

Demographical expansion of *Handroanthus ochraceus* in the Cerrado during the Quaternary: implications for the genetic diversity of Neotropical trees

LUCIANA CRISTINA VITORINO¹, MATHEUS S. LIMA-RIBEIRO²,
LEVI CARINA TERRIBILE² and ROSANE G. COLLEVATTI^{1*}

¹Laboratório de Genética & Biodiversidade, Instituto de Ciências Biológicas (ICB), Universidade Federal de Goiás (UFG), Goiânia, GO 74001-970, Brazil

²Laboratório de Macroecologia, Instituto de Biociências, Universidade Federal de Goiás (UFG), Regional Jataí, Cx.P. 03, Jataí, GO 75801-615, Brazil

Received 23 November 2017; revised 12 December 2017; accepted for publication 12 December 2017

Phylogeographical studies provide insights into the dispersal dynamics of species needed to understand the effects of Quaternary climate changes on the spatial patterns of genetic diversity. We used a multi-model inference approach coupling ecological niche modelling (ENM) with a relaxed random walk model to reconstruct the spatio-temporal history of lineage dispersal of the Neotropical tree species *Handroanthus ochraceus*. We sampled 24 populations throughout the Cerrado Biome and analysed polymorphisms at three intergenic chloroplast regions and ITS. Coalescent analyses revealed demographical expansion since *c.* 380 ka. Although ENM predicted no range expansion, coalescent simulations reinforce the pattern of range expansion because demographical expansion was the most likely scenario able to produce the observed spatial pattern of genetic diversity of *H. ochraceus*. Its most recent common ancestor dated from ~1.9 Ma, and lineages cyclically dispersed from the Southeast and West towards Central–West Brazil. Most dispersal events occurred from populations at the edges of the Cerrado towards Central Brazil. Populations at the edge of the historical refugium show higher genetic diversity and were the source of migrants to central populations. A wide historical climatic refugium through time for *H. ochraceus* may have allowed dispersal of lineages among populations of Central Brazil, maintaining their historical connectivity and genetic diversity.

ADDITIONAL KEYWORDS: Bignoniaceae – biogeography – Cerrado – coalescence – ecological niche modelling – Neotropical savanna – phylogeography – Quaternary climate changes – random walk model.

INTRODUCTION

Neotropical vegetation was affected by Palaeogene–Neogene events that might have driven demographical dynamics of species from savannas and adjacent rainforests (Pennington *et al.*, 2004; Hoorn *et al.*, 2010; Rull, 2011). For instance, the Andean orogeny in the Palaeogene increased the latitudinal temperature gradient in South America, and uplift of the Central Brazilian Plateau in the Miocene–Pliocene transition caused the geographical reorganization of the Central Brazil ecosystems that may have favoured species diversification by vicariance (e.g. Luebert & Weigend, 2014; Luebert & Muller, 2015). Quaternary climate

oscillations also affected species distributions and diversity through space and time in the Neotropics (Rull *et al.*, 2013; Rull, Vegas-Vilarrúbia & Montoya, 2015). However, due to differences in climatic tolerances and niche characteristics, different species responded individually to the common drier and colder conditions of glacial periods (Cheng *et al.*, 2013; Summerhayes & Charman, 2014).

The pollen fossil record suggests that south-eastern Brazil was cooler and drier during the glacial periods than it is at present, and savannas shifted in distribution northward due to drier and colder conditions (Behling, 2002, 2003). The grasslands and subtropical Araucaria forest spread northward, replacing the Atlantic Forest and savannas to latitudes up to 20°S and retreating to the south only after the glaciation (Salgado-Labouriau,

*Corresponding author: E-mail: rosanegc68@hotmail.com.

1997; Behling, 2003). Modern wet climatic conditions in central Brazil, with short dry periods, and in all main Neotropical biomes became established during the late Holocene (Behling, 2003). Phylogeographical analyses of Neotropical species show the effect of Quaternary climatic change, with demographical expansion of some seasonally dry tropical forest species (e.g. Caetano *et al.*, 2008; Collevatti *et al.*, 2012a, 2016; Vieira *et al.*, 2015), but retraction of others (e.g. Melo *et al.*, 2016). Savanna species also show contrasting phylogeographical patterns in the Neotropics, with demographical retraction during the Last Glacial Maximum (LGM) with multiple refugia (e.g. Collevatti *et al.*, 2012b, 2015a; Lima *et al.*, 2014), or stability over time (e.g. Souza *et al.*, 2017; Lima *et al.*, 2017). Nevertheless, the response of Neotropical plant species to Quaternary climate changes remains poorly understood due to the low number of studies explicitly testing it (e.g. Collevatti *et al.*, 2012b, 2015a; Lima *et al.*, 2014, 2017; Souza *et al.*, 2017). Moreover, the wide variation in response among species hinders a synthesis of the biogeography of the Neotropics and the effects of Quaternary climate changes in species distribution dynamics. Thus, study of the demographical history of plant species is important for understanding the role of Quaternary climate changes to unravel biogeographical patterns in Neotropical species, as we address here for a Neotropical savanna species.

Range expansion may lead to spatial genome sorting as the species tracks suitable environments, resulting in lower genetic diversity in new colonizing areas (Hewitt, 1996). In addition, the spread of a low-frequency allele that migrates on the wave of advance of a population in expansion (allele surfing; Excoffier & Ray, 2008; Arenas *et al.*, 2012), and density-dependent processes due to fast colonization and founder events may cause low genetic diversity in new colonizing areas (see Excoffier, Foll & Petit, 2009; Waters, Fraser & Hewitt, 2013). Reconstructing historical demography is essential to elucidate the evolution and divergence of tropical plant lineages (Knowles & Maddison, 2002). For the Neotropics, this is often compromised because of the lack of fossil records for most species (but see an example in Lima *et al.*, 2014 for a swamp palm species). Thus, integrating direct spatio-temporal reconstruction of lineage diffusion (Lemey *et al.*, 2009, 2010) in a multi-model framework (see Collevatti *et al.*, 2015a, b) may give clues to the pathways of lineage dispersal across the Neotropics through the Quaternary.

Here, we reconstruct the demographical history and dispersal dynamics of *Handroanthus ochraceus* (Cham.) Standl. (Bignoniaceae) through the Quaternary using a multi-model inference approach following Collevatti

et al. (2015a, b) to understand how the Quaternary climate changes affected the spatial pattern in genetic diversity. Because Neotropical savanna species showed range retraction during the LGM followed by further expansion with multiple refugia we expect that *H. ochraceus* will show a pattern of range and demographical retraction during glacial phases and range expansion from the LGM to the present day. Because of population isolation in refugia we expect low historical genetic connectivity among populations leading to high population differentiation. We also hypothesize that the south-eastern and western Cerrado were important refugia for savanna species, as shown for a related species, *Tabebuia aurea* (Collevatti *et al.*, 2015a), and suggested by palaeovegetation reconstruction (Ab'Saber, 2000). Thus, *H. ochraceus* will have the dispersal origin of extant lineages in the south-eastern and western Cerrado.

MATERIAL AND METHODS

POPULATION SAMPLING AND DNA EXTRACTION

Handroanthus ochraceus (syn. *Tabebuia ochracea*) is a widespread savanna tree species, occurring in seasonal savannas and seasonally dry forests in Brazil, Paraguay, northern Argentina, Bolivia, Peru, Colombia, Venezuela and Mesoamerica. Locally it is distributed in well-delimited patches at high density. The species is hermaphroditic and pollinated mainly by large bees such as bumblebees (*Bombus* spp.), carpenter bees (*Xylocopa* spp.) and *Centris* spp., and its winged seeds are wind-dispersed (Gibbs & Bianchi, 1993).

We extensively sampled *H. ochraceus* throughout the Brazilian savanna (Fig. 1; see also Supporting Information Table S1 in Appendix S1 and Fig. S1 in Appendix S2) and collected 468 individuals from 24 populations. We focused our sampling efforts on the Cerrado biome, especially in savannas of Central–West Brazil, due to the higher abundance of the species there and our focus on savanna biogeography. In addition, although occurrence maps (Fig. S1) show records of *H. ochraceus* in north-eastern Brazil in the Caatinga biome and in the south-east in the Atlantic Forest biome, populations in these regions are rare. North-east Brazil is dominated by xeromorphic steppe vegetation with seasonally dry forests and some islands of savanna where *H. ochraceus* may occur, but the high level of anthropic disturbance has removed most natural vegetation. In the Atlantic Forest of south-east Brazil most savanna enclaves disappeared due to agriculture and urban expansion. We sampled only adult individuals to avoid the effects of high kinship on estimation of genetic parameters. Vouchers

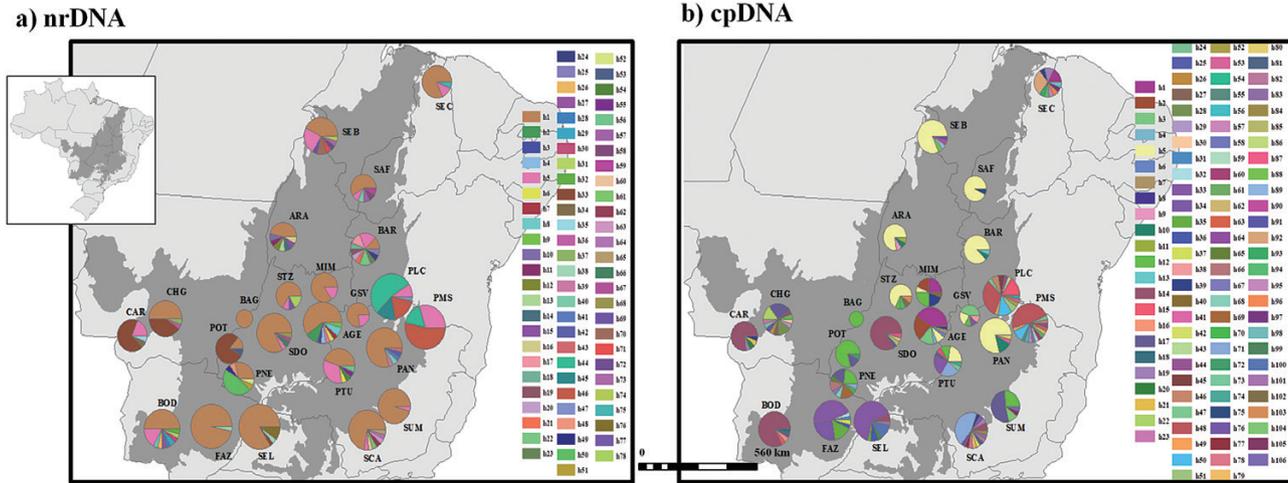


Figure 1. Geographical distribution of haplotypes of *Handroanthus ochraceus* for (A) ITS and (B) cpDNA, based on the sequencing of 468 individuals from 24 populations from the Brazilian savanna. Different colours were assigned for each haplotype according to the figure legend. The circle size represents the sample size in each population and the circle sections represent the haplotype frequency in each sampled population. For details on population codes and localities see Table S1 in Appendix S1.

