

RESEARCH ARTICLE

Open Access



# Demographical history and palaeodistribution modelling show range shift towards Amazon Basin for a Neotropical tree species in the LGM

Luciana Cristina Vitorino<sup>1</sup>, Matheus S. Lima-Ribeiro<sup>2</sup>, Levi Carina Terribile<sup>2</sup> and Rosane G. Collevatti<sup>1\*</sup>

## Abstract

**Background:** We studied the phylogeography and demographical history of *Tabebuia serratifolia* (Bignoniaceae) to understand the disjunct geographical distribution of South American seasonally dry tropical forests (SDTFs). We specifically tested if the multiple and isolated patches of SDTFs are current climatic relicts of a widespread and continuously distributed dry forest during the last glacial maximum (LGM), the so called South American dry forest refugia hypothesis, using ecological niche modelling (ENM) and statistical phylogeography. We sampled 235 individuals of *T. serratifolia* in 17 populations in Brazil and analysed the polymorphisms at three intergenic chloroplast regions and ITS nuclear ribosomal DNA.

**Results:** Coalescent analyses showed a demographical expansion at the last c. 130 ka (thousand years before present). Simulations and ENM also showed that the current spatial pattern of genetic diversity is most likely due to a scenario of range expansion and range shift towards the Amazon Basin during the colder and arid climatic conditions associated with the LGM, matching the expected for the South American dry forest refugia hypothesis, although contrasting to the Pleistocene Arc hypothesis. Populations in more stable areas or with higher suitability through time showed higher genetic diversity. Postglacial range shift towards the Southeast and Atlantic coast may have led to spatial genome assortment due to leading edge colonization as the species tracks suitable environments, leading to lower genetic diversity in populations at higher distance from the distribution centroid at 21 ka.

**Conclusion:** Haplotype sharing or common ancestry among populations from Caatinga in Northeast Brazil, Atlantic Forest in Southeast and Cerrado biome and ENM evince the past connection among these biomes.

**Keywords:** Bignoniaceae, Dry forest refugia, Ecological niche modelling, Phylogeography, Pleistocene arc hypothesis, Quaternary climatic changes

## Background

Recent phylogeographical works indicate expansion of South American seasonally dry forests (SDTF hereafter) across Quaternary glaciations [1, 2], agreeing with the idea of a unique and continuously distributed SDTF bordering Amazon Basin and northern Andes during glacial phases, a scenario also known as the Pleistocene Arc hypothesis

(PLAH; [3]). The alternative scenario of range expansion towards the interior of Amazon Basin (the Pennington, Prado and Pendry hypothesis, PPPH hereafter; [4]) is also supported by phylogeographical patterns of other species, like the widely distributed *Tabebuia impetiginosa* [2]. Both hypotheses predict that the current disjunct distribution of SDTFs across South America is the result of vicariance (i.e. fragmentation) of a formerly more widespread and continuously distributed dry forest during the Last Glacial Maximum (LGM), which is known as the dry forest refugia hypothesis [5, 6].

\* Correspondence: rosanegc68@hotmail.com

<sup>1</sup>Laboratório de Genética & Biodiversidade, Instituto de Ciências Biológicas, (ICB), Universidade Federal de Goiás (UFG), Cx.P. 131, Goiânia, GO 74001-970, Brazil

Full list of author information is available at the end of the article



However, the response of South American SDTF species to Quaternary climate change is still poorly understood, as revealed by contrasting phylogeographical patterns and multiple distribution dynamic response of species to Quaternary climate changes [1, 2, 7]. Actually, an increasing body of evidence does not support the generalized dry forest refugia hypothesis as originally proposed [7, 8]. For instance, the Bolivian Chiquitano dry forest was established only during the Holocene as a consequence of population expansions from southern Amazon rain forest [5, 6]. In Northeast Brazil, in Caatinga biome, the fossil records indicate that current climatic and vegetation conditions have been established only after 4.8 ka (thousands of years before present) [9], reaching as late as 1.0 ka in some regions [10]. Studies using ecological niche modelling (ENM) also show species-specific responses to the Quaternary climate changes [11]. Phylogenetic studies show an ancient origin of SDTFs in Mesoamerica, dated from 20 to 30 Ma (millions of years before present) [12], and to ~17 Ma in Brazilian Caatinga [13], whereas the fossil records suggest that SDTFs have originated around 13 to 12 Ma, at least in some part of its distribution [14, 15]. Dated phylogenies indicate a more ancient origin for Neotropical rain forests in the Cretaceous [16], although fossil records do not show evidence for Neotropical rain forests before the early Tertiary (~60 Ma) [15]. Such contrasting information reinforces the need for comprehensive studies about the dynamics of SDTF as an important source of evidence for the effect of climate change on species distribution.

SDTFs are one of the most threatened ecosystems in the world [17]. Although they originally occupied ~42 % of the tropical and subtropical forest regions [18], currently most of their remaining areas are in South America (~54.2 %), mainly in Northeast and Central Brazil and in Southeast Bolivia, Paraguay and Northern Argentina (Additional file 1: Figure S1), representing ~22 % of the forested area in South America [18]. In Brazil, the SDTFs are distributed from the Northeast, in Caatinga (Additional file 1: Figure S1), towards the Southwest in Misiones nucleus (that includes the northeastern Argentina and eastern Paraguay), and also are scattered throughout other vegetation types such as Amazon rain forest and savannas in Central Brazil [19] in areas of eutrophic and oligotrophic soils with neutral or quasi-neutral pH values and low levels of aluminium [20]. Most remaining areas of SDTFs in Brazil are threatened mainly by agricultural expansion, harvesting for wood products and the increase of fire frequency due to agricultural practices [11].

The distribution of SDTF species and the contrasting palaeoscenarios raise questions about the long-term stability and dynamics of the SDTF communities and

the role of the Quaternary climate oscillations in the current patterns of genetic diversity and geographical distribution of plant species, that may be answered using multi-model inference approach [2, 21–23], as we address here for the widely distributed Neotropical tree *Tabebuia serratifolia* (Vahl) Nichols. (Bignoniaceae). *Tabebuia serratifolia* has a disjunct distribution and local low abundance throughout the SDTFs of South America (see Additional file 1: Figure S2). It occurs from the fragments of SDTFs in Northeast Brazil, in the Caatinga biome, towards the Misiones nucleus (Southwest Brazil, Bolivia), Paraguay and Peru, and in SDTFs of Amazon Basin and of Atlantic Forest. It is also scattered throughout the fragments of SDTFs in Central Brazil and the east slopes of Andes. *Tabebuia serratifolia* and *T. impetiginosa* are the most logged species in Brazil and the second most expensive timber, popularly known as 'pau d'arco' [24].

Here we studied the phylogeography of *T. serratifolia* and tested the dry forest refugia hypothesis concerning specific climatic oscillations during the last glacial cycle. Our analyses followed the framework proposed by Collevatti et al. [2, 21, 23], which is based on coalescence simulations of alternative demographical hypotheses based on biogeographical *a priori* hypotheses (PLAH and PPPH) and other two demographical expectations predicted by the ecological niche models (ENM). Hypotheses from the ENMs include the combined dynamics of both PLAH and PPPH hypotheses and range retraction at the LGM (instead of range expansion, a formerly widespread and continuously distributed SDTFs, as expected by the previous hypotheses). These hypotheses were tested elsewhere for *T. impetiginosa* [2].

## Results

### Genetic diversity and population structure

The sequencing of the chloroplast intergenic spacers *psbA-trnH*, *trnC-ycf6* and *trnS-trnG* generated combined data with 1,742 bp sites (excluding microsatellites and coding indels as one evolutionary step), 157 polymorphic sites and 79 different haplotypes for the 257 individuals of *T. serratifolia*. For ITS1 + 5.8S + ITS2 (ITS) we obtained a fragment of 518 bp with 23 polymorphic sites and 21 different haplotypes.

Chloroplast genome showed higher haplotype diversity ( $h = 0.873$ ) than nuclear ITS ( $h = 0.772$ ), but nucleotide diversity at cpDNA ( $\pi = 0.0032$ ,  $SD = 0.0017$ ) was lower than for nuclear genome ( $\pi = 0.0058$ ,  $SD = 0.0034$ , Table 1). Higher genetic diversity was found for populations ALT, CRA, PNI and SAB.

