



Relaxed random walk model coupled with ecological niche modeling unravel the dispersal dynamics of a Neotropical savanna tree species in the deeper Quaternary

Rosane G. Collevatti^{1*}, Levi C. Terribile², Suelen G. Rabelo¹ and Matheus S. Lima-Ribeiro²

OPEN ACCESS

Edited by:

Guo-Bo Chen,
University of Queensland, Australia

Reviewed by:

Octavio Salgueiro Paulo,
Universidade de Lisboa, Portugal
Yessica Rico,
Royal Ontario Museum, Canada

*Correspondence:

Rosane G. Collevatti,
Instituto de Ciências Biológicas,
Universidade Federal de Goiás,
CP 131, 74001-970 Goiânia, Brasil
rosanegc68@hotmail.com

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Plant Science

Received: 28 April 2015

Accepted: 07 August 2015

Published: 25 August 2015

Citation:

Collevatti RG, Terribile LC, Rabelo SG and Lima-Ribeiro MS (2015) Relaxed random walk model coupled with ecological niche modeling unravel the dispersal dynamics of a Neotropical savanna tree species in the deeper Quaternary. *Front. Plant Sci.* 6:653. doi: 10.3389/fpls.2015.00653

¹ Laboratório de Genética & Biodiversidade, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, Brasil, ² Laboratório de Macroecologia, Universidade Federal de Goiás, Jataí, Brasil

Understanding the dispersal routes of Neotropical savanna tree species is an essential step to unravel the effects of past climate change on genetic patterns, species distribution and population demography. Here we reconstruct the demographic history and dispersal dynamics of the Neotropical savanna tree species *Tabebuia aurea* to understand the effects of Quaternary climate change on its current spatial patterns of genetic diversity. We sampled 285 individuals from 21 populations throughout Brazilian savannas and sequenced all individuals for three chloroplast intergenic spacers and ITS nrDNA. We analyzed data using a multi-model inference framework by coupling the relaxed random walk model (RRW), ecological niche modeling (ENM) and statistical phylogeography. The most recent common ancestor of *T. aurea* lineages dated from $\sim 4.0 \pm 2.5$ Ma. *T. aurea* lineages cyclically dispersed from the West toward the Central-West Brazil, and from the Southeast toward the East and Northeast Brazil, following the paleodistribution dynamics shown by the ENMs through the last glacial cycle. A historical refugium through time may have allowed dispersal of lineages among populations of Central Brazil, overlapping with population expansion during interglacial periods and the diversification of new lineages. Range and population expansion through the Quaternary were, respectively, the most frequent prediction from ENMs and the most likely demographic scenario from coalescent simulations. Consistent phylogeographic patterns among multiple modeling inferences indicate a promising approach, allowing us to understand how cyclical climate changes through the Quaternary drove complex population dynamics and the current patterns of species distribution and genetic diversity.

Keywords: Bignoniaceae, brownian diffusion, coalescent simulation, phylogeography, Quaternary climate changes, *Tabebuia aurea*

Introduction

Spatial displacements in species distributions due to Quaternary climate changes have played an important role in shaping the genetic diversity of many species across space (e.g., Taberlet et al., 1998). Postglacial range expansion, for instance, may have led to spatial genome assortment due to leading edge colonization as the species tracks suitable environments (Hewitt, 1996). The spreading from the leading edge may lead to bottlenecks of the colonizing genome, decreasing genetic diversity in some new colonizing areas. In addition, allele surfing, i.e., the spread and frequency increase of a low-frequency allele that migrates on the wave of advance of a population in expansion (Excoffier and Ray, 2008; Arenas et al., 2012), and density-dependent processes due to the fast colonization and founder events may also cause patches and sectoring in genetic diversity (see Excoffier et al., 2009; Waters et al., 2013 for reviews). Investigating these aspects of paleodistribution dynamics has been a key point to understand the effects of past colonization and to unravel the role of Quaternary climate changes in shaping the current spatial pattern of genetic diversity worldwide (e.g., Petit et al., 2002).

A recurrent approach to reconstruct past species distributions and generate independent paleoscenarios of demographical history have been coupling ecological niche modeling (ENM) with fossil data and paleoclimatic simulations in an ecological and biogeographical context (e.g., Lima et al., 2014; Metcalf et al., 2014). Subsequently, the alternative hypotheses of demographical histories are tested using coalescence analysis (e.g., Carstens and Richards, 2007; Knowles et al., 2007; Collevatti et al., 2013a). In the context of multi-model inference, ENM and coalescence modeling have provided a promising and valuable tool by considering important sources of uncertainty during historical reconstructions, such as ENM algorithms and paleoclimatic simulations uncertainties (Collevatti et al., 2012a,b), and stochastic variance in coalescent models (Kuhner, 2008). However, such an approach does not directly recover the spatio-temporal lineage dispersal process, which has a central role on the evolutionary dynamics and structure of populations. For instance, spatial inferences about dispersal events are often limited to the indirect interpretation of evolutionary histories by analyzing the predicted shifts in species' geographical range over time and the spatial locations of the sampled populations. As a consequence, the source, pathways, and routes of migration remain uncertain, and indirect inferences limit the scope of hypotheses that could be tested in phylogeography. Thus, together with ENM and coalescence modeling, we propose that a direct spatio-temporal reconstruction of lineage dispersal (Lemey et al., 2009, 2010) should integrate the context of multi-model inference to better understand the climate footprint on demographic history, shaping the spatial pattern in genetic diversity (see also Collevatti et al., 2015).

A multi-model inference approach is particularly useful to investigate the complex dynamics and current patterns of genetic diversity in tree species with a long generation time and life span. This is the case of *Tabebuia aurea* (Bignoniaceae), which is widely distributed through the Neotropics, occurring in seasonal savannas and wet-savanna grasslands (Figure S1 in Appendix

S2). Locally it is distributed in well-delimited patches with high density. The species is hermaphroditic and pollinated mainly by large-size bees such as bumblebees (*Bombus* spp.), carpenter bees (*Xylocopa* spp.) and *Centris* spp., and its winged seeds are wind dispersed.

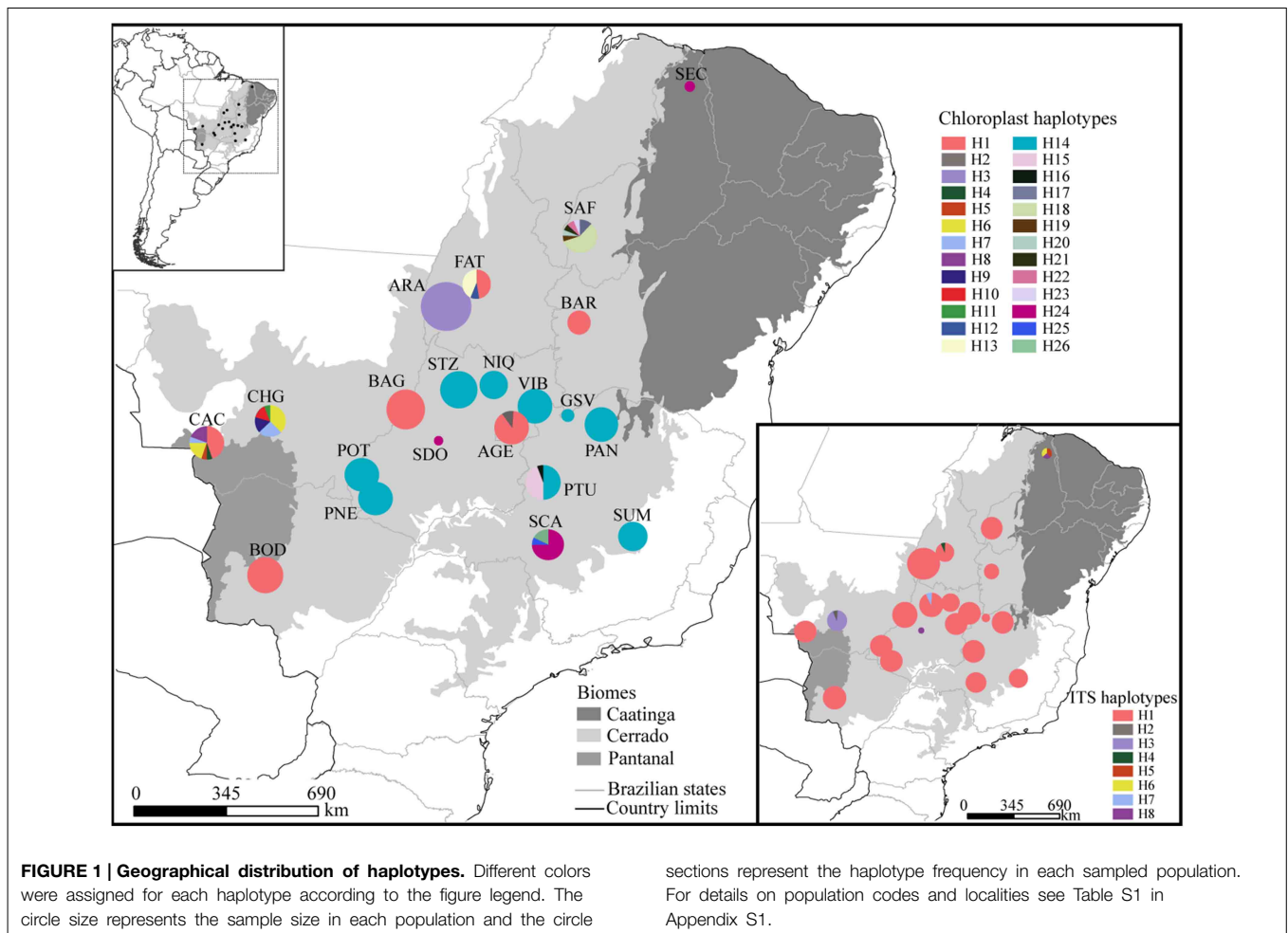
During the Last Glacial Maximum (LGM), for instance, the climate was drier in most of South America, leading to a retraction in geographical range of many arboreal savanna taxa and expansion of grasslands (Salgado-Labouriau, 1997; Behling, 2003). Because glaciations were recurrent throughout the Quaternary, the cycles of range retraction and expansion during glacial and interglacial phases may have left genetic signatures in Neotropical savannas species (e.g., Collevatti et al., 2012b). Moreover, understanding historical processes like dispersal routes in the Neotropics 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) may give clues to the pathways of lineage dispersal across the Neotropics through the Quaternary.

