Molecular Phylogenetics and Evolution 95 (2016) 11–19

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Anonymous nuclear markers reveal taxonomic incongruence and long-term disjunction in a cactus species complex with continental-island distribution in South America ☆

Manolo F. Perez [a], Bryan C. Carstens [b], Gustavo L. Rodrigues [a], Evandro M. Moraes [a,*]

[a] Departamento de Biologia, Universidade Federal de São Carlos, Rodovia João Leme dos Santos km 110, Sorocaba, São Paulo 18052780, Brazil
[b] Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, Columbus, OH, USA

ARTICLE INFO

Article history:
Received 27 May 2015
Revised 5 November 2015
Accepted 6 November 2015
Available online 12 November 2015

Keywords:
Species tree
Next generation sequencing
Molecular markers
Phylogeography
Non-model species
Climatic Niche Modelling

ABSTRACT

The *Pilosocereus aurisetus* complex consists of eight cactus species with a fragmented distribution associated to xeric enclaves within the Cerrado biome in eastern South America. The phylogeny of these species is incompletely resolved, and this instability complicates evolutionary analyses. Previous analyses based on both plastid and microsatellite markers suggested that this complex contained species with inherent phylogeographic structure, which was attributed to recent diversification and recurring range shifts. However, limitations of the molecular markers used in these analyses prevented some questions from being properly addressed. In order to better understand the relationship among these species and make a preliminary assessment of the genetic structure within them, we developed anonymous nuclear loci from pyrosequencing data of 40 individuals from four species in the *P. aurisetus* complex. The data obtained from these loci were used to identify genetic clusters within species, and to investigate the phylogenetic relationship among these inferred clusters using a species tree methodology. Coupled with a palaeodistributional modelling, our results reveal a deep phylogenetic and climatic disjunction between two geographic lineages. Our results highlight the importance of sampling more regions from the genome to gain better insights on the evolution of species with an intricate evolutionary history. The methodology used here provides a feasible approach to develop numerous genealogical molecular markers throughout the genome for non-model species. These data provide a more robust hypothesis for the relationship among the lineages of the *P. aurisetus* complex.

1. Introduction

The limitations of using one or a few genes to assess phylogenetic/phylogeographic relationships and demographic history of species is extensively documented (Edwards and Beerli, 2000; Edwards, 2009; Andrew et al., 2013), and the use of datasets consisting of sequence data from multiple genomic regions can improve inferences by accounting for the stochasticity in the coalescent process (Knowles and Maddison, 2002; Knowles, 2009; Carstens et al., 2013).

The development of molecular markers in non-model species has been facilitated in recent years by new sequencing technologies, which make it possible to quickly develop genomic datasets (Glenn, 2011; Lemmon and Lemmon, 2013; Garrick et al., 2015). Moreover, these new strategies allow for simultaneous marker development and polymorphism genotyping, and these large datasets are useful to study recent species radiations (McCormack et al., 2012). Only a few studies in the Cactaceae have collected data from nuclear genes, and/or microsatellites (Edwards et al., 2005; Majure et al., 2012; Ritz et al., 2012; Franck et al., 2013; Bonatelli et al., 2014), while the majority of molecular systematics in this family has relied on plastid DNA (Nyffeler and Eggli, 2010 and references therein; Ritz et al., 2007; Arakaki et al., 2011; Hernández-Hernández et al., 2011, 2014; Calvente et al., 2011).

The *Pilosocereus aurisetus* complex, which consists of eight columnar cactus species associated exclusively with the rocky savannas in eastern Brazil, has been defined on the basis of morphological characters (Zappi, 1994; Taylor and Zappi, 2004; Hunt et al., 2006). The *P. aurisetus* complex has two species with broad distributions, *P. aurisetus* in southeastern Brazil, along the

☆ This paper was edited by the Associate Editor Quan Wang Xiao.

* Corresponding author. Fax: +55 15 32295982.
E-mail address: emarsola@ufscar.br (E.M. Moraes).

http://dx.doi.org/10.1016/j.ympev.2015.11.005
1055-7903/

Espinhaço Mountain Range, and *P. machrisii*, with populations scattered in central and southeastern Brazil. The species *P. jauruensis* and *P. vilaboensis* are narrowly distributed, with populations located in mountains from central-western Brazil. Four species (*P. parvus*, *P. pusillibaccatus*, *P. aureispinus* and *P. bohlei*) have a single or very few known populations, located in the borders of the complex distribution (Fig. 2). Species within this complex are differentiated from each other by several lines of evidence: on the basis of their geographic ranges, by vegetative and reproductive morphological characters, principally the plant habit, number of ribs, spination patterns, flower color, ovary locules, as well as a number of seed and fruit characteristics (Hunt et al., 2006). Taxonomy in these species has not been stable; several described species have been synonymized in recent years because they exhibit considerable intraspecific polymorphism, and there remains overlapping morphologic variation in the interspecific level (Zappi, 1994), a characteristic common for the Cactaceae family (Gibson and Nobel, 1986). In addition, the taxonomy is similarly complex in other co-distributed taxa, particularly in plants such as *Vellozia* (Barbosa et al., 2012) and orchids (Antonelli et al., 2010), which has prompted authors to attribute these patterns to population expansion and retraction events, coupled with secondary contact.

Previous studies in the *Pilosocereus aurisetus* complex using allozymes (Moraes et al., 2005), as well as microsatellites, cpDNA sequences, and a nuclear gene (Bonatelli et al., 2014), suggest that the diversification is very recent, paralleling the observed pattern for the entire Cactaceae family (Arakaki et al., 2011; Hernández-Hernández et al., 2011). Due to low resolution in these genetic markers, these studies were not able to recover some of the relationships between the observed lineages. Specifically, these studies uncovered the existence of the two distinct lineages in the widely distributed species *P. machrisii*, but failed to indisputably recover the population composition of each lineage, and its relationships with the other species of the complex (Moraes et al., 2005; Bonatelli et al., 2014). It was also not possible to assess the relationship of the *P. aurisetus* northern populations, which were recovered as distinct from the other conspecific populations, and showed distinct clustering patterns with cpDNA and microsatellite data. Furthermore, the species *P. jauruensis* that shows the westernmost distribution of the complex, was more related to *P. vilaboensis* populations in central Brazil for the nuclear gene *PhyC*, but also showed a closer relationship to the southern *P. machrisii* populations in the cpDNA data (Bonatelli et al., 2014).

Therefore, the development and validation of polymorphic nuclear markers for the *Pilosocereus* genus represents an intriguing strategy to address the limitations of previous studies. Moreover, integrating molecular data with methods for estimating climatic niches from current occurrence data can also be useful to assess the phylogeographic history of these species. Recent advances in this field have allowed researchers to compare niches of different lineages (Zellmer et al., 2012; Joly et al., 2014), as well as address several ecological questions, such as defining biogeographical hypothesis (Carnaval and Moritz, 2008; Collevatti et al., 2012).