Nuclear ITS showed four widespread haplotypes shared by populations from the Northeast and the Central Brazil (Fig. 1a, b) and cpDNA showed only two widespread haplotypes. Haplotype H4 (nrDNA ITS) and H2 (cpDNA) were

**Table 1** Genetic diversity based on Arlequin Ver 3.11 software and demographical parameters based on coalescent analysis performed with Lamarc 2.1.9 software for 17 populations of *Tabebuia serratifolia*

Pop	N	cpDNA				ITS				Combined cpDNA and ITS nrDNA			
		k	h	$\pi$	SD	k	h	$\pi$	SD	$\theta$	$\theta$ - 95 % CI	Ne	Ne - 95 % CI
North - Northeast													
CCM	20	2	0.100	0.0001	0.0001	1	0.000	0.0000	0.0000	0.0001	0.0001 – 0.0004	270.34	270.34 – 1081.37
CRA	09	9	1.000	0.0065	0.0037	2	0.389	0.0075	0.0047	0.0071	0.0023 – 0.0231	19194.38	6217.90 – 62449.31
POF	27	8	0.752	0.0012	0.0008	3	0.145	0.0003	0.0005	0.0009	0.0005 – 0.0024	2433.09	1351.72 – 6488.24
SEC	04	2	0.500	0.0009	0.0008	3	0.833	0.0035	0.0030	–	–	–	–
Central – Central West													
ALT	32	16	0.845	0.0022	0.0012	2	0.405	0.0008	0.0008	0.0071	0.0001 – 0.0229	19194.38	270.34 – 61908.62
ARA	13	4	0.423	0.0008	0.0006	2	0.385	0.0007	0.0008	0.0006	0.0001 – 0.0019	1622.06	270.34 – 5136.52
BOD	10	4	0.533	0.0005	0.0004	1	0.000	0.0000	0.0000	0.0003	0.0001 – 0.0011	811.03	270.343 – 2973.78
GSV	03	1	0.000	0.0000	0.0000	2	1.000	0.0039	0.0036	–	–	–	–
LUZ	30	4	0.356	0.0008	0.0006	2	0.131	0.0003	0.0004	0.0006	0.0002 – 0.0013	1622.06	540.69 – 3514.46
MIM	13	7	0.846	0.0031	0.0018	5	0.756	0.0090	0.0004	0.0024	0.0009 – 0.0068	6488.24	2433.09 – 18383.35
POT	02	2	1.000	0.0040	0.0043	1	0.000	0.0000	0.0000	–	–	–	–
PNA	04	2	0.500	0.0003	0.0004	1	0.000	0.0000	0.0000	–	–	–	–
Southeast													
CAP	23	8	0.629	0.0010	0.0006	1	0.000	0.0000	0.0000	0.0008	0.0005 – 0.0014	2162.75	1351.72 – 3784.81
PNI	21	11	0.871	0.0032	0.0018	6	0.758	0.0080	0.0048	0.0038	0.0014 – 0.0124	10273.05	3784.81 – 33522.57
SAB	06	4	0.866	0.0047	0.0029	2	0.533	0.0010	0.0011	0.0012	0.0004 – 0.0117	3244.12	1081.37 – 31630.17
SCA	21	7	0.500	0.0009	0.0006	2	0.095	0.0007	0.0008	0.0009	0.0004 – 0.0016	2433.09	1081.37 – 4325.49
SUM	19	3	0.205	0.0009	0.0006	3	0.298	0.0022	0.0017	0.0005	0.0002 – 0.0955	1351.72	540.69 – 258177.90
Mean	–	5.53	0.584	–	–	2.29	0.337	–	–	–	–	–	–
SD	–	3.99	0.306	–	–	1.40	0.335	–	–	–	–	–	–
Overall	257	79	0.873	0.0032	0.0017	21	0.772	0.0058	0.0034	0.0275	0.0221–0.0355	74344.42	59745.88 – 95971.88

N sample size, k - number of haplotypes, h haplotype diversity,  $\pi$  nucleotide diversity, SD standard deviation,  $\theta$  coalescent or mutation parameter, Ne effective population size; 95 % CI is the credibility interval at 95 %.  $\theta$  and Ne were estimated for combined cpDNA and ITS data

shared by populations CRA, SEC, CCM, POF, from northeast, PNA and ARA, from central and POT and BOD, from Central-West Brazil (Fig. 1a and b). Despite haplotype sharing, the network showed evidence of genetic structure with high differentiation between populations from Northeast and Southeast Brazil (Additional file 1: Figure S3).

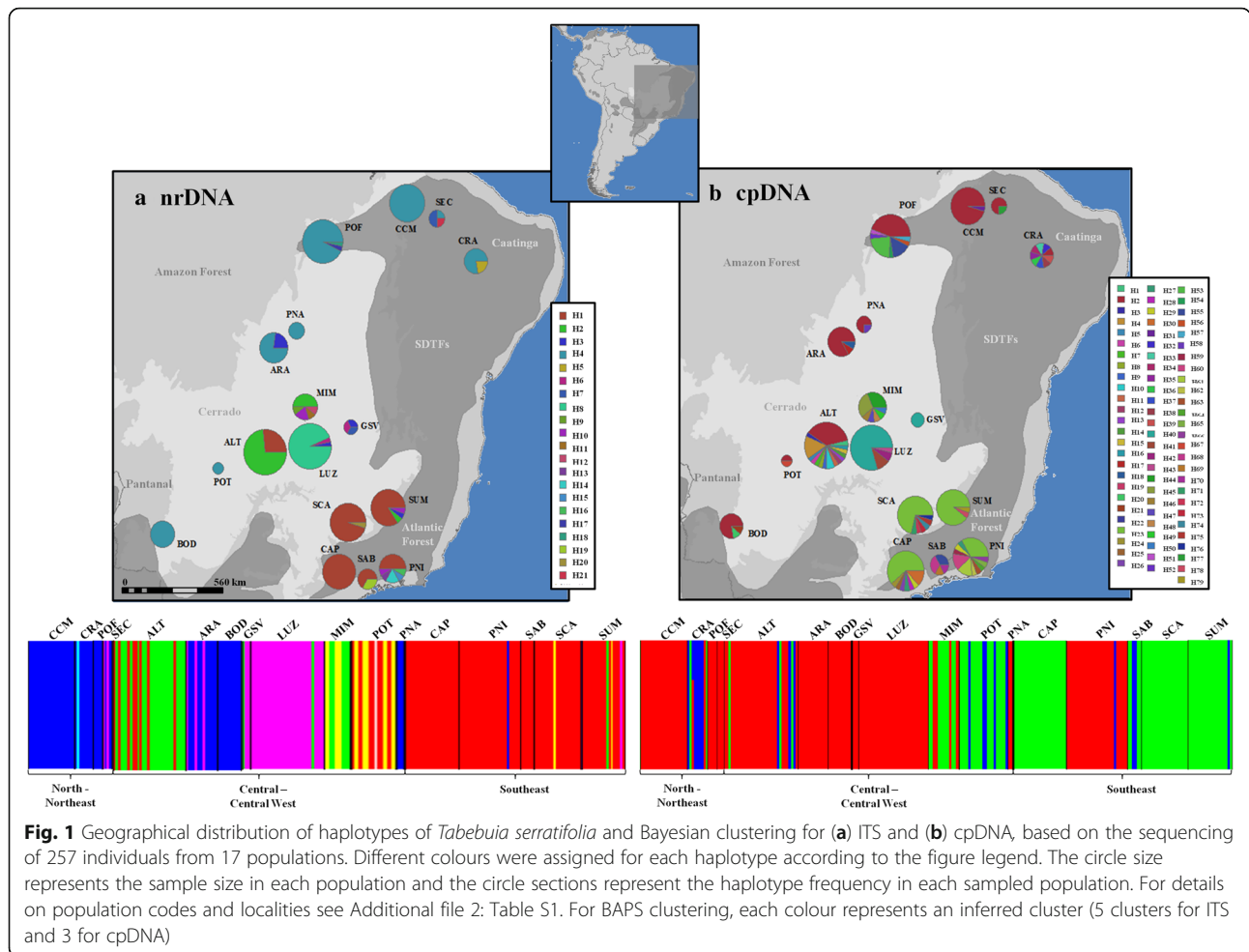
Analysis of Molecular Variance also showed a high differentiation among populations for both chloroplast ( $\phi_{ST} = 0.528$ ,  $p < 0.001$ ) and ITS ( $\phi_{ST} = 0.742$ ,  $p < 0.001$ ). Pairwise  $\phi_{ST}$  was high between almost all population pairs for both cpDNA and ITS (Additional file 2: Table S6).

Bayesian clustering for cpDNA indicated an optimal partition of 3 groups for cpDNA and 5 groups for ITS (Fig. 1a, b). For cpDNA (Fig. 1b), populations from Northeast and Central-West were clustered together in one group (red cluster, ALT, ARA, BOD, CCM, GSV, LUZ, POF, PNA, SEC) with population PNI from Southeast. Populations from Southeast and Central-East formed other cluster (green, CAP, MIM, POT, SAB, SCA, SUM) with some individual from population ALT and CRA, from the

Northeast. The third cluster (blue) was formed mainly by a population from the Northeast (CRA) and individuals from ALT, PNI, POT and PNA. ITS Bayesian clustering (Fig. 1a) split the cpDNA cluster 1 (red) in two. The first cluster comprised populations from North–Northeast and Central-West (blue, ARA, BOD, CCM, CRA, PNA, POF, SEC). The second cluster was formed by populations from Central-West Brazil (green, ALT, MIM). Populations from the Southeast were grouped in one cluster (red, CAP, PNI, SAB, SCA, SUM) with some individuals from populations of Central-West (ALT, POT). Population LUZ, from Central-West formed one cluster (pink) with individuals from populations ARA, SEC and SUM. Another cluster (black) was formed by individuals from different populations from the three geographical regions (Fig. 1a).

#### Demographical history and time to most recent common ancestor

Extended Bayesian Skyline Plot (EBSP) analysis showed a demographical expansion of *T. serratifolia* at c. 130 ka



(Fig. 2a) steadying after the LGM. We also found low values of mutation parameter  $\theta$  for all populations (Table 1) and overall population ( $\theta = 0.0275$ ). Most populations had high effective population sizes (Table 1). Gene flow among all population pairs was negligible (less than 1 migrant per generation, Additional file 2: Table S7 and S8).

Haplotypes of *T. serratifolia* showed an ancient time to most recent common ancestor (TMRCA, Fig. 2b), ~3.4 Ma (95 % CI = 1.9 – 6.8 Ma), which coincides with the coalescence of haplotypes from populations ALT and MIM with all other haplotypes. Most divergences occurred after c. 1.5 Ma (Fig. 2b) and resulted in incomplete lineage sorting, with geographically distant populations sharing haplotypes and common ancestors.

