MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 5323–5336 doi: 10.1111/mec.12937

Strong spatial structure, Pliocene diversification and cryptic diversity in the Neotropical dry forest spider *Sicarius cariri*

IVAN L. F. MAGALHAES,*† UBIRAJARA OLIVEIRA,* FABRÍCIO R. SANTOS,‡ TEOFÂNIA H. D. A. VIDIGAL,* ANTONIO D. BRESCOVIT§ and ADALBERTO J. SANTOS*

Departamento de Zoologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, 31270-901, Belo Horizonte, Minas Gerais, Brazil, †División Aracnología, Museo Argentino de Ciencias Naturales 'Bernardino Rivadavia', Av. Angel Gallardo 470, C1405DJR, Buenos Aires, Argentina, ‡Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, 31270-901, Belo Horizonte, Minas Gerais, Brazil, §Laboratório Especial de Coleções Zoológicas, Instituto Butantan, Av. Vital Brazil 1500, 05503-900, São Paulo, Brazil

Abstract

The Brazilian Caatinga is part of the seasonally dry tropical forests, a vegetation type disjunctly distributed throughout the Neotropics. It has been suggested that during Pleistocene glacial periods, these dry forests had a continuous distribution, so that these climatic shifts may have acted as important driving forces of the Caatinga biota diversification. To address how these events affected the distribution of a dry forest species, we chose *Sicarius cariri*, a spider endemic to the Caatinga, as a model. We studied the phylogeography of one mitochondrial and one nuclear gene and reconstructed the paleodistribution of the species using modelling algorithms. We found two allopatric and deeply divergent clades within *S. cariri*, suggesting that this species as currently recognized might consist of more than one independently evolving lineage. *Sicarius cariri* populations are highly structured, with low haplotype sharing among localities, high fixation index and isolation by distance. Models of paleodistribution, Bayesian reconstructions and coalescent simulations suggest that this species experienced a reduction in its population size during glacial periods, rather than the expansion expected by previous hypotheses on the paleodistribution of dry forest taxa. In addition to that, major splits of intraspecific lineages of *S. cariri* took place in the Pliocene. Taken together, these results indicate *S. cariri* has a complex diversification history dating back to the Tertiary, suggesting the history of dry forest taxa may be significantly older than previously thought.

Keywords: biogeography, Caatinga, distribution modelling, glaciations, molecular clock, phylogeography, Pleistocene, seasonally dry tropical forests, Sicariidae

Received 29 June 2014; revision received 18 September 2014; accepted 19 September 2014

Introduction

The Neotropical region has an astonishing biological diversity and variety of vegetation types, including tropical and temperate rainforests, savannahs, deserts, montane grasslands, mangroves and dry tropical forests. Seasonally dry tropical forests (SDTFs), in particular, are a poorly studied system which has only recently been recognized as a cohesive phytogeographic unit (Prado & Gibbs 1993). This biome receives 300–1600 mm of rainfall annually and is characterized by having a marked dry season lasting for 6–9 months, thus being a xeric formation (Werneck 2011). It has a strongly disjunct distribution and its several nuclei are recognized as a single biological entity due to the sharing of endemic taxa, including plants (Prado & Gibbs 1993; Pennington *et al.* 2000; Prado 2000) and animals

Correspondence: Ivan L. F. Magalhaes, Fax +54 11 4982 6670 int. 172; E-mail: magalhaes@macn.gov.ar

(Ramos & Melo 2010; Werneck *et al.* 2012; Magalhães *et al.* 2013). The Caatinga, located in north-eastern Brazil, is the largest and most diverse nucleus of Neotropical SDTF (Prado 2000; Werneck 2011). Species composition is highly heterogeneous across the region (Santos *et al.* 2012), and some areas are known for harbouring endemic species, such as the fossil dunes of the middle São Francisco river (Rodrigues 1996). Yet, the Caatinga is very poorly known for most biological groups, especially invertebrates (Silva *et al.* 2004).

It has been claimed that the distribution range of the Caatinga has been affected by past climatic events. Prado & Gibbs (1993) postulated that SDTFs were widely and continuously distributed in the dry, cold periods of the Pleistocene. In this scenario, the present-day range of SDTF nuclei would be a disjunct relict of this once wider distribution. This idea became known as the Pleistocene Arc hypothesis (PAH). However, empirical evidence for this hypothesis is controversial (reviewed in Werneck 2011). Werneck *et al.* (2011) used paleodistribution modelling validated by paleopalynological data to estimate the distribution of the Caatinga during the Pleistocene and found no evidence of an expansion of its range. Some of the few phylogenetic/phylogeographic studies of SDTF taxa indicate a pre-Pleistocene diversification (Pennington *et al.* 2004; Werneck *et al.* 2012), while others reveal patterns like those expected by the PAH (Pennington *et al.* 2004; Caetano *et al.* 2008; Collevatti *et al.* 2012). The Caatinga has also been subject to climatic changes other than glaciations. In particular, the region experienced humid periods linked to distant climatic anomalies (Auler *et al.* 2004). In these events, the dry forest was probably replaced by rainforest in some points, as supported by evidence from geology, paleopalynology and fossil mammals (Cartelle & Hartwig 1996; de Oliveira *et al.* 1999; reviewed in Auler *et al.* 2004) and genetic signatures in rainforest taxa (Carnaval & Bates 2007). However, the extent to which these Pleistocene climatic changes might have shaped the genetic diversity of dry forest-associated taxa is not well understood.

The Caatinga is not the only SDTF nucleus in Brazil. Several smaller nuclei can be found in calcareous, nutrient-rich soil patches in the Cerrado, a tropical savannah which differs from dry forests both taxonomically and functionally (Werneck 2011). The largest of these savannah-associated nuclei is located at the Paranã River Valley, in Central Brazil (Werneck 2011; fig. 2). This region presents several endemic species (see Guadanucci 2011; Werneck 2011) and unique genetic lineages of widespread SDTF taxa (Caetano *et al.* 2008; Collevatti *et al.* 2012; Werneck *et al.* 2012). The PAH predicts that these smaller nuclei are remnant of a wider distribution of SDTFs in the past. Thus, testing whether the timing of diversification of their biota coincides with the Pleistocene would be useful for testing the PAH.

Phylogeographic and phylogenetic studies may be helpful in testing hypotheses on the evolution of environments (e.g. Byrne *et al.* 2008; Collevatti *et al.* 2012; Werneck *et al.* 2012). Recently, phylogeographic studies have been successfully combined with paleodistribution modelling, which is useful for generating explicit a priori hypotheses that can be tested with genetic data (e.g. Collevatti *et al.* 2012; Werneck *et al.* 2012). This approach is particularly interesting for testing events that took place in the Pleistocene, because (i) there are publicly available paleoclimatic models for this period and (ii) most of the genetic signatures caused by these events can still be observed because the Pleistocene is a relatively recent period. Additionally, phylogeographic studies are helpful for delimiting species and uncovering hidden diversity in taxonomically difficult taxa (e.g. Werneck *et al.* 2012; Satler *et al.* 2013). Such studies are particularly useful for studying groups whose morphology is simple and/or homogeneous, and clear-cut characters for defining species are hard to find (e.g. some haplogyne spiders; Macías-Hernandez *et al.* 2010).

There is a lack of phylogeographic studies for the Caatinga species (Turchetto-Zolet *et al.* 2013). This study focuses on the phylogeography of the spider *Sicarius cariri* as a model for inferring the recent history of a dry forest-associated taxon. *Sicarius* is a genus of medium to large-bodied soil spiders that do not spin webs, but rather bury themselves in fine, dry sediment to ambush their prey. They are strictly associated with xeric habitats and four species are endemic to the Caatinga. *Sicarius cariri* is the species with the largest distribution range in this region and also occurs in the SDTF enclaves in the Paranã River Valley (Magalhães *et al.* 2013). Using phylogeography coupled with paleodistribution modelling, we investigated questions related to demographic changes, the timing of diversification and geographic distribution of *S. cariri* haplotypes and their relation with Pleistocene climatic changes.

