CrossMark

ORIGINAL ARTICLE

Gene flow between vicariant tree species: insights into savanna-forest evolutionary relationships

Luciana C. Resende-Moreira[1] · Ana Carolina S. Ramos[2] · Marília O. Scliar[1] ·
Reinaldo M. Silva[3] · Vânia C. R. Azevedo[4] · Ana Y. Ciampi[4] · José P. Lemos-Filho[5] ·
Maria Bernadete Lovato[1]

Abstract Studying the genetic structure of vicariant species (i.e., closely related species that occupy ecologically distinct yet adjacent habitats) can shed light on the evolution and divergence of species with different ecological requirements. A previous phylogeographic study identified chloroplast DNA haplotype sharing between two vicariant tree species, one from forest (*Hymenaea courbaril*) and one from savanna (*H. stigonocarpa*) habitats. These species co-occur in the Brazilian Cerrado, a biome that encompasses forest patches and riverine forests within a savanna matrix. In order to investigate the evolutionary processes involved in the genetic divergence of these trees, we used nuclear microsatellite markers, statistical methods including approximate Bayesian computation (ABC), and leaf morphology to analyze neighboring and distant populations. Bayesian analysis revealed admixture between the species. ABC analysis supported the scenarios with the occurrence of gene flow between species during the Last Glacial Maximum or from the Holocene to the present, when compared to alternative scenarios of no gene flow or constant gene flow since divergence. However, putative hybrids did not exhibit intermediate leaflet morphology, which could be related to distinct selective pressures maintaining species integrity even in the face of gene flow. Our results suggest that despite morphological differences between savanna and forest species, interspecific barriers to gene flow might not be fully developed between vicariant tree species and that interspecific hybridization in trees from Cerrado biome may be an underdiagnosed process.

Luciana C. Resende-Moreira and Ana Carolina S. Ramos contributed equally to this work.

Communicated by S. C. González-Martínez

✉ Maria Bernadete Lovato
lovatomb@icb.ufmg.br; bernadete.lovato@gmail.com

1 Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, CP 486, Belo Horizonte, MG 31270-901, Brazil

2 Associação Educacional de Vitória, FAESA, Av. Vitória, 2200, Vitória, ES 29053-360, Brazil

3 Laboratório de DNA, Instituto de Criminalística, Polícia Civil de Minas Gerais, Rua Juiz de Fora, 400, Belo Horizonte, MG 30800-060, Brazil

4 Laboratório de Genética Vegetal, EMBRAPA Recursos Genéticos e Biotecnologia, CP 02372, Brasília, DF 70770-917, Brazil

5 Departamento de Botânica, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, CP 486, Belo Horizonte, MG 31270-901, Brazil

Keywords Cerrado · Hybridization · *Hymenaea* · Leaf morphology · Microsatellites

Introduction

Comparison of the genetic structure of closely related species occurring in distinct but geographically adjacent habitats might shed light on the evolution and the divergence of species with different ecological requirements. The Cerrado and the Atlantic forest are two neighboring biomes in Eastern Tropical South America (ETSA), considered "hotspots" for conservation priorities (Mittermeier et al. 2004). The Cerrado biome is the world's most species-rich tropical savanna comprising many physiognomies with different tree densities. In addition to savanna physiognomies, the Cerrado biome includes riverine forests along the watercourses and deciduous and semidecidous forests in areas with more fertile soils

Springer

(Ratter et al. 2006). Climatic conditions, especially those related to water availability such as rainfall and dry season length, and fire frequency seem to be associated with the occurrence of each biome (Eiten 1972; Rizzini 1997; Oliveira-Filho and Ratter 2002). Compared with the Cerrado, the Brazilian Atlantic forest usually displays lower light accessibility, higher water availability, and soil that is richer in nutrients and has lower aluminum content (Eiten 1972; Furley 1992; Ratter et al. 2006). Species occurring in these biomes display morphological and physiological features resulting from adaptation to characteristics of each environment, mainly related to the presence of fire in savanna (Hoffmann and Franco 2003; Goulart et al. 2011). Some of the characteristics related to distinct selective pressures include tree architecture and leaf morphology (Ackerly et al. 2000; Archibald and Bond 2003; Hovenden and Vander Schoor 2005; Barros et al. 2012).

Recent phylogenetic studies have suggested that Cerrado species are mainly derived from lineages of forest ancestry, such as those present in the Atlantic forest, in the Amazon forest, and in seasonally dry forests (Simon et al. 2009; Duarte et al. 2011; Fiaschi and Plunkett 2011; Simon and Pennington 2012). In addition, some of these studies show that the Cerrado is a recent biome, with diversification of most of its lineages dating from less than 10 Myr (Simon et al. 2009; Simon and Pennington 2012). The close phylogenetic relationship among many forest and savanna species and the strong adaptive differences between these habitats suggest that ecological speciation might be an important process related to tree diversification in the Cerrado biome (Pennington et al. 2006; Simon and Pennington 2012). Pairs of closely related species that occupy distinct habitats like savanna and forest have been called vicariant species by botanists (Rizzini 1997). Vicariant species may co-occur in some areas within the Cerrado biome, especially where forest patches and riverine forests are present, providing a large geographic interface, or contact zone, between savanna and forest species.

Although the geographical proximity of closely related species may allow interspecific gene flow if reproductive barriers are not fully developed, only a handful of natural hybridization events between native tree species have been described at ETSA. Interspecific hybridization between Cerrado trees has been suggested previously for some pairs of species such as *Handroanthus chrysotrichus* and *H. ochraceus* (Bittencourt and Moraes 2010) and also for *Kielmeyera coriaceae* and *K. grandiflora* (Caddah et al. 2013). In addition, studies have suggested a putative hybrid origin for some Cerrado tree species such as *Eriotheca pubescens* (Forni-Martins et al. 1995), *Tocoyena formosa* (Silberbauer-Gottsberger et al. 1992), and *Anemopaegma arvense* s.l. species complex (Firetti-Leggieri et al. 2011). Most of these species are polyploid, a trait present in many other Cerrado species (Forni-Martins and Martins 2000), which could suggest hybrid origin for other Cerrado

trees. Also, a few molecular studies have reported intraspecific gene flow between highly divergent lineages, evidencing that hybridization might occur at different stages of divergence in trees present in this biome (Cavallari et al. 2010; Garcia et al. 2011; Lacerda et al. 2002; Warwick and Lewis 2003).

Environmental differences could result in divergent lineages either in sympatric or allopatric populations. In such cases, barriers to gene flow usually evolve as a by-product of adaptations driven by different environmental selective pressures (Schluter 2001). If selection is strong enough and gene flow is not very common, reproductive isolation might slowly arise in sympatry, but this process often starts in allopatry, and if the range distribution of different lineages changes over time allowing secondary contact, gene flow among them might promote hybridization. In the contact zone, hybridization could break down the barriers to gene exchange (Grant and Grant 2014), maintain the current lineage divergence through persistent balance between selection and hybridization, or even strengthen divergence, ultimately resulting in different species (reviewed in Abbott et al. 2013). While some hybrids are less competitive than parental species, others might display mechanisms responsible for hybrid persistence such as asexual reproduction, polyploidy, and occupation of new niches (reviewed in Vallejo-Marín and Hiscock 2016).

Hybridization events are usually described when morphologically intermediate individuals are found. However, hybrids may also express a high proportion of characters of a single parent (Rieseberg et al. 1993), and hybridization may therefore be underestimated if molecular markers are not employed. Recently diverged species usually share haplotypes/alleles, which might be related to hybridization and introgression or ancestral polymorphism (incomplete lineage sorting) (Muir and Schlotterer 2005; Lexer et al. 2006; Wang et al. 2011). In ETSA, haplotype sharing was found between *Hymenaea stigonocarpa* and *H. courbaril*, a pair of vicariant trees (Ramos et al. 2007, 2009). These species shared the most common chloroplast DNA (cpDNA) haplotypes and displayed similar phylogeographic structure, but the authors could not distinguish whether haplotype sharing was due to ancestral polymorphism or hybridization between these species (Ramos et al. 2009). The former is endemic to the Cerrado biome, and the latter is frequently found in the Atlantic forest and in riverine forests and forest patches within the Cerrado. Despite the lack of phylogenetic studies including all 17 species belonging to this genus (Lee and Langenheim 1975; Ribeiro et al. 2015), Fougère-Danezan et al. (2007) analyzed cpDNA of six species and observed that *H. courbaril*, *H. eriogyne*, and *H. stigonocarpa* are part of a monophyletic, closely related, and unresolved clade within the genus. In addition, an early review of the genus suggested that savanna species like *H. stigonocarpa* might have derived from forest *Hymenaea* lineages (Lee and Langenheim 1975).

In order to investigate the evolutionary processes that have shaped the divergence of vicariant Neotropical tree species of savanna and forest, we used *H. stigonocarpa* and *H. courbaril* as a case study. We analyzed neighboring and distant populations of these species for nuclear microsatellite markers (nSSRs) and leaf morphology. Specifically, we addressed the following questions regarding their interspecific relationship: (1) Do nSSR markers reveal genetic evidence of hybridization between these vicariant species? (2) If they do, when did the hybridization happen? (3) Is there any evidence of hybridization between the two species based on leaf morphology? By answering these questions we attempted to unravel the causes of polymorphism sharing in vicariant ETSA tree species in order to further understand the evolutionary relationship between savanna and forest vegetation.