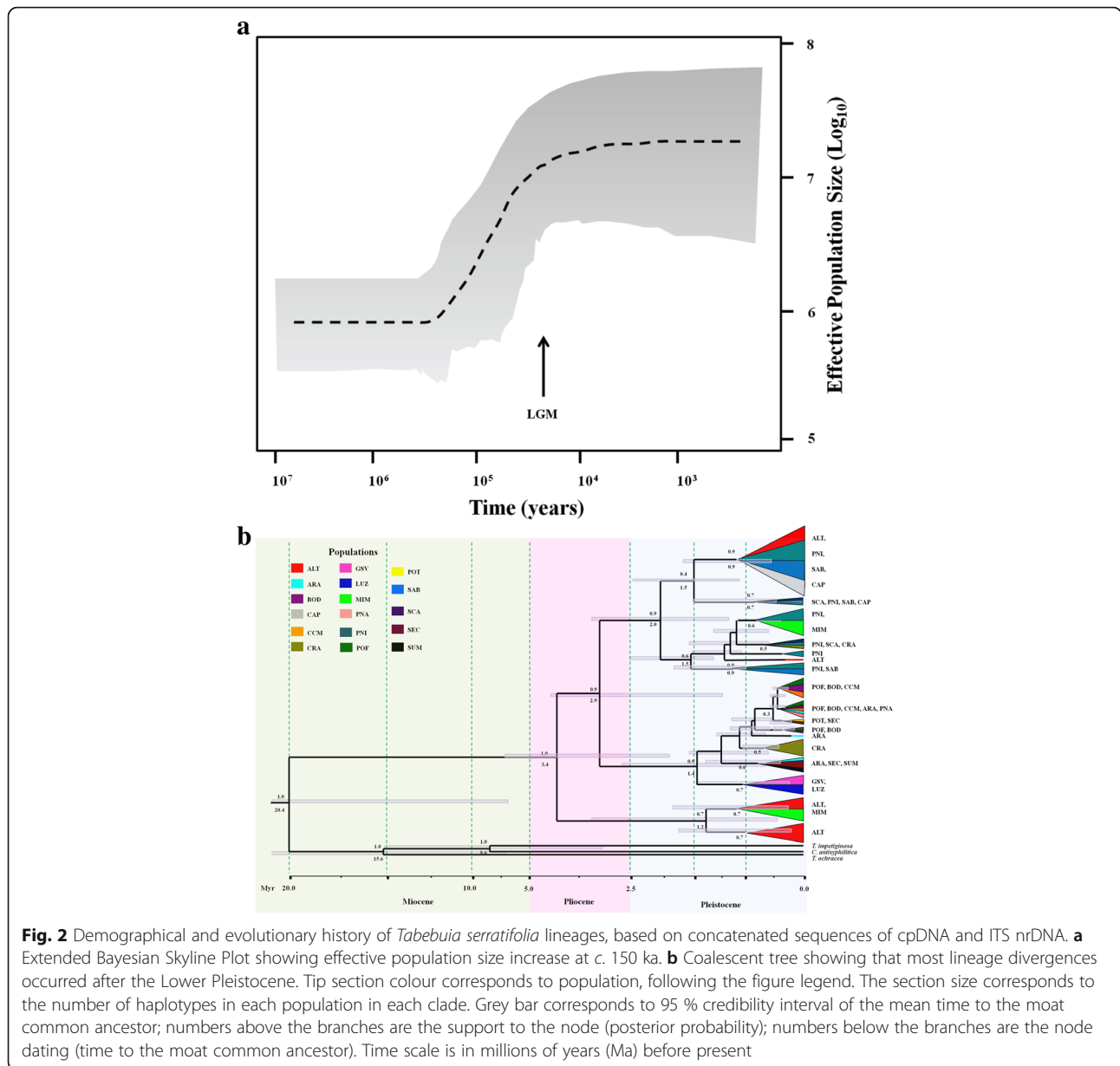
**Setting-up demographical hypotheses**

**Species palaeodistribution modelling**

The ENM showed that the range size of *T. serratifolia* at the LGM (Fig. 3a) was greater than at the mid-Holocene (Fig. 3b), but similar to the present-day (Fig. 3c, see also

Additional file 1: Figure S4). During the LGM, the species was predicted to occur across the Amazon Basin, Brazilian Cerrado and western Caatinga. A range shift towards the Atlantic coast at the Northeast and the Southeast was predicted from the LGM to the mid-Holocene (Fig. 3, Additional file 1: Figure S4). In the present-day, *T. serratifolia* remained in the Atlantic Forest, but expanded its distribution towards the west Brazil, Bolivia and Peru, mainly bordering the Amazon Basin (Fig. 3, Additional file 1: Figure S4). The ensemble predicted a historical refugium extending from Amazon Basin towards (Fig. 3d) the Central and Southeast Brazil.

The analysis of uncertainty using hierarchical ANOVA showed lower proportional variance through time than among model methods (Additional file 2: Table S9). However, variation was highly spatially structured (Additional file 1: Figure S5) with higher proportion of variation from time component and low from ENM and AOGCMs (atmosphere-ocean general circulation models) within the geographical range of *T. serratifolia*, indicating that the ENMs were able to detect the effects



of climate changes on the distribution dynamics of *T. serratifolia* through the last glaciation, despite the AOGCM variation (Additional file 1: Figure S5).

#### Inferring demographical hypotheses and model selection

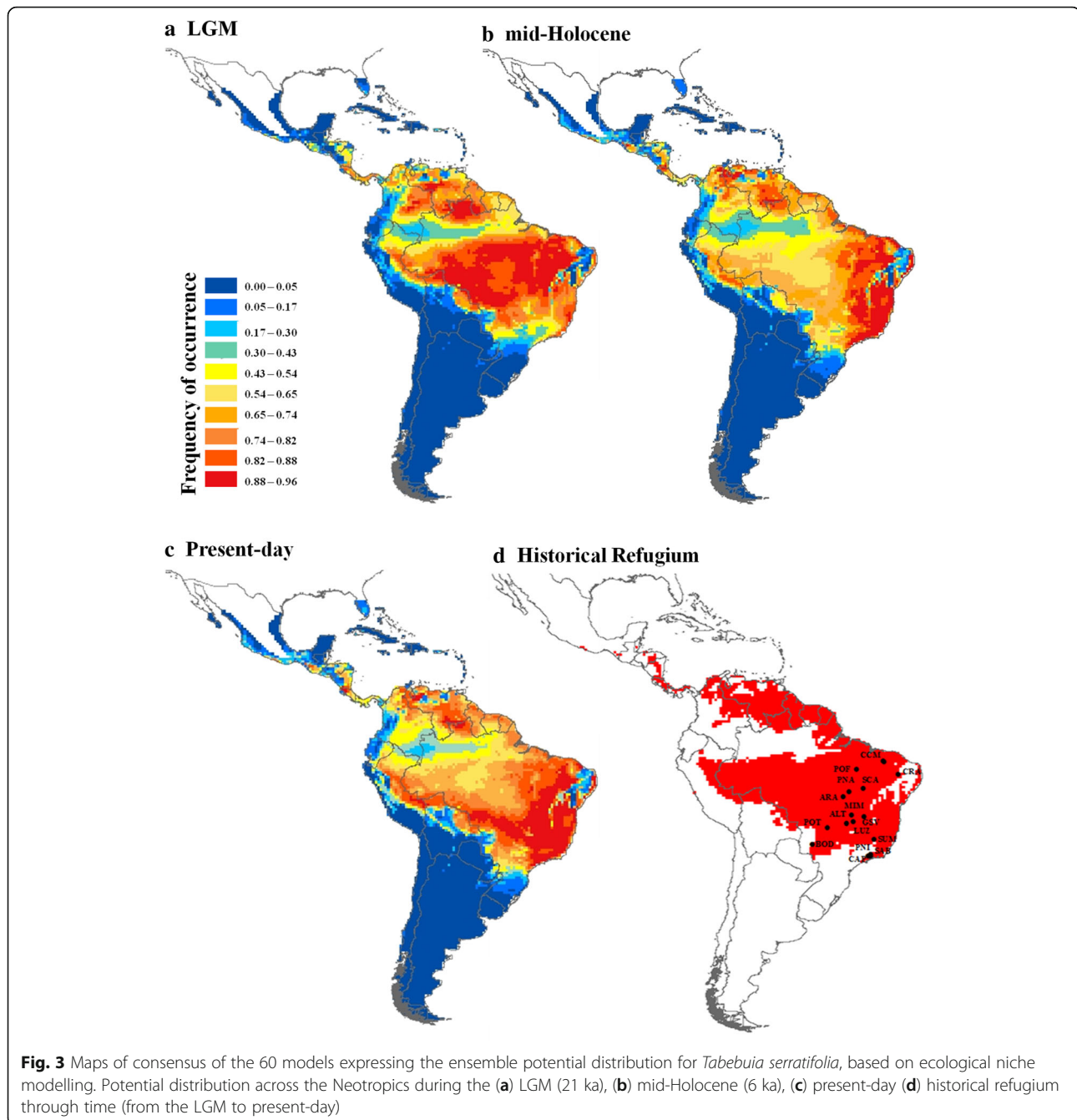
Among the 60 palaeodistribution maps the pattern supported was PPPH (67 %), followed by “Both” hypothesis (PLAH + PPPH) with 15 % of maps, and “Range Retraction” (10 %). No ENM prediction supported the scenario expected by PLAH hypothesis (Additional file 2: Table S10) and 8 % of the maps matched none of the hypotheses.

