Annals of Botany **119**: 239–252, 2017
doi:10.1093/aob/mcw096, available online at www.aob.oxfordjournals.org

ANNALS OF BOTANY
Founded 1887

PART OF A SPECIAL ISSUE ON ENDEMISM HOTSPOTS

A large historical refugium explains spatial patterns of genetic diversity in a Neotropical savanna tree species

Helena Augusta Viana e Souza[1,†], Rosane Garcia Collevatti[2,†], Matheus S. Lima-Ribeiro[3], José Pires de Lemos-Filho[4] and Maria Bernadete Lovato[1,*]

[1]Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Caixa Postal 486, 31270-901 Belo Horizonte, MG, Brazil, [2]Laboratório de Genética & Biodiversidade, ICB, Universidade Federal de Goiás, Caixa Postal 131, 74001-970 Goiânia, GO, Brazil, [3]Laboratório de Macroecologia, Universidade Federal de Goiás, Campus Jataí, 75801-615 Jataí, GO, Brazil and [4]Departamento de Botânica, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Caixa Postal 486, 31270-901 Belo Horizonte, MG, Brazil
**For correspondence. E-mail lovatomb@icb.ufmg.br*
[†]These authors contributed equally to this work.

Received: 5 November 2015 Returned for revision: 2 December 2015 Accepted: 29 March 2016 Published electronically: 16 June 2016

• **Background and Aims** The relative role of Pleistocene climate changes in driving the geographic distribution and genetic diversity of South American species is not well known, especially from open biomes such as the Cerrado, the most diverse tropical savanna, encompassing high levels of endemism. Here the effects of Quaternary climatic changes on demographic history, distribution dynamics and genetic diversity of *Dimorphandra mollis*, an endemic tree species widely distributed in the Cerrado, were investigated.
• **Methods** A total of 38 populations covering most of the distribution of *D. mollis* were analysed using internal transcribed spacer (ITS) sequences and nuclear microsatellite variation [simple sequence repeats (SSRs)]. The framework incorporated statistical phylogeography, coalescent analyses and ecological niche modelling (ENM).
• **Key Results** Different signatures of Quaternary climatic changes were found for ITS sequences and SSRs corresponding to different time slices. Coalescent analyses revealed large and constant effective population sizes, with high historical connectivity among the populations for ITS sequences and low effective population sizes and gene flow with recent population retraction for SSRs. ENMs indicated a slight geographical range retraction during the Last Glacial Maximum. A large historical refugium across central Brazil was predicted. Spatially explicit analyses showed a spatial cline pattern in genetic diversity related to the paleodistribution of *D. mollis* and to the centre of its historical refugium.
• **Conclusions** The complex genetic patterns found in *D. mollis* are the result of a slight geographical range retraction during the Last Glacial Maximum followed by population expansion to the east and south from a large refugium in the central part of the Cerrado. This historical refugium is coincident with an area predicted to be climatically stable under future climate scenarios. The identified refugium should be given high priority in conservation polices to safeguard the evolutionary potential of the species under predicted future climatic changes.

Key words: Cerrado, climatic changes, coalescent analysis, demographic history, *Dimorphandra mollis*, ecological niche modelling, genetic diversity, Neotropical savanna, phylogeography, refugium.

INTRODUCTION

The relative role of Pleistocene climate changes in driving the geographical distribution and genetic diversity of South American species, especially those from open biomes, is not yet well known (Turchetto-Zolet *et al*., 2013). Recovering historical population dynamics and investigating how genetic variation was spatially structured are important for understanding species responses to Quaternary climate changes (Knowles, 2009). Ecological niche modelling (ENM; *sensu* Peterson *et al*., 2011) has allowed the exploration of the geographical context of species dynamics through time by hindcasting suitable climatic conditions currently occupied by a species. Together with coalescent models, ENMs provide the framework to reconstruct species' demographic history (e.g. Collevatti *et al*., 2012*a*, 2013*a*) and to better understand the effects of past colonization and the role of Quaternary climate changes in shaping the current spatial pattern of genetic diversity, especially for species with incomplete or no fossil records (e.g. Collevatti *et al*., 2013*a*, 2015*a*). The identification of areas with low climate change velocity through time (Quaternary refugia) is important to establish adequate conservation measures to ensure the maintenance of biodiversity in the face of ongoing climate change.

The Cerrado is the most diverse tropical savanna in the world and is considered a global biodiversity 'hotspot', with high levels of endemism and threat (Myers *et al*., 2000). Occupying more than 20° of latitude in central Brazil and an altitudinal range from sea level to 1800 m, the Cerrado region extends from the margin of the Amazon Forest in northern Brazil, to the Araucaria forest in southern Brazil and Chaco in Bolivia and Paraguay (Ratter *et al*., 1997). The climate is characterized by a strong dry season during the winter with annual temperatures

ranging between 18 and 28 °C and precipitation of 800–2000 mm (Ratter *et al*., 2006). The Cerrado includes a wide variety of physiognomic types ranging from treeless grasslands to dense woodlands. The most frequent physiognomy is the cerrado *sensu stricto*, associated with oligotrophic soils and subject to fire action (Furley and Ratter, 1988). It is a savanna characterized by a community of trees and shrubs, with contorted trunks, thick corky bark, sclerophyllous leaves and crown cover of 10–60 %, below which there is a well-developed grassy ground layer (Ratter *et al*., 1997). The total number of vascular plant species in the Cerrado biome is estimated to be about 10 000 (Ratter *et al*., 1997; Myers *et al*., 2000). Ratter *et al*. (2003) estimated the presence of 951 trees and large shrubs in the cerrado *sensu latu* (savannic element) and, of the 300 most common of these woody species, 35 % are endemic.

The South American climate was spatially and temporally complex in both open and forest biomes during the Quaternary (Ab'Saber, 2000). Palynological data suggest that south-eastern Brazil was cooler and drier during the glacial periods than it is at present, and savannas expanded both north and south of the equator due to markedly drier conditions (Behling, 2002). The sub-tropical grasses in southern Brazil spread northward, replacing the Atlantic Forest and savannas to latitudes up to 20°S and retreating to the south only after the glaciation (Behling, 2002). Phylogeographical studies with Cerrado species have shown that Pleistocene climatic changes have shaped the demographic history of populations and the geographical range of species of plants (e.g. Collevatti *et al*., 2003, 2013*b*; Ramos *et al*., 2007; Novaes *et al*., 2010, 2013), as well as animals (e.g. Prado *et al*., 2012; Santos *et al*., 2014; González *et al*., 2015). The few studies on phylogeography and demographic history performed with Cerrado plant species have shown some congruencies. High levels of genetic differentiation among populations, evidence of reduction in species range during the Last Glacial Maximum (LGM), and recent colonization of the southern Cerrado from northern sources were found (Collevatti *et al*., 2003, 2013*b*; Ramos *et al*., 2007; Novaes *et al*., 2010, 2013). Most species have also shown a striking phylogeographical structure, with a somewhat similar subdivision between groups of populations (eg. Ramos *et al*., 2007; Novaes *et al*., 2010, 2013).

Here we studied the demographical history of *Dimorphandra mollis* Benth. (Fabaceae), a widely distributed tree species, endemic to savannas of the Cerrado biome that occurs across central, south-east and south-west Brazil, west Bolivia and Paraguay. The *Dimorphandra* genus is made up of 40 species, all endemics of tropical South America, and most occurring in Amazonia (Silva, 1986). Most *Dimorphandra* species have restricted distributions, with only a few widely distributed. *Dimorphandra mollis*, popularly known as faveiro and fava d'anta, is used as raw material by the pharmaceutical industry for flavonoid rutin extraction. It is pollinated by small insects, its seeds are dispersed by mammals and its mating system is not known (Gonçalves *et al*., 2010).

We address the effect of the Quaternary climate changes on the genetic diversity and distribution of *D. mollis*, coupling coalescent analyses and ecological niche modelling (ENM). We used ENM to hindcast the species distribution at the LGM (21 ka) and mid-Holocene (6 ka), to infer palaeodistribution dynamics and to identify putative climatically stable areas. Genetic diversity was assessed using internal transcribed spacer (ITS) sequencing and nuclear microsatellites in a large number of populations covering most of the geographical distribution of the species. Considering that areas of low-velocity climatic change through time potentially shelter large and more stable populations, we expected to find higher genetic diversity in these areas. In contrast, more unstable areas would exhibit lower genetic diversity due to a population bottleneck. Furthermore, because the geographical distribution of *D. mollis* is included in the geographic range of other Cerrado legume tree species with phylogeographical patterns which are already known (Ramos *et al*., 2007; Novaes *et al*., 2010, 2013; Colevatti *et al*., 2013*b*), we expected a similar phylogeographical structure for *D. mollis*. Since savanna species have shown evidence of range retraction during the LGM followed by expansion after climate amelioration (e.g. Colevatti *et al*., 2012*a*, *b*, 2013*b*, 2015*b*), we expected that *D. mollis* would have a similar pattern of historical distribution shifts. Such shifts would be reflected by high genetic differentiation among populations with low historical gene flow and lower genetic diversity in areas recently colonized. Based on data obtained here for *D. mollis*, and on the projected future distribution of 18 tree species widely distributed in the Cerrado (Terribile *et al*., 2012), we discussed the consequences of ongoing climate change on the persistence and maintenance of evolutionary potential in *D. mollis*, and preservation of the Cerrado biodiversity.

MATERIALS AND METHODS

Population sampling and DNA isolation

We sampled 38 populations (Figs 1 and 2; Supplementary Data Table S1) throughout the geographical distribution of *D. mollis* (Fig. S1). For ITS analyses, 32 populations (155 individuals) were studied, and 19 populations were studied for microsatellites (371 individuals). The distance between population pairs ranged from 69 km to approx. 1400 km. For each population, a voucher was collected and deposited at the Herbarium of the Botany Department of Universidade Federal de Minas Gerais (BHCB). Leaves were collected and DNA was extracted following Souza *et al*. (2012*a*).