In this study, we reconstructed the demographical history and dispersal dynamics of the Neotropical savanna tree species *T. aurea* through the Quaternary using a multi-model inference approach to unravel the climate change effects on its current spatial pattern of genetic diversity. Because Neotropical savanna species have been showing range retraction during the LGM followed by expansion (e.g., Collevatti et al., 2012b, 2013b), we expect that *T. aurea* had a range and demographic retraction during glacial phases, leading to high population differentiation and low genetic connectivity among populations.

Materials and Methods

Population Sampling

We sampled 21 populations (285 individuals) of *T. aurea* throughout the Brazilian savanna (Figure 1, see also Table S1 in Appendix S1). Distance between population pairs ranged from 69.0 to ~3800 km. We sampled only adult individuals to avoid effects of high kinship on estimation of genetic parameters. In populations with more than 16 adult individuals, we sampled and sequenced 16 individuals and in populations smaller than this we sampled all individuals (Table 1). We focused our sampling efforts on the savannas of Central-West Brazil due to the current higher abundance of *T. aurea* and our focus on savanna biogeography. In addition, although occurrence map (see Figure S1 in Appendix S2) show records of *T. aurea* in the Northeastern Brazil (Caatinga biome), populations in these places are barely found (we could find only one or two individuals in most places we visited, such as population SEC). The Northeast Brazil is dominated by a xeromorphic steppe vegetation with seasonally dry forest and some island of savanna where *T. aurea* may occur but the high level of anthropogenic disturbance had removed most natural vegetation. We also sampled and sequenced individuals of *Tabebuia impetiginosa*, *Tabebuia chrysotrica*, and *Cybistax antisiphilitica* to include as outgroups in coalescence analyses (Table S1).



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 (ITS hereafter) from nuclear ribosomal DNA (nrDNA) (Desfeux and Lejeune, 1996). PCR conditions, amplifications and sequencing followed Collevatti et al. (2012a).

Consensus sequences were obtained using the software SeqScape v2.6 (Applied Biosystems, CA) and aligned with the software ClustalΩ (Sievers et al., 2011). Polymorphisms at mononucleotide microsatellites were excluded due to ambiguous alignment and to higher mutation rates. Long indels (usually with more than 5 bp) were coded as one evolutionary event (one character).

Genetic Diversity and Structure of Populations

To investigate genetic diversity and structure of populations, chloroplast and nuclear ITS were analyzed separately. Nucleotide (π) and haplotype (h) diversity were estimated for each population and overall populations following Nei (1987) using the software ArlequinVer 3.11 (Excoffier et al., 2005). To understand the phylogenetic relationships among haplotypes, intraspecific phylogenies for chloroplast and ITS data were

inferred using median-joining network analysis implemented in the software Network 4.6.1.0 (Forster et al., 2004). To test the hypothesis of population differentiation, we performed an analysis of molecular variance (AMOVA, Excoffier et al., 1992) and estimated F_{ST} and tested whether genetic differentiation across *T. aurea* populations (linearized F_{ST}) is correlated with geographical distance (logarithm) among populations using the Mantel test. Analyses were performed using the software ArlequinVer 3.11 and statistical significance was tested by a non-parametric permutation test (10,000 permutations).

Phylogeographic Reconstruction

Population Demography

All coalescent analyses were performed with chloroplast and nrDNA ITS partitions concatenated, but giving separate priors. No evidence of heterozygous individuals was found when ITS sequences were analyzed using SeqScape v2.6 (Applied Biosystems, CA). Thus, recombination was neglected in all coalescent analyses. To set the priors, evolutionary model selection for both chloroplast and ITS regions was performed using Akaike Information Criterion implemented in the software jModelTest2 (Darriba et al., 2012). For chloroplast regions, the model HKY+G was selected, with gamma shape

TABLE 1 | Genetic diversity and demographic parameters for 21 populations of *Tabebuia aurea* in Brazil for combined cpDNA data and for ITS nrDNA.

| Pop | cpDNA | | | | | ITS | | | | | Combined cpDNA and ITS nrDNA | | | | |
|---------|-------|-----|-------|-----------------|------------|-------|-----------------|------------|-----------------|------------------------|------------------------------|---------------------|-------------------|----------------|--|
| | N | k | h | π (SD) | π (SD) | k | h | π (SD) | θ | θ -95% interval | N_e | N_e -95% interval | g | g-95% interval | |
| AGE | 16 | 2 | 0.441 | 0.0127 (0.0067) | 1 | 0.000 | 0.0000 | 0.00117 | 0.00031-0.00218 | 1629.7 | 425.9-3020.8 | 114.78 | -491.64 to 898.71 | | |
| ARA | 14 | 1 | 0.000 | 0.0000 | 1 | 0.000 | 0.0000 | 0.00019 | 0.00009-0.00038 | 276.2 | 37.2-520.3 | 259.48 | -442.66 to 976.39 | | |
| BAG | 18 | 1 | 0.000 | 0.0000 | 1 | 0.000 | 0.0000 | 0.00012 | 0.00001-0.00059 | 165.7 | 14.7-819.7 | 398.88 | -356.14 to 990.39 | | |
| BAR | 12 | 1 | 0.000 | 0.0000 | 1 | 0.000 | 0.0000 | 0.00006 | 0.00001-0.00014 | 82.8 | 14.7-187.7 | 205.66 | -429.11 to 996.42 | | |
| BOD | 17 | 1 | 0.000 | 0.0000 | 1 | 0.000 | 0.0000 | 0.00005 | 0.00001-0.00011 | 68.5 | 14.5-155.7 | 288.64 | -435.94 to 990.31 | | |
| CAC | 16 | 6 | 0.542 | 0.0842 (0.0428) | 1 | 0.000 | 0.0000 | 0.00263 | 0.00058-0.00606 | 3646.5 | 802.1-8419.0 | 142.59 | 40.54 to 847.62 | | |
| CHG | 15 | 5 | 0.617 | 0.0334 (0.0172) | 3 | 0.514 | 0.0011 (0.0001) | 0.00409 | 0.00087-0.00735 | 5683.2 | 1209.5-10202.8 | 118.99 | 98.60 to 426.76 | | |
| FAT | 22 | 3 | 0.834 | 0.0099 (0.0053) | 3 | 0.256 | 0.0319 (0.0016) | 0.00277 | 0.00146-0.00514 | 3845.3 | 2028.2-7143.2 | 30.04 | 9.86 to 237.16 | | |
| GSV | 8 | 1 | 0.000 | 0.0000 | 1 | 0.000 | 0.0000 | 0.00062 | 0.00009-0.00166 | 865.5 | 135.9-2309.7 | 309.98 | -479.87 to 895.55 | | |
| NIQ | 14 | 1 | 0.000 | 0.0000 | 1 | 0.000 | 0.0000 | 0.00006 | 0.00001-0.00027 | 76.8 | 14.9-367.9 | 116.62 | -496.46 to 936.36 | | |
| PAN | 16 | 1 | 0.000 | 0.0000 | 1 | 0.000 | 0.0000 | 0.00011 | 0.00001-0.00028 | 148.9 | 14.8-394.2 | 197.63 | -487.72 to 902.18 | | |
| PNE | 16 | 1 | 0.000 | 0.0000 | 1 | 0.000 | 0.0000 | 0.00027 | 0.00004-0.00059 | 378.3 | 48.9-820.2 | -328.35 | -498.69 to -57.59 | | |
| POT | 16 | 1 | 0.000 | 0.0000 | 1 | 0.000 | 0.0000 | 0.00004 | 0.00001-0.00009 | 60.4 | 14.0-129.9 | 264.05 | -468.77 to 936.29 | | |
| PTU | 16 | 3 | 0.792 | 0.0274 (0.0141) | 1 | 0.000 | 0.0000 | 0.00230 | 0.00107-0.00488 | 3199.9 | 1482.5-6782.5 | 90.38 | 3.53 to 204.53 | | |
| SAF | 16 | 7 | 0.771 | 0.0137 (0.0072) | 1 | 0.000 | 0.0000 | 0.01660 | 0.00811-0.02999 | 23106.9 | 11268.5-41533.3 | 168.56 | 36.96 to 280.83 | | |
| SCA | 15 | 3 | 0.628 | 0.0148 (0.0077) | 1 | 0.000 | 0.0000 | 0.00472 | 0.00205-0.00882 | 6561.8 | 2850.7-12254.4 | 338.21 | 9.09 to 752.49 | | |
| SDO | 03 | 1 | 0.000 | 0.0000 | 1 | 0.000 | 0.0000 | - | - | - | - | - | - | | |
| SEC | 04 | 1 | 0.000 | 0.0000 | 3 | 0.500 | 0.0019 (0.0003) | - | - | - | - | - | - | | |
| SUM | 14 | 1 | 0.000 | 0.0000 | 1 | 0.000 | 0.0000 | 0.00149 | 0.00068-0.00234 | 2070.7 | 947.8-3245.8 | -30.15 | -98.82 to 41.06 | | |
| STZ | 17 | 1 | 0.000 | 0.0000 | 2 | 0.471 | 0.0009 (0.0001) | 0.00024 | 0.00004-0.00054 | 334.9 | 53.3-753.6 | 96.21 | -498.38 to 920.54 | | |
| VIB | 16 | 1 | 0.000 | 0.0000 | 2 | 0.125 | 0.0001 (0.0000) | 0.00024 | 0.00005-0.00054 | 333.5 | 70.6-747.7 | -269.59 | -419.58 to -98.29 | | |
| Mean | - | 2.1 | 0.220 | 0.0093 (0.0197) | 1.4 | 0.089 | 0.0006 (0.0025) | - | - | 2537.4 | - | - | - | | |
| SD | - | 1.9 | 0.329 | - | 0.4 | 0.158 | - | - | - | 5103.8 | - | - | - | | |
| Overall | 285 | 26 | 0.933 | 0.0466 (0.0223) | 8 | 0.267 | 0.0186 (0.0096) | 0.02940 | 0.02430-0.03510 | 40788.9 | 33813.9-48715.3 | 70.49 | 7.22 to 822.84 | | |

N, sample size; *k*, number of haplotypes; *h*, nucleotide diversity; π , nucleotide diversity; π , standard deviation; θ , coalescent parameter; N_e , effective population size; *g*, exponential growth parameter (all values are not significantly different from zero based on the credibility interval around the estimates); 95% CI is the credibility interval at 95%.

equal to 1.87. For ITS, the evolutionary model JC was selected.

