

Unravelling the genetic differentiation among varieties of the Neotropical savanna tree *Hancornia speciosa* Gomes

Rosane G. Collevatti^{1,*}, Eduardo E. Rodrigues¹, Luciana C. Vitorino², Matheus S. Lima-Ribeiro³,
Lázaro J. Chaves⁴ and Mariana P. C. Telles^{1,5}

¹Laboratório de Genética & Biodiversidade, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, GO 74001-970, Brasil, ²Instituto Federal Goiano, Rio Verde, GO 74001-970, Brasil, ³Laboratório de Macroecologia, Instituto de Biociências, Universidade Federal de Goiás (UFG), Regional Jataí, Cx.p. 03, Jataí, GO 75801-615, Brasil, ⁴Escola de Agronomia, Universidade Federal de Goiás, Goiânia, GO 74001-970, Brasil and ⁵Escola de Ciências Agrárias e Biológicas, Pontifícia Universidade Católica de Goiás, Goiânia, GO 74605-010, Brasil

*For correspondence. E-mail rosanegc68@hotmail.com

Received: 20 September 2017 Returned for revision: 8 February 2018 Editorial decision: 29 March 2018 Accepted: 4 April 2018
Published electronically 20 April 2018

- **Background and Aims** Spatial distribution of species genetic diversity is often driven by geographical distance (isolation by distance) or environmental conditions (isolation by environment), especially under climate change scenarios such as Quaternary glaciations. Here, we used coalescent analyses coupled with ecological niche modelling (ENM), spatially explicit quantile regression analyses and the multiple matrix regression with randomization (MMRR) approach to unravel the patterns of genetic differentiation in the widely distributed Neotropical savanna tree, *Hancornia speciosa* (Apocynaceae). Due to its high morphological differentiation, the species was originally classified into six botanical varieties by Monachino, and has recently been recognized as only two varieties by Flora do Brasil 2020. Thus, *H. speciosa* is a good biological model for learning about evolution of phenotypic plasticity under genetic and ecological effects, and predicting their responses to changing environmental conditions.
- **Methods** We sampled 28 populations (777 individuals) of Monachino's four varieties of *H. speciosa* and used seven microsatellite loci to genotype them.
- **Key Results** Bayesian clustering showed five distinct genetic groups ($K = 5$) with high admixture among Monachino's varieties, mainly among populations in the central area of the species geographical range. Genetic differentiation among Monachino's varieties was lower than the genetic differentiation among populations within varieties, with higher within-population inbreeding. A high historical connectivity among populations of the central Cerrado shown by coalescent analyses may explain the high admixture among varieties. In addition, areas of higher climatic suitability also presented higher genetic diversity in such a way that the wide historical refugium across central Brazil might have promoted the long-term connectivity among populations. Yet, F_{ST} was significantly related to geographic distances, but not to environmental distances, and coalescent analyses and ENM predicted a demographical scenario of quasi-stability through time.
- **Conclusions** Our findings show that demographical history and isolation by distance, but not isolation by environment, drove genetic differentiation of populations. Finally, the genetic clusters do not support the two recently recognized botanical varieties of *H. speciosa*, but partially support Monachino's classification at least for the four sampled varieties, similar to morphological variation.

Key words: Apocynaceae, Bayesian clustering, Cerrado, coalescence, ecological niche modelling, hierarchical AMOVA, isolation by distance, isolation by environment, microsatellites, multiple matrix regression with randomization, quantile regression.

INTRODUCTION

It has long been known that species may undergo genetic differentiation due to restricted gene flow among distant populations (isolation by distance; Wright, 1943). However, spatial distribution of genetic diversity may also be caused by interactions with the environment, resulting in patterns of isolation by environment in which genetic and environmental distances are positively correlated, independent of geographical distance (Wang and Bradburd, 2014). Species with a wide geographical distribution may experience high environmental heterogeneity,

thus isolation by environment may be important in shaping genetic differentiation. Spatial structure in genetic diversity may be highly affected by ecological factors as a result of selection against immigrants or hybrids, or non-random mating among populations due to local adaptation (Sexton *et al.*, 2014)

Additionally, spatial displacements of a species' geographical range in response to Quaternary glaciations (i.e. habitat tracking) might also have affected the spatial distribution of genetic diversity and genetic differentiation among populations (Hewitt, 1996; Excoffier *et al.*, 2009). The cycles of range retraction in glacial periods and subsequent expansion

TABLE 1. Genetic diversity of the 28 populations (776 individuals) of *Hancornia speciosa* based on seven microsatellite loci

Population	<i>N</i>	<i>A</i>	<i>A_r</i>	<i>H_e</i>	<i>H_o</i>	<i>f</i>	θ	95 % CI	<i>N_e</i>	95 % CI
<i>H. speciosa</i> var. <i>cuyabensis</i>	154	8.9 (1.1)	4.02 (0.32)	0.689 (0.051)	0.591 (0.046)	0.144 (0.053)				
HCCAG	32	8.7	4.002	0.685	0.616	0.102	0.010	0.008–0.012	223.97	185.20–261.67
HCCHG	32	9.7	4.023	0.674	0.619	0.082	0.013	0.011–0.016	286.06	232.84–357.02
HCDIA	33	10.1	4.554	0.778	0.631	0.192	0.012	0.010–0.014	272.75	230.62–301.58
HCGCA	25	8.3	3.734	0.646	0.519	0.200	0.006	0.005–0.008	124.18	103.11–187.77
HCTAS	32	7.6	3.789	0.666	0.571	0.144	0.008	0.008–0.009	190.93	169.23–213.03
<i>H. speciosa</i> var. <i>gardinerii</i>	379	9.3 (2.2)	4.27 (0.48)	0.700 (0.065)	0.639 (0.093)	0.118 (0.049)				
HGBRA	32	13.1	4.81	0.712	0.661	0.073	0.014	0.013–0.015	306.02	279.41–339.28
HGCOC	14	8.9	4.61	0.763	0.721	0.057	0.013	0.012–0.015	292.71	274.97–337.06
HGFAN	18	10.3	4.73	0.761	0.690	0.096	0.026	0.019–0.032	567.68	410.24–700.73
HGFRO	32	6.9	3.26	0.559	0.482	0.139	0.008	0.006–0.011	187.44	126.48–237.27
HGJAR	32	9.5	4.39	0.726	0.652	0.104	0.022	0.019–0.024	496.72	410.24–523.33
HGMAT	32	10.9	4.48	0.769	0.757	0.016 ^{n.s.}	0.009	0.007–0.010	207.84	164.51–230.62
HGMCA	19	9.0	4.33	0.715	0.584	0.187	0.008	0.006–0.013	184.62	140.18–281.62
HGMUT	6	4.6	4.57	0.765	0.821	–0.087 ^{n.s.}	0.031	0.014–0.049	685.21	301.58–1106.53
HGNAT	32	11.3	4.37	0.674	0.591	0.124	0.013	0.011–0.014	290.49	248.36–314.89
HGPER	32	9.3	3.68	0.603	0.519	0.142	0.008	0.007–0.010	183.99	161.66–221.75
HGPIR	33	9.7	4.29	0.678	0.628	0.074	0.010	0.009–0.012	223.97	193.33–270.54
HGPOT	32	7.1	3.47	0.641	0.615	0.040 ^{n.s.}	0.011	0.009–0.012	248.36	193.31–272.75
HGPNA	32	12.3	4.71	0.751	0.683	0.092	0.011	0.010–0.013	243.93	223.97–290.49
HGSEL	33	8.1	4.11	0.689	0.542	0.215	0.010	0.009–0.012	226.19	197.3–263.88
<i>H. speciosa</i> var. <i>pubescens</i>	146	10.6 (1.5)	4.64 (0.17)	0.737 (0.045)	0.682 (0.079)	0.103 (0.066)				
HPAPA	23	9.7	4.62	0.693	0.557	0.200	0.039	0.023–0.061	871.48	505.59–1359.33
HPCOC	18	10.0	4.62	0.766	0.738	0.038 ^{n.s.}	0.014	0.012–0.018	314.89	257.23–396.93
HPFAN	14	8.4	4.56	0.761	0.712	0.066	0.023	0.017–0.034	514.46	368.11–751.73
HPJAR	32	12.4	4.89	0.776	0.759	0.022 ^{n.s.}	0.032	0.022–0.042	698.51	487.85–926.92
HPMUT	26	11.1	4.76	0.760	0.717	0.059	0.015	0.010–0.018	330.41	221.75–405.80
HPNIQ	33	11.7	4.41	0.667	0.611	0.085	0.022	0.014–0.040	496.72	299.36–887.00
<i>H. speciosa</i> var. <i>speciosa</i>	97	9.9 (1.9)	4.16 (0.64)	0.677 (0.074)	0.604 (0.055)	0.164 (0.156)				
HSJAP	32	7.9	3.47	0.599	0.617	–0.031 ^{n.s.}	0.007	0.004–0.010	150.14	96.2–221.75
HSPIP	33	11.4	4.73	0.746	0.544	0.274	0.018	0.014–0.024	403.59	314.89–527.77
HSROV	32	10.7	4.29	0.687	0.651	0.054	0.008	0.006–0.011	178.24	136.46–243.93
Mean overall	–	9.6 (1.9)	4.29 (0.45)	0.704 (0.059)	0.636 (0.083)	0.126 (0.062)	9.973	9.881–10.004	2810.88	2784.94–2819.62