Similarly, the coalescent simulations of alternative demographical hypotheses suggested that the hypothesis

PPPH (Table 2) was the most likely scenario explaining the current pattern of genetic diversity at the chloroplast genome of *T. serratifolia* compared to the other competing hypotheses, using either two-tailed probability or Akaike criteria ( $\Delta AIC$  and  $AICw$ ; Table 2). For ITS, although hypothesis PPPH was the most likely for haplotype and nucleotide diversities ( $\Delta AIC = 0.000$  for both), hypotheses “Both” and “Range Retraction” could not be rejected for haplotype diversity (Table 2).

#### Spatial patterns of genetic diversity

Genetic differentiation among populations and geographical distance were not significantly correlated for



cpDNA (Mantel Test,  $r^2 = 0.0223$ ,  $p = 0.560$ ) but significantly correlated for ITS ( $r^2 = 0.1892$ ,  $p = 0.02$ ).

Quantile regressions showed effects of climate changes on the genetic diversity for ITS. Populations in more climatically suitable areas and closer to the centroid of the distribution during the LGM showed higher haplotype ( $h$ ) diversity (Fig. 4, see also Additional file 1: Figures S6 and S7). However, opposite relationships were observed for the mutation parameter  $\theta$  (Additional file 1: Figures S5 and S6). In addition, populations in more unstable areas showed lower genetic diversity and lower

$\theta$  (Additional file 1: Figure S8). Relationships for chloroplast genome were not significant.

## Discussion

Our findings from palaeodistribution modelling, phylogeographical analyses and coalescent simulations supported the hypothesis of dry forest refugia in South America for *T. serratifolia*. The pattern of genetic diversity is consistent with a demographical scenario of range expansion at the LGM, also supported by ENM, which predicted the scenario PPPH as the most likely;

**Table 2** Comparison of the demographical hypotheses in retrieve the haplotype ( $h$ ) and nucleotide ( $\pi$ ) diversities observed

	Hypothesis	$P$	$h$		$\pi$		
			$\Delta AIC$	$AICw$	$\Delta AIC$	$AICw$	
cpDNA	PLAH	0.040	6.680	0.024	0.020	2.681	0.164
	PPPH	0.360	0.000	0.680	0.100	0.000	0.629
	Both	0.080	2.000	0.277	0.020	2.756	0.158
	Retraction	0.200	7.187	0.019	0.002	5.095	0.049
ITS	PLAH	0.194	2.852	0.111	0.040	2.000	0.241
	PPPH	0.498	0.000	0.465	0.140	0.000	0.587
	Both	0.194	1.287	0.243	0.020	3.428	0.106
	Retraction	0.378	1.876	0.181	0.040	4.382	0.066

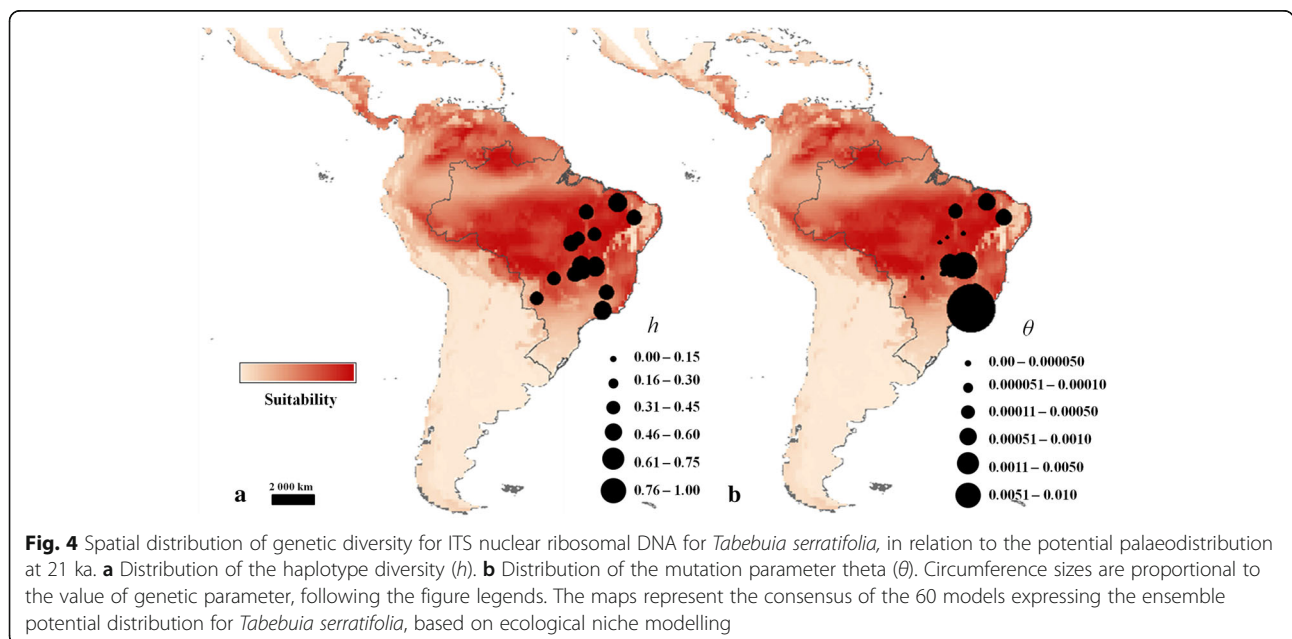
Data obtained from 2,000 simulations using the software BayeSSC. ( $P$ ) two-tailed probability.  $\Delta AIC$  and  $AICw$  Akaike Information Criterion delta and weights. PLAH, Pleistocene Arc Hypothesis, PPPH, Amazon SDF Hypothesis; Both, pattern matching both hypotheses PLAH + PPPH; Retraction, pattern matching a strong range retraction. See Fig. 5 for details about the demographical scenarios.  $h$  haplotype diversity,  $\pi$  nucleotide diversity

i.e. a range expansion and shift towards the Amazon Basin. Thus, the current disjunct distribution of *T. serratifolia* is most likely due to the range contraction of an ancient wider distribution, representing a climatic relict of drier and colder ice ages in Central-Southeast Brazil and Amazon Basin.

In fact, our results suggest that *T. serratifolia* was distributed northeastward during the LGM. Some typical species of SDTF are now widely distributed in Amazon Basin, occurring at low frequency in areas of more fertile soils [4], like *T. serratifolia* (see Additional file 1: Figure S2). The expansion of SDTFs throughout the Amazon Basin may have been favoured by the reduction in the

sea level during the Pleistocene glaciations causing a decrease in the level of the Amazon Basin rivers, exposing areas for the colonization by SDTF species [4]. In addition, *T. serratifolia* is a more generalist species, currently occurring in forests of Amazon Basin, SDTFs and riparian forests of the Cerrado and Caatinga biomes and in Atlantic Forest. The suitable climatic conditions for *T. serratifolia* at present-day clearly show its preference for hot climates but with high variation in precipitation. Such climatic conditions were less available during the LGM at the Central Brazil (e.g. [25]) than during the mid-Holocene and the present-day. The temperature decrease during the LGM potentially lead to a range shift and expansion towards the Amazon Basin in response to the decreasing availability of suitable conditions in certain regions, such as the Southeast Brazil. In addition, although Amazon Basin remained forested during the LGM, the forest structure and species composition might have changed due to the lower temperatures, precipitation and atmospheric CO<sub>2</sub> concentrations [26]. The southern Amazon forest became more seasonal, opening the opportunity for seasonally dry forest species colonization.

Despite the geographic and demographical expansion during glaciations, some populations had low haplotype and nucleotide diversity, contrary to the expected for a species with wide and continuous distribution during the LGM and with high effective population sizes. This is most likely due to the cycles of range shifts towards the Amazon Basin during recurrent glaciations that may have caused the extinction of haplotypes in some populations and a spatial assortment decreasing genetic diversity [27–29]. Indeed, quantile regressions showed a



**Fig. 4** Spatial distribution of genetic diversity for ITS nuclear ribosomal DNA for *Tabebuia serratifolia*, in relation to the potential palaeodistribution at 21 ka. **a** Distribution of the haplotype diversity ( $h$ ). **b** Distribution of the mutation parameter theta ( $\theta$ ). Circumference sizes are proportional to the value of genetic parameter, following the figure legends. The maps represent the consensus of the 60 models expressing the ensemble potential distribution for *Tabebuia serratifolia*, based on ecological niche modelling

cline variation in the genetic diversity for ITS, because populations in more stable areas through time and with higher suitability had higher genetic diversity. Postglacial range shift towards the southeast and east may have led to spatial genome assortment due to leading edge colonization, as the species tracked suitable environments [29]. The leading edge colonization may have triggered the lower genetic diversity in populations at higher distance from the centroid of the geographical distribution at 21 ka, i.e. areas with lower climatic stability. The spreading from the leading edge may lead to bottlenecks of the colonizing genome, decreasing genetic diversity in some new colonizing areas [29]. 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 [27, 28], and density-dependent processes due to the fast colonization and founder events may also cause patches and sectoring in genetic diversity [30, 31].

