npg

ORIGINAL ARTICLE

Demographic history and the low genetic diversity in *Dipteryx alata* (Fabaceae) from Brazilian Neotropical savannas

RG Collevatti[1], MPC Telles[1], JC Nabout[2], LJ Chaves[3] and TN Soares[1]

Genetic effects of habitat fragmentation may be undetectable because they are generally a recent event in evolutionary time or because of confounding effects such as historical bottlenecks and historical changes in species' distribution. To assess the effects of demographic history on the genetic diversity and population structure in the Neotropical tree *Dipteryx alata* (Fabaceae), we used coalescence analyses coupled with ecological niche modeling to hindcast its distribution over the last 21 000 years. Twenty-five populations (644 individuals) were sampled and all individuals were genotyped using eight microsatellite loci. All populations presented low allelic richness and genetic diversity. The estimated effective population size was small in all populations and gene flow was negligible among most. We also found a significant signal of demographic reduction in most cases. Genetic differentiation among populations was significantly correlated with geographical distance. Allelic richness showed a spatial cline pattern in relation to the species' paleodistribution 21 kyr BP (thousand years before present), as expected under a range expansion model. Our results show strong evidences that genetic diversity in *D. alata* is the outcome of the historical changes in species distribution during the late Pleistocene. Because of this historically low effective population size and the low genetic diversity, recent fragmentation of the Cerrado biome may increase population differentiation, causing population decline and compromising long-term persistence.
Heredity (2013) **111**, 97–105; doi:10.1038/hdy.2013.23; published online 17 April 2013

INTRODUCTION

It is well established that population extinction risk due to environmental, demographic or genetic stochastic processes is linked to population size (Gilpin and Soulé, 1986; Frankham, 1996; Willi and Hoffmann, 2009). Genetic stochasticity may increase the probability of extinction of small populations due to the enhancement of deleterious effects of genetic drift and inbreeding (Gilpin and Soulé, 1986; Frankham, 1996) and loss of genetic variability may lead to a reduction in individual fitness and population evolutionary potential (Lande, 1988). When population fragmentation reduces effective size many ecological processes may be disrupted, such as pollination and seed dispersal, reducing contemporary gene flow and augmenting the risk of extinction due to multiple Allee effects (see Berec *et al.* (2007) for a review).

The effective population size (N_e) refers to the size of an ideal population experiencing the same rate of random genetic drift over time as the real population under consideration (Wright, 1931). Because real populations deviate from the idealized models, effective sizes are typically much smaller than census sizes (see Palstra and Ruzzante (2008), for a review). The effectiveness of gene flow in counteracting genetic stochasticity and fostering population genetic connectivity is strongly correlated to N_e, especially when populations are subdivided (Palstra and Ruzzante, 2008; Lowe and Allendorf,

2010). Although it has been accepted that very small numbers of migrants per generation ($N_e m = 1.0$) may be sufficient to counteract the deleterious effects of genetic drift and inbreeding (Wright, 1931; Slatkin, 1985), enhancing genetic connectivity among populations, moderate allele frequency divergence is expected even when $N_e m$ is greater than 1.0 due to genetic stochasticity (Lenormand *et al.*, 2009; Lowe and Allendorf, 2010). Therefore, effective population size is an important parameter to evaluate population persistence probability in species conservation and management planning and to understand species evolution (Wang, 2004; Luikart *et al.*, 2010).

The effects of forest fragmentation on genetic diversity in Neotropical tree species in a broad geographical context have been reported recently (for example, Lowe *et al.*, 2005; Soares *et al.*, 2008; Moreira *et al.*, 2009). However, because anthropogenic fragmentation is generally a recent event in evolutionary time and when considering the life cycle of these species, genetic effects of habitat fragmentation may be still undetectable (Collevatti *et al.*, 2001). Genetic signatures resulting from historical bottlenecks may confound patterns of genetic diversity caused by the effects of recent fragmentation (see Ewers and Didham, 2006, for a review). Shifts in species' distribution because of climate changes may also cause a decrease in the genetic variation and an increase in homozygosity (Hewitt 1996). The process of colonization into newly available habitat due to range

[1]Departamento de Biologia Geral, Laboratório de Genética & Biodiversidade, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, Brazil; [2]Unidade de Ciências Exatas e Tecnológica (UnUCET), Universidade Estadual de Goiás, Anápolis, Brazil and [3]Escola de Agronomia e Engenharia de Alimentos, UFG, Goiânia, Brazil
Correspondence: Professor RG Collevatti, Departamento de Biologia Geral, Laboratório de Genética & Biodiversidade, Instituto de Ciências Biológicas, Universidade Federal de Goiás, CP 131, 74001-970, Goiânia, Brasil.
E-mail: rosanegc68@hotmail.com
Received 30 October 2011; revised 8 March 2013; accepted 11 March 2013; published online 17 April 2013

shifts may also lead to a gradient in genetic diversity, but the phenomenon of allele 'surfing' may lead to a high frequency of an allele and induce the structuring of newly colonized areas into distinct sectors of low genetic diversity (Excoffier 2004; Excoffier and Ray, 2008). The estimation of demographic parameters based on coalescence theory (Kingman, 1982) may give clues to the different roles of historical demographic processes on current patterns of genetic diversity. Migration rates and effective population sizes may be jointly estimated for a set of populations weighted by mutation rate (Beerli and Felsenstein, 2001) and reflects an average over a time period,that is, the changes that become established in the subpopulation gene pool. This kind of model is particularly valuable to clarify whether shared polymorphisms result from incomplete lineage sorting or migration when population differentiation is recent or ongoing (Broquet and Petit, 2009). In the context of shared polymorphism due to persistence of ancestral polymorphisms, migration will be low or negligible.

The Cerrado biome in Central Brazil is one of the world's top 25 biodiversity hot spots (Myers et al., 2000). In this biome, the distribution and composition of plant communities are highly complex, with several types of forests and xerophytic communities found in mosaics with Neotropical savannas (see Silva et al., 2006, for a detailed description). This pattern is caused mainly by the high environmental heterogeneity due to geology, geomorphology, diversity in soil types and the climate. However, less than 50% of the original Cerrado area still remains because of agricultural expansion, and this habitat loss has occurred in 50 years (Klink and Machado, 2005). Moreover, the distribution and composition of vegetation in the Central Brazil were deeply influenced by the climatic oscillations of the Tertiary and Quaternary. The last glacial maximum (LGM) coincides with a drier climate in Southern and Central Cerrado domain. During the early Holocene period (until ca 6000– 5000 ^{14}C yr BP), the climate was drier in most of the South American savannas and distribution of savanna-like vegetation in Central and Southeast Brazil was more extensive in the early compared with the late Holocene (Salgado-Labouriau et al., 1998; Behling and Hooghiemstra, 2000; Behling, 2003). In southeastern Brazil, drier climates lasted until 8000–8500 ^{14}C yr BP or later (Behling, 2003) and in the fossil pollen record, arboreal and palm swamp elements are rare and savanna landscapes were dominated by grasslands (Salgado-Labouriau et al., 1998; Behling, 2003). It is possible that many arboreal savanna taxa became restricted to areas with moist climatic conditions (Collevatti et al., 2003; 2012a). After 7000 yr BP, the climate became moister and arboreal pollen dominates the savanna vegetation record. After 4000 yr BP, cerrado and semi-deciduous forests attained their modern distribution and species composition and seasonality continued to increase until reaching the current pattern (Salgado-Labouriau et al., 1998).