Materials and methods

Studied species, sampling, and DNA extraction

Hymenaea courbaril and *H. stigonocarpa* are a pair of vicariant tree species (Heringer et al. 1976). The first species occurs mainly in the Atlantic forest but is also found in forest patches and riverine forest within the Cerrado biome, while the second species is restricted to the Brazilian Cerrado (Lee and Langenheim 1975). These species are closely related (Fougère-Danezan et al. 2007) and share many life-history traits: bats are the main pollinators (Gibbs et al. 1999; Dunphy et al. 2004), and agoutis are probably the main seed dispersers (Asquith et al. 1999). Eight populations of *H. courbaril* located in the Brazilian Atlantic forest or forest sites in the Cerrado biome were sampled, with a total of 142 individuals, and 11 populations of *H. stigonocarpa* were sampled in the Cerrado, with a total of 210 individuals (Table 1 and Fig. 1). In order to facilitate the sampling identification, abbreviations of all populations of *H. courbaril* ended with the letter M and abbreviations of *H. stigonocarpa* ended with the letter C (following Ramos et al. 2007, 2009). Whenever possible, one population of each species was sampled in close localities, and therefore, they have similar abbreviations with the exception of the last letter. Leaves were collected, identified, and ice frozen before permanent storage in −20 °C. Extraction of total DNA followed modifications of the CTAB protocol.

Nuclear microsatellite loci

Seven nSSR loci developed for *H. courbaril* and transferred to *H. stigonocarpa* (Ciampi et al. 2008) were used for the population genetic analyses in both species. Microsatellite loci were amplified using polymerase chain reaction (PCR) following the conditions established by Ciampi et al. (2008). Amplifications were confirmed with 3.5% agarose gel electrophoresis. Amplified products were analyzed in an ABI-Prism 377XL sequencer (Applied Biosystems, Life Technologies Ltd., Paisley, UK) with a ROX fluorescent-labeled internal size standard. The size of amplified fragments (in bp) was obtained with the program GeneScan (Applied Biosystems) and analyzed by the software Genotyper (Applied Biosystems).

Genetic data analyses

Genetic diversity levels for each population were estimated using the following statistics: mean number of alleles per locus ($\hat{A}$), mean-observed (H_O) and mean-expected (H_E) heterozygosities with the program Arlequin v. 3.5.1.2 (Excoffier and Lischer 2010), and allelic richness (A_R) using FSTAT v. 2.9.3.2 (Goudet 2001). Deviations from Hardy-Weinberg equilibrium (HWE) were assessed with allelic permutations among individuals within populations, and linkage disequilibrium between pairs of loci was evaluated using the program FSTAT. The presence of null alleles at each locus was tested by Monte Carlo simulation (1000 bootstraps) using MICRO-CHECKER v. 2.2.3 (Van Oosterhout et al. 2004).

Genetic structure was assessed using several different approaches. Analysis of molecular variance (AMOVA) based on the sum of squared size differences implemented in Arlequin was computed to describe the degree of genetic structure between species, among populations within species, and within populations. Differentiation among pairs of populations was estimated by pairwise F_{ST} using significance tests at the $P < 0.05$ level followed by a Bonferroni correction in the software FSTAT. In addition, pairwise Cavalli-Sforza and Edwards chord distances (Cavalli-Sforza and Edwards 1967) were calculated for all populations using Microsatellite Analyzer (MSA) (Dieringer and Schlotterer 2003) with 1000 randomized permutations to assess statistical support. Trees were constructed using the neighbor-joining algorithm, and a majority rule bootstrap consensus tree was constructed with the programs Neighbor and Consense in PHYLIP v. 3.695 (http://evolution.genetics.washington.edu/phylip.html) and visualized with the software TreeView v. 1.6.6 (Page 1996). Genetic structure was also evaluated by a principal components analysis (PCA) using the *adegenet* and *ade4* (Jombart 2008) packages in R v. 3.0.1 (R Foundation for Statistical Computing, http://www.R-project.org).

Additionally, Bayesian clustering analyses were performed to identify clusters (K) of genetically similar individuals and to look for evidence of interspecific admixture using the software STRUCTURE v. 2.3.4 (Pritchard et al. 2000). Three

Table 1 Sampling localities, information about hybrids, genetic diversity, and inbreeding coefficient across microsatellite loci for each population sampled

Population code[a]	Locality	State	N	N_H	%H	$\hat{A}$	A_R	H_E	H_O	F_{IS}
H. courbaril										
PAM	Chapada Diamantina National Park	BA	20	1	5	6.71	4.89	0.684	0.521	0.242*
RLM	Reserva de Linhares	ES	10	0	0	5.29	4.54	0.628	0.557	0.118
CPM	Rio Doce State Park	MG	20	5	25	8.43	6.01	0.804	0.739	0.082
FBMI	Fazenda Brejão I	MG	20	2	10	6.00	4.46	0.623	0.469	0.253*
FBMII	Fazenda Brejão II	MG	21	0	0	4.29	3.39	0.536	0.495	0.087
FUM	Furnas	MG	21	0	0	7.86	5.79	0.795	0.739	0.072
RPM	São Gonçalo do Rio Preto	MG	10	3	30	5.57	5.04	0.745	0.490	0.356*
SCM	Serra do Cipó National Park	MG	20	0	0	5.71	4.22	0.618	0.466	0.251*
Overall			142	11	7.75	6.23	4.79	0.679	0.560	0.171*
H. stigonocarpa										
MUC	Chapada Diamantina National Park	BA	20	4	20	4.86	4.16	0.574	0.411	0.290*
RCC	Rio de Contas	BA	21	2	9.5	4.29	3.68	0.550	0.366	0.340*
ADC	Abadia dos Dourados	MG	12	3	25	4.29	4.20	0.587	0.421	0.292*
BIC	Biribiri	MG	17	0	0	4.00	3.48	0.496	0.379	0.243*
FBC	Fazenda Brejão	MG	22	1	4.5	4.29	3.55	0.526	0.480	0.090
FUC	Furnas	MG	21	3	14.3	5.57	4.46	0.540	0.400	0.264*
MCC	Montes Claros	MG	20	3	15	5.00	3.97	0.438	0.354	0.197*
PEC	Vale do Peruaçu National Park	MG	20	1	5	5.14	4.28	0.550	0.309	0.445*
RPC	Rio Preto State Park	MG	17	1	5.9	4.86	4.03	0.547	0.451	0.180
SCC	Serra do Cipó National Park	MG	20	1	5	4.43	3.73	0.557	0.370	0.342*
VPC	Várzea da Palma	MG	20	0	0	4.71	3.63	0.443	0.317	0.290*
Overall			210	19	9.05	4.83	3.93	0.507	0.366	0.272*

N number of individuals, N_H number of putative hybrids according to the *STRUCTURE* analysis ($0.2 \le Q \le 0.8$), *%H* percentage of putative hybrids, $\hat{A}$ mean number of alleles per locus, A_R allelic richness, H_E expected heterozygosity, H_O observed heterozygosity, F_{IS} inbreeding coefficient

[a] The third letter of each population code represents the species. M = *H. courbaril* (forest species) and C = *H. stigonocarpa* (savanna species) following Ramos et al. (2007, 2009)

*Significant values (*P* values smaller than the adjusted nominal level—5%—after allelic permutations among individuals within samples)

different analyses were conducted, one for each species separately and one including all sampled individuals of both species. Analyses were carried out with 10 independent runs of 2×10^5 burn-in-length Markov Chain Monte Carlo (MCMC) iterations and 5×10^5 additional simulations, testing the probable number of clusters ranging from one to the number of sampled populations in each analysis plus one, except when both species were included and the specified *K* ranged from one to six. Parameter sets assumed the admixture model without prior population information and allele frequency correlation among populations. The optimal number of clusters was determined following the recommendations of Evanno et al. (2005). We used STRUCTURE to identify putative hybrids because the overall performance of this software in identifying hybridization is satisfactory even when a small number of loci are used (Vähä and Primmer 2006). The admixture proportion (*Q*) obtained with STRUCTURE analysis was used to classify individuals as putative hybrids with a threshold of *Q* = 0.8, as a compromise between efficiency and accuracy (Vähä and Primmer 2006). Hybrids were identified
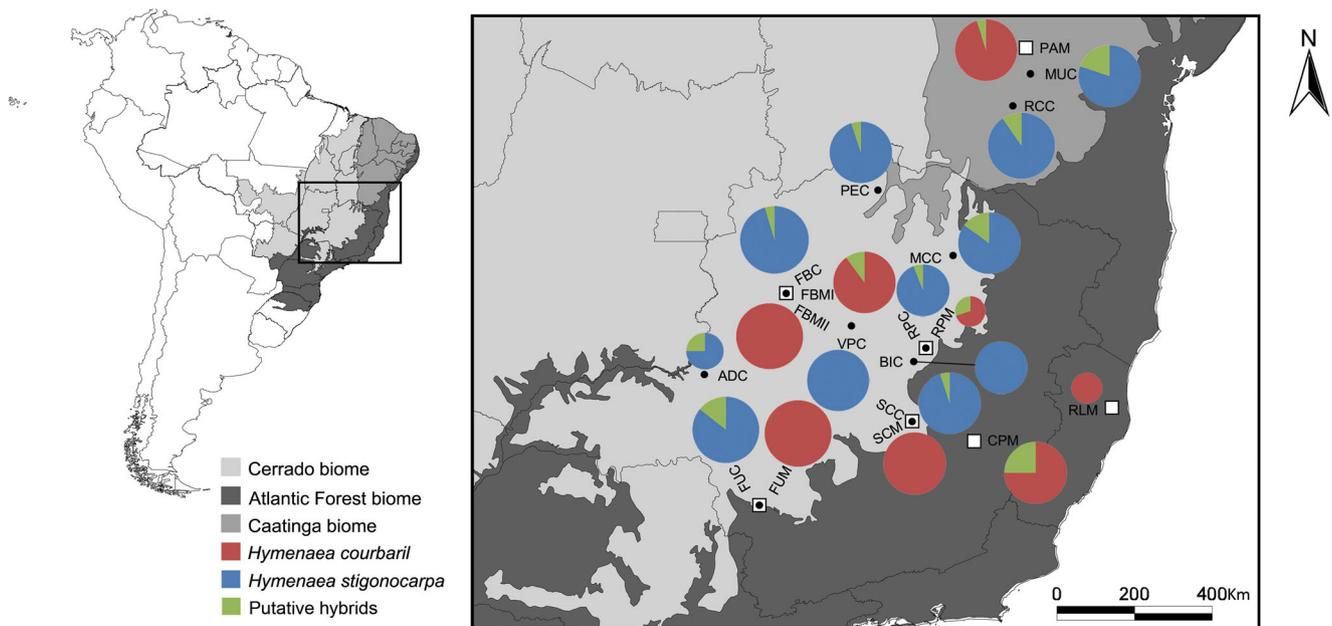
Fig. 1 Geographical location of sampled populations of *Hymenaea courbaril* (*white squares*) and *H. stigonocarpa* (*black dots*) and distribution of the individuals classified as belonging to one of each species or putative hybrids. Hybrids were identified based on assignment probabilities ($0.2 \leq Q \leq 0.8$) for each individual according to STRUCTURE results for $K = 3$ in the combined dataset of both species. *Circle sizes* are proportional to the number of sampled individuals. For population codes, see Table 1

conservatively because morphologically intermediate individuals between the two species have not been described.