N, number of individuals analysed; *A*, mean number of alleles per locus; *A_r*, allelic richness; *H_e*, expected heterozygosity based on the Hardy–Weinberg equilibrium; *H_o*, observed heterozygosity; *f*, inbreeding coefficient (values followed by n.s. are not significantly different from zero, *P* > 0.05). The s.d. is shown in parentheses.

Average values of *A_r*, *H_e* and *f* are not significantly different between varieties (*P* > 0.05). θ , coalescent or mutation parameter; *N_e*, effective population size; 95 % CI is the credibility interval at 95%.

of genes (Coelho and Valva, 2001; Chaves, 2006). However, some varieties are sympatric or parapatric, and contact zones among two or more varieties, as well as individuals with intermediate phenotype in hybrid zones, can be observed (Chaves, 2006). In fact, varieties can cross-pollinate under experimental conditions (Collevatti et al., 2016), and no significant effects of heterosis or exogamic depression were detected in hybrids for early growth components, but maternal contribution significantly affected plant height (Collevatti et al., 2016). These results raise the question of which mechanisms maintain the morphological differentiation among *H. speciosa* varieties and whether the differentiation has been driven by genetic drift or ecological adaptations. Differentiation in phenotypic traits among populations may be due to an environmental effect such

as phenotypic plasticity (e.g. Hoffmann et al., 2005) or a genetic effect such as local adaptation or drift (e.g. Mitchell-Olds and Schmitt, 2006; Leinonen et al., 2008). By answering these questions, *H. speciosa* becomes a good biological model for learning about the evolution of phenotypic plasticity and predicting its responses to changing environmental conditions.

Here, we analyse the genetic diversity and differentiation among *H. speciosa* varieties based on microsatellite loci. Specifically we test whether historical demographical dynamics may explain patterns of differentiation among varieties using coalescent analyses and Bayesian clustering. By coupling spatially explicit analyses and ENM, we also test whether differentiation among populations better responds to isolation by distance or isolation by environment.

MATERIALS AND METHODS

Populations and sampling

We sampled 28 local populations (777 adult individuals) of *H. speciosa* (Fig. 1, see also Supplementary Data Table S1). Vouchers were collected in each population to check variety identification using leaf morphology following Monachino (1945) and compared with herbarium material from the Federal University of Goiás (Universidade Federal de Goiás) in Goiânia. Our sampling included four different varieties following Monachino (1945): *H. speciosa* var. *cuyabensis*, *H. speciosa* var. *gardneri*, *H. speciosa* var. *pubescens* and *H. speciosa* var. *speciosa*. Following the new classification, our sampling included the two accepted varieties (Flora do Brasil 2020): *H. speciosa* var. *speciosa* (including *H. speciosa* var. *cuyabensis* and *H. speciosa* var. *gardneri*) and *H. speciosa* var. *pubescens*. We focused our sampling efforts on savannas of central-west Brazil due to the current higher abundance of *H. speciosa* and our focus on biogeography of Neotropical savannas. In addition, all varieties of *H. speciosa* can be found in Cerrado biome and, although *H. speciosa* occurs in the Amazon basin (see Supplementary Data Fig. S1), its populations in these sites are found in low numbers only in enclaves of savannas (we could find only one or two individuals in most places we visited). Because of differences in plant density and fragment sizes, the number of sampled individuals ranged from six to 33 (Table 1). In four localities we found sympatric populations of *H. speciosa* var. *gardnerii* and *H. speciosa* var. *pubescens* (Fig. 1; see also Supplementary Data Table S1).

Genetic analysis

To obtain genetic data, we genotyped all sampled individuals using seven microsatellite loci (Rodrigues et al., 2015). DNA was extracted from leaves using the 2 % cetyltrimethylammonium bromide (CTAB) protocol (Doyle and Doyle, 1987). Forward primers were marked with fluorescent dyes 6-FAM, HEX and NED (Applied Biosystems, CA, USA). PCR was performed in a 15 µL reaction with 5 ng of template DNA, 1× reaction buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl₂), 250 µM of each dNTP, 333 µg of bovine serum albumin (BSA), 2.16 µM of each primer and 1 U of *Taq* DNA polymerase (Phonutria, Brazil). Amplifications were performed in a GeneAmp PCR System 9700 (Applied Biosystems) with one cycle at 95 °C for 5 min, 30 cycles at 95 °C for 1 min, annealing temperature for 1 min (specific for each primer pair, see Rodrigues et al., 2015), 72 °C for 1 min and a final step at 72 °C for 30 min to enforce 3' *Taq* adenylation. PCR fragments were electrophoresed with GeneScan 500 internal lane standard (ROX, Applied Biosystems) in an ABI Prism 3500 Genetic Analyzer (Applied Biosystems) and automatically sized using the software GeneMapper® v4.1 software (Applied Biosystems).

Some individuals were genotyped twice to control for genotyping errors, genotypes were visually reviewed and then Micro-Checker software (van Oosterhout et al., 2004) was used to detect errors due to stutter bands, allele drop-out and null alleles. We found no evidence of genotyping errors due to

stutter bands or allele drop-out when the raw data were analysed. We also found no evidence of null alleles.

Genetic differentiation among varieties and populations

To understand the genetic differentiation among varieties, we first performed a Bayesian clustering simulation to assess the number of discrete genetic clusters (K) and unravel genetic admixture among varieties using the software Structure 2.3.4 (Pritchard et al., 2000). To minimize the effect of the starting configuration, simulations were performed with a burn-in period of 100 000 repetitions followed by 1 000 000 MCMC (Markov Chain Monte Carlo) repetitions of data collection. To evaluate the consistency among results, we performed five independent runs for each K using the admixture model of ancestry and correlated allele frequencies. Post-processing of data was performed using STRUCTURE HARVESTER (Earl and von Holdt, 2012) and Clumpp (Jakobsson and Rosenberg, 2007). We used the ΔK method (Evanno et al., 2005) to assess the most likely number of clusters supported by the data.