Dipteryx alata Vogel (Fabaceae) is a widely distributed tree species in the Cerrado Biome, but is restricted to seasonal savannas that grow in eutrophic and drained soils (cerradão). The species is hermaphroditic and pollination is mainly performed by large and medium size bees. Seeds have a very woody endocarp with an edible nut that are eaten and dispersed by mammals such as bats and monkeys, and are source of raw material for small- and middle-sized food industries, playing an important role in the local economy of Central Brazil (Collevatti et al., 2010). Because of its naturally patchy distribution, species restricted to this kind of vegetation, such as *D. alata*, are suitable to test hypotheses of how a naturally patchy distribution and historical changes in geographical distribution of species affect genetic diversity. We predict that the demographic history of *D. alata* may have led to high population differentiation and low genetic diversity.

Here we studied historical changes in effective population size and gene flow in *D. alata* to understand how demographic history has affected the geographic distribution of genetic diversity. To achieve this goal, we analyzed genetic diversity and structure for the species throughout its distribution, based on eight polymorphic nuclear microsatellite loci. To study historical changes on species' range because of the Pleistocene climatic oscillations, we estimated demographic parameters (effective population size, migration and growth) using coalescent analyses. Further, we used spatially explicit analyses to establish the correspondence between species' demography and environmental climatic changes over the last 21 000 years BP. We first obtained a theoretical distribution of climatically suitable areas for the species, both for current and past (6 and 21 kyr BP) climatic data, using ecological niche modeling (ENM). We then performed a Mantel test and linear regression analyses of genetic diversity and allelic richness on habitat suitability and stability obtained from the ENM, and a regression to the center of the potential paleodistribution to test whether historical changes in species' range generated a cline spatial pattern in genetic diversity and allelic richness, because of range shift.

MATERIALS AND METHODS

Populations and sampling

We sampled 25 populations (644 individuals) throughout the geographical distribution of *D. alata* (Figure 1a). Populations were georeferenced using a Geographical Positional System. Distance between pair of populations ranged from ∼20 to ∼1350 km. All populations, but RAGO and RAMT (see Figure 1a), were in fragments of savanna vegetation surrounded by pasture or crop areas. The populations RAGO and RAMT were in continuous, non-fragmented area of savanna. Because of differences in plant density and fragment size, the number of individuals sampled was different in each locality, ranging from 12 to 32 (Table 1).

Genetic analyses

Genomic DNA extraction followed a standard CTAB procedure (Doyle and Doyle, 1987). Eight microsatellite loci, previously developed for *D. alata* (Soares et al., 2012), were used to genotype the sampled individuals. Genotyping procedures were performed as described elsewhere (Soares et al., 2012). All individuals were genotyped at least twice in independent PCR amplifications and polyacrylamide gels to avoid genotyping error and afterwards, errors due to stutter bands; allele dropout and null alleles were checked using Micro-Checker software (van Oosterhout et al., 2004).

Population genetic diversity and population genetic structure

Population polymorphism was characterized as the number of alleles per locus and allelic richness, based on rarefaction analysis (Mousadik and Petit, 1996). Genetic diversity was estimated based on the expected heterozygosity under Hardy–Weinberg equilibrium (Nei, 1978). The inbreeding coefficient (f) was also estimated to test for deviation from Hardy–Weinberg equilibrium. Analyses were performed with FSTAT 2.9.3.2 software (Goudet, 2002) and randomization-based tests were performed to check for deviations from Hardy–Weinberg expectations and linkage disequilibrium.

Population genetic differentiation was assessed by Wright's F-statistics, F_{IT}, F_{ST} and F_{IS} (Wright, 1951), obtained from an analysis of variance of allele frequencies (Weir and Cockerham, 1984), implemented in the software FSTAT 2.9.3.2 (Goudet, 2002). To verify the contribution of stepwise-like mutations to the genetic differentiation, we estimated Slatkin's R_{ST} (Slatkin, 1995) obtained from an analysis of variance of allele size. Then, we tested the hypothesis that $F_{ST} = R_{ST}$ based on Hardy et al. (2003), using the software SPAGeDI (Hardy and Vekemans, 2002). The comparison of F_{ST} and R_{ST} provides insights into the role of drift and mutation in population differentiation, because R_{ST} is expected to be larger than F_{ST} under stepwise-like mutations, but equal when differentiation is caused solely by drift (see Hardy et al., 2003).

Further, we used a Bayesian clustering simulation implemented in the software STRUCTURE 2.3.3 (Pritchard et al. 2000) to assess the number of
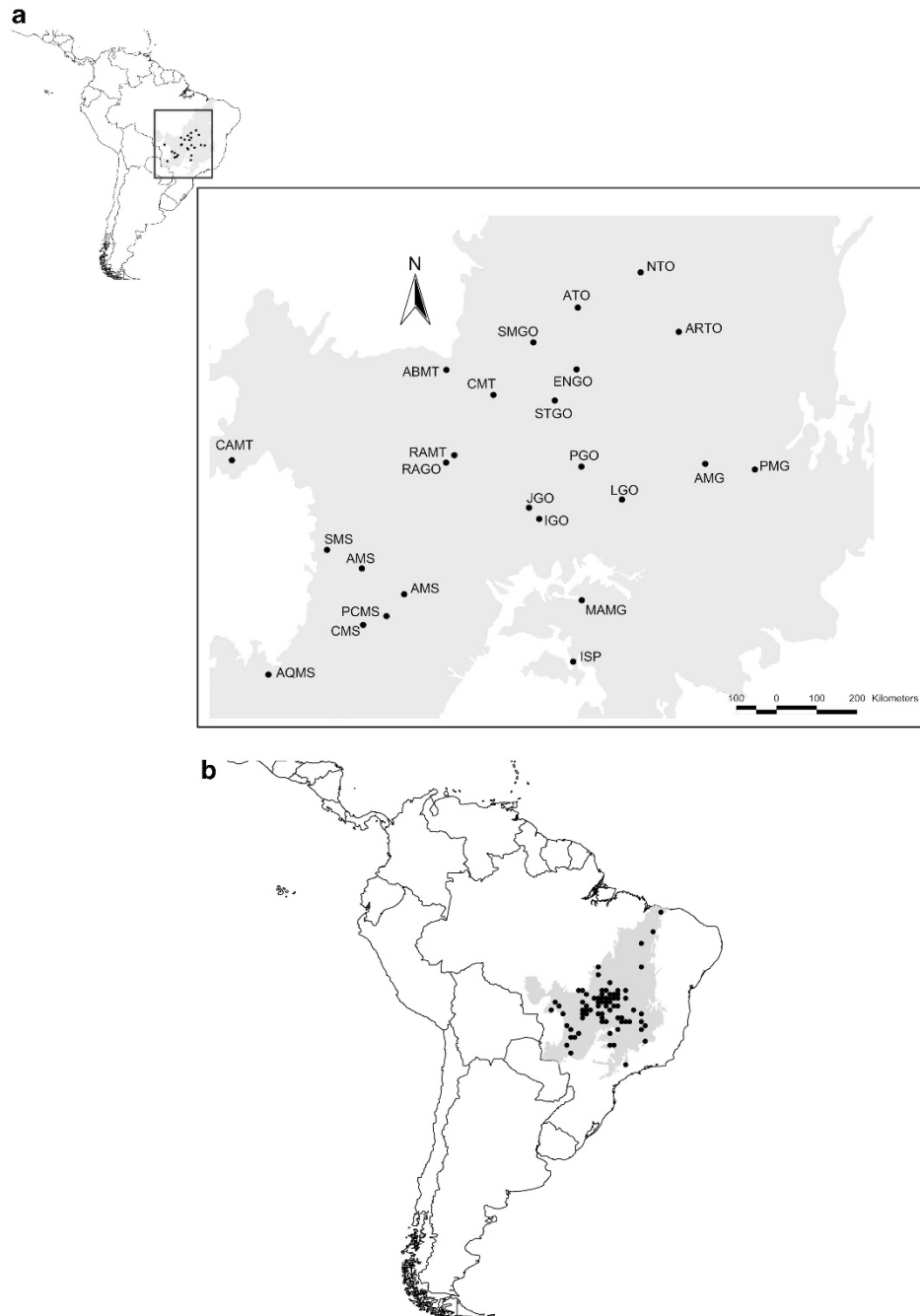
Figure 1 Map of the Cerrado region of Central Brazil in South America (gray area) showing the sample locations of the 25 populations of *D. alata* used for genetic analyses (**a**), and the occurrences for species distribution modeling (**b**). ABMT, Água Boa, MT; AMG, Arinos, MG; AMS, Alcinópolis, MS; AQMS, Aquidauana, MS; ARTO, Arraias, TO; ATO, Alvorada, TO; CAMT, Cáceres, MT; CMS, Camapuã, MS; CMT, Cristalino, MT; ENGO, Estrela do Norte, GO; IGO, Indiara, GO; ISP, Icém, SP; JGO, Jandaia, GO; LGO, Luzilania, GO; MAMG, Monte Alegre, MG; NTO, Natividade, TO; PCMS, Paraíso, Camapuã, MS; PGO, Pirenópolis, GO; PMG, Pintópolis, MG; PMS, Paraíso, MS; RAGO, Aragarças, GO; RAMT, Barra do Garça, MT; SMGO, São Miguel do Araguaia, GO; SMS, Sonora, MS; STGO, Santa Terezinha, GO. For details on populations see Supporting Information.