In order to assess STRUCTURE's ability to identify hybrids correctly, we used HYBRIDLAB (Nielsen et al. 2006) to simulate a scenario of complete isolation between the three groups identified with STRUCTURE, two groups within *H. courbaril* and one for *H. stigonocarpa* (see "Results" section). Alleles were randomly drawn from each group to generate three populations of 300 "pure" individuals, two populations of *H. courbaril* (one using individuals from FBMI, FBMII, and PAM and one with individuals from the other populations), and one of *H. stigonocarpa*. The simulated datasets were then analyzed together in STRUCTURE with the same settings used previously. The probability of misclassification of hybrids was estimated from the percentage of simulated individuals with the admixture proportion of $0.2 \leq Q \leq 0.8$. This method has been used successfully in other studies, and the inclusion of putative admixed individuals is considered a conservative approach (Vähä and Primmer 2006; Marino et al. 2013).

As the inference of K can be sensitive to the presence of isolation by distance (IBD) in STRUCTURE (Pritchard et al. 2000), we performed a Mantel test in the program Arlequin using the correlation between the matrix of pairwise Slatkin's linearized F_{ST} (Slatkin 1995) and the matrix of the natural logarithm of approximate geographic distances (in km). The significance of the Mantel test was assessed with 1000 randomizations.

Test of evolutionary scenarios using approximate Bayesian computation

To further investigate gene flow between *H. courbaril* and *H. stigonocarpa*, we used approximate Bayesian computation (ABC). ABC is a statistical framework within which one can test the relative probability of a set of alternative evolutionary models of interest and estimate the parameters of the best model. This approach has been successfully used in other studies to investigate the occurrence and time of gene flow (Camargo et al. 2012; Marino et al. 2013; Turchetto et al. 2014). Here, we tested four alternative models assuming different schemes of gene flow or no gene flow between the two species after divergence from an ancestral population (Fig. 2). Specifically, the NoMig model assumes isolation after divergence, the ConstMig model assumes constant gene flow since divergence, and the RecMig and AncMig models assume recent gene flow beginning in Holocene (10,000 years ago to present) and ancient gene flow during the Last Glacial Maximum (LGM; 23,000 to 19,000 years ago), respectively. In all models, the two populations were allowed to change their effective population sizes (N) since time of divergence (t) (Online Resource 1 Fig. S1).

We performed the analysis in a dataset of 10 randomly chosen individuals from each of the 19 populations of both species (dataset ALLPOP). Also, considering that STRUCTURE suggested the existence of two groups within *H. courbaril* (see "Results" section), each model was tested
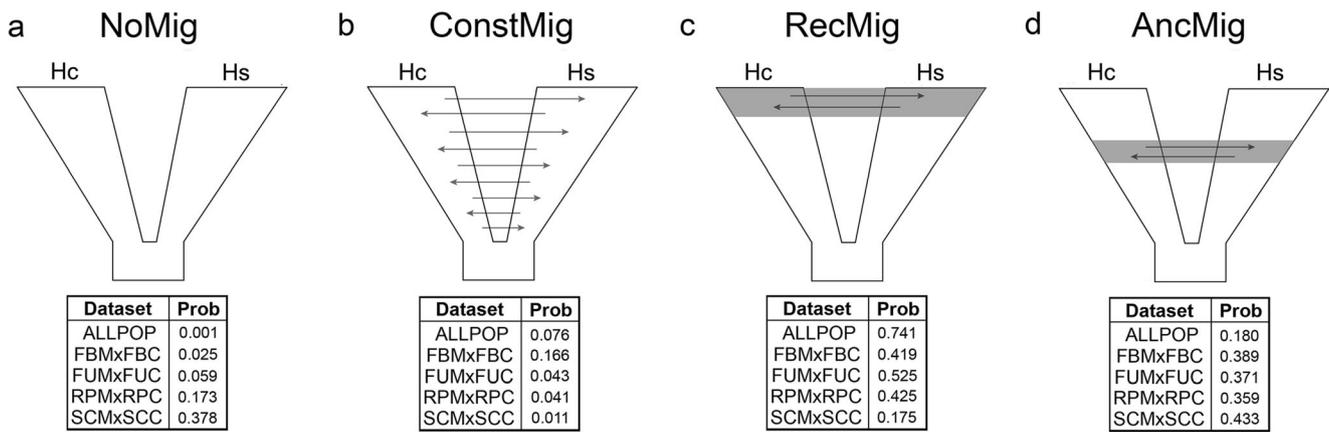
Fig. 2 Four alternative evolutionary models tested in approximate Bayesian computation analysis. Models assume **a** isolation after divergence—NoMig, **b** constant gene flow since divergence—ConstMig, **c** recent gene flow beginning in Holocene (10,000 years ago to present)—RecMig, and **d** ancient gene flow during the Last Glacial Maximum (LGM; 23,000 to 19,000 years ago)—AncMig. *Tables* display the posterior probabilities (Prob) for each model and dataset obtained with 20,000 retained simulations. See the text for dataset description. Hc and Hs stand for *Hymenaea courbaril* and *H. stigonocarpa*, respectively

using two different datasets for this species, one comprising only populations belonging to the first group (FBMI, FBMII, and PAM—dataset FBM_PAM × Stigo) and the other excluding these populations (dataset otherM × Stigo). In addition, to investigate the pattern of gene flow between neighbor populations of different species, we performed the ABC for the following four available pairs of co-located populations: FBM × FBC, FUM × FUC, RPM × RPC, and SCM × SCC. Since FBMI and FBMII are closely located, we pooled the two populations to constitute the FBM population.

For each combination of dataset and model, we did 1.5 million simulations of 7 nSSRs with the same sample size of each dataset using the program fsc25 (Excoffier and Foll 2011; Excoffier et al. 2013) within the ABCToolBox (Wegmann et al. 2010). The values of current and ancestral N were drawn from uniform prior distributions ranging from 100 to 25,000. A uniform prior distribution was also used for the t parameter, ranging from 6667 to 366,667 generations. To define this interval, we estimated the time to the most recent common ancestor (T_{MRCA}) in BEAST using cpDNA data from Ramos et al. (2007, 2009). Two coalescent models (constant population size and Bayesian skyline) and two cpDNA evolutionary rates described for angiosperms at neutral sites (0.0011 and 0.0029 sub/site/ Myr; Wolfe et al. 1987) were used under a strict clock model. For each evolutionary model and rate, we conducted three independent runs based on 30^7 generations, sampling every 10^3 generations. Convergence of results and effective sample sizes (ESSs) larger than 200 were observed in TRACER v1.5 (Rambaut and Drummond 2009). We used a combination of the minimum and maximum values obtained for T_{MRCA} from all different models as priors for t in the ABC analyses. Based on ex situ growth experiments, we assumed 15 years/generation (Sano and Fonseca 2003; Costa et al. 2011). For the

migration rates (m), we used log-uniform prior distributions ranging from 0.00001 to 0.01. The average mutation rate (μ) across loci was drawn from a log-uniform prior distribution ranging from 0.000001 to 0.001 (Thuillet et al. 2002; Vigouroux et al. 2002; Leonard et al. 2003), and single-locus mutation rates were drawn from a gamma distribution with mean μ and a shape parameter of 2 (Excoffier et al. 2005; Neuenschwander et al. 2008; Marino et al. 2013). The number of repeats by which a new mutant allele differ from its ancestor were generated following a generalized stepwise mutation (GSM) model (Estoup et al. 2002), with uniform distribution of the mean geometric parameter p ranging from 0 to 0.8 (Neuenschwander et al. 2008; Marino et al. 2013). Prior distributions are given in Online Resource 2 Table S1.

Summary statistics (SuSt) used to compare observed and simulated data in the ABC analysis were F_{ST} between species and the mean and standard deviation over loci for each population of the parameters: number of alleles (K), expected heterozygosity (H), and allelic range (R). All these SuSt were calculated with arlsumstat (Excoffier and Lischer 2010) inside ABCToolBox.

We used the regression approach (Beaumont 2008), testing different numbers of retained simulations (20,000, 40,000, and 80,000), to obtain the posterior probabilities of each model. The quality of model selection was evaluated by estimating the rate of true and false positives for each model. To do this, we first simulated 1000 datasets from the same prior distribution of parameters used for each combination of dataset and model considered. Each of these simulated datasets was then used as a pseudo-observed dataset (PODS), and the posterior probability of each model was estimated with the same ABC approach applied to the real dataset. The fraction of true positives for each model was assessed by the proportion of the 1000 PODSs simulated under a specific model that presented

the highest posterior probability for this specific model, and the fraction of false positives was the proportion of the remaining 3000 PODSs simulated under the other three different models that wrongly recovered this specific model.