*Tabebuia serratifolia* showed slightly lower genetic diversity than *T. impetiginosa*, most likely due to different response to Quaternary climate changes. *Tabebuia impetiginosa*, showed a larger range expansion (PLAH + PPPH) during glacial periods (see [2]) as compared to *T. serratifolia* that showed a range shift towards the Amazon Basin (PPPH). Despite differences in effective population size and inheritance, chloroplast genome showed higher diversity than ITS nrDNA, which has four times the effective size of chloroplast genome and higher mutation rates. This may be due to concerted evolution in nrDNA that may homogenize copies decreasing genetic variation [32]. In addition, different genetic diversity signatures for ITS and chloroplast may be also due to different evolutionary rates corresponding to different time slices in the species evolution. The differences may also be due to the sequences sizes (1,743 versus 518 bp, for cpDNA and ITS respectively) because nucleotide diversity, which is corrected for sequence size, was slightly higher for ITS than for cpDNA.

Genetic diversity and number of haplotypes were higher in Central Brazil, in SDTFs from Cerrado biome (ALT, MIM), although some populations at Atlantic Forest also presented high genetic diversity (PNI). In fact, the results on ENM and quantile regressions showed that populations in more climatically suitable areas during the LGM presented higher genetic diversity (see Fig. 3) suggesting that stable areas at the Central Brazil were important refugia for *T. serratifolia*. Haplotype sharing between populations from Northeast Brazil (for instance, CRA in Caatinga) and populations in Atlantic Forest in Southeast Brazil and in Cerrado biome, and the Bayesian clustering of populations from these biomes (see Fig. 1) show the past connection between Caatinga and Cerrado and Atlantic Forest.

Indeed, BAPS showed a more shallow genetic structure for cpDNA than for ITS, that may reflect differences in seed and pollen dispersal. In fact, genetic differentiation was higher for ITS ( $\phi_{ST} = 0.742$ ) than for cpDNA ( $\phi_{ST} = 0.528$ ), showing that restriction in pollen dispersal contributes more to differentiation than seed dispersal. BAPS showed almost the same result for both sequences for the Southeast cluster (green cluster for cpDNA and red cluster for ITS) except for population CRA in Northeast, that was not included in ITS red cluster. The pollen fossil record shows warmer and wetter climatic periods in the Late Pleistocene that may have caused the expansion of riparian forests in Cerrado biome connecting Atlantic and Amazon rainforests [33, 34]. Indeed, the ENM predictions for *T. serratifolia* show a high connection between the Atlantic Forest, Caatinga and Cerrado. However, the connection through the SDTF of the 'dry diagonal' may also have occurred [9, 35–37]. In Northeast Brazil, wet periods occurred during the last 210 ka, due to a southward displacement of the Intertropical Convergence Zone [37] and may have affected rainforest distribution [38]. The fossil record shows a forest expansion during the intermittent wet intervals that may have linked Amazon and Atlantic rainforests [39]. In addition, there are evidences of a drier climate in Atlantic Forest domain, in Southeast Brazil, c. 20 ka to 14 ka with establishment of species from Atlantic Forest only after that [40]. In other localities in Southeast Brazil, drier and colder climate may have persisted c. 48 – 18 ka [40, 41].

*Tabebuia serratifolia* lineages from Central Brazil (ALT, MIM) started to diverge first, in the Pliocene c. 3.4 Ma. However, major divergences occurred in the Lower Pleistocene (Calabrian Stage, 1.8 Ma to 781 ka [42]). We hypothesize that favourable climatic conditions for *T. serratifolia* (i.e. hot and wetter climates) were spatially more restricted before the Pleistocene leading to smaller effective population sizes as predicted by EBSP. Because the probability of coalescence is inversely related to the number of gene copies [43], most coalescence may occur just before demographical expansion, when populations have smaller effective population sizes (see [30] for a review). In fact, most coalescence occurred before the last glaciation coinciding with low population sizes showed by EBSP. Thus, the range expansion during the LGM, indicated by the palaeodistribution modelling and by the simulation of demographical hypotheses, matches the regional process of differentiation indicated by the coalescent tree with no later secondary contact. This result suggests an ancient process of incomplete lineage sorting due to expansion followed by a more regional population demographical expansion circumventing secondary contact and haplotype sharing, most likely due to restriction in soil suitability [2].



## Conclusions

In conclusion, our analyses based on coalescent simulation and ecological niche modelling strongly support a demographical and a geographical range expansion of *T. serratifolia* during the LGM as expected by dry forest refugia hypothesis. We also found a range shift towards the Amazon Basin, as expected by the prediction of PPPH scenario, with an important effect on the spatial pattern of genetic diversity. Finally, we showed here that phylogeographical analyses coupled with ecological niche modelling and coalescent simulations can be a very powerful framework for evaluating alternative hypotheses and potentially useful for disentangling mechanisms involved in the origin of the disjunct distribution of SDTFs.

## Methods

### Population sampling

We sampled 17 populations (257 individuals) of *T. serratifolia* mainly in SDTFs from Caatinga, Cerrado, Amazon and Atlantic Forest biomes (Fig. 1, see also the Additional file 2: Table S1). Distance between population pairs ranged from 13 (CCM and SEC) to ~2.276 km (CRA and BOD). In all populations we sampled expanded leaves or cambium from adult individuals for DNA extraction. Because of the high level of anthropic disturbs in the Brazilian SDTFs, some regions had limited amount of living individuals, resulting in different sample sizes among populations (Table 1). The DNA from leaves and cambium was extracted following the standard CTAB procedure [44].

We used samples of *Tabebuia ochracea*, *T. impetiginosa* and *Cybistax antisiphilitica* as outgroups for divergence dating (Additional file 2: Table S1). Sampling was not performed in conservation units and thus did not require any license. Vouchers were compared to herbarium material from the Federal University of Goiás (Universidade Federal de Goiás), in Goiânia, Brazil.

### DNA sequencing

Polymorphism was assessed using sequences of three chloroplast DNA (cpDNA) intergenic spacers, *psbA-trnH*, *trnC-ycf6* and *trnS-trnG* [45], and the nrDNA ITS1 + 5.8S + ITS2 (primers 75 and 92, [46]). DNA was amplified by PCR following the same conditions described in Collevatti et al. [2] for *T. impetiginosa*. PCR products were sequenced on a GS 3500 Genetic Analyzer (Applied Biosystems, CA) using the BigDye<sup>®</sup> Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems), according to the manufacturer instructions. All fragments were sequenced in forward and reverse directions.

Sequences were analysed and edited using the software SeqScape v3.0 (Applied Biosystems, CA) and final alignments were obtained using the software ClustalΩ [47].

Polymorphisms at mononucleotide microsatellites in cpDNA were excluded due to ambiguous alignment and to higher mutation rates. Long indels were coded as one evolutionary event (one character) and each base pair were equally weighted before analyses. The sequences of the three chloroplast regions were concatenated for all analyses.

### Genetic diversity and population structure

To understand the relationships among haplotypes of *T. serratifolia* we inferred intraspecific phylogeny for chloroplast and ITS data using median-joining network analysis implemented in the software Network 4.6.1.0 [48]. Genetic diversity for each population and overall populations were estimated based on nucleotide ( $\pi$ ) and haplotype ( $h$ ) diversities [49] using the software ArlequinVer 3.11 [50].

The hypothesis of population differentiation was tested based on an analysis of molecular variance (AMOVA, [51]) using the software ArlequinVer 3.11 [50], that estimates  $\phi_{ST}$  analogous to  $F_{ST}$  using information on the allelic content of haplotypes, as well as their frequencies [51]. Population structure was also assessed using Bayesian clustering implemented in the software BAPS v6.0 [52]. cpDNA and ITS were analysed as separate partitions with linkage model for sequences. We performed population admixture analysis based on mixture clustering with estimated number of clusters (K) with an upper limit of K = 17.

### Demographical history and time to most recent common ancestor

To trace the dynamics in effective population size we performed an Extended Bayesian Skyline Plot (EBSP) analysis [53] implemented in BEAST 1.8.3 [54], which calculates the effective population size ( $N_e$ ) through time combining data from different partitions. Chloroplast and nrDNA ITS data were combined in one analysis, but separate priors were given for each partition. No evidences of heterozygous individuals were found when sequences were analysed 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 (AIC), implemented in the software jModelTest2 [55]. For chloroplast regions, the model HKY + G was selected (-lnL = 3576.1766), with kappa = 1.5 and gamma shape equal to 0.0170. For ITS, the evolutionary model TIM1 + G was selected (-lnL = 905.4461) with gamma shape equal to 0.010. We used 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 the taxonomic

related species, *T. impetiginosa* [2] and *T. aurea* [56]. 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 [57]. We considered the results only when  $ESS \geq 200$  (effective sample size).

Further, we estimated the mutation or coalescent parameter  $\theta = 2\mu N_e$  (mutation parameter,  $\theta = 4\mu N_e$  for diploid genome) based on a Bayesian modelling using Markov Chain Monte Carlo (MCMC) approach [58] implemented in Lamarc 2.1.9 software [59]. For this analysis, we excluded populations with less than 5 individuals (see Table 1). The analyses were run with 20 initial chains of 10,000 steps and three final chains of 100,000 steps. The chains were sampled every 100 steps. We used the default settings for the initial estimate of theta. The program was run three times to certify for convergence and validate the analyses using Tracer v1.6 [57] and combined results were then generated. We considered results when  $ESS \geq 200$  and when marginal posterior probability densities were unimodal and converged among runs. The effective population size was estimated from the mutation parameter  $\theta$  using a generation time of 15 years (based on flowering time on permanent plots; RG Collevatti, unpublished data) and the same mutation rate used to related species [2, 56]. To study the past connectivity among populations we also estimated immigration parameter,  $M = 2N_e m / \theta$  (immigration rate,  $M = 4N_e m / \theta$  for diploid genome), using Lamarc 2.1.9 software.