ITS sequencing

We sequenced the region ITS1 + 5.8S + ITS2 (hereafter, 'ITS') from nuclear ribosomal DNA (nrDNA) using the primer pair cy1–cy3 (Wright *et al*., 2006). PCR was performed in a 25 µL volume containing 10 ng of template DNA, 1·0 U of *Taq* DNA polymerase (Phoneutria, BR), 1× reaction buffer (10 mM Tris–HCl, pH 8·3, 50 mM KCl, 2·0 mM $MgCl_2$), 2 % dimethylsulphoxide (DMSO), 1 M betaine, 0·2 ng of bovine serum albumin (BSA), 200 µM of each dNTP and 0·5 µM of each primer. The amplifications were carried out in an Eppendorf Thermocycler (Eppendorf, Germany) with an initial step at 94 °C for 3 min, followed by 34 cycles at 94 °C for 1 min, 54 °C for 1 min and 72 °C for 1 min, and a final extension at 72 °C for 10 min. Automated sequencing was carried out in a MegaBACE 1000 automated sequencer (GE Healthcare,

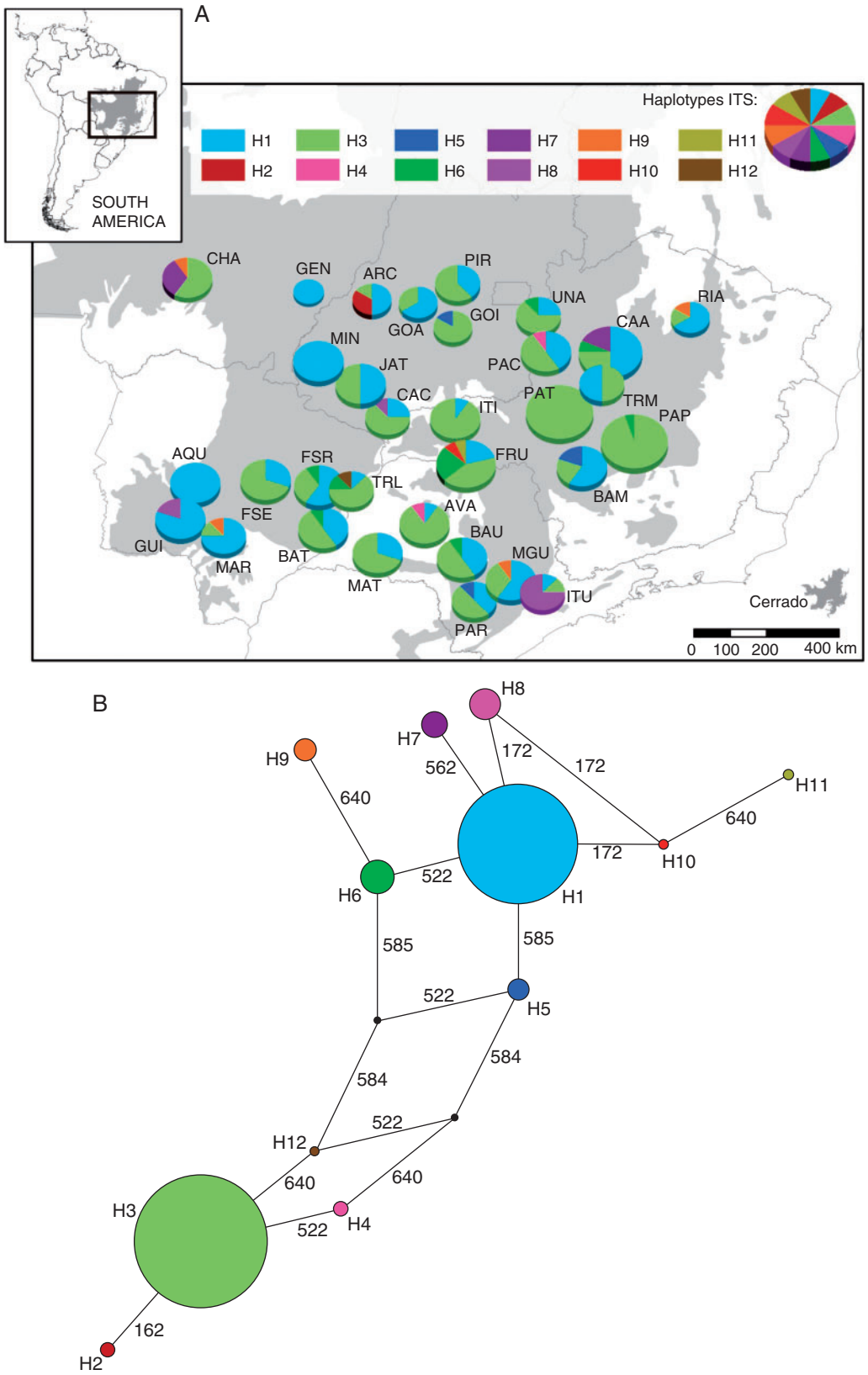

FIG. 1. Haplotype phylogenetic relationships and geographical distribution for ITS nrDNA based on the sequencing of 155 individuals of *Dimorphandra mollis* from 32 populations. (A) Geographical distribution of haplotypes. Different colours were assigned for each haplotype according to the key. The size of the circle represents the sample size in each population and the circle sections represent the haplotype frequency in each sampled population. (B) Phylogenetic relationships among haplotypes using a median-joining network. The size of the circumference is proportional to the haplotype frequency. Mutation sites are shown along lines in the network; small black circles are the median vectors. Different colours were assigned for each population according to the key. For details on population codes and localities Table S1.

Sweden) using the DYEnamic™ ET terminator sequencing kit (GE Healthcare), following the manufacturer's instructions.

Nuclear microsatellites

Seven loci previously isolated for *D. mollis* (Souza *et al.*, 2012*b*) were used to genotype 371 individuals. Amplifications were performed in a 13 µL reaction volume containing 10–20 ng of template DNA, 1× reaction buffer (50 mM KCl, 10 mM Tris–HCl, pH 8·4, 0·1 % Triton X-100; Phoneutria, Brazil), 0·86 mM $MgCl_2$, 0·25 mM of each dNTP, 0·05 µM forward primer, 0·10 µM reverse primer, 0·10 µM M13 fluorescent-labelled primer and 1 U of *Taq* polymerase (Phoneutria). PCRs were performed on a Eppendorf Thermocycler under the following conditions: initial denaturation at 94 °C for 5 min, followed by 34 cycles of 94 °C for 1 min, 57 or 63 °C for 1 min (primer annealing temperature), 72 °C for 1 min, eight cycles at 94 °C for 30 s, 53 °C for 45 s, 72 °C for 45 s, and a final extension at 72 °C for 10 min. Loci were genotyped on a MegaBACE 1000 automated sequencer, using 0·1 % Tween-20 and ROX-500 Size Standard (GE Healthcare). Alleles were sized using the MegaBACE Fragment Profiler version 1.2 software (GE Healthcare).

Errors due to null alleles, stuttering or allele drop-out were tested using Micro-Checker software (van Oosterhout *et al.*, 2004). Allele frequencies were corrected for loci with evidence of null alleles using the Brookfield method (Brookfield, 1996) implemented in Micro-Checker.

Nuclear ITS data analysis

To generate consensus sequences, the nrDNA sequences were assembled using the PHRED/PHRAP/CONSED package. The nrDNA sequences were deposited in GenBank under accession numbers KT970387–KT970398. The consensus sequences were aligned using CLUSTAL-W (Thompson *et al.*, 1994) implemented in the MEGA ver. 5 program (Tamura *et al.*, 2011), followed by manual adjustment. All variable sites were carefully checked in the original electropherogram, and double peaks (>25 % overlap) were identified as heterozygous nucleotide positions. Putative heterozygote individuals were re-sequenced independently to confirm heterozygous sites. We used a Bayesian approach implemented in the program PHASE ver. 2.1 (Stephens *et al.*, 2001) to identify heterozygous haplotypes. Analyses were performed allowing for multiallelic loci (-d option) for 10 000 iterations. In a first run, one multiallelic site showed homoplasy and was removed from further analyses. Next, only haplotypes recovered with posterior probabilities >0·90 were used in subsequent analyses, following Garrick *et al.* (2010).

Nucleotide (π) and haplotype (h) diversities were estimated for each population and overall populations using the software Arlequin ver. 3.11 (Excoffier *et al.*, 2005). An intraspecific phylogeny was inferred using the median-joining algorithm (Bandelt *et al.*, 1999) implemented in the software Network ver. 4.6.1.0 (Forster *et al.*, 2004).

To analyse the genetic differentiation among populations, we performed an analysis of molecular variance (AMOVA; Excoffier *et al.*, 1992) using the software Arlequin ver. 3.11 (Excoffier *et al.*, 2005). We also performed a Bayesian analysis of population structure using the software BAPS ver. 5 (Corander *et al.*, 2008). We ran a population mixture analysis five times to check for convergence with the maximum number of genetically diverged groups $K = 32$.

Analysis of nuclear microsatellite data

Population genetic diversity was assessed by estimating the number of alleles per locus (A), the expected heterozygosity under Hardy–Weinberg equilibrium (H_E), the observed heterozygosity (H_O) and the inbreeding coefficient (f, corrected for null alleles) using Arlequin ver. 3.11 (Excoffier *et al.*, 2005). We also tested for deviation from Hardy–Weinberg expectations and linkage disequilibrium. Allelic richness for a reference sample size of five individuals was analysed with the software FSTAT ver. 2.9.3.2 (Goudet, 2002).

To test for population differentiation, we estimated 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 Arlequin ver. 3.11 (Excoffier *et al.*, 2005). We also estimated Slatkin's R_{ST} (Slatkin, 1995) to verify the contribution of stepwise-like mutations to the genetic differentiation, and tested the hypothesis that $F_{ST} = R_{ST}$ based on Hardy *et al.* (2003), using the software SPAGeDI (Hardy and Vekemans, 2002).