Material and methods

Sequence data collection

We collected specimens of *Sicarius* in the field and stored them in 96% ethanol at −20 °C. We also used museum samples collected from 2006 on and stored in 75% ethanol at room temperature. We extracted DNA by taking one leg from each specimen, removing its dirt with an ultrasonic cleaner, excising muscle tissue and using a Wizard® Genomic DNA purification kit (Promega) following the manufacturer's instructions. We amplified one or two loci for 224 individuals (169 with

both loci, not considering out-groups; Tables S1 and S2, Supporting information): a 715-bp fragment of the mitochondrial gene coding for cytochrome-C oxidase, subunit 1 (COI; 211 individuals) using primers C1-N-2568 (5′-GCT-ACA-ACA-TAA-TAA-GTA-TCA-TG-3′) and C1-J-1751 'SPID' (5′-GAG-CTC-CTG-ATA-TAG-CTT-TTC-C-3′) (Hedin 1997), and a 327-bp fragment of the nuclear gene coding for histone 3, subunit a (H3; 183 individuals, Tables S1 and S2, Supporting information) using primers H3aF (5′-ATG-GCT-CGT-ACC-AAG-CAG-ACV-GC-3′) and H3aR (5′-ATA-TCC-TTR-GGC-ATR-ATR-GTG-AC-3′) (Colgan *et al.* 1998). Sequences of six *S. tropicus*, six *S. ornatus* and five *S. diadorim* specimens were obtained for use as out-groups. We amplified the fragments through the following PCR conditions: 5 μL 5× buffer (Phoneutria IVB® or Promega Gotaq® Flexi buffer), 2 μL 25 mM $MgCl_2$, 2.5 μL dNTP mix at 2 μM (Promega), 2.5 μL of each primer at 5 μM, 1.25 μL of dimethyl sulfoxide (for H3 only), 0.25 unit of Taq DNA polymerase (Phoneutria or Promega Gotaq®) and 1 μL of extracted DNA, for a final volume of 25 μL. PCR cycles consisted of 10′ at 94 °C, 35–40 × (30″ at 95 °C, 45″ at 52 °C and 45″ at 72 °C) and 7–10′ at 72 °C. Alternative PCR programmes (including five initial cycles at 46 °C, COI, or 48 °C, H3) were used in the few cases (e.g. museum samples) in which PCRs were not successful. We checked PCR success in a 1% agarose gel stained with GelRed® (Biotium) and purified it using a 20% 8000 polyethyleneglycol + NaCl solution followed by washes in 80% ethanol. Sequencing reactions consisted of 4 μL of PCR product, 1 μL of either forward or reverse primers, 1.6 μL 5× sequencing buffer and 0.8 μL BigDye® Terminator Cycle Sequencing Kit (Applied Biosystems) for a 10 μL final volume. Automated sequencing was carried out on an ABI 3130xl Genetic Analyzer (Applied Biosystems) following the manufacturer's instructions. We checked chromatograms by eye in SeqScape® 2.6.0 (Applied Biosystems). Alignment was trivial due to lack of insertions/deletions and was performed in Muscle (Edgar 2004) as implemented in MEGA v6 (Tamura *et al.* 2011). We phased H3 sequences in DNASP 5.1 (Librado & Rozas 2009) using default parameters because several exploratory analysis yielded identical results.

Gene trees and divergence times

The COI and H3 gene trees were estimated along with calculations of divergence times in a Bayesian framework using BEAST 1.7.5 (Drummond *et al.* 2012). Models of DNA evolution were determined with MRMODELTEST2 (Nylander 2004). Analyses in BEAST were performed using GTR+I with equal base frequencies for H3, and GTR+I+G with estimated base frequencies for COI. Setting priors for Bayesian inference is not always straightforward; thus, we ran several preliminary analyses combining different clock and population models (Table S3, Supporting information). For each gene/model, we ran two (H3) or three (COI) independent Markov chains for 10–20 million generations, sampling every 1000–2000 generations. We checked stationarity and convergence of the chains using TRACER 1.6. We used LOGCOMBINER 1.7.5 to join trees from independent chains and to remove the initial 20% of the sampled states as burn-in, TREEANNOTATOR 1.7.5 to obtain the maximum credibility tree and FIGTREE 1.4 to visualize the final gene trees. To determine the best-fit clock and population models, we compared runs using Markov chain Monte Carlo Akaike information criterion (AICM) using TRACER 1.6 (part of the BEAST package; see Baele *et al.* 2012) (Table S3, Supporting information), assessing SD using 100 bootstrap replicates. Although the main results are strongly robust to different combinations of clock/population models (data no shown), only analyses using the best-fitting priors were further considered (see Table S3, Supporting information). There are few H3 haplotypes (Table S4, Supporting information), and thus for this marker we estimated the gene tree for unique haplotypes to save computational time; the gene tree for COI was estimated using all obtained sequences. We calibrated molecular clocks using mutation rates obtained by Bidegaray-Batista & Arnedo (2011) for a genus of dysderid spiders (ucld.mean = 0.0013 for H3, 0.0199 for COI). We took this approach because the only known sicariid fossil is a very distantly related Caribbean *Loxosceles* from the Miocene, and using biogeographic calibrations would result in circular reasoning. We are aware of the limitations of calibrating the molecular clock with rates estimated for different taxa, and our estimates of divergence times should be taken cautiously. On the other hand, the dysderid rates were provided as normally distributed priors (not fixed values) for our analyses and were allowed to vary along the Markov chains. The rates estimated for our data set are higher than the priors, and enforcing the dysderid rates to our data set would probably yield even older divergence times. All Markov chain analyses using BEAST were performed remotely using the CIPRES platform (Miller *et al.* 2010). Additionally, we built parsimony-based median-joining networks (Bandelt *et al.* 1999) for each gene using NETWORK 4.6 (http://www.fluxus-engineering.com).

Species limits

As phylogenetic analysis of both genes indicated the presence of two, highly divergent clades within *S. cariri* (see Results), we tested whether genetic data support

the hypothesis of cryptic diversity in this species. We carried this analysis because, if those two clades act as independently evolving lineages, it would be more interesting to conduct distribution modelling and phylogeographic analyses for each of them separately. We compared two taxonomic hypotheses: (i) *S. cariri* as currently defined is a single species or (ii) the two main clades recovered for both COI and H3 represent distinct species (see 'Gene trees and divergence times' below). To test which of these scenarios is more likely, we used *BEAST 1.7.5 to generate species trees under each of these assumptions. Species trees were estimated under a Yule prior, using data from both genes simultaneously and time-calibrated using estimated mutation rates for each gene. For each hypothesis, four independent Markov chains were run for 50 million generations. We used TRACER 1.6 to check stationarity and convergence of the chains and combined BEAST log files of each hypothesis using LOGCOMBINER 1.7.5, removing 20% of the initial samples of each chain as burn-in. We calculated differences in AICM between the combined log files of each competing hypotheses (SD based on 100 bootstrap replicates). Additionally, we calculated Kimura 2-parameter (K2-P) genetic distances within and among species under the two hypotheses above. Genetic distances were calculated in MEGA v6, assessing standard deviations through 100 bootstrap replicates.

Demographic changes

We estimated ancestral population sizes of *S. cariri* in a Bayesian framework using extended Bayesian skyline plots (EBSP) calculated using both genes simultaneously in BEAST 1.7.5. For each *S. cariri* clade, we ran one Markov chain for 200 million generations, under an exponential relaxed molecular clock calibrated as described above and under an EBSP coalescent prior. We checked convergence and stationarity of the chains using TRACER 1.6. We ran demographic reconstructions and all subsequent analyses for each of the two *S. cariri* main clades separately, because we have evidence that they act as evolutionary independent lineages (see 'Species limits' below).