Next, we performed hierarchical analyses of variance (ANOVA; Weir and Cockerham, 1984) apportioning the total variance into variance components and calculating fixation indices (Wright, 1951). We first performed an analysis based on the four sampled Monachino's varieties to apportion the total variance and obtain the fixation index among varieties (F_{CT} , among populations within varieties (F_{SC}), among individuals within populations (F_{IS}) and among individuals in the total population (F_{IT}). We also performed a hierarchical analysis based on four groups of Bayesian structure clusters (see the Results). Analyses were performed using hierarchical analysis of molecular variance (AMOVA) implemented in the software Arlequin 3.5 (Excoffier et al., 2005).

Demographical history

Population demography and historical connectivity. We assessed population demographical history by performing coalescent analyses (Kingman, 1982) implemented in the software Lamarc 2.1.8 (Kuhner, 2006). We obtained the demographical parameter theta, $\theta = 4\mu N_e$ (coalescent or mutation parameter for a diploid genome, where N_e is the effective population size), and explored changes in effective population size by estimating the demographical parameter g (exponential growth rate), where $\theta_t = \theta_{\text{now}} \exp(-gt)$, 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, $M = 4N_e m/\theta$, where m is the migration rate. Because of the high number of populations, we constrained migration (maintained migration constant) to estimate the growth rate, and migration was estimated in independent runs. The analyses were run with 20 initial chains of 10 000 steps and three final chains of 100 000 steps. The chains were sampled every 100 steps following a burn-in of 10 000 steps. The analyses were repeated four times, under a stepwise mutation model, and convergence and stability of the outcome were verified using Tracer v1.5 (Rambaut and

Drummond, 2007) and combined using LogCombiner 1.7.1 (Rambaut and Drummond, 2007). The results were considered reliable only when the effective sample size (ESS) was ≥ 200 .

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 5 years. Generation time was estimated based on flowering time data from a common garden (see Collevatti et al., 2016). We used the mutation rate reported for microsatellite markers in plants, 8.87×10^{-4} (s.e. = 2.57×10^{-4}) mutations per allele per generation (Marriage et al., 2009), often quoted in the range of 10^{-2} – 10^{-4} per locus per generation (Udupa and Baum, 2001; Thuillet et al., 2002; Marriage et al., 2009). We chose a low value based on the low genetic diversity (see the Results).

Species palaeodistribution. We used ENM to estimate the current and past potential distribution of *H. speciosa* in the mid-Holocene [6000 years ago (ka)] and the LGM (21 ka). We obtained occurrence records of *H. speciosa* across Neotropics from the online databases Species Link (<http://www.splink.org.br>), ‘Tropicos’ (<http://www.tropicos.org>) and ‘Flora Integrada da Região Centro-Oeste’ (<http://www.florescer.unb.br>). All records were examined for probable errors and duplicates, and the nomenclature was checked for synonymies. The occurrence records were mapped in a grid of cells of $0.5^\circ \times 0.5^\circ$ (longitude \times latitude) encompassing the entire Neotropical region (Supplementary Data Fig. S1; Table S2), and resulted in a matrix with 900 occurrence records used to build the ecological niche models.

The environmental layers were represented by five bioclimatic variables (annual mean temperature, mean diurnal range, isothermality, precipitation of the wettest month and precipitation of the driest month) and sub-soil pH (30–100 cm) from Harmonized World Soil Database – version 1.1 (FAO, 2009). The bioclimatic variables were selected from the highest loading on each of the five first factors after performing a factorial analysis with Varimax rotation (the number of factors was established from Scree plot) using the full set of 19 bioclimatic variables (<http://www.worldclim.org/bioclim>). While sub-soil pH (30–100 cm) was used as a constraint variable in all time periods to improve the palaeodistribution models, the bioclimatic variables were obtained for pre-industrial (representing current climate conditions), mid-Holocene and the LGM to characterize the climate oscillation through the last glacial cycle. The bioclimatic variables were obtained at a 0.5° spatial resolution from the ecoClimate database (www.ecoclimate.org; Lima-Ribeiro et al., 2015) for five coupled Atmosphere–Ocean General Circulation Models (AOGCMs): CCSM4, CNRM-CM5, MIROC-ESM, MPI-ESM-P and MRI-CGCM3 (Supplementary Data Table S3).

Estimates of the current and past potential distribution of *H. speciosa* were obtained using ten presence only, presence–background and presence–absence ENM algorithms (Supplementary Data Table S4; see also Franklin 2009 for details of the methods). Because absence data for *H. speciosa* are not available, we randomly selected pseudo-absences across the Neotropical grid cells keeping prevalence equal to 0.5 (Stokland et al., 2011) to calibrate the ecological niche models based on presence–absence and presence–background observations. The selection of pseudo-absences and all ecological

niche models were run in the integrated computational platform BIOENSEMBLES (Diniz-Filho et al., 2009).

The procedures for niche modelling using the ensemble approach in BIOENSEMBLES have been discussed extensively elsewhere (see Diniz-Filho et al., 2009; Collevatti et al., 2012a, 2013a). For each algorithm, models were built using a calibration sub-set of 75 % of the presence of cells selected at random and then evaluated against the remaining 25 %, repeating this procedure 50 times. During this initial phase, models with poor performance were eliminated based on True Skill Statistics (TSS; Allouche et al., 2006) and the area under the receiver operating characteristic (ROC) curve (AUC; Fielding and Bell, 1997), such that only models with high performance (i.e. TSS > 0.70 and AUC > 0.75) were combined to generate the consensus maps (an average weighted by the TSS value of each model) for both current and past climatic scenarios (see TSS and AUC values for all models in Supplementary Data Table S5). The resulting models were used to generate occurrence maps based on thresholds established by the ROC curve, from which the frequency of occurrence in each Neotropical grid cell (i.e. the suitability) was obtained for each algorithm, AOGCM and time period. This frequency of occurrence is used here as a measure of suitability for *H. speciosa*’s occurrence in each Neotropical grid cell.

The combination of all ecological niche models and AOGCMs resulted in 50 predictive maps (10 ENMs \times 5 AOGCMs) for each time period. We applied a hierarchical ANOVA using the predicted suitability from all models (10 ENMs \times 5 AOGCMs \times 3 times) as the response variable to disentangle and map the uncertainties in the potential distribution due to modelling components, ENMs, AOGCMs, and to investigate the distribution shifts through time. The ENM and AOGCM components were nested into the time component, but crossed by a two-way factorial design within each time period (see Terribile et al., 2012 for details of hierarchical design). The final ensemble at each time period was obtained using the predictive performance (TSS) from each model to compute a weighted mean. Moreover, the ensemble maps were used to delimit the historical refugia (suitable areas through time), i.e. all grid cells with suitability values ≥ 0.5 in the three time periods.

Finally, we computed the range for each time interval and range shift (difference in predicted range size between current and 6 ka, 6 and 21 ka, and current and 21 ka) and classified the 50 predictive maps according to three general demographical scenarios: (1) ‘range stability’, no difference in range size through time; (2) ‘range expansion’, range size was higher in the past (at 6 or 21 ka) than at present; and (3) ‘range retraction’, range size was lower in the past (at 6 or 21 ka) than at present.

Spatial patterns in genetic diversity

We estimated the number of alleles per locus (A) and allelic richness (A_r), based on rarefaction analysis (Mousadik and Petit, 1996). We also estimated genetic diversity using expected heterozygosity (H_e) under Hardy–Weinberg equilibrium (Nei, 1978). Observed heterozygosity (H_o) and the inbreeding coefficient (f) were also estimated to test for deviation from Hardy–Weinberg equilibrium using randomization-based tests. Analyses were performed with the software FSTAT 2.9.3.2 (Goudet, 2002).