The work presented here is intended to address specific questions about the diversification of *P. aurisetus* complex: (1) Are the northern *P. aurisetus* populations more related to the other conspecific populations in the Espinhaço Mountain range or to population from other species in Central Brazil, as shown by cpDNA data? (2) Is the currently recognized *P. machrisii* species composed of two distinct lineages? (3) What is the relationship of *P. jauruensis* with the other species of the complex? To answer these specific questions, we recovered the main structure and a species tree for the populations of the *P. aurisetus* complex using the developed anonymous nuclear markers (Thomson et al., 2010). Further, we also tested climatic niche differences between the observed geographic lineages.

2. Methods

2.1. Library preparation and pyrosequencing

Amplicon genomic libraries were prepared for 40 *Pilosocereus* samples from 4 species belonging to the *P. aurisetus* species group and from *P. gounellei* species, a species belonging to a distinct subgenus (*Gounellea* Zappi) of *P. aurisetus* which was used as an outgroup (Table 1). This subset of species was selected because they have the widest distributions, and showed more complex and unresolved phylogenetic relationships in a previous work (Bonatelli et al., 2014). We followed the AFLP protocol from Vos et al. (1995), with modifications for pyrosequencing developed by other authors (McCormack et al., 2012; Zellmer et al., 2012). Briefly, we extracted total DNA using the Qiagen DNeasy Plant Mini Kit (Qiagen, Hilden, Germany), these extractions were then purified using Zymo Genomic DNA Clean & Concentrator Kit (Zymo Research, Irvine, CA, USA). We digested and ligated adaptors to 250 ng of DNA in a 11 μL reaction containing 1.1 μL T4 ligase buffer, 1.1 μL of 0.5 mM NaCl, 5 U *Eco*RI, 5 U *Mse*I, 0.55 μL of 1 μg/μL BSA, 5 U T4 ligase and 1.0 μL of 10 μM of *Mse*I and *Eco*RI adaptors (Vos et al., 1995). We then amplified the fragments in a 20 μL PCR reaction with 10 μL of digest-ligation reaction diluted 10X, 5.4 μL water, 2 μL of 25 mM $MgCl_2$, 2 μL of 10X buffer, 0.4 μL of 10 mM dNTPs, 0.06 μL of 100 μM concentration adaptor-specific primer (*Mse*I: 5′ GATGAGTCCTGAGTAA and *Eco*RI: 5′ GACTGCGTACCAATTC), and 0.08 μL of 5 U/μL Phusion high-fidelity *Taq* (New England Biosciences, Ipswich, MA, USA). The conditions of the PCR reactions were 2 min at 72 °C; followed by 15 cycles of 98 °C for 30 s, 56 °C for 30 s, and 72 °C for 2 min; followed by 72 °C for 10 min. The PCR products were visualized individually on an

Table 1
Population samples from *P. aurisetus* complex used to prepare pyrosequencing libraries.

Species	Population	Code	Number of samples
P. machrisii (E.Y. Dawson) Backeb.	Delfinópolis	DEL*	5
	Cristalina	CRI*	5
	Alto Paraíso de Goiás	APA*	5
	Aurora do Tocantins	ART*	5
P. aurisetus (Werderm.) Byles & G.D. Rowley	Grão Mogol	GMII*	5
	Mendanha	MEN*	5
P. vilaboensis (Diers & Esteves) P.J. Braun	Pirenópolis	PIR*	5
P. jauruensis (Buining & Brederoo) P.J. Braun	Rio Verde do Mato Grosso	RVE*	4
P. gounellei (F.A.C. Weber) Byles & G.D. Rowley	Milagres	GO-1078*	1

* Vouchers were deposited at the Universidade Federal de São Carlos, Campus Sorocaba Herbarium: DEL (HUFS636), CRI (HUFS643), APA (HUFS645), ART (SORO3620), GMII (HUFS640), MEN (SORO3619), PIR (HUFS641), RVE (SORO3617), GO-1078 (SORO3618).

agarose gel prior to excision of a section containing 400–550 bp fragments. The excised gel sample was purified using a QIAquick gel extraction kit (Valencia, CA, USA), and eluted with 50 μL volume. A second round of PCR was performed with 2.5 μL of purified PCR product, 4.56 μL water, 1 μL of 25 mM $MgCl_2$, 1 μL 10X buffer, 0.2 μL 10 mM dNTPs, 0.06 μL of *Mse*I reverse fusion primer (Roche B Fusion/*Mse*I Sequence/Selective Bases) at 100 μM, containing 2 selective base pairs (CC) at its 3′ end, as suggested by Zellmer et al. (2012), and 0.08 μL of 5 U/μL Phusion *Taq*. To each sample, we separately added 0.6 μL of *Eco*RI forward fusion primer with indexes at 10 μM concentration (Roche A Fusion/index/*Eco*RI Sequence; see McCormack et al., 2012 for indexes sequences). For this PCR, we used a touchdown profile beginning with 94 °C for 2 min; then 10 cycles of 94 °C for 30 s, 65 °C for 30 s (reducing temperature by 0.7 °C in each cycle), 72 °C for 60 s; then 10 cycles of 94 °C for 30 s, 56 °C for 30 s, and 72 °C for 60 s; followed by 10 min at 72 °C. Final DNA concentration was quantified using a combination of Picogreen (Molecular Probes, Eugene, OR, USA) fluorescent dye assay and a LabChip assay on an Agilent BioAnalyzer. The samples were sequenced in two Roche 454 runs with 20 individuals each at the Center for Advanced Technologies (CATG-USP, São Paulo, SP, Brazil).

2.2. Bioinformatic analysis

The package pyRAD (Eaton, 2014) was used to process sequence reads and identify variable loci across the *P. aurisetus* species and populations. Details of the filtering steps are available in Perez et al. (in press: Table 1). In order to validate the results and minimize the amount of missing data, primers were developed for the obtained loci (Table 3 in Perez et al., in press). These primers were designed at positions of the flanking regions without observable nucleotide variation in the obtained sample. PCR reactions were carried out in a total volume of 25 μL, containing 1X Sigma Buffer, 10 mM dNTPs, 10 μM of each primer, 0.2 U of Sigma *Taq* Polymerase and 1 μL of genomic DNA, with the following thermocycler conditions: 94 °C for 2 min; 35 cycles of 94 °C for 30 s, specific Tm (Table 1) for 30 s, 72 °C for 30 s; 72 °C for 2 min. All sequencing reactions were performed in Macrogen Inc. (Seoul, South Korea) facility. The resulting loci were quality-controlled for recombination using the DSS method (McGuire and Wright, 2000) as implemented in the software package TOPALi v2 (Milne et al., 2009). We also tried to detect loci under selection using Tajima's D, Fu and Li's D^* and F^* in DNAsp (Librado and Rozas, 2009).