collected were compared and deposited in herbarium UFG (Universidade Federal de Goiás, exsiccate J.A Rizzo 4471, UFG 4491). The distance between population pairs ranged from 20 km to 2393 km. In populations with fewer than 30 individuals we sampled and sequenced all of them (Table 1). We also sampled and sequenced individuals of *Handroanthus impetiginosus* (exsiccate N.R. Cunha 221, UFG 27029), *Handroanthus serratifolius* (exsiccate J.M. Rezende 535, UFG 24877) and *Cybistax antisiphylitica* (exsiccate L.F. Souza 3227, UFG 1020) to include as outgroups in order to estimate the divergence time to the most recent common ancestor (Table S1). Genomic DNA was extracted from leaves or cambium tissues following the standard 2% CTAB protocol (Doyle & Doyle, 1990).

GENETIC DATA

To generate genetic data, we sequenced three intergenic spacers of chloroplast DNA (cpDNA): *psbA-trnH*, *trnC-ycf6* and *trnS-trnG* (Shaw *et al.*, 2005) and the region ITS1 + 5.8S + ITS2 (hereafter ITS) from nuclear ribosomal DNA (nrDNA) (Desfeux & Lejeune, 1996). PCR conditions and amplifications followed Collevatti *et al.* (2012a). PCR fragments were sequenced using the BigDye Terminator v3.1 kit (Applied Biosystems), and sequenced in the forward and reverse directions using a GS 3500 Genetic Analyzer (Applied Biosystems), according to the manufacturer's instructions.

Sequences were analysed, and consensus sequences were obtained using the software SeqScape v2.7 (Applied Biosystems) and aligned with the software

ClustalΩ (Sievers *et al.*, 2011). Polymorphisms at mononucleotide microsatellites were excluded due to ambiguous alignment and higher mutation rates. Long indels (usually > 5 bp) were coded as one evolutionary event (one character).

GENETIC DIVERSITY AND POPULATION STRUCTURE

We estimated nucleotide (π) and haplotype (h) diversity for each population and the populations overall following Nei (1987) using the software ArlequinVer 3.11 (Excoffier, Laval & Schneider, 2005a). To test for population differentiation, we performed an analysis of molecular variance (AMOVA, Excoffier, Smouse & Quattro, 1992) and estimated F_{ST} values. Analyses were performed using the software Arlequin v3.11 and statistical significance was tested via a non-parametric permutation test (10 000 permutations). Population structure was also assessed using Bayesian clustering implemented in the software BAPS v6.0 (Corander & Marttinen, 2006; Corander *et al.*, 2008). The cpDNA and ITS partitions were analysed separately. We performed population admixture analysis based on mixture clustering with 'not fixed number of clusters (K)' with an upper limit of $K = 24$.

PHYLOGEOGRAPHICAL RECONSTRUCTION

Population demography

All coalescent analyses were performed with cpDNA and nrDNA ITS partitions concatenated, but giving separate priors. No evidence of heterozygous individuals was found when ITS sequences were

Table 1. Genetic diversity based on Arlequin v3.11 software and demographical parameters based on coalescent analysis performed with Lamarc 2.1.9 software for 24 populations of *Handroanthus ochraceus*

Population	N	k	cpDNA		SD	k	ITS	π	SD	Combined cpDNA and ITS data			
			h	π						h	θ	95%CI	Ne
AGE	27	8	0.781	0.0014	0.0009	9	0.618	0.0023	0.0016	0.0017	0.0004-1.0497	6.9E04	1.8E04-4.3E07
ARA	13	4	0.423	0.0012	0.0008	8	0.808	0.0068	0.0066	0.0077	0.0042-2.8443	3.1E05	1.7E05-1.2E08
BAG	05	1	0.000	0.0000	0.0000	1	0.000	0.0000	0.0000	0.0000	0.0000-0.0004	1.8E03	4.1E02-1.6E04
BAR	16	3	0.242	0.0001	0.0002	11	0.961	0.0019	0.0032	0.0090	0.0000-1.2457	3.7E04	7.2E02-50515031.09
BOD	24	5	0.249	0.0002	0.0002	10	0.652	0.0749	0.0396	0.0006	0.0001-5.3013	2.5E04	5.0E03-2.2E08
CAR	16	4	0.350	0.0010	0.0008	4	0.295	0.0022	0.0016	0.0006	0.0002-0.0013	2.5E04	8.8E03-5.5E04
CHG	20	12	0.916	0.0004	0.0004	4	0.616	0.0017	0.0013	0.0011	0.0006-0.0038	4.3E04	2.5E04-1.5E05
FAZ	31	7	0.723	0.0008	0.0005	2	0.064	0.0001	0.0003	0.0007	0.0002-0.0013	2.7E04	6.9E03-5.4E04
GSV	9	5	0.806	0.0009	0.0007	2	0.000	0.0000	0.0000	0.0004	0.0001-0.0015	1.7E04	3.9E03-6.2E04
MIM	13	5	0.808	0.0015	0.0010	2	0.000	0.0000	0.0000	0.0005	0.0001-0.0011	1.9E04	5.6E03-4.6E04
PAN	24	4	0.370	0.0002	0.0002	6	0.380	0.0013	0.0011	0.0005	0.0004-1.0401	2.1E04	1.5E04-4.2E07
PLC	32	13	0.846	0.0035	0.0020	5	0.603	0.0014	0.0011	0.0013	0.0008-0.0018	5.1E04	3.4E04-7.5E04
PMS	30	15	0.784	0.0023	0.0013	3	0.381	0.0007	0.0007	0.0022	0.0003-0.0043	8.9E04	1.2E04-1.8E05
PNE	17	9	0.904	0.0022	0.0014	6	0.721	0.0024	0.0018	0.0036	0.0013-0.0099	1.5E05	5.3E04-4.0E05
POT	14	3	0.385	0.0001	0.0002	3	0.385	0.0007	0.0008	0.0002	0.0001-0.0007	1.0E04	2.5E03-2.7E04
PTU	18	8	0.860	0.0034	0.0020	6	0.576	0.0023	0.0017	0.0021	0.0007-0.0054	8.5E04	3.0E04-2.2E05
SAF	12	2	0.167	0.0003	0.0003	6	0.618	0.0022	0.0035	0.0007	0.0004-2.9817	2.8E04	1.6E04-1.2E08
SCA	25	14	0.813	0.0042	0.0024	8	0.446	0.0018	0.0014	0.0032	0.0019-0.0077	1.3E05	7.7E05-3.1E06
SDO	23	5	0.324	0.0002	0.0002	5	0.260	0.0037	0.0024	0.0010	0.0004-0.0019	4.2E04	1.8E04-7.9E04
SEB	21	5	0.338	0.0003	0.0003	8	0.692	0.0071	0.0042	0.0012	0.0007-0.7750	4.7E04	2.8E04-3.1E07
SEC	16	10	0.897	0.0012	0.0013	3	0.143	0.0018	0.0014	0.0008	0.0003-0.0020	3.1E04	1.2E04-7.9E04
SEL	31	7	0.574	0.0012	0.0008	4	0.295	0.0006	0.0007	0.0006	0.0003-0.0018	2.6E04	1.1E04-7.4E04
STZ	12	5	0.576	0.0011	0.0008	4	0.473	0.0009	0.0009	0.0004	0.0001-0.0010	1.7E04	5.8E03-4.2E04
SUM	19	5	0.672	0.0001	0.0002	2	0.000	0.0000	0.0000	0.0001	0.0000-0.0009	4.3E03	1.6E03-3.8E04
Mean	-	6.6	0.575	-	-	5.1	0.416	-	-	-	-	-	-
SD	-	3.8	0.275	-	-	2.8	0.280	-	-	-	-	-	-
Overall	468	106	0.927	0.0044	0.0024	78	0.637	0.0063	0.0036	0.0401	0.0337-0.0430	1.6E06	1.4E06-1.7E06

N, sample size; k, number of haplotypes; h, haplotype diversity; π , nucleotide diversity; π , standard deviation; SD, standard deviation; θ , coalescent or mutation parameter; Ne, effective population size; 95% CI, 95% credibility interval.

analysed using SeqScape v2.7 (Applied Biosystems). Thus, recombination was neglected in all coalescent analyses. To set the priors, evolutionary model selection for both chloroplast and ITS regions was performed based on the Akaike Information Criterion implemented in the software jModelTest2 (Darriba *et al.*, 2012). For chloroplast regions, the model HKY+G+I ($-\ln L = 3625.8751$, wBIC = 0.5466) was selected, with gamma shape equal to 0.067, kappa equal to 1.95 and pinv = 0.404. For ITS, the evolutionary model TPM3uf+G ($-\ln L = 1755.8913$, wBIC = 0.3426) was selected with gamma shape equal to 0.558.

