correlation between each variable and the geographic distance of the populations to the centre of the Cerrado biome (which largely coincides with the centre of distribution of *C. brasiliense*).

Relative divergences among central populations and among peripheral populations were compared with a t -test on the R_{ST} values (see Eckert et al. 2008). To some extent, we arbitrarily defined central populations as those for which there is at least one population between it and the range edge, along a Gabriel network (Legendre and Legendre 1998; Fig. 1). However, R_{ST} values were spatially patterned (see “Results”) and therefore, they cannot be considered independent within these categories. We tested the spatial dependence of R_{ST} values with a Mantel test of matrix correlation (Manly 1985; Legendre and Legendre 1998; see also Collevatti et al. 2001), and the statistical significance of the resultant correlation was determined by a Monte Carlo procedure with 999 permutations. Thus, to correct for differences in distances among populations (by definition larger between pairs of peripheral populations than between pairs of central populations) we performed an analysis of covariance (ANCOVA) (Sokal and Rohlf 1995), comparing R_{ST} within categories (pairs of central and peripheral populations), while keeping the pairwise geographical distances statistically constant.

We used simple and multiple ordinary least-squares (OLS) regression models to relate genetic parameters to explanatory variables (Legendre and Legendre 1998). Residuals were checked for spatial autocorrelation using Moran's I correlograms and their statistical significance was tested with 1,000 random permutations (Diniz-Filho et al. 2003; Diniz-Filho and Telles 2002). Presence of significant autocorrelation in OLS residuals inflate Type I errors and therefore, conclusions derived from null-hypothesis significance tests may be misleading (Hawkins et al. 2007). Recently, some of us also found that even model selection procedures based on AIC are also subject to potential flaws (Diniz-Filho et al. 2008b). Because of high multicollinearity among explanatory variables and the low statistical power, we modelled the genetic variables (H_e , A and F) using different OLS models, separating the four explanatory variables into two sets: first, the "ecological" effects were indicated by the variables J and S , and, second, the effects of human occupation were expressed by R and HF. We then used a partial regression approach to decouple the effects of human occupation and ecological variables in the genetic variability, based on adjusted R^2 (see Legendre and Legendre 1998). All regression and autocorrelation analyses were performed in SAM 3.0 (Rangel et al. 2006), freely available at www.ecoevol.ufg.br/sam.

Results

Central–peripheral patterns in genetic and ecological data

Out of the three local genetic parameters estimated, only H_e showed a clear central–peripheral pattern according to the improvement in the quadratic TSA model, with a gain in coefficient of determination (ΔR^2) equal to 0.509 (Table 1). The negative correlation between H_e and distance from the centre of the biome is not statistically significant, but it indicates that higher values of H_e are found in central populations, with very low values in south and west edges

(Fig. 3a). No clear pattern is observed for A and, indeed, the explanatory power of both linear and quadratic models is low (Table 1), although reduction of values toward the west appears in the map (Fig. 3b). The pattern in f , on the other hand, is much better described by a north–south cline than by a central–peripheral pattern, and higher inbreeding was found in populations more distant from the centre of the biome in the southern border (Fig. 3c). Thus, peripheral populations tend to have lower diversity and, simultaneously, higher levels of inbreeding are found in the southern border of the geographic range. The first eigenvector of the cpDNA F_{ST} matrix, expressing "genealogical" connections between populations, was not correlated with the distance from the geographic centre of the Cerrado ($r=0.301$; $P=0.402$), showing thus no central–periphery pattern. There is a strong quadratic trend ($R^2=0.934$, improved over the linear model with R^2 of 0.541), and also positively correlated with inbreeding values f (0.624; $P=0.054$) (see below).

The average R_{ST} among central populations ($R_{ST}=0.164\pm 0.082$ SD) and among peripheral populations ($R_{ST}=0.305\pm 0.218$ SD) were not statistically significant ($t=1.91$; $P=0.08$). Difference between variances was, however, highly significant and much larger in the periphery than in the centre ($P\ll 0.001$). Since a significant spatial pattern in R_{ST} is observed (Mantel's $r=0.518$; $P=0.032$), this can explain pairwise differences in R_{ST} among central and among peripheral populations. Indeed, after taking geographic distances among local populations into account (ANCOVA), the difference was even lower (adjusted mean difference in R_{ST} between peripheral and central population=0.07). Historical divergence expressed by pairwise F_{ST} from cpDNA is not correlated with population divergence in nuclear markers measured by R_{ST} (Mantel's $r=0.0154$; $P=0.496$).