For parameter estimation, we calculated the Euclidian distance between the observed and simulated SuSt of the best model and retained 0.5% of the total simulations corresponding to the shortest distances. Posterior probability for each parameter was estimated using a weighted local regression (Beaumont et al. 2002), after a log-tan transformation was applied to prevent the posterior distribution from exceeding the limits of the prior distribution (Hamilton et al. 2005). These steps were done using a modified version of the makepd4 R script (Scliar et al. 2014). We assessed the potential for a parameter to be correctly estimated by estimating the coefficient of determination (R^2—the proportion of parameter variance explained by the summary statistics) computed through all simulations.

Leaflet morphometric analysis

In their review of systematics of the *Hymenaea* genus, Lee and Langenheim (1975) made a detailed description of the morphology of each species. According to those authors, *H. courbaril* trees found in the area sampled in this study belong to the var. *stilbocarpa*. In addition, *H. courbaril* var. *stilbocarpa* and *H. stigonocarpa* present distinct tree architecture. *H. courbaril* trees are higher (12–20 m) and have smooth silvery gray bark (rarely rough with deep fissures and grooves), whereas *H. stigonocarpa* is shorter (5–12 m); the bark is very rough, deeply fissured, and grooved or rarely smooth or silvery gray. In addition, the two species differ in leaflet morphology. *H. courbaril* presents medium-sized and sub-falcate leaflets, and *H. stigonocarpa* presents large and obavate-oblong leaflets (occasionally falcate). Considering these overall distinctive characteristics, leaflet morphological data were collected from individuals present in some co-located populations (FU, SC, and FB) in order to test if there is any morphological evidence of hybridization. Seven leaflet characters were measured on mature leaflets and two derived variables were also obtained, for a total of nine variables analyzed. We acknowledge that derived variables are correlated with the original variables, but we considered them informative to our analyses. A detailed description of each variable is shown in Online Resource 1 Fig. S2. Five to 15 leaflets of each individual were measured, ranging from 9 to 17 individuals per population, in a total of 94 individuals analyzed.

Analyses of variance (ANOVAs) were performed in BioEstat v.5.0 (http://www.mamiraua.org.br/pt-br/downloads/programas/bioestat-versao-53/) in order to test for statistical morphological differences between pairs of neighboring populations. Other ANOVAs using all sampled individuals were performed to evaluate differences between the two species. In addition, a principal component analysis (PCA) was carried out using the *adegenet* and *ade4* packages in R to check for multivariate separation between species based on the morphological characters measured. Prior to each analysis, number and proportion variables were transformed (square root and arc-sine of the square root, respectively).

Results

Genetic diversity and intraspecific structure

No significant linkage disequilibrium was found between any pair of loci ($P > 0.002$—adjusted P value for 5% nominal level). Genetic diversity and fixation indices (F_{IS}) per population for each species are displayed in Table 1. *H. courbaril* populations harbor higher genetic diversity than *H. stigonocarpa*, even when neighboring populations are compared. The most diverse populations of *H. courbaril* were CPM, FUM, RPM, and PAM. For *H. stigonocarpa*, higher diversity was found in the populations ADC, PEC, FUC, and MUC. All sampled populations showed H_O values lower than H_E in both species. *H. stigonocarpa* populations showed higher F_{IS} values than *H. courbaril*, with values ranging from 0.090 to 0.445 and from 0.082 to 0.356, respectively. The F_{IS} values were significant in 9 out of 11 populations of *H. stigonocarpa* and in 4 out of 8 populations of *H. courbaril*. There was no evidence of null alleles at any locus. Considered together, these results suggest the occurrence of inbreeding in both species.

AMOVA indicated higher genetic structuring in *H. courbaril* ($F_{ST} = 0.299$) than in *H. stigonocarpa*, with F_{ST} value of 0.197 (Online Resource 2 Table S2). Significant pairwise F_{ST} ranged from 0.077 to 0.307 between populations of *H. courbaril* and from 0.045 to 0.240 between populations of *H. stigonocarpa* (Online Resource 2 Table S3).

After corrections according to Evanno et al. (2005), Bayesian analysis using the software STRUCTURE indicated that the most likely number of genetic clusters in *H. courbaril* was two, one cluster formed by populations PAM, FBMI, and FBMII and the other by the remaining populations (Fig. 3a and Online Resource 1 Fig. S3). The neighbor-joining consensus tree based on pairwise Cavalli-Sforza and Edwards chord distances also reveals two clusters in the *H. courbaril* clade with robust bootstrap support (Online Resource 1 Fig. S4). *H. stigonocarpa* populations were separated into three clusters but with highly admixed individuals and populations (Fig. 3b). Correlations between genetic and geographic distances (Mantel tests) were not significant for *H. stigonocarpa* ($r = 0.24$, $P = 0.07$) nor *H. courbaril* ($r = 0.21$, $P = 0.12$), suggesting that geographic distance was not an important factor influencing genetic structure in these species.
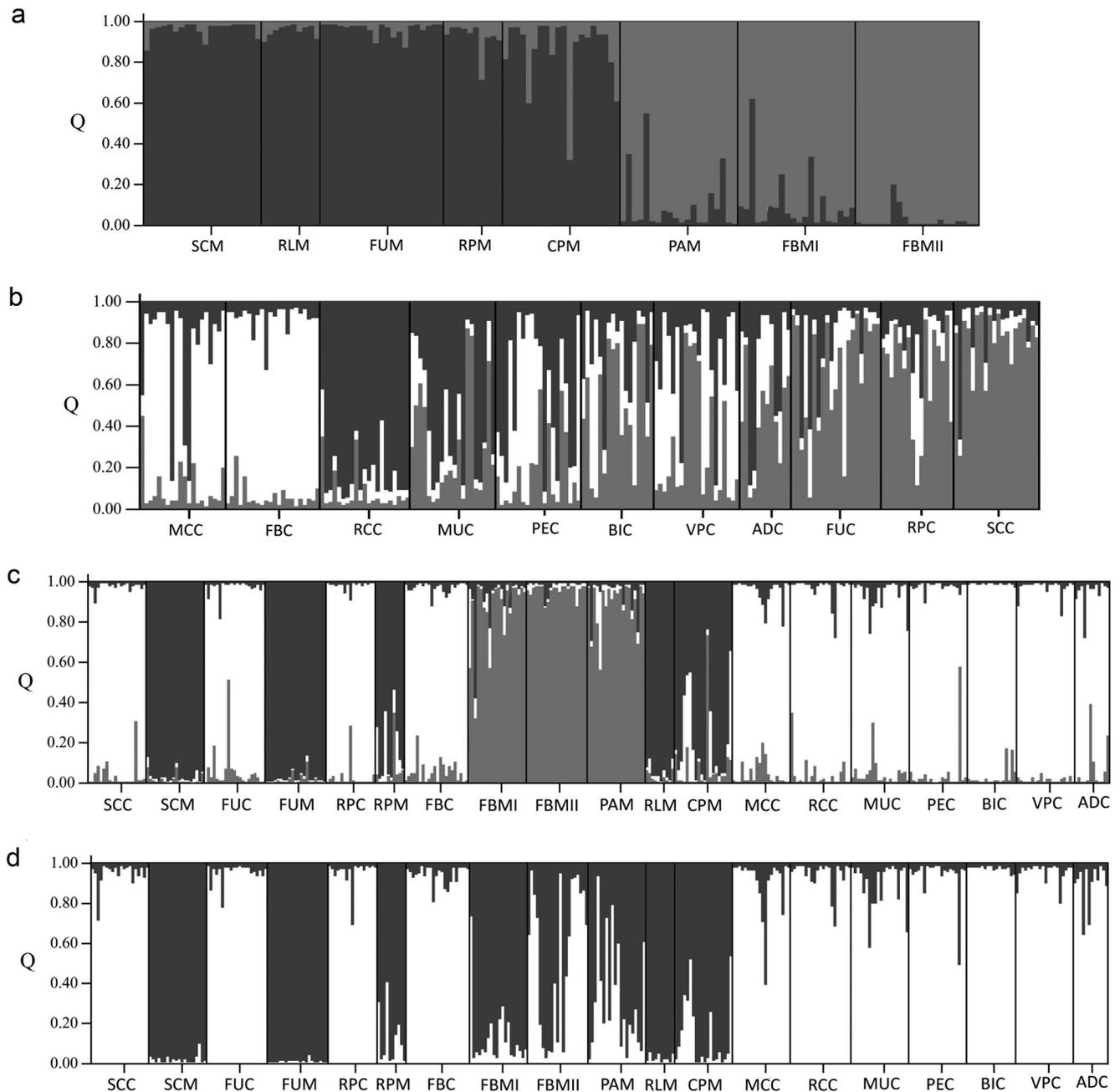
Fig. 3 Bayesian clustering analysis in STRUCTURE. The most probable number of clusters was inferred in *Hymenaea courbaril* (**a**), *H. stigonocarpa* (**b**), and in the combined dataset of both species (**c**, **d**). Each *vertical line* represents an individual, which is partitioned into the number of inferred clusters according to the probability of assignment *Q*. The number of inferred clusters for the combined dataset is presented for $K = 3$ (**c**) and $K = 2$ (**d**). Co-located populations of *H. courbaril* and *H. stigonocarpa* were plotted side by side in **c** and **d** to facilitate the comparison between them. For population codes, see Table 1

Interspecific relationships

AMOVA considering the two species together showed that 26.9% of the total genetic variation are due to differences among the two species ($F_{CT} = 0.269$; Online Resource 2 Table S2). Interspecific pairwise F_{ST} yielded higher values (ranging from 0.136 to 0.372) than intraspecific comparisons for both species (Online Resource 2 Table S3). However, the

genetic divergence among the two clusters of *H. courbaril* is high (mean pairwise $F_{ST} = 0.204$) and similar to genetic divergence among *H. courbaril* cluster 2 and *H. stigonocarpa* (mean pairwise $F_{ST} = 0.205$). Bayesian analysis identified three genetic clusters when all individuals were included (Fig. 3c), two of which coincide exactly with those observed when only populations of *H. courbaril* were analyzed (Fig. 3a), and the remaining cluster included all

H. stigonocarpa populations. Thus, STRUCTURE analysis seems to retrieve *H. courbaril* structure even when *H. stigonocarpa* data is included. In the results from an arbitrary value of $K = 2$, populations of *H. courbaril* and *H. stigonocarpa* are separated into distinct clusters (Fig. 3d). Neighbor-joining consensus tree, pairwise F_{ST}, and PCA also suggested strong genetic structure among populations of *H. courbaril*. Therefore, we acknowledge the existence of three groups while investigating the interspecific relationship between *H. courbaril* and *H. stigonocarpa* (Fig. 1).