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 genome, $\theta = 4\mu N_e$, for diploid genome), the exponential growth rate (g , where $\theta_t = \theta_{now} \exp(-gt\mu)$ and t is time in mutational unit), and the immigration rate among all population pairs ($M = 2N_e m/\theta$, for haploid genome, $M = 4N_e m/\theta$, for diploid genome). Estimations were based on Bayesian model using the Markov Chain Monte Carlo (MCMC) approach implemented in Lamarc 2.1.9 software (Kuhner, 2006). Because of the high number of populations, to estimate growth rate, we constrained migration (maintained migration constant), and migration was estimated in independent runs. 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 the 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, and combined results were then generated. Results were considered when ESS \geq 200 (effective sample size) and when marginal posterior probability densities were unimodal and converged among runs. Because of the low sample size, demographic parameters were not estimated for populations SDO and SEC. Effective population size was obtained from θ (Kingman, 1982) using a generation time of 12 years (based on flowering time on permanent plots; RG Collevatti, unpublished data).

Then, we performed a Coalescent Extended Bayesian Skyline Plot (EBSP) analysis (Heled and Drummond, 2008) implemented in BEAST 1.8.0 (Drummond and Rambaut, 2007) which calculates the effective population size (N_e) through time to better understand changes in population size, 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. Mutation rates for both chloroplast and ITS regions were the same used for a taxonomic related species *Tabebuia impetiginosa* (Collevatti et al., 2012a). Four independent analyses were run for 30 million generations. Convergence and stationarity were checked, and the independent runs were combined using the software Tracer v1.6. Results were considered when ESS \geq 200. We also inspected lineage (haplotype) diversification using Lineage Through Time (LTT) analysis to provide clues of major divergence timing. We performed the Fu's FS tests of neutrality implemented in Arlequin 3.11. Negative values for FS indicate an excess of rare alleles or new mutations in the genealogy resulting from either population expansion or selective sweeps.

Lineages Dispersal

To reconstruct the spatio-temporal history of lineage dispersal, we used the relaxed random walk model (RRW, Lemey et al., 2009, 2010) implemented in the software BEAST that analyses molecular sequence evolution, demographic model, and lineage dispersal in space and time simultaneously. RRW infers continuous phylogeographical diffusion while simultaneously reconstructing the evolutionary history in time

(Lemey et al., 2010). More specifically, the approach identifies the phylogeographical diffusion processes by stochastically selecting a diffusion rate scalar on each branch of the rooted phylogeny from an underlying discretized rate distribution while running a MCMC. Because this framework is based on stochastic models, it naturally accesses the uncertainties along the ancestral state reconstructions and the underlying phylogeographical process (Lemey et al., 2009, 2010).

To build RRW, we used both sequence partitions with unlinked priors but sharing the same location rate matrix (Lemey et al., 2009, 2010), including the rate for Bayesian Stochastic Search Variable Selection procedure (BSSVS), which considers a limited number of rates (at least $k-1$) to explain the phylogenetic diffusion process. The sampling locality was added as discrete character states ($k = 21$ localities). For the diffusion process, we used the 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 tree priors, we used the Coalescent GMRF Bayesian Skyride model (Minin et al., 2008) and for location state rate the prior CTMC Rate Reference (Ferreira and Suchard, 2008). Four runs were performed with 30 million generations, and stability was analyzed using Tracer 1.6, and results were considered when ESS \geq 200. The annotate tree (maximum clade credibility tree) was generated with 10% of burnin. The spatio-temporal reconstruction was performed using SPREAD 1.0.6 (Bielejec et al., 2011). We also analyzed the well-supported transition rates using the Bayes factors (BF) test implemented in SPREAD. Transitions rates between localities were considered only for BF $>$ 8.0.

Coalescent Tree and Time to Most Recent Common Ancestor

We obtained a coalescent tree based on Bayesian coalescent analysis implemented in the software BEAST 1.8.0 to estimate time to most recent common ancestor (TMRCA). For this analysis, we included the outgroup sequences from *T. impetiginosa*, *T. chrysotrica*, and *C. antisiphilitica*. We used the same priors of EBSP, except that for tree prior we used Coalescent Expansion Growth based on the results of EBSP and ENM (see Results below). MCMC conditions and number of runs also remained unchanged. The independent runs were analyzed using Tracer 1.6 and results were considered when ESS \geq 200. We also ran an empty alignment (sampling only from priors) to verify the sensitivity of results to the given priors. The analysis showed that our data is informative because posterior values (e.g., posterior probability) were different from those obtained from empty alignment.

Ecological Niche Models

Occurrence Records and Environmental Layers

We obtained 237 occurrence records of *T. aurea* across Neotropics from the online databases GBIF (Global Biodiversity Information Facility <http://www.gbif.org/>) and Species Link (<http://splink.cria.org.br/>), which were mapped in a grid of cells of $0.5^\circ \times 0.5^\circ$ (longitude \times latitude), corresponding to 55 km

at the equator, encompassing the entire Neotropics (Table S2 in Appendix S1, Figure S1 in Appendix S2).

The environmental layers were represented by five bioclimatic variables (annual mean temperature, mean diurnal 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 in the EcoClimate database (www.ecoclimate.org), along with subsoil pH (30–100 cm; from Harmonized World Soil Database—version 1.1, FAO/IIASA/ISRIC/ISS-CAS/JR, 2009). We assume subsoil pH to be constant between LGM and pre-industrial, which was used in ENMs as a “constraint variable” to better model the environmental preferences of *T. aurea*. The bioclimatic variables were obtained for pre-industrial (representing current climate conditions), mid-Holocene (6 ka) and Last Glacial Maximum (LGM; 21 ka) from four coupled Atmosphere-Ocean General Circulation Models (AOGCM): CCSM4, CNRM-CM5, MIROC-ESM, and MRI-CGCM3 (Table S3).

Palaeodistribution Modeling

Estimates of the current and past potential distribution of *T. aurea* were obtained using 13 algorithms by modeling its ecological niche (Table S4). When needed, we randomly selected pseudo-absences across the Neotropical grid cells keeping prevalence equal to 0.5 (see Stokland et al., 2011). The ENMs were then built on 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 and New, 2006; Diniz-Filho et al., 2009).

For each algorithm and AOGCM, models were built 50 times. After eliminating the models with poor performance (True Skill Statistics—TSS - < 0.5, Allouche et al., 2006), we computed the consensus maps using thresholds established by the ROC curve and the frequency of occurrence was used here as a measure of suitability for *T. aurea* across the Neotropical grid cells.

Next, we applied a hierarchical ANOVA using the predicted suitability from all models (13 ENMs × 4 AOGCMs × 3 Times) as a response variable to disentangle the effects of climate change on species distribution through the time from predictive uncertainties in the potential distribution due to modeling 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 × ENM; Sokal and 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 ≥ 0.5 during the three time periods). The threshold of 0.5 corresponds to the 10th percentile of the distribution of suitability values

related to the species occurrence records, i.e., we excluded the lowest 10% suitability values from the species occurrence records. We performed sensitivity analyses using different thresholds, 0.4, 0.6, and 0.7, corresponding to the 5th, 12th, and 17th percentiles, respectively, and the results were virtually the same as using threshold 0.5. The result differed only for a threshold > 0.7, resulting in a small refugium in central-west Brazil.

Setting Demographic Scenarios

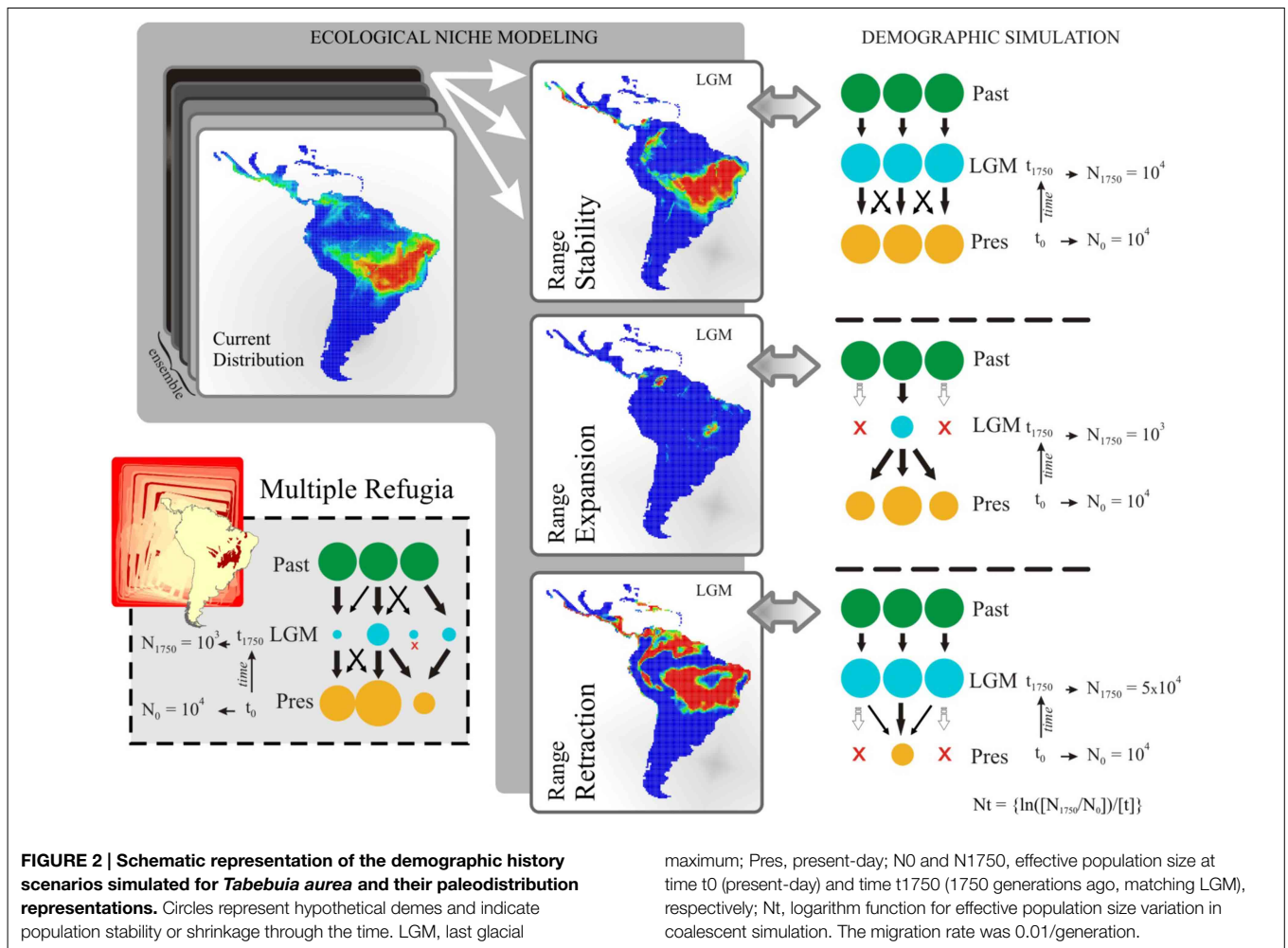
The ENMs resulted in 52 predictive maps (13 ENMs × 4 AOGCMs) for each time period, which were used to support alternative demographic scenarios. By keeping constant the ENM and AOGCM across time periods, we computed the range shift across the last glacial cycle (i.e., the difference in predicted range size between current and LGM—21 ka), and classified the 52 predictive maps according to three general demographic scenarios: (i) “Range Stability,” no difference in range size through time; (ii) “Range Expansion,” range size was lower at LGM than in present-day; (iii) “Range Retraction,” range size was larger at LGM than in present-day. Map classification for coalescent simulation was performed only for the time slice from LGM to present-day due to the TMRCA for the analyzed genome regions (~4.4 Ma; see results below).