The values of ecological suitability derived from all niche modelling techniques do not showed a clear central–peripheral pattern, although the improvement in quadratic TSA fit is relatively high due to a non-linear increase in suitable areas towards southeastern region of Cerrado (Table 1). Indeed, the correlation with distance from the

Table 1 Analysis of spatial patterns in genetic, ecological and human occupation variables, including linear correlation with distance from the centre of the biome ($r(\text{centre})$) and adjusted R^2 from linear and quadratic trend surface models

Variable	$r(\text{centre})$	adj. R^2 (linear)	adj. R^2 (quadratic)	ΔR^2
Genetic				
H_e	−0.466	0.477	0.986	0.509**
A	−0.223	0.179	0.007	−0.172
f	0.589	0.441	0.622	0.181
Ecological				
S	−0.055	0.741	0.924	0.183*
J	−0.559	0.432	0.894	0.462*
Human occupation				
HF	0.492	0.583	0.444	−0.139
R	−0.714	0.807	0.804	−0.003

The improvement in quadratic model (ΔR^2) indicates a non-linear pattern and was tested by the statistical significance of partial regression of polynomial terms

* $P<0.05$; ** $P<0.01$

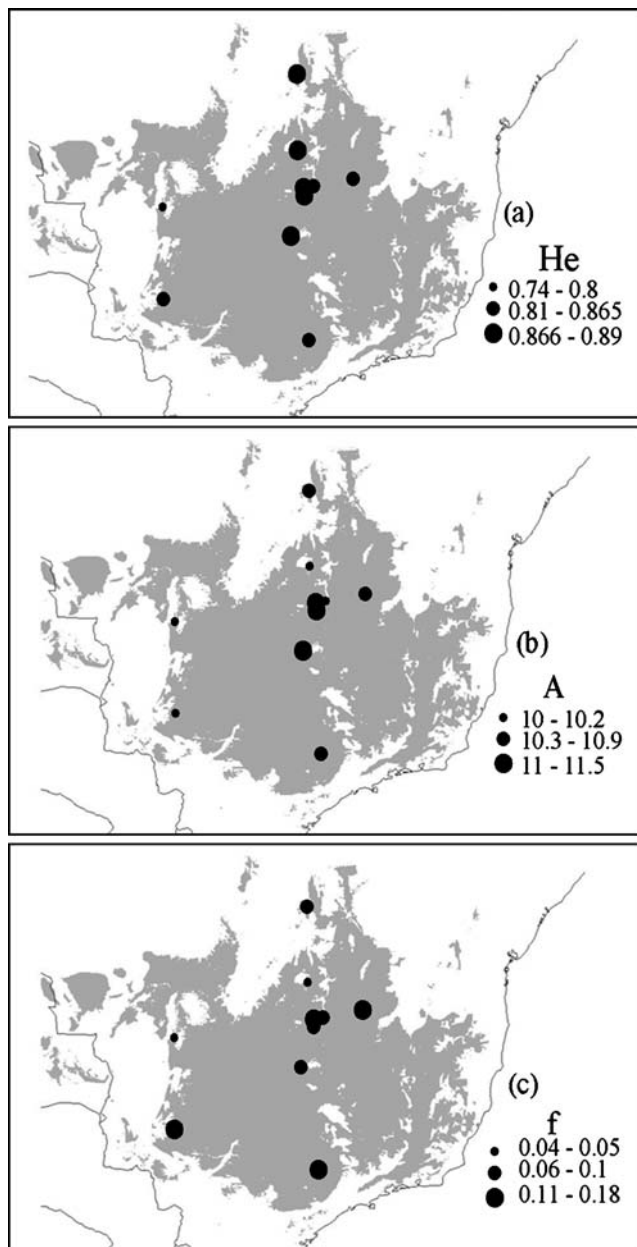


Fig. 3 Geographical patterns of genetic parameters estimated within populations, H_e (a), A (b) and f (c), overlapping the potential distribution of *Caryocar brasiliense*

centre of the biome is close to zero (Fig. 1). All distribution models possessed high fit according to the AUC criterion (i.e., $AUC > 0.9$ for all five methods), so that using the AUC weighted average probability of occurrence to express

ecological suitability did not qualitatively change these interpretations.