discrete genetic clusters (K). The method groups individuals into clusters that minimize Hardy–Weinberg and gametic disequilibrium without *a priori* assumptions about population boundaries. Ten independent runs for each K were performed to evaluate consistency of the results using the admixture model of ancestry and correlated allele frequencies. The simulations were performed with a burn-in period of 10 000 repetitions, to minimize the effect of the starting configuration, followed by 100 000 MCMC (Markov chain Monte Carlo) repetitions of data collection. We used the mean data likelihood L(D|K) among runs for the same K to assess the most likely number of clusters supported by the data.

Population demography and gene flow

To study population demographic history, we estimated effective population size (N_e) and number of migrants per generation ($N_e m$) using the coalescent (Kingman, 1982). The demographic parameters $\theta = 4\mu N_e$ (coalescent or

mutation parameter for a diploid genome) and $M = 4N_em/\theta$ (migration rate) were estimated using a MCMC approach (Beerli and Felsenstein, 2001) implemented in Lamarc 2.0.2 (Kuhner, 2006). The analyses were run with 10 initial chains of 10 000 steps and two final chains of 100 000 steps. The chains were sampled every 100 steps following a burn-in of 10 000 steps. We used the default settings for the initial estimate of θ. Also, to improve run performance, we added heating (MCMCMC – Metropolis-Coupled MCMC approach) with one heated chain and temperature of 1.2. The program was run four times to assess convergence. Then, we generated combined results using Tracer v1.4.1 (Rambaut and Drummond, 2007) and considered the results only when ESS $\geqslant 200$ (effective sample size). Most probable estimates were obtained, that is, the highest point on the posterior probability curve for a given parameter, which is the best solution found by a Bayesian run, and also the credibility interval around the estimate of each parameter (Kuhner and Smith, 2007). The effective population size was estimated from the mutation parameter θ. For this, we used the lowest mutation rate reported for microsatellite marker in plants, 2.4×10^{-4} mutation per allele per generation (95% CI = $(1.4 \times 10^{-4}; 4.2 \times 10^{-4})$; Thuillet *et al.*, 2002), often quoted in the range of 10^{-3} to 10^{-4} per locus per generation (Thuillet *et al.*, 2002; Marriage *et al.*, 2009). We chose the lowest value based on the comparison of F_{ST} and R_{ST} (see results below). Time to most recent common ancestor was estimated from overall θ using the above mutation rate and a generation time of 15 years (TN Soares, unpublished data).

We also explored changes in effective population size by estimating the demographic parameter g (exponential growth rate, $\theta_t = \theta_{now} \exp(-gt)$, where t, the time to coalescence is the mutational unit (Kuhner and Smith, 2007), obtained from the coalescent analyses described above, performed in Lamarc 2.0.2 (Kuhner, 2006).

The effect of reduction in population size on population diversity because of bottlenecks was analyzed using the Wilcoxon sign-rank test implemented in Bottleneck version 1.2.02 (Cornuet and Luikart, 1997). Populations that have experienced recent reduction in effective population size may present higher observed heterozygosity (H_o) than heterozygosity expected under mutation-drift equilibrium (H_{eq}) for a given number of alleles in the population (Maruyama and Fuerst, 1985). The program computes the distribution of the H_{eq} based on the simulation of a coalescent process of n genes under a mutation model and the average H_{eq}, which is compared with H_o to test for heterozygosity excess or deficit. The analyses were performed under Infinite Allele Mutation Model, SMM (Stepwise Mutation Model) and TPM (Two-Phase Mutation Model) to verify the sensitivity of the analysis to the mutation model. Because allele frequency distributions for the eight microsatellite loci were unimodal, which may indicate that the loci evolves following a stepwise mutation (Valdes *et al.*, 1993); for TPM, we ran the analyses considering only 30% two phase and 70% SMM. We also run the analyses considering 70% TPM and 30% SMM, but in general, results were not changed.

Ecological niche modeling

We used ENM to obtain the environmental suitability of *D. alata* in the Cerrado biome in the present, and hindcasted for past environments, 6 and 21 kyr BP, corresponding to the mid-Holocene and to the LGM, when a drier climate prevailed in Central Brazil. For this analysis, we obtained 75 points of occurrence of *D. alata* (Figure 1b) based on data from the literature and the Reference Center of Environmental Information (see Nabout *et al.*, 2010, for details).

Three climatic variables were used in ENMs (mean annual rainfall, average temperature of the warmest and coldest months) to project the model into the past climate for 6 and 21 kyr BP, from one Atmospheric-Oceanic Global Circulation Models, the Community Climate System Model version 3.0 from the Paleoclimate Modelling Intercomparison Project Phase II database (http://pmip2.lsce.ipsl.fr/). Current climate was represented by pre-industrial data and was obtained from the same Atmospheric-Oceanic Global Circulation Model (PMIP2). The spatial resolution of these data ranged from $2.2° \times 3.7°$ to $5.4° \times 5.6°$ (longitude $\times$ latitude). From these data, we proceeded to generate climate data at $0.5°$ of resolution, covering the New World (see details of downscaling in http://www.worldclim.org/downscaling; see also Martínez-Meyer *et al.* (2004) and Hijmans and Graham (2006)).

We adopted Araújo and New (2007) approach and niche models and their associated geographic projections were built using a series of combinations of methods, variable combinations and the climatic scenario defined above. Four different methods were used for modeling the species' distribution: BIOCLIM, Mahalanobis Distances, Genetic Algorithm for Rule Set Production and Maximum Entropy.