Along with the three hypotheses supported by ENMs (see above), a fourth biogeographic hypothesis of “Multiple Refugia” was derived from paleoclimatic (e.g., Ab’Sáber, 2000) and paleovegetational reconstructions (e.g., Salgado-Labouriau, 1997).

Demographic History Simulation

The demographic scenarios were simulated based on coalescent analysis (Kingman, 1982) implemented in the software ByeSSC (Excoffier et al., 2000), following the framework described in Collevatti et al. (2012a, 2013b). For model calibration, we used the demographic parameters estimated with Lamarc software and the same priors for sequence evolution used in Bayesian analyses (see above). The number of generations until LGM (21 ka) was calculated using the generation time of *T. aurea* (12 years, RG Collevatti, unpublished data).

Population dynamics were simulated backwards, with 21 demes, from t0 (present) to t1750 generations ago (at 21 ka), with sizes $N_t = \{\ln(N_1/N_0)/t\}$. At t0, all demes had the same population size N_0 and reached N_{1750} according to our theoretical expectation for each demographic scenario (see **Figure 2** for details). Given the high variation in *T. aurea* effective population sizes (see **Table 1**), we performed simulations with different initial deme sizes, $N_0 = 100$, $N_0 = 1000$, and $N_0 = 10,000$ for all scenarios. Because coalescent simulations run backward through time, negative growth implies population expansion from the past, and a positive growth a population smaller now than in the past. The demographic scenario predicted by “Range Expansion” was then simulated by applying an exponentially negative population growth from present to 21 ka, reaching $N_{1750} = 1000$, if $N_0 = 10,000$, or $N_{1750} = 100$, if $N_0 = 1000$, and $N_{1750} = 10$, if $N_0 = 100$. In opposition, the “Range Retraction” scenario was simulated



applying exponentially positive population growth during the same period, attaining $N_{1750} = 50,000$, if $N_0 = 10,000$, or $N_{1750} = 5000$, if $N_0 = 1000$ and $N_{1750} = 500$, if $N_0 = 100$. 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/generation chance that each lineage in deme x will migrate to deme 1. We also simulated different values of migration rate. Values < 0.01 were not sufficient to show any demographic variation at the time scale we are working and values > 0.1 retrieved equal likelihoods for all models. For the “Range Retraction” hypothesis, we considered that each lineage in deme x will migrate to deme 1 and then shrink until extinction. For the “Multiple Refugia” scenario we considered a finite island model; i.e., all current populations are descendant from lineages originally in the demes at t_{1750} generations ago, meaning that while the tree builds back through time, there is a 0.01/generation chance that each lineage will migrate to deme x .

The simulated haplotype and nucleotide diversities for the four alternative demographic scenarios, across 2000 simulations, were compared with the empirical haplotype and nucleotide diversity (mean for the 21 populations).

One-tailed probability (P) and Akaike Information Criterion (AIC) were estimated for each comparison. The log-likelihood was estimated as the product of the height of the empirical frequency distribution at the observed value of diversity by the maximum height of the distribution (see BayeSSC website www.stanford.edu/group/hadlylab/ssc/index.html). AIC was transformed into AIC weight of evidence ($AICw$), given by $\exp[-0.5(AIC - AIC_{min})]$ (see Burnham and Anderson, 2002), from which we obtained ΔAIC ; i.e., the difference of $AICw$ between each model and the best model. Models with $\Delta AIC < 2$ were considered as equally plausible to explain the observed pattern (Zuur et al., 2009).

Results

Genetic Diversity and Structure of Populations

The combined data of chloroplast intergenic spacers generated 2330 bp and the nrDNA ITS, 567 bp (see Appendix S3 for GenBank accession numbers), resulting in 2627 bp when microsatellites and ambiguous alignment were eliminated (2100 bp for chloroplast and 527 for ITS). Although 26 different chloroplast and 8 ITS haplotypes were found for the

285 individuals of *T. aurea*, most populations presented low haplotype and nucleotide diversities (Table 1). Two chloroplast haplotypes were very widespread (Figures 1, 3), H1 and H14, and one ITS haplotype (H1) was found in all populations, but CHG and SDO (Figures 1, 3). The analysis of molecular variance showed significant genetic differentiation among populations for both cpDNA ($F_{ST} = 0.773$; $p < 0.001$) and nuclear ITS ($F_{ST} = 0.966$; $p < 0.001$). Although pairwise genetic differentiation was significant among all pairs of populations for cpDNA, differentiation for nuclear ITS was mainly due to populations CHG, SDO, and SEC (Table S6). The Mantel test showed weak (chloroplast, $r^2 = 0.069$, $p = 0.026$) or no association between the geographical distance and the genetic differentiation among pairs of populations (ITS, $r^2 = 0.032$, $p = 0.159$).

Phylogeographic Reconstruction Population Demography and Time to Most Recent Common Ancestor

Coalescent analyses performed with Lamarc software showed low values of mutation parameter θ for each and overall population ($\theta = 0.029$, Table 1) and negligible gene flow among all population pairs (less than 1.0 migrant per generation, Tables S7, S8). Our results showed population growth through time, with positive values of g for some populations and overall populations (overall $g = 70.49$; Table 1).

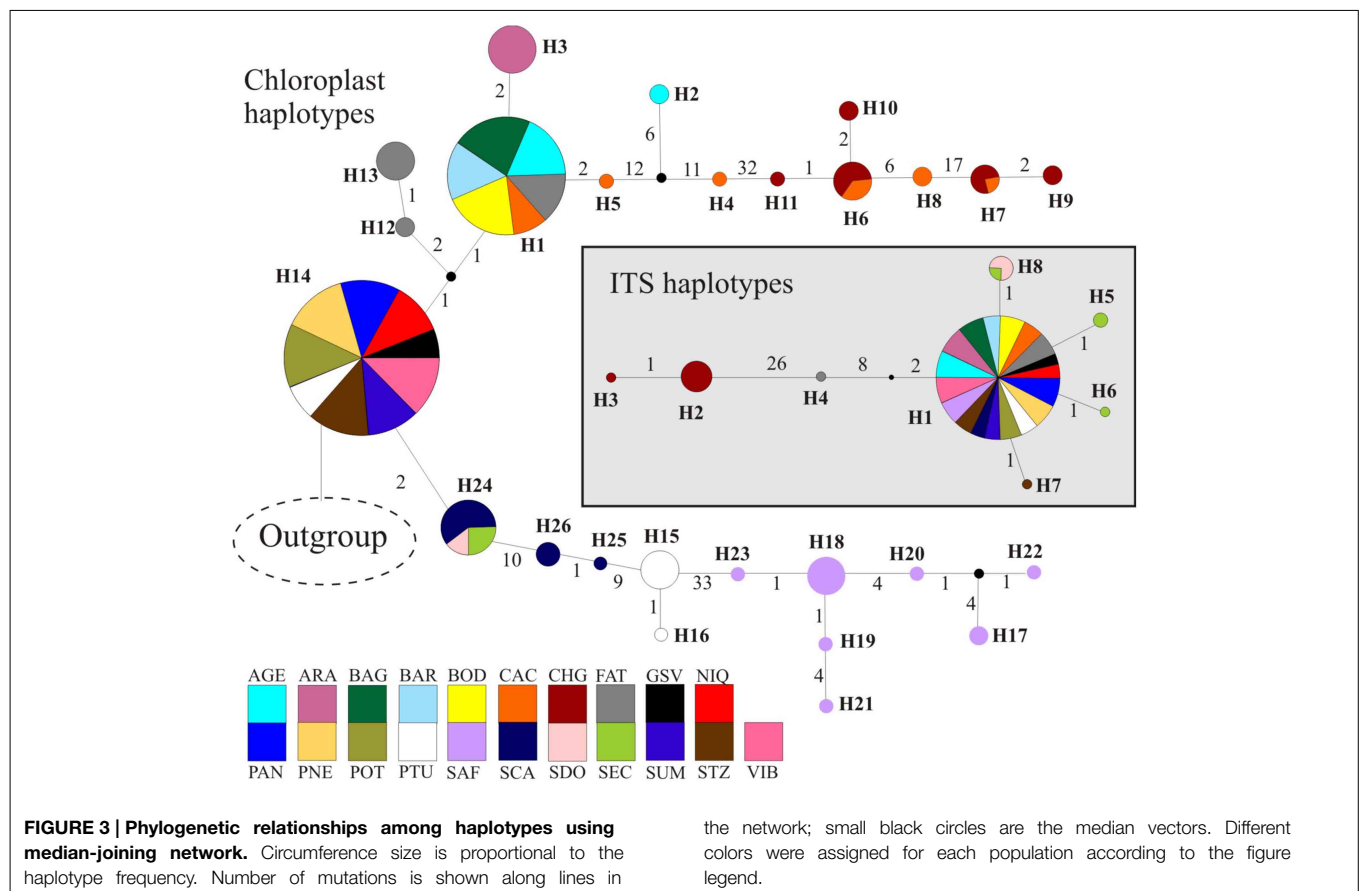
The number of lineages increased almost linearly with a higher rate in the last ~ 1.0 Ma (Figure S2). Lineages of *T. aurea*