Population statistics and geographic isolation

We calculated Tajima's D and Fu and Li's D* and F* for each *S. cariri* clade/gene, and number of haplotypes, haplotype diversity, number of polymorphic sites and nucleotide diversity for each sampled population using DNASP 5.1. For each clade, we assessed population differentiation by calculating the F_{ST} index in DNASP 5.1. For clade 1, we performed analyses of molecular variance (AMOVA; Excoffier *et al.* 1992) in ARLEQUIN 3.5.1.2 (Excoffier *et al.* 2005) separating populations by sampling points and testing three different deme schemes (see Table S5, Supporting information): (i) populations grouped by Caatinga ecoregions (see Werneck 2011) and SDTF enclaves from the Cerrado (nine demes), (ii) populations grouped by each bank of the São Francisco River and SDTF enclaves from the Cerrado (three demes) and (iii) populations grouped by COI haplogroups (13 demes); these were defined based on the natural breaks among well-supported groups of mtDNA sequences (Fig. 1, letters A–M). Significance of F_{ST}, F_{SC} and F_{CT} was tested based on 10 000 permutations. To test whether there is isolation by distance in *S. cariri*, we performed a Mantel test to check the correlation between a geographic distance matrix and a genetic K2-P distance matrix between sampling points for each clade separately. We used BARRIER 2.2 (Manni *et al.* 2004) to visualize regions where genetic dissimilarity among populations is higher than expected by distance alone.

Paleodistribution modelling

Analysis of paleodistribution was conducted for *S. cariri* clade 1 only, as we do not have enough occurrence points of clade 2 to ensure a reliable distribution modelling. To reconstruct climatically stable areas of *S. cariri* clade 1, we tested eight algorithms that use different logics for modelling distributions: BioClim, SVM, MaxEnt, Mahalanobis distance, Euclidian distance, GARP, ENFA and Domain. All algorithms were run on OPENMODELLER 1.3 (Muñoz *et al.* 2011), except for MAXENT 3.3.1, which is a stand-alone software (Phillips *et al.* 2006). Variables were selected aiming at reducing the collinearity between them in the study area. Thus, all variables that presented a correlation lower than $R = 0.7$ among them were considered. We used the following variables from the Worldclim database (Hijmans *et al.* 2005; http://www.worldclim.org): annual mean temperature, mean diurnal range, isothermality, annual mean precipitation, precipitation of driest month and altitude. We applied the minimum presence threshold to all algorithms to generate presence/absence maps. To evaluate algorithm performance, we used ten pseudoreplicate data sets consisting of 54 presence points of *S. cariri*, randomly divided into 38 training and 19 testing points, and 89 absence points. We considered absences as (i) points where other species of *Sicarius* occur, but not *S. cariri* (ii) localities visited during our fieldwork or localities where extensive inventories have been carried out and *S. cariri* has not been detected. We measured algorithm performance using adjusted accuracy (see legend on Fig. S9, Supporting information), sensitivity, specificity and AUC. We compared the performance of the algorithms using analyses of variance,

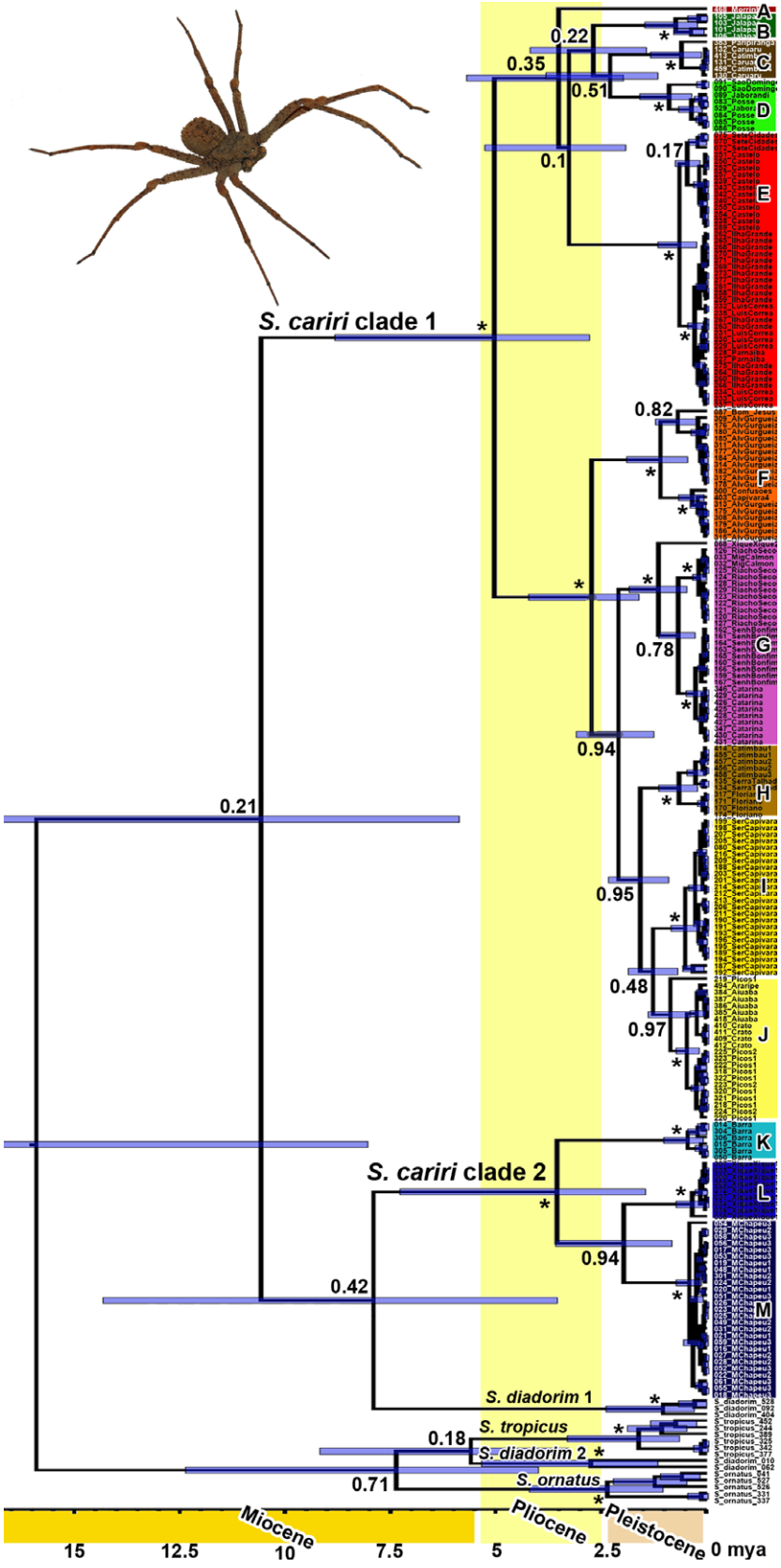

Fig. 1 Gene tree for COI haplotypes of *Sicarius cariri* and related species, estimated under Bayesian inference. Divergence times were estimated based on a molecular clock calibrated using mutation rates. Main well-supported haplogroups indicated by capital letters; haplogroup colours refer to the map in Fig. 3. Blue bars represent 95% highest posterior density intervals of node ages. Node posterior probabilities indicated on branches (stars indicate PP >0.99).

with each of the ten pseudoreplicate data sets as samples. GARP, ENFA and Bioclim showed the best overall performance in the validation (see Results), while other algorithms performed poorly. Thus, we used these algorithms to generate present-day distribution models using all presence points. Then, we projected these models onto paleoclimatic layers obtained from WorldClim and representing the climate from 21 000 years ago (last glacial maximum; LGM). We used two different models for LGM climate (CCSM and MIROC) due to uncertainties on the knowledge of past climates. For each algorithm and LGM paleoclimate model, we summed present and past distribution models to generate consensus maps representing stable areas for the occurrence of *S. cariri* clade 1.