For each of the climates (that is, current time, 6 and 21 kyr BP), a total of 1400 different models were generated, using distinct combinations of data set partitioning and variable selection (for detailed description of the analyses see Supplementary Information 1). The ENMs were processed in the integrated computing platform BioEnsembles (Diniz-Filho *et al.*, 2009). Following Allouche *et al.* (2006), the evaluation of models was done by the true skill statistics. Models with true skill statistics <0.5 were not used to obtain final ensemble projection (Table 2). The final variable analyzed for each SDM was then the frequency with which *D. alata* is predicted in each cells, called here Estimated Occurrence Frequencies (EOF). This EOF indicates the more climatically suitable areas for the species. Each of the EOF vectors was then averaged and weighted (by true skill statistics) and consensus of the frequencies was produced.

Spatial patterns in genetic diversity

To assess spatial patterns, we first tested whether differentiation is correlated to spatial distance. Pairwise F_{ST} among pairs of populations were estimated and correlated with a geographical distance matrix by a Mantel test (Mantel, 1967), and statistical significance of matrix correlations was established using 10 000 random permutations, also using the software FSTAT 2.9.3.2 (Goudet, 2002).

To analyze the effect of habitat suitability and stability over the last 21 kyr BP in genetic diversity and allelic richness, we calculated the difference on the current and past (6 and 21 kyr BP) suitabilities as surrogates of habitat stability. The genetic diversity and allelic richness were then regressed on suitability and stability variables using the software SAM 4.0 (Rangel *et al.*, 2010). To verify if historical changes in species' range generated a cline spatial pattern in genetic diversity and allelic richness due to range expansion and contraction, we calculated the centroids of the potential current and paleodistributions and performed linear regression analyses also using the software SAM 4.0 (Rangel *et al.*, 2010).

RESULTS

Population genetic diversity and population genetic structure

Pairs of loci did not deviate significantly from linkage equilibrium (all $P > 0.10$) but loci deviated significantly from Hardy–Weinberg expectations for many populations. Analysis of raw data showed no significant evidence of genotyping errors or null alleles (results not shown). Populations of *D. alata* displayed low levels of polymorphism, with low number of alleles, ranging from 2.00 to 4.25, low genetic diversity, ranging from 0.257 to 0.599, and high and significant inbreeding (Table 1). Private alleles were uncommon (six alleles) and were found only in populations AMS (one allele), CAMT (two), CMT (one) and RAMT (two).

Populations are geographically heterogeneous ($F_{ST} = 0.267$, s.d. = 0.036, $P < 0.0001$) with high amounts of inbreeding within populations ($F_{IS} = 0.172$, s.d. = 0.070, $P < 0.0001$) and non-random mating among populations ($F_{IT} = 0.393$, s.d. = 0.066, $P < 0.0001$). R_{ST} also showed a high level of genetic differentiation ($R_{ST} = 0.236$, s.d. = 0.083, $P < 0.0001$), and was not significantly different from F_{ST} ($P = 0.139$), indicating that genetic drift is an important evolutionary force shaping the genetic structure of *D. alata*.

The Bayesian analyses showed a maximal value of L(D|K) with $K = 14$ (Supplementary Information S2). However, the L(D|K) never reached a plateau and the values decreased and became more variable among runs with $K > 14$ and the proportion of individuals in each inferred cluster for $K > 14$ were evenly distributed. Considering $K = 14$, the clusters (Supplementary Information 3) were in agreement with the expected by the patterns of migration obtained from

Table 1 Genetic characterization and demographic parameters for 25 populations of *Dipteryx alata* based on eight microsatellite loci

Pop	N	A	Ar	Ho	He	f	θ	θ–95% interval	N_e	N_e–95% interval	g	g–95% interval
ABMT	32	2.375	2.223	0.291	0.382	0.238**	0.0091	0.0073–0.0100	9.4764	7.6042–10.4167	−115.3400	−425.385 to −15.375
AMG	32	2.250	1.973	0.293	0.310	0.055n.s.	0.0118	0.0100–0.0144	12.2771	10.4167–15.0000	−205.0470	−499.354 to −39.510
AMS	32	4.250	3.609	0.349	0.476	0.267**	0.0100	0.0085–0.0105	10.4406	8.8542–15.6250	−392.7980	−492.838 to −286.160
AQMS	31	3.375	2.817	0.442	0.459	0.038n.s.	0.0124	0.0100–0.0138	12.8979	10.4167–14.3750	−121.8130	−370.743 to −100.082
ARTO	15	2.250	2.205	0.381	0.372	−0.025n.s.	0.0151	0.0107–0.0177	15.7656	11.1458–18.4375	−23.9830	−49.446 to −8.564
ATO	32	3.000	2.747	0.382	0.488	0.217**	0.0104	0.0065–0.0127	10.8146	6.7708–13.2292	−293.1090	−498.616 to −72.646
CAMT	30	3.750	3.152	0.278	0.466	0.405**	0.0093	0.0090–0.0100	9.7172	9.3750–10.4167	−11.3490	−25.267 to −2.656
CMS	13	2.875	2.861	0.466	0.480	0.029n.s.	0.0086	0.0039–0.0123	8.9595	4.0625–12.8125	−131.9720	−367.293 to −10.103
CMT	32	2.625	2.502	0.249	0.322	0.229**	0.0126	0.0083–0.0161	13.1000	8.6458–16.7708	−310.7240	−436.613 to −160.586
ENGO	12	2.500	2.478	0.408	0.474	0.139*	0.0100	0.0084–0.0107	10.4365	8.7500–11.1456	−15.6501	−15.750 to −10.651
IGO	13	2.500	2.447	0.350	0.453	0.226*	0.0089	0.0065–0.0100	9.2599	6.7708–10.4167	−44.9981	−106.804 to −17.811
ISP	32	2.375	2.212	0.322	0.455	0.291**	0.0095	0.0087–0.0100	9.9048	9.0625–10.4167	−271.9271	−485.037 to −144.445
JGO	32	2.625	2.480	0.638	0.599	−0.065n.s.	0.0091	0.0085–0.0101	9.4821	8.8542–10.5208	−253.1099	−496.762 to −104.047
LGO	32	2.500	2.425	0.424	0.509	0.166*	0.0097	0.0067–0.0126	10.0926	6.979–13.1250	−51.0387	−264.664 to −11.594
MAMG	32	2.875	2.582	0.448	0.514	0.128*	0.0063	0.0052–0.0100	6.5318	5.4167–10.4167	−226.9082	−478.219 to −97.0519
NTO	12	2.000	2.093	0.458	0.393	−0.165n.s.	0.0083	0.0069–0.0100	8.5979	7.1875–10.4167	−221.8085	−467.893 to −13.7505
PCMS	13	2.875	2.808	0.169	0.345	0.509**	0.0214	0.0141–0.0260	22.2813	14.6875–27.0833	−243.3480	−452.302 to −62.4119
PGO	32	2.500	2.164	0.259	0.426	0.336**	0.0097	0.0092–0.0120	10.1507	9.5833–12.5000	−379.1954	−483.455 to −210.108
PMG	32	2.375	2.103	0.176	0.257	0.314**	0.0121	0.0106–0.0142	12.6094	11.0417–14.7917	−7.7678	−18.667 to −0.9497
PMS	13	3.125	3.059	0.215	0.410	0.476**	0.0062	0.0046–0.0081	6.4302	4.7917–8.4375	−119.3957	−234.269 to −23.438
RAGO	32	3.250	2.865	0.473	0.462	−0.024n.s.	0.0136	0.0124–0.0157	14.1989	12.9167–16.3542	−359.0560	−495.905 to −85.843
RAMT	32	3.375	2.976	0.308	0.462	0.333**	0.0110	0.0100–0.0124	11.4729	10.4167–12.9167	−30.1285	−31.694 to −28.850
SMGO	32	2.625	2.409	0.354	0.336	−0.052n.s.	0.0106	0.0098–0.0112	11.0146	10.2083–11.6667	−73.6528	−115.900 to −19.734
SMS	32	3.625	3.213	0.422	0.482	0.125*	0.0135	0.0084–0.0220	14.0729	8.7500–22.9167	−189.9883	−357.588 to −101.352
STGO	12	2.875	2.823	0.356	0.387	0.080n.s.	0.0197	0.0114–0.0219	20.5333	11.8750–22.8125	−42.9045	−209.513 to −3.067
Overall	644	2.835	2.609	0.356	0.429	0.171**(0.070)	0.0809	0.0776–0.0844	84.2708	80.8333–87.9167	−5.2505	−24.226 to −1.071