TMRCAs were estimated based on Bayesian coalescent analysis implemented in the software BEAST 1.8.3 [54]. For both chloroplast and ITS, a relaxed molecular clock (uncorrelated lognormal) was assumed. 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 the ucl.d.stdev was greater than 1.25 and the coefficient of variation frequency histogram viewed in Tracer abutted against zero (~1.6 to 2.5) showing heterogeneity among branches. We assumed population expansion, based on the Extended Bayesian Skyline Plot (EBSP) analysis [60]. Prior  $N_e$  was set to assume a lower bound from zero to infinity upper bound with exponential distribution. Four independent analyses were run for 100 million generations. Mutation rates for both chloroplast and ITS regions were the same used for a taxonomic related species [2, 56]. 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 are informative because posterior values (e.g. posterior probability) were different from those obtained from empty alignment (priors only).

## Setting-up demographical hypotheses

### Species palaeodistribution modelling

Occurrence records of *T. serratifolia* across Neotropics (Additional file 1: Figure S2, Additional file 2: Table S1) were obtained from GBIF (Global Biodiversity Information Facility <http://www.gbif.org/>). All records were examined for probable errors and duplicates, and the nomenclature was checked for synonymies. The records were mapped in a grid of cells of  $0.5^\circ \times 0.5^\circ$  (longitude x latitude) encompassing the Neotropical region to generate a matrix of 698 presences (cells with occurrence records, Additional file 2: Table S2) used for distribution modelling (see below).

We also generate environmental layers as predictors for ENM using 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) and subsoil pH (30-100 cm). These five bioclimatic variables present low multicollinearity and were selected by factorial analysis with Varimax rotation using the 19 bioclimatic variables obtained in the EcoClimate database ([www.ecoclimate.org](http://www.ecoclimate.org); [61]). The climate predictors present  $0.5^\circ$  of spatial resolution and were obtained for LGM (21 ka), mid-Holocene (6 ka) and pre-industrial (expressing the current climate) periods, using simulations from four atmosphere-ocean general circulation models (AOGCM): CCSM4, CNRM-CM5, MIROC-ESM, MPI-ESM-P and MRI-CGCM3 (Additional file 2: Table S3). Subsoil pH (30–100 cm) was obtained from Harmonized World Soil Database (version 1.1, FAO/IIASA/ISRIC/ISS-CAS/JRC 2009). We assumed subsoil pH to be constant through time (from LGM to pre-industrial) and used in ENM as a “constraint variable” to better model the environmental preferences of *T. serratifolia*.

The distribution of *T. serratifolia* was modelled using 12 methods encompassing both presence-only, presence-background and presence-absence algorithms (Additional file 2: Table S4). Because absence data are not available for *T. serratifolia*, we randomly selected pseudo-absences throughout the Neotropical grid cells (excepting cells with presences) keeping prevalence equal to 0.5 to calibrate the ENM based on presence-absence observations [62, 63]. This approach was based on studies suggesting that the extent of the geographical region in which the pseudo-absence points are taken have important influences for prediction and performance of ENM [64, 65]. Thus, selecting pseudo-absences throughout the distribution of *T. serratifolia* (i.e. the Neotropical region) essentially represents a compromise between generating models that do not generalize well, not produce over predictions of distribution areas ignoring important spatial structure associated with finer scale environmental gradients [65].

The distribution of *T. serratifolia* was first modelled for current (i.e. pre-industrial) climate and then projected onto LGM (21 ka) and mid-Holocene (6 ka) palaeoclimatic conditions. All ENMs were ran in the integrated computational platform BIOENSEMBLES [66] following the ensemble approach [67]. The procedures for modelling using the ensemble approach were extensively discussed elsewhere [2, 66] and just a brief description will be presented below. For each species distribution model, the occurrence points were randomly partitioned into two subsets (training and testing) comprising 75 and 25 % of dataset, respectively, and this procedure was repeated 50 times. Initial models were evaluated by True Skill Statistics (TSS, [68]); models with poor performance ( $TSS < 0.5$ ) were eliminated (TSS values for all models are provided in Additional file 2: Table S5). Remaining models were combined (a weighted average by TSS value of each model) to generate the frequency of models supporting the occurrence of the species in each cell of Neotropical grid (i.e. consensus maps), for both current and past climatic layers. Next, the predictive maps for LGM, mid-Holocene and present-day were obtained by using the 10<sup>th</sup> percentile lowest presence threshold, i.e. excluding 10 % of the lowest consensus values linked to a occurrence record used to build the ENM.

We applied a hierarchical ANOVA using the predicted suitability from all models (12 ENM  $\times$  5 AOGCMs  $\times$  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 modelling components (i.e. ENM, 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 [69] for details about hierarchical design).

### Inferring demographical hypotheses

The 60 predicted palaeodistribution maps (estimated frequencies of occurrence for the 12 ENMs  $\times$  5 AOGCMs) were visually inspected by two of us (RGC and LCV), using a double-blind experimental design, and classified as supporting the alternative scenarios following Collevatti et al. [2]: i) the 'Pleistocene Arc Hypothesis', PLAH hypothesis [3], an expansion throughout the Central and Southwest Brazil; ii) the 'Amazon SDTF Hypothesis', PPPH hypothesis [4], a westward range shift, toward the Amazon Basin; iii) PLAH + PPPH ("Both"), i.e. a prediction for the past distribution as expected by both hypotheses, resulting in an expansion throughout the Central and Southwest Brazil and also towards the interior of Amazon Basin; iv) "Range Retraction", a retraction in geographical range in Central Brazil but without range shift. Although PLAH, PPPH and "Both"

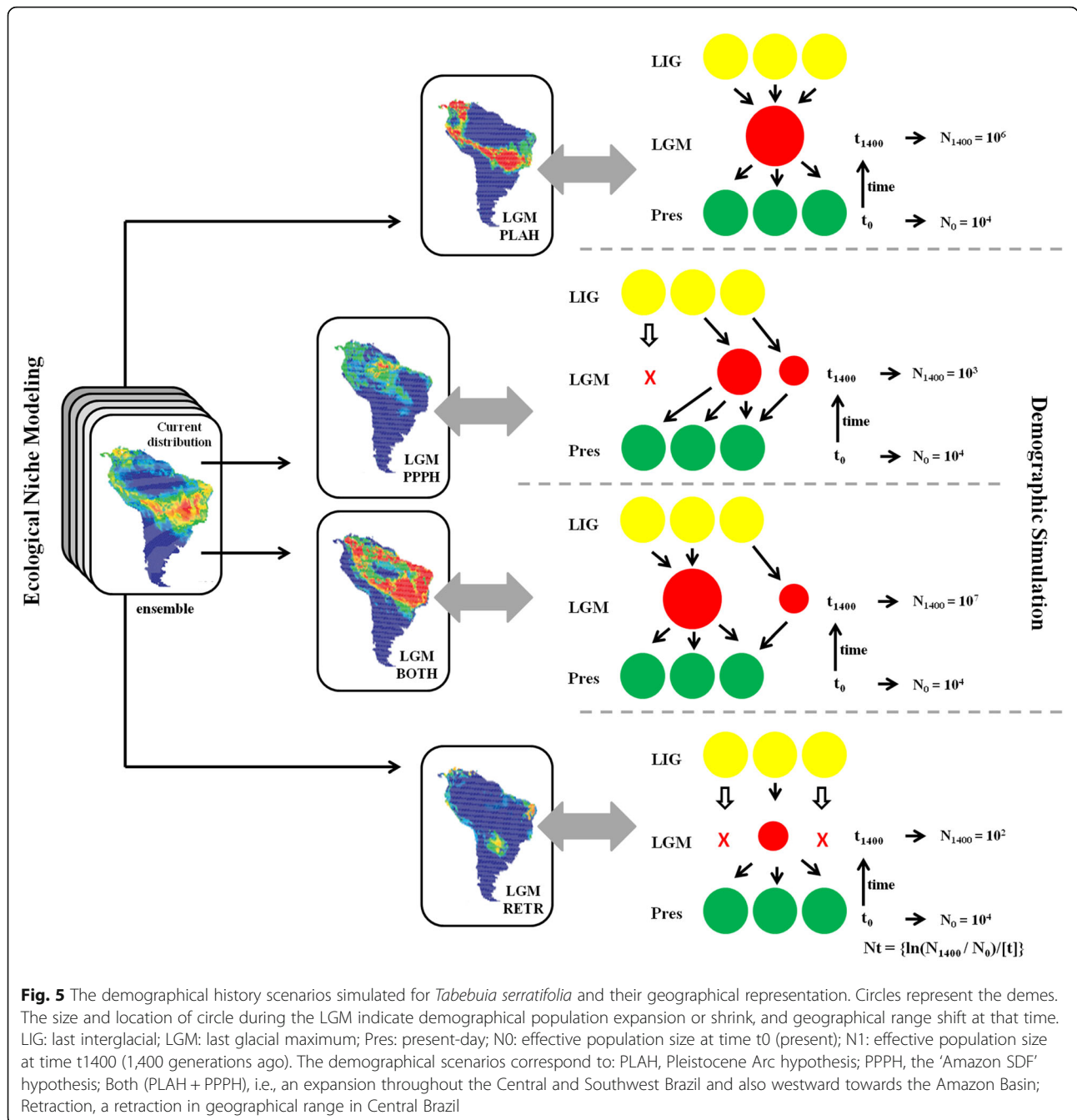
scenarios show different distribution dynamics for the SDTF in South America, they are all compatible with the dry forest refugia hypothesis.