Based on STRUCTURE results, individuals with Q 0.80 were considered *H. stigonocarpa*, individuals with Q 0.20 were considered *H. courbaril* (clusters 1 and 2), and individuals with $0.2 \leq Q \leq 0.8$ were considered putative hybrids. According to this criterion, 19 hybrids were observed in *H. stigonocarpa*, ranging from 0 (BIC and VPC) to 4 (MUC) per population, whereas the number of hybrids observed in *H. courbaril* was smaller, a total of 11 hybrids, ranging from 0 (FBMII, FUM, RLM, and SCM) to 5 (CPM) per population (Table 1 and Figs. 1 and 3c). The hybrids' admixture proportions resulting from Bayesian analysis are presented in Online Resource 2 Table S4. Although the total number of hybrids observed is larger in *H. stigonocarpa* than in *H. courbaril*, the percentage of hybrids in each species is similar (Table 1). Analysis of "pure" individuals simulated with HYBRIDLAB revealed that hybrid identification using STRUCTURE is not a spurious result because the probability of obtaining $0.2 \leq Q \leq 0.8$ was only 0.014. Therefore, most of the 30 individuals considered hybrids are likely to have mixed ancestry, as only 5 individuals from our dataset would have been misclassified as hybrids if there was no gene flow.

The first principal component (PC1) of PCA analysis using nSSR was able to separate most populations of *H. courbaril* from populations of *H. stigonocarpa* (Fig. 4a, b), despite the majority of alleles observed (75%) being shared between species. Out of 90 alleles observed in both species, only 22 are found in only 1 species (usually in low frequencies), with 77% of private alleles being present only in *H. courbaril* (Online Resource 2 Table S5). If STRUCTURE results are considered ($K = 3$), less than 14% of private alleles are present only in cluster 2 of *H. courbaril*, formed by populations FBMI, FBMII, and PAM. In fact, cluster 2 of *H. courbaril* is genetically closer to *H. stigonocarpa* than cluster 1 (Fig. 4a). This result is clearer in the PCA that included only neighboring populations: FUM, RPM, and SCM are clearly separated from their congeneric pairs (FUC, RPC, and SCC), while populations FBMI and FBMII are in an intermediate position, overlapping with some individuals from population FBC (Fig. 4b).

ABC analysis also supported hybridization between species. Six out of the seven datasets used in the ABC analysis favored the scenarios including either recent gene flow (from 10,000 years ago until present; RecMig model) or past gene flow (from 23,000 to 19,000 years ago; AncMig model)

between the species (Fig. 2 and Online Resource 2 Table S6). Specifically, for dataset ALLPOP, regardless of the number of retained simulations, model RecMig had the highest posterior probability (0.74) and the model NoMig was clearly rejected. The analyses performed with each cluster of *H. courbaril* (FBM_PAM × Stigo and OtherM × Stigo) yielded results similar to dataset ALLPOP and are presented only in Online Resource 2 Table S6. For the datasets FBM × FBC, FUM × FUC, and RPM × RPC, models RecMig and AncMig displayed high but similar posterior probabilities, meaning that it is not possible to reliably choose the best model between them. Only the pair SCM × SCC displayed similar posterior probabilities for models AncMig and NoMig. The analysis of true and false positives based on PODS simulated under each model showed that the NoMig and ConstMig models present higher proportion of true positives than AncMig model and especially the RecMig model (Online Resource 2 Table S7), indicating that the two last models (particularly RecMig) are difficult to identify. On the other hand, the RecMig model is the model with the lowest false positive rate, being wrongly chosen as the true model only in approximately 9% of the cases.

The parameter estimates of the best model(s) for each dataset are in Online Resource 1 Fig. S5 and Online Resource 2 Table S8. In general, the chosen SuSt contained enough information to estimate model parameters, as the R^2 obtained for most parameters were higher than 0.20 (Online Resource 2 Table S8). Migration rates were significant between the two species, with the number of migrants ($m \times$ current N) per generation ranging between 1.0 and 30.6 in different datasets. Although rates were generally higher for migration into *H. courbaril*, there was high overlapping between confidence intervals. Furthermore, the current effective size of *H. courbaril* (Nc) was consistently higher than the current effective size of *H. stigonocarpa* (Ns) for all datasets and models. Parameters with R^2 smaller than 0.05–0.10 were considered to be unreliably estimated since SuSt explain little of their variability (Neuenschwander et al. 2008). Ancestral effective size of *H. courbaril* (Nac) and *H. stigonocarpa* (Nas) after divergence were considered unreliable estimates because all datasets and models yielded $R^2 < 0.01$ and therefore are not presented.

Considering the occurrence of gene flow, we expected a considerable number of morphologically intermediate individuals. However, leaflet morphometric ANOVA revealed that both species are significantly different for eight out of nine measured variables, considering all individuals sampled. All neighboring populations analyzed were significantly different ($P < 0.05$) for at least seven leaf traits (Table 2). Additionally, PCA performed with these morphological traits for neighboring populations clearly distinguished individuals of *H. courbaril* from individuals of *H. stigonocarpa*, with PC1 and PC2 explaining over 80% of the observed variation. Only
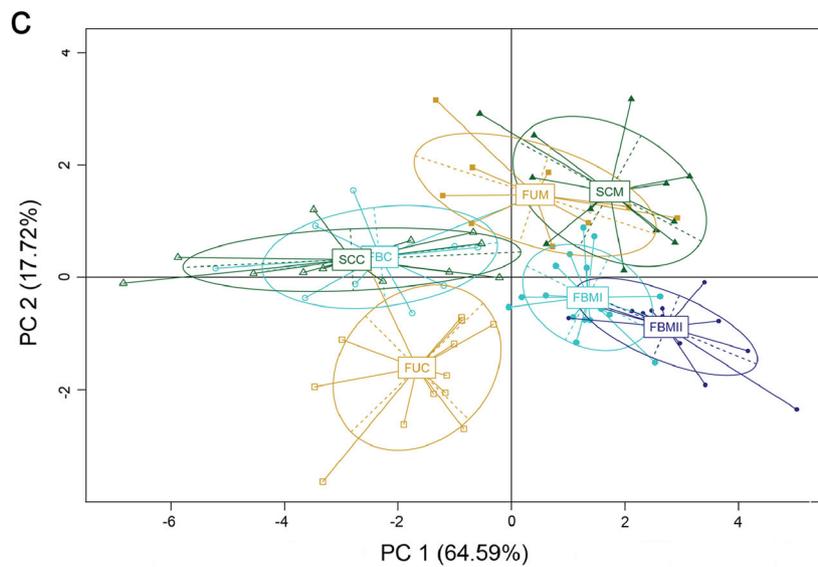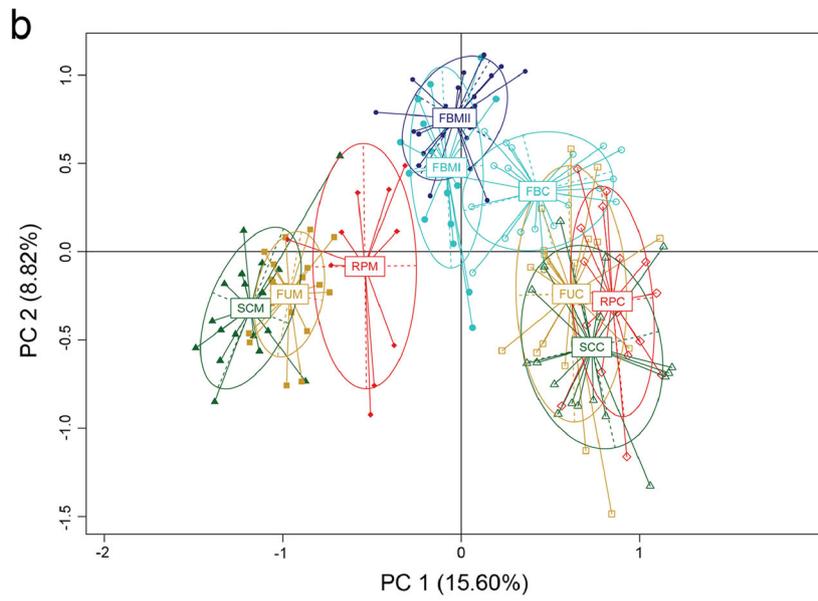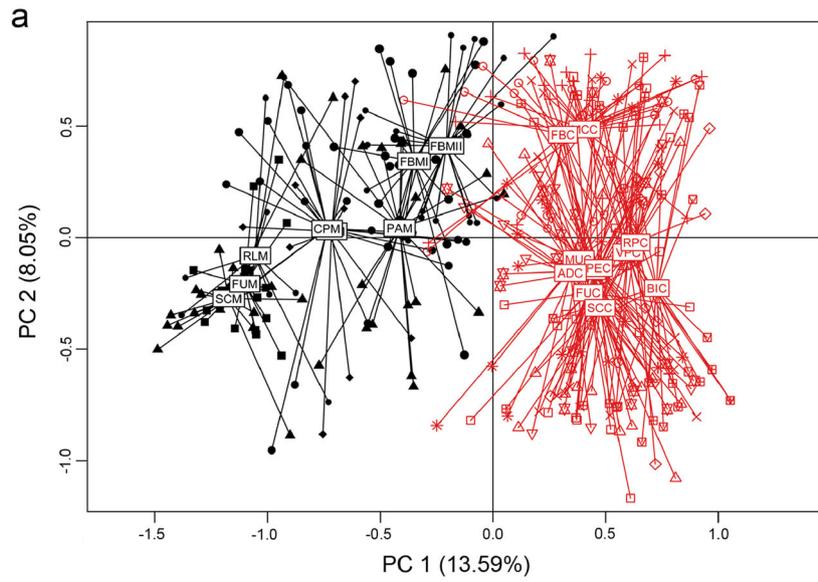
◀ **Fig. 4** Genetic structure (**a**, **b**) and leaflet morphological structure (**c**) were estimated by principal component analysis (PCA). Genetic structure was evaluated through PCA of microsatellite data of all populations (**a**) and of co-located populations (**b**). Percentages of the variation of the first two axes (PC1 and PC2) are displayed in each graph. Individuals of *Hymenaea courbaril* are represented by *filled symbols*, and individuals of *H. stigonocarpa* are represented by *open symbols*

