

ORIGINAL ARTICLE

# Model-based analysis supports interglacial refugia over long-dispersal events in the diversification of two South American cactus species

MF Perez<sup>1,3</sup>, IAS Bonatelli<sup>1,3</sup>, EM Moraes<sup>1</sup> and BC Carstens<sup>2</sup>

*Pilosocereus machrisii* and *P. aurisetus* are cactus species within the *P. aurisetus* complex, a group of eight columnar cactus species that are restricted to rocky habitats within the Neotropical savannas of eastern South America. Previous studies have suggested that diversification within this complex was driven by distributional fragmentation, isolation leading to allopatric differentiation, and secondary contact among divergent lineages. These events have been associated with quaternary climatic cycles, leading to the hypothesis that the xerophytic vegetation patches that presently harbor these populations operate as refugia during the current interglacial. However, owing to limitations of the standard phylogeographic approaches used in these studies, this hypothesis was not explicitly tested. Here we use Approximate Bayesian Computation to refine the previous inferences and test the role of different events in the diversification of two species within *P. aurisetus* group. We used molecular data from chloroplast DNA and simple sequence repeats loci of *P. machrisii* and *P. aurisetus*, the two species with broadest distribution in the complex, in order to test if the diversification in each species was driven mostly by vicariance or by long-dispersal events. We found that both species were affected primarily by vicariance, with a refuge model as the most likely scenario for *P. aurisetus* and a soft vicariance scenario most probable for *P. machrisii*. These results emphasize the importance of distributional fragmentation in these species, and add support to the hypothesis of long-term isolation in interglacial refugia previously proposed for the *P. aurisetus* species complex diversification. *Heredity* (2016) **00**, 1–8. doi:10.1038/hdy.2016.17

## INTRODUCTION

The investigation into the effect of historical climatic changes on the distribution and population dynamics of species has been a major focus of phylogeography (Hewitt, 1996; Avise *et al.*, 1998; Taberlet *et al.*, 1998; Soltis *et al.*, 2006; Garzón-Orduña *et al.*, 2014). Most of these studies focused in the Northern Hemisphere (Beheregaray, 2008), and tried to identify putative species refugia by locating regions of the species' range that contain high levels of genetic diversity and were also ice-free during glaciation (Keppel *et al.*, 2012). The paucity of data for South American species, particularly for those adapted to open and/or dry habitat types, has greatly hindered the understanding about diversification mechanisms in the continent (Hughes *et al.*, 2013; Turchetto-Zolet *et al.*, 2013). Despite the controversy about Haffer's (1969) forest refuge hypothesis (for example, Hoorn *et al.*, 2010; Rull 2011), many species associated with mesic vegetation display similar patterns (Turchetto-Zolet *et al.*, 2013; Garzón-Orduña *et al.*, 2014) that are consistent with the tropical refugial hypothesis. This is not true for species occurring in xeric and open environments as the rocky savanna habitats in Eastern South America, which show more idiosyncratic patterns (Turchetto-Zolet *et al.*, 2013). For instance, distinct responses have been found in species distribution ranges during the same paleoclimatic phases such as signals of expansion in the *Drosophila*

*buzatti* cluster, which exclusively use decaying stems of cactus as breeding sites (Moraes *et al.*, 2009), shrink in a tree species of Cerrado (savanna) biome, *Dipteryx alata* Vogel (Collevatti *et al.*, 2013) and stability in the Neotropical coastal species of orchid *Epidendrum fulgens* (Pinheiro *et al.*, 2011), and in the gecko *Phyllopezus pollicaris*, a specialist species of rocky outcrop habitats (Werneck *et al.*, 2012).