To study the dynamics in effective population size and genetic connectivity among populations, we estimated the mutation or coalescent parameter theta ($\theta = 2\mu N_e$ for haploid genomes, $\theta = 4\mu N_e$ for diploid genomes), and the immigration rate among all population pairs ($M = 2N_e m/\theta$ for haploid genomes, $M = 4N_e m/\theta$ for diploid genomes). Estimations were based on a Bayesian model using the Markov chain Monte Carlo (MCMC) approach implemented in Lamarc 2.1.10 software (Kuhner, 2006). Each analysis was 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 default settings for the initial estimate of theta. The program was run four times for each parameter to certify convergence among runs and validate the analyses using Tracer v1.6 (Rambaut & Drummond, 2007), and combined results were then generated using LogCombiner 1.8.2 (Drummond *et al.*, 2012). Results were considered when the effective sample size (ESS) was ≥ 200 (and when marginal posterior probability densities were unimodal and converged among runs. Effective population size was obtained from θ (Kingman, 1982) using a generation time of 15 years based on flowering time measured in permanent plots established for demography dynamics studies (R. G. Collevatti, unpublished data).

To understand changes in population size through time we performed a Coalescent Extended Bayesian Skyline Plot (EBSP) analysis (Heled & Drummond, 2008) implemented in BEAST 1.8.3 (Drummond & Rambaut, 2007), combining data from different partitions. We used the substitution models reported above and the relaxed molecular clock model (uncorrelated lognormal) for both chloroplast and ITS regions. The ucl.d.stdev parameter (standard deviation of the uncorrelated lognormal relaxed clock) and the coefficient of variation were inspected for among-branch rate heterogeneity within the data. In all runs ucl.d.stdev was greater than 1.25 and the coefficient of variation frequency histogram viewed in Tracer spanned zero (~ 1.6 – 2.5) showing heterogeneity among branches. Mutation rates for both chloroplast and ITS regions were the same as used for two taxonomically

related species, *Handroanthus impetiginosus* (Collevatti *et al.*, 2012a) and *Tabebuia aurea* (Collevatti *et al.*, 2015a). Four independent analyses were run for 30 million generations. Convergence and stationarity were checked, and the independent runs were combined. Results were considered when $ESS \geq 200$. We performed Fu's F_s tests of neutrality implemented in Arlequin 3.11. Negative F_s values indicate an excess of rare alleles or new mutations in the genealogy resulting from either population expansion or selective sweeps.

Lineage dispersal

To reconstruct the spatio-temporal history of *H. ochraceus* lineage dispersal, we used the relaxed random walk model (RRW; Lemey *et al.*, 2009, 2010) implemented in the software BEAST 1.8.3, which analyses molecular sequence evolution, the demographical model and lineage dispersal in space and time simultaneously (Lemey *et al.*, 2010). We used the two sequence partitions with unlinked priors but sharing the same location rate matrix (Lemey *et al.*, 2009, 2010). We used Bayesian Stochastic Search Variable Selection (BSSVS), which considers a limited number of rates (at least $k - 1$, $k = 24$ localities) to explain the phylogenetic diffusion process and a Symmetric Substitution Model that uses a standard continuous-time Markov chain (CTMC) in which the transition rates between locations are reversible. Priors for sequence evolution were the same as described above for EBSP. For tree priors, we used the Coalescent GMRF Bayesian Skyride model (Minin, Bloomquist & Suchard, 2008) and for location state rate the prior CTMC Rate Reference (Ferreira & Suchard, 2008). Four runs were performed with 30 million generations, stability was analysed using Tracer 1.6 and results were considered when $ESS \geq 200$. The annotated tree (maximum clade credibility tree) was generated with 50% burnin and the spatio-temporal reconstruction was performed using SPREAD 1.0.6 (Bielejec *et al.*, 2011). We also analysed the well-supported transition rates using the Bayes factor (BF) test implemented in SPREAD. Transition rates between localities were considered only for $BF > 8.0$. We ran the analysis for each sequence region, cpDNA and ITS, to detect different contributions by seed and pollen dispersal.

Coalescent tree and time to most recent common ancestor

We obtained a coalescent tree based on Bayesian coalescent analysis implemented in the software BEAST 1.8.3 to estimate time to the most recent common ancestor (TMRCA). For this analysis, we included the outgroup sequences of *H. impetiginosus*,

H. serratifolius and *C. antisyphillitica*. We used the same priors of EBSP, except that for tree prior we used Coalescent Expansion Growth based on the results of EBSP (see Results below). To date the tree we used the same mutation rates used for EBSP, i.e. the same mutation used for the related species *Handroanthus impetiginosus* (Collevatti *et al.*, 2012a) and *Tabebuia aurea* (Collevatti *et al.*, 2015a). MCMC conditions and number of runs also remained unchanged. The independent runs were analysed using Tracer 1.6 and the results were considered when ESS \geq 200. We also ran an empty alignment (sampling only from priors) to verify the sensitivity of the results to the given priors. The analysis showed that our data were informative because posterior values (e.g. posterior probability) were different from those obtained from empty alignment.

SETTING DEMOGRAPHICAL SCENARIOS

Palaeodistribution modelling

We obtained 888 occurrence records of *H. ochraceus* across the Neotropics from the online databases GBIF (Global Biodiversity Information Facility, <http://www.gbif.org/>) and Species Link (<http://smlink.cria.org.br/>), which were mapped in a grid of cells of $0.5^\circ \times 0.5^\circ$ (longitude \times latitude) (Table S2, Fig. S1). The resolution of $0.5^\circ \times 0.5^\circ$ was the same as for the climatic variables used in the ecological niche modelling (ENM) approach.

The climatic layers were represented by five bioclimatic variables (annual mean temperature, mean diurnal temperature range, isothermality – mean diurnal range/temperature annual range, precipitation of wettest month and precipitation of driest month), selected by factorial analysis with Varimax rotation from 19 bioclimatic variables obtained from the ecoClimate database (www.ecoclimate.org; Lima-Ribeiro *et al.*, 2015). We also obtained subsoil (30–100 cm) data related to soil fertility from the Harmonized World Soil Database (version 1.2, FAO/IIASA/ISRIC/ISS-CAS/JRC 2009, available at <http://www.fao.org/docrep/018/aq361e/aq361e.pdf>). Using factorial analysis we selected subsoil pH (30–100 cm). We assume subsoil pH to be constant between the LGM and pre-industrial age, which was used in ENMs as a ‘constraint variable’ to better model the environmental preferences of *H. ochraceus*. The bioclimatic variables were obtained for pre-industrial (representing current climate conditions), mid-Holocene (6 ka) and LGM (21 ka) from five coupled Atmosphere–Ocean General Circulation Models (AOGCMs) (Table S3).

Estimates of the current and past potential distribution of *H. ochraceus* were obtained by modelling its ecological niche using 12 algorithms (Table S4). When needed, we randomly selected pseudo-absences across the Neotropical grid cells keeping prevalence

equal to 0.5 (see Stokland, Halvorsen & Stoa, 2011). The ENMs were then built on the current climatic scenario and projected onto the climatic conditions during both the mid-Holocene (6 ka) and LGM (21 ka). All ENM procedures were run in the integrated computational platform BIOENSEMBLES, which provides predictions based on the ensemble approach (see Araújo & New, 2007; Diniz-Filho *et al.*, 2009).

For each algorithm and AOGCM, models were built using 75% of occurrences and tested against the remaining records (25%). To avoid modelling bias, this procedure was repeated 50 times for each combination of ENM and AOGCM by randomly splitting training (75%) and test data (25%). After eliminating the models showing poor performance [True Skill Statistics (TSS) $<$ 0.5, Allouche Tsoar & Kadmon, 2006, see Table S5], we computed the binary maps using thresholds established by the area under the receiver operating characteristic (ROC) curve (maximizing specificity + sensitivity). The frequency of a predicted presence was used as a measure of consensus suitability from 50 initial models.

Next, we applied a hierarchical ANOVA using the predicted suitability from all models (12 ENMs \times 5 AOGCMs \times 3 Times) as the response variable to disentangle the effects of climate change on species distribution through the time from predictive uncertainties in the potential distribution due to modelling components (i.e. ENMs, AOGCMs). For this, 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 about hierarchical design). Because we do not have repetitions, the residual term from ANOVA represents the interaction between AOGCM and ENM (AOGCM \times ENM; Sokal & Rohlf, 1995). Finally, the full ensemble was obtained for each time period using the predictive performances (TSS) to compute a weighted mean of suitabilities (Table S5), from which historical refugia were mapped (i.e. all grid cells with suitability values \geq 0.5 during the three time periods). The threshold of 0.5 corresponds to the suitability value after excluding the lowest 10th percentile of suitability values related to the species occurrence records. We performed sensitivity analyses using thresholds of 0.4, 0.6 and 0.7, corresponding to the 5th, 12th and 17th percentiles, respectively, and the results were virtually indistinguishable. The result differed only for a threshold $>$ 0.7, resulting in a small refugium in Central–West Brazil.

Demographical scenarios

The ENMs resulted in 60 predictive maps (12 ENMs \times 5 AOGCMs) for each time period, which were used to support alternative demographical scenarios.

By keeping the ENM and AOGCM constant across time periods, we computed the range shift across the last glacial cycle (i.e. the difference in predicted range size between the present and LGM – 21 ka), and classified the 60 predictive maps according to three general demographical scenarios using a threshold difference of ± 200 cells: (1) ‘Range Stability’, no difference in range size through time; (2) ‘Range Expansion’, range size was smaller at the LGM than at present; and (3) ‘Range Retraction’, range size was larger at the LGM than at present.

Along with the three hypotheses supported by ENMs (see above), a fourth biogeographical hypothesis of ‘Multiple Refugia’ was derived from palaeoclimatic (e.g. Ab’Saber, 2000) and palaeovegetation reconstructions (e.g. Salgado-Labouriau, 1997).