Coalescent simulations

To test which demographic scenario is more likely to have occurred in *S. cariri* clade 1, we used a coalescent-based simulation approach. Based on previous works and our own results, we hypothesized five scenarios for the demographic history of *S. cariri* (see Results below). We carried out 10 000 simulations for each gene and scenario using BAYESSC (http://www.stanford.edu/group/hadlylab/ssc/; see Excoffier *et al.* 2000). Mutation and transversion rates and gamma parameters were taken from BEAST results. We estimated N_e based on θ estimates obtained with MIGRATE-N 3.3.1 (Beerli & Palczewski 2010) (four increasingly heated chains run for 5×10^7 generations, discarding 10^6 as burn-in). We evaluated simulated scenarios using three summary statistics: number of segregating sites, haplotype diversity and nucleotide diversity. The likelihood density (probability of observing the values from the empirical data set) for each of these was estimated using the SSC.Like function for R 3.1 (R Core Team 2014) using a 0.05 acceptance kernel. Some of the scenarios have more free parameters than others; thus, we also compared them by calculating their AIC values using the aic.ssc function for R (both functions depend on the packages lattice, akima and locfit and on a script available from the BayesSC website).

Results

Gene trees and divergence times

Sicarius cariri is divided into two very well-supported main clades in both COI and H3 gene trees (Figs 1 and 2), which are also present in haplotype networks (Figs S1 and S2, Supporting information). However, none of the gene trees provide evidence that these two clades form a single monophyletic group (Figs 1 and 2). The first clade comprises most of *S. cariri* populations (Figs 1 and 2), including the species' type locality (Serra da Capivara: Fig. 3, population 14) and populations from SDTF enclaves in the Cerrado (Fig. 3, populations 23–26). The second clade comprises populations around the fossil dunes of the São Francisco River (Figs 1 to 3, populations 28–30). For both clades, COI haplogroups are strongly structured geographically (F_{ST} = ~0.9): each sampled location is always composed of individuals belonging to a single haplogroup, except for populations 11 and 14, which have a single outlier individual (Fig. 3). Also, haplogroups and haplotypes which are shared among locations generally occur in adjacent populations. Finally, all sampled populations have unique haplotypes, except for populations 12 and 22, which are subsets of populations 11 and 17, respectively. On the other hand, H3 haplogroups are much less diversified and geographically structured, with more sharing of haplotypes among locations (F_{ST} = ~0.6; Fig. 3). However, they do not appear to be randomly distributed in space (IBD = 0.82), as some haplotypes are more common in certain geographic regions.

Both gene trees imply that *S. cariri* has an old history of diversification, although confidence intervals are somewhat large. Estimates suggest clade 1 started diversifying around 5 million years ago (Ma), while clade 2 started diversifying around 3.5 Ma (Figs 1 and 2). Most of current *S. cariri* haplogroups appear to have been established during early to mid-Pleistocene (Figs 1 and 2). Both gene trees support the view that *S. cariri* populations from SDTF enclaves have haplotypes that diverged from their sister lineages around 2.5 Ma.

Species limits

The hypothesis that each of the main clades of *S. cariri* can be recognized as a different species (AICM = 16 088 ± 0.6) is favoured in relation to the hypothesis that *S. cariri* as presently recognized is a single species (AICM = 16 104.5 ± 4) due to its lower AICM value. Also, K2P distances between clade 1 and clade 2 are comparable to those among other species pairs (around 15.5%; Fig. S3, Supporting information). Within-species distances vary from 4 to 8% in *S. tropicus*, *S. ornatus* and *S. cariri* clades 1 and 2 and are higher in *S. diadorim* and *S. cariri*, the two species for which we had no evidence of monophyly (as currently recognized—11 and 10%, respectively) (Fig. S3, Supporting information). Both results are congruent with the facts that these two clades were strongly recovered in both gene trees but do not form a single monophyletic lineage in any of them.

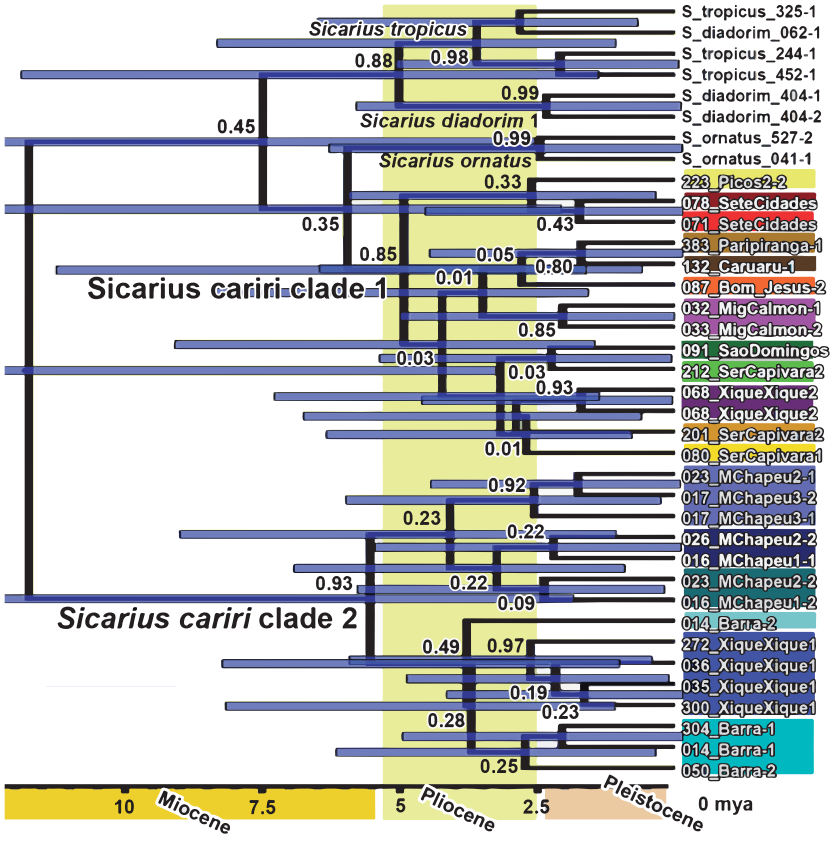

Fig. 2 Gene tree for H3 subunit A haplotypes of *Sicarius cariri* and related species, estimated under Bayesian inference. Divergence times were estimated based on a molecular clock calibrated using mutation rates. Haplogroup colours refer to the map in Fig. 3. Blue bars represent 95% highest posterior density intervals of node ages. Node posterior probabilities indicated on branches.

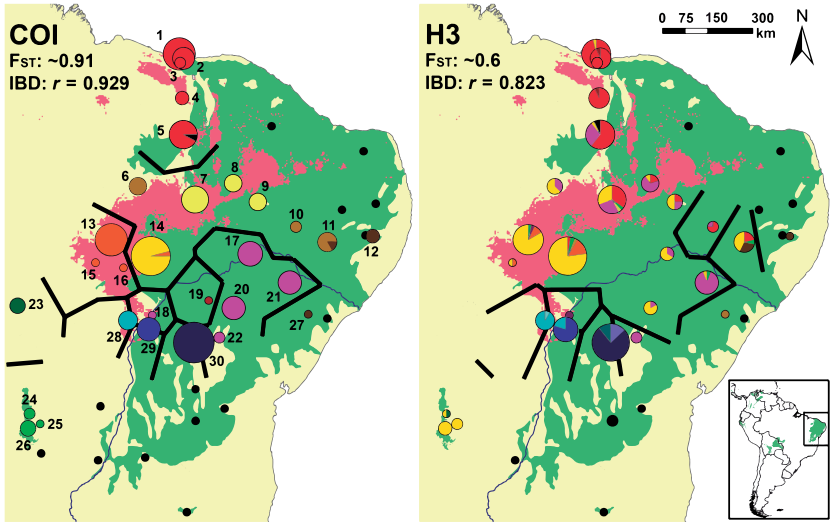

Fig. 3 Geographic distribution of haplogroups of COI and H3 subunit A of *Sicarius cariri* (coloured locations) and related species (black locations). Circle size is proportional to number of sampled individuals per locality. Haplogroup colours are the same as in Figs 1 and 2. The blue line represents the São Francisco River, black lines represent areas of genetic discontinuity identified by BARRIER 2.2, the green area represents current Caatinga limits, and the red area represents the consensus of historically stable areas for occurrence of *S. cariri*, as estimated through paleodistribution modelling (see text for details). Population numbers refer to Table S3 (Supporting information).