The majority of analyses in phylogeographic studies to date have been limited to *post hoc* inferences derived from summaries of genetic variation (for example,  $F_{ST}$ ,  $\theta_w$ ) or parameters estimated under specified models (for example,  $\theta = 4N_e\mu$ ,  $\tau$  and  $m$  under an isolation with migration model). Phylogeographic inferences that are inherently qualitative are a limitation in complex systems where multiple historical processes operating on different temporal and geographic scales can produce the same general phylogeographical pattern (Riddle and Hafner, 2006). To avoid the qualitative inferences, researchers have turned to explicit tests of phylogeographic models, either in a frequentist (for example, Knowles, 2001), Bayesian (for example, Fagundes *et al.*, 2007), or information-theoretic (for example, Carstens *et al.*, 2009) framework. Model-based analysis such as Approximate Bayesian Computation (ABC) can refine the findings of classic phylogeographic methods, allowing a quantitative evaluation of the

<sup>1</sup>Departamento de Biologia, Universidade Federal de São Carlos, Sorocaba, São Paulo, Brazil and <sup>2</sup>Department of Evolution, Ecology and Organismal Biology, The Ohio State University, Columbus, OH, USA

Correspondence: MF Perez or IAS Bonatelli, Department of Biology, Federal University of Sao Carlos, Rod Joao Leme dos Santos km 110, Sorocaba, 18052780 São Paulo, Brazil. E-mail: manolofperez@gmail.com or belbonatelli@gmail.com

<sup>3</sup>These authors contributed equally to this work.

Received 16 August 2015; revised 26 January 2016; accepted 2 February 2016

demographic history by contrasting demographic models defined a priori and estimating relevant parameters even in complex systems.

Traditionally, the main question in many biogeographic and phylogeographic studies has been the dualism of vicariance versus dispersal (for example, Ronquist, 1997; Nason *et al.*, 2002; Yoder and Nowak, 2006). Empirical studies have increasingly applied ABC methods to define the role of these different events in the evolution of the target species, and estimate their effects in the genetic diversity and demographic parameters even in complex scenarios such as those encountered in comparative phylogeographic studies (Becquet and Przeworski, 2007; Palero *et al.*, 2009). Hickerson and Meyer (2008) have dealt with this situation in a study with marine communities of cowrie gastropod in the Marquesas and Hawaiian islands of the Indo-Pacific region. The implementation of ABC analyses have allowed the distinction among models involving variance and dispersal processes, resulting in evidences of isolation by colonization events in co-distributed cowrie gastropods of the Marquesas and evidences of isolation by vicariant events in a subset of taxa in the Hawaiian islands. Other empirical studies demonstrated the benefits of an ABC approach to comparative systems by evaluating the same demographic models to show that co-distributed species responded differently to shared climatic events (for example, Espindola *et al.*, 2014). In either case, the use of customized phylogeographic models enabled system-specific hypotheses to be tested.

*Pilosocereus* (Cactaceae) is comprised of 41 recognized species and is subdivided into five informal taxonomic groups based on morphological and geographical clusters (Zappi, 1994). *Pilosocereus machrisii* and *P. aurisetus* belong to a species complex containing eight cactus species (*P. aurisetus*, *P. machrisii*, *P. vilaboensis*, *P. jauruensis*, *P. aureispinus*, *P. parvus*, *P. bohlei*, *P. pusilibaccatus*) grouped by morphological characteristics (Zappi, 1994; Taylor and Zappi, 2004; Hunt *et al.*, 2006). Morphologically, this group is defined by the shrubby habit of the species, ground level branching and differentiated flower-bearing areolas that produce abundant bristles and sometimes wool. Flowers are usually pink or red and the fruits are white-pulped in most cases. Within the group, the species are differentiated mostly by their size, color of the epiderm, number and depth of the ribs in the cladodium, disposition and color of the spines, as well as color and format of the seeds. In some cases the habitat is also used to delimit species, with some species occurring only at specific environments. For example, *P. aurisetus* only occurs above quartzitic rock outcrops, whereas *P. machrisii* can occur on quartzitic, arenitic and limestone outcrops. Pollination and dispersal are poorly known, but moths, hummingbirds and bats are indicated as putative pollinators, and seeds are dispersed by bats, birds and lizards of Cerrado. Zappi (1994) suggested that pollen can be dispersed farther than seeds in *Pilosocereus* species, a suggestion that follows the observation of low levels of shared chloroplast DNA (cpDNA) haplotypes among proximate populations (Bonatelli *et al.*, 2014). The species of the *P. aurisetus* complex display a disjunct distribution, ecologically restricted to drought-stressed conditions, characteristic of somewhat open or quite bare rock outcrops (usually non-calcareous), associated with campo rupestre vegetation patches in eastern Brazil. The patchy distribution of their habitat can be compared with oceanic islands, and make the species of *P. aurisetus* complex a suitable system to investigate the evolutionary dynamics of interglacial refugia.