Simulation of demographical scenarios

The demographical scenarios were simulated based on coalescent analysis (Kingman, 1982) implemented in DIYABC v2.1.0 (Cornuet *et al.*, 2008; Cornuet, Ravigne & Estoup, 2010), following the framework described by Collevatti *et al.* (2012a, 2013a, 2015b). For model calibration, we used the demographical parameters estimated with Lamarc software and the same priors for sequence evolution used in Bayesian analyses (see above). The number of generations until the LGM (21 ka) was calculated using the generation time of *H. ochraceus* (15 years, R. G. Collevatti, unpublished data).

Population dynamics were simulated backwards, with 24 demes, from t_0 (present) to t_{1400} generations

ago (at 21 ka). At t_0 , all demes had the same population size $N_0 = 100\,000$ and reached N_1 , the effective population size at t_{1400} (21 ka), according to our theoretical expectation for each demographical scenario (see Figure 2 for details). Given the high variation in *H. ochraceus* effective population sizes (Table 1), we performed simulations with different initial deme sizes, $N_0 = 100, 1000, 10,000$ and $100\,000$ for all scenarios. However, we detected differences among scenarios only when $N_0 = 100\,000$. The demographical scenario predicted by ‘Range Expansion’ was then simulated by applying an exponentially negative population growth from the present to 21 ka, reaching $N_{1400} = 10\,000$, if $N_0 = 100\,000$. By contrast, the ‘Range Retraction’ scenario was simulated applying exponentially positive population growth during the same period, reaching $N_{1400} = 1000\,000$, if $N_0 = 100\,000$. The scenario of ‘Range Stability’ predicted no change in effective population size through time, i.e. $N_{1400} = 100\,000$ if $N_0 = 100\,000$. The ‘Multiple Refugia’ scenario considers that savannas shifted in range and retracted during the LGM, forming fragments of different sizes (Ab’Saber, 2000). Thus, to model ‘Multiple Refugia’, we considered that current populations (at t_0) are descendant from lineages in different demes at t_{1400} generations ago (LGM) with different effective population sizes: $N_{1400} = 1000\,000, 100\,000, 10\,000, 1000, 100$ or 10 . Migration was simulated considering all current deme descendants from lineages originally in deme 1 at t generations ago, meaning that while the coalescent tree builds back through time, there is a 0.01 per generation chance that each lineage in deme x will migrate to deme 1. We also simulated different

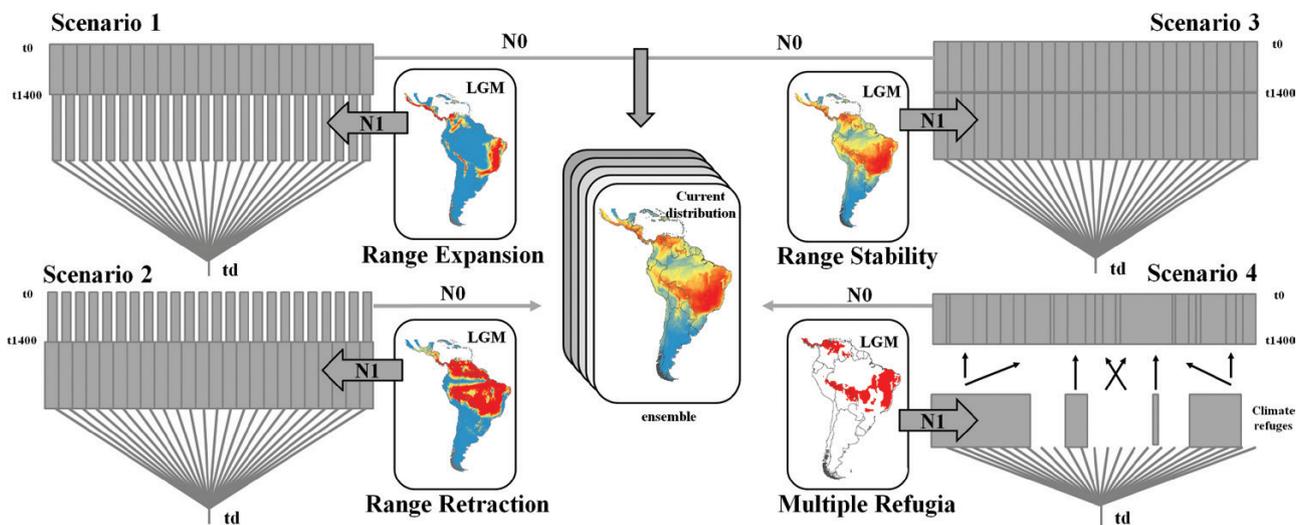


Figure 2. Demographical scenarios simulated for *Handroanthus ochraceus* using the software DIYABC 2.1.0 and their geographical representation. The simulations were performed for 24 demes. LIG, last interglacial; LGM, last glacial maximum; Pres, present day; N_0 , effective population size at time t_0 (present); N_1 , effective population size at time t_{1400} (1400 generations ago). t_d is the time to most recent common ancestor of all lineages.

migration rates. Values < 0.01 were not sufficient to show any demographical variation at the time scale we are working and values > 0.1 retrieved equal likelihoods for all models.

The relative fits of the models were calculated using approximate Bayesian computation (ABC; Excoffier, Estoup & Cornuet, 2005b; Beaumont, 2008) implemented in the software DIYABC v2.1.0. The number of distinct haplotypes, number of segregating sites, mean and variance of the pairwise differences, and mean and variance of the number of the rarest nucleotide of segregating sites, retrieved across 4 million simulations, were compared with those of the observed data sets, for cpDNA and ITS separately. In the software DYABC, the posterior probability of each scenario was retrieved to compare different scenarios that can explain the observed data set. We used the relative proportion of each scenario in the simulated data set closest to the observed data set (hereafter 'direct approach') and the logistic regression (hereafter 'logistic approach') of each scenario probability on the deviation between simulated and observed summary statistics (Cornuet *et al.*, 2008; see also the DIYABC website, <http://www1.montpellier.inra.fr/CBGP/diyabc/>).

SPATIAL PATTERNS 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 cpDNA and ITS. We first tested whether genetic differentiation across *H. ochraceus* populations (linearized F_{ST}) correlates with their geographical distance (on a log scale) using the Mantel test. The analysis was performed using the software Arlequin v3.11 and statistical significance was tested across 10 000 random permutations.

Next, we used quantile regressions to analyse the relationships of climatic suitability and stability through time with genetic diversity (Cade & Noon, 2003). For this, we calculated the difference of ensemble suitability between the LGM and the present day as a measure of climate stability through time. We then analysed whether historical changes in the geographical range of a species generated a cline spatial pattern in haplotype (h) and nucleotide (π) diversities, as well as in effective population size (N_e), due to expansion of climatically suitable conditions. We obtained the distance between each sampled population and the centroid of historical refugium and of the predicted distribution at 21 ka and the present day, and then performed quantile regressions of genetic parameters against these spatial distances and against climatic stability.

RESULTS

GENETIC DIVERSITY AND POPULATION STRUCTURE

The combined data of chloroplast intergenic spacers generated 2133 bp and the nrDNA ITS generated 610 bp, resulting in 2299 bp when microsatellites and ambiguous alignments were eliminated (1772 bp for chloroplast and 527 for ITS). We found 106 different chloroplast and 78 ITS haplotypes for the 468 individuals of *H. ochraceus*, respectively (Fig. 1). Some populations showed low haplotype and nucleotide diversities (Table 1), despite the high number of individuals sampled. Most haplotypes were exclusive of one population (Fig. 1, see also Table S6) and two chloroplast and ITS haplotypes were very widespread, H1 and H5 (Fig. 1).

The AMOVA showed significant genetic differentiation among populations for both cpDNA ($F_{ST} = 0.742$; $P < 0.001$) and nuclear ITS ($F_{ST} = 0.544$; $P < 0.001$). Bayesian clustering showed an optimal partition of four clusters for both ITS and cpDNA (Fig. 3), but with no clear geographical pattern. For ITS (Fig. 3A), almost all populations were clustered together in one group (cluster I, in green), but individuals from populations ARA and SEB from Northwest, SCA from Southeast, and PAN and BAR from Central-West were also grouped in cluster II (red). The chloroplast genome (Fig. 3B) showed a higher admixture, with most populations grouped in cluster III (light green), but with admixture with cluster I (blue) and cluster IV (orange). Populations PLC and PMS formed a different cluster for both ITS (cluster IV, blue) and chloroplast data (cluster II, red). Overall, we found high genetic differentiation among populations but with high admixture indicating historical connectivity.

PHYLOGEOGRAPHICAL RECONSTRUCTION

Population demography

We found low values of θ for each population and for the populations as a whole (Table 1), but high effective population sizes. However, migration was negligible (less than one migrant per generation) between all pair of populations (Tables S7 and S8). The EBSP showed population demographical expansion (Fig. S2a) at *c.* 380 ka, with lower growth rates after *c.* 150 ka. Fu's F_s test also showed significant signal of population expansion following a retraction for both ITS ($F_s = -25.520$, $P = 0.0001$) and cpDNA ($F_s = -24.848$, $P = 0.0002$). However, the negative F_s values potentially indicate selective sweeps. Overall, coalescent analyses showed demographical expansion through time and high effective population sizes.