In addition, we used a Bayesian clustering simulation implemented in the software STRUCTURE 2.3.3 (Pritchard *et al.*, 2000) to verify the number of discrete genetic clusters (K) comprised by *D. mollis* samples. Ten independent runs for each K (1–19) were performed to evaluate the consistency of the results using the admixture model of ancestry and correlated allele frequencies. The simulations were performed with a burn-in of 100 000 repetitions, to minimize the effect of the starting configuration, followed by 1 000 000 Markov chain Monte Carlo (MCMC) repetitions of data collection. To assess the most likely number of clusters supported by the data, we used ΔK statistics following Evanno *et al.* (2005).

Demographic history and historical connectivity

To study demographic history, we used coalescent analyses implemented in the software Lamarc 2.1.10 (Kuhner, 2006) to analyse both data sets (ITS and nuclear microsatellites). The demographic parameter theta, $\theta = 4\mu Ne$ (coalescent or mutation parameter for a diploid genome, where Ne is the effective population size), was estimated using an MCMC approach (Beerli and Felsenstein, 2001). To set the priors for ITS, evolutionary model selection was performed using the Akaike Information Criterion implemented in the software jModelTest2 (Darriba *et al.*, 2012). The model HKY + G + I was selected, with gamma shape equal to 0·014 and pinv = 0·954. The analyses were run with 20 initial chains of 4000 steps and three final chains of 50 000 steps. The chains were sampled every 100 steps. We used the default settings for the initial estimate of theta. The program was run four times to certify convergence and validate the analyses using Tracer ver. 1.6 (Rambaut and Drummond, 2007), and combined results were generated. Results were considered when the effective sample size (ESS) was ≥200. We also explored

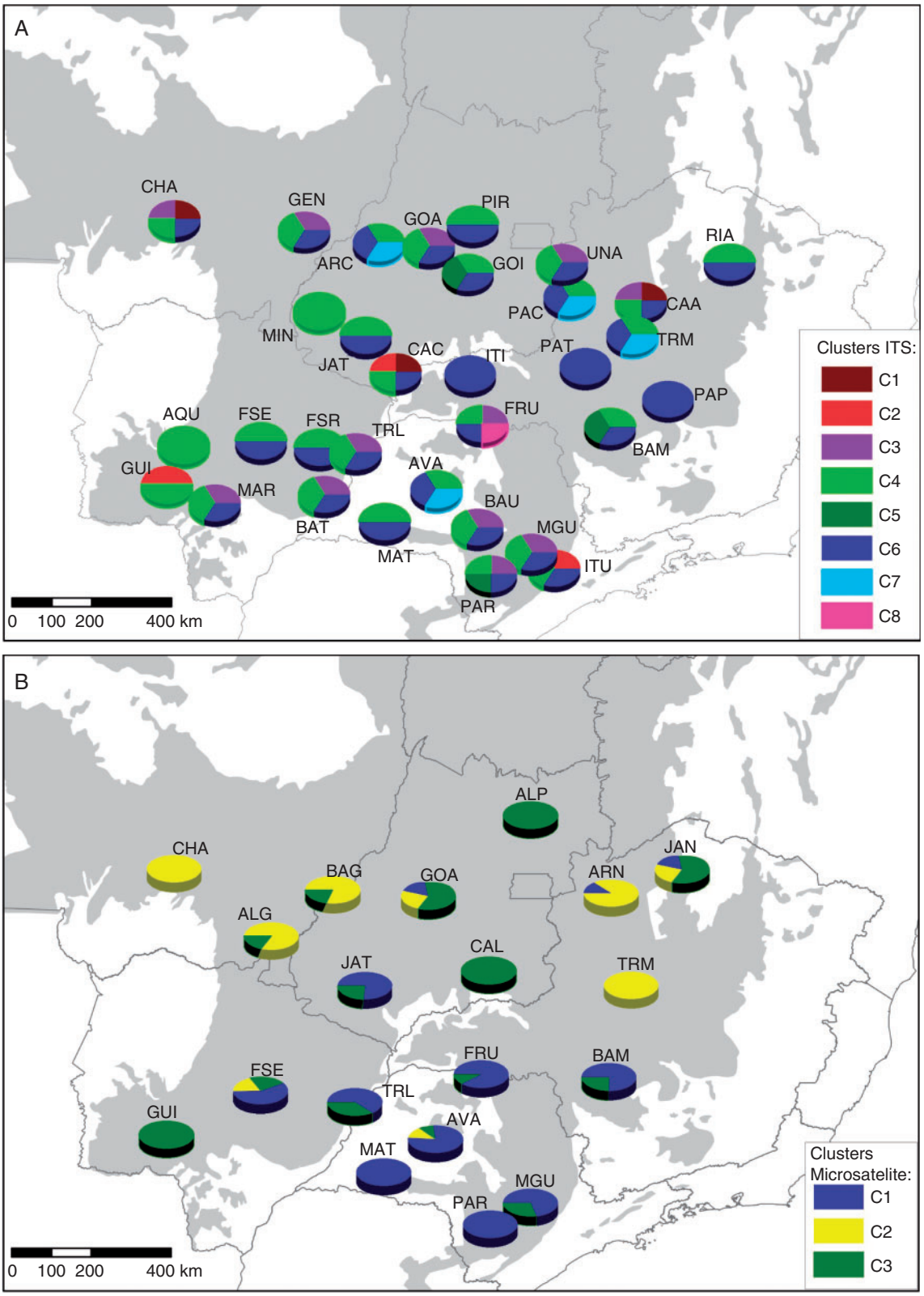

FIG. 2. Bayesian cluster of individuals of *Dimorphandra mollis* for ITS nrDNA (A) and nuclear microsatellites (B). Different colours were assigned to different clusters following the key.

changes in effective population size by estimating the demographical parameter g (exponential growth rate), where $\theta t = \theta now \exp(-gt)$, and t is the time to coalescence in mutational units (Kuhner and Smith, 2007). To assess historical genetic connectivity, we estimated the number of migrants per generation from the scaled migration rate, $M = 4Nem/\theta$, where m is the migration rate. Time to most recent common ancestor (TMRCA) and effective population size were estimated from the mutation parameter θ [Kingman (1982), see also Hein *et al.* (2005)] using a generation time of 20 years (based on flowering time; unpubl. data). Based on the comparison of F_{ST} and R_{ST} (see the Results), we used the lowest mutation rate reported for a microsatellite marker in plants, $2{\cdot}4 \times 10^{-4}$ [95 % confidence interval (CI) = $1{\cdot}4 \times 10^{-4}$; $4{\cdot}2 \times 10^{-4}$] mutations per allele per generation (Thuillet *et al.*, 2002). For ITS, we used the substitution rate reported for another Fabaceae genus, *Inga*, $2{\cdot}34 \times 10^{-9}$ substitutions per site per year (Richardson *et al.*, 2001).

For ITS data we tested the hypothesis of pure demographic population expansion (Rogers and Harpending, 1992) and population range expansion with high levels of migration between neighbouring demes (Ray *et al.*, 2003; Excoffier, 2004) using pairwise mismatch distribution implemented in Arlequin ver. 3.11. We also performed Fu's neutrality tests (Fu, 1997) using Arlequin ver 3.11. Negative values for *Fs* indicate an excess of rare alleles or new mutations in the genealogy resulting from either population expansion or selective sweeps.

Population size dynamics over time for ITSs was also investigated using Bayesian Skyline Plot (BSP; Drummond *et al.*, 2005), implemented in BEAST 1.8.0 (Drummond and Rambaut, 2007). We used the substitution model reported above and the relaxed molecular clock model (uncorrelated lognormal) and the substitution rate reported above. Four independent analyses were run for 30 million generations. Convergence and stationarity were checked, and the independent runs were combined using the software Tracer ver. 1.6. Results were considered when ESS was $\geq$200.

Ecological niche modelling

Occurrence records of *D. mollis* across the Neotropics (Supplementary Data Fig S1, Table S2) were obtained from NeoTropTree (Oliveira-Filho, 2014), a database that consists of tree species checklists with geo-referenced sites compiled from the literature and herbarium specimen records. All records were examined for probable errors and duplicates. These records were mapped in a grid of cells of $0{\cdot}5^\circ \times 0{\cdot}5^\circ$ (longitude × latitude) encompassing the Neotropical region to generate a matrix of 189 presences (cells with unique occurrence records) used for niche modelling (see below).

We also generated environmental layers as predictors for ENM using five bioclimatic variables (annual mean temperature, Bio1; mean diurnal range, Bio2; isothermality, Bio3; precipitation of wettest quarter, Bio16; and precipitation of driest quarter, Bio17) and sub-soil pH (30–100 cm). The bioclimatic variables were obtained from the ecoClimate database (www.ecoclimate.org; Lima-Ribeiro *et al.*, 2015) for pre-industrial (representing current climate conditions), mid-Holocene (6 ka) and the LGM (21 ka) from five coupled Atmosphere–Ocean General Circulation Models (AOGCMs): CCSM4, CNRM–CM5, MIROC–ESM, MPI–ESM-P and MRI–CGCM3 (Supplementary Data Table S3). Sub-soil pH data were obtained from the Harmonized World Soil Database (version 1.1, FAO/IIASA/ISRIC/ISS-CAS/JRC 2009) and assumed to be constant between LGM and pre-industrial, which was used in ENMs as a 'constraint variable' to better model the environmental preferences of *D. mollis*.

Next, the niche of *D. mollis* was modelled using both presence-only, presence–background and presence–absence algorithms (Supplementary Data Table S4): (1) Bioclimatic envelopes (Bioclim), Ecological Niche Factor Analysis (ENFA), Euclidian Distance, Gower distances, Mahalanobis distances; (2) Maximum Entropy (Maxent); and (3) Flexible Discriminant Analysis (FDA), Generalized Linear Models (GLMs), Generalized Additive Models (GAMs), Multivariate Adaptive Regression Splines (MARS), Neural Networks (NNETs) and Random Forest [RNDFOR; see Franklin (2009) for details of the methods].