*P. machrisii* (EY Dawson) Backeb. is the most broadly distributed species in the complex, occurring from northeastern to southeastern Brazil. *Pilosocereus aurisetus* (Werderm.) Byles & GD Rowley has the second largest distribution, occurring along the southern Espinhaço Mountain Range in southeastern Brazil (Figure 1), and consists of the

subspecies *P. aurisetus aurisetus* and the microendemic subspecies *P. aurisetus aurilanatus*, which has only one known population that is restricted to an isolated mountain in the western portion of the Espinhaço Range. Previous findings based on the variation of noncoding regions of cpDNA and nuclear microsatellite markers (simple sequence repeats, SSR), coupled with palaeodistributional estimates, suggested that the diversification of these species occurred during early to middle Pleistocene, when a broad ancestral distribution was fragmented, leading to isolation and allopatric differentiation. Based on these findings, Bonatelli *et al.* (2014) hypothesized that diversification in the *P. aurisetus* complex was caused by the formation of xeric microrefugia during the interglacial phases of the quaternary climatic cycles. Although these results strongly support an allopatric scenario of diversification, equating present-day isolation with refugial diversification is difficult because patchy distributions may be connected by episodic long-distance dispersal (Stewart *et al.*, 2010).

Here, we apply model-based approaches using a data set consisting of two types of molecular marker (cpDNA and SSR) that provide insight into the relative influence of ancient and recent processes impacting the diversification of *P. aurisetus* and *P. machrisii*. This approach allows us to refine our understanding of the phylogeographic history in these species in a manner that accounts for the uncertainty in parameter estimation as well as the models used to estimate these parameters. We elaborate and test models, representing putative divergence scenarios, to investigate the relative role of vicariance and dispersal in the evolutionary history of these two species, and demonstrate how model-based analyses complement more traditional means of phylogeographic inference.

## MATERIALS AND METHODS

### Population sampling and genetic data sets

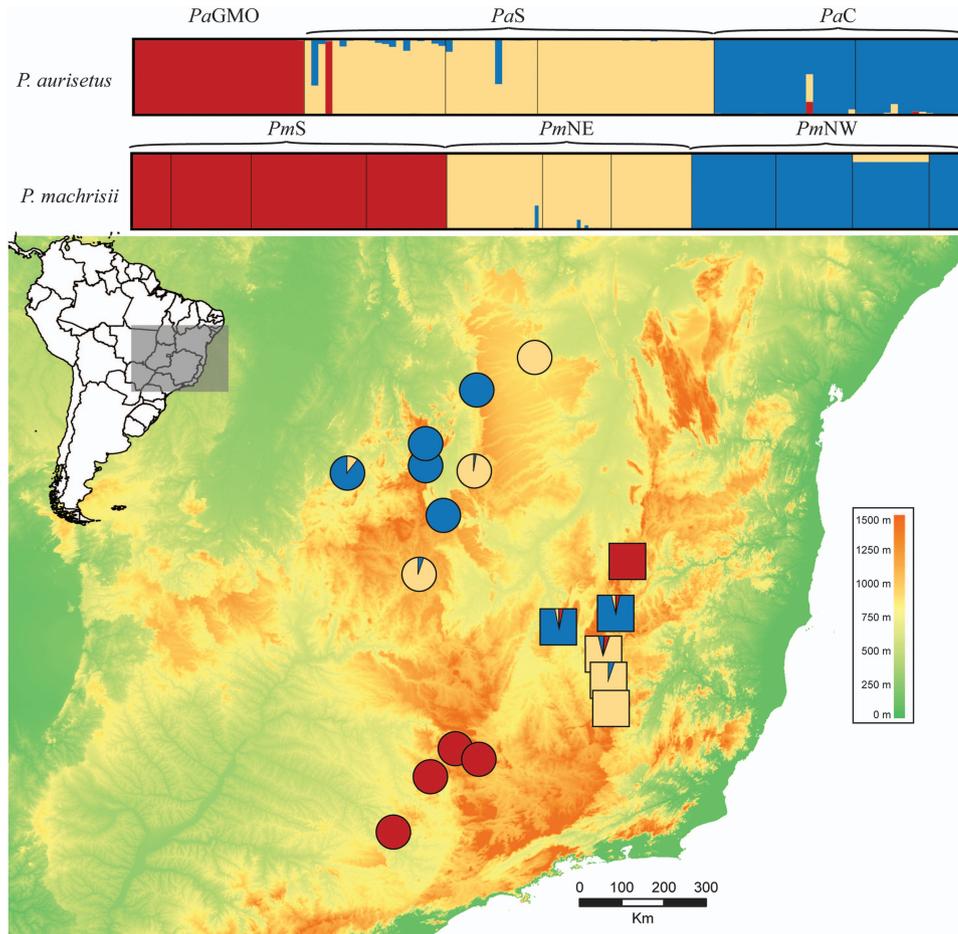
We used a partial SSR and cpDNA data set previously generated by Bonatelli *et al.* (2014). The present data set consists of 10 markers in *P. aurisetus* in a sample of 117 individuals from six sampling localities, and 11 SSR markers in *P. machrisii* for 241 individuals from 11 sampling localities. Two cpDNA intergenic spacers, *trnT-trnL* and *trnS-trnG*, were used for the same populations in 61 individuals for *P. machrisii* and 30 for *P. aurisetus*. Our sample covers all of the known localities described for these species (Figure 1).