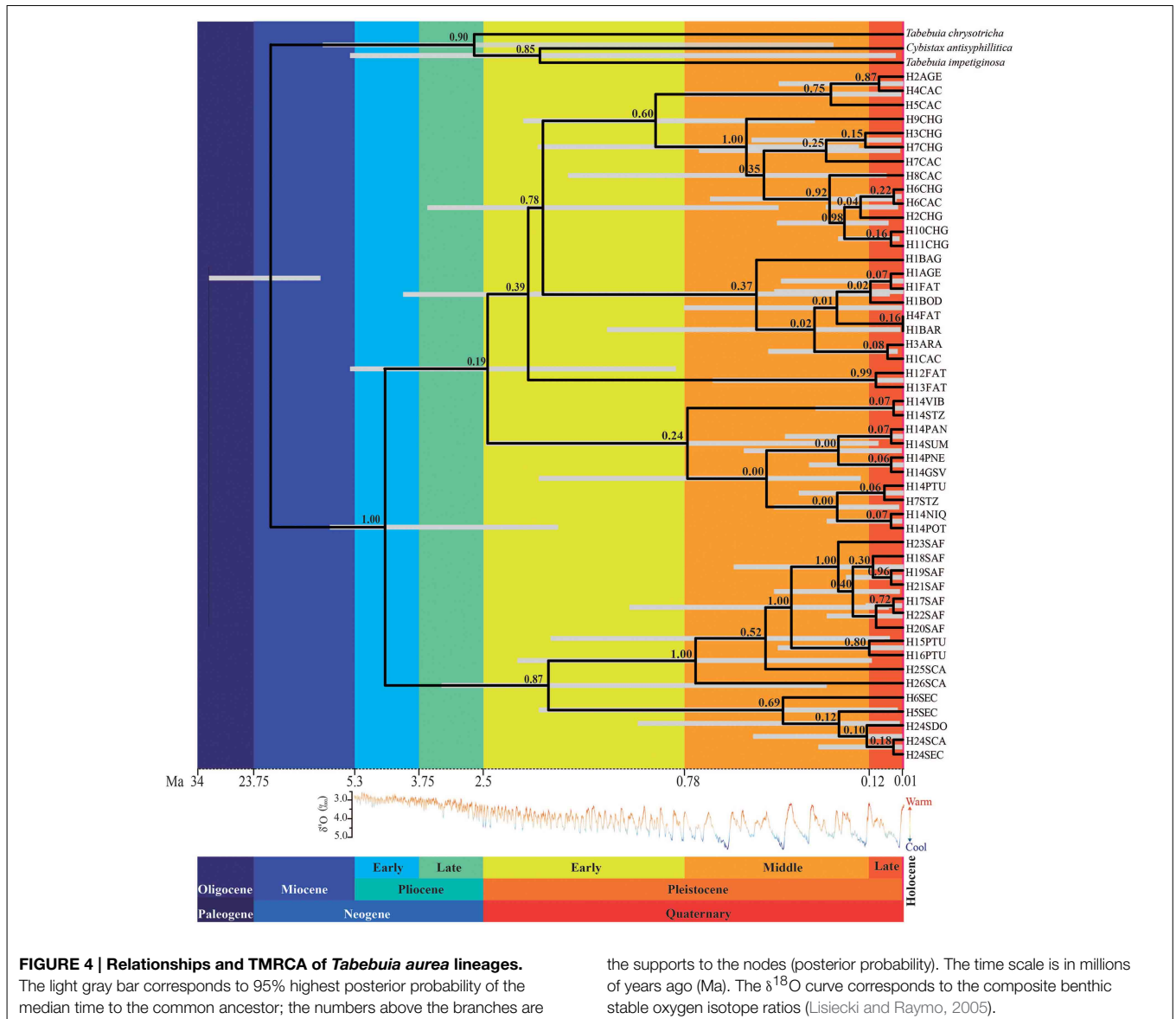
started to diverge at $\sim 4.4 \pm 2.6$ Ma (Figure 4), although major divergence events dated from the Pliocene/Pleistocene transition ($\sim 2.5 \pm 2.3$ Ma) with increasing divergence rates after ~ 1.2 Ma, coinciding with the increasing in lineage diversification (Figure S2).

The Extended Bayesian Skyline Plot also showed increasing population growth through the last 1.2 Ma, especially in the last 100 ka (Figure 5). Fu' test was significant ($FS = -5.978$, $p = 0.007$) and potentially indicates a demographic expansion or departure from neutrality.

Lineages Dispersal

The phylogeographic reconstruction of lineage dispersal showed the two most probable ancestral locations for *T. aurea* lineages, one in the west (CAC) and other in the southeast Brazil (SCA) (Figure 5). These lineages started to spread at ~ 1.55 – 1.25 Ma in two main directions: from CAC, mainly to the central-west Brazil, and from SCA toward the east and northeast (Figure 6). However, most dispersal events occurred after 900 ka overlapping the temporal increase in population expansion, as shown by EBSP and the diversification timing of new lineages. From 25 dispersal events, 18 occurred during interglacial or warming periods and just 7 in glacial or cooling periods (see Figure S4). Most dispersal events during interglacial and warming periods ($n = 14$) occurred toward populations in northern Brazil, corresponding to the spatial displacement in suitable climatic conditions as showed by ENMs (see below). The mean rate of





dispersal was 1.17 km/yr ($SD = 0.978$, 95% HPD interval 0.0005–3.517 km/yr), and the Bayes factor showed that most links among localities are well-supported ($BF > 8.0$).

Ecological Niche Models

ENMs predicted that *T. aurea* was potentially distributed across central and northeast Brazil through the last glaciation (Figure 7A). The highest levels of suitability were restricted to central Brazil, followed by a spatial displacement toward northeast Brazil during the mid-Holocene (6 ka, Figure 7B), and also expanding more recently toward the west (present-day, Figure 7C). Besides the spatial displacements, range size increased through time (see Figure S3). In addition, a wide region across central Brazil probably acted as historical refugium maintaining populations of *T. aurea* during the climate changes throughout the last glaciation (Figure 7D).

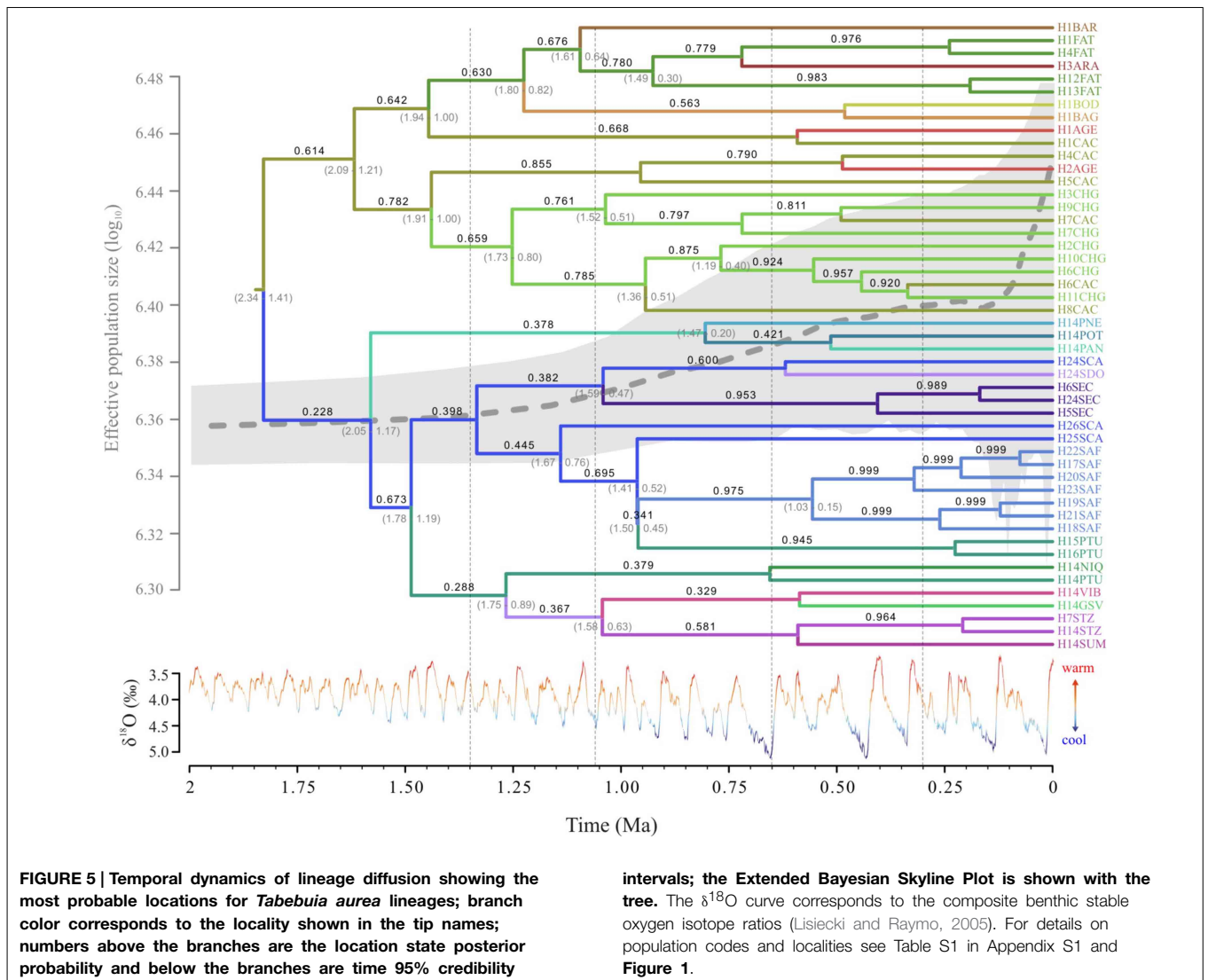
the supports to the nodes (posterior probability). The time scale is in millions of years ago (Ma). The $\delta^{18}\text{O}$ curve corresponds to the composite benthic stable oxygen isotope ratios (Lisiecki and Raymo, 2005).

The models were well evaluated from TSS (Table S5). The analysis of uncertainty using hierarchical ANOVA showed lower proportional variance from modeling method (ENMs) than time component (Table S9, Figure S5) indicating that the ENMs were able to detect the effects of climate changes on the distribution dynamics of *T. aurea* through the last glaciation, despite the AOGCM variation.

The scenario of “Range Expansion” was the most frequent hypothesis from ENM predictions (65.3% of the 52 maps, Table 2, see also Table S10), followed by “Range Stability” (30.7%) and “Range Retraction” (4%). Moreover, the average range shift predicted by ENMs systematically follows what is expected by general demographic scenarios (Figure S6).