2.3. Population structure

Sequences were reduced to alleles by hand prior to genetic clustering analysis. The Bayesian analysis implemented in STRUCTURE 2.3.4 (Pritchard et al., 2000) was performed to infer the most likely number of interbreeding groups (K) in our dataset (excluding the outgroup) without prior information on the sampling sites for each sample. Five independent runs for each K from 1 to 9 (the total number of population + 1) were performed. For each simulation, 2 000 000 interactions were performed, with the initial 500 000 discarded as burn-in. Because the dataset contains more than one population from *P. aurisetus* and *P. machrisii* species, the correlated allele frequencies and linkage model was utilized. DISTRUCT v1.1 (Rosenberg, 2004) was used to provide a graphical representation of the STRUCTURE results. The most likely K was determined with the following criteria: (1) stability in the clustering patterns of different runs for the same K; (2) smaller value of K after which the posterior probability values reach a plateau (Pritchard et al., 2000); (3) ΔK statistics (Evanno et al., 2005); (4) absence of "virtual" groupings, that is, groups containing only individuals with genomes scattered in more than one cluster.

2.4. Species tree

Once groups were identified using STRUCTURE, we treated these as operational taxonomic units (OTUs) and estimated a species tree using BEAST 2 (Bouckaert et al., 2014). We performed this analysis using a Yule speciation prior, with the most likely model of sequence evolution obtained in jModeltest2 (Darriba et al., 2012). We used either a strict or a relaxed lognormal clock at each locus, selected after comparing the marginal likelihoods of runs using each model with a Path Sampling analysis with 8 steps and 500 000 generation after a 50% burn-in. The species tree was obtained after two independent runs of 100 000 000 MCMC generations each, with a 10% burn-in, and sampling trees every 5000 steps.

2.5. Climatic Niche Modelling

Species occurrence data was obtained by GPS measurements during field trips across the range of the complex, and also by available occurrences from Global Biodiversity Information Facility (GBIF http://www.gbif.org/). The later were manually inspected to verify that the specific locality was reasonable based on the known distribution of each taxon. The sample sizes were generally small for each species, reflecting the relative rarity and range restriction of the species of this complex: *P. aurisetus* 51, *P. machrisii* 34, *P. vilaboensis* 5, *P. jauruensis* 7. Because the occurrence data were relatively sparse, overfitting was unlikely. In order to test climatic divergence, in addition to the genetic differences observed, we grouped the occurrences according to the main genetic lineages recovered in the analyzes described above. To test the effects of past climatic oscillations on the niche of each lineage, we fitted the models in present, 21 kya (LGM), and 120 kya (LIG) scenarios using three different algorithms: GARP, Support Vector Machines (SVM hereon) and neural networks implemented in OpenModeller (Muñoz et al., 2011), and the Maximum Entropy algorithm, implemented in Maxent 3.3.3 (Phillips et al., 2006). We selected six bioclimatic variables (temperature seasonality, mean temperature of warmest quarter, precipitation of wettest month, precipitation of wettest quarter, precipitation of warmest quarter, precipitation of coldest quarter) that showed low correlation and high informativeness after a jackknife procedure on the 19 BIOCLIM variables downloaded from the WorldClim data set (Hijmans et al., 2005). We used the data delimited for South America region with a 2.5 arc-min. To identify suitable areas for each period (Present, LGM and LIG), we converted the continuous outputs into presence/absence maps by selecting threshold values at which sensitivity (proportion of true positive predictions versus number of actual positives) is equal to specificity (proportion of true negative predictions versus number of actual negatives), and then overlapped the four presence/absence projections for each spatial definition. Suitable areas were defined with overlap of at least three projections. We also defined the climatic stable areas, which were established for the overlapping stable areas during the three periods (Fig. 3).

3. Results

3.1. Pyrosequencing

The two Pyrosequencing runs resulted in 2 282 266 sequences with 483 bp average length. After quality controlling the reads

using FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), 1 471 135 high quality (90% with $Q \geqslant 20$) barcoded sequences with more than 100 base pairs were retained. We recovered a mean of 892 loci per individual, and then clustered loci from different individuals, resulting in a total of 223 loci occurring in at least 10 individuals. Those loci were 461.74 base pairs in average, with a mean of 9.17 SNPs per locus. After a manual inspection following the guidelines proposed by Hird et al. (2011), loci with no variation and potential paralogs were removed based on the amount of variation between and within individuals. After this proceeding, we observed 36 loci present in all populations. Blastn searches in those loci resulted in 4 of them matching cpDNA sequences, 3 matching sequences on the mtDNA, 2 similar to retrotransposon sequences, and 13 sequences with significant similarity with plant nuclear genomes. All loci that matched cytoplasmatic sequences (cp and mtDNA) and retrotransposons were discarded, resulting in 26 loci in all populations sampled. Primers for these 26 loci showed specific amplification in at least one sample, but one marker was discarded from further analysis owing to amplification and sequencing problems in the outgroup. In order to improve analysis relying on population parameters, Sanger sequencing reactions were obtained for 117 sequences (containing both strands), selected to assure data for at least two individuals for each locus. After combining sequences from both Sanger and pyrosequencing for the 25 loci, a total of 687 sequences over 40 individuals were obtained (Supplementary Table in Perez et al., in press), with a total of 367 SNPs. The analysis of recombination did not suggest any significant results, and selection was not detected for any locus (Table 3 in Perez et al., in press).

3.2. Population structure

STRUCTURE results suggested a main population structure in five groups ($K = 5$), but Evanno's statistics also showed a peak for $K = 3$ and $K = 7$, indicating hierarchical population structure. Considering $K = 5$ (Fig. 1), each sampling locality was allocated in a single group, except for the two *P. aurisetus* localities (GMO and MEN), and for a group clustering three localities sampled in central Brazil (group GO), two of them from *P. machrisii* (APA and CRI) and one from *P. vilaboensis* (PIR).

3.3. Species tree

The species tree analyzes were conducted using the sequence evolution and clock models recovered for each marker (Table S2). All parameters demonstrated convergence (ESS > 100) in TRACER