The niche of *D. mollis* was first modelled in the current climatic scenario (i.e. pre-industrial) and then projected onto palaeoclimatic conditions of the mid-Holocene (6 ka) and the LGM (21 ka) to infer its spatial distribution through the last glaciation. All ENMs used in this study were run in the integrated computational platform BIOENSEMBLES (Diniz-Filho *et al.*, 2009), which implements the ensemble approach (Araujo and New, 2007). For each ENM, the occurrence points were randomly partitioned into two sub-sets comprising 75 and 25 % of the data set for model training and testing, respectively. To avoid biases in data partition, the procedure was repeated 50 times, selecting different combinations of training/testing data sets through model building. The 50 initial models were evaluated by True Skill Statistics (TSS; Allouche *et al.*, 2006), such that only models with high performance (i.e. TSS $>0{\cdot}5$) were combined to generate the consensus maps (an average weighted by TSS value of each model) for both current and past climatic scenarios (see TSS values for all models in Supplementary Data Table S5). The selected models were geographically classified for presence and absence of species by using the threshold that maximizes specificity plus sensibility, and then the consensus maps were computed from the frequency of initial models supporting the occurrence of the species in each cell of the grid. The consensual frequency is then interpreted and used in further analyses as a measure of environmental suitability [see details of the ensemble approach in Diniz-Filho *et al.* (2009) and Collevatti *et al.* (2012*a*)]. The combinations of all modelling components (12 algorithms ×5 AOGCMs) resulted in 60 consensual predictive maps for each time period. The final consensual maps were generated by computing the mean suitability across all 60 models and reflect the resulting potential climate change effects on the distribution dynamics of the species. The consensual maps were also used to identify historical refugia for *D. mollis*; a cell was considered climatically stable if the species was predicted to be present (considering frequencies $>0{\cdot}5$) in that cell during the three time periods (the LGM, the mid-Holocene and the present).

Finally, we applied a hierarchical ANOVA using the predicted suitability from all models (12 ENMs × 5 AOGCMs × 3 time periods) as the response variable to disentangle and map the uncertainties in the potential distribution due to modelling components (i.e. ENMs and AOGCMs) and the effects of climate change on the species distribution through time [i.e, the

'ecological signal'; see Terribile *et al.* (2012) for details of the hierarchical design].

Setting demographic scenarios

The combination of all ENMs and AOGCMs resulted in 60 predictive maps (12 ENMs × 5 AOGCMs) for each time period, which were used to support alternative demographical scenarios. We computed the range shift across the last glacial cycle, and classified all 60 predictive maps according to three general demographic scenarios: (1) 'range stability', no difference in range size over time; (2) 'range expansion', range size was lower in the past than in the present; and (3) 'range retraction', range size was higher in the past than in the present. Classifications were first made by visual inspection and then by computing the difference in range size between combinations of time slices: 21–6 ka, 6 ka–present and 21 ka–present), but keeping constant the combination of algorithms and AOGCMs in the three time periods.

Spatial pattern in genetic diversity

We used spatially explicit analyses to detect spatial patterns in observed genetic diversity in response to late Quaternary climate oscillations, for both ITS and nuclear microsatellite loci. To test for spatial structures, the linearized F_{ST} and geographical distances (in logarithm scale) among pairs of sampled populations were correlated using a Mantel test, and statistical significance was established from 10 000 random permutations using the software Arlequin 3.11.

We used quantile regressions to analyse the relationships of climatic suitability and stability through time with gene diversity (H_E) and allelic richness (Ar) for microsatellite data and haplotype (h) and nucleotide (π) diversities for ITS data (Cade and Noon, 2003). For this, we calculated the difference in ensembled suitabilities between the LGM and the present as a measure of climate stability through the time. Next, we analysed whether historical changes in the species' geographical range generated a spatial cline in genetic parameters due to expansion of climatically suitable conditions. For this, we obtained the distance between each analysed population and the centroid of the historical refugium, and then we performed quantile regressions of genetic parameters against this spatial distance.

RESULTS

Nuclear ITS genetic diversity and population structure

The ITS sequencing generated fragments of 803 bp. We found seven polymorphic sites and 12 haplotypes in the total sample (Table 1). Most populations were polymorphic and only four (AQU, GEN, MIN and PAT) were monomorphic (Table 1). Two haplotypes (H1 and H3) were very widespread and with high frequency (Fig. 1A, B). The other haplotypes found were rare or very rare. Haplotypes H2, H10, H11 and H12 were found in only one population (ARC, FRU and TRL; Fig. 1A). Phylogenetic relationships among haplotypes were unresolved (Fig. 1B). Bayesian analysis of population structure (BAPS) showed eight population clusters (Fig. 2A). Neither network nor BAPS analyses showed any phylogeographical structure, with distant populations sharing the same haplotype and assigned to the same cluster (Figs 1B and 2A).

The AMOVA showed significant genetic differentiation among populations ($F_{ST} = 0{\cdot}299$; $P < 0{\cdot}001$). Populations AQU, GEN, GUI, ITU, MIM, PAP and PAT showed high and significant differentiation with almost all populations (see Supplementary Data Table S6).

Genetic diversity and structure of nuclear microsatellites

Populations of *D. mollis* had moderate genetic diversity across its geographical range (Table 2). From the 19 populations, five had significant inbreeding. Populations were genetically differentiated ($F_{ST} = 0{\cdot}179$, $P < 0{\cdot}0001$). R_{ST} also showed high differentiation among populations ($R_{ST} = 0{\cdot}133$, $P < 0{\cdot}0001$) and was not significantly different from F_{ST} ($P = 0{\cdot}217$), indicating that genetic drift is an important evolutionary force shaping the genetic structure of *D. mollis*. All pairs of populations showed significant genetic differentiation (Supplementary Data Table S7).

The Bayesian analyses showed a maximal value of L(D|K) with $K = 14$ (Supplementary Data Fig, S2A). However, the L(D|K) never reached a plateau and the highest L″(D|K) and delta K were obtained with $K = 3$ (Fig. S2B, C), although they were multimodal and the values decreased and became more variable among runs with $K > 3$. Considering $K = 3$, the clustering (Fig. 2B; see also Supplementary Table S8) showed some geographical pattern. Cluster C1 is located mainly in the southern part of the distribution, whereas clusters C2 and C3 are found in the remaining part of the distribution of the species.

Demographic history and historical connectivity

Coalescent analyses supported large effective population sizes for all populations for ITS (Table 1). For nuclear microsatellites (Table 2), effective population sizes were low. Using the mutation parameter over all populations ($\theta = 0{\cdot}432$), TMRCA for ITS dated to 3·696 Ma (95 % CI = 2·24–6·33 Ma). For nuclear microsatellites, TMRCA dated to 813·5 ka (95 % CI = 757·2–826·7 ka).

For ITS, coalescent analyses showed constant population sizes for most populations and overall population (total population) (Table 1), and the Bayesian Skyline Plot did not detect variation in population size for *D. mollis* in the last approx. 10 Ma (Supplementary Data Fig. S3). Fu's test showed no evidence of population expansion for each population (all $P > 0{\cdot}05$) or overall populations ($Fs = -0{\cdot}262$, $P = 0{\cdot}607$). However, tests of sudden demographic (raggedness index = 0·287, $P = 0{\cdot}056$) or spatial expansion (raggedness index = 0·287, $P = 0{\cdot}404$) did not reject the population expansion assumption for the total population. Furthermore, the mismatch test for individual populations also showed that the observed distributions were not different from those expected according to sudden demographic and range expansion (all raggedness indexes with $P > 0{\cdot}05$). However, highly negative values of the growth parameter (g) for nuclear microsatellites for most populations, as overall population (Table 2), suggest recent population retraction.

TABLE 1. *Genetic diversity and demographic parameters for 32 populations of* Dimorphandra mollis *in Brazil for ITS nrDNA*

Population	*N*	*K*	*h*	π	θ	θ 95 % CI	*Ne*	θ 95 % CI	*g*	*g* 95 % CI
ARC	3	3	0·733	0·0037	0·084	0·009–0·309	8961·0	947·9–33 034·2	–194·07	–485·29 to 30·73
AQU	5	1	0·000	0·0000	0·011	0·003–0·019	1133·0	278·2–2006·1	–51·00	–352·84 to 58·14
AVA	5	3	0·378	0·0012	0·004	0·002–0·006	378·3	256·5–674·5	–70·37	–276·51 to 1·00
BAM	5	3	0·622	0·0019	1·248	0·016–4·334	13 335·0	167·4–46 304·7	–218·52	–452·49 to 217·04
BAT	5	3	0·644	0·0027	0·041	0·006–0·086	4371·7	600·9–9231·9	33·73	–325·97 to 109·88
BAU	5	3	0·644	0·0027	0·508	0·007–2·663	54 273·0	765·1–84 541·9	–208·37	–466·14 to 77·01
CAA	8	4	0·692	0·0025	0·015	0·005–0·028	1586·1	509·3–2979·6	–219·31	–491·48 to 106·98
CAC	4	3	0·607	0·0029	1·341	0·016–7·411	14 328·2	1759·1–79 176·2	–176·71	–463·08 to 137·62
CHA	5	3	0·600	0·0031	0·005	0·002 – 0·013	556·4	254·6–1352·9	–155·60	–482·65 to –5·85
GEN	3	1	0·000	0·0000	1·411	0·010 – 8·209	15071·1	1088·3–87 687·9	–114·47	–445·13 to 100·04
FRU	7	5	0·769	0·0029	0·021	0·007 – 0·043	2253·7	697·6–4584·3	–235·04	–459·11 to 12·82
FSE	5	2	0·467	0·0023	0·009	0·006 – 0·022	1011·1	635·8 – 2379·5	–291·69	–490·43 to 21·57
FSR	5	3	0·600	0·0024	0·751	0·014 – 3·067	80 256·1	15 317·3–327 649·7	–111·27	–452·03 to 262·84
GOA	3	2	0·533	0·0027	2·606	0·025 – 8·155	27838·5	2647·5–87 120·7	–145·31	–465·38 to 110·44
GOI	3	2	0·333	0·0013	1·294	0·019–4·797	13 290·9	2048·1–51 248·3	–178·00	–492·76 to 141·63
GUI	5	2	0·356	0·0004	0·012	0·007–0·039	1312·5	701·8–4151·6	–251·39	–451·29 to –49·86
ITI	5	2	0·200	0·0010	0·005	0·001–0·016	556·0	68·3–1735·7	–286·94	–480·37 to –134·13
ITU	4	3	0·464	0·0018	0·006	0·004–0·012	6264·0	3863·7–12 305·6	–220·24	–420·27 to –51·71
JAT	5	2	0·556	0·0028	0·006	0·005–0·011	679·0	494·9–1125·9	115·34	–80·70 to 132·14
MAR	4	3	0·464	0·0017	2·137	0·042–7·724	22 826·3	4449·6–82 522·6	–159·91	–494·54 to 79·15
MAT	5	2	0·467	0·0023	0·827	0·044–4·319	88 354·0	46 831·2–161 474·6	65·10	–298·89 to 204·61
MGU	5	3	0·600	0·0025	1·148	0·030–4·899	66 175·2	31 733·9–82 339·4	–241·15	–477·85 to 2·41
MIN	5	1	0·000	0·0000	0·010	0·001–0·026	1078·7	58·9–2788·9	–233·24	–495·28 to –43·16
PAC	5	3	0·644	0·0027	0·936	0·009–4·458	10 320·5	9861·9–76 324·9	–65·17	–469·10 to 189·60
PAP	9	2	0·111	0·0004	0·009	0·004–0·015	1008·5	456·6–1563·3	–95·15	–346·01 to 13·65
PAR	4	3	0·679	0·0028	0·290	0·003–0·133	309 61·4	366·2–14252·1	–189·55	–493·15 to –6·73
PAT	9	1	0·000	0·0000	0·002	0·001–0·006	240·3	105·9–693·9	–137·24	–443·66 to –27·51
PIR	4	2	0·536	0·0027	1·470	0·013–7·080	15 707·5	1375·0–75 639·7	–180·90	–466·53 to 121·85
RIA	3	3	0·600	0·0022	0·273	0·003–0·660	29 113·4	30 31·5–70 470·9	–233·05	–493·15 to 23·71
TRL	4	4	0·643	0·0021	0·737	0·006–4·280	78 707·5	6861·1–97211·4	–117·63	–485·73 to 328·64
TRM	4	2	0·571	0·0029	1·670	0·020–6·172	17 837·6	2148·9–65 939·3	5·34	–482·71 to 327·61
UNA	4	3	0·607	0·0025	0·381	0·005–1·635	40 705·3	5776·1–74 626·7	–46·58	–426·80 to 221·95
Mean (s.d.)	4·8 (1·5)	2·6 (0·9)	0·119 (0·097)	0·0019 (0·0011)	–	–			–	–
Overall	155	12	0·622 (0·017)	0·3030 (0·1872)	0·432	0·262–0·740	46 197·5	28 020·8–79 097·5	–384·45	–517·00 to 244·13