Demographic History Simulations

Simulations using $N_0 = 100$ presented all values of haplotype and genetic diversity lower than those observed for *T. aurea*



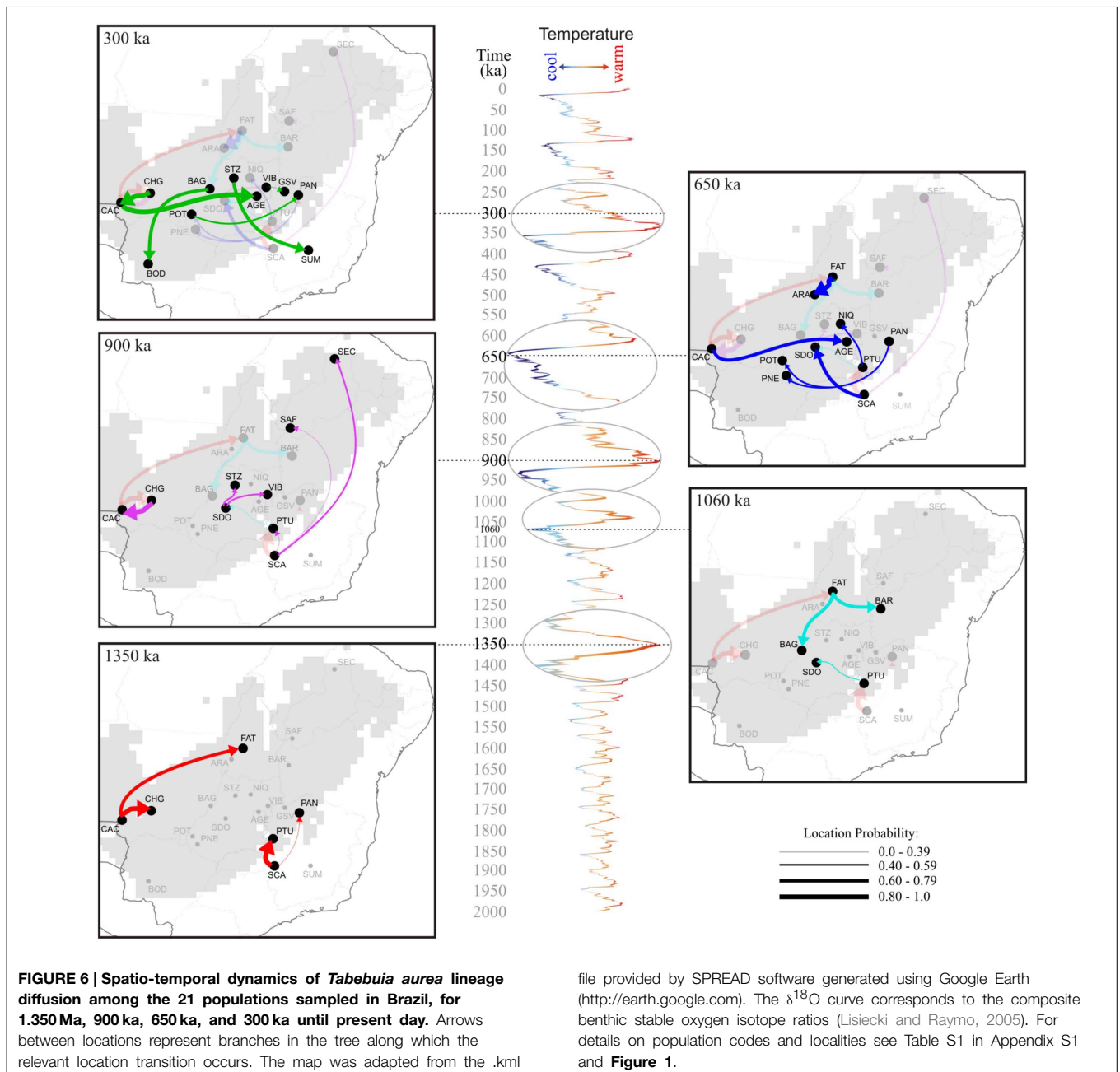
for all demographical scenarios. Simulations using $N_0 = 1000$ and $N_0 = 10,000$ retrieved the same final results; i.e., the same most likely scenario, thus we show the results only for $N_0 = 10,000$. The scenario of “Range Expansion” was the most likely hypothesis to predict the observed genetic parameters of *T. aurea*, using either one-tailed probability or AIC_w criteria (Table 2). The simulations following the “Range Retraction” scenario retrieved all values lower (see Table S11 for details) than the observed mean haplotype and nucleotide diversities for the 21 populations of *T. aurea* (Table 1), meaning that AIC and P could not be estimated. In contrast, the scenarios of “Range Stability” and “Multiple Refugia” retrieved higher values of genetic diversity than those observed for most populations (Table S11).

Discussion

Our integrative approach shows clear convergence among phylogeographic inferences from multiple models. The reconstruction of lineages dispersal inferred using the RRW, the

distribution dynamics retrieved by ENMs and the coalescent simulation predicted similar demographic and dispersal dynamics for *T. aurea* through the Quaternary.

Our results show that *T. aurea* lineages cyclically dispersed from the West toward Central-West Brazil, and from the Southeast toward the East and Northeast Brazil, following the paleodistribution dynamics showed by ENMs through the last glacial cycle. The dispersal results indicated that populations currently occurring in northern and northeastern Brazil received migrants mainly from the southwest and southeast, respectively (e.g., see the arrow direction at 1350 and 900 ka in Figure 6). Such dispersal routes have most likely occurred because suitable climatic conditions were available across a wider region during warming phases, allowing spatial displacements of the species’ geographical distribution toward northeastern Brazil. Diffusion models suggested that an inverse dynamic occurred during cooling phases (e.g., 1060 and 650 ka, Figure 6), when lineages dispersed in the opposite direction, from north-northeast toward central hot and drier climate in the central-west Brazil, thus



retaining more suitable niches for *T. aurea* during the cooler phases (see Collevatti et al., 2014). Yet, in intermediate warmer phases, the findings showed that lineages dispersed in both directions, mainly across central Brazil (300 ka), in response to slight changes in climatic suitability, like ENM predictions between mid-Holocene and present-day. A potential continuous historical refugium across the Neotropical savannas may be the key factor allowing uninterrupted lineage dispersal among populations of central Brazil, as predicted by the RRW (at least since the middle Pleistocene; last 900 ka). It is important to note that the dispersal of lineages already present in a population would not be detected as a new gene flow event. Thus, population

connectivity could be maintained after 300 ka but our results show no dispersal of new lineages.

Concurrently, coalescent simulations showed population expansion through time (i.e., from LGM to present-day) as the most likely general scenario among alternative demographic hypotheses, matching our theoretical prediction, ENMs and extended Bayesian Skyline Plot results. Because favorable climatic conditions for *T. aurea* (i.e., hot and drier climates, see Collevatti et al., 2014) were spatially more restricted during glacial than interglacial periods in Neotropics, smaller effective population sizes during the cold phases of the last glacial cycle is a consequence of the retraction in geographical range when species

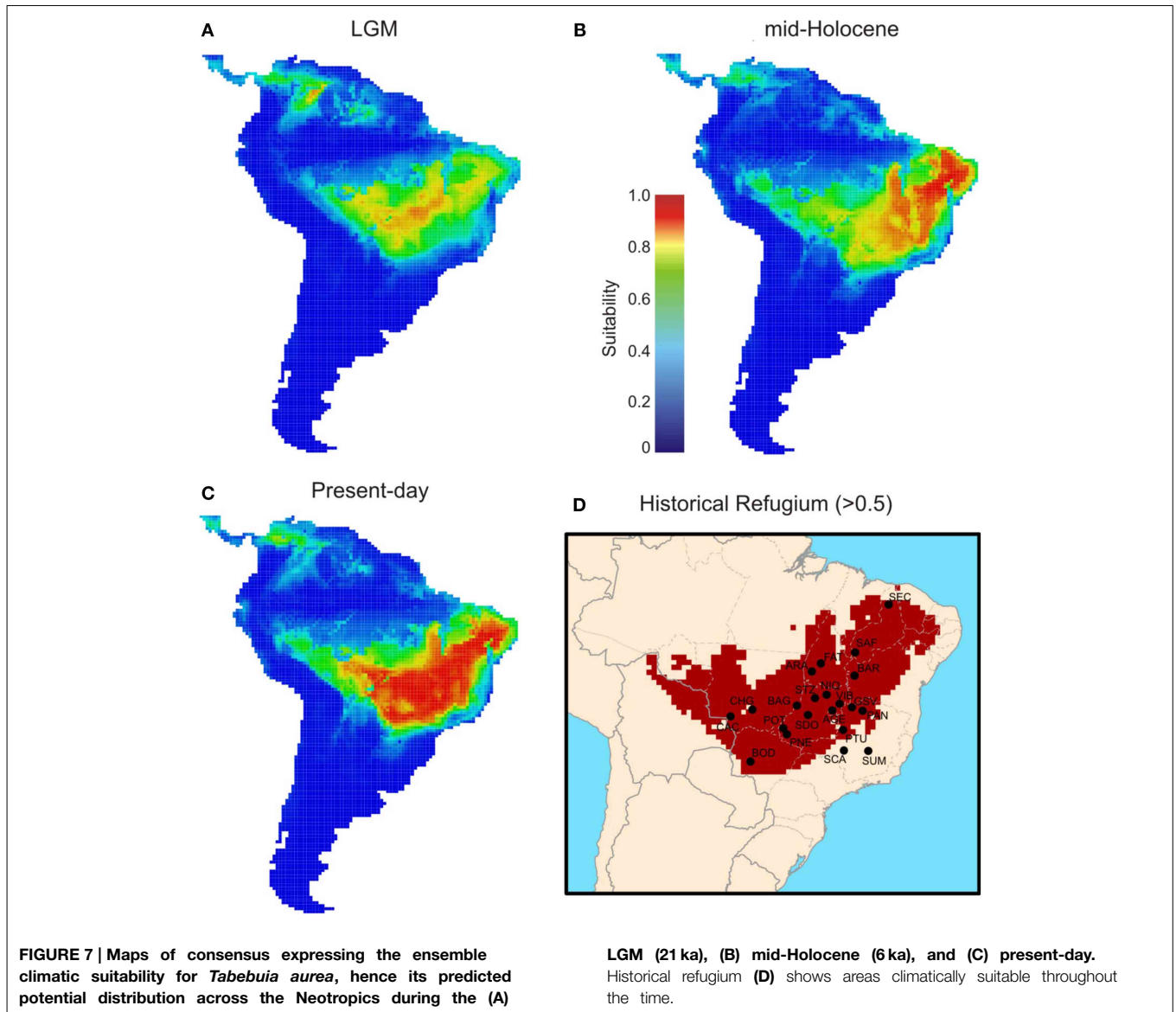


TABLE 2 | Comparison of the four demographic scenario models in retrieving the haplotype (h) and nucleotide (π) diversity observed for *Tabebuia aurea*, obtained from 2000 simulations using the software BayeSSC.