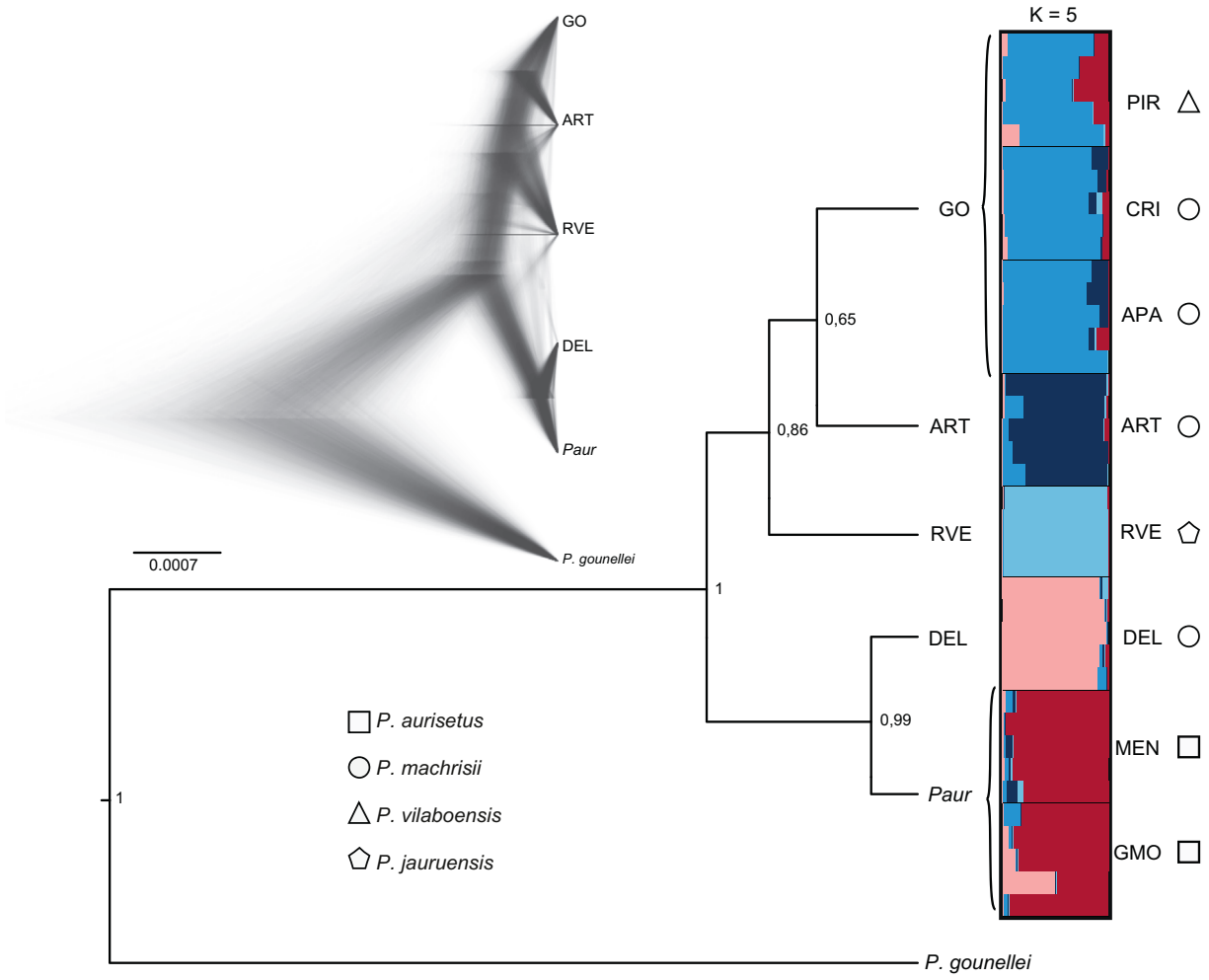

Fig. 1. Estimated population structure and species tree relationships among STRUCTURE genetic clusters. Each individual is represented as a vertical bar in the STRUCTURE results, showing the proportion of its genome assigned to each of the five clusters. Black lines separate individuals of different sampled populations. Posterior probabilities are shown for each branch in the species tree. The insert shows a density tree, with the retained trees during the BEAST analysis. Symbols denoting each species are shown for each cluster.

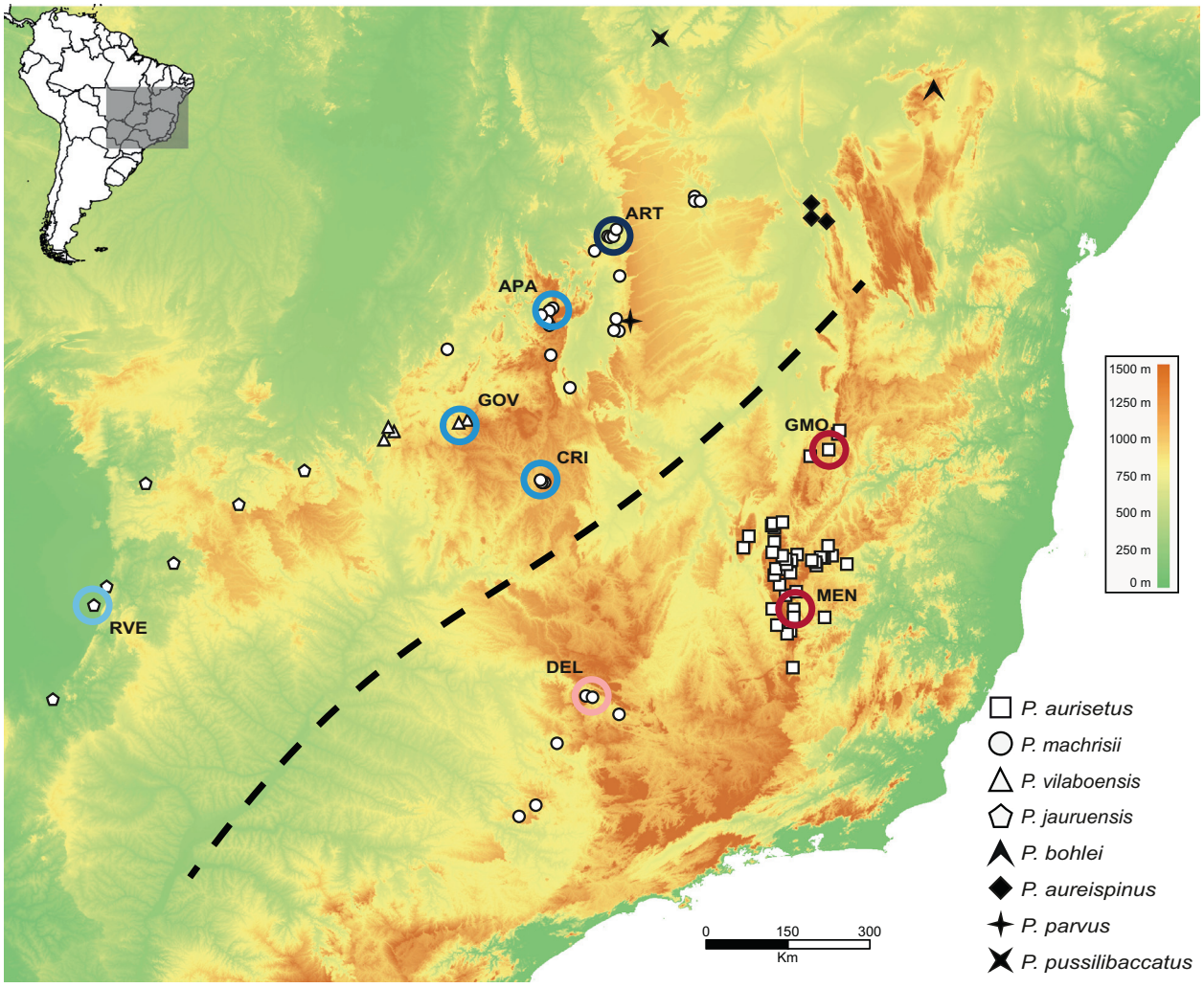

Fig. 2. Distribution of *Pilosocereus aurisetus* complex species. Colored circles represent populations used for pyrosequencing, and colors follow STRUCTURE clusters in Fig. 1. Dashed line split the main clades recovered by the species tree. All occurrence points used in the Climatic Niche Modelling analyzes are shown with white symbols. Black symbols represent species from the *P. aurisetus* complex that were not analyzed in the study.