Demographic changes

Extended Bayesian skyline plots reconstructions indicate that the two *S. cariri* main clades had contrasting demographic histories (Fig. 4). Clade 1 has experienced a steady increase in population size from 5 to 0.86 Ma, when population size started to decrease at a fast pace. After a threefold reduction, population size started to grow again around 20 000–10 000 years ago. On the other hand, clade 2 population size has remained constant throughout most of its history, suddenly increasing almost sevenfold around 150 000 years ago.

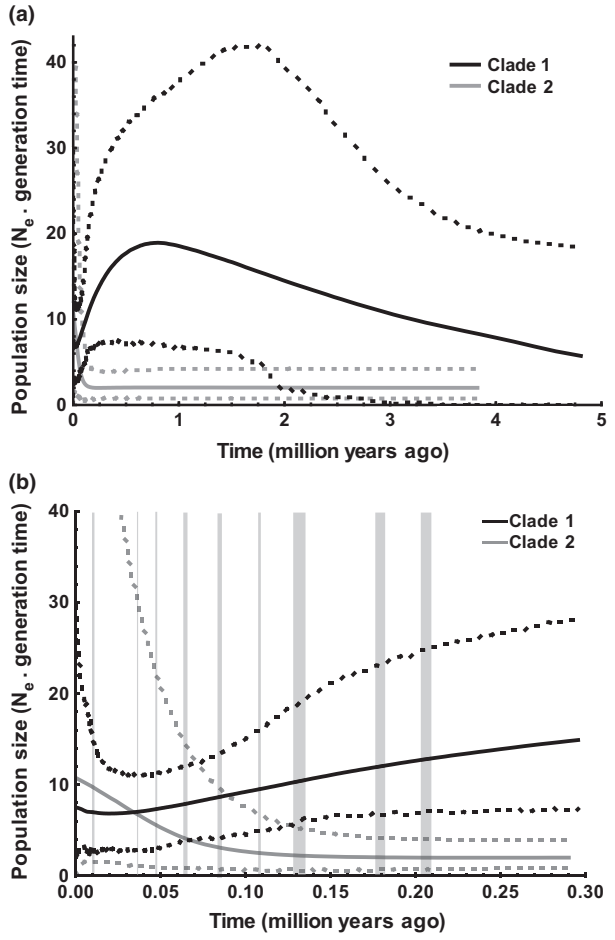

Fig. 4 Past demographic history of the two main clades of *Sicarius cariri* as revealed by extended Bayesian skyline plots. (a) Reconstruction of the last 5 Myr. (b) Detail of the last 300 000 years. Shaded areas indicate periods of increased rainfall in the Caatinga region (adapted from Auler *et al.* 2004). Solid line = median estimated population size, dashed lines = upper and lower bounds of 95% highest posterior density.

Population statistics and geographic isolation

Neutrality tests indicate that both COI and H3 variation is neutral (nonsignificant Tajima's D and Fu and Li's F* and D* for both clades/genes). COI is substantially more diverse than H3, having a higher number of polymorphic sites and haplotypes and higher values of nucleotide diversity for all populations (Table S4, Supporting information). This, in turn, leads to a greater geographic structuration of COI diversity: F_{ST} values are higher for this gene (clade 1 = 0.912, clade 2 = 0.925) than for H3 (clade 1 = 0.577, clade 2 = 0.603). Although both genes present evidence of isolation by distance, the signal is stronger for COI ($r = 0.929$, $P < 0.001$) than for H3 ($r = 0.823$, $P < 0.001$) (Fig. S4, Supporting information). Most of the regions of phylogeographic breaks identified by BARRIER 2.2 do not coincide between the two genes, except for the ones separating populations from *S. cariri* clade 1 and clade 2 (Fig. 3). While H3 variation is more evenly distributed in space, COI has several phylogeographic breaks separating populations from different geographic regions (Fig. 3). AMOVA results indicate that COI variation is essentially distributed among populations (only ~10% of the variation is within populations), while H3 shows the converse pattern (~45% of genetic variation is within populations) (Table S5, Supporting information). Grouping populations by ecoregions or COI haplogroups led to higher F_{CT} indexes than grouping by banks of the São Francisco River. Regardless of grouping scheme, results indicate strong structuring for both COI (F_{SC} >0.6 and F_{ST} >0.9) and H3 (F_{SC} >0.25 and F_{ST} >0.5).

Paleodistribution modelling

The ENFA algorithm showed the best results regarding its adjusted accuracy, specificity and AUC values, although sensitivity values are low (Figs S7 to S10, Supporting information). MaxEnt showed similar results to ENFA regarding its specificity, but had lower sensitivity, accuracy and AUC. BioClim and GARP predicted larger areas, leading to lower omission but higher overprediction and lower AUC and accuracy. The Euclidean distance, Mahalanobis distance, Domain and SVM algorithms were highly sensitive, but this led to unreasonable overprediction rates and a corresponding low accuracy, making these models less useful (Figs S7 to S10, Supporting information). Considering these results and visual inspections of present-day models (Figs S5 and S6, Supporting information), we chose only ENFA, BioClim and GARP for projection onto paleoclimatic layers and further consideration.

Projection of the models onto paleoclimatic scenarios revealed three possible histories for *S. cariri* clade 1. Consensus calculated for ENFA (MIROC) and BioClim (MIROC) indicate multiple refuges for this species throughout the Pleistocene (Fig. 5a, c). The consensus calculated for GARP (MIROC) (Fig. 5e) indicates that the species range remained stable throughout the Pleistocene. Consensus calculated for ENFA (CCSM), BioClim (CCSM) and GARP (CCSM) (respectively, Figs. 5b, d, f) indicate that the species was absent from the eastern part of its current range during the LGM.

Coalescent simulations

We compared five different histories: (i) a stable population (a simple working hypothesis assuming no changes in population size), (ii) a population suffering a bottleneck during glaciations (as suggested by

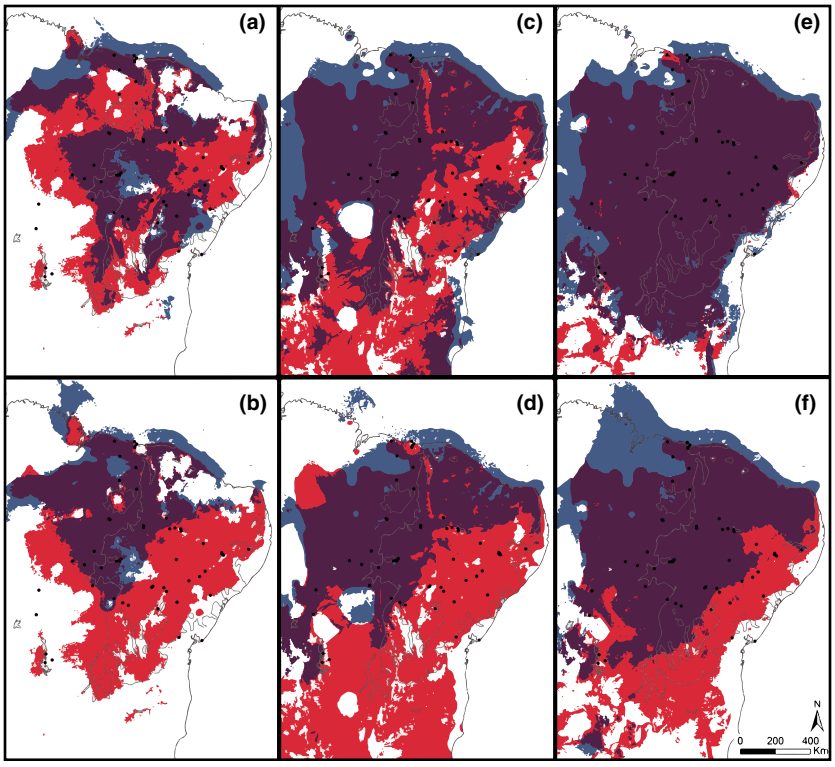

Fig. 5 Models for present (red) and past (last glacial maximum, 21 000 years ago; blue) distributions of *Sicarius cariri*. Stable areas (predicted as suitable areas in both present and past models) are represented in purple. (a) ENFA, MIROC. (b) ENFA, CCSM. (c) BioClim, MIROC. (d) BioClim, CCSM. (e) GARP, MIROC. (f) GARP, CCSM. Black dots are *S. cariri* presence points used to generate the models.