We compared genetic diversity and allelic richness among varieties using a permutation test implemented in FSTAT 2.9.3.2 with 10 000 permutations. Then, we tested whether genetic differentiation among populations may be explained by isolation by distance or isolation by environment using spatially explicit analyses. We obtained linearized F_{ST} among pairs of populations using the software Arlequin 3.11 (Excoffier *et al.*, 2005), the geographical distance matrix (in logarithm scale) and the environmental distance matrices. We built environmental distance matrices based on the suitability obtained from ENM (difference in suitability among pairs of populations) and four bioclimatic variables (temperature seasonality, mean temperature of the driest quarter, precipitation seasonality and precipitation of the driest quarter). The linearized F_{ST} values among pairs of populations were correlated with either geographical or environment distance matrices using a multiple regression analysis on matrices of genetic, geographical and environmental distances based on the MMRR approach (multiple matrix regression with randomization; Wang, 2013). Statistical significance of matrix correlations was established from 10 000 random permutations. We also performed an autocorrelation analyses using Moran's I to understand the effect of spatial scale in genetic differentiation implemented in the software SAM (Spatial Analysis in Macroecology; Rangel *et al.*, 2010).

To understand how potential changes in species distribution through time affected the spatial distribution in genetic diversity, we analysed the relationship of climate suitability and stability through time (the difference of suitability between time intervals; i.e. 0–6 and 6–21 ka) with genetic diversity (H_e), allelic richness (A_e), θ and effective population size (N_e) using quantile regression (Cade and Noon, 2003). We also analysed whether historical changes in species' geographical range generated a cline spatial pattern in genetic diversity and allelic richness due to expansion, contraction and spatial displacements of climatically suitable conditions. For this, we obtained, for each population, the distance from the centroids of the potential distributions at present, and at 6 and 21 ka, as well as to the centroid of the historical refugium. Then, we performed quantile regressions of genetic parameters against these spatial distances from the centroids.

RESULTS

Genetic differentiation among varieties and populations

Bayesian clustering showed an optimum of five genetic groups ($K = 5$; Supplementary Data Fig S2). Individuals of *H. speciosa* var. *speciosa* from populations HSJAP and HSROV were clustered together (cluster 3, blue, Fig. 1; see also Supplementary Data Table S6), but individuals from population HSPJP showed high admixture with clusters 1 (red) and 4 (yellow) similar to populations of *H. speciosa* var. *gardneri*. Populations of *H. speciosa* var. *gardneri* from the north-west (HGNAT, HGPNA and HGCOC) were grouped in the same cluster (clusters 4, yellow, Fig. 1) with a sympatric population of *H. speciosa* var. *pubescens* (HPCOC). Populations of *H. speciosa* var. *cuyabensis* were grouped mainly in cluster 5 (pink HCGCA, HCDIA and HCTAS) with a population of *H. speciosa* var. *gardneri* from the west (HGPO). However, a population of *H. speciosa* var. *cuyabensis*, HCCHG, showed high admixture with clusters 2 (green) and 1 (red), whereas population HCCAG was assigned

to cluster 2 (green) with populations of *H. speciosa* var. *gardneri*, HGSEL, HGMCA, HGFRO and HGPER. Even though Structure inferred $K = 5$ as the most likely number of groups, clusters 1 (red) and 4 (yellow) comprised mainly sympatric populations of *H. speciosa* var. *gardneri* + *H. speciosa* var. *pubescens* from the central Cerrado. Indeed, Bayesian clustering for $K = 4$ assigned these populations mainly to cluster 4 (see Supplementary Data Table S7) and the other clusters were similar for both $K = 4$ and $K = 5$.

Genetic differentiation among Manachino's four varieties was low but significant ($F_{CT} = 0.027$, $P = 0.0006$), whereas differentiation among populations within varieties was higher ($F_{SC} = 0.107$, $P < 0.001$). The within-population fixation index ($F_{IS} = 0.103$, $P < 0.001$) showed non-random mating within populations, leading to a high total fixation index ($F_{IT} = 0.203$, $P < 0.001$). We performed a hierarchical AMOVA based on four Bayesian groups (Fig. 1): cluster 3, corresponding to populations from the north-eastern Cerrado (HSROV and HSJAP); cluster 1+ cluster 4, comprising populations from the central Cerrado (*H. speciosa* var. *gardneri* + *H. speciosa* var. *pubescens*); cluster 5, comprising populations of the western Cerrado (HCTAS, HCDIA, HCCHG, HCGCA and HGPO); and cluster 2, including populations from the southern and south-western Cerrado (HCCAG, HGSEL, HGFRO, HGPER and HGMCA). The analysis showed higher variation among groups ($F_{CT} = 0.058$, $P < 0.001$) compared with the analysis considering Manachino's four varieties ($F_{CT} = 0.027$). In addition, differentiation among populations within groups ($F_{SC} = 0.087$, $P < 0.001$) and the within-population fixation index ($F_{IS} = 0.083$, $P < 0.001$) were more similar. Pairwise F_{ST} was high for most comparisons (Supplementary Data Table S8), ranging from $F_{ST} = 0.004$ (between HPAPA and HGBRA) and $F_{ST} = 0.290$ (between HSJAP and HCCHG).

Demographical history

Population demography and historical connectivity. All populations of *H. speciosa* had low θ (mutation parameter) but high effective population sizes (Table 1). The overall mutation parameter was high ($\theta = 9.973$, Table 1) and TMRCA dated from 56.218 ka [95 % confidence interval (CI) 55.698–56.392 ka], and effective population size was $N_e = 2810.88$ (95 % CI 2784.94–2819.62). We also found no evidence of historical demographic retraction for *H. speciosa* because the growth parameter (g) was not statistically different from zero for all populations (credibility intervals including zero; results not shown).

Gene flow was low among most population pairs ($N_m < 1.00$, Supplementary Data Table S9), indicating low historical genetic connectivity among them. However, in general, populations from the central Cerrado showed higher connectivity ($N_m > 1.00$), such as populations HGFAN, HGJAR, HGMUT, HPAPA, HPFAN, HPJAR, HPMUT, HPNIQ and HSPJP.

Species palaeodistribution. The climatic conditions currently occupied by *H. speciosa* clearly show its preference for hot and drier climates, matching the current general conditions of Neotropical savannas (Supplementary Data Fig. S1). By modelling the climatic preferences, ecological niche models predicted the highest levels of suitability for *H. speciosa* across central and north-east Brazil through the last glaciation (Fig. 2), overlapping

the core area of its known distribution in the biomes Cerrado (central Brazil) and Caatinga (north-east Brazil). During the LGM (21 ka; Fig. 2), the potential distribution of *H. speciosa* had a spatial displacement towards north-west Brazil, and then returning to central Brazil, expanding also towards the south-east during the mid-Holocene and at present (Fig. 2).

Besides the spatial displacements, such distribution dynamics show that the range size (in the number of grid cells) increased from the LGM to the mid-Holocene and then to the present day (Supplementary Data Fig. S3a). The highest range shift occurred between the LGM and mid-Holocene (Supplementary Data Fig. S3b). In addition, the region across central Brazil was probably a historical refugium maintaining populations of *H. speciosa* during the climate changes throughout the last glacial cycle (Fig. 2).

The hierarchical ANOVA showed that time and AOGCM components had the highest proportion of modelling uncertainties (time = 0.25; AOGCM = 0.35). The ENM component presented the lowest proportion of mean squares (ENM = 0.18) and with higher uncertainty outside the areas where *H. speciosa* was predicted to occur (Supplementary Data Fig. S4). Such a pattern suggests that the ENM algorithms converged in their main prediction about the distribution of the species through time. Moreover, higher uncertainty from the time component was distributed in central, south-east and north-east Brazil (Supplementary Data Fig. S4), indicating that the ecological niche models were able to recover the effects of climate changes through the last glacial cycle on *H. speciosa*'s distribution dynamics.