1.6 (Rambaut et al., 2014), and a Maximum Clade Credibility (MCC) tree was generated in TreeAnotator (Drummond and Rambaut, 2007), by combining the trees from the two runs. The obtained MCC tree suggested that two main geographic lineages (Fig. 2) with high support (PP = 1) are present. The first lineage (SE) was distributed in the southeastern part of the *P. aurisetus* complex known occurrence, and contained the *P. aurisetus* species OTU (MEN and GMO) and the *P. machrisii* population DEL. The second main lineage (NW) was composed by the STRUCTURE's group GO, as well as ART from *P. machrisii*, and RVE from *P. jauruensis*, all of them located in the northwest part of the species complex distribution (Fig. 2). All branches within these main clades are highly supported, except for the one that unites ART and GO, which has a posterior probability of 0.65. This low support apparently results from differences in the gene trees topologies (Fig. 1); this is a common feature in species trees (McCormack et al., 2011).

3.4. Climatic Niche Modelling

The climatic niche distributions estimated for each of the two main geographic lineages (SE and NW) in the species tree (Figs. 3 and 4) show a very small overlap across all the three time periods (Present, LGM and LIG). Furthermore, the niche overlap was even smaller when the stable areas were considered (Fig. 3). Taken as a whole, the climatic niche does not appear to have changed dramatically over the three time periods considered here, suggesting that the distributions of the two lineages have not been severely affected by range shifts.

4. Discussion

The development and sequencing of a set of anonymous nuclear loci in the recently diverged species from the *P. aurisetus* complex allowed us to assess the phylogeographic patterns and to address questions raised in previous studies that used classical markers with limited resolution (Moraes et al., 2005; Bonatelli et al., 2014). While in most cases, the increased resolution provided by these data affords a more developed view of the recent evolutionary history in this clade, in some cases the results contradict those from earlier investigations. We explore each in turn.

Our analyses identify two well-differentiated groups within the morphologically defined *P. machrisii* species. One of these forms a clade with southern populations of *P. aurisetus*, while two northern *P. machrisii* populations form a clade with the remaining species (Fig. 1). *Pilosocereus machrisii* is the most widely distributed species in the *P. aurisetus* complex (Zappi, 1994; Fig. 2), and its taxonomic boundaries have been recently reviewed. As a result of this review, some local populations have been raised to the species status

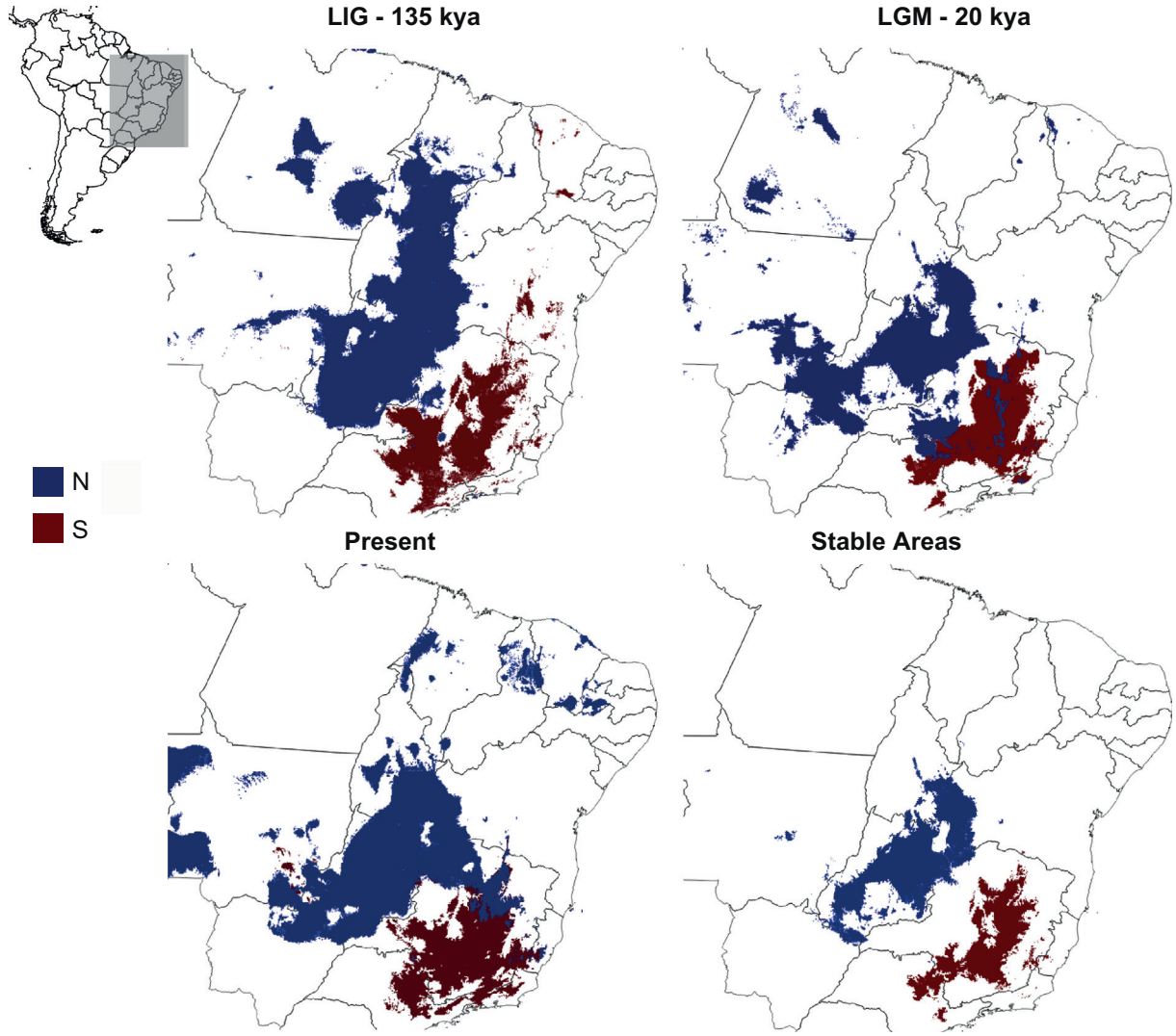

Fig. 3. Estimated climatic niche for the two main lineages (NW and SE) recovered in the species tree analysis. Maps are shown for each of the three periods tested (Present, LGM and LIG) and for the stable area, recovered in all periods.