Abbreviations: θ, coalescent parameter; A, total number of alleles; Ar, allelic richness; g, exponential growth parameter; Ho, observed heterozygosity; He, expected heterozygosity; f, inbreeding coefficient; N, number of individuals sampled; N_e, effective population size; ns, not significant, $P > 0.05$; Pop, population.
*, $P < 0.05$, **, $P < 0.001$, standard error for overall f in parenthesis.
95% Credibility interval around the estimate.
Exponential growth parameter (all values are significantly different from zero based on the credibility interval around the estimate.

Table 2 Accuracy, sensitivity, specificity and TSS for the four SDM

SDM	Accuracy	Sensitivity	Specificity	TSS
BIOCLIM	0.89	0.86	0.90	0.88
Mahalanobis	0.92	0.88	0.92	0.90
MAXENT	0.98	0.47	0.98	0.72
GARP	0.77	0.68	0.77	0.73

Abbreviations: GARP, Genetic Algorithm for Rule Set Production; MAXENT, maximum entropy; SDM, species distribution modeling; TSS, true skill statistics.

coalescence analysis (see results below). For instance, populations AQMS and SMS were assigned to the same cluster (cluster 8, see Supplementary Information 2) and populations that were isolated or were only sources of migrants, such as CAMT, MAMG, ATO and SMGO were also assigned alone to a unique cluster (see also Figure 2).

Population demography and gene flow
Coalescent analysis showed that all populations of *D. alata* have low mutation parameter (θ) leading to a low-effective population size (Table 1), also suggesting an important effect of genetic drift. Using the coalescent parameter over all populations (θ = 0.0809), we were able to estimate the time to most recent common ancestor, which dated to 5056 yr BP (95% CI = (8667–2889 yr BP)).

Most populations received negligible gene flow ($N_em < 0.50$). A few populations were estimated to receive more than 1.0 migrant per generations, but gene flow was generally not symmetrical (Figure 2, see also Supplementary Information 4 and 5). These results are not an artifact of sample size, because two populations really had the larger sample sizes (32, RAMT and RAGO, see Table 1), but three of them had the smallest (ENGO and STGO, 12, and PCMS, 13).

Coalescent analyses also showed highly negative values of the growth parameter (g) for most populations (Table 1), suggesting that populations are shrinking. The analysis of population reduction using Bottleneck showed significant excess of H_o for all populations (all $P < 0.02$, see Supplementary Information 6) but not for populations AQMS, PMG, SMS, SMGO. The results were highly consistent among different mutation models.

Ecological niche modeling
The ENM supported the highest levels of suitability of occurrence for *D. alata* throughout the central portion of Brazil (Figure 3a). The region of high environmental suitability coincides with the core area of known occurrences in the Cerrado biome (Figure 3a). The potential distribution area of *D. alata* at 21 kyr BP (the LGM) was smaller than the potential current distribution or at 6 kyr BP (Figures 3b,c). Considering only cells with more than a suitability value of 0.5, the current distribution corresponds to 853 cells, 6 kyr BP to 969 cells and 21 kyr BP, to 239 cells. Moreover, the distribution at 21 kyr BP shifted toward Northwest Brazil, currently occupied by tropical rain forest (Amazonia) and in the boundaries of Cerrado/ Amazonia (Figure 3c). Also, a secondary region could be identified in the Northeast Brazil (Figure 3c), but with lower suitability than the
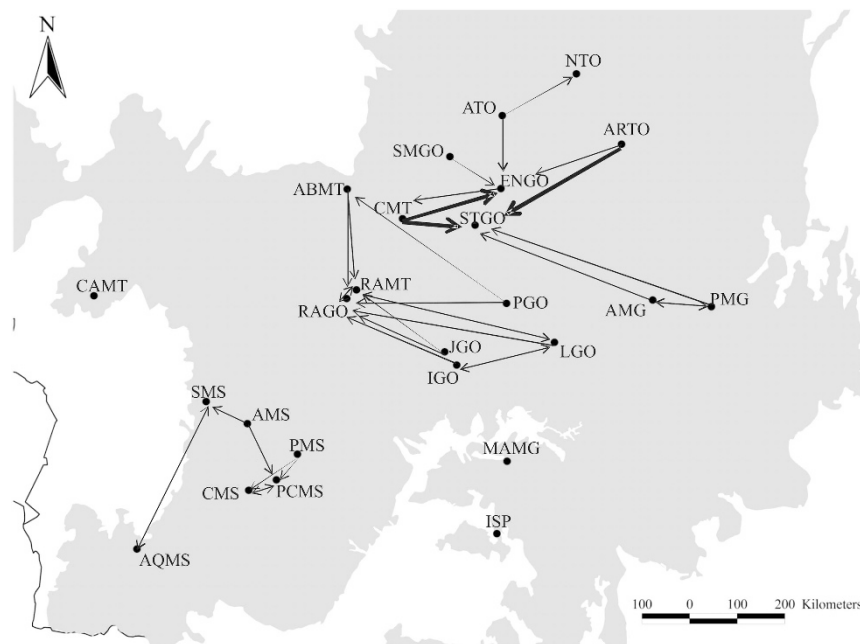
Figure 2 Migration among populations of *D. alata* sampled in the Central Brazil. The arrows connect populations only when population received $N_em > 1.0$. The arrow heads from the source to the receipt population. Thin arrows represent $N_em > 1.0$, thick arrows $N_em > 2$. For all N_em values and credibility intervals see Supplementary Information 4 and 5.

Northwest. The differences of the current and past (6 and 21 kyr BP) suitabilities showed a strong instability for populations in the Central Brazil during the LGM.

Spatial patterns in genetic diversity
Geographical distance explained 20.79% of the differentiation among pairs of populations (Mantel test, $r = 0.456$, $P = 0.001$). The genetic diversity and allelic richness were neither correlated with suitability nor with stability ($P > 0.05$). Genetic diversity was not correlated with geographical distance from the center of paleodistribution at 6 and 21 kyr BP. However, allelic richness showed a clinal spatial pattern, with a decrease in allelic richness with geographical distance from the centroid of the paleodistribution in Northwest Brazil ($R^2 = 0.496$, $b = -0.002$, $t = -4.754$, $P < 0.001$) and an increase from the paleo-distribution in Northeast Brazil ($R^2 = 0.494$, $b = 0.001$, $t = 4.738$, $P < 0.001$). There was also a negative and significant effect of distance from the center of the current distribution of *D. alata* on allelic richness ($R^2 = 0.607$, $b = -0.002$, $t = -5.966$, $P < 0.001$).