Most maps predicted a range expansion through time (45.3 % of maps), i.e. a geographical range size in LGM smaller than at present (Supplementary Data Fig. S5), followed by range stability (32.7 %), i.e. the geographical range size was constant through time. Range retraction, i.e. the geographical range size in the LGM being larger than at present (Supplementary Data Fig. S5), was predicted by 22 % of maps.

Spatial patterns in genetic diversity

Most populations showed high polymorphism, with the mean number of alleles ranging from 4.6 to 13.1 (Table 1), and genetic diversity (H_e) ranging from 0.559 to 0.778. Most populations showed a significant reduction in heterozygosity due to inbreeding (Table 1). The four varieties showed no significant difference in mean allelic richness (A_r), genetic diversity (H_e) or inbreeding (Table 1).

Genetic differentiation amongst pairs of populations (MMRR, full model $r^2 = 0.122$; $P = 0.068$) was significantly correlated with geographical distance (MMRR, $r^2 = 0.099$; $P = 0.001$), although only 9.9 % of the variation in population pairwise F_{ST} could be explained by geographical distance. Genetic differentiation among populations was not significantly correlated either with climatic suitability ($r^2 = 0.001$, $P = 0.762$) or with climatic variables, such as temperature seasonality ($r^2 = 0.009$, $P = 0.102$), mean temperature of the driest quarter ($r^2 = 0.009$, $P = 0.186$), precipitation seasonality ($r^2 = 0.009$, $P = 0.354$) and precipitation of the driest quarter ($r^2 = 0.001$, $P = 0.741$). The results were similar when we used logarithmic transformation for geographical and environmental distances. Autocorrelation

analysis showed a significant relationship between genetic differentiation and geographical distance for populations up to 400 km ($P = 0.006$; Supplementary Data Fig. S6).

Quantile regressions showed positive correlation of genetic diversity (H_e) and allelic richness (A_r) with suitability at present, during the mid-Holocene and during the LGM (Fig. 3; Supplementary Data Fig. S7). However, H_e and A_r were negatively correlated with climate stability through time (Supplementary Data Fig. S8). The increase in suitability from 21 ka and 6 ka to the present was associated with higher values of both H_e and A_r (Supplementary Data Fig. S8). Moreover, the distances from the centroids of current, 6 ka and 21 ka distributions, as well as from the centroid of the historical refugium were poorly correlated with both H_e and A_r (Supplementary Data Fig. S9). Similarly, θ and N_e increase with suitability (Fig. 3) at present, during the mid-Holocene and during the LGM (Supplementary Data Fig. S10), but were not correlated with stability (Supplementary Data Fig. S11). In addition, both θ and N_e were poorly correlated with the distance from the centroids of current, 6 ka and 21 ka distributions, as well as from the centroid of the historical refugium (Supplementary Data Fig. S12).

DISCUSSION

Our findings show clear evidence that the genetic differentiation among *H. speciosa* populations in neutral markers is mainly due to demographical history and isolation by distance, but not isolation by environment. The high admixture among Monachino's four varieties, especially among spatially closer populations in central Brazil, indicates that the genetic variation among populations does not necessarily match the morphological variation and thus the assignment of individuals to the four varieties proposed by Monachino (1945). Genetically, populations from the western Cerrado were clustered together, independent of taxonomy (Fig. 1), such as HCTAS, HCDIA, HCCHG and HCGCA from *H. speciosa* var. *cuyabensis* and HGPOT from *H. speciosa* var. *gardneri*. Similarly, populations from the south and south-west Cerrado were clustered together, such as HCCAG from *H. speciosa* var. *cuyabensis*, and HGSEL, HGFRO, HGPOR and HGMCA from *H. speciosa* var. *gardneri*. Morphologically, plant architecture (more closed canopy) and leaf size (slightly smaller) in populations of *H. speciosa* var. *gardneri* from the south and south-west Cerrado are slightly different from those of central Brazil (pers. obs.). Likewise, population HSPIP has morphological characteristics similar to *H. speciosa* var. *speciosa*, but with slightly larger leaves, and is genetically more similar to population HGBRA, that has morphological characteristics of *H. speciosa* var. *gardneri*. Thus, we hypothesize that the eastern Cerrado may be a contact zone between *H. speciosa* var. *gardneri* and *H. speciosa* var. *speciosa*.

Indeed, cross-pollination between these two varieties and similar fitness of hybrids compared with non-hybrids were observed in a previous study (Collevatti et al., 2016). Therefore, populations HSPIP and HGBRA may have hybrid individuals between the two varieties. On the other hand, although some individuals showed low admixture with other varieties, populations HSROV and HSJAP were clustered together in C3, a

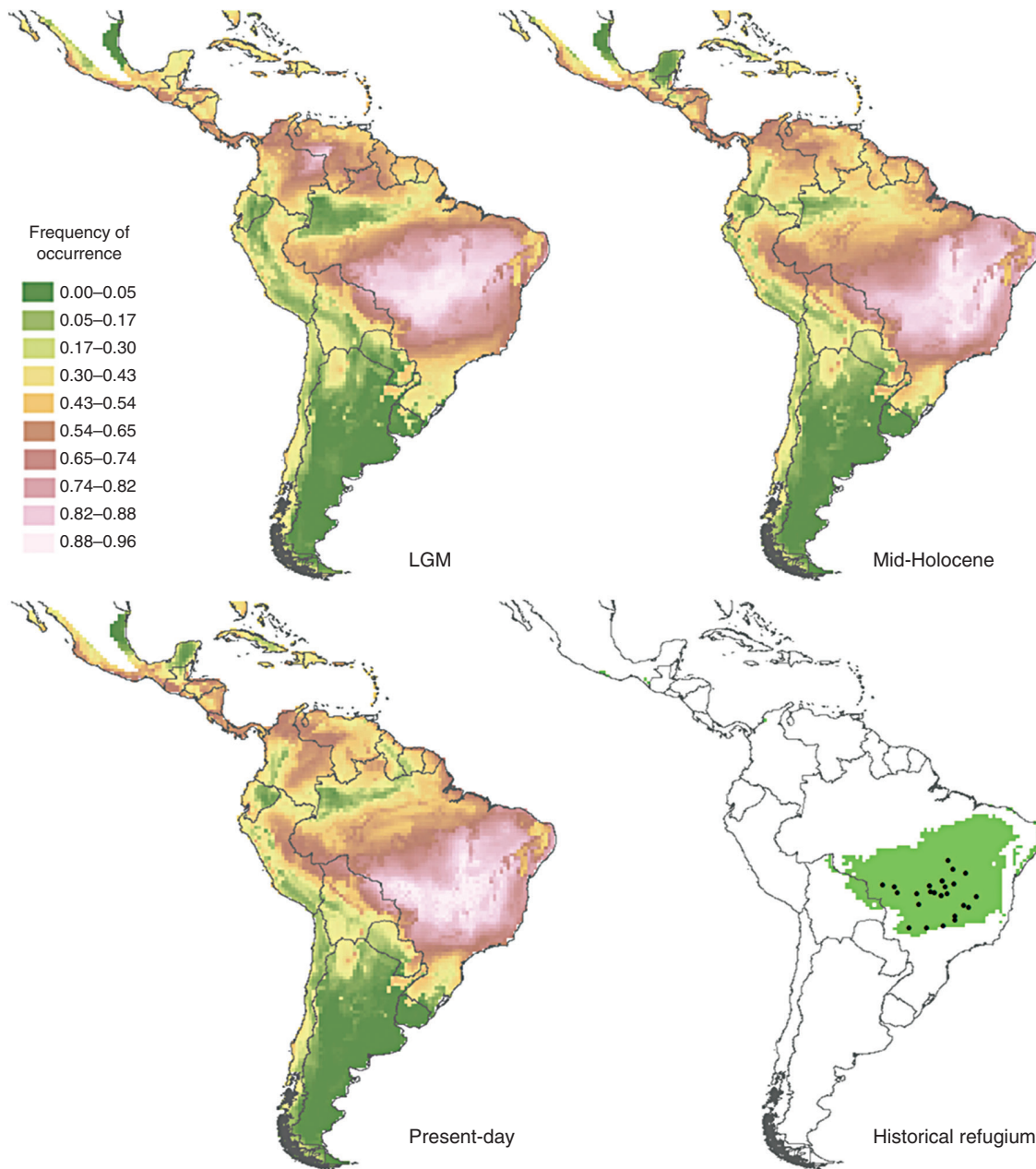


FIG. 2. Environmental suitability for *Hancornia speciosa* in the Neotropics expressed by the ensemble of 50 ENMs for the LGM (21 ka), mid-Holocene (6 ka), the present and the historical refugium through time (suitability >0.5).