(Hunt et al., 2006), while other studies have suggested more unrecognized variation is present within this taxon (Moraes et al., 2005; Bonatelli et al., 2014). Our results obtained suggest that the taxonomy of the *P. aurisetus* complex, which was developed solely on the basis of morphological characters, is incomplete. For example, *P. machrisii* southern samples cluster with those from *P. aurisetus*, while northern *P. machrisii* localities apparently share a close relationship with *P. vilaboensis*, a result previously suggested by allozyme markers (Moraes et al., 2005). Moreover, *P. jauruensis* is closely related to the central Brazil populations, but not related to the southern populations of *P. machrisii*, as suggested by microsatellites and cpDNA (Bonatelli et al., 2014). The multilocus sequence data presented appear to offer greater resolution of the relationships within this complex.

In one notable case, results presented here are in conflict with previous work on this group. A previous analysis of plastid data (Bonatelli et al., 2014) suggested that northern and southern *P. aurisetus* populations are genetically isolated, but this finding is not supported by analysis of the nuclear dataset (Fig. 1), where both northern and southern *P. aurisetus* populations are included in a single cluster by the STRUCTURE analysis (Fig. 1). Bonatelli et al. (2014) additionally detected a geographic overlap between cpDNA lineages from northern *P. aurisetus* and the other central-northern populations of the *P. aurisetus* complex and interpreted this as evidence of secondary contact. However, the results presented here do not support the secondary contact hypothesis. The different patterns seen between the two datasets could be attributed to a number of causes: (i) incomplete lineage sorting of the cpDNA variation, (ii) to introgression events that might have affected the cpDNA genome of northern *P. aurisetus* populations, or (iii) to differences in inheritance patterns of these markers. There is no information on the plastid inheritance pattern for *Pilosocereus*, but the results gathered here, and in a previous work from our group (Bonatelli et al., 2014) points to a stronger structure in plastid compared to the nuclear markers. Therefore, we suspect that the dispersal pattern of the cytoplasmic markers might have a maternal inheritance. Pollination and dispersal in these species are poorly known, but it has been suggested that hummingbirds and bats are important pollen vectors (Zappi, 1994; Rodrigues and Rodrigues, 2014), and that pollen can disperse farther than seeds (Zappi, 1994).

An important aspect of our study is the potential increase in resolution afforded by the anonymous nuclear markers. In species with very recent diversification, like the *P. aurisetus* complex (Bonatelli et al., 2014), such markers allow us to obtain more resolved results in our phylogeographic inferences. In order to

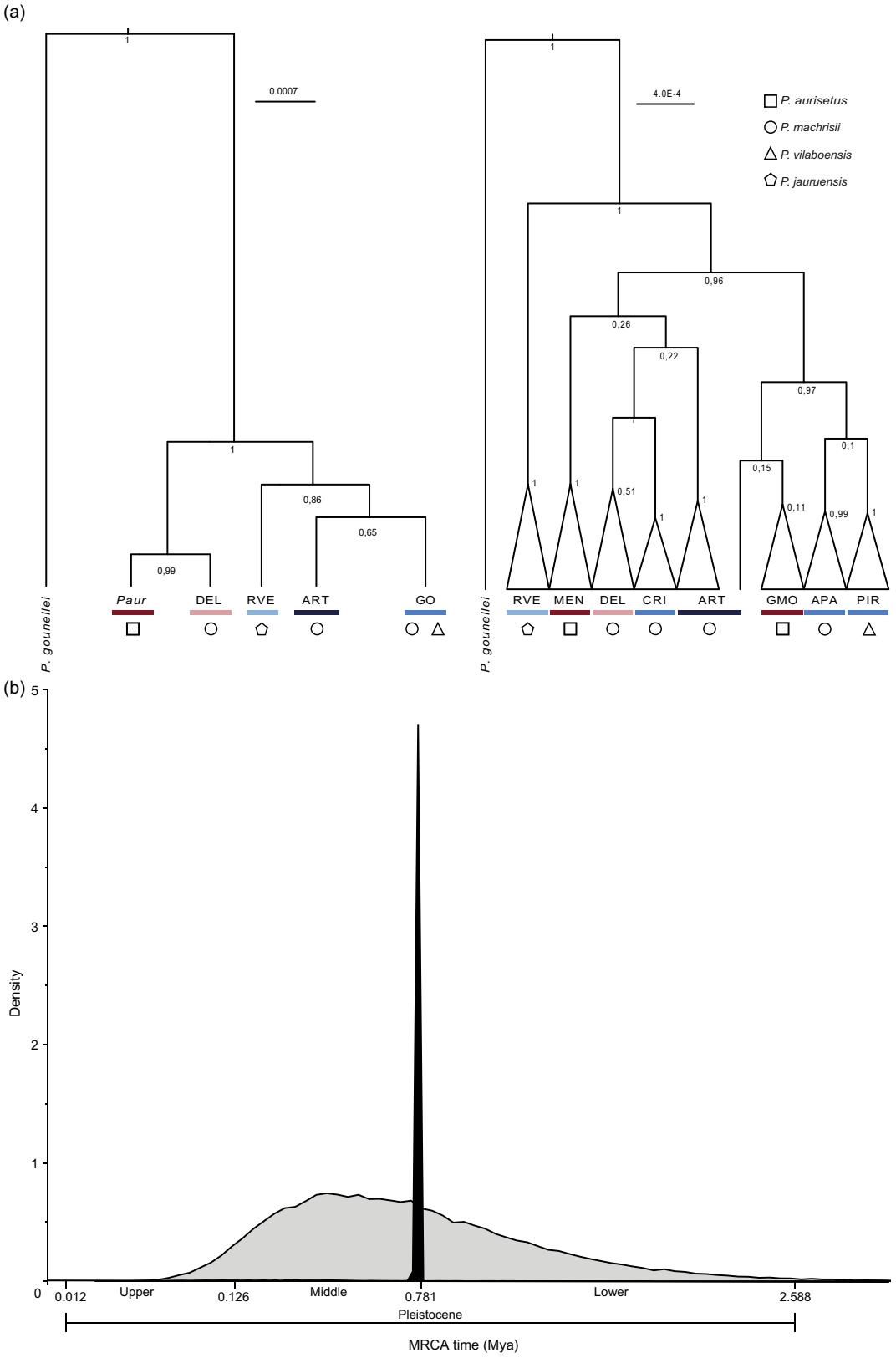

Fig. 4. Comparison of the plastid (partial *trnT-trnL* and *trnS-trnG* data from Bonatelli et al., 2014) and the combined multilocus datasets. (a) Maximum Clade Credibility trees for the species tree analysis in the nuclear data (left) and for a BEAST phylogenetic analysis with a relaxed lognormal clock in the plastid data (right). Posterior probabilities are shown in each branch of the tree, and populations are color coded according to the STRUCTURE results in Fig. 1. (b) Comparison of the divergence time (Mya) between the two main lineages (NW and SE), estimated by setting them as monophyletic and calculating the time to the Most Recent Common Ancestor (TMRCA) using BEAST for the plastid dataset (gray curve) and the combined multilocus dataset, including the plastid data (black curve). Because of the lack of substitution rates for the nuclear markers developed, we estimated relative rates to the plastid marker, by using a prior distribution including the minimum and maximum substitution rates observed in the chloroplast sequences of angiosperms (Wolfe et al., 1987).

explicitly show how the use of multiple markers can improve the resolution of our phylogeographic insights, we contrasted the phylogenetic inferences and divergence time estimates from our dataset and those from cpDNA (Fig. 4). Our results showed different topologies between the phylogenetic trees, with the plastid tree showing a low resolution in several branches, and with the multilocus data showing a more resolved estimate of divergence times for the two main groups obtained, with a much narrower HPD95% (Table S3, Fig. 4). Furthermore, by including a distantly related species in our marker development approach (*P. gounellei*, belonging to a distinct subgenus of *P. aurisetus* species), we believe that the set of anonymous nuclear loci described here can be applied across the *Pilosocereus* genus.