On the other hand, the frequency of juveniles J shows a relatively high correlation with distance to centre ($r = -0.56$; $P = 0.098$), and there was a significant gain in the adjusted R^2 by extending the linear model to a quadratic model ($\Delta R^2 = 0.462$), reflecting a centre–periphery pattern in which central populations have more juveniles than peripheral populations.

Human footprint (HF) and percentage of remnants (R) did not show clear central–peripheral patterns when evaluated isolated. Instead, north–south and east–west linear gradients are, respectively, observed (Table 1). HF tends to increase towards the west portion of the biome, whereas there is a north–south gradient in the percentage of remnants. Indeed, there is a significant negative correlation between R and distance from the centre of the biome ($r = -0.714$; $P = 0.024$), mainly driven by high fragmentation in the southern border of Cerrado due to increase economical activities in this region (see Rangel et al. 2006). Thus, there is a central–peripheral pattern of overall human occupation in the biome, expressed by a combination of north–south and east–west gradients in the two variables analysed.

Quantifying the effects of ecological suitability and human occupation variables on population genetics

In general, average number of alleles per locus (A) was not well explained by the ecological and human occupation variables, with R^2 values lower than 0.1. For both H_e and f , the effects of human occupation were higher than the effects of ecological variables (Table 2), but with different relative magnitudes. Partial regression results indicated that almost 46% of the variation in f can be explained by effects of human occupation alone, whereas for H_e the effects of human occupation and ecological variables overlap to a great extent (Table 2). There is no residual autocorrelation in these models, according to Moran’s I for the first distance class (about 150 km).

The first eigenvector extracted from the cpDNA F_{ST} matrix has only a minor individual effect on H_e and A , much smaller than those observed for ecological factors and human occupation (correlations lower than 0.25 in both cases). However, there is a relatively high correlation between the first eigenvector and the inbreeding levels (f)

Table 2 Partial regression analysis of ecological and human occupation effects on within-population genetic variation (as measured by H_e , A and f) as a function of geographical patterns

	R^2 (full)	R^2 (ecology)	R^2 (humans)	Ecology	Overlap Partials	Humans
H_e	0.318	0.201	0.284	0.034	0.167	0.117
A	0.098	0.009	0.081	0.017	-0.008	0.089
f	0.643	0.186	0.496	0.147	0.039	0.457

($r=0.625$; $P=0.054$). After allowing for the effects of variables expressing human occupation and ecological conditions, however, the relationship between the first eigenvector and f was cancelled out ($P=0.074$).

Discussion

Our analyses reveal some central–peripheral patterns in within-population genetic diversity of *C. brasiliense* in the Cerrado region of Central Brazil, despite the small number of local populations analysed. However, the complex correlation patterns found between different genetic estimates with ecological and human impact variables indicate that it may be difficult to disentangle the processes involved in the origin and maintenance of these patterns and they do not corroborate that these patterns arise under simple abundant-centre models.

Indeed, due to the predominance of studies focusing on the northern limit of species in the temperate zone (see Eckert et al. 2008), it has been difficult to disentangle the effects of historical processes of colonisation, related to glaciation cycles in the extreme northern hemisphere, from the influences of current processes of recent population isolation and small effective population size (e.g., Howes and Lougheed 2008). In our study system, although we cannot rule out the effects of Pleistocene climate change related to glacial refugia (see below), these effects were quite probably much smaller than those in temperate regions. Our analyses in the middle of the tropical region thus support central–peripheral patterns in some genetic parameters that may be related to variation in demographic parameters and to human effects in the southern border of the species' distribution.

First of all, to corroborate the classical central–peripheral model based on the abundant-centre model it would be important to demonstrate a progressive decline in abundance and an increase in the level of geographical isolation from central regions towards the range limits of the species of interest. Although we do not have data on abundance, we assessed, based on geographic occurrences of local populations, if peripheral populations were located in less “suitable” regions of the niche space (Anderson and Martinez-Meyer 2004). This variable can be a macro-ecological surrogate for abundances and, at the same time, can ensure that the logic underlying the abundant-centre model could be applied even if more complex spatial patterns in abundance are found in nature (because high genetic diversity could appear in the centre of ecological space, even when this does not coincide with species' range geographical centre). Even so, small populations may still occur in regions with relatively high ecological suitability due to factors creating non-equilibrium between distribu-

tion and climate and thus unrelated to environmental niche such as, for instance, lack of dispersion, biotic interactions and human-induced effects, which would also affect other demographic processes (Holt and Keit 2005; Araújo and Pearson 2005). In our analyses, however, we did not find a clear central–peripheral pattern in the ecological suitability for the populations analysed, based on multiple niche models, which suggest that more complex processes may be involved in the origin of central-periphery pattern in genetic variability, beyond a simple abundant-centre model (both in ecological and geographical spaces).