two individuals, one from FBC and one from FUM, appeared within the population radius of the other species (Fig. 4c). Therefore, these results for leaf morphology support the taxonomic delimitation of the taxa studied.

Discussion

Evidences and timing of hybridization

Analysis with nSSR markers generally allowed the separation of the species *H. courbaril* and *H. stigonocarpa*. However, most alleles observed are shared between species, and genetically intermediate individuals were found in most populations analyzed with Bayesian analysis and PCA, suggesting the occurrence of hybridization and/or incomplete lineage sorting between those species. In order to separate the two processes, we performed ABC analysis with different evolutionary models including and excluding gene flow between *H. courbaril* and *H. stigonocarpa*. If incomplete lineage sorting explained most of the genetic similarity observed between species, we would expect higher support for the model without interspecies gene flow (NoMig). However, our ABC analysis showed that a model excluding gene flow was only supported for SCM × SCC (Fig. 2 and Online Resource 2 Table S6). Together, these results suggest that gene flow between *H. courbaril* and *H. stigonocarpa* is widespread, playing an important role in the evolution of these vicariant trees even if there is still incomplete lineage sorting between them.

In a previous study with both species, Ramos et al. (2009) revealed similar geographical distribution of shared cpDNA haplotypes and low genetic differentiation between these species. The number of hybrids that could be inferred from cpDNA haplotype sharing between *H. courbaril* and *H. stigonocarpa* (Ramos et al. 2009) was larger than the number of hybrids inferred herein with nSSR, although some populations such as CPM, RPM, ADC, and FUC displayed a large proportion of putative hybrids in both cpDNA and nSSR analyses. While cpDNA markers are associated with limited seed dispersal by agoutis (Asquith et al. 1999), nSSRs are biparentally inherited markers also associated with potentially long-distance pollen dispersal by bats (Gibbs et al. 1999; Dunphy et al. 2004). This result is in agreement with studies involving simulation data which predicted that markers

associated with limited dispersion will more promptly reveal genetic signals of hybridization (Petit and Excoffier 2009). The cpDNA haplotype sharing previously observed and the nSSR admixture here found between the two *Hymenaea* species could be related either to past hybridization in glacial refugia common to both species during LGM or recent hybridization after climate amelioration during Holocene, as has been suggested in other studies (Heuertz et al. 2006; Marino et al. 2013). Therefore, we also used the ABC approach to test the occurrence of gene flow in each of these periods. Our analysis, however, did not distinguish between ancient gene flow, during the LGM, and recent gene flow, during the Holocene.

Some evidence suggests recent hybridization (i.e., from the Holocene to the present), and others suggest that gene flow might have occurred during the LGM. In the ABC analysis, the RecMig model displayed the highest posterior probabilities for the dataset that included all sampled populations, and for both datasets that included one *H. courbaril* cluster at a time, suggesting recent gene flow. In fact, forest patches and riverine forests are common in the Cerrado biome, and there is a large ecotonal area between Cerrado and Atlantic forest allowing hybridization between the two species. Hybridization between these species is biologically possible because they share similar life-history traits (mainly bat-pollinated) (Gibbs et al. 1999; Dunphy et al. 2004) and display synchronization in flowering events possibly triggered by photoperiod (Souza and Funch 2016). On the other hand, ancient gene flow was equally probable to recent gene flow in three of the four datasets that included only pairs of neighboring populations (FBM × FBC, FUM × FUC, and RPM × RPC). Additionally, the CPM population, located in the semidecidous Atlantic forest, approximately 100 km apart from the ecotone with the Cerrado (Fig. 1), is one of the most admixed populations, with 25% of individuals considered hybrids. Thus, recent gene flow alone could not explain this pattern. According to the paleopalynological data (Behling and Lichte 1997; Behling and Negrelle 2001), during the LGM, the Cerrado expanded eastwards, where the semidecidous Atlantic forest is present currently. During this period, these species might have co-occurred in the area where the CPM population is currently found, allowing ancient hybridization events. Although gene flow between these species in other periods cannot be discarded, our results indicate that gene flow occurred during the LGM and/or the Holocene.

Evolutionary consequences of hybridization

Our genetic data supports the evidence that *H. courbaril* and *H. stigonocarpa* are closely related species, as have been shown by cpDNA studies (Fougère-Danezan et al. 2007; Ramos et al. 2009). The genetic differentiation observed between these species is as high as that observed between some

Table 2 Mean values of morphological characters for each species and population analyzed. Leaflet morphological characters of neighboring populations of *Hymenaea courbaril* (forest) and *H. stigonocarpa* (savanna) were measured, and the mean values were tested for statistical difference between pairs of populations and between species (one-way analysis of variance followed by Tukey's test)

Population[a]	N_I	LPV	MWL	WLS	WRS	NSV	DLB	L2SV	LA	LF
SCC	13	9.204**	5.5714**	2.2502**	3.3211**	20.0929	0.2924**	3.0962**	26.1133**	0.6039**
SCM	13	7.008	3.1225	1.2548	1.8678	21.3362	0.9747	2.1741	11.2863	0.4472
FUC	14	7.9915	4.6321**	1.9407**	2.6915*	14.4056**	1.081**	4.1477**	18.7303	0.5947**
FUM	9	8.3372	3.6907	1.4254	2.2653	20.9198	0.3475	2.7675	15.7933	0.4449
FBC	13	9.355**	5.3043**	2.2078**	3.0965**	21.3136**	0.6978**	3.2864	25.0531**	0.5711**
FBMI	17	6.965	3.3348	1.3569	1.9779	17.7483	0.379	2.9353	11.7030	0.4826
FBC	13	9.355**	5.3043**	2.2078**	3.0965**	21.3136**	0.6978**	3.2864	25.0531**	0.5711**
FBMII	15	5.4946	2.7423	1.2041	1.5382	17.1468	0.3359	2.3405	7.7410	0.4982
Mean *H. stigonocarpa*	40	8.8727**	5.1834**	2.1379**	3.0455**	18.4991	0.6801*	3.4933**	23.1847**	0.5900**
Mean *H. courbaril*	54	6.7956	3.1784	1.3013	1.8772	18.9735	0.5052	2.5588	11.1839	0.4721

N_I number of individuals measured in each population, *LPV* length of primary venation, *MWL* maximal width of leaflet lamina, *WLS* width of left side of leaflet lamina, *WRS* width of right side of leaflet lamina, *NSV* total number of secondary veins, *DLB* distance between lamina insertion on leaflet base, *L2SV* length of second secondary vein on right side of leaflet lamina, *LA* leaflet area, *LF* leaflet form (for measurement details, see Online Resource 1 Fig. S2)

[a] For population codes, see Table 1. The third letter of each population code represents the species. *M* = *H. courbaril* (forest species) and *C* = *H. stigonocarpa* (savanna species)

*$P < 0.05$; **$P < 0.01$

populations of *H. courbaril*. This cryptic genetic variation in *H. courbaril* reinforces the necessity of more taxonomic and phylogenetic investigation in the genus *Hymenaea*. Due to evidence of hybridization between *H. courbaril* and *H. stigonocarpa*, individuals with intermediate morphological characteristics were expected, but analysis of leaflet morphology clearly separated both species, with the exception of only two individuals (Fig. 4c). Leaf characters are known to be related to plant adaptation to different environmental conditions (Ackerly et al. 2000; Hovenden and Vander Schoor 2005). Thus, morphological differences observed herein could result from distinct selective pressures related to the large ecological differences between savanna and forest habitats (Eiten 1972; Furley 1992), in spite of gene flow. Our results revealed that the evolutionary scenario of continuous gene flow between *H. courbaril* and *H. stigonocarpa* since their divergence is less probable (Fig. 2 and Online Resource 2 Table S6), suggesting that divergence between these species seems to have started allopatrically. Later, after their divergence, the savanna-forest interface allowed hybridization between species, revealing incomplete reproductive barriers. Molecular evidence of interspecific gene flow could be maintained for long periods if gene exchange among conspecific populations was limited, as purging of heterospecific alleles would be a slow process (Currat et al. 2008; Petit and Excoffier 2009). In fact, our data showed significant inbreeding for both species, revealing limited intraspecific gene flow (Table 1).