N, sample size; *k*, number of haplotypes; *h*, haplotype diversity; π, nucleotide diversity; θ, coalescent parameter; *Ne*, effective number; *g*, exponential growth parameter; 95 % CI is the confidence interval at 95 %.

For ITS, most populations received high gene flow ($N_em \geq 1{\cdot}00$, see Supplementary Data Tables S9 and S10) and also were a source of migrants to other populations, such as ARC, BAM, BAT, BAU, CAC, FSR, GEN, GOA, GOI, MAR, MAT, MGU, PAC, PAR, PIR, RIA, TRL, TRM and UNA. However, for microsatellite loci, gene flow was low among almost all population pairs (see Tables S11 and S12). For microsatellites, only a few populations (MAT, ARN, BAM, TRL and JAN) were a source of migrants to several populations (see Tables S11 and S12).

Ecological niche modelling

The potential distribution of *D. mollis* extends over a widely continuous area through central–west Brazil since the LGM (Fig. 3), matching the Brazilian Cerrado limits, with higher variation in the peripheral areas. The spatio-temporal dynamics of species range show an increase in suitability over time (Fig. 3), with a slightly smaller range size in the LGM (see Supplementary Data Fig. S4). The scenario of 'range expansion' from the LGM to the present was favoured in the most predictive maps (43 %; Table S13), although 'range stability' was almost equally supported (40 %). A large historical refugium of *D. mollis* (area of high climatic suitability throughout time) was predicted from ENMs to occur in central–northern and north-western portions of its current geographical range, where most populations occur currently (Fig. 4). After the LGM, there was an eastward and southward expansion of the range.

Considering the predictions across all 60 maps (12 algorithms × 5 AOGCMs) through the three time periods, the greatest predictive uncertainty (see Supplementary Data Fig. S5) came from AOGCMs [proportional sum of squares (SS) = 0·30], followed by the time component (SS = 0·26) and algorithms (SS = 0·24). Moreover, their higher variances from the time component (i.e. the ecological signal) are distributed across central and north-east Brazil, whereas the methodological uncertainties from AOGCM and algorithm components occur mainly outside central Brazil (Fig. S5). Such results indicate that our ENMs were able to capture reliably the effect of climate change on geographical distribution of *D. mollis* (i.e. ecological signal), despite some methodological noise.

Spatial patterns in genetic diversity

We found a weak relationship between the geographical distance and the genetic differentiation among pairs of populations for both microsatellite loci (Mantel test, $R^2 = 0{\cdot}036$, $P = 0{\cdot}061$)

TABLE 2. *Genetic diversity and demographic parameters for 17 populations of* Dimorphandra mollis *in Brazil based on seven nuclear microsatellite loci*

Population	*N*	*A*	*Ar*	H_E	H_O	*f*	θ	θ 95 % CI	*Ne*	*Ne* 95 % CI	*g*	*g* 95 % CI
ALG	17	5·0	3·7	0·619	0·541	0·071	0·015	0·006–0·027	15·6	5·8–27·9	–264·23	–475·15 to 1·00
ALP	19	4·7	3·3	0·582	0·557	0·019	0·014	0·008–0·026	14·2	8·1–26·9	–118·13	–439·55 to –3·33
ARN	21	5·1	3·7	0·628	0·523	–0·023	0·024	0·004–0·060	24·5	4·7–62·2	–258·38	–497·46 to –41·18
AVA	21	5·1	3·6	0·602	0·455	0·100*	0·013	0·002–0·031	13·2	2·6–32·1	–324·01	–498·77 to –105·24
BAM	20	3·4	2·7	0·481	0·423	0·054	0·019	0·011–0·034	20·2	11·7–35·9	–209·39	–456·14 to –29·38
BAG	21	4·9	3·5	0·622	0·485	0·142*	0·009	0·005–0·012	9·1	5·6–12·9	–249·05	–452·57 to –64·14
CAL	22	3·6	2·5	0·417	0·399	0·043	0·015	0·008–0·028	15·3	8·6–28·9	–288·58	–459·23 to –71·04
CHA	15	5·1	3·8	0·625	0·509	0·126*	0·015	0·002–0·029	15·9	1·8–30·2	–252·84	–497·36 to –48·12
FRU	21	4·7	3·5	0·605	0·541	0·087	0·013	0·004–0·032	13·8	3·8–33·3	–306·54	–488·86 to –70·96
FSE	20	6·1	3·9	0·659	0·616	0·068	0·014	0·007–0·025	14·7	7·6–26·4	–152·65	–459·89 to 9·64
GOA	21	6·4	4·1	0·672	0·622	0·008	0·018	0·006–0·057	18·5	6·0–59·2	–209·47	–470·29 to –25·47
GUI	17	4·0	3·3	0·606	0·511	0·044	0·020	0·006–0·038	20·4	6·2–39·9	–249·24	–481·09 to –66·01
JAN	22	6·9	4·5	0·723	0·608	0·075	0·024	0·005–0·049	25·5	5·5–50·8	–88·25	–424·70 to 22·96
JAT	20	4·6	3·5	0·595	0·510	0·065	0·010	0·003–0·016	10·5	3·4–16·8	–222·17	–464·15 to –19·61
MAT	17	5·1	3·4	0·575	0·487	0·154*	0·083	0·010–0·124	86·4	10·2–129·1	5·71	–338·30 to 104·46
MGU	16	3·4	2·7	0·468	0·431	0·081	0·015	0·003–0·036	16·0	3·4–37·4	–284·75	–497·55 to –46·21
PAR	26	4·7	3·0	0·522	0·514	0·021	0·009	0·003–0·018	8·9	2·6–18·9	–227·30	–474·88 to –0·45
TRL	20	6·0	3·9	0·681	0·610	0·044	0·023	0·009–0·048	24·2	9·1–50·2	–169·86	–431·16 to 7·36
TRM	15	4·1	3·4	0·607	0·492	0·124*	0·018	0·005–0·052	18·4	5·5–53·8	–204·42	–475·57 to 1·00
Mean (SD)	19·5 (2·8)	4·9 (0·9)	3·5 (0·5)	0·594 (0·076)	0·518 (0·066)	0·069 (0·047)	–	–				
Overall	371	–	–	0·610 (0·059)	0·568 (0·069)	0·054 (0·200)	9·762	9·086–9·921	10 168·8	9464·9–10 334·42	–411·33	–537·42 to –48·91

N, sample size; *A*, mean number of alleles per locus; *Ar*, allelic richness; H_E, expected heterozygosity under Hardy–Weinberg equilibrium; H_O, observed heterozygosity; *f*, inbreeding coefficient; θ, coalescent parameter; *g*, exponential growth parameter; *Ne*, effective number; 95 % CI, is the confidence interval at 95 %.

*Significant at the 0·05 level.