How can these findings be linked to central–peripheral patterns and clines observed for different metrics used to express within-population genetic variability? Although differences between geographical patterns found for H_e , A and f have been usually explained in “statistical” terms (see Eckert et al. 2008), we believe that they are ecologically and genetically interpretable. Mainly for H_e , there is a relatively clear central–peripheral pattern, but this is not well explained neither by ecological or human impact factors (the full R^2 is around 30%). Due to the relatively short time for population differentiation and long life cycle of the species (Collevatti et al. 2001), it is more likely that patterns appear in H_e , indicating that reduction in genetic diversity would occur mainly in the quantitative component (i.e., dominance) of one allele over the others, so that patterns observed would be “transitory” in time. Indeed, low H_e values are found in peripheral populations, mainly at south and west range edges, probably as a result of genetic drift due to isolation-by-distance and reduction in population effective size (which may be at least in part independent of current habitat suitability, at least as expected by niche models used here). At the same time, Mantel tests and ANCOVA show that diversity among the peripheral populations is not larger than among central ones. This reinforces the short time for divergence, since although there is a reduction in H_e in each population in the periphery, the time was not enough to generate more difference among them than expected by their geographic distances, following then an isolation-by-distance process in genetic similarity.

The percentage of juveniles (J) of *C. brasiliense* shows a much clearer central–peripheral pattern than ecological suitability and this could be in principle also associated with genetic diversity by different mechanisms (see Lönn and Prentice 2002). At least, our analyses show (Table 2) that there is a shared geographic effect between ecological (mainly J) and genetic diversity, and H_e is the genetic parameter with highest (although not statistically significant) correlation with J . This association may be generated by assuming that low values of J indicate failure in recruitment and thus population might be experiencing population reduction. The association could also indicate much recent “time-lag” effects, since genetic parameters

estimated in a population with low J are actually reflecting processes that happened in previous generations, which in the case of large trees such as “pequi” can be 50 years ago. This is before the intensification of human occupation in the Cerrado, and in this case we expect that genetic parameters in populations with low J are not reflecting, for example, recent human impacts on abundance and isolation. However, we observed that there may be a negative correlation J and human footprint ($r=-0.602$; $P=0.065$), so that populations with more young individuals are found in regions with low human impacts and this may suggest that the first demographic explanation above (not simple time-lag effects) are more likely.

On the other hand, a much clearer pattern is observed for f , forming a strong north–south clinal. Although there is a relatively high correlation between f and distance from the centre of the biome, these f -values are much higher in the southern edge of the biome (see Fig. 2). This geographical pattern can be explained by different (but not mutually exclusive) mechanisms and they are supported by different correlations found here. Of course, it is difficult to disentangle direct causal links based on these observational data, but a more likely scenario can be possibly built-up.

First, by considering what is known about recent occupation in Cerrado (Klink and Moreira 2002; Klink and Machado 2005) and the partial central–peripheral patterns observed for human occupation, it is possible to infer, in a landscape genetics context, that peripheral populations of *C. brasiliense* in the southern border of the biome would have higher levels of inbreeding due to isolation and reduction in abundance caused by human occupation driving habitat loss and fragmentation (see Telles et al. 2007; Soares et al. 2008). It is particularly difficult to decouple effects of drift and inbreeding because the two processes are linked to a common factor (reduction in effective population size) and because human occupation creates not necessarily reductions in population abundance. *C. brasiliense* is a tree used by humans for several purposes, so these trees are preserved in agricultural fields or pastures because their fruits are much appreciated and because they provide shadows for the herd. Thus, except under broad-scale deforestation processes that indeed occurred in the southern part of the biome, these trees are preserved in human occupied landscapes. At small scales, habitat loss may even improve the gene flow for tree species that are maintained in an area, but this may be quite problematic at larger geographical scales because connection among local populations will depend mainly on rare long-distance dispersal events. Under this reasoning, the north–south clines observed in f and better explained by human occupation make even more sense. Also, due to extensive human occupation, J may be also affected and amplify the patterns observed in the southern border of the biome.