Despite the observed hybrid individuals, many cpDNA haplotypes (Ramos et al. 2009) and nSSR alleles (Online Resource 2 Table S5) were found exclusively in only one species, and nuclear data generally allowed species separation, indicating that species integrity is revealed not only by morphological traits but also by genetic traits. Intrinsic factors like interspecific pollen-pistil incompatibility could be inhibiting extensive gene flow between *H. courbaril* and *H. stigonocarpa* (Hiscock and Dickinson 1993). Interspecific incompatibility seems likely if we consider that both *H. courbaril* and *H. stigonocarpa* were suggested as self-incompatible (Bawa 1974; Gibbs et al. 1999) and that crosses between self-incompatible species are often related to unilateral or reciprocal incompatibility (Hiscock and Dickinson 1993). In addition, extrinsic factors like strong selective pressures associated with each environment could be responsible for maintaining *H. courbaril* and *H. stigonocarpa* as separate entities even in the face of interspecific gene flow (Lexer et al. 2006; Palma-Silva et al. 2011). The extent of introgression among these species and the mechanisms involved in promoting and/or preventing it deserve to be further investigated as well as the possible effects of hybridization on adaptation to different environments and speciation.

Reports of interspecific hybridization are rare for woody species from the Cerrado biome, but floral morphology, pollination syndromes, and breeding systems are similar between vicariant tree species of forest and savanna areas inside the Cerrado biome in a variety of genera, as in *Vochysia* (Oliveira

and Gibbs 1994) and *Hymenaea* (Oliveira and Gibbs 2002). The recent evolution of the Cerrado, probably from forest lineages (Simon et al. 2009; Simon and Pennington 2012); the large number of genera of woody plants that present congeners in rain forest habitats (Hoffmann and Franco 2003; Pennington et al. 2006); the low levels of genetic divergence observed in the few studies that investigated species from genera that belong to the Cerrado and rain forest habitats (Collevatti et al. 2009; Ramos et al. 2009); and evidence of hybridization between one pair of vicariant species presented here support the close relationship between forest and Cerrado species.

Conclusions

Our study shows the occurrence of hybrids between two vicariant ETSA tree species, *H. courbaril*, from the Atlantic forest and Cerrado forested habitats, and *H. stigonocarpa*, from savanna habitats. However, hybrids did not exhibit intermediate leaflet morphology, more probably due to strong selective pressures associated with different environments. Morphological hybrids have been rarely reported in the literature for woody species from the two biomes, despite the large number of floristic studies available. Therefore, interspecific hybridization between tree species in the Cerrado biome may be a process underdiagnosed. In addition, our results highlight the importance of further morphological and genetic studies with other vicariant pairs of species in order to improve our understanding about the role of gene flow and adaptation in the evolution of the Cerrado biome.

References

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman JW, Brelsford A, Buerkle CA, Buggs R, Butlin RK, Dieckmann U, Eroukhmanoff F, Grill A, Cahan SH, Hermansen JS, Hewitt G, Hudson AG, Jiggins C, Jones J, Keller B, Marczewski T, Mallet J, Martinez-Rodriguez P, Most M, Mullen S, Nichols R, Nolte AW, Parisod C, Pfennig K, Rice AM, Ritchie MG, Seifert B, Smadja CM, Stelkens R, Szymura JM, Vainola R,

Wolf JBW, Zinner D (2013) Hybridization and speciation. J Evol Biol 26:229–246

Ackerly DD, Dudley SA, Sultan SE, Schmitt J, Coleman JS, Linder CR, Sandquist DR, Geber MA, Evans AS, Dawson TE, Lachowicz MJ (2000) The evolution of plant ecophysiological traits: recent advances and future directions. Bioscience 50:979–995

Archibald S, Bond WJ (2003) Growing tall vs growing wide: tree architecture and allometry of *Acacia karroo* in forest, savanna, and arid environments. Oikos 102:3–14

Asquith NM, Terborgh J, Arnold AE, Riveros CM (1999) The fruits the agouti ate: *Hymenaea courbaril* seed fate when its disperser is absent. J Trop Ecol 15:229–235

Barros FV, Goulart MF, Telles SBS, Lovato MB, Valladares F, Lemos-Filho JP (2012) Phenotypic plasticity to light of two congeneric trees from contrasting habitats: Brazilian Atlantic forest versus Cerrado (savanna). Plant Biol 14:208–215

Bawa KS (1974) Breeding systems of tree species of a lowland tropical community. Evolution 28:85–92

Beaumont MA (2008) Joint determination of topology, divergence time, and immigration in population trees. In:Matsumura S, Forster P, Renfrew C (eds) simulations, genetics, and human prehistory. McDonald Institute for Archaeological Research, Cambridge, pp 135–154

Beaumont MA, Zhang WY, Balding DJ (2002) Approximate Bayesian computation in population genetics. Genetics 162:2025–2035

Behling H, Lichte M (1997) Evidence of dry and cold climatic conditions at glacial times in tropical southeastern Brazil. Quat Res 48:348–358

Behling H, Negrelle RRB (2001) Tropical rain forest and climate dynamics of the Atlantic lowland, southern Brazil, during the late quaternary. Quat Res 56:383–389

Bittencourt NS Jr, Moraes CIG (2010) Self-fertility and polyembryony in South American yellow trumpet trees (*Handroanthus chrysotrichus* and *H. ochraceus*, Bignoniaceae): a histological study of postpollination events. Plant Syst Evol 288:59–76

Caddah MK, Campos T, Zucchi MI, Souza AP, Bittrich V, Amaral MCE (2013) Species boundaries inferred from microsatellite markers in the *Kielmeyera coriacea* complex (Calophyllaceae) and evidence of asymmetric hybridization. Plant Syst Evol 299:731–741

Camargo A, Morando M, Avila LJ, Sites JW (2012) Species delimitation with ABC and other coalescent-based methods: a test of accuracy with simulations and an empirical example with lizards of the *Liolaemus darwinii* complex (Squamata: Liolaemidae). Evolution 66:2834–2849

Cavallari MM, Gimenes MA, Billit C, Torres RB, Zucchi MI, Cavalheiro AJ, Bouvet JM (2010) Population genetic relationships between *Casearia sylvestris* (Salicaceae) varieties occurring sympatrically and allopatrically in different ecosystems in south-east Brazil. Ann Bot-London 106:627–636

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis models and estimation procedures. Am J Hum Genet 19:233–257

Ciampi AY, Azevedo VCR, Gaiotto FA, Ramos AC, Lovato MB (2008) Isolation and characterization of microsatellite loci for *Hymenaea courbaril* and transferability to *Hymenaea stigonocarpa*, two tropical timber species. Mol Ecol Resour 8:1074–1077

Collevatti RG, Leoi LCT, Leite SA, Gribel R (2009) Contrasting patterns of genetic structure in *Caryocar* (Caryocaraceae) congeners from flooded and upland Amazonian forests. Biol J Linn Soc 98:278–290

Costa WS, Souza AL, Souza PB (2011) Ecologia, manejo, silvicultura e tecnologia de espécies nativas da Mata Atlântica. Espécies Nativas da Mata Atlântica 2 http://www.sifloresta.ufv.br/bitstream/handle/123456789/11139/Jatoba_Ecologia,%20manejo,%20silvicultura%20e%20tecnologia%20de%20especies%20nativas%20da%20Mata%20atlantica%20.pdf?sequence=1&isAllowed=y. Accessed 10 November 2015

Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: massive introgression by local genes. Evolution 62: 1908–1920

Dieringer D, Schlotterer C (2003) Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Mol Ecol Notes 3:167–169

Duarte MC, Esteves GL, Salatino MLF, Walsh KC, Baum DA (2011) Phylogenetic analyses of *Eriotheca* and related genera (Bombacoideae, Malvaceae). Syst Bot 36:690–701

Dunphy BK, Hamrick JL, Schwagerl J (2004) A comparison of direct and indirect measures of gene flow in the bat-pollinated tree *Hymenaea courbaril* in the dry forest life zone of southwestern Puerto Rico. Int J Plant Sci 165:427–436

Eiten G (1972) The Cerrado vegetation of Brazil. Bot Rev 38:201–341

Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Mol Ecol 11:1591–1604

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Excoffier L, Dupanloup I, Huerta-Sanchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. PLoS Genet 9:e1003905

Excoffier L, Estoup A, Cornuet JM (2005) Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. Genet 169 (3):1727-1738

Excoffier L, Foll M (2011) Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. Bioinformatics 27:1332–1334

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567

Fiaschi P, Plunkett GM (2011) Monophyly and phylogenetic relationships of neotropical *Schefflera* (Araliaceae) based on plastid and nuclear markers. Syst Bot 36:806–817

Firetti-Leggieri F, Costa IR, Lohmann LG, Semir J, Forni-Martins ER (2011) Chromosome studies in Bignonieae (Bignoniaceae): the first record of polyploidy in *Anemopaegma*. Cytologia 76:185–191

Forni-Martins ER, Martins FR (2000) Chromosome studies on Brazilian cerrado plants. Genet Mol Biol 23:947–955

Forni-Martins ER, Pinto-Maglio CAF, Cruz ND (1995) Chromosome numbers in Brazilian cerrado plants. Rev Bras Genet 18:281–288

Fougère-Danezan M, Maumont S, Bruneau A (2007) Relationships among resin-producing Detarieae s.L. (Leguminosae) as inferred by molecular data. Syst Bot 32:748–761

Furley PA (1992) Edaphic changes at the forest savanna boundary with particular reference to the Neotropics. In: Furley P, Proctor J, Ratter J (eds) Nature and dynamics of forest-savanna boundaries. Chapman & Hall, London, pp 91–117