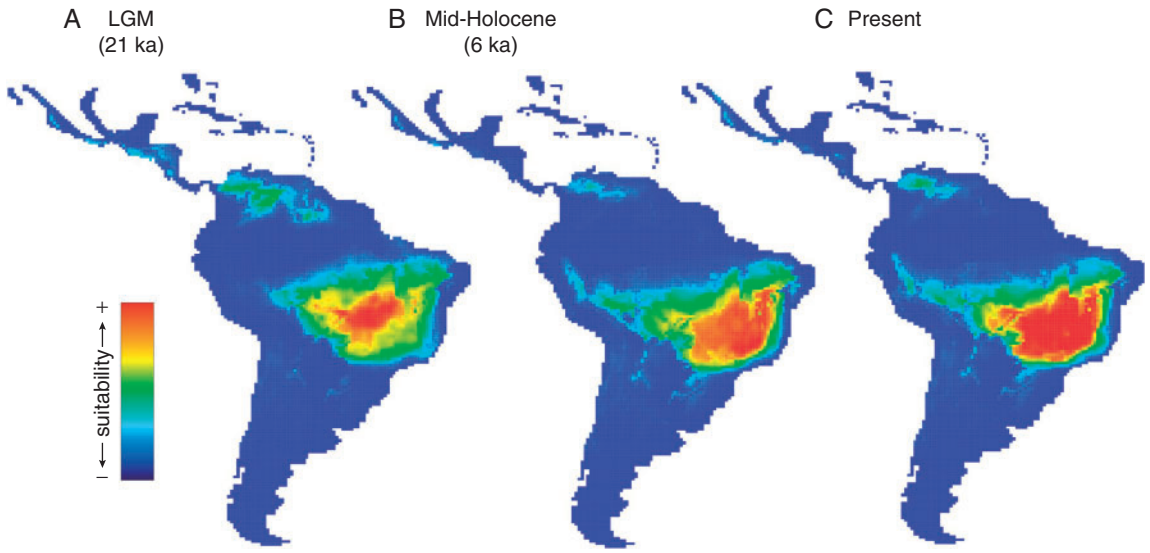

FIG. 3. Maps of consensus expressing the ensemble climatic suitability for *Dimorphandra mollis*, and hence its predicted potential distribution across the Neotropics during the (A) LGM (21 ka), (B) mid-Holocene (6 ka) and (C) the present.

and ITS (Mantel test, $R^2 = 0{\cdot}018$, $P = 0{\cdot}022$). Although the relationship was significant for ITS, the determination coefficient was very low for both markers. Quantile regressions showed clear effects of climate change on the genetic structure of *D. mollis*. Relationships resembling triangular-shaped envelopes showed that populations in more climatically suitable areas during the LGM, mid-Holocene and the present presented high allelic richness (*Ar*) and expected heterozygosity (H_E), especially for higher quantiles (see Supplementary Data Figs S6 and S7). Inbreeding (*f*), nucleotide (π) and haplotype (*h*) diversity were not significantly related to suitability.

Likewise, higher values of both H_E and *Ar* are associated with populations in more stable areas, where climatic suitability have been unchanged since the LGM (Fig. 4 and Supplementary Data Fig. S8). In addition, both H_E and *Ar* decreased with distances from the centroid of the refugium (higher quantiles) and *f* increased (lower quantiles; see Supplementary Data Fig. S9).

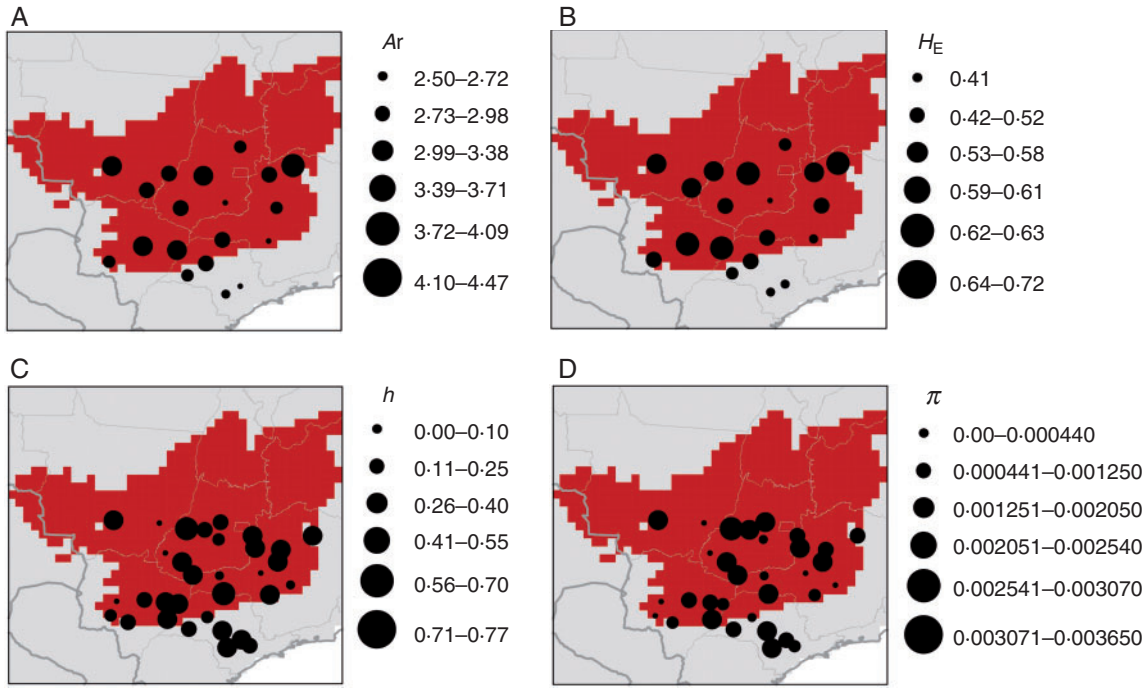

FIG. 4. Spatial distribution of genetic diversity of *Dimorphandra mollis* in relation to the historical refugium, i.e. areas climatically suitable throughout time (in red). (A) Allelic richness (*Ar*). (B) Expected heterozygosity (H_E). (C) Haplotype diversity (*h*). (D) Nucleotide diversity (π). Circumference sizes are proportional to the value of the genetic parameter, following the key.

DISCUSSION

Our study coupling phylogeographical and genetic analyses with markers with different evolutionary rates and ENM provided evidence that the current pattern of *D. mollis* genetic diversity is the result of a decrease in suitable area for the species, and a slight geographical range retraction during the LGM, with maintenance of a large stable area as predicted by ENMs. Our results showed different signatures for ITS and nuclear microsatellites corresponding to different time slices in the evolution of *D. mollis*, most probably due to differences in evolutionary rates. Signs of Quaternary climate changes could not be found in ITS sequences. ITS haplotypes had a more ancient TMRCA (approx. 3·7 Ma) dated to the Pliocene and showed that populations of *D. mollis* had constant sizes at least in the last 10 Ma, supported by BSP and coalescent analyses, with large effective populations sizes. However, more recently, the population suffered a demographical retraction evidenced by coalescent analyses based on microsatellite data, leading to small effective population sizes. Because of the lower evolutionary rate of ITS, the cycles of population retraction in glacial periods and subsequent expansions across cyclic interglacial periods may have hindered the signal of more ancient bottleneck events.

Connectivity among populations in deeper time was higher than more recently. Coalescent analyses showed high migration among almost all populations for ITS leading to haplotype sharing among populations and the assignment of individuals from distant populations to the same cluster and absence of phylogeographical structure. For microsatellites, migration was low (N_em <1·0), and only a few populations were sources of migrants (e.g. MAT, ARN, TRL and JAN). Besides the moderate genetic diversity, MAT was an important source of migrants for all populations. This population had the highest effective size for microsatellite loci even though ENM showed that MAT remained in unsuitable areas during the glacial periods, outside the historical refugium. It is possible that MAT remained in microrefugia not detected by the ENM due to the data resolution for niche modelling, and maintained a relatively large effective size.

The slight reduction in suitable area during the LGM may have caused population extinction or shrinkage in some parts of the *D. mollis* distribution, leading to low effective population size and, as a consequence, relatively low genetic diversity. These results indicate that genetic drift was an important factor shaping the current distribution of the genetic diversity of *D. mollis*, and that historical changes in the species' range may have affected the patterns of species genetic diversity. Therefore, the genetic patterns found may be the outcome of a bottleneck because of a decrease in suitability across species distribution and a slight range retraction in southern and eastern parts due to harsh climatic conditions during the LGM followed by range expansion after climate amelioration, mainly after the mid-Holocene (6·0 ka). These results are consistent with palaeo-palynological studies. Pollen records from central Brazil indicated a long drier period between about 19 ka and about 7 ka (from the LGM until the mid-Holocene), suggesting that the current savanna areas were under severe water stress during this period (Salgado-Labouriau *et al.*, 1998). Also in south-eastern Brazil, pollen records show that during the LGM savannas were mostly replaced by grassland (Behling, 2002). From about 7 to 5 ka, rainfall started to increase and since approx 4·6 ka vegetation in central Brazil did not suffer

significant changes, having the same geographical pattern as the present (Salgado-Labouriau *et al.*, 1998).

Indeed, spatially explicit analyses showed a spatial cline pattern in allelic richness and expected heterozygosity, related to the palaeodistribution at 21 ka, 6 ka and the present time. In addition, the significant relationship with the centre of distribution of the historical refugium reinforces the spatial cline pattern and shows higher genetic diversity in more stable populations near the centre of the historical refugium. The inbreeding was also higher in peripheral populations (Table 2; Fig. S9). This pattern is consistent with the abundant-centre model, i.e. loss of genetic variability due to genetic drift and inbreeding in peripheral populations with low effective population sizes and higher diversity in central populations (Soulé, 1973).

Despite the slight range retraction during the LGM, our results showed that the populations of *D. mollis* have moderate genetic diversity for both ITS sequences and microsatellites. Low to moderate diversity has already been documented for this species with other markers (Souza and Lovato, 2010). Indeed, ENMs revealed that most populations are inside the historical glacial refugium, i.e. in areas that have had high climatic suitability throughout the last glaciation. Such evidence suggests that even with the aridity registered in central Brazil by pollen data at the end of the Pleistocene (Salgado-Labouriau *et al.*, 1998), a wide climatically stable area was still available throughout the last glacial cycle, favouring population persistence and genetic connectivity. Populations outside the historical refugium (MGU, PAR, BAM, MAT, GUI and TRM), which were potentially disconnected from the others in some periods, presented the lowest values of both expected heterozygosity (H_E) and allelic richness (*Ar*). Hence, east and south of the present distribution of *D. mollis* was most probably colonized by populations that remained in the refugium and expanded by sending migrants after the climate became moister and warmer after the LGM. Other evidence of recent colonization is the fact that southern populations have different frequencies of simple sequence repeat (SSR) clusters as compared with northernmost areas, a signal of a founder effect. This pattern was also found in another savanna legume tree, *Dalbergia miscolobium* (Novaes *et al.*, 2013).