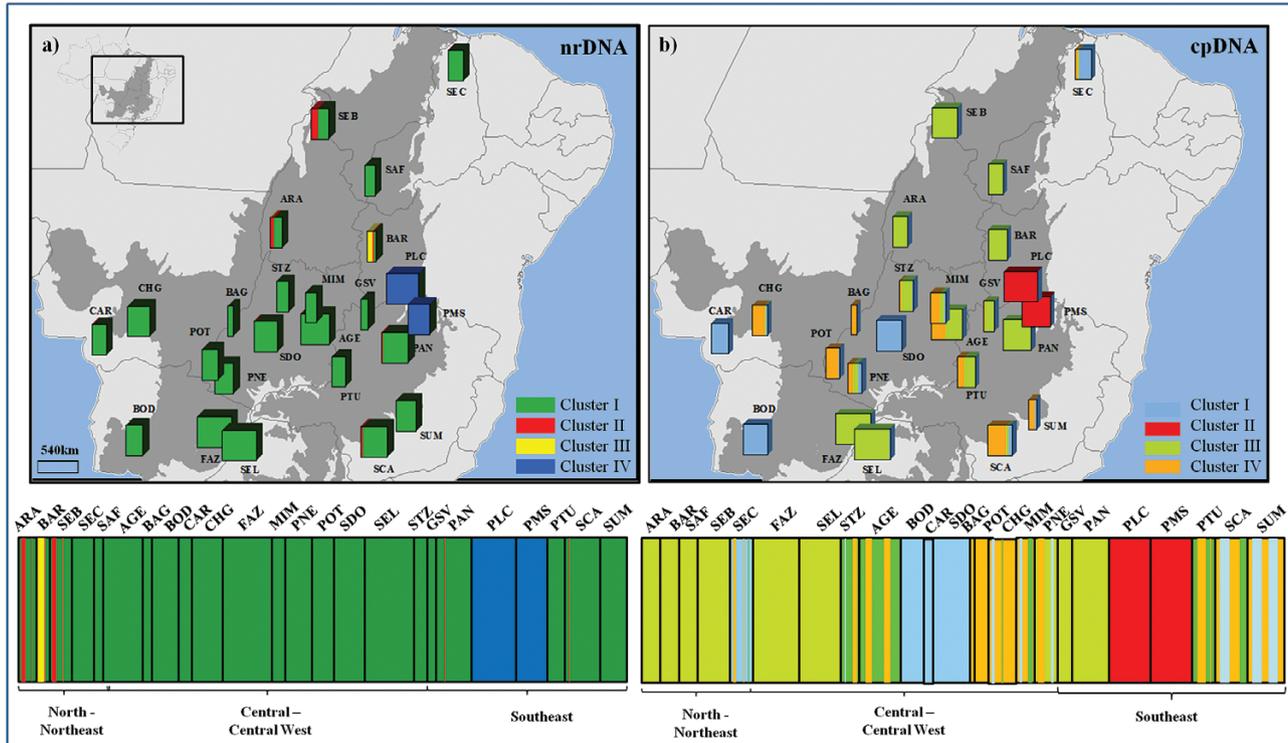


Figure 3. Geographical distribution of Bayesian clustering of *Handroanthus ochraceus* for (A) ITS and (B) cpDNA, based on the sequencing of 468 individuals from 24 populations from the Brazilian savanna. Each colour represents an inferred cluster (four clusters for ITS and cpDNA), also represented in the bar graphs below the maps. In the bar graphs, each individual bar represents an individual plant and colour indicates the cluster. The dark grey area represents the limits of the Brazilian Cerrado Biome. For details on population codes and localities see Table S1 in Appendix S1.

Lineage dispersal

We recovered one most probable ancestral location for *H. ochraceus* haplotypes (Fig. S2b) in population SCA in Southeast Brazil. Lineages started to disperse at *c.* 520 ka (Fig. S2b) to populations at the eastern (PLC and PAN), central (MIM and AGE) and western Cerrado (POT and FAZ). Most dispersal events (28 of 41) occurred in the Middle Pleistocene, after *c.* 400 ka, matching the demographical expansion revealed by EBSP (Figure 4). In addition, most dispersal events occurred during interglacial periods (47 of 71 dispersal events). The Bayes factor analysis showed that all links among localities are well supported (all BF > 8.0). The mean dispersal rate was 1.01 km/year [SD = 0.939, 95% highest posterior density interval 0.019–2.946 km/year]. The RRW analyses revealed historical connectivity among populations that coincides with the timing of population demographical expansion.

Time to most recent common ancestor

Haplotypes of *H. ochraceus* showed a recent TMRCA (Fig. 5), *c.* 1.9 Ma [95% credibility interval (CI) = 0.1–2.3 Ma], which coincides with the coalescence of

haplotypes from populations PLC and PMS with all other haplotypes. Most divergences occurred after *c.* 1.1 Ma (Fig. 5) and geographically distant populations (e.g. SEC and BOD) shared haplotypes and common ancestors.

SETTING DEMOGRAPHICAL SCENARIOS

Palaeodistribution modelling

Handroanthus ochraceus shows a clear preference for hot and dry climates (Fig. S3), matching the general current conditions of Neotropical savannas. Such climate space was equally available at the LGM, mid-Holocene and the present day (Fig. 6A). ENMs predicted only a slightly difference in the potential distribution of *H. ochraceus* through time (Fig. 6, see also Fig. S4), with a slightly higher range size at the LGM (Fig. 6A) compared to both the mid-Holocene (Fig. 6B) and the present day (Fig. 6C). In addition, a wide region across Central–West and Northeast Brazil probably acted as an historical refugium maintaining populations of *H. ochraceus* during the climate changes of the last glaciation (Fig. 6D). Overall, the ENMs showed quasi-stability in the geographical range of *H. ochraceus* through time.

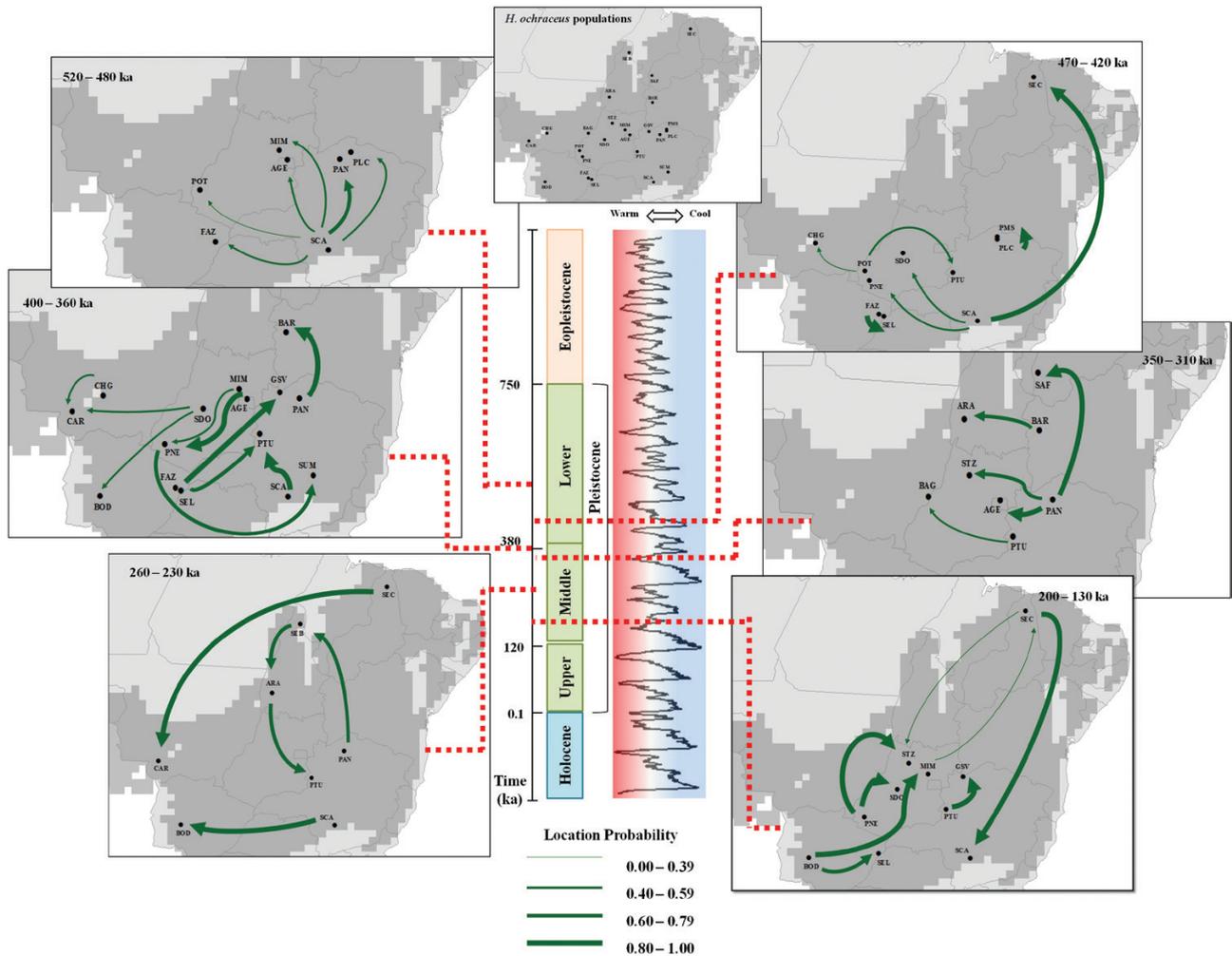


Figure 4. Spatio-temporal dynamics of *Handroanthus ochraceus* lineage diffusion among the 24 populations sampled in the Brazilian savanna, for 1.350 Ma, 900 ka, 650 ka and 300 ka until the present day. Arrows between locations represent branches in the tree along which the relevant location transition occurs. The map was adapted from the .kml file provided by SPREAD software generated using Google Earth (<http://earth.google.com>). The $\delta^{18}\text{O}$ curve corresponds to the composite benthic stable oxygen isotope ratios (Lisiecki and Raymo, 2005). For details on population codes and localities see Table S1 in Appendix S1 and Figure 1.

The models were well evaluated from TSS (Table S5). Analysis of uncertainty using a hierarchical ANOVA showed high proportional variance from the modelling method (ENMs) and AOGCM (Table S9, Fig. S5), but variation was spatially structure and higher outside the occurrence range of *H. ochraceus*, indicating that the ENMs were able to detect the effects of climate changes on the distribution dynamics of *H. ochraceus* through the last glaciation, despite the AOGCM variation.