DISCUSSION
Demographic parameters estimated by coalescent analyses supports an inference that the low genetic diversity of *D. alata* is most likely due to demographical history. All populations had small effective population size and most populations showed a significant signal of a bottleneck. The coalescent parameter for all populations of *D. alata* was lower than 1.0, suggesting either that effective population size or mutation rate is very low (Kimura and Crow, 1964; Hedrick, 1999). As all loci showed low heterozygosity and stepwise mutation was not important in shaping population structure ($F_{ST} = R_{ST}$), one can infer that both were low. Actually, assuming the lowest value of mutation rate reported so far for plants (see Thuillet *et al.*, 2002; Marriage *et al.*, 2009), N_e was also very low in all populations (Table 1). As the effective population size was small, a substantial effect of demographic

events is expected. We acknowledge a potential limitation of the method used here to estimate N_e and time to most recent common ancestor, because we used a mutation rate estimated for other species. We were conservative and used the lowest value published so far. Higher mutation rates may lead to lower N_e and an overestimation of the splitting dates. However, it is possible that the mutation rate for *D. alata* is lower than the value used here, which may lead to a larger N_e estimate.

Hindcasting of paleodistribution corroborates the results based on coalescence analyses. The strong reduction in suitability during the LGM may have caused population extinction or shrinkage in most parts of the *D. alata* distribution in Central Brazil. Our results show that drift is an important factor shaping the current genetic diversity distribution in *D. alata* and that historical changes in species' range because of Pleistocene climatic oscillations may have affected the patterns of species genetic diversity. Therefore, the observed patterns may be the outcome of a bottleneck because of a range shift during the LGM in the Pleistocene, followed by population expansion after ~6.0 kyr BP.

Notwithstanding, the analyses of habitat stability showed that populations of *D. alata* are localized in unstable areas in relation to the LGM, because only the western region of the potential distribution of *D. alata* remained stable during this period (Figure 3). The present distribution of suitable places was attained after the LGM, when a moister and warmer climate was established. Studies based on the pollen record indicate a drier period during the last glacial for the Central and Southeast Brazil, until 6000–5000 yr BP (Behling, 2003). As a consequence, the expansion of species adapted to drier climates was favored and the distribution of dry savanna like vegetation, dominated by grass and herbs, was more extensive in the early compared with the late Holocene (Behling and Hooghiemstra, 2000). The low genetic diversity and effective population size may be the outcome of a lack of long-term stable areas in the present
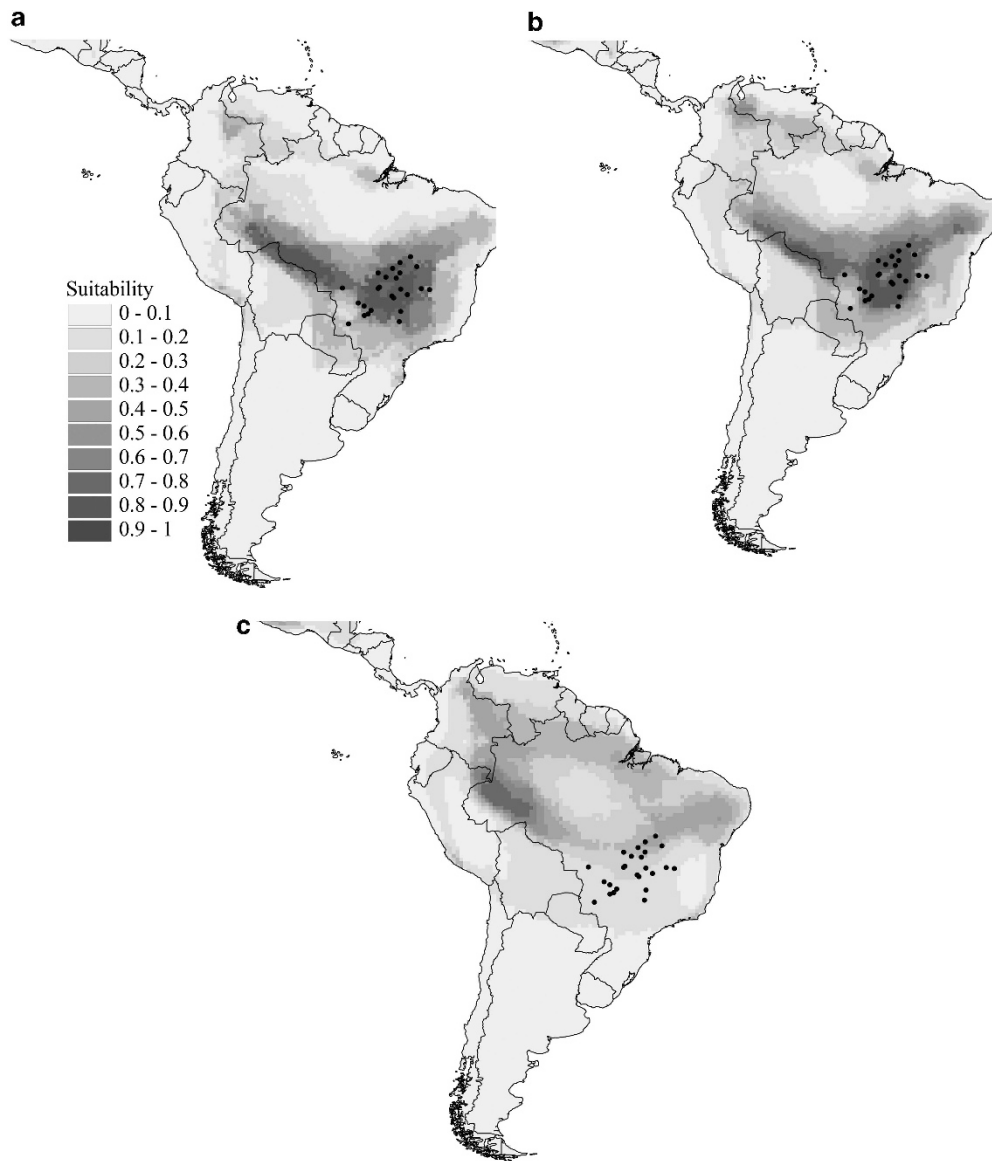
Figure 3 Geographic distribution of environmental suitabilities modeled for *D. alata* in present time (**a**), projected for 6 kyr BP (**b**) and 21 kyr BP (**c**). Black dots are populations sampled for genetic analyses, following Figure 1.

distribution of *D. alata* in Central Brazil. Hence, Central Brazil was most likely colonized by populations from Southwest Amazonia and Northeast Brazil after the climate became moister and warmer, at ~6000–5000 yr BP, as a result of migration of populations as though tracking climate changing in a trail-leading population expansion (Hewitt, 1996). Repeated cycles of extinction and range shift due to glaciation may have caused loss of genetic diversity and high differentiation among populations.

Indeed, spatially explicit analyses showed a spatial cline pattern in allelic richness, related to the paleodistribution at 21 kyr BP, which corroborates the pattern expected under a range shift (Excoffier *et al.*, 2009) and indicates that historical factors may confound patterns of genetic diversity caused by the effects of recent fragmentation. The significant relationship with the center of present distribution reinforces the cline spatial pattern obtained for the paleodistribution, because the center of the current *D. alata* distribution is in West Brazil, between the CAMT and RAGO populations. This may also

explain the pattern of migration, mainly from populations at the edge to populations at the center of the *D. alata* distribution (see Figure 2) and the lack of correlation between stability and genetic diversity and allelic richness. This pattern of expansion may lead to complex patterns of differentiation and also to a loss of genetic diversity (Excoffier *et al.*, 2009).