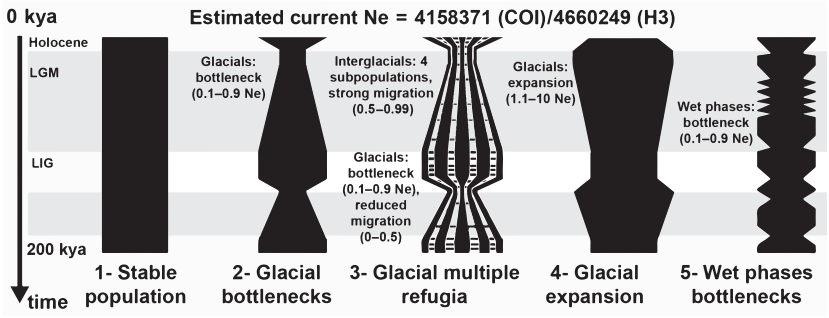

Fig. 6 Schematic representation of the five hypothetical demographic scenarios for *Sicarius cariri* which have been simulated. Simulations start at present and run backwards in time. Grey highlighting indicates glacial periods in the last 200 thousand years. Width of the bars indicates changes in population size (not to scale, for illustration purposes only). Horizontal bars in model 3 indicate migration among subpopulations.

paleodistribution models; Fig. 5b, d, f), (iii) a structured population suffering a bottleneck and reduced migration during glaciations (multiple refugia, as suggested by paleodistribution models; see Fig. 5a, c), (iv) a population expanding during glaciations (Prado & Gibbs 1993 hypothesis) and (v) a population suffering bottlenecks in Caatinga wet phases (see Auler *et al.* 2004) (see Fig. 6; Table 1). In general, the simulations generated diversity values well above those observed for *S. cariri* (Table 1). For COI, estimated likelihood density and AIC values suggest the most likely scenario to be glacial bottlenecks (scenario 2), while for H3 the best-fitting scenario is glacial multiple refugia (scenario 3; Fig. 6, Table 1). The least likely scenario for both genes and by both criteria is population expansion at the LGM.

Discussion

We demonstrate that *S. cariri* is a species with very strong geographic structuring regarding its genetic diversity. The genetic relationships among *Sicarius* from the Caatinga suggest that *S. cariri* as currently recognized is composed by two independent lineages that might merit species status. Even considering each clade separately, haplotypes are deeply divergent and major splits pre-date the Pleistocene, including those leading to lineages that colonized SDTF enclaves in Central Brazil. Both paleodistribution modelling and genetic evidence indicate that clade 1, encompassing most of the species' known range (see Figs 1 and 3), has suffered a reduction in its population size during the Pleistocene. This suggests that *S. cariri* has responded to glaciations

Table 1 Number of segregating sites, haplotypic diversity and nucleotide diversity observed in our empirical data set and those expected by our coalescent simulations under five different demographic histories (SP, stable population; GB, glacial bottleneck; GMR, glacial multiple refugia; GE, glacial expansion and WPB, wet phases bottleneck). *P* = probability of observing the empiric value, AIC = Akaike information criterion value for each demographic scenario. The best-fitting models/values are marked in bold

Scenario	SegSites	*P*	HapDiv	*P*	NuclDiv	*P*	AIC
COI							
Observed	203		0.986		0.07238		
SP	279.7–407.9	0	0.983–0.99	0.114	0.053–0.185	0.064	26.439
GB	230.1–401.5	0	0.961–0.996	**0.192**	0.048–0.182	**0.08**	**23.525**
GMR	110.7–265.6	**0.134**	0.945–0.996	0.014	0.012–0.092	0.06	24.428
GE	316.7–432.2	0	0.988–0.993	0	0.057–0.185	0.052	28.223
WPB	249.1–402.3	0	0.976–0.991	0.176	0.05–0.182	0.077	26.893
H3							
Observed	13		0.719		0.00318		
SP	8.3–32.8	0.033	0.583–0.969	0.065	0–0.022	0.065	23.909
GB	6–29.1	**0.064**	0.541–0.974	0.087	0–0.021	0.078	24.346
GMR	0.9–14.3	0.033	0.152–0.931	**0.167**	0–0.008	**0.138**	**22.905**
GE	12.9–37.1	0	0.606–0.973	0.06	0–0.022	0.06	25.537
WPB	7.2–30	0.054	0.555–0.969	0.079	0–0.021	0.074	26.185

differently than hypothesized by previous research for other dry forest taxa (e.g. Prado & Gibbs 1993).

Genetic variation in *S. cariri* is strongly correlated with geography for markers in the mitochondrial and nuclear genomes. Deep divergences and strong geographical structuring are common for spiders (e.g. Cooper *et al.* 2011; Planas & Ribera 2014), even for large species with good dispersal capabilities (Kuntner & Agnarsson 2011). This, allied to the simple and variable morphology typical of spiders with haplogyne genitalia (such as sicariids), makes cryptic species common in this group of organisms (Bond *et al.* 2001; Duncan *et al.* 2010; Satler *et al.* 2013). We here show that patterns of genetic variation in *S. cariri* are consistent with this nominal species being composed of at least two independently evolving lineages. This is supported by the facts that (i) there is no evidence of these two lineages forming a single, monophyletic lineage, (ii) genetic distances among populations of *S. cariri* as currently delimited are well above those of *S. tropicus* and *S. ornatus*, and (iii) a species tree with these two lineages considered separately fits the data better than a species tree considering them as part of the same species. For these reasons, we chose to conduct phylogeographical analysis for each clade separately. Further studies, including a closer examination of the genital morphology of these populations, would be needed to fully ascertain the species limits in *S. cariri*.

Geographical structuring is very strong even among populations of each clade of nominal *S. cariri*. This is reflected in its high fixation indexes, and in the fact that each sampling point is composed almost exclusively of COI haplotypes which are very closely related (Fig. 3). There are only a few studies of genetic variation of SDTF-associated organisms (Monteiro *et al.* 2004; Caetano *et al.* 2008; Collevatti *et al.* 2012; Werneck *et al.* 2012). Although some of these account for similarly high levels of geographic structuring, the scarcity of taxa studied makes the finding of common phylogeographic breaks difficult. In fact, even the phylogeographic barriers found for *S. cariri* are generally not congruent between COI and H3. We believe that this spatial incongruence is due to a lower resolution of this nuclear gene. The weaker geographical structuring observed for H3 is probably a consequence of the much slower mutation rate of this marker in relation to mtDNA (between 10 and 20 times slower), combined with a four times larger effective size (being a diploid, biparentally inherited autosomal gene). This makes the fixation of any new haplotype in a population much slower, so one should expect a weaker geographic structure for H3 (as we have observed). Finally, the one thing that appears common to *S. cariri* and other organisms is that populations from SDTF enclaves in the Cerrado domain represent unique genetic lineages. In addition, this region presents several narrowly endemic species (Guadanucci 2011; Werneck 2011). This reinforces the special biological significance of these enclaves and, along with the fact that they are under different kinds of threats (Werneck 2011), calls for conservation actions and further research in this region.