Demographical scenarios and simulations

The scenario of 'Range Stability' was the most frequent hypothesis from ENM predictions (45%, see Table

S10) followed by 'Range Retraction' (40%) and 'Range Expansion' (15.0%).

Simulations of demographical scenarios showed only a slight difference using a direct approach (Fig. S6) for both cpDNA (Fig. S6a) and ITS (Fig. S6c). However, using a logistic approach, scenario 4 (Multiple Refugia) was the most likely one to predict the observed parameters for cpDNA (posterior probability = 1.00, 95% CI = 1.00–1.0, Fig. S6b) and scenario 1 (Range Expansion) was the most likely one for ITS (posterior probability = 0.91, 95% CI = 0.84–0.97, Fig. S6d, see also Table S11 for posterior probabilities). The simulations thus showed that *H. ochraceus* had a smaller range size at the LGM than at the present day for both molecular markers.

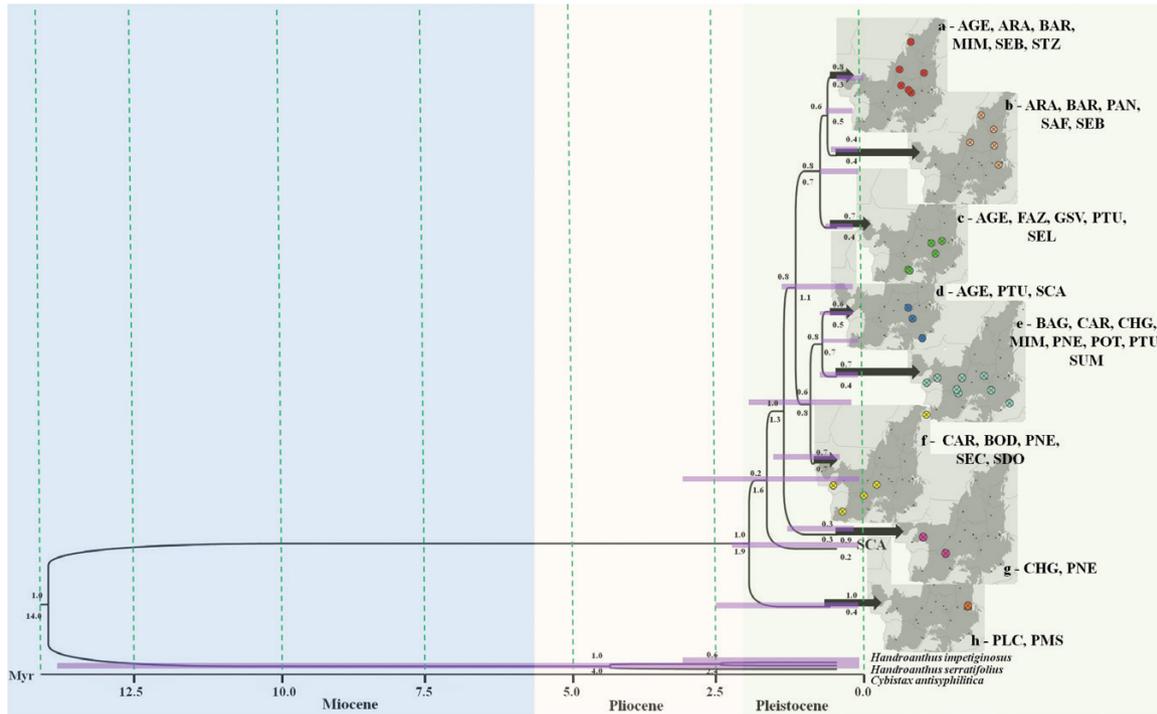


Figure 5. Relationships and TMRCA of *Handroanthus ochraceus* lineages. The light pink bar corresponds to the 95% highest posterior probability of the median time to the common ancestor; numbers above branches are support values for the nodes (posterior probability) and numbers below are the TMRCA. The time scale is in millions of years ago (Ma).

SPATIAL PATTERNS IN GENETIC DIVERSITY

Pairwise differentiation was high for almost all population pairs (Table S12), but was not correlated with geographical distance for either chloroplast ($r^2 = 0.078$, $P = 0.180$) or ITS ($r^2 = 0.054$, $P = 0.670$).

Areas with higher climatic instability through time (higher delta suitability 0–21 ka) showed lower nucleotide and haplotype diversity (Fig. S7). In addition, populations at greater distances from the centroid of the climatic refugium showed higher nucleotide and haplotype diversities (Fig. S8). Effective population size was not significantly related to the distance from the centroid or to climatic suitability.

DISCUSSION

Our results show that *H. ochraceus* lineages dispersed cyclically from Southeast and West towards Central–West Brazil. Most dispersal events occurred from populations at the edges of the Cerrado biome towards Central Brazil. Populations at the edge of the historical refugium showed higher genetic diversity and were the source of migrants to populations at the centre of the refugium. The palaeodistribution dynamics showed a large historical refugium throughout Central Brazil and range stability through time. The slight

difference in range size through the last glacial cycle is due mainly to range expansion towards the Amazon Basin at the LGM. All sampled populations remained in environmentally suitable areas through the last glacial cycle, inside the large climatic refugium, which may have favoured high historical connectivity among populations. In fact, the climatic niche space of *H. ochraceus* indicates its preference for hot and dry conditions that prevailed during the last glacial cycle in its geographical range. The lack of dispersal of new lineages after *c.* 130 ka, the low number of migrants per generation ($Nm < 1.0$) and the high number of haplotypes with restricted geographical distribution show that, despite the high historical connectivity, recent connectivity is low. The lack of recent connectivity may be due to widespread fragmentation due to anthropogenic disturbance. Although RRW predicts spatial displacements occurring earlier than ENMs, they both converge to consistent patterns, showing that our approach is reliable in reconstructing the phylogeography of *H. ochraceus*.

The dispersal results also indicated that populations currently found in Northern and Northeastern Brazil received migrants mainly from the Southeast (e.g. see arrow direction in Fig. 4). Such dispersal routes have probably occurred because suitable climatic conditions were available across a wider region during warming

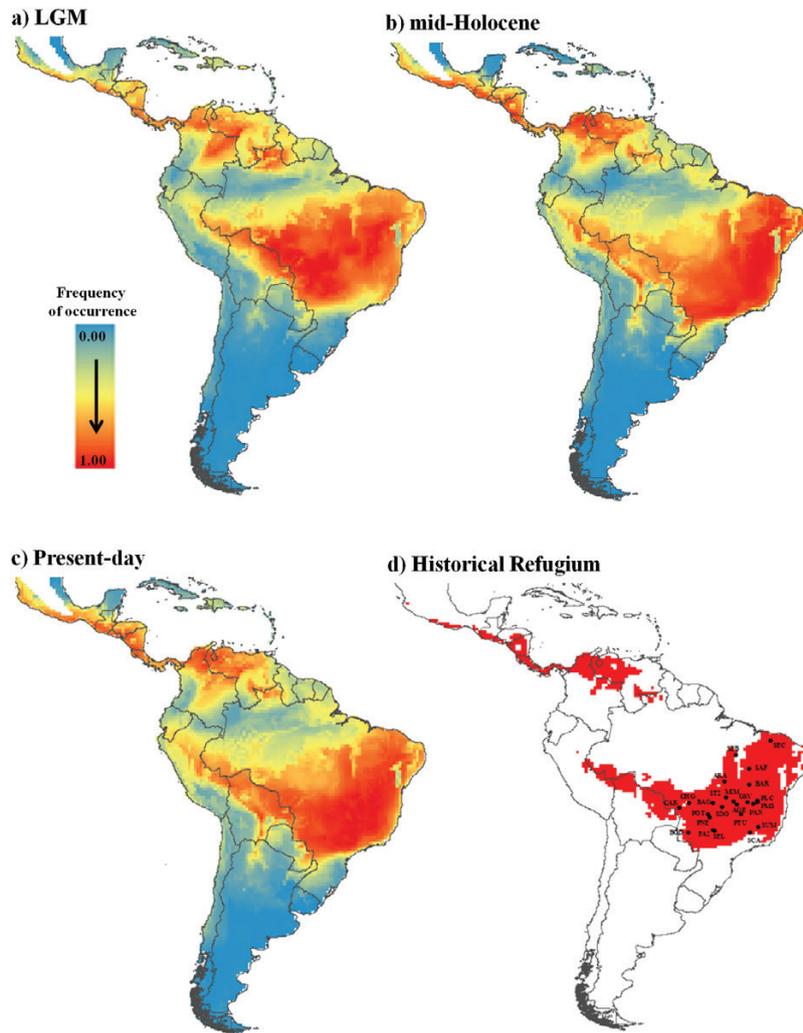


Figure 6. Consensus maps of the 60 models expressing the potential distribution for *Handroanthus ochraceus*, based on ecological niche modelling during the (A) LGM (21 ka), (B) mid-Holocene (6 ka) and (C) present day, and (D) for historical refugium through time (from the LGM to the present day). Black dots indicate the 24 populations sampled.

phases, allowing the connectivity of populations. Diffusion models suggested that inverse dynamics occurred during cooling phases (e.g. 200 and 260 ka), when lineages dispersed in the opposite direction, from North-Northeast toward Central and Central-West Brazil. This region was hotter and drier, thus retaining more suitable niches for *H. ochraceus* during the cooler phases. A potential continuous historical refugium across Neotropical savannas may be the key factor allowing uninterrupted lineage dispersal among populations of Central Brazil, as predicted by the RRW model. Note that a similar result was reported for a related species, *Tabebuia aurea* (Collevatti *et al.*, 2015a), reinforcing the idea that dispersal from Southeast to Northeast might have been an important route for savanna plant species in the Quaternary. In addition, population SCA was also one of the ancestral

localities of *T. aurea* lineages (together with CAC, see Collevatti *et al.*, 2015a). However, the dispersal of lineages present in a population would not be detected as a new gene flow event. Thus, population connectivity could be maintained after 130 ka but our results show no dispersal of new lineages.