Results on the *D. mollis* demographic history and phylogeographical structure are different from those obtained for most of the other Fabaceae trees from the Neotropical savanna. In contrast to *D. mollis*, strong phylogeographical structure and high genetic divergence among populations were found with ITS sequences in *D. miscolobium* (Novaes *et al.*, 2013), as well as in two other legume tree species with chloroplast DNA, *Hymenaea stigonocarpa* (Ramos *et al.*, 2007) and *Plathymenia reticulata* (Novaes *et al.*, 2010). These differences can be due to different effects of Quaternary climatic changes on *D. mollis* in comparison with these species, since they showed genetic evidence of strong range retraction and recent population expansion, whereas *D. mollis* presented only slight range retraction, with a large stable area maintained since the LGM. The distribution dynamics through the last glacial cycle of *D. mollis* were also different from those of *Dypterix alata* (Collevatti *et al.*, 2013*b*). Although both species presented range retraction during the LGM, *D. alata* showed lower genetic diversity and stronger genetic differentiation. In contrast to *D. mollis*, the ENM showed a strong instability for populations of *D. alata* in central Brazil during the LGM and a range shift towards north-west and north-east Brazil, with a clear spatial cline pattern in genetic diversity. Life-history attributes do not appear to explain the differences found in *D. mollis* in relation to other legume trees. Both *D. mollis* and *D. alata* are pollinated by insects and have their seeds dispersed by mammals, in spite of differences in genetic diversity and distribution during the LGM. Still, the genetic diversity and genetic differentiation found for *D. mollis* was lower than for two other widely distributed non-legume tree species from central Brazil, *Caryocar brasiliense* (Collevatti *et al.*, 2001) and *Tabebuia aurea* (Collevatti *et al.*, 2014). In fact, there is evidence of 'range retraction' with multiple refugia during the LGM for *C. brasiliense* and a wide historical refugium across central Brazil for both *C. brasiliense* and *T. aurea* (Collevatti *et al.*, 2012*b*, 2014), similar to what we found for *D. mollis*. These differences in demographic history and phylogeographical patterns between savanna tree species highlight the complexity of the biogeographical processes in the Cerrado biome. Regardless of these singularities, most species show the highest genetic diversity in the central portion of the Cerrado.

An ENM study (Terribile *et al.*, 2012) involving past and predicted future climate (end of the 21st century) of 18 Cerrado tree species, including *D. mollis*, projected a reduction of suitable area in relation to current distributions. In particular, the study projected southern and eastern displacement of the core area, with loss of suitability in most of the northern and western parts of the biome. A single climatically stable area was identified in the central area of the Cerrado for most species through time (the LGM, the present and the future). The large historical refugium that we identified for *D. mollis* includes, mainly in its eastern part, the climatically stable area identified by Terribile *et al.* (2012). This climatically stable area encompasses parts of three floristic regions of the Cerrado biome (central–western, central and south-eastern, and north and north-eastern), which harbour high species diversity (Ratter *et al.*, 2003). The highest genetic diversity found in *D. mollis* in the refugium is concordant with phylogeographic studies with other tree species, which also found that the central Cerrado harbours high genetic diversity, encompassing most species' genetic lineages (Ramos *et al.*, 2007; Novaes *et al.*, 2010, 2013; Ribeiro *et al.*, 2016). The evolutionary potential of a species is dependent on genetic diversity. Thus, the climatic refugium of the Cerrado, which is predicted to be maintained under future climate scenarios, should be given high priority for conservation in order to maintain the evolutionary potential of the species present. Conservation efforts for this area of low-velocity climatic change in the central Cerrado could also favour rare endemic species, as pointed out by Petit *et al.* (2015), as well as various ecological networks, including pollinators and disperses.

Over the past few decades, the Cerrado has suffered substantial destruction due to agricultural expansion, which is expected to continue over the coming years. The net loss of natural vegetation in the Cerrado from 1990 to 2010 was estimated to be 0·6 % per year (Beuchle *et al.*, 2015). Besides this rapid destruction and fragmentation, the absence of a consistent fire management policy, which has transformed several areas of savanna vegetation into forests (Durigan and Ratter, 2016), threatens the biodiversity of the Cerrado. In addition, only 8·3 % of the area of the biome is protected by reserves (Ganem *et al.*,

2013). All of these factors should be taken into account when considering conservation policies to safeguard the role of the climatic refugium in the maintenance of biodiversity under climate change.

SUPPLEMENTARY DATA

Supplementary data are available online at www.aob.oxfordjournals.org and consist of the following. Table S1: populations of *Dimorphandra mollis* sampled for ITS and nuclear microsatellite analyses. Table S2: contemporary occurrence records (189) of *D. mollis* represented by the centroid of grid cells across the Neotropics used in the ecological niche modelling. Table S3: details on the palaeoclimatic simulations (AOGCMs) used in the ecological niche modelling of *D. mollis*. Table S4: ecological niche modelling methods used to estimate *D. mollis* potential distribution. Table S5: values of True Skill Statistics (TSS), with mean and confidence interval, for all ENM × AOGCM combinations from ecological niche modelling of *D. mollis*. Table S6: pairwise F_{ST} for 32 populations of *D. mollis* based on ITS sequencing. Table S7: pairwise F_{ST} for 19 populations of *D. mollis* based on seven nuclear microsatellite loci. Table S8: proportion of membership (Q) of each pre-defined population of *Dimorphandra mollis* in each of the three clusters inferred by Bayesian analyses implemented in STRUCTURE software, based on seven nuclear microsatellite loci. Table S9: number of migrants per generation for 32 populations of *D. mollis*, based on ITS sequencing. Table S10: credibility interval (95 %) of the number of migrants per generation for 32 populations of *D. mollis*, based on ITS sequencing. Table S11: number of migrants per generation for 19 populations of *D. mollis*, based on seven nuclear microsatellites. Table S12: credibility interval (95 %) of the number of migrants per generationfor 19 populations of *D. mollis*, based on seven nuclear microsatellites. Table S13: demographic scenarios resulting from the palaeodistribution dynamics of *D. mollis* as predicted by the 60 maps (12 ENMs × 5 AOGCMs) obtained from the ecological niche modelling. Figure S1: current geographical distribution of *D. mollis* across the Neotropics based on 189 occurrence records used for ecological niche modelling. Figure S2: posterior probability graphs of the Bayesian clustering simulation implemented in the software STRUCTURE 2.3.3 for 371 individuals of *D. mollis*, as a function of K averaged over five independent runs. Figure S3. Bayesian Skyline Plot for *D. mollis*, based on the sequencing of ITS. Figure S4: range size and shift (range difference among time periods) for *D.a mollis* at present-day (0 ka), mid-Holocene (6 ka) and LGM (21 ka). Figure S5: maps of uncertainty (relative sum of squares) for the modelling components of *D. mollis*. Figure S6: quantile regression between allelic richness and suitability from the present day, mid-Holocene and the LGM for *D. mollis*. Figure S7: quantile regression between genetic diversity and suitability from the present day, mid-Holocene and the LGM for *D. mollis*. Figure S8: quantile regression of environmental stability since the LGM with allelic richness and genetic diversity of *D. mollis*. Figure S9: quantile regression of distance from the centroid of the historical refugium with allelic richness, genetic diversity and inbreeding of *D. mollis*.

ACKNOWLEDGEMENTS

This work was supported by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). We thank the Sistema de Autorização e Informação em Biodiversidade (SISBio) do Instituto Chico Mendes (ICMBio), which provided research facilities for the fieldwork. We also thank Thiago F. Rangel for providing access to the computational platform BIOENSEMBLES and the World Climate Research Programmer's Working Group on Coupled Modelling (Table S3) for making available their model outputs. R.G.C., J.P.L.-F. and M.B.L. are continuously supported by grants and fellowships from CNPq and CAPES, which we gratefully acknowledge.

LITERATURE CITED

Ab'Saber AN. 2000. Spaces occupied by the expansion of dry climates in South America during the Quaternary ice ages. *Revista do Instituto Geológico* **21**: 71–78.

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

Araújo MB, New M. 2007. Ensemble forecasting of species distributions. *Trends in Ecology and Evolution* **22**: 42–47.

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**: 37–48.

Beerli P, Felsenstein J. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences, USA* **98**: 4563–4568.

Behling H. 2002. South and southeast Brazilian grasslands during Late Quaternary times: a synthesis. *Palaeogeography, Palaeoclimatology, Palaeoecology* **177**: 19–27.

Beuchle R, Grecchi RC, Shimabukuro YE, *et al.* 2015. Land cover changes in the Brazilian Cerrado and Caatinga biomes from 1990 to 2010 based on a systematic remote sensing sampling approach. *Applied Geography* **58**: 116–127.

Brookfield JFY. 1996. A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology* **5**: 453−455.

Cade BS, Noon BR. 2003. A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment* **1**: 412–420.

Collevatti RG, Grattapaglia D, Hay JD. 2001. Population genetic structure of the endangered tropical tree species *Caryocar brasiliense*, based on variability at microsatellite loci. *Molecular Ecology* **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. *Molecular Ecology* **12**: 105–115.

Collevatti RG, Terribile LC, Lima-Ribeiro MS, *et al.* 2012*a*. A coupled phylogeographical and species distribution modelling approach recovers the demographical history of a Neotropical seasonally dry forest tree species. *Molecular Ecology* **21**: 5845–5863.

Collevatti RG, Lima-Ribeiro MS, Souza-Neto AC, Franco AA, Terribile LC. 2012*b*. Recovering the demographical history of a Brazilian Cerrado tree species *Caryocar brasiliense*: coupling ecological niche modeling and coalescent analyses. *Natureza e Conservação* **10**: 169–176.

Collevatti RG, Terribile LC, Oliveira G, *et al.* 2013*a*. Drawbacks to palaeodistribution modelling: the case of South American seasonally dry forests. *Journal of Biogeography* **40**: 345–358.