Alternatively, it is possible to consider that patterns in f and genetic diversity are also related to historical processes, which could also explain the relationship between f and the first eigenvector from cpDNA distances. *C. brasiliense* is endemic to the Cerrado vegetation type, despite its wide geographic distribution in Central South America. Thus, the delimitation of its spatial edges must be necessarily linked with definitions of habitat and ecological requirements at the biome level. This is usually attributed to environmental components, such as seasonal precipitation, fire regime, soil fertility and drainage, as well as historical fluctuations due to climate changes (Oliveira-Filho and Ratter 2002; Ledru 2002). Because of the intrinsic association between a species’ range and the ecological characteristics at the biome level, it is expected that part of the genetic variation could be explained by shifts at these scales and would be linked with Pleistocene climate fluctuations that affected the limits of Cerrado, Amazon and Atlantic forest in the last 10,000 years (see Toledo and Bush 2007; Caetano et al. 2008). Although these fluctuations can help understand phylogeographic patterns within the species and eventually its origin from other species (probably in the Amazonian region), they are probably not strongly related to clinal and central–peripheral patterns discussed here. For instance, populations in the southern border, in São Paulo state, would be formed by populations from different parts of the centre of the species’ distribution, in Goiás and Mato Grosso (Collevatti et al. 2003). Thus, current populations in the southern border of the biome would have an even lower level of inbreeding, as they are composed by lineages from different and more central populations. This explains why the effect of the first eigenvector of cpDNA similarity matrix is cancelled out after taking human occupation into account, reinforcing the role of isolation.

By coupling demographic data, niche modelling and processes of human occupation, we were able to build a coherent scenario explaining how different genetic parameters vary from central towards peripheral local populations. Thus, our analyses suggests that isolation-by-distance induced by human effects, and the consequent increase in inbreeding levels, are beginning to affect the genetic variability within local populations at the southern border of the biome. In general, all these processes accentuate the central–marginal patterns and, at the same time, create directional patterns in genetic variability. This can trigger a contraction of geographic range, which will have unknown long-term consequences for the conservation of this species in the future.

Acknowledgments We thank an anonymous reviewer for useful suggestions that improved the manuscript and clarified our ideas about this issue. Our programme in conservation biogeography and molecular ecology in the Brazilian Cerrado is supported by a PRONEX project from CNPq/SECTEC-GO (Proc. No. 23234156) for establishing conservation priorities in the Cerrado region and by

several grants and fellowships from CNPq, CAPES and FUNAPE-UFJF. Cooperation between genetic laboratories of MPCT and RGC are supported by a PROCAD-NF/CAPES programme, whereas work by JAFDF, LMB and PDM on distribution modelling have been supported by BIOIMPACT project for evaluating effect of climate change in the Neotropics, coordinated by M. B. Araújo.