Despite the difference in demographical scenarios between cpDNA (Multiple Refugia) and ITS (Range Expansion), both molecular markers predicted population expansion through time (i.e. from the LGM to present day) as the most likely general demographical scenario for *H. ochraceus*, matching our theoretical prediction. The difference between the two molecular markers is probably due to evolutionary rates. Besides substitution rates, which are higher in ITS than in chloroplast intergenic regions, and concerted evolution in nrDNA that may decrease genetic variation, ITS has

four times the effective size of the chloroplast genome and may be differentially affected by drift. However, ENMs revealed quasi-stability in geographical range, with a slightly larger range at the LGM compared to the present day. Note that *H. ochraceus* showed different geographical range dynamics through time compared to other taxonomically related species such as *T. aurea* (Collevatti *et al.* 2015a) and *T. roseoalba* (Melo *et al.*, 2016) that showed range retraction during the LGM. Moreover, *H. ochraceus* has a more recent common ancestor (*c.* 1.9 Ma) than other related species, such as *T. aurea* (4.0 ± 2.5 Ma, Collevatti *et al.* 2015a), *T. roseoalba* (4.9 ± 1.9 Ma, Melo *et al.*, 2016) and *H. impetiginosus* (4.7 ± 1.1 Ma, Collevatti *et al.*, 2012a), probably due to differences in demographical history.

Because favourable climatic conditions for *H. ochraceus* (i.e. hotter and drier) were more widely available during glacial than interglacial periods in the Neotropics, effective population size has increased in the last *c.* 380 ka, despite glacial cycles. Besides, the increasing number of lineages, especially in the last *c.* 500 ka, may have resulted from the demographical expansions across cyclical glacial periods, during which wider areas of suitability were available for colonization.

Lineage dispersal was mainly from populations SCA and PAN (see Fig. 4). This diffusion pattern explains the widespread geographical distribution of some haplotypes (e.g. H1 and H5 for ITS, and H1 and H5 for cpDNA, see Fig. 1) and the restricted geographical distribution of other haplotypes that diverged more recently (e.g. H76, H77 and H78 for ITS, and H104, H105 and H106 for cpDNA, see Fig. 1), leading to high population differentiation. The analyses also indicated *in situ* diversification of haplotypes in many populations, such as PLC, PMS, PNE, SEB, SEC, SEL, STZ, SUM (see Fig. S2b), and a higher diversification after *c.* 200 ka, which coincides with the demographical population growth. The lack of dispersal events after *c.* 130 ka, also indicated by coalescent analyses in Lamarc software, may have led to the restricted geographical distribution of many haplotypes. Although our findings show possible long-distance dispersal between distant populations such as SCA and SEC and SEC and CAR, it is important to note that these events are probably due to stepping-stone dispersal and would be detected only with continuous spatial sampling. However, continuous large-scale sampling is hindered by the fragmentation of savanna in Brazil. Besides, mean rate of dispersal estimates of 1.01 km/year using RRW is congruent with the findings for the related species *T. aurea* (1.17 km/yr, Collevatti *et al.*, 2015a). *Handroanthus ochraceus* and *T. aurea* share the same pollinator species and seed dispersal

syndrome, and they co-occur in many localities. Thus, we expect that long-distance dispersal be due mainly to pollen dispersal (see Braga & Collevatti, 2011), which may also explain the widespread distribution of ITS haplotypes.

Despite the demographical expansions through time, ENMs suggested that all populations are within the historical refugium, i.e. areas of high climatic suitability throughout the study period (see Fig. 6D), favouring population persistence and genetic connectivity. If this stable area was maintained during the recurrent glaciations of the Quaternary, we can predict long-term genetic connectivity among *H. ochraceus* populations, which is also suggested by a high density of well-supported links among populations inside this historical refugium. Thus, the low haplotype diversity within some populations may be an effect of selective sweeps, as indicated by the negative F_s values (e.g. Kapralov & Filatov, 2007). In addition, despite differences in effective population size and inheritance, the chloroplast genome showed higher diversity than ITS nrDNA, although the latter has four times the effective size. This may be due to concerted evolution in nrDNA that may decrease genetic variation (Ohta, 1984). Moreover, recent anthropogenic disturbance may also have caused bottlenecks, decreasing genetic diversity in *H. ochraceus* populations because some populations, such as BAG, POT, CAR, which are isolated remnants in pasture or crop plantations, have low genetic diversity. We believe that our results are not an artefact of sampling effort because populations with larger sample sizes had the same number of haplotypes as populations with smaller sample sizes.

The pattern of lineage spread and extinction may explain the low genetic diversity in many populations, even in the centre of the species' geographical range. In fact, populations in more unstable areas had lower genetic diversity and populations further from the centroid, at the edge of the climatic refugium, showed higher genetic diversity. This reinforces the multiple refugia demographical scenario. We hypothesize that these populations, at the edge of historical refugia, may represent 'stable rear edges', i.e. relict populations that have persisted in suitable local habitats across the Quaternary glaciation cycles (see Petit *et al.*, 2003; Hampe & Petit, 2005). Note that due to the coarse resolution of climatic layers ($0.5^\circ \times 0.5^\circ$) the ENM may not be able to predict species presence on the local scale (small refugia).

Finally, the scenario of demographical expansion through time for *H. ochraceus* was also supported for other Brazilian savanna species such as *Tabebuia aurea* (Collevatti *et al.*, 2015a), *Caryocar brasiliense* (Collevatti *et al.*, 2012b) and *Dipteryx alata* (Collevatti *et al.*, 2013b) and for a Brazilian savanna swamp

palm species, *Mauritia flexuosa* (Lima *et al.*, 2014). Phylogeographical studies of savanna species (e.g. Ramos *et al.*, 2007; Novaes *et al.*, 2010, 2013; Collevatti *et al.*, 2012b) have shown high levels of genetic differentiation among populations and evidence of colonization of Southeastern Brazilian savannas during warmer interglacial periods of the Quaternary. However, our results for *H. ochraceus* and *T. aurea* (Collevatti *et al.* 2015a) do not support that populations from Southeast Brazil were recently colonized because population SCA, in Southeast Brazil, was the ancestral locality for extant lineages of the two species and had old lineages. Our results also match the spatial pattern in genetic diversity for other savanna tree species, such as *C. brasiliense* (Collevatti *et al.*, 2012b) and *T. aurea* (Collevatti *et al.*, 2015a), and reinforces the idea that relict populations may have persisted in suitable local habitats across the Quaternary glaciation cycles. Nevertheless, there is high variation in spatial patterns of genetic diversity in savanna trees, and some species show a central–peripheral pattern, such as *Eugenia dysenterica* (Lima *et al.*, 2017) and *Dimorphandra mollis* (Souza *et al.*, 2017).

In conclusion, our findings suggest that the pattern of genetic diversity in *H. ochraceus* may be the result of range stability and population demographical expansion through the Quaternary. The general dynamics of lineage dispersal followed suitable climatic conditions, towards the Northern and Central–West regions during warming periods in the early Quaternary, and an opposite direction during glacial phases. The same pattern was also found for a taxonomically related species, *T. aurea* (Collevatti *et al.* 2015a). We also provide evidence of a general pattern of distribution dynamics and lineage dispersal in savanna tree species and emphasize that the southeastern and western Cerrado were important refugia for savanna species. Our study also shows that our multi-model inference approach allows us to infer the different processes affecting phylogeographical patterns through time.

ACKNOWLEDGMENTS

This work was supported by grants to the research network GENPAC (Geographical Genetics and Regional Planning for natural resources in Brazilian Cerrado) supported by CNPq/MCT/CAPES/FAPEG (project nos. 564717/2010-0, 563727/2010-1 and 563624/2010-8), and Rede Cerrado CNPq/PPBio (project no. 457406/2012-7). We 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.

We also thank three anonymous reviewers for their helpful comments.

REFERENCES

- 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.
- 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.
- Beaumont MA. 2008.** Joint determination of topology, divergence time, and immigration in population 5 trees. In: Matsumura S, Renfrew PFC, eds. *Simulation, genetics and human prehistory*. Cambridge: McDonald Institute for Archeological Research, 134–154.
- Behling H. 2003.** Late glacial and Holocene vegetation, climate and fire history inferred from Lagoa Nova in the southeastern Brazilian lowland. *Vegetation History and Archaeobotany* **12**: 263–270.
- Bielejec F, Rambaut A, Suchard MA, Lemey P. 2011.** SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics* **27**: 2910–2912.
- Braga AC, Collevatti RG. 2011.** Temporal variation in pollen dispersal and breeding structure in a bee pollinated Neotropical tree. *Heredity* **106**: 911–919.
- Cade BS, Noon BR. 2003.** A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment* **1**: 412–420.
- Caetano S, Prado D, Pennington RT, Beck S, Oliveira-Filho A, Spichiger R, Naciri Y. 2008.** The history of seasonally dry tropical forests in eastern South America: inferences from the genetic structure of the tree *Astronium urundeuwa* (Anacardiaceae). *Molecular Ecology* **17**: 3147–3159.
- Cheng H, Sinha A, Cruz FW, Wang X, Edwards RL, d'Horta FM, Ribas CC, Vuille M, Stott LD, Auler AS. 2013.** Climate change patterns in Amazonia and biodiversity. *Nature Communications* **4**: 1–6.
- Collevatti RG, Terribile LC, Lima-Ribeiro MS, Nabout JC, Oliveira G, Rangel TF, Rabelo SG, Diniz-Filho JAF. 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**: 1–8.
- Collevatti RG, Telles MPC, Nabout JC, Chaves LJ, Soares TN. 2013a.** Demographic history and the low