A picture of how SDTF-associated taxa responded to Pleistocene climatic shifts is still far from complete. It has been advocated that the present distribution of SDTF taxa is consistent with a wider distribution during the Pleistocene (Prado & Gibbs 1993). Paleodistribution

modelling suggests that SDTFs in general (Werneck *et al.* 2011) and some organisms in particular (this study) have not experienced a range expansion during the LGM. On the other hand, the SDTF-associated tree *Tabebuia impetiginosa* expanded its distribution in the same period (Collevatti *et al.* 2012). In the case of *S. cariri*, paleomodels suggest three exclusive hypotheses: (i) constant range size (Fig. 5e), (ii) multiple refugia (Fig. 5a, c) and (iii) extinction from eastern parts of the range (Fig. 5b, d, f). The results of Werneck *et al.* (2011) also indicate that the eastern parts of the Caatinga are not stable for SDTF occurrence. However, molecular data suggest that complete extinction of *S. cariri* from the eastern part of the range seems unlikely, as this region presents unique haplogroups for both COI and H3 (Fig. 3), and some eastern populations present high genetic diversity (e.g. population 11 from Fig. 3; see Table S4, Fig. S11, Supporting information). On the other hand, demographic reconstructions demonstrate that *S. cariri* (clade 1) populations suffered a reduction in size in the last 100 000 years (Fig. 4), with a slight recovery in the Holocene. Also, coalescent simulations indicate that LGM reduction/fragmentation of *S. cariri* populations are more probably to have generated observed genetic diversity values than stable or expanding populations during the same period. These results suggest that, rather than standing large and stable, *S. cariri* populations have been reduced (and possibly fragmented) during the Pleistocene. Wet periods in north-eastern Brazil caused replacement of Caatinga by humid forests (Auler *et al.* 2004), which left genetic signatures in rainforest-associated frogs (Carnaval & Bates 2007). *Sicarius cariri* is presently absent from humid forests, and it is reasonable to expect some of its populations to have been reduced or extinct during these wet periods. However, our coalescent simulation results indicate that reductions in population size during glacial periods are more probably to have occurred than reductions during Caatinga wet phases. This finding is very interesting, as it suggests that glacial periods took a heavier toll in the populations of this dry forest species than increases in rainfall—very much in disagreement with the initial hypothesis of Prado & Gibbs (1993).

Our results suggest that *Sicarius* spiders have been diversifying in the Caatinga for at least the past 10 Myr and species divergences seem to be no younger than 3–5 Myr old. *Sicarius cariri* major splits seem to have occurred in early to mid-Pleistocene, including those leading to the lineages occurring in SDTF enclaves in the Cerrado. One of the few phylogeographic studies of SDTF taxa found significant structuring among SDTF nuclei, but did not date the divergences (Caetano *et al.* 2008). Other studies show that divergence times coincide with the Pleistocene (Collevatti *et al.* 2012), with the Miocene (Werneck *et al.* 2012) or both (Pennington *et al.* 2004). The *S. cariri* present distribution is disjunct, with some populations located at SDTF enclaves in the Cerrado. The strong geographic structure of the genetic variation in this species suggests it is a poor disperser. This makes long-dispersal unlikely in this species and suggests that its presence in SDTF enclaves in the Cerrado is due to an ancient vicariant event. However, contrary to the expected from the Pleistocene Arc Hypothesis (Prado & Gibbs 1993), the paleodistribution modelling of *S. cariri* does not show evidence of distribution range expansion during the LGM. This is congruent with our coalescent simulations and demographic reconstruction analysis, which indicate that this species experienced a reduction in its population size during glacial periods. This, together with the old diversification history of this species, suggests that *S. cariri* colonized SDTF enclaves in the Cerrado in a more remote time. The geological events (if any) that might have allowed for this species to reach these regions, however, remain unclear.

The Caatinga in particular, and SDTFs in general, have been neglected as research systems (Werneck 2011; Turchetto-Zolet *et al.* 2013). As more organisms from these regions are studied, we will be able to paint a better picture of the diversification of this fascinating biome. *Sicarius* spiders seem to be suitable model species for studying SDTF biogeography, as they are strongly associated with this system, seem to be poor dispersers and have an ancient history of diversification in this biome. We hope that the study of other species of *Sicarius*, including a complete phylogeny of the genus, will shed more light on the evolutionary history of SDTFs.

Acknowledgements

We are grateful to the following researchers and institutions for lending museum specimens or collecting and sending samples: A.B. Kury (MNRJ), D.M.B. Battesti (IBSP), L.S. Carvalho (CNHUFPI), P.C. Motta (DZUB), R.L. Ferreira (UFLA) and T.J. Porto. We are deeply indebted to L.S. Carvalho, J.L. Chavari, G.F.B. Pereira, B.T. Faleiro and A.A.S. Monteiro for providing vital help during fieldwork. G.H.F. Azevedo and the crew of Laboratório de Biodiversidade e Evolução Molecular kindly helped I.L.F.M. to obtain part of the sequences used in this study. We thank Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio), SEMA-Bahia, COPAM-Ceará, and SUDEMA-Paraíba for conceding collecting permits and the staff of several Brazilian conservation units for providing help in the field. Earlier versions of this study were greatly improved by fruitful discussions with D.R.M. Neves, comments from F.P. Werneck, A.R. Pepato and G.H.F. Azevedo, and a thorough revision by R. Gillespie, M. Kuntner, I. Agnarsson and two anonymous reviewers. This is a

contribution from Pós-Graduação em Ecologia, Conservação e Manejo da Vida Silvestre—UFMG (I.L.F.M.) and Pós-Graduação em Zoologia—UFMG (U.O.). This work was financially supported by a CNPq scholarship to I.L.F.M. (Proc. 132086/2011-5), a CAPES/REUNI scholarship to U.O., and by grants from FAPEMIG (APQ-01991-09; PPM-00553-11, PPM-00335-13), CNPq (Procs. 474680/2010-0; 308072/2012-0) and Instituto Nacional de Ciência e Tecnologia dos Hymenoptera Parasitóides da Região Sudeste Brasileira (http:/www.hympar.ufscar.br/) to A.J.S., from FAPEMIG to T.H.D.A.V., from CNPq and FAPEMIG to F.R.S. and from FAPESP (Proc. 2011/50689-0) and CNPq (Proc. 301776/2004-0) to A.D.B.

References

Auler AS, Wang X, Edwards RL *et al.* (2004) Quaternary ecological and geomorphic changes associated with rainfall events in presently semi-arid northeastern Brazil. *Journal of Quaternary Science*, **19**, 693–701.

Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution*, **29**, 2157–2167.

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Beerli P, Palczewski M (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics*, **185**, 313–326.

Bidegaray-Batista L, Arnedo M (2011) Gone with the plate: the opening of the Western Mediterranean basin drove the diversification of ground-dweller spiders. *BMC Evolutionary Biology*, **11**, 1–15.

Bond JE, Hedin MC, Ramírez M, Opell BD (2001) Deep molecular divergence in the absence of morphological and ecological change in the Californian coastal dune endemic trapdoor spider *Aptostichus simus*. *Molecular Ecology*, **10**, 899–910.

Byrne M, Yeates D, Joseph L *et al.* (2008) Birth of a biome: insights into the assembly and maintenance of the Australian arid zone biota. *Molecular Ecology*, **17**, 4398–4417.

Caetano S, Prado DE, Pennington RT *et al.* (2008) The history of Seasonally Dry Tropical Forests in eastern South America: inferences from the genetic structure of the tree *Astronium urundeuva* (Anacardiaceae). *Molecular Ecology*, **17**, 3147–3159.

Carnaval AC, Bates JM (2007) Amphibian DNA shows marked genetic structure and tracks Pleistocene climate change in Northeastern Brazil. *Evolution*, **61**, 2942–2957.

Cartelle C, Hartwig WC (1996) A new extinct primate among the Pleistocene megafauna of Bahia, Brazil. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 6405–6409.

Colgan DJ, McLauchlan A, Wilson GDF *et al.* (1998) Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolution. *Australian Journal of Zoology*, **46**, 419–437.

Collevatti RG, Terribile LC, Lima-Ribeiro MS *et al.* (2012) A coupled phylogeographical and species distribution modelling approach recovers the demographical history of a Neotropical seasonally dry forest tree species. *Molecular Ecology*, **21**, 5845–5863.