cluster comprising only individuals that match the morphology of *H. speciosa* var. *speciosa*. Hybridization and similar fitness between hybrids and non-hybrids (Collevatti *et al.*, 2016) may also explain the similarity of individuals from sympatric populations of *H. speciosa* var. *gardneri* and *H. speciosa* var. *pubescens*, and between parapatric populations of *H. speciosa* var. *gardneri* and *H. speciosa* var. *cuyabensis*.

Therefore, although we recovered a continuum of genetic variation correlated with geographical distance, our findings support four different clusters that match at least the sampled morphological variation corresponding partially to the botanical varieties proposed by Monachino (1945): a cluster

of *H. speciosa* var. *speciosa* (cluster 3, Fig. 1) corresponding to populations from the north-eastern Cerrado (HSROV and HSJAP); a cluster of *H. speciosa* var. *gardneri* + *H. speciosa* var. *pubescens* (clusters 1 and 4, Fig. 1) corresponding to populations from the central Cerrado; *H. speciosa* var. *cuyabensis* + *H. speciosa* var. *gardneri* (cluster 5, Fig. 1), corresponding to populations of the western Cerrado (HCTAS, HCDIA, HCCHG, HCGCA, HGPOT); and a cluster of *H. speciosa* var. *cuyabensis* + *H. speciosa* var. *gardneri* (cluster 2, Fig. 1), corresponding to populations from the southern and south-western Cerrado (HCCAG, HGSEL, HGFRO, HGPOR and HGMCA). These four clusters were also supported by hierarchical AMOVA that

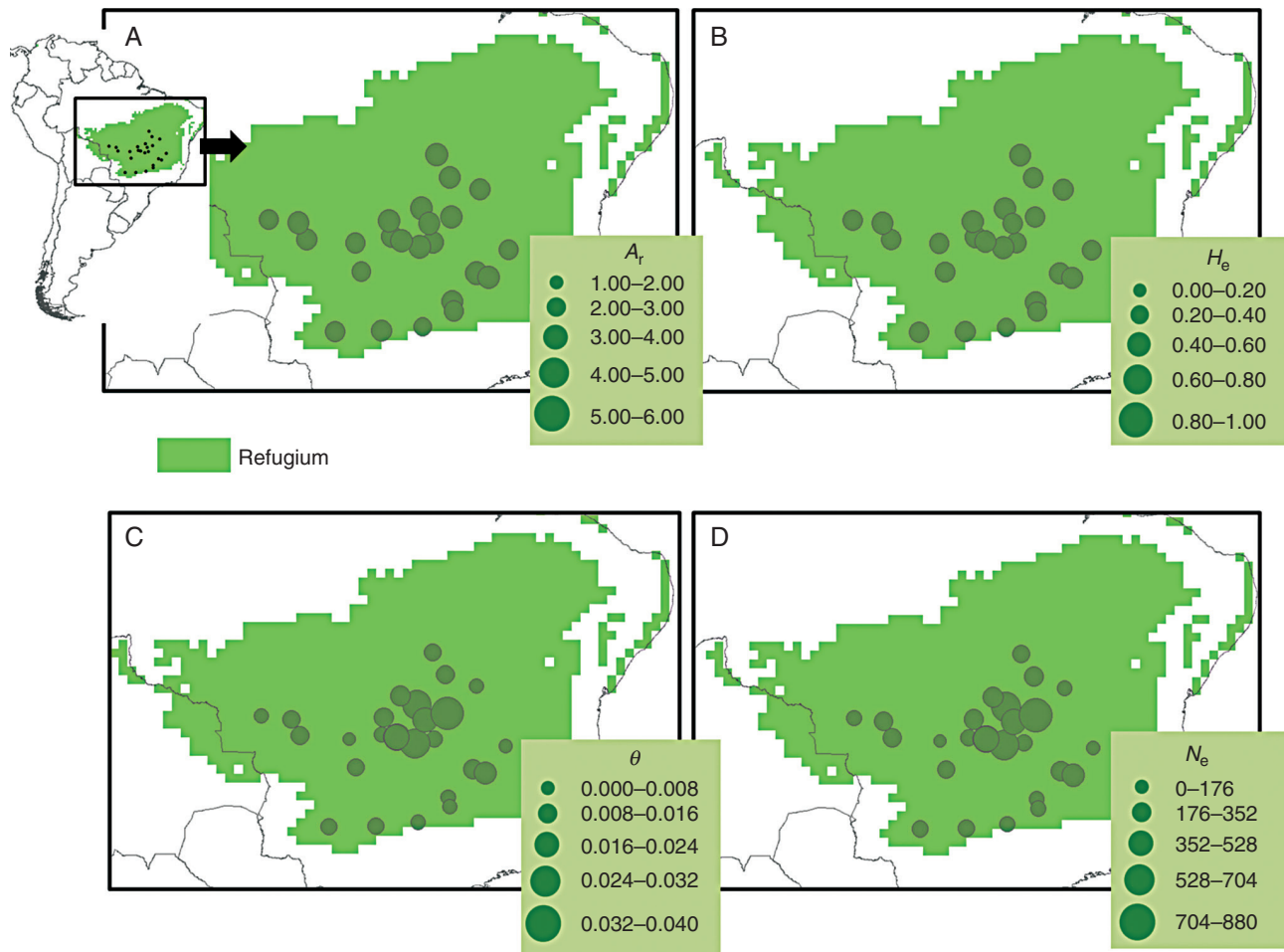


FIG. 3. Spatial distribution of genetic diversity of *Hancornia speciosa* in relation to the historical refugium, i.e. areas climatically suitable throughout the time (in green). (A) Allelic richness (A_r). (B) Genetic diversity (H_e). (C) θ . (D) Effective population size (N_e). Circumference sizes are proportional to the value of genetic parameter, according to the keys.

showed higher differentiation among the four genetic clusters than among Monachino's four varieties, meaning that the co-ancestry among individuals within our four Bayesian clusters is higher than among individuals within Monachino's varieties.

Moreover, our data do not support the classification in two varieties, *H. speciosa* var. *speciosa* and *H. speciosa* var. *pubescens* (Flora do Brasil 2020, available at <http://floradobrasil.jbrj.gov.br/reflora/floradobrasil/>), due to the high admixture between these two varieties and because individuals from *H. speciosa* var. *pubescens* were clustered with *H. speciosa* var. *gardneri*, mainly in clusters 1 and 4. In addition, coalescent analyses showed high historical gene flow among populations of both varieties and sympatric populations. Thus morphological variation may be due to quantitative variation and may have no effect on individual fitness.

Thus, our results suggest that morphological variation among Monachino's varieties of *H. speciosa* are most likely to be due to selection in loci controlling morphological characteristics (see Chaves, 2006). Both the cross-pollination among varieties shown in a previous study (see Collevatti et al., 2016) and the admixture revealed by our findings suggest no reproductive isolation. In fact, populations of recently diverged species

or varieties may have different levels of reproductive isolation and gene flow (Riesenberg et al., 2004). The low differentiation among Monachino's varieties ($F_{CT} = 0.027$) may be due to the high historical effective population size counterbalancing the effect of genetic drift and the recency of TMRCA. In fact, hierarchical ANOVA in allele frequency showed that genetic differentiation of *H. speciosa* is due to non-random mating among Monachino's varieties. Moreover, the variation among structure groups ($F_{CT} = 0.058$) was much higher than among Monachino's varieties, showing that our clustering supports more homogenous groups.