- genetic diversity in *Dipteryx alata* (Fabaceae) from Brazilian Neotropical savannas. *Heredity* **111**: 97–105.
- Collevatti RG, Terribile LV, Oliveira G, Lima-Ribeiro MS, Nabout JC, Rangel TF, Diniz-Filho JAF. 2013b.** Drawbacks to palaeodistribution modelling: the case of South American seasonally dry forests. *Journal of Biogeography* **40**: 345–358.
- Collevatti RG, Terribile LV, 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**: 1–15.
- 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**: 1–8.
- Corander J, Marttinen P. 2006.** Bayesian identification of admixture events using multi-locus molecular markers. *Molecular Ecology* **15**: 2833–2843.
- Corander J, Marttinen P, Sirén J, Tang J. 2008.** Enhanced Bayesian modeling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* **9**: 1–14.
- Cornuet JM, Santos F, Beaumont MA, Robert CP, Marin JM, Balding DJ, Guillemaud T, Estoup A. 2008.** Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* **24**: 2713–2719.
- Cornuet JM, Ravigne V, Estoup A. 2010.** Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics* **11**: 401.
- Darriba D, Taboada GL, Doallo R, Posada D. 2012.** jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**: 772.
- Desfeux C, Lejeune B. 1996.** Systematics of Euro Mediterranean *Silene* (Caryophyllaceae): evidence from a phylogenetic analysis using ITS sequence. *Academic Science of Paris* **319**: 351–358.
- Diniz-Filho JAF, Bini LM, Rangel TF, Loyola RD, Hof C, Noguez-Bravo D, Araujo MB. 2009.** Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography* **32**: 897–906.
- Doyle JJ, Doyle JL. 1990.** Isolation of plant DNA from fresh tissue. *Focus* **12**: 13–15.
- Drummond AJ, Rambaut A. 2007.** BEAST: Bayesian Evolutionary Analysis by Sampling Trees. *BMC Evolutionary Biology* **7**: 214.
- Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012.** Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**: 1969–1973.
- Excoffier L, Smouse P, Quattro J. 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. 2005a.** Arlequin ver. 3.0: na integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**: 47–50.
- Excoffier L, Estoup A, Cornuet JM. 2005b.** Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics* **169**: 1727–1738.
- Excoffier L, Ray N. 2008.** Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology and Evolution* **23**: 347–351.
- Excoffier L, Foll M, Petit RJ. 2009.** Genetic consequences of range expansions. *Annual Review of Ecology and Systematics* **40**: 481–501.
- Ferreira MAR, Suchard MA. 2008.** Bayesian analysis of elapsed times in continuous-time Markov chains. *Canadian Journal of Statistics* **36**: 355–368.
- Gibbs PE, Bianchi M. 1993.** Post-pollination events in species of *Chorisia* (Bombacaceae) and *Tabebuia* (Bignoniaceae) with late acting self-incompatibility. *Acta Botanica Brasiliica* **106**: 64–71.
- Hampe A, Petit RJ. 2005.** Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters* **8**: 461–467.
- Heled J, Drummond AJ. 2008.** Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology* **8**: 289.
- 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.
- Hoorn C, Wesselingh FP, ter Steege H, Bermutez MA, Mora A, Sevink J, Sanchez-Mesenguer A, Anderson CL, Figueiredo JP, Jaramillo C, Riff D, Negri FR, Hooghiemstra H, Lundberg J, Stadler T, Sarkanen T, Antonelli A. 2010.** Amazonia through time: Andean uplift, climate change, landscape evolution and biodiversity. *Science* **330**: 927–931.
- Huber O. 1987.** Neotropical savannas: Their flora and vegetation. *Trends in Ecology & Evolution* **3**: 67–71.
- Kapralov MV, Filatov DA. 2007.** Widespread positive selection in the photosynthetic Rubisco enzyme. *BMC Evolutionary Biology* **7**: 73.
- Kingman JFC. 1982.** The coalescent. *Stochastic Processes and their Applications* **13**: 235–248.
- Klink CA, Machado RB. 2005.** Conservation of the Brazilian Cerrado. *Conservation Biology* **19**: 707–713.
- Knowles LL, Maddison WP. 2002.** Statistical phylogeography. *Molecular Ecology* **11**: 2623–2635.
- Kuhner MK. 2006.** Lamarc 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* **22**: 768–770.
- Lemey P, Rambaut A, Drummond AJ, Suchard MA. 2009.** Bayesian phylogeography finds its roots. *PLoS Computational Biology* **5**: 1–16.
- Lemey P, Rambaut A, Welch JJ, Suchard MA. 2010.** Phylogeography takes a relaxed random walk in continuous space and time. *Molecular Biology and Evolution* **27**: 1877–1885.
- 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-Ribeiro MS, Varela S, González-Hernández 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.
- Lima JS, Telles MPC, Chaves LJ, 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.
- Lisiecki EL, Raymo ME. 2005.** A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}\text{O}$ records. *Paleoceanography* **20**: PA1003.
- Luebert F, Weigend M. 2014.** Phylogenetic insights into Andean plant diversification. *Frontiers in Ecology and Evolution* **2**: 1–17.
- Luebert F, Muller AH. 2015.** Effects of mountain formation and uplift on biological diversity. *Frontiers in Genetics* **6**: 1–2.
- Marris E. 2005.** Conservation in Brazil: The forgotten ecosystem. *Nature* **437**: 944–945.
- Melo WA, Lima-Ribeiro MS, Terribile LC, Collevatti RG. 2016.** Coalescent simulation and paleodistribution modeling for *Tabebuia rosealba* do not support south american dry forest refugia hypothesis. *PLoS One* **11**: e0159314.
- Minin VN, Bloomquist EW, Suchard MA. 2008.** Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular Biology and Evolution* **25**: 1459–1471.
- Nei M. 1987.** *Molecular evolutionary genetics*. New York: Columbia University Press.
- 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.
- Ohta T. 1984.** Population genetics theory of concerted evolution and its application to the immunoglobulin V gene tree. *Journal of Molecular Evolution* **20**: 274–280.
- Pennington RT, Lavin M, Prado DE, Pendry CA, Pell SK, Butterworth CA. 2004.** Historical climate change and speciation: neotropical seasonally dry forest plants show patterns of both Tertiary and Quaternary diversification. *Philosophical Transactions of the Royal Society B: Biological Sciences* **359**: 515–538.
- Petit RJ, Aguinagalde I, de Beaulieu JL, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Muller-Starck G, Demesure-Musch G, Palmé A, Martín JP, Rendell S, Vendramin GG. 2003.** Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300**: 1563–1565.
- Rambaut A, Drummond AJ. 2008.** *Tracer v1.4*. Available at: <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.
- Rull V. 2011.** Neotropical biodiversity: timing and potential drivers. *Trends in Ecology & Evolution* **26**: 508–513.
- Rull V, Montoya E, Nogué S, Vegas-Vilarrúbia T, Safont E. 2013.** Ecological palaeoecology in the neotropical Gran Sabana region: Long-term records of vegetation dynamics as a basis for ecological hypothesis testing. *Perspectives in Plant Ecology, Evolution and Systematics* **15**: 338–359.
- Rull V, Vegas-Vilarrúbia T, Montoya E. 2015.** Neotropical vegetation responses to Younger Dryas climates as analogs for future climate change scenarios and lessons for conservation. *Quaternary Science Reviews* **115**: 28–38.
- Salgado-Labouriau ML. 1997.** Late Quaternary paleoclimate in the savannas of South America. *Journal of Quaternary Science* **12**: 371–379.
- Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL. 2005.** The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* **92**: 142–166.
- Sievers F, Wilm A, Dineen DG, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011.** Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**: 1–6.
- Sokal RR, Rohlf FJ. 1995.** *Biometry: The Principles and Practice of Statistics in Biological Research, 3rd Edn*. New York, NY: W. H. Freeman and Company.
- Souza HAV, Collevatti RG, Lima-Ribeiro MS, Lemos-Filho JP, Lovato MB. 2017.** A large historical refugium explains spatial patterns of genetic diversity in a Neotropical savanna tree species. *Annals of Botany* **119**: 239–252.
- Spera SA, Galford GL, Coe MT, Macedo MN, Mustard JF. 2016.** Land-use change affects water recycling in Brazil's last agricultural frontier. *Global Change Biology* **22**: 3405–3413.
- 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.
- Summerhayes C, Charman D. 2014.** Introduction to Holocene climate change: new perspectives. *Journal of the Geological Society* **172**: 251–253.
- Terribile LC, Lima-Ribeiro MS, Araújo MB, Bizão N, Collevatti RG, Dobrovolski R, Franco AA, Guilhaumon F, Lima JS, Murakami DM, Nabout JC, Oliveira G, Oliveira LK, Rabelo SG, Rangel TF, Simon LM, Soares TN, Telles MPC, Diniz-Filho JAF. 2012.** Areas of climate stability of species ranges in the Brazilian Cerrado: disentangling uncertainties through time. *Natureza & Conservação* **10**: 152–159.
- Vieira FA, Novaes RML, Fajardo CG, Santos RM, Almeida HS, Carvalho D, Lovato MB. 2015.** Holocene southward expansion in seasonally dry tropical forests in South America: phylogeography of *Ficus bonijesulapensis* (Moraceae). *Botanical Journal of the Linnean Society* **177**: 189–2011.
- Waters JM, Fraser CI, Hewitt GM. 2013.** Founder takes all: density-dependent processes structure biodiversity. *Trends in Ecology and Evolution* **28**: 78–85.

SUPPORTING INFORMATION

Supplementary data are available online at www.aob.oxfordjournals.org and consist of the following.

Appendix S1. Supplementary tables (Tables S1–S11) with sampling locations and details on ecological niche modelling and demographical scenarios.

Appendix S2. Supplementary figures (Figs S1–S8) with details on ecological niche modelling, demographical simulations and spatial genetic diversity.