References

- Alleaume-Benharira M, Pen IR, Ronce O (2006) Geographical patterns of adaptation within a species' range: interactions between drift and gene flow. *J Evol Biol* 19:203–215
- Anderson RP, Martinez-Meyer E (2004) Modeling species' geographic distributions for preliminary conservation assessments: an implementation with the spiny pocket mice (*Heteromys*) of Ecuador. *Biol Conserv* 116:167–179
- Araújo MB, Pearson RG (2005) Equilibrium of species' distributions with climate. *Ecography* 28:693–695
- Araújo MB, Guisan A (2006) Five (or so) challenges for species distribution modelling. *J Biogeogr* 33:1677–1688
- Araújo MB, New M (2007) Ensemble forecasting of species distributions. *Trends Ecol Evol* 22:42–47
- Avise JC (2004) Molecular markers, natural history and evolution. Sinauer Associates, Sunderland
- Brown JH (1984) On the relationship between abundance and distribution of species. *Am Nat* 124:255–279
- Caetano S, Prado D, Pennington RR, Oliveira-Filho A, Spichiger R, Naciri Y (2008) The history of seasonally dry tropical forests in eastern South America: inferences from the genetic structure of the tree *Astronium urundeuva* (Anacardiaceae). *Mol Ecol* 17:3147–3159
- Collevatti RG, Brondani RPV, Grattapaglia D (1999) Development and characterization of microsatellite markers for genetic analysis of a Brazilian endangered species *Caryocar brasiliense*. *Heredity* 83:748–756
- Collevatti RG, Grattapaglia D, Hay JD (2001) Population genetic structure of the endangered tropical tree species *Caryocar brasiliense*, based on variability at microsatellite loci. *Mol Ecol* 10:349–356
- Collevatti RG, Grattapaglia D, Hay JD (2003) Evidences for multiple maternal lineage origin of *Caryocar brasiliense* populations in Brazilian Cerrado based on the analysis chloroplast DNA sequences and microsatellites haplotypes. *Mol Ecol* 12:105–115
- Diniz-Filho JAF, Telles MPC (2002) Spatial autocorrelation analysis and the identification of operational units for conservation in continuous populations. *Conserv Biol* 16:924–935
- Diniz-Filho JAF, Sant'ana CER, Bini LM (1998) An eigenvector method for estimating phylogenetic inertia. *Evolution* 52:1247–1262
- Diniz-Filho JAF, Fuchs S, Arias MC (1999) Phylogeographic autocorrelation of phenotypic evolution in honey bees (*Apis mellifera*). *Heredity* 83:671–680
- Diniz-Filho JAF, Bini LM, Hawkins BA (2003) Spatial autocorrelation and red herrings in geographical ecology. *Global Ecol Biogeogr* 12:53–64
- Diniz-Filho JAF, Bini LM, Pinto MP, Rangel TFLVB, Carvalho P, Bastos RP (2006) Anuran species richness, complementarity and conservation conflicts in Brazilian Cerrado. *Acta Oecol* 29:9–15
- Diniz-Filho JAF, Telles MPC, Bonatto SL et al (2008a) Mapping the evolutionary twilight zone: molecular markers, populations and geography. *J Biogeogr* 35:753–763
- Diniz-Filho JAF, Rangel TFLVB, Bini LM (2008b) Model selection and information theory in geographical ecology. *Global Ecol Biogeogr* 17:479–488
- Eckert CG, Samis KE, Loughheed SC (2008) Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Mol Ecol* 17:1170–1188
- Elith J, Graham CH, Anderson RP et al (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129–151
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491
- Farber O, Kadmon R (2003) Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. *Ecol Model* 160:115–130
- Gaston KJ (2003) The structure and dynamics of geographic ranges. Oxford University Press, Oxford
- Gaston KJ, Blackburn TM (2000) Pattern and process in macroecology. Blackwell, London
- Ginzburg LR, Golenberg EM (1985) Lectures in theoretical population biology. Prentice-Hall, Englewood Cliffs
- Hawkins BA, Diniz-Filho JAF, Bini LM, De Marco P, Blackburn TM (2007) Red herrings revisited: spatial autocorrelation and parameter estimation in geographical ecology. *Ecography* 30:375–384
- Hijmans RJ, Guarino L, Bussink C, Barrantes I, Rojas E (2005) DIVA-GIS, Version 5.2. <http://www.diva-gis.org> (last accessed 14 July 2008)
- Hoffman AA, Blows MW (1994) Species borders: ecological and evolutionary perspectives. *Trends Ecol Evol* 9:223–227
- Holderegger R, Wagner HH (2006) A brief guide to landscape genetics. *Landsc Ecol* 21:793–796
- Holt RD, Keitt TH (2005) Species' borders: a unifying theme in ecology. *Oikos* 18:3–6
- Howes BJ, Loughheed SC (2008) Genetic diversity across the range of a temperate lizard. *J Biogeogr* 35:1269–1278
- Jakob SS, Ihlow A, Blattner FR (2007) Combined ecological niche modelling and molecular phylogeography revealed the evolutionary history of *Hordeum marinum* (Poaceae)—niche differentiation, loss of genetic diversity and speciation in Mediterranean Quaternary refugia. *Mol Ecol* 16:1713–1727
- Kidd DM, Ritchie MG (2006) Phylogeographic information systems: putting the geography into phylogeography. *J Biogeogr* 33:1851–1865
- Kirkpatrick M, Barton NH (1997) Evolution of a species' range. *Am Nat* 150:1–23
- Klink CA, Moreira AG (2002) Past and current human occupation, and land use. In: Oliveira PS, Marquis RJ (eds) The Cerrados of Brazil. Columbia University Press, New York, pp 69–88
- Klink CA, Machado RB (2005) Conservation of the Brazilian Cerrado. *Conserv Biol* 10:710–713
- Knowles LL, Carstens BC, Keat ML (2007) Coupling genetic and ecological-niche models to examine how past population distributions contribute to divergence. *Curr Biol* 17:940–946
- Ledru MP (2002) Late Quaternary history and evolution of the cerrados as revealed by palynological records. In: Oliveira PS, Marquis RJ (eds) The Cerrados of Brazil. Columbia University Press, New York, pp 35–50
- Legendre P, Legendre L (1998) Numerical ecology. Elsevier, Amsterdam
- Lenormand T (2002) Gene flow and the limits to natural selection. *Trends Ecol Evol* 17:183–189
- Lobo JM, Jimenez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecol Biogeogr* 17:145–151
- Lönn M, Prentice HC (2002) Gene diversity and demographic turnover in central and peripheral populations of the perennial herb, *Gypsophila fastigiata*. *Oikos* 99:489–498