Collevatti RG, Telles MPC, Nabout JC, Chaves LJ, Soares TN. 2013*b*. Demographic history and the low genetic diversity in *Dipteryx alata* (Fabaceae) from Brazilian Neotropical savannas. *Heredity* **111**: 105.

Collevatti RG, Lima-Ribeiro MS, Terribile LC, Guedes LB, Rosa FF, Telles MP. 2014. Recovering species demographic history from multi-model inference: the case of a Neotropical savanna tree species. *BMC Evolutionary Biology*. **14**: 213. doi:10.1186/s12862-014-0213-0.

Collevatti RG, Terribile LC, Diniz-Filho JA, Lima-Ribeiro MS. 2015*a*. Multi-model inference in comparative phylogeography: an integrative

approach based on multiple lines of evidence. *Frontiers in Genetics* **6**: 31. doi: 10.3389/fgene.2015.00031.

Collevatti RG, Terribile LC, Rabelo SG, Lima-Ribeiro MS. 2015*b*. Relaxed random walk model coupled with ecological niche modelling unravel the dispersal dynamics of a Neotropical savannah tree species in the deeper Quaternary. *Frontiers in Plant Science* **6**: 653. doi:10.3389/fpls.2015.00653.

Corander J, Marttinen P, Siren J, Tang J. 2008. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* **9**: 539.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**: 772.

Diniz-Filho JAF, Bini LM, Rangel TF, *et al.* 2009. Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography* **32**: 897–906.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**: 214. doi:10.1186/1471-2148-7-214.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* **22**: 1185–1192.

Durigan G, Ratter JA. 2015. The need for a consistent fire policy for Cerrado conservation. *Journal of Applied Ecology* **53**: 11–15.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**: 2611–2620.

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

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.

Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**: 47–50.

Forster P, Bandelt HJ, Röhl A. 2004. *Network 4.2.0.1*. Available online at: http://www.fluxus-engineering.com (last accessed 10 March 2015).

Franklin J. 2009. *Mapping species distributions: spatial inference and predictions*. Cambridge: Cambridge University Press.

Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**: 915–925.

Furley PA, Ratter JA. 1988. Soil resources and plant communities of the Central Brazilian cerrado and their development. *Journal of Biogeography* **15**: 97–108.

Ganem RS, Drummond JA, Franco JLA. 2013. Conservation policies and control of habitat fragmentation in the Brazilian Cerrado biome. *Ambiente e Sociedade* **16**: 99–117.

Garrick RC, Sunnucks P, Dyer RJ. 2010. Nuclear gene phylogeography using PHASE: dealing with unresolved genotypes, lost alleles, and systematic bias in parameter estimation. *BMC Evolutionary Biology* **10**: 118.

Gonçalves AC, Reis CAF, Vieira FA, Carvalho D. 2010. Spatial genetic structure in natural populations of *Dimorphandra mollis* (Fabaceae) in the north of Minas Gerais State, Brazil. *Revista Brasileira de Botânica* **33**: 325–332.

González S, Cosse M, Franco, MR, *et al.* 2015. Population structure of mtDNA variation due to Pleistocene fluctuations in the South American maned wolf (*Chrysocyon brachyurus*, Illiger, 1815): management units for conservation. *Journal of Heredity* **106**: 459–468.

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 (last accessed 15 March 2015).

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. *Molecular Ecology Notes* **2**: 618–620.

Hein J, Schierup M, Wiuf C. 2005. *Gene genealogies, variation and evolution: a primer in coalescent theory*. Oxford: Oxford University Press.

Kingman JFC. 1982. The coalescent. *Stochastic Processes and their Applications* **13**: 235–248.

Knowles LL. 2009. Statistical phylogeography. *Annual Review of Ecology, Evolution, and Systematics* **40**: 593–612.

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.

Lima-Ribeiro MS, Varela S, González-Hernández J, *et al.* 2015. ecoClimate: a database of climate data from multiple models for past, present, and future for Macroecologists and Biogeographers. *Biodiversity Informatics* **10**: 1–21.

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

Novaes RML, Lemos-Filho JP, Ribeiro RA, Lovato MB. 2010. Phylogeography of *Plathymenia reticulata* (Leguminosae) reveals patterns of recent range expansion towards northeastern Brazil and southern Cerrados in Eastern Tropical South America. *Molecular Ecology* **19**: 985–998.

Novaes RML, Ribeiro RA, Lemos-Filho JP, Lovato MB. 2013. Concordance between phylogeographic and biogeographical patterns in the Brazilian Cerrado: diversification of the endemic tree *Dalbergia miscolobium* (Fabaceae). *PLoS One* **8**: e82198. doi:10.1371/journal.pone.0082198.

Oliveira-Filho AT. 2014. *NeoTropTree, Flora arbórea da Região Neotropical: um banco de dados envolvendo biogeografia, diversidade e conservação*. Universidade Federal de Minas Gerais. http://www.icb.ufmg.br/treeatlan/ (last accessed 18 December 2014).

Peterson AT, Soberón J, Pearson RG, *et al.* 2011. *Ecological niches and geographic distributions*. Princeton, NJ: Princeton University Press.

Petit RJ, Lafontaine G, Ouayjan A, Corcket E, Ducousso A, Hampe A. 2015. Quaternary refugia in centers of endemism. In: *Ecological implications of endemism hotspots as climate change refugia. Symposium. The 2015 ESA Annual Meeting*. https://eco.confex.com/eco/2015/webprogram/Paper50704.html.

Prado CPA, Haddad, CFB, Zamudio KR. 2012. Cryptic lineages and Pleistocene population expansion in a Brazilian Cerrado frog. *Molecular Ecology* **21**: 921–941.

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 v1.4*. Free available from http://beast.bio.ed.ac.uk/Tracer.

Ramos ACS, Lemos-Filho JP, Ribeiro RA, Santos FR, Lovato MB. 2007. Phylogeography of the tree *Hymenaea stigonocarpa* (Fabaceae: Caesalpinioideae) and the influence of Quaternary climate changes in the Brazilian Cerrado. *Annals of Botany* **100**: 1219–1228.

Ratter JA, Ribeiro JF, Bridgewater S. 1997. The Brazilian cerrado vegetation and threats to its biodiversity. *Annals of Botany* **80**: 223–230.

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 Journal of Botany* **60**: 57–109.

Ratter JA, Bridgewater S, Ribeiro JF. 2006. Biodiversity patterns of the woody vegetation of the Brazilian cerrado. In: Pennington RT, Lewis GP, Ratter JA, eds. *Neotropical savannas and seasonally dry forests: plant diversity, biogeography and conservation*. Boca Raton, FL: CRC Press, 31–66.

Ray N, Currat M, Excoffier L. 2003. Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution* **20**: 76–86.

Ribeiro PC, Lemos-Filho JP, Buzatti RSO, Lovato MB, Heuertz M. 2016. Species-specific phylogeographic patterns and Pleistocene east-west divergence in *Annona* (Annonaceae) trees in the Brazilian Cerrado. *Botanical Journal of the Linnean Society* **181**: 21–36.

Richardson JE, Pennington RT, Pennington TD, Hollingsworth PM. 2001. Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science* **293**: 2242–2245.

Rogers AR, Harpending HC. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9**: 552–569.

Salgado-Labouriau ML, Barberi M, Ferraz-Vicentini KR, Parizzi MG. 1998. A dry climatic event during the late Quaternary of tropical Brazil. *Review of Palaeobotany and Palynology* **99**: 115–129.

Santos MG, Nogueira C, Giugliano LG, Colli GR. 2014. Landscape evolution and phylogeography of *Micrablepharus atticolus* (Squamata, Gymnophthalmidae), an endemic lizard of the Brazilian Cerrado. *Journal of Biogeography* **41**: 1506–1519.

Silva MF. 1986. *Flora Neotrópica: Dimorphandra (Caesalpiniaceae)*. The New York Botanical Garden, New York.

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

Soulé ME. 1973. The epistasis cycle: a theory of marginal populations. *Annual Review of Ecology and Systematics* **4**: 165–187.

Souza HAV, Lovato MB. 2010. Genetic diversity of the critically endangered tree *Dimorphandra wilsonii* and the widespread in the Brazilian Cerrado *D. mollis*: implications for conservation. *Biochemical Systematics and Ecology* **38**: 49–56.

Souza HAV, Muller LAC, Brandão RL, Lovato MB. 2012*a*. Isolation of high quality and polysaccharide-free DNA from leaves of *Dimorphandra mollis* (Leguminosae), a tree from the Brazilian Cerrado. *Genetics and Molecular Research* **11**: 756–764.

Souza HAV, Collevatti RG, Lemos-Filho JP, Santos FR, Lovato MB. 2012*b*. Development of microsatellite markers for *Dimorphandra mollis* (Leguminosae), a widespread tree from the Brazilian Cerrado. *American Journal of Botany* **99**: e102–e104.

Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68**: 978–989.

Tamura K, Peterson D, Peterson N, *et al.* 2011. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**: 2731–2739.

Terribile LC, Lima-Ribeiro MS, Araújo MB, *et al.* 2012. Areas of climate stability in the Brazilian Cerrado, disentangling uncertainties through time. *Natureza e Conservação* **10**: 152–159.

Thompson J, Higgins D, Gibson T. 1994. ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**: 4673–4690.

Thuillet AC, Bru D, David J, *et al.* 2002. Direct estimation of mutation rate for 10 microsatellite loci in durum wheat, *Triticum turgidum* (L.) Thell. ssp durum desf. *Molecular Biology and Evolution* **19**: 122–125.

Turchetto-Zolet AC, Pinheiro F, Salgueiro F, Palma-Silva C. 2013. Phylogeographic patterns shed light on evolutionary process in South America. *Molecular Ecology* **22**: 1193–213.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**: 535–538.

Weir BSCockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.

Wright S. 1951. The genetic structure of populations. *Annals of Eugenic* **15**: 323–354.

Wright S, Keeling J, Gillman L. 2006. The road of Santa Rosalia: a faster tempo of evolution in tropical climates. *Proceedings of the National Academy of Sciences, USA* **103**: 7718–7722.