Populations of *D. alata* also showed high genetic differentiation, contrasting with other Cerrado tree species. *Caryocar brasiliense*, a widely distributed savanna tree species, shows high genetic diversity, no sign of a bottleneck and low differentiation among population ($F_{ST} = 0.07$; Collevatti *et al.*, 2001) and also *Euterpe edulis* ($F_{ST} = 0.06$; Gaiotto *et al.*, 2003). It is possible that climate changes have not caused a strong range shift in those species because they are more continuously distributed than *D. alata*, which is restricted to more fertile soils. In fact, there are evidences of multiple refugia for *C. brasiliense* in the Central Brazil, leading to a high genetic diversity (Collevatti *et al.*, 2001, 2003, 2012a). Nevertheless, genetic diversity

and effective population size are low for other species with restricted distribution in the Central Brazil, for example, *Tibouchina papyrus* (Collevatti *et al.*, 2012b) and *Lychnophora ericoides* (Collevatti *et al.* 2009), both restricted to quartzite outcrops in the Cerrado biome, most likely due to range contraction during the interglacial periods in the Quaternary.

Our results also showed a low gene flow among populations. Despite the long-distance seed dispersal expected for *D. alata*, performed by bats and large mammals such as tapirs (*Tapirus terrestris*, Tapiridae) in Central Brazil, we could not detect high levels of gene flow. It is important to note that the current landscape in the studied region is comprised of small fragments or isolated individuals in a matrix of pastures and crops, such as soybean and maize, and the few estimates of gene flow above 1.0 migrants per generation detected here are most likely due to gene flow before anthropogenic fragmentation and disturbance ($\sim$60 years ago). The only populations connected by continuous savanna like vegetation were RAGO and RAMT, which are along Araguaia River and showed symmetrical gene flow (see Figure 2). It follows that population demographic history may have caused the small population effective size, leading to fixation and loss of alleles due to drift, and consequently, to low genetic diversity within populations and high differentiation among them. Current fragmentation may augment isolation and cease connectivity among populations, despite the very short period of time of habitat loss. Although Mantel tests showed a significant spatial pattern, suggesting migration-drift equilibrium, the model explanatory power was low ($R^2 = 0.2079$) implying that other factors than spatial distance also affected differentiation and gene flow among populations, such as habitat suitability and landscape level processes. The lack of, or low, gene flow between geographically closer populations reinforces this conclusion (see Figure 2). The Lamarck and Bottleneck results on population reduction also suggest that populations are not in equilibrium. Under migration-drift equilibrium with spatially structured gene flow, a much stronger spatial pattern with higher gene flow between closer populations should have been detected. Results based on different analytical approaches support our findings on low gene flow. Fine-scale genetic structure showed that *D. alata* has smaller neighborhood size and stronger spatial structure than other Cerrado tree species (Collevatti *et al.*, 2010). Besides, Soares *et al.* (2008) reported, using RAPD markers, a high genetic differentiation ($\phi_{ST} = 0.15$) and discontinuity among 10 populations of *D. alata*, from the 25 analyzed in the present study.

In addition, besides the relatively recent population coalescence ($\sim$5.0 kyr BP, 95% CI = (8667–2889 yr BP)), differentiation among populations was very high, considering both, F_{ST} and Bayesian analyses performed in Structure software, suggesting that drift has driven population evolution but that more recent landscape process may have affected population connectivity leading to population differentiation. Despite the relatively short time since anthropogenic fragmentation ($\sim$50 years), compared with the generation time of the species (15 years), the rate of fragmentation is one of the highest reported so far for Brazilian ecosystems (Klink and Machado, 2005). Thus, despite population expansion after the LGM because of the increasing in suitable areas, populations may be shrinking because of an Allee effect caused by recent anthropogenic fragmentation.

These findings have important consequences in the conservation of *D. alata* in the Brazilian Cerrado. Under the present scenario of low gene flow, low genetic diversity and high population differentiation, decisions made on the conservation of *D. alata* should include allelic diversity and regional patterns of landscape fragmentation to establish corridors and improve connectivity among populations and

potentially increasing effective population size (Diniz-Filho *et al.*, 2012). Because allelic richness tends to decrease with distance from West Brazil, where one can find populations with the highest levels of allelic richness (for example, AMS, AQMS, CAMT, RAMT, SMS, Table 1), this region should be a priority in conservation planning, together with the RAGO and RAMT populations, where the vegetation is continuous. These populations have also been included in systematic conservation planning, to establish optimum strategies for *in situ* conservation of *D. alata*, as the smallest set of local population that should be conserved to represent the known genetic diversity (Diniz-Filho *et al.*, 2012).

In conclusion, our results strongly suggest that the pattern of low genetic diversity and allelic richness in *D. alata* may be the outcome of habitat instability during the late Pleistocene, and might reflect range shifts and a restriction in distribution during the last glacial period, with subsequent expansion, leading to its present-day distribution, high population differentiation and clinal distribution in allelic richness. Current fragmentation may be serving to enhance historical process, decrease population size and augment isolation.

DATA ARCHIVING
Data deposited in the Dryad repository: doi:10.5061/dryad.1cd80.

Allouche O, Tsoar A, Kadmon R (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J Appl Ecol* **43**: 1223–1232.

Araújo MB, New M (2007). Ensemble forecasting of species distributions. *Trends Ecol Evol* **22**: 42–47.

Beerli P, Felsenstein J (2001). Maximum likelihood estimation of a migration matrix and effective population sizes in subpopulations by using a coalescent approach. *Proc Natl Acad Sci USA* **98**: 4563–4568.

Behling H (2003). Late glacial and Holocene vegetation, climate and fire history inferred from Lagoa Nova in the southeastern Brazilian lowland. *Veg Hist Archaeobot* **12**: 263–270.

Behling H, Hooghiemstra H (2000). Neotropical savanna environments in space and time: Late Quaternary. In: Markgraf V (ed.). *Interhemispheric climate linkages*. Academic Press: San Diego, pp 307–323.

Berec L, Angulo E, Courchamp F (2007). Multiple Allee effects and population management. *Trends Ecol Evol* **22**: 185–191.

Broquet T, Petit EJ (2009). Molecular estimation of dispersal for ecology and population genetics. *Annu Rev Ecol Evol Syst* **40**: 193–216.

Collevatti RG, Rabelo SG, Vieira RF (2009). Phylogeography and disjunct distribution in *Lychnophora ericoides* (Asteraceae), an endangered cerrado shrub species. *Ann Bot* **104**: 655–664.

Collevatti RG, Castro TG, Lima JS, Telles MPC (2012b). Phylogeography of *Tibouchina papyrus* (Pohl) Toledo (Melastomataceae), an endangered tree species from rocky savannas, suggests bidirectional expansion due to climate cooling in the Pleistocene. *Ecol Evol* **2**: 1024–1035.

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 lineages of *Caryocar brasiliense* populations in the Brazilian Cerrado based on the analysis of chloroplast DNA sequences and microsatellite haplotype variation. *Mol Ecol* **12**: 105–115.