### Demographical history simulation

The demographical history of *T. serratifolia* was modelled and simulated based on coalescent analysis [45] implemented in the software ByeSSC [70, 71]. We modelled four demographical scenarios (Fig. 5) according to the hypotheses supported by ecological niche modelling and biogeographical hypotheses (PLAH, PPPH, "Both", or "Range Retraction"), following the framework described in Collevatti et al. [2, 11]. For each demographical scenario, we run 2,000 independent simulations for each sequence region. Model calibration was based on parameters estimated with Lamarc 2.1.9 software and the molecular evolution of the chloroplast non-coding and ITS regions, i.e. the same evolutionary model, sequence size (in base pair) and mutation rates. The number of generations until the LGM (at 21 kyr BP) was calculated using a generation time of 15 years (RG Collevatti unpublished data).

Demographical hypotheses were simulated backward, with 17 demes from  $t_0$  (present) to  $t_{1400}$  generations ago (at the LGM). Population sizes at time  $t$  were calculated from  $N_t = \{\ln(N_1/N_0)/[t]\}$ , where  $N_0 = 10,000$  was the same for all scenarios, and  $N_1$  shifted among them according to our theoretical expectation (see Fig. 5 for details). In BayeSSC negative growth implies population expansion, because coalescent simulations run backward through time. Thus, a negative growth rate implies a population larger now than in the past, and a positive growth, a population smaller now than in the past. Because of the high variation in effective population sizes for *T. serratifolia* (see Table 1), we performed simulations with different initial deme sizes for  $N_0 = 1,000$ ,  $N_0 = 10,000$  and  $N_0 = 100,000$  for all scenarios. Simulations using  $N_0 = 1,000$  presented all values of haplotype and genetic diversity lower than those observed for *T. serratifolia* for all demographical hypotheses and  $N_0 = 100,000$  retrieved almost all values higher than the observed for *T. serratifolia*. Henceforth, we used simulations for  $N_0 = 10,000$ .

To simulate migration we considered a finite island model in which all current demes are descendants from lineages originally in deme 1 at  $t$  generations ago, meaning that as the 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 but 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. For PPPH and 'Range Retraction' we



considered that each lineage in deme x will migrate to deme 1 and than shrink until extinction.

**Model selection**

Simulated alternative models were compared based on the distribution of haplotype and nucleotide diversities in the 2,000 simulations for both chloroplast and ITS sequences. We estimated two-tailed probabilities as twice the number of diversity estimates that were higher than the observed, divided by the number of simulations, so that a high P-value indicates failure to reject the model.

We also estimated Akaike information criterion (AIC) for model choice. The log-likelihood, ln(L), 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. AIC (-2Ln(L) + 2 K, where K is the number of free parameters, 2 for all models) was transformed into AIC weight of evidence (AICw), given by exp[-0.5(AIC - AICmin)] [72], from which we obtained ΔAIC; i.e. the difference of AICw between each model and the best model. Models with ΔAIC < 2 were considered as equally plausible to

explain the observed pattern [73].  $AIC_w$  was expressed as a relative value among models (see [72]).

### Spatial patterns in genetic diversity

Because ENM and simulation supported range shift towards Amazonia Basin (see Results) for *T. serratifolia*, 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. Spatial expansion may lead to gradients in genetic diversity because of allele surfing during the colonization of new areas and “lead trail” expansion (see [30, 31] for reviews).

We first tested if differentiation is an effect of isolation by distance [72]. Pairwise linearized  $\phi_{ST}$  (analogous to  $F_{ST}$ ) among pairs of populations were estimated for both cpDNA and ITS and correlated with a geographical distance matrix (logarithm of geodesic distance) by a Mantel test using ArlequinVer 3.11 [52] and statistical significance was tested by a non-parametric permutation test (10,000 permutations).

We used quantile regressions to analyse the relationships of climatic suitability and stability through time with genetic diversity [73]. For this, we calculated the difference of ensemble suitability between LGM, mid-Holocene and present-day as a measure of climate stability through time. Next, we analysed whether historical changes in species’ geographical range generated a cline spatial pattern in genetic parameters, haplotype ( $h$ ) and nucleotide ( $\pi$ ) diversities,  $N_e$  and  $\theta$  due to expansion of climatically suitable conditions. For this, we obtained the distance between each sampled population and the centroid of historical refugium and of the predicted distribution at 21 ka, 6 ka and present-day, and then we performed quantile regressions of genetic parameters against these spatial distances and against climatic stability.

### Additional files

**Additional file 1: Figures S1–S8.** with details on ecological niche modelling, networks and quantile regressions. (DOCX 3553 kb)

**Additional file 2: Tables S1–S11.** with sampling locations and details on ecological niche modelling and demographical parameters. (DOCX 143 kb)

### 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 (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 Modelling (Additional file 2: Table S1 and S3) for making available their model outputs.

### Availability of data and materials

Additional accessibility data is provided as Additional file.

### Authors’ contributions

RGC conceived the general idea of the study; LCV generated the genetic data; LCV, MSL-R, LCT and RGC performed the analyses and wrote the manuscript. All authors approved the final version of the text.

### Competing interests

The authors declare that they have no competing interests.

### Ethics approval and consent to participate

Not applicable.

### Author details

<sup>1</sup>Laboratório de Genética & Biodiversidade, Instituto de Ciências Biológicas, (ICB), Universidade Federal de Goiás (UFG), Cx.P. 131, Goiânia, GO 74001-970, Brazil. <sup>2</sup>Laboratório de Macroecologia, Universidade Federal de Goiás (UFG), Campus Jataí, Cx.P. 03, Jataí, GO 75801-615, Brazil.

Received: 18 April 2016 Accepted: 29 September 2016

Published online: 13 October 2016

### References

- Caetano S, Prado D, Pennington RT, Beck S, Oliveira-Filho A, Spichiger R, Naciri Y. The history of seasonally dry tropical forests in eastern South America: inferences from the genetic structure of the tree *Astronium urundeuva* (Anacardiaceae). *Mol Ecol*. 2008;17:3147–59.
- Collevatti RG, Terribile LC, Lima-Ribeiro MS, Nabout JC, Oliveira G, Rabelo SG, et al. A coupled phylogeographical and species distribution modelling approach recovers the demographical history of a Neotropical seasonally dry forest tree species. *Mol Ecol*. 2012;21:5845–63.
- Prado DE, Gibbs PE. Patterns of species distributions in the dry seasonal forests of South America. *Ann Miss Bot Gard*. 1993;80:902–27.
- Pennington RT, Prado DE, Pendry CA. Neotropical seasonally dry forests and quaternary vegetation changes. *J Biogeogr*. 2000;27:261–73.
- Mayle FE, Beerling DJ, Gosling WD, Bush MB. Responses of Amazonian ecosystems to climatic and atmospheric carbon dioxide changes since the last glacial maximum. *Phil Trans R Soc Lond B*. 2004;359:499–514.
- Mayle FE. The late quaternary biogeographical history of South American seasonally dry tropical forests: insights from Palaeo-ecological data. In: Pennington R, Lewis GP, Ratter JA, editors. *Neotropical savannas and seasonally dry forests: plant diversity, biogeography, and conservation*. New York: CRC press; 2006. p. 395–416.
- Vieira FA, Novaes RML, Fajardo CG, dos Santos RM, Almeida HS, Carvalho D, Lovato MB. Holocene southward expansion in seasonally dry tropical forests in South America: phylogeography of *Ficus bonijesulapensis* (Moraceae). *Bot J Linn Soc*. 2015;177:189–201.
- Côrtes ALA, Rapini A, Daniel TF. The Tetramerium lineage (Acanthaceae: Justiceaeae) does not support the Pleistocene Arc hypothesis for South American seasonally dry forests. *Am J Bot*. 2015;102:1–16.
- Oliveira PE, Barreto AMF, Suguio K. Late Pleistocene Holocene climatic and vegetational history of the Brazilian caatinga: the fossil dunes of the middle São Francisco River. *Palaeogeogr Palaeoclim Palaeoecol*. 1999;145:319–37.
- Nascimento LRS. Dinâmica vegetacional e climática holocênica da Caatinga, na região do Parque Nacional do Camtimbau, Buíque – PE. Masters’ Thesis, CTG, Universidade Federal de Pernambuco; Recife, Brazil. 2008.
- Collevatti RG, Lima-Ribeiro MS, Diniz-Filho JAF, Oliveira G, Dobrovolski R, Terribile LC. Stability of Brazilian seasonally dry forests under climate change: Inferences for long-term conservation. *Am J Plant Sci*. 2013;4:792–805.
- Becerra JX. Timing the origin and expansion of the Mexican tropical dry forest. *Proc Natl Acad Sci U S A*. 2005;102:10919–23.
- Queiroz LP, Lavin M. *Coursetia* (Leguminosae) from eastern Brazil: nuclear ribosomal and chloroplast DNA sequence analysis reveal the monophyly of three caatinga-inhabiting Species. *Syst Bot*. 2011;36:69–79.
- Burnham RJ, Graham A. The history of Neotropical vegetation: new developments and status. *Ann Missouri Bot Gard*. 1999;86:546–89.
- Burnham RJ, Johnson KR. South American palaeobotany and the origins of Neotropical rainforests. *Phil Trans R Soc B*. 2004;359:1595–610.
- Davis CC, Webb CO, Wurdack KJ, Jaramillo CA, Donoghue MJ. Explosive radiation of malpighiales supports a mid-cretaceous origin of modern tropical rain forests. *Amer Nat*. 2004;165:E36–65.