Garcia MG, Silva RS, Carniello MA, Veldman JW (2011) Molecular evidence of cryptic speciation, historical range expansion, and recent intraspecific hybridization in the neotropical seasonal forest tree *Cedrela fissilis* (Meliaceae). Mol Phylogenet Evol 61:639–649

Gibbs PE, Oliveira PE, Bianchi MB (1999) Postzygotic control of selfing in *Hymenaea stigonocarpa* (Leguminosae-Caesalpinioideae), a bat-pollinated tree of the Brazilian cerrados. Int J Plant Sci 160:72–78

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). http://www.unil.ch/izea/softwares/fstat.html. Acessed 30 Oct 2015

Goulart MF, Lovato MB, Barros FD, Valladares F, Lemos JP (2011) Which extent is plasticity to light involved in the ecotypic differentiation of a tree species from savanna and forest? Biotropica 43:695–703

Grant PR, Grant BR (2014) Evolutionary biology: speciation undone. Nature 507:178–179

Hamilton G, Currat M, Ray N, Heckel G, Beaumont M, Excoffier L (2005) Bayesian estimation of recent migration rates after a spatial expansion. Genetics 170:409–417

Heringer EP, Barroso GM, Rizzo JA, Rizzini CT (1976) A Flora do Cerrado. In: Ferri MG, coord. IV Simpósio sobre o Cerrado: bases para utilização agropecuária. São Paulo: USP. pp 211–231

Heuertz M, Carnevale S, Fineschi S, Sebastiani F, Hausman JF, Paule L, Vendramin GG (2006) Chloroplast DNA phylogeography of European ashes, *Fraxinus sp* (Oleaceae): roles of hybridization and life history traits. Mol Ecol 15:2131–2140

Hiscock SJ, Dickinson HG (1993) Unilateral incompatibility within the Brassicaceae: further evidence for the involvement of the self-incompatibility (*S*)-locus. Theor Appl Genet 86:744–753

Hoffmann WA, Franco AC (2003) Comparative growth analysis of tropical forest and savanna woody plants using phylogenetically independent contrasts. J Ecol 91:475–484

Hovenden MJ, Vander Schoor JK (2005) The response of leaf morphology to irradiance depends on altitude of origin in *Nothofagus cunninghamii*. New Phytol 169:291–297

Jombart T (2008) Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405

Lee YT, LAngenheim JH (1975) Systematics of the genus *Hymenaea* L. (Leguminosae, Caesalpinioideae, Detarieae). U Calif Publ Bot 69:1–109

Lacerda DR, Lemos Filho JP, Acedo MDP, Lovato MB (2002) Molecular differentiation of two vicariant neotropical tree species, *Plathymenia foliolosa* and *P. reticulata* (Mimosoideae), inferred using RAPD markers. Plant Syst Evol 235:67–77

Leonard JM, Bollmann SR, Hays JB (2003) Reduction of stability of *Arabidopsis* genomic and transgenic DNA-repeat sequences (microsatellites) by inactivation of AtMSH2 mismatch-repair function. Plant Physiol 133:328–338

Lexer C, Kremer A, Petit RJ (2006) Comment: shared alleles in sympatric oaks: recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. Mol Ecol 15:2007–2012

Marino IAM, Benazzo A, Agostini C, Mezzavilla M, Hoban SM, Patarnello T, Zane L, Bertorelle G (2013) Evidence for past and present hybridization in three Antarctic icefish species provides new perspectives on an evolutionary radiation. Mol Ecol 22:5148–5161

Mittermeier RA, Gil PR, Hoffman M, Pilgrim J, Brooks T, Mittermeier CG, Lamoreux J, Fonseca GAB (2004) Hotspots revisited: Earth's biologically richest and most threatened terrestrial ecoregions. CEMEX, Mexico City

Muir G, Schlotterer C (2005) Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus* spp.). Mol Ecol 14:549–561

Neuenschwander S, Largiader CR, Ray N, Currat M, Vonlanthen P, Excoffier L (2008) Colonization history of the Swiss Rhine basin by the bullhead (*Cottus gobio*): inference under a Bayesian spatially explicit framework. Mol Ecol 17:757–772

Nielsen EE, Bach LA, Kotlicki P (2006) HYBRIDLAB (version 1.0): a program for generating simulated hybrids from population samples. Mol Ecol Notes 6:971–973

Oliveira PE, Gibbs PE (1994) Pollination and breeding systems of six *Vochysia* species (Vochysiaceae) in central Brazil. J Trop Ecol 10:509–522

Oliveira PE, Gibbs PE (2002) Pollination and reproductive biology in Cerrado plant communities. In: Oliveira PS, Marquis RJ (eds) The Cerrados of Brazil: ecology and natural history of a neotropical savanna. Columbia University Press, New York, pp 329–347

Oliveira-Filho AT, Ratter JA (2002) Vegetation physiognomies and woody flora of the Cerrado biome. In: Oliveira PS, Marquis RJ (eds) The Cerrados of Brazil: ecology and natural history of a neotropical savanna. Columbia University Press, New York, pp 91–120

Page RDM (1996) Tree view: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12:357–358

Palma-Silva C, Wendt T, Pinheiro F, Barbara T, Fay MF, Cozzolino S, Lexer C (2011) Sympatric bromeliad species (*Pitcairnia* spp.) facilitate tests of mechanisms involved in species cohesion and reproductive isolation in neotropical inselbergs. Mol Ecol 20:3185–3201

Pennington RT, Richardson JE, Lavin M (2006) Insights into the historical construction of species-rich biomes from dated plant phylogenies, neutral ecological theory and phylogenetic community structure. New Phytol 172:605–616

Petit RJ, Excoffier L (2009) Gene flow and species delimitation. Trends Ecol Evol 24:386–393

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Rambaut A, Drummond AJ (2009) Tracer Version:1.5 http://beast.bio.ed.ac.uk/Tracer. Accessed 5 October 2015

Ramos ACS, Lemos-Filho JP, Lovato MB (2009) Phylogeographical structure of the neotropical forest tree *Hymenaea courbaril* (Leguminosae: Caesalpinioideae) and its relationship with the vicariant *Hymenaea stigonocarpa* from Cerrado. J Hered 100:206–216

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

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

Ribeiro RD, Cardoso D, de Lima HC (2015) A new species of *Hymenaea* (Leguminosae: Caesalpinioideae) with a revised identification key to the genus in the Brazilian Atlantic forest. Syst Bot 40:151–156

Rieseberg LH, Ellstrand NC, Arnold M (1993) What can molecular and morphological markers tell us about plant hybridization? Crit Rev Plant Sci 12:213–241

Rizzini CT (1997) Tratado de fitogeografia do Brasil: aspectos ecológicos, sociológicos e florísticos. Âmbito Cultural, Rio de Janeiro

Sano SM, Fonseca CEL (2003) Taxa de Sobrevivência e Frutificação de Espécies Nativas do Cerrado. Boletim de Pesquisa e Desenvolvimento. http://www.cpac.embrapa.br/baixar/91/t. Accessed 10 Nov 2015

Schluter D (2001) Ecology and the origin of species. Trends Ecol Evol 16:372–380

Scliar MO, Gouveia MH, Benazzo A, Ghirotto S, Fagundes NJR, Leal TP, Magalhaes WCS, Pereira L, Rodrigues MR, Soares-Souza GB, Cabrera L, Berg DE, Gilman RH, Bertorelle G, Tarazona-Santos E (2014) Bayesian inferences suggest that Amazon Yunga natives diverged from Andeans less than 5000 ybp: implications for South American prehistory. BMC Evol Biol 14:174. doi:10.1186/s12862-014-0174-3

Silberbauer-Gottsberger L, Gottsberger G, Ehrendorfer F (1992) Hybrid speciation and radiation in the neotropical woody genus *Tocoyena* (Rubiaceae). Plant Syst Evol 181:143–169

Simon MF, Grether R, de Queiroz LP, Skema C, Pennington RT, Hughes CE (2009) Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. Proc Natl Acad Sci U S A 106:20359–20364

Simon MF, Pennington T (2012) Evidence for adaptation to fire regimes in the tropical savannas of the Brazilian cerrado. Int J Plant Sci 173:711–723

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

Souza IM, Funch LS (2016) Synchronization of leafing and reproductive phenological events in *Hymenaea* L. species (Leguminosae, Caesalpinioideae): the role of photoperiod as the trigger. Braz J Bot. doi:10.1007/s40415-016-0314-7

Thuillet AC, Bru D, David J, Roumet P, Santomi S, Sourdille P, Bataillon T (2002) Direct estimation of mutation rate for 10 microsatellite loci in durum wheat, *Triticum turgidum* (L.) Thell. ssp. *durum* desf. Mol Biol Evol 19:122–125

Turchetto C, Fagundes NJR, Segatto ALA, Kuhlemeier C, Neffa VGS, Speranza PR, Bonatto SL, Freitas LB (2014) Diversification in the South American pampas: the genetic and morphological variation of the widespread *Petunia axillaris* complex (Solanaceae). Mol Ecol 23:374–389

Vähä JP, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. Mol Ecol 15:63–72

Vallejo-Marín M, Hiscock SJ (2016) Hybridization and hybrid speciation under global change. New Phytol 211:1170–1187

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

Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WF, Smith JSC, Doebley J (2002) Rate and pattern of mutation at microsatellite loci in maize. Mol Biol Evol 19:1251–1260

Wang J, Abbott RJ, Peng YL, Du FK, Liu JQ (2011) Species delimitation and biogeography of two fir species (*Abies*) in central China: cytoplasmic DNA variation. Heredity 107:362–370

Warwick MC, Lewis GP (2003) Revision of *Plathymenia* (Leguminosae—Mimosoideae). Edinb J Bot 60:111–119

Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABCToolBox: a versatile toolkit for approximate Bayesian computations. BMC Bioinformatics 11:116

Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci U S A 84:9054–9058