The two deeply diverged lineages observed here were split in the STRUCTURE analysis, recovered in different main clades in the species tree, and showed different climatic requirements for the three periods tested (Figs. 1 and 3). The time estimate for the splitting of these two main clades, calculated with the multilocus dataset, suggested an early Pleistocene (1.69 Mya, HPD 95% = 0.9131–1.766 Mya) divergence between the lineages. While divergence dating should be interpreted with caution, this estimate overlaps with previous estimates based solely on plastid markers (Table S3; Fig. 4), as well as with the date presented by Bonatelli et al. (2014). Taken in total, these results suggest that these two main groups have long been isolated. Similar results have been also observed in codistributed taxa, for example in frogs (Prado et al., 2012), a group of snakes (Machado et al., 2014), and a rocky savanna plant (Collevatti et al., 2009).

Taxonomic boundaries in the *P. aurisetus* complex should be reviewed in light of our results. However, we are opposed to making a formal taxonomic revision based on molecular data alone, without considering additional data sources, such as morphological features. In addition, our results should be treated with caution because we analyzed only a subset of the known taxonomic variation for the *P. aurisetus* complex. A more detailed analysis on the whole spectrum of morphological variation (Schaal et al., 1998) for these species, using other sources of data in an integrative framework (Carstens et al., 2013), is warranted to provide a better taxonomic arrangement in the *P. aurisetus* complex.

Data accessibility

DNA sequences are deposited as GenBank accession numbers KU161695–KU162858.

Sequence alignments are deposited in Perez et al. (in press).

Acknowledgments

This work was supported through grants from the Fundação de Amparo à Pesquisa do Estado de São Paulo – Brazil (FAPESP) (2012/22943-1 to MFP and 2012/22477-0 to EMM) and CAPES – Brazil (BEX: 5638/13-5 to MFP). We thank Isabel Bonatelli and Heidi Utsunomiya for laboratory assistance, Felipe Braga for assistance with Roche platform, and Marcelo Menezes for kindly provide us with cpDNA sequences of *P. gounellei*. We also thank two anonymous reviewers for valuable comments on an earlier version of the manuscript.

References

Andrew, R.L., Bernatchez, L., Bonin, A., Buerkle, C.A., Carstens, B.C., Emerson, B.C., Garant, D., Giraud, T., Kane, N.C., Rogers, S.M., Slate, J., Smith, H., Sork, V.L., Stone, G.N., Vines, T.H., Waits, L., Widmer, A., Rieseberg, L.H., 2013. A road map for molecular ecology. Mol. Ecol. 22, 2605–2626.

Antonelli, A., Verola, C., Parisod, C., Gustafsson, A.L., 2010. Climate cooling promoted the expansion and radiation of a threatened group of South American orchids (Epidendroideae: Laeliinae). Biol. J. Linn. Soc. 100, 597–607.

Arakaki, M., Christin, P.A., Nyffeler, R., Lendel, A., Eggli, U., Ogburn, R.M., Spriggs, E., Moored, M.J., Edwards, E.J., 2011. Contemporaneous and recent radiations of the world's major succulent plant lineages. PNAS USA 108, 8379–8384.

Barbosa, A.R., Fiorini, C.F., Silva-Pereira, V., Mello-Silva, R., Borba, E.L., 2012. Geographical genetic structuring and phenotypic variation in the *Vellozia hirsuta* (Velloziaceae) ochlospecies complex. Am. J. Bot. 99, 1477–1488.

Bonatelli, I.A.S., Perez, M.F., Peterson, A.T., Taylor, N.P., Zappi, D.C., Machado, M.C., Koch, I., Pires, A.H.C., Moraes, E.M., 2014. Interglacial microrefugia and diversification of a cactus species complex: phylogeography and palaeodistributional reconstructions for *Pilosocereus aurisetus* and allies. Mol. Ecol. 23, 3044–3063.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10 (4), e1003537.

Calvente, A., Zappi, D.C., Forest, F., Lohmann, L.G., 2011. Molecular phylogeny of tribe Rhipsalideae (Cactaceae) and taxonomic implications for *Schlumbergera* and *Hatiora*. Mol. Phylogenet. Evol. 58, 456–468.

Carnaval, A.C.O.Q., Moritz, C.M., 2008. Historical climate change predicts current biodiversity patterns in the Brazilian Atlantic rainforest. J. Biogeogr. 35, 1187–1201.

Carstens, B.C., Pelletier, T.A., Reid, N.M., Satler, J.D., 2013. How to fail at species delimitation. Mol. Ecol. 22, 4369–4383.

Collevatti, R.G., Rabelo, S.G., Vieira, R.F., 2009. Phylogeography and disjunct distribution in *Lychnophora ericoides* (Asteraceae), an endangered cerrado shrub species. Ann. Bot. 104, 655–664.

Collevatti, R.G., Terribile, L.C., Lima-Ribeiro, M.S., Nabout, J.C., de Oliveira, G., Rangel, T.F., Rabelo, S.G., Diniz-Filho, J.A.F., 2012. A coupled phylogeographical and species distribution modelling approach recovers the demographical history of a Neotropical seasonally dry forest tree species. Mol. Ecol. 21, 5845–5863.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. JModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9 (8), 772.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.

Eaton, D.A.R., 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics 30 (13), 1844–1849.

Edwards, S.V., 2009. Is a new and general theory of molecular systematics emerging? Evolution 63, 1–19.

Edwards, S.V., Beerli, P., 2000. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution 54 (6), 1839–1854.

Edwards, E.J., Nyffeler, R., Donoghue, M.J., 2005. Basal cactus phylogeny: implications of *Pereskia* paraphyly for the transition to the cactus life form. Am. J. Bot. 92 (7), 1177–1188.

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.

Franck, A.R., Cochrane, B.J., Garey, J.R., 2013. Phylogeny, biogeography, and infrageneric classification of *Harrisia* (Cactaceae). Syst. Bot. 38 (1), 210–223.