17. Miles L, Newton AC, De Fries RS, Ravilious C, May I, Blyth S, et al. A global overview of the conservation status of tropical dry forests. *J Biogeogr.* 2006;33:491–505.
18. Murphy PG, Lugo AE. Dry forests of central America and the Caribbean. In: Bullock SH, Mooney HA, Medina E, editors. *Seasonally dry tropical forests*. Cambridge: Cambridge University Press; 1995. p. 9–34.
19. Oliveira-Filho AT, Ratter JA. Vegetation physiognomies and woody flora of the Cerrado biome. In: Oliveira OS, Marquis RJ, editors. *The Cerrados of Brazil: ecology and natural history of a Neotropical savannah*. New York: Columbia University Press; 2002. p. 91–120.
20. Furlley PA, Ratter JA. Soil resources and plant communities of the central Brazilian cerrado and their development. *J Biogeogr.* 1988;15:97–108.
21. Collevatti RG, Terribile LC, Lima-Ribeiro MS, Nabout JC, Rangel TF, Diniz-Filho JAF. Drawbacks to palaeodistribution modelling: the case of South American seasonally dry forests. *J Biogeogr.* 2013;40:345–58.
22. Collevatti RG, Lima-Ribeiro MS, Terribile LC, Guedes LB, Rosa FF, Telles MP. Recovering species demographic history from multi-model inference: the case of a Neotropical savanna tree species. *BMC Evol Biol.* 2014;14:213. doi:10.1186/s12862-014-0213-0.
23. Collevatti RG, Terribile LC, Diniz-Filho JAF, Lima-Ribeiro MS. Multi-model inference in comparative phylogeography: an integrative approach based on multiple lines of evidence. *Front Genet.* 2015;6:1–8.
24. Schulze M, Grogan J, Uhi C, Lentini M, Vidal E. Evaluating ipê (Tabebuia, Bignoniaceae) logging in Amazonia: sustainable management or catalyst for forest degradation? *Biol Cons.* 2008;141:2071–85.
25. Behling H. Late glacial and Holocene vegetation, climate and fire history inferred from Lagoa Nova in the southeastern Brazilian lowland. *Veg Hist Archaeobot.* 2003;12:263–70.
26. Mayle FE. Assessment of the Neotropical dry forest refugia hypothesis in the light of palaeoecological data and vegetation model simulations. *J Quat Sci.* 2004;19:713–20.
27. Arenas M, Ray N, Currat M, Excoffier L. Consequences of range contractions and range shifts on molecular diversity. *Mol Biol Evol.* 2012;29:207–18.
28. Excoffier L, Ray N. Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol Evol.* 2008;23:347–51.
29. Hewitt GM. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol J Linn Soc.* 1996;58:247–76.
30. Excoffier L, Foll M, Petit RJ. Genetic consequences of range expansions. *Ann Rev Ecol Evol Syst.* 2009;40:481–501.
31. Waters JM, Fraser CJ, Hewitt GM. Founder takes all: density-dependent processes structure biodiversity. *Trends in Ecol Evol.* 2013;28:78–85.
32. Ohta T. Some models of gene conversion for treating the evolution of multigene families. *Genetics.* 1984;106:517–28.
33. Auler AS, Wang X, Edwards RL, Cheng H, Cristalli PS, Smart PL, et al. Quaternary ecological and geomorphic changes associated with rainfall events in presently semi-arid northeastern Brazil. *J Quat Sci.* 2004;19:693–701.
34. Santos AMM, Cavalcanti DR, da Silva JMC, Tabarelli M. Biogeographical relationships among tropical forests in northeastern Brazil. *J Biogeogr.* 2007;34:437–46.
35. Batalha-Filho H, Fjeldsa J, Fabre P-H, Miyaki. Connections between the Atlantic and the Amazonian forest avifaunas represent distinct historical events. *J Ornithol.* 2013;154:41–50.
36. Costa LP. The historical bridge between the Amazon and the Atlantic forest of Brazil: a study of molecular phylogeography with small mammals. *J Biogeogr.* 2003;30:71–86.
37. Wang X, Auler AS, Edwards RL, Cheng H, Cristalli PSP, Smart L, et al. Wet periods in northeastern Brazil over the past 210 kyr linked to distant climate anomalies. *Nature.* 2004;432:740–3.
38. Hoon C, Wesselingh FP, ter Steege H, Bermutez MA, Mora A, Sevink J, et al. Amazonia through time: andean uplift, climate change, landscape evolution and biodiversity. *Science.* 2010;330:927–31.
39. Saia SEMG, Pessenda LCR, Gouveia SEM, Aravena R, Bendassollia JA. Last glacial maximum (LGM) vegetation changes in the Atlantic Forest, southeastern Brazil. *Quat Int.* 2008;184:195–201.
40. Behling H. South and southeast Brazilian grasslands during late quaternary times: a synthesis. *Palaeogeogr Palaeoclimatol Palaeoecol.* 2002;177:19–27.
41. Behling H. Tropical mountain forest dynamics in Mata Atlantica and northern Andean biodiversity hotspots during the late quaternary. *Biodivers Ecol.* 2008;2:25–33.
42. Gibbard P, van Kolfschoten T. The Pleistocene and Holocene Epochs. In: Gradstein FM, Ogg JG, Smith AG, editors. *A geologic time scale 2004*. Cambridge: Cambridge University Press; 2005. p. 441–52.
43. Kingman JFC. The coalescent. *Stoc Proc Appl.* 1982;13:235–48.
44. Doyle JJ, Doyle JL. Isolation of plant DNA from fresh tissue. *Focus.* 1987;12:13–5.
45. Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, et al. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am J Bot.* 2005;92:142–66.
46. Desfeux C, Lejeune B. Systematics of Euro Mediterranean *Silene* (Caryophyllaceae): evidence from a phylogenetic analysis using ITS sequence. *Acad Sci Paris.* 1996;319:351–8.
47. Sievers F, Wilm A, Dineen DG, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 2011;7:539–44.
48. Forster P, Bandelt HJ, Röhl A. Network 4.2.0.1. 2004. Available from [www.fluxus-engineering.com](http://www.fluxus-engineering.com). Fluxus Technology Ltd.
49. Nei M. *Molecular evolutionary genetics*. New York: Columbia University Press; 1987. p. 512.
50. Excoffier L, Laval G, Schneider S. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol Bioinform.* 2005;1:47–50.
51. Excoffier L, Smouse P, Quattro J. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics.* 1992;131:479–91.
52. Corander J, Marttinen P, Sirén J, Tang J. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinform.* 2008;9:1–14.
53. Heled J, Drummond AJ. Bayesian inference of population size history from multiple loci. *BMC Evol Biol.* 2008;8:1–15.
54. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology And Evolution.* 2012;29:1969–73.
55. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods.* 2012;9:772.
56. Collevatti RG, Terribile LC, Rabelo SG, Lima-Ribeiro MS. 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.* 2015;6:1–15.
57. Rambaut A, Suchard MA, Xie D & Drummond AJ. Tracer v1.6. 2014. Available from <http://beast.bio.ed.ac.uk/Tracer>.
58. Beerli P, Felsenstein J. Maximum likelihood estimation of a migration matrix and effective population sizes in subpopulations by using a coalescent approach. *Proc Natl Acad Sci U S A.* 2001;98:4563–8.
59. Kuhner MK. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics.* 2006;22:768–70.
60. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol.* 2005;22:1185–92.
61. Lima-Ribeiro MS, Varela S, González-Hernández J, Oliveira G, Diniz-Filho JAF, Terribile LC. The EcoClimate Database. 2015. Available from <http://ecoclimate.org>. Accessed in 20 Feb 2015.
62. Thuiller W, Brotons L, Araújo MB, Lavorel S. Effects of restricting environmental range of data to project current and future species distributions. *Ecography.* 2004;27:165–72.
63. VanderWal J, Shoo LP, Graham C, Williams SE. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? *Ecol Model.* 2009;22:589–94.
64. Diniz-Filho JAF, Bini LM, Rangel TF, Loyola RD, Hof C, Nogués-Bravo D, et al. Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography.* 2009;32:897–906.
65. Araújo MB, New M. Ensemble forecasting of species distributions. *Trends Ecol Evol.* 2007;22:42–7.
66. Allouche O, Tsoar A, Kadmon R. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J Appl Ecol.* 2006;43:1223–32.
67. Terribile LC, Lima-Ribeiro MS, Araújo MB, Bizão N, Collevatti RG, Dobrovolski R, et al. Areas of climate stability in the Brazilian Cerrado, disentangling uncertainties through time. *Nat Conserv.* 2012;10:152–9.
68. Excoffier L, Novembre J, Schneider S. SIMCOAL: a general coalescent program for simulation of molecular data in interconnected populations with arbitrary demography. *J Hered.* 2000;91:506–9.

69. Anderson CNK, Ramakrishnan U, Chan YL, Hadly EA. Serial SimCoal: A population genetic model for data from multiple populations and points in time. *Bioinformatics*. 2005;21:1733–4.
70. Burnham KP, Anderson DR. Model selection and multimodel inference: a practical information-theoretic approach. New York: Springer; 2002. p. 488.
71. Zuur AF, Ieno EN, Walker N, Saveliev AA, Smith GM. Mixed effects models and extensions in ecology with R. In: Gail M, Krickeberg m, Samet JM, Tsiatis A, Wong W, editors. *Statistics for Biology and Health*. New York: Springer; 2009. p. 261–93.
72. Wright S. Isolation by distance. *Genetics*. 1943;28:114–38.
73. Cade BS, Noon BR. A gentle introduction to quantile regression for ecologists. *Front Ecol Environ*. 2003;1:412–20.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