| Models | Chloroplast | | | | | | ITS | | | | | | ENM % |
|------------------|--------------|-------|--------|-------|-------|-------|--------------|-------|--------|-------|-------|-------|-------|
| | ΔAIC | | $AICw$ | | P | | ΔAIC | | $AICw$ | | P | | |
| | h | π | h | π | h | π | h | π | h | π | h | π | |
| Stability | 11.9 | 9.2 | 0.002 | 0.005 | 0.005 | 0.001 | 7.3 | 4.2 | 0.022 | 0.083 | 0.001 | 0.002 | 30.7 |
| Retraction | – | – | – | – | – | – | – | – | – | – | – | – | 4.0 |
| Expansion | 0.0 | 0.0 | 0.789 | 0.571 | 0.991 | 0.959 | 0.0 | 0.0 | 0.865 | 0.671 | 0.971 | 0.876 | 65.3 |
| Multiple Refugia | 2.7 | 0.5 | 0.209 | 0.424 | 0.435 | 0.354 | 4.1 | 2.3 | 0.113 | 0.243 | 0.743 | 0.623 | – |

ΔAIC , the difference of $AICw$ between each model and the best model; $AICw$, Akaike Information Criterion (AIC) weights; P , one-tailed probability of not rejecting the model. ENM %, the percentage of the 52 paleodistribution maps that represent each demographic scenario. See **Figure 3** for details about the demographic scenarios. h , haplotype diversity; π , nucleotide diversity.

spatially tracked suitable habitats. Besides, the increasing number of lineages, especially in the last ~ 1.25 Ma, may have resulted from subsequent population expansions across cyclic interglacial periods, during which wider suitable areas were available for colonization. Because the probability of coalescence is inversely related to the number of gene copies (Kingman, 1982), most coalescence would occur just before demographic expansion, when populations had smaller effective population sizes (see (Excoffier et al., 2009) for a review). After diversification, some lineages became widespread through dispersal during interglacial periods (Figure 6), most likely due to range expansion.

Lineage dispersal was mainly from three populations: CAC, SCA, FAT, and PTU (see Figure 5). This diffusion pattern explains the widespread geographical distribution of some haplotypes (e.g., H1 and H14, see Figures 1, 3) and the restricted geographical distribution of other haplotypes that diverged more recently (e.g., H2, H17, H18, see Figures 1, 3), leading to high population differentiation. Our results also evidenced *in situ* diversification of many haplotypes in SAF (see Figure 5) with restricted geographical distribution. This population is at the edge of the *T. aurea* historical refugium and is now an isolated savanna remaining in the transition between savanna and caatinga woodlands (Caatinga biome). Although our findings show possible connections among SAF and other populations (wide historical refugium and dispersal of H25 from SCA to SAF with posterior extinction), we hypothesize that the relative isolation may have precluded the dispersal of lineages from SAF to other populations. It is important to note that events of long distance dispersal, such as from SCA to SAF, are most likely due to stepping-stone dispersal and would be detected only with continuous spatial sampling. However, continuously sampling in large scale is hindered by the high savanna fragmentation in Brazil. Besides, estimates of the mean rate of dispersal equal to 1.17 km/yr using RRW is congruent with the pollen dispersal distance reported for *T. aurea* (~ 2.0 km, Braga and Collevatti, 2011). Gene dispersal estimate (~ 300 m, Braga and Collevatti, 2011) based on spatial genetic structure was lower than pollen dispersal. Although this method may underestimate seed dispersal distance, long-distance dispersal is mainly due to pollen dispersal, which may also explain widespread distribution of ITS haplotypes (see also the discussion below).

The consistence among predictions is still evidence of genetic connectivity. Despite the range and demographical expansions through the time, ENMs suggested that all populations, except for SUM and SCA, are inside the historical refugium; i.e., areas with high climatic suitability throughout time (Figure 7D). Such evidence suggests that even with the smaller geographical range during LGM and spatial displacement during warming phases, a wide climatically stable area was still available throughout the last glacial cycle, favoring 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 *T. aurea* populations, which is still supported by a high density of well-supported links among populations inside that historical refugium (Figure 6). Genetic connectivity may have also been maintained by the long-distance pollen and seed dispersal of *T. aurea* (Braga and Collevatti,

2011). Thus, the low haplotype diversity within a population may be an effect of selective sweeps. For instance, Fu's neutrality test was significant, and negative *FS* values potentially indicate selective sweeps of chloroplast genomes, as reported for other plant species (e.g., Kapralov and Filatov, 2007). In addition, despite differences in effective population size and inheritance, chloroplast genome showed higher diversity than ITS nrDNA, that has four times the effective size of chloroplast genome. This may be due to concerted evolution in nrDNA that may decrease genetic variation (Ohta, 1984). Complementarily, the cycles of range retraction and expansion during recurrent glaciations, may have caused the extinction of haplotypes in some populations and a spatial assortment decreasing genetic diversity, as suggested by similar distribution dynamics along correspondent glacial and interglacial phases from older glacial cycles, (see Hewitt, 1996; Excoffier and Ray, 2008; Arenas et al., 2012). For instance, the findings from RRW showed that the chloroplast haplotype H14 (ITS H1) dispersed from population PTU to SDO (~ 1250 – 900 ka, see Figure 6) and became extinct in SDO after dispersion to VIB and STZ (~ 1050 – 900 ka). Extinction of this haplotype may also have occurred in SCA, after dispersion to PTU and PAN (~ 1350 ka). In addition, recent anthropogenic disturbs may also have caused bottlenecks, decreasing genetic diversity in *T. aurea* populations because some populations, such as BAR, STZ, ARA, which are isolated remnant individuals in pasture or crop plantations have low genetic diversity. We believe that our results are not an artifact of sampling effort because populations with larger sample sizes had the same number of haplotypes as populations with smaller sample sizes.

Actually, the pattern of lineage spread and extinction may unravel the low genetic diversity in many populations, even in the center of species geographical range. In fact, populations in more instable areas in eastern and western edges of the historical refugium (CAC, CHG, PTU, SCA) have high genetic diversity. We hypothesize that these populations, at the edge of historical refugium 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 and Petit, 2005). Moreover, along with high genetic diversity, CAC, CHG, PTU, and SCA also present the oldest lineages sampled in this study (see Figures 4, 6).

Finally, the scenario of range and demographic expansion through time for *T. aurea* was also supported for other Brazilian savanna species such as *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). The few phylogeographical studies performed so far with savanna species (e.g., Ramos et al., 2007; Novaes et al., 2010, 2013; Collevatti et al., 2012b) have also shown high levels of genetic differentiation among populations and evidences of colonization of southeastern Brazilian savannas during warming interglacial periods of the Quaternary.

In conclusion, our findings suggest that the pattern of genetic diversity in *T. aurea* may be the outcome of population expansion through warming periods of the Quaternary, with recurrent events of spatial displacements along the cyclic events of glaciations. We are aware of the differences in the time

scale of predictions from different approaches, given that the RRW predicts spatial displacements occurring in a deeper time than ENMs. However, all approaches converge to consistent patterns: general dynamics of *T. aurea* following suitable climatic conditions toward the northern regions during warming periods in the early Quaternary, with retraction of populations during glacial phases, and new dispersal events at the middle and end of the Quaternary. Such consistent patterns from different methods and analyses indicate that our multi-model inference approach is dynamic and flexible, allowing the inclusion of new components to directly infer processes affecting phylogeographic patterns through time, and promises advances in hypothesis testing in phylogeography.

Author Contributions

RC and ML conceived the overall study. SR generated the genetic data. RC, LT, and ML analyzed the data and wrote the manuscript. All authors approved the final version of the manuscript.

References

- Ab'Sáber, A. N. (2000). Spaces occupied by the expansion of dry climates in South America during the Quaternary ice ages. *Rev. Inst. Geol.* 21, 71–78. doi: 10.5935/0100-929X.20000006
- Allouche, O., Tsoar, A., and Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* 43, 1223–1232. doi: 10.1111/j.1365-2664.2006.01214.x
- Araújo, M. B., and New, M. (2006). Ensemble forecasting of species distributions. *Trends Ecol. Evol.* 22, 42–47. doi: 10.1016/j.tree.2006.09.010
- Arenas, M., Ray, N., Currat, M., and Excoffier, L. (2012). Consequences of range contractions and range shifts on molecular diversity. *Mol. Biol. Evol.* 29, 207–218. doi: 10.1093/molbev/msr187
- Behling, H. (2003). Late glacial and Holocene vegetation, climate and fire history inferred from Lagoa Nova in the southeastern Brazilian lowland. *Veg. Hist. Archaeobot.* 12, 263–270. doi: 10.1007/s00334-003-0020-9
- Bielejec, F., Rambaut, A., Suchard, M. A., and Lemey, P. (2011). SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics* 27, 2910–2912. doi: 10.1093/bioinformatics/btr481
- Braga, A. C., and Collevatti, R. G. (2011). Temporal variation in pollen dispersal and breeding structure in a bee-pollinated Neotropical tree. *Heredity* 106, 911–919. doi: 10.1038/hdy.2010.134
- Burnham, K. P., and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach, 2nd Edn.* New York, NY: Springer.
- Carstens, B. C., and Richards, C. L. (2007). Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution* 61, 1439–1454. doi: 10.1111/j.1558-5646.2007.00117.x
- Collevatti, R. G., Lima-Ribeiro, M. S., Souza-Neto, A. C., Franco, A. A., and Terribile, L. V. (2012b). Recovering the demographical history of a Brazilian Cerrado tree species *Caryocar brasiliense*: coupling ecological niche modeling and coalescent analyses. *Nat. Conserv.* 10, 169–176. doi: 10.4322/natcon.2012.024
- Collevatti, R. G., Lima-Ribeiro, M. S., Terribile, L. C., Guedes, L. B., Rosa, F. F., Telles, M. P. et al. (2014). Recovering species demographic history from multi-model inference: the case of a Neotropical savanna tree species. *BMC Evol. Biol.* 14:213. doi: 10.1186/s12862-014-0213-0
- Collevatti, R. G., Telles, M. P. C., Nabout, J. C., Chaves, L. J., and Soares, T. N. (2013b). Demographic history and the low genetic diversity in *Dipteryx alata* (Fabaceae) from Brazilian Neotropical savannas. *Heredity* 111, 105. doi: 10.1038/hdy.2013.23

Acknowledgments

This work was supported by a competitive grant from CNPq to RGC (project no. 471085/2009-0), the research network GENPAC (CNPq/MCT/CAPES/FAPEG projects no. 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 Modeling (Table S3 in Appendix S1) for making available their model outputs. We also acknowledge Professor P. Lemey for advice on the use of random walk analyses and SPREAD software.

Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2015.00653>

- Collevatti, R. G., Terribile, L. C., Diniz-Filho, J. A., and Lima-Ribeiro, M. S. (2015). Multi-model inference in comparative phylogeography: an integrative approach based on multiple lines of evidence. *Front. Genet.* 6:31. doi: 10.3389/fgene.2015.00031
- Collevatti, R. G., Terribile, L. V., Lima-Ribeiro, M. S., Nabout, J. C., Oliveira, G., Rangel, T. F., et al. (2012a). A coupled phylogeographic and species distribution modeling approach recovers the demographic history of a Neotropical seasonally dry forest tree species. *Mol. Ecol.* 21, 5845–5863. doi: 10.1111/mec.12071
- Collevatti, R. G., Terribile, L. V., Oliveira, G., Lima-Ribeiro, M. S., Nabout, J. C., Rangel, T. F., et al. (2013a). Drawbacks to palaeodistribution modelling: the case of South American seasonally dry forests. *J. Biogeogr.* 40, 345–358. doi: 10.1111/jbi.12005
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772. doi: 10.1038/nmeth.2109
- Desfeux, C., and Lejeune, B. (1996). Systematics of Euro Mediterranean *Silene* (Caryophyllaceae): evidence from a phylogenetic analysis using ITS sequence. *Acad. Sci. Paris* 319, 351–358.
- Diniz-Filho, J. A. F., Bini, L. M., Rangel, T. F., Loyola, R. D., Hof, C., Nogués-Bravo, D., et al. (2009). Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography* 32, 897–906. doi: 10.1111/j.1600-0587.2009.06196.x
- Drummond, A. J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214. doi: 10.1186/1471-2148-7-214
- Excoffier, L., Foll, M., and Petit, R. J. (2009). Genetic consequences of range expansions. *Annu. Rev. Ecol. Syst.* 40, 481–501. doi: 10.1146/annurev.ecolsys.39.110707.173414
- Excoffier, L., Laval, G., and Schneider, S. (2005). Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1, 47–50.
- Excoffier, L., Novembre, J., and Schneider, S. (2000). SIMCOAL: a general coalescent program for simulation of molecular data in interconnected populations with arbitrary demography. *J. Hered.* 91, 506–509. doi: 10.1093/jhered/91.6.506
- Excoffier, L., and Ray, N. (2008). Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol. Evol.* 23, 347–351. doi: 10.1016/j.tree.2008.04.004
- Excoffier, L., Smouse, P. E., and Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131, 479–491.

- FAO/IIASA/ISRIC/ISS-CAS/JR. (2009). *Harmonized World Soil Database (version 1.1)*. Rome; Laxenburg: Food and Agriculture Organization of the United Nations (FAO); International Institute for Applied Systems Analysis (IIASA). Available online at: <http://www.fao.org/nr/land/soils/harmonized-world-soil-database/en/>
- Ferreira, M. A. R., and Suchard, M. A. (2008). Bayesian analysis of elapsed times in continuous-time Markov chains. *Can. J. Stat.* 36, 355–368. doi: 10.1002/cjs.5550360302
- Forster, P., Bandelt, H. J., and Röhl, A. (2004). *Network 4.2.0.1*. Available online at: <http://www.fluxus-engineering.com>
- Hampe, A., and Petit, R. J. (2005). Conserving biodiversity under climate change: the rear edge matters. *Ecol. Lett.* 8, 461–467. doi: 10.1111/j.1461-0248.2005.00739.x
- Heled, J., and Drummond, A. J. (2008). Bayesian inference of population size history from multiple loci. *BMC Evol. Biol.* 8:289. doi: 10.1186/1471-2148-8-289
- Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol. J. Linn. Soc. Lond.* 58, 247–276. doi: 10.1111/j.1095-8312.1996.tb01434.x
- Kapralov, M. V., and Filatov, D. A. (2007). Widespread positive selection in the photosynthetic Rubisco enzyme. *BMC Evol. Biol.* 7:73. doi: 10.1186/1471-2148-7-73
- Kingman, J. F. C. (1982). The coalescent. *Stochastic Processes Appl.* 13, 235–248. doi: 10.1016/0304-4149(82)90011-4
- Knowles, L. L., Carstens, B. C., and Keat, M. L. (2007). Coupled genetic and ecological-niche models to examine how past population distributions contribute to divergence. *Curr. Biol.* 17, 1–7. doi: 10.1016/j.cub.2007.04.033
- Kuhner, M. K. (2006). LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* 22, 768–770. doi: 10.1093/bioinformatics/btk051
- Kuhner, M. K. (2008). Coalescent genealogy samplers: windows into population history. *Trends Ecol. Evol.* 24, 86–93. doi: 10.1016/j.tree.2008.09.007
- Lemey, P., Rambaut, A., Drummond, A. J., and Suchard, M. A. (2009). Bayesian phylogeography finds its roots. *PLoS Comput. Biol.* 5:e1000520. doi: 10.1371/journal.pcbi.1000520
- Lemey, P., Rambaut, A., Welch, J. J., and Suchard, M. A. (2010). Phylogeography takes a relaxed random walk in continuous space and time. *Mol. Biol. Evol.* 27, 1877–1885. doi: 10.1093/molbev/msq067
- Lima, N. E., Lima-Ribeiro, M. S., Tinoco, C. F., Terribile, L. C., and Collevatti, R. G. (2014). Phylogeography and ecological niche modelling, coupled with the fossil pollen record, unravel the demographic history of a Neotropical swamp palm through the Quaternary. *J. Biogeogr.* 41, 673–686. doi: 10.1111/jbi.12269
- Lisiecki, E. L., and Raymo, M. E. (2005). A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}O$ records. *Paleoceanography* 20, PA1003. doi: 10.1029/2004PA001071
- Metcalfe, J. L., Prost, S., Nogués-Bravo, D., DeChaine, E. G., Anderson, C., Batra, P., et al. (2014). Integrating multiple lines of evidence into historical biogeography hypothesis testing: a *Bison bison* case study. *Proc. R. Soc. B* 281, 2013–2782. doi: 10.1098/rspb.2013.2782
- Minin, V. N., Bloomquist, E. W., and Suchard, M. A. (2008). Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol. Biol. Evol.* 25, 1459–1471. doi: 10.1093/molbev/msn090
- Nei, M. (1987). *Molecular Evolutionary Genetics*. New York, NY: Columbia University Press.
- Novaes, R. M., De Lemos Filho, J. P., Ribeiro, R. A., and Lovato, M. B. (2010). Phylogeography of *Plathymenia reticulata* (Leguminosae) reveals patterns of recent range expansion towards northeastern Brazil and southern Cerrados in Eastern Tropical South America. *Mol. Ecol.* 19, 985–998. doi: 10.1111/j.1365-294X.2010.04530.x
- Novaes, R. M., Ribeiro, R. A., Lemos-Filho, J. P., and Lovato, M. B. (2013). Concordance between phylogeographical and biogeographical patterns in the Brazilian Cerrado: diversification of the endemic tree *Dalbergia miscolobium* (Fabaceae). *PLoS ONE* 8:e82198. doi: 10.1371/journal.pone.0082198
- Ohta, T. (1984). Some models of gene conversion for treating the evolution of multigene families. *Genetics* 106, 517–528.
- Petit, R. J., Aguinagalde, I., de Beaulieu, J. L., Bittkau, C., Brewer, S., Cheddadi, R., et al. (2003). Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* 300, 1563–1565. doi: 10.1126/science.1083264
- Petit, R. J., Brewer, S., Bordács, S., Burg, K., Cheddadi, R., Coart, E., et al. (2002). Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecol. Manage.* 156, 49–74. doi: 10.1016/S0378-1127(01)00634-X
- Ramos, A. C. S., Lemos-Filho, J. P., Ribeiro, R. A., Santos, F. R., and Lovato, M. B. (2007). Phylogeography of the tree *Hymenaea stigonocarpa* (Fabaceae: Caesalpinioideae) and the influence of Quaternary climate changes in the Brazilian Cerrado. *Ann. Bot.* 100, 1219–1228. doi: 10.1093/aob/mcm221
- Salgado-Labouriau, M. L. (1997). Late Quaternary paleoclimate in the savannas of South America. *J. Q. Sci.* 12, 371–379.
- Shaw, J., Lickey, E. B., Beck, J. T., Farmer, S. B., Liu, W., Miller, J., et al. (2005). The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* 92, 142–166. doi: 10.3732/ajb.92.1.142
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539–544. doi: 10.1038/msb.2011.75
- Sokal, R. R., and Rohlf, F. J. (1995). *Biometry: The Principles and Practice of Statistics in Biological Research, 3rd Edn*. New York, NY: W. H. Freeman and Company.
- Stokland, J. N., Halvorsen, R., and Stoa, B. (2011). Species distribution modeling - effect of design and sample size of pseudo-absence observations. *Ecol. Modell.* 222, 1800–1809. doi: 10.1016/j.ecolmodel.2011.02.025
- Taberlet, P., Fumagalli, L., Wust-Saucy, A. G., and Cosson, J. F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Mol. Ecol.* 7, 453–464. doi: 10.1046/j.1365-294x.1998.00289.x
- Terribile, L. C., Lima-Ribeiro, M. S., Araújo, M. B., Bizão, N., Collevatti, R. G., Dobrovolski, R., et al. (2012). Areas of climate stability of species ranges in the Brazilian Cerrado: disentangling uncertainties through time. *Nat. Conserv.* 10, 152–159. doi: 10.4322/natcon.2012.025
- Waters, J. M., Fraser, C. I., and Hewitt, G. M. (2013). Founder takes all: density-dependent processes structure biodiversity. *Trends Ecol. Evol.* 28, 78–85. doi: 10.1016/j.tree.2012.08.024
- Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A., and Smith, G. M. (2009). *Mixed Effects Models and Extensions in Ecology with R*. New York, NY: Springer.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Collevatti, Terribile, Rabelo and Lima-Ribeiro. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.