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, Lima-Ribeiro MS, Souza-Neto AC, Franco AA, Terrible LV (2012a). Recovering the demographical history of a Brazilian Cerrado tree species *Caryocar brasiliense*: coupling ecological niche modeling and coalescent analyses. *Nat Conserv* **10**: 169–176.

Cornuet JM, Luikart G (1997). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**: 2001–2014.

Diniz-Filho JAF, Bini LM, Rangel TF, Loyola RD, Hof C, Nogués-Bravo D et al. (2009). Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography* **32**: 897–906.

Diniz-Filho JAF, Melo DB, Oliveira G, Collevatti RG, Soares TN, Nabout JC et al. (2012). Planning for optimal conservation of geographical genetic variability within species. *Conserv Genet* **13**: 1085–1093.

Doyle JJ, Doyle JL (1987). Isolation of plant DNA from fresh tissue. *Focus* **12**: 13–15.

Ewers RM, Didham RK (2006). Confounding factors in the detection of species responses to habitat fragmentation. *Biol Rev* **81**: 117–142.

Excoffier L (2004). Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Mol. Ecol* **13**: 853–864.

Excoffier L, Foll M, Petit RJ (2009). Genetic consequences of range expansions. *Annu Rev Ecol Evol Syst* **40**: 481–501.

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

Frankham R (1996). Relationship of genetic variation to population size in wildlife. *Conserv Biol* **10**: 1500–1508.

Gaiotto FA, Grattapaglia D, Vencovsky R (2003). Genetic structure, mating system, and long-distance gene flow in heart of palm (*Euterpe edulis* Mart.). *J Hered* **94**: 399–406.

Gaston KJ (2003). *The Structure and Dynamics of Geographic Ranges*. Oxford University Press: Oxford.

Gilpin M, Soulé ME (1986). Minimum viable populations: process of species extinction. In: Soulé ME (ed.). *Conservation Biological, the Science of Scarcity and Diversity*. Sinauer Associates Inc.: Sunderland, pp 19–34.

Goudet J (2002). FSTAT, a program to estimate and test gene diversities and fixation indices. Version 2.9.3.2. Available from: http://www2.unil.ch/popgen/softwares/fstat.htm

Hardy OJ, Charbonnel N, Fréville H, Heuertz M (2003). Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics* **163**: 1467–1482.

Hardy OJ, Vekemans X (2002). SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* **2**: 618–620.

Hedrick PW (1999). Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**: 313–318.

Hewitt GM (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol J Linn Soc Lond* **58**: 247–276.

Hijmans RJ, Graham CH (2006). The ability of climate envelope models to predict the effect of climate change on species distributions. *Glob Change Biol* **12**: 2272–2281.

Kimura M, Crow JF (1964). The number of alleles that can be maintained in a finite population. *Genetics* **49**: 725–738.

Kingman JFC (1982). The coalescent. *Stoch Proc Appl* **13**: 235–248.

Klink CA, Machado RB (2005). Conservation of the Brazilian Cerrado. *Conserv Biol* **19**: 707–713.

Kuhner MK (2006). LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* **22**: 768–770.

Kuhner MK, Smith LP (2007). Comparing likelihood and Bayesian coalescent estimation of population parameters. *Genetics* **175**: 155–165.

Lande R (1988). Genetics and demography in biological conservation. *Science* **241**: 1455–1460.

Lenormand T, Roze D, Rousset F (2009). Stochasticity in evolution. *Trends Ecol Evol* **24**: 157–165.

Lowe AJ, Boshier D, Ward M, Bacles CFE, Navarro C (2005). Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees. *Heredity* **95**: 255–273.

Lowe WH, Allendorf FW (2010). What can genetics tell us about population connectivity? *Mol Ecol* **19**: 3038–3051.

Luikart G, Ryman N, Tallmon DA, Schwartz MK, Allendorf FW (2010). Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conserv Genet* **11**: 355–373.

Mantel NA (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res* **27**: 209–220.

Marriage TN, Hudman S, Mort ME, Orive ME, Shaw RG, Kelly JK (2009). Direct estimation of the mutation rate at dinucleotide microsatellite loci in *Arabidopsis thaliana* (Brassicaceae). *Heredity* **103**: 310–317.

Martínez-Meyer E, Peterson AT, Hargrove WW (2004). Ecological niches as stable distributional constraints on mammal species, with implications for Pleistocene extinctions and climate change projections for biodiversity. *Global Ecol Biogeogr* **13**: 305–314.

Maruyama T, Fuerst PA (1985). Population bottlenecks and non equilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. *Genetics* **111**: 675–689.

Moreira PA, Fernandes GW, Collevatti RG (2009). Fragmentation and spatial genetic structure in *Tabebuia ochracea* (Bignoniaceae) a seasonally dry Neotropical tree. *For Ecol Manage* **258**: 2690–2695.

Mousadik A, Petit RJ (1996). High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theor Appl Genet* **92**: 832–839.

Myers N, Mittermeier RA, Mittermeier CG, Fonseca GAB, Kent J (2000). Biodiversity hotspots for conservation priorities. *Nature* **403**: 853–858.

Nabout JC, Soares TN, Diniz-Filho JAF, Marco P, Telles MPC, Naves RV et al. (2010). Combining multiple models to predict the geographical distribution of the Baru tree (*Dipteryx alata* Vogel) in the Brazilian Cerrado. *Braz J Biol* **70**: 911–919.

Nei M (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583–590.

Oosterhout CV, Hutchinson WF, Wills DPM, Shipley P (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* **4**: 535–538.

Palstra FP, Ruzzante DE (2008). Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Mol Ecol* **17**: 3428–3447.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Rambaut A, Drummond AJ (2007). *Tracer version 1.4*. Free available from: http://beast.bio.ed.ac.uk/Tracer

Rangel TF, Diniz-Filho JAF, Bini LM (2010). SAM: a comprehensive application for spatial analysis in macroecology. *Ecography* **33**: 46–50.

Salgado-Labouriau ML, Barberi M, Ferraz-Vicentini KR, Parizzi MG (1998). A dry climatic event during the late Quaternary of tropical Brazil. *Rev Paleob Palyn* **99**: 115–129.

Silva JF, Farinas MR, Felfili JM, Klink CA (2006). Spatial heterogeneity, land use and conservation in the Cerrado region of Brazil. *J Biogeogr* **33**: 536–548.

Slatkin M (1985). Gene flow in natural populations. *Annu Rev Ecol Evol Syst* **16**: 393–430.

Slatkin M (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**: 457–462.

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 et al. (2012). Development of microsatellite markers for the Neotropical tree species *Dipteryx alata* (Fabaceae). *Am J Bot* **99**: e72–e73.

Thuillet AC, Bru D, David J, Roumet P, Santoni S, Sourdille P et al. (2002). Direct estimation of mutation rate for 10 microsatellite loci in durum wheat, *Triticum turgidum* (L.) Thell. ssp durum desf. *Mol Biol Evol* **19**: 122–125.

Valdes AM, Slatkin M, Freimer NB (1993). Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics* **133**: 737–749.

Wang J (2004). Application of the one-migrant-per-generation rule to conservation and management. *Conserv Biol* **18**: 332–343.

Weir BS, Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.

Willi Y, Hoffmann AA (2009). Demographic factors and genetic variation influence population persistence under environmental change. *J Evol Biol* **22**: 124–133.

Wright S (1931). Evolution in Mendelian populations. *Genetics* **16**: 97–159.

Wright S (1951). The genetic structure of populations. *Ann Eug* **15**: 323–354.