We also found no evidence of isolation by environment when climatic variables related to temperature, precipitation and climatic suitability were analysed, suggesting no differentiation due to local adaptation to climate. Indeed, hybrid offspring show no heterosis or exogamic depression in a common garden experiment (Collevatti et al., 2016), but it is important to note that our analysis of isolation by environment was performed at a coarse spatial resolution (climate layers with $0.5^\circ \times 0.5^\circ$ latitude \times longitude). Thus, we cannot absolutely discard local adaptation to microclimate or to another unknown environmental variable. However, the sympatric

coexistence of different varieties, such as *H. speciosa* var. *gardneri* and *H. speciosa* var. *pubescens*, and the coexistence of populations with intermediate morphological traits, such as HSPIP, suggest no reproductive barrier among varieties but that natural selection may lead to morphological differentiation with maintenance of hybrid zones where hybrids may be as fit as the non-hybrids (Hewitt, 1988; Barton and Hewitt, 1989). The non-random mating within and among populations may facilitate the fixation of alleles leading to morphological differentiation.

Despite the high differentiation in morphological traits, the low genetic differentiation in neutral markers among varieties may also be due to the low length of time since divergence of the lineages. The recent time to the most common ancestor of approx. 56.218 ka demonstrates a very recent lineage differentiation. The evolution in quantitative morphological traits may be faster than that in neutral markers (McKay and Latta, 2002); thus, despite morphological differentiation, it is likely that time since divergence was not long enough to accumulate high divergence in neutral microsatellite markers.

Hancornia speciosa show quasi-stability in geographical range through time, which may have allowed a high connectivity among populations, the demographical stability and the high effective population size. In addition, all populations showed no evidence of contraction (growth parameter). Thus, the high genetic diversity may be the result of demographical and range stability through time. The high historical connectivity, mainly between closer populations, together with the large effective population size may also have hindered the differentiation among varieties, leading to an isolation by distance pattern of differentiation, but no isolation by environment based on neutral markers. Moreover, we found evidence of historical connectivity among populations of *H. speciosa* but mainly among populations of central Brazil, despite the variety, explaining the high genetic differentiation (F_{SC}) among populations. In fact, all populations of *H. speciosa* analysed in this work were always in areas of high suitability over the last 21 ka (≥ 0.5 , see Fig. 3). The connectivity among closer population may have led to the mixed ancestry of individuals from geographically closer localities revealed by Bayesian clustering. Our results also indicate no recent gene flow among populations because no individual was assigned to a different population. *Hancornia speciosa* is self-incompatible (Darrault and Schlindwein, 2005; Collevatti et al., 2016), and high proportions of biparental inbreeding have been observed (Collevatti et al., 2016), which may lead to high inbreeding within population and a high fixation index (F_{IS}). The scenario of quasi-stability in geographical range (slight range retraction in the LGM) and high connectivity among populations was also seen for other savanna tree species, such as *Dimorphandra mollis* (Souza et al., 2016), *Eugenia dysenterica* (Lima et al., 2017) and *Handroanthus ochraceus* (Vitorino et al., 2018). These species also showed a slight retraction in the south-eastern Cerrado during the LGM, which might explain the high differentiation of populations from this region. The scenario of demographical expansion through time and evidence of colonization of south-eastern Brazilian savannas during warmer interglacial periods of the Quaternary (see Fig. 2) was also supported for other Brazilian savanna species (e.g. Novaes et al., 2010, 2013;

Collevatti et al. 2012b, 2013b; Lima et al., 2014). Indeed, populations of *H. speciosa* var. *gardneri* from the south-eastern Cerrado were clustered in a different group (cluster 2, Fig. 1), and tended to show high genetic differentiation (Supplementary Data Table S8) and a low number of migrants per generation (Supplementary Data Table S9) from populations of the central Cerrado.

In conclusion, our results strongly suggest that the pattern of genetic differentiation among populations of *H. speciosa* is mainly due to demographical history and isolation by distance. The range stability through time leading to a wide historical climatic refugium allowed high connectivity among populations and high effective population sizes, thus maintaining a relatively high level of genetic diversity and allelic richness in most populations. The high historical connectivity may also have caused high admixture among populations, leading to a spatial continuum of genetic variation due to isolation by distance. Despite that, our findings based on neutral molecular markers support four different groups of *H. speciosa*, but do not agree with the recent classification into only two botanic varieties (Flora do Brasil 2020). The four groups found in the present work partially support Monachino's varieties, with modifications to better embody both the genetic and morphological variation. Additionally, the present study reinforces the usefulness of the framework coupling coalescent simulation and ENM (see Collevatti et al., 2012a, 2013a, 2015b) for understanding the mechanisms involved in the origin and maintenance of genetic diversity of species.

SUPPLEMENTARY DATA

Supplementary data are available online at <https://academic.oup.com/aob> and consist of the following. Tables S1–S9: populations sampled, results of structure analyses and coalescent analysis. Figures S1–S12: structure analyses, ecological niche modelling and quantile regressions.

ACKNOWLEDGEMENTS

This work was supported by grants to the research network GECER (Núcleo de Excelência em Genética e Conservação de Espécies do Cerrado), supported by FAPEG (Goias Foundation for Research) and CNPq (Brazilian Council of Science and Technology) (PRONEX/FAPEG/CNPq project no. CH07-2009), the network GENPAC ('Geographical genetics and regional planning for natural resources in Brazilian Cerrado', CNPq/MCT/CAPES/FAPEG projects nos 564717/2010-0, 563727/2010-1 and 563624/2010-8) and the network Rede Cerrado PPBio (CNPq project no. 457406/2012-7). Fieldwork was supported by Systema Naturae for Environmental Consulting. E.E.R. received a scholarship from CAPES. R.G.C., M.P.C.T. and L.J.C. have continuously been supported by CNPq grants and scholarships whose assistance we gratefully acknowledge. Current research is funded by the National Institute for Science and Technology (INCT) in Ecology, Evolution and Biodiversity Conservation (MCTIC/CNPq/FAPEG project no. 465610/2014-5). We also acknowledge the helpful comments of two anonymous reviewers.