Cooper S, Harvey M, Saint K, Main BY (2011) Phylogeographic structuring of populations of the trapdoor spider *Moggridgea tingle* (Migidae) from south-western Australia: evidence for long-term refugia within refugia. *Molecular Ecology*, **20**, 3219–3236.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUTI and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Duncan RP, Rynerson MR, Ribera C, Binford GJ (2010) Diversity of *Loxosceles* spiders in Northwestern Africa and molecular support for cryptic species in the *Loxosceles rufescens* lineage. *Molecular Phylogenetics and Evolution*, **55**, 234–248.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Excoffier L, Smouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Excoffier L, Novembre J, Schneider S (2000) SIMCOAL: a general coalescent program for simulation of molecular data in interconnected populations with arbitrary demography. *Journal of Heredity*, **91**, 506–509.

Excoffier L, Laval G, Schneider S (2005) ARLEQUIN (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Guadanucci JPL (2011) Cladistic analysis and biogeography of the genus *Oligoxystre* Vellard 1924 (Araneae: Mygalomorphae: Theraphosidae). *Journal of Arachnology*, **39**, 320–326.

Hedin MC (1997) Molecular phylogenetics at the population/species interface in cave spiders of the southern Appalachians (Araneae: Nesticidae: *Nesticus*). *Molecular Biology and Evolution*, **14**, 309–324.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Kuntner M, Agnarsson I (2011) Phylogeography of a successful aerial disperser: the golden orb spider *Nephila* on Indian Ocean islands. *BMC Evolutionary Biology*, **11**, 1–9.

Librado P, Rozas J (2009) DNASP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Macías-Hernandez N, Oromí P, Arnedo MA (2010) Integrative taxonomy uncovers hidden species diversity in woodlouse hunter spiders (Araneae, Dysderidae) endemic to the Macaronesian archipelago. *Systematics and Biodiversity*, **8**, 531–553.

Magalhães ILF, Brescovit AD, Santos AJ (2013) The six-eyed sand spiders of the genus *Sicarius* (Araneae: Haplogynae: Sicariidae) from the Brazilian Caatinga. *Zootaxa*, **3599**, 101–135.

Manni F, Guerard E, Heyer E (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. *Human Biology*, **76**, 173–190.

Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE)*, 14 November 2010, New Orleans, Louisiana, pp. 1–8.

Monteiro FA, Donnelly MJ, Beard CB, Costa J (2004) Nested clade and phylogeographic analyses of the Chagas disease vector *Triatoma brasiliensis* in Northeast Brazil. *Molecular Phylogenetics and Evolution*, **32**, 46–56.

Muñoz MES, Giovanni R, Siqueira MF *et al*. (2011) OpenModeller: a generic approach to species' potential distribution modelling. *Geoinformatica*, **15**, 111–135.

Nylander JAA (2004) *MrModeltest v2*. Available from http://www.abc.se/~nylander/mrmodeltest2/mrmodeltest2.html. Last accessed January 7, 2013.

de Oliveira PE, Barreto AMF, Suguio K (1999) Late Pleistocene/Holocene climatic and vegetational history of the Brazilian caatinga: the fossil dunes of the middle São Francisco River. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **152**, 319–337.

Pennington RT, Prado DE, Pendry CA (2000) Neotropical seasonally dry forests and Quaternary vegetation changes. *Journal of Biogeography*, **27**, 261–273.

Pennington RT, Lavin M, Prado DE, Pendry CA, Pell SK, Butterworth CA (2004) Historical climate change and speciation: Neotropical seasonally dry forest plants show patterns of both Tertiary and Quaternary diversification. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, **359**, 515–537.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Planas E, Ribera C (2014) Uncovering overlooked island diversity: colonization and diversification of the medically important spider genus *Loxosceles* (Arachnida: Sicariidae) on the Canary Islands. *Journal of Biogeography*, **41**, 1255–1266.

Prado DE (2000) Seasonally dry forests of tropical South America: from forgotten ecosystems to a new phytogeographic unit. *Edinburgh Journal of Botany*, **57**, 437–461.

Prado DE, Gibbs PE (1993) Patterns of species distributions in the dry seasonal forests of South America. *Annals of the Missouri Botanical Garden*, **80**, 902–927.

R Core Team (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available from http://www.R-project.org. Last accessed June 29, 2014.

Ramos KS, Melo GA (2010) Taxonomic revision and phylogenetic relationships of the bee genus *Parapsaenythia* Friese (Hymenoptera, Apidae, Protandrenini), with biogeographic inferences for the South American Chacoan subregion. *Systematic Entomology*, **35**, 449–474.

Rodrigues MT (1996) Lizards, snakes, and amphisbaenians from the quaternary sand dunes of the middle Rio São Francisco, Bahia, Brazil. *Journal of Herpetology*, **30**, 513–523.

Santos RM, Oliveira-Filho AT, Eisenlohr PV, Queiroz LP, Cardoso DBOS, Rodal MJN (2012) Identity and relationships of the Arboreal Caatinga among other floristic units of seasonally dry tropical forests (SDTFs) of north-eastern and Central Brazil. *Ecology and Evolution*, **2**, 409–428.

Satler JD, Carstens BC, Hedin M (2013) Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae, *Aliatypus*). *Systematic Biology*, **62**, 805–823.

Silva JMC, Tabarelli M, Fonseca MT, Lins LV (2004) *Biodiversidade da Caatinga: áreas e ações prioritárias para a conservação*, 382 p. Ministério do Meio Ambiente, Brasília.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Turchetto-Zolet AC, Pinheiro F, Salgueiro F, Palma-Silva C (2013) Phylogeographical patterns shed light on evolutionary process in South America. *Molecular Ecology*, **22**, 1193–1213.

Werneck FP (2011) The diversification of eastern South American open vegetation biomes: historical biogeography and perspectives. *Quaternary Science Reviews*, **30**, 1630–1648.

Werneck FP, Costa GC, Colli GR, Prado DE, Sites JW (2011) Revisiting the historical distribution of Seasonally Dry Tropical Forests: new insights based on palaeodistribution modelling and palynological evidence. *Global Ecology and Biogeography*, **20**, 272–288.

Werneck FP, Gamble T, Colli GR, Rodrigues MT, Sites JW Jr (2012) Deep diversification and long-term persistence in the South American 'dry diagonal': integrating continent-wide phylogeography and distribution modeling of geckos. *Evolution*, **66**, 3014–3034.

A.J.S. and I.L.F.M. conceived the initial idea of the study. I.L.F.M., A.J.S. and A.D.B. obtained samples from the field and museums and identified specimens. I.L.F.M. obtained genetic data. U.O. and I.L.F.M. designed and performed analyses of distribution modelling. I.L.F.M., A.J.S., T.H.D.A.V. and F.R.S. discussed and performed genetic analyses. All authors discussed the main results. I.L.F.M. and A.J.S. lead the writing of the first version of the manuscript. All authors commented and improved the final version of the manuscript.

Data accessibility

DNA sequences: deposited in GenBank (accession nos. KM582417-KM582599). Voucher, geographic and haplotypic data of individuals, original, unphased alignments, presence and absence data for generating and validating distribution models, Beast XML input files, annotated tree files and inputs for BayesSC deposited at the Dryad repository (doi:10.5061/dryad.ps12h).

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Voucher and locality information.

Table S2 GenBank accession nos. for sequences used in this study.

Table S3 Information on preliminary runs of BEAST for setting priors.

Table S4 Summary statistics of population diversity of *Sicarius cariri*.

Table S5 Results of the molecular analysis of variance.

Fig. S1 Median-joining network of COI haplotypes of *Sicarius cariri*.

Fig. S2 Median-joining network of H3 haplotypes of *Sicarius cariri*.

Fig. S3 Intra- and interspecific K2P genetic distances among Brazilian species of *Sicarius* for COI.

Fig. S4 Relationship between geographic and K2P genetic distances for populations of *Sicarius cariri* clade 1.

Figs S5–S6 Present-day distribution models for *Sicarius cariri* using different algorithms.

Figs S7–S10 Measures of accuracy for the models of distribution of *Sicarius cariri*.

Fig. S11 Spatial distribution of nucleotide diversity in *Sicarius cariri*.