Garrick, R.C., Bonatelli, I.A.S., Hyseni, C., Morales-García, A., Pelletier, T.A., Perez, M. F., Rice, E., Satler, J.D., Symula, R.E., Thomé, M.T.C., Carstens, B.C., 2015. The evolution of phylogeographic datasets. Mol. Ecol. 24, 1164–1171.

Gibson, A.C., Nobel, P.S., 1986. The Cactus Primer. Harvard University Press, Cambridge.

Glenn, T.C., 2011. Field guide to next-generation DNA sequencers. Mol. Ecol. Res. 11 (5), 759–769.

Hernández-Hernández, T., Hernández, H.M., De-Nova, J.A., Puente, R., Eguiarte, L.E., Magallón, S., 2011. Phylogenetic relationships and evolution of growth form in Cactaceae (Caryophyllales, Eudicotyledoneae). Am. J. Bot. 98, 44–61.

Hernández-Hernández, T., Brown, J.W., Schlumpberger, B.O., Eguiarte, L.E., Magallón, S., 2014. Beyond aridification: multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. New Phytol. 202, 1382–1397.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Clim. 25, 1965–1978.

Hird, S.M., Brumfield, R.T., Carstens, B.C., 2011. PRGmatic: an efficient pipeline for collating genome-enriched second generation sequencing data using a "provisional-reference genome". Mol. Ecol. Res. 11 (4), 743–748.

Hunt, D., Taylor, N.P., Charles, G., 2006. The New Cactus Lexicon: Atlas & Text. DH Books, Milborne Port.

Joly, S., Hennan, P.B., Lockhart, P.J., 2014. Species radiation by niche shifts in New Zealand's rockcresses (*Pachycladon*, Brassicaceae). Syst. Biol. 63 (2), 192–202.

Knowles, L.L., 2009. Statistical phylogeography. Annu. Rev. Ecol. Evol. Syst. 40, 593–612.

Knowles, L.L., Maddison, W.P., 2002. Statistical phylogeography. Mol. Ecol. 11, 2623–2635.

Lemmon, E.M., Lemmon, A.R., 2013. High-throughput genomic data in systematics and phylogenetics. Annu. Rev. Ecol. Evol. Syst. 44, 99–121.

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

Machado, T., Silva, V.X., Silva, M.J.J., 2014. Phylogenetic relationships within *Bothrops neuwiedi* group (Serpentes, Squamata): geographically highly-structured lineages, evidence of introgressive hybridization and Neogene/Quaternary diversification. Mol. Phylogenet. Evol. 71, 1–14.

Majure, L.C., Puente, R., Griffith, M.P., Judd, W.S., Soltis, P.S., Soltis, D.S., 2012. Phylogeny of *Opuntia* s.s. (Cactaceae): clade delineation, geographic origins, and reticulate evolution. Am. J. Bot. 99, 847–864.

McCormack, J.E., Heled, J., Delaney, K.S., Peterson, A.T., Knowles, L.L., 2011. Calibrating divergence times on species trees versus gene trees: implications for speciation history of *Aphelocoma* jays. Evolution 65, 184–202.

McCormack, J.E., Maley, J.M., Hird, S.M., Derryberry, E.P., Graves, G.R., Brumfield, R. T., 2012. Next-generation sequencing reveals population genetic structure and a species tree for recent bird divergences. Mol. Phylogenet. Evol. 62, 397–406.

McGuire, G., Wright, F., 2000. TOPALi 2.0: improved detection of mosaic sequences within multiple alignments. Bioinformatics 16 (2), 130–134.

Milne, I., Lindner, D., Bayer, M., Husmeier, D., McGuire, G., Marshall, D.F., Wright, F., 2009. TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. Bioinformatics 25 (1), 126–127.

Moraes, E.M., Abreu, A.G., Andrade, S.C.S., Sene, F.M., Solferini, V.N., 2005. Population genetic structure of two columnar cacti with a patchy distribution in eastern Brazil. Genetica 125, 311–323.

Muñoz, M.E.S., Giovanni, R., Siqueira, M.F., Sutton, T., Brewer, P., Pereira, R.S., Canhos, D.A.L., Canhos, V.P., 2011. OpenModeller: a generic approach to species' potential distribution modelling. GeoInformatica 15 (1), 111–135.

Nyffeler, R., Eggli, U., 2010. A farewell to dated ideas and concepts: molecular phylogenetics and a revised suprageneric classification of the family Cactaceae. Schumannia 6, 109–149.

Perez, M.F., Carstens, B.C., Rodrigues, G.L., Moraes, E.M., in press. Anonymous nuclear markers data supporting species tree phylogeny and divergence time estimates in a cactus species complex in South America. Data Brief.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190, 231–259.

Prado, C.P.A., Haddad, C.F.B., Zamudio, K.R., 2012. Cryptic lineages and Pleistocene population expansion in a Brazilian Cerrado frog. Mol. Ecol. 21, 921–941.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

Rambaut, A., Suchard, M.A., Xie, D., Drummond, A.J., 2014. Tracer v1.6. <http://beast.bio.ed.ac.uk/Tracer>.

Ritz, C.M., Martins, L., Mecklenburg, R., Goremykin, V., Hellwig, F.H., 2007. The molecular phylogeny of *Rebutia* (Cactaceae) and its allies demonstrates the influence of paleogeography on the evolution of South American mountain cacti. Am. J. Bot. 94, 1321–1332.

Ritz, C.M., Reiker, J., Charles, G., Hoxey, P., Hunt, D., Lowry, M., Stuppy, W., Taylor, N. P., 2012. Molecular phylogeny and character evolution in terete-stemmed Andean opuntias (Cactaceae-Opuntioideae). Mol. Phylogenet. Evol. 65, 668–681.

Rodrigues, L.C., Rodrigues, M., 2014. Flowers visited by hummingbirds in the open habitats of the southeastern Brazilian mountaintops: species composition and seasonality. Braz. J. Biol. 74 (3), 659–676.

Rosenberg, N.A., 2004. Distruct: a program for the graphical display of population structure. Mol. Ecol. Notes 4, 137–138.

Schaal, B.A., Hayworth, D., Olsen, A.K.M., Rauscher, J.T., Smith, W.A., 1998. Phylogeographic studies in plants problems and prospects. Mol. Ecol. 7, 465–474.

Taylor, N.P., Zappi, D.C., 2004. Cacti of Eastern Brazil. Royal Botanic Gardens, Kew.

Thomson, R.C., Wang, I.J., Johnson, J.R., 2010. Genome-enabled development of DNA markers for ecology, evolution and conservation. Mol. Ecol. 19, 2184–2195.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 11, 4407–4414.

Wolfe, K.H., Li, W.-H., Harp, P.M.S., 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA 84, 9054–9058.

Zappi, D.C., 1994. *Pilosocereus* (Cactaceae). The Genus in Brazil. Royal Botanic Gardens, Kew.

Zellmer, A.J., Hird, S.M., Koopman, M.M., Carstens, B.C., 2012. Deep phylogeographic structure and environmental differentiation in the carnivorous plant *Sarracenia alata*. Syst. Biol. 61 (5), 763–777.