- Luoto M, Poyry J, Heikkinen RK, Saarinen K (2005) Uncertainty of bioclimate envelope models based on the geographical distribution of species. *Global Ecol Biogeogr* 14:575–584
- Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol Evol* 18:189–197
- Manly BFJ (1985) The statistics of natural selection. Chapman & Hall, London
- Maurer BA (1994) Geographical population analysis: tools for the analysis of biodiversity. Blackwell Scientific, Oxford
- Oliveira-Filho AT, Ratter JA (2002) Vegetation physiognomies and woody flora of the cerrado Biome. In: Oliveira PS, Marquis RJ (eds) *The Cerrados of Brazil*. Columbia University Press, New York, pp 91–120
- Peterson AT (2001) Predicting species' geographical distribution based on ecological niche modelling. *Condor* 103:599–605
- Phillips S, Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 190:231–259
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecol Model* . doi:10.1016/j.ecolmodel.2005.03.026
- Rangel TFLVB, Diniz-Filho JAF, Bini LM (2006) Towards an integrated computational tool for spatial analysis in macroecology and biogeography. *Global Ecol Biogeogr* 15:321–327
- Ratter JA, Bridgewater S, Ribeiro JF (2003) Analysis of the floristic composition of the Brazilian Cerrado vegetation. III: comparison of the woody vegetation of 376 areas. *Edinburgh J Bot* 60:57–109
- Sagarin RD, Gaines SD (2002) The 'abundant centre' distributions: to what extent is it a biogeographical rule? *Ecol Lett* 5:137–147
- Sagarin RD, Gaines SD, Gaylord B (2006) Moving beyond assumptions to understand abundance distributions across the range of species. *Trends Ecol Evol* 21:524–530
- Sanderson EW, Jaiteh M, Levy MA, Redford KH, Wannebo AV, Woolmer G (2002) The human footprint and the last of the wild. *Bioscience* 52:891–904
- Sano EE, Rosa R, Brito JLS, Ferreira LG (2008) Mapeamento semidetalhado do uso da terra do Bioma Cerrado. *Pesq Agropec Bras* 43:153–156
- Segurado P, Araújo MB (2004) A evaluation of methods for modelling species distributions. *J Biogeogr* 31:1555–1568
- 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
- Sokal RR, Rohlf FJ (1995) *Biometry*, 3rd edn. Freeman, San Francisco
- Soulé M (1973) The epistasis cycle: a theory of marginal populations. *Annu Rev Ecol Syst* 4:165–187
- Stockwell DRB, Peterson AT (2002) Effects of sample size on accuracy of species distribution models. *Ecol Model* 148:1–13
- Telles MPC, Diniz-Filho JAF, Bastos RP, Soares TN, Guimarães LD, Lima LP (2007) Landscape genetics of *Physalaemus cuvieri* in Brazilian Cerrado: correspondence between population structure and patterns of human occupation and habitat loss. *Biol Conserv* 139:37–46
- Toledo MB, Bush MA (2007) A mid-holocene environmental change in Amazonian savannas. *J Biogeogr* 34:1313–1326
- Tsoar A, Allouche O, Steinitz O, Rotem D, Kadmon R (2007) A comparative evaluation of presence-only methods for modelling species distributions. *Divers Distrib* 13:397–405
- Walter BMT, Guarino ESG (2006) Comparação do método de parcelas com o "levantamento rápido" para amostragem da vegetação arbórea do Cerrado sentido restrito. *Acta Bot Braz* 20:285–297