LITERATURE CITED

- 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.
- Arenas M, Ray N, Currat M, Excoffier L. 2012. Consequences of range contractions and range shifts on molecular diversity. *Molecular Biology and Evolution* 29: 207–218.
- Barton NH, Hewitt GM. 1989. Adaptation, speciation and hybrid zones. *Nature* 341: 497–503.
- Buzatti RSO, Lemos-Filho JP, Bueno ML, Lovato MB. 2017. Multiple Pleistocene refugia in the Brazilian cerrado: evidence from phylogeography and climatic niche modelling of two *Qualea* species (Vochysiaceae). *Biological Journal of the Linnean Society* 185: 307–320.
- Cade BS, Noon BR. 2003. A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment* 1: 412–420.
- Chaves LJ. 2006. Recursos genéticos no Cerrado. In: Silva-Junior JF, Ledo AS, eds. *A cultura da mangaba*. Aracaju: Embrapa Tabuleiros Costeiros, 75–84.
- Coelho ASG, Valva FDO. 2001. Processo evolutivo e o melhoramento de plantas. In: Nass LL, Valois ACC, Melo IS, Valadares-Inlis MC, eds. *Recursos genéticos e melhoramento de plantas*. Rondonópolis: Fundação MT, 58–78.
- Collevatti RG, Terribile LC, Lima-Ribeiro MS, et al. 2012a. 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, Oliveira G, Terribile LC. 2012b. Recovering the demographical history of a Brazilian cerrado tree species *Caryocar brasiliense*: coupling ecological niche modeling and coalescent analyses. *Natureza & Conservação* 10: 169–176.
- Collevatti RG, Terribile LC, Oliveira G, et al. 2013a. Drawbacks to palaeo-distribution 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. 2013b. Demographic history and the low genetic diversity in *Dipteryx alata* (Fabaceae) from Brazilian Neotropical savannas. *Heredity* 111: 97–105.
- Collevatti RG, Terribile LC, Rabelo SG, Lima-Ribeiro MS. 2015a. Relaxed random walk model coupled with ecological niche modeling unravel the dispersal dynamics of a Neotropical savanna tree species in the deeper Quaternary. *Frontiers in Plant Science* 6: e653.
- Collevatti RG, Terribile LC, Diniz-Filho JAF, Lima-Ribeiro MS. 2015b. Multi-model inference in comparative phylogeography: an integrative approach based on multiple lines of evidence. *Frontiers in Genetics* 6: 31.
- Collevatti RG, Olivatti AM, Telles MPC, Chaves LJ. 2016. Gene flow among *Hancornia speciosa* (Apocynaceae) varieties and hybrid fitness. *Tree Genetics & Genomes* 12: 74.
- Darrault RO, Schlindwein C. 2005. Limited fruit production in *Hancornia speciosa* (Apocynaceae) and pollination by nocturnal and diurnal insects. *Biotropica* 37: 381–388.
- Diniz-Filho JAF, Bini ML, Rangel FT, et al. 2009. Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography* 32: 897–906.
- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.
- Earl DA, von Holdt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359–361.
- 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, Ray N. 2008. Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology and Evolution* 23: 347–351.
- 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.
- Excoffier L, Foll M, Petit R. 2009. Genetic consequences of range expansions. *Annual Review of Ecology and Systematics* 40: 481–501.
- FAO/IIASA/ISRIC/ISS-CAS/JRC. 2009. *Harmonized world soil database (version 1.1)*. Food and Agriculture Organization of the United Nations (FAO), Rome, and International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria. Available at: http://www.fao.org/fileadmin/templates/nr/documents/HWSD/HWSD_Documentation.pdf.
- Fielding AH, Bell JF. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24: 38–49.
- Goudet J. 2002. *FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3.2)*. <http://www.unil.ch/izea/software/fstat.html>
- Hein J, Schierup M, Wiuf C. 2005. *Gene genealogies, variation and evolution: a primer in coalescent theory*. Oxford: Oxford University Press.
- Hewitt GM. 1988. Hybrids zones – natural laboratories for evolutionary studies. *Trends in Ecology and Evolution* 3: 158–167.
- Hewitt GM. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 58: 247–276.
- Hoffmann AA, Shirriffs J, Scott M. 2005. Relative importance of plastic vs genetic factors in adaptive differentiation: geographical variation for stress resistance in *Drosophila melanogaster* from eastern Australia. *Functional Ecology* 19: 222–227.
- Jakobsson M, Rosenzberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 14: 1801–1806.
- Kingman JFC. 1982. The coalescent. *Stochastic Processes and their Applications* 13: 235–248.
- Knowles LL, Carstens BC, Keat ML. 2007. Coupled genetic and ecological-niche models to examine how past population distributions contribute to divergence. *Current Biology* 17: 1–7.
- Kuhner MK. 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* 22: 768–770.
- Kuhner M, Smith L. 2007. Comparing likelihood and Bayesian coalescent estimation of population parameters. *Genetics* 175: 155–165.
- Leinonen T, O'Hara RB, Cano JM, Merilä J. 2008. Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. *Journal of Evolutionary Biology* 21: 1–17.
- Lima NE, Lima-Ribeiro MS, Tinoco CF, Terribile LC, Collevatti RG. 2014. Phylogeography and ecological niche modelling, coupled with the fossil pollen record, unravel the demographic history of a Neotropical swamp palm through the Quaternary. *Journal of Biogeography* 41: 673–686.
- Lima JS, Telles MPC, Chaves JL, Lima-Ribeiro MS, Collevatti RG. 2017. Demographic stability and high historical connectivity explain the diversity of a savanna tree species in the Quaternary. *Annals of Botany* 119: 645–657.
- Lima-Ribeiro MS, Varela S, Gonzalez-Hernandez J, Oliveira G, Diniz-Filho JAF, Terribile LC. 2015. ecoClimate: a database of climate data from multiple models for past, present, and future for macroecologists and biogeographers. *Biodiversity Informatics* 10: 1–21.
- 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.
- McKay JK, Latta RG. 2002. Adaptive population divergence: markers, QTL and traits. *Trends in Ecology and Evolution* 17: 285–291.
- Mitchell-Olds T, Schmitt J. 2006. Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis*. *Nature* 22: 947–952.
- Monachino J. 1945. A revision of *Hancornia* (Apocynaceae). *Lilloa* 11: 19–48.
- 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. *Theoretical and Applied Genetics* 92: 832–839.
- Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583–590.
- Novaes RML, Lemos-Filho JP, Ribeiro RA, Lovato MB. 2010. Phylogeography of *Plathymentia 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 phylogeographical and biogeographical patterns in the Brazilian Cerrado: diversification of the endemic tree *Dalbergia miscolobium* (Fabaceae). *PLoS One* 8: e82198.
- van Oosterhout CV, 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.
- Pritchard J, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Rambaut A, Drummond AJ. 2007. *Tracer v1.6*. Available at: <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.
- Rieseberg LH, Church SA, Morjan CL. 2004. Integration of populations and differentiation of species. *New Phytologists* **161**: 59–69.
- Rodrigues AJL, Yamaguishi AT, Chaves LJ, Coelho ASG, Lima JS, Telles MPC. 2015. Development of microsatellites markers for *Hancornia speciosa* Gomes (Apocynaceae). *Genetics and Molecular Research* **14**: 7274–7278.
- Sexton JP, Hangartner SB, Hoffmann AA. 2014. Genetic isolation by environment or distance; which pattern of gene flow is most common? *Evolution* **68**: 1–15.
- Souza HAV, Collevatti RG, Lima-Ribeiro MS, Lemos-Filho JP, Lovato MB. 2016. A large historical refugium explains spatial patterns of genetic diversity in a Neotropical savanna tree species. *Annals of Botany* **119**: 239–252.
- Stokland JN, Halvorsen R, Stoa B. 2011. Species distribution modeling – effect of design and sample size of pseudo-absence observations. *Ecological Modelling* **222**: 1800–1809.
- Terribile LC, Lima-Ribeiro MS, Araujo MB, et al. 2012. Areas of climate stability in the Brazilian Cerrado, disentangling uncertainties through time. *Natureza & Conservação* **10**: 152–159.
- 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.
- Udupa S, Baum M. 2001. High mutation rate and mutational bias at (TAA)_n microsatellite loci in chickpea (*Cicer arietinum* L.). *Molecular Genetics and Genomics* **265**: 1097–1103.
- Vitorino LC, Lima-Ribeiro MS, Terribile LC, Collevatti RG. 2018. Demographical expansion of *Handroanthus ochraceus* in the Cerrado during the Quaternary: implications for the genetic diversity of Neotropical trees. *Biological Journal of the Linnean Society* **123**: 561–577.
- Wang IJ. 2013. Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution* **16**: 175–182.
- Wang IJ, Bradburd GS. 2014. Isolation by environment. *Molecular Ecology* **23**: 5649–5662.
- Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.
- Wright S. 1943. Isolation by distance. *Genetics* **28**: 114–138.
- Wright S. 1951. The genetic structure of populations. *Annual Eugenics* **15**: 323–354.