### Population structure

In order to estimate the number of interbreeding groups ( $K$ ) in the SSR data set, we implemented a Bayesian analysis with STRUCTURE 2.3.4 (Pritchard *et al.*, 2000) without prior information on the sampling sites for each sample. Ten independent runs for each  $K$  from 1 to 7 in *P. aurisetus* and from 1 to 14 in *P. machrisii* (the total number of sampled populations+1) were performed. For each simulation, 1 000 000 interactions were carried on, discarding the first 500 000 (burn-in). The admixture model with correlated allele frequencies was used. For the graphical representation of STRUCTURE results we used DISTRUCT v1.1 (Rosenberg, 2004). 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)  $\Delta K$  statistics (Evanno *et al.*, 2005); (4) absence of 'virtual' groupings, that is, groups containing only individuals with genomes splitted in more than one cluster.

### Model definition

Based on the results from the STRUCTURE analysis, we developed specific models to test the performance of concurrent demographic history in *P. aurisetus* and *P. machrisii*. The demographic models defined were designed to test specifically if the diversification within each species was primarily driven either by vicariance or by long-dispersal events. It is important to highlight here that the subspecies level was not considered in the model design, as we only consider the genetic clusters recovered in STRUCTURE. Specifically, in order to



**Figure 1** (a) STRUCTURE results with each individual represented as a vertical bar showing the proportion of its genome assigned to each cluster. Black lines separate individuals of different sampled localities. (b) Elevational map showing the geographic distribution of *P. machrisii* populations (circles) in the southern (*PmS*), northeastern (*PmNE*) and northwestern (*PmNW*) and *P. aurisetus* populations (squares) for the GMO, southern (*PaS*) and center (*PaC*) distribution of the species analyzed in this study. Lineages are shown with different colors defined by the STRUCTURE analysis.

test the importance of vicariance, we included three models simulating different forms for vicariant divergence. The simulated models include a simple vicariance (model 1) with populations retaining their sizes after the split; vicariance within refuges (model 2), with size reductions that accompanied the range fragmentation; and soft vicariance (model 3), characterized by drastic reduction of gene flow (according to the model proposed by Hickerson and Meyer (2008)). To test long-distance dispersal events, we designed four models that consisted in one population serving as source and serial founder effects in the other population, with strong size reduction followed by exponential growth (models 4–7). We tested all possibilities of founder and being founded populations; these combinations are shown in Figure 2.

### Simulation data and summary statistics

The program *ms* (Hudson, 2002) was used to simulate genetic data under the defined demographic models for the *P. aurisetus* and *P. machrisii* systems. A Python script (available at: <http://dx.doi.org/10.5061/dryad.8h6k1>) was used to draw values for the parameters related to divergence times ( $\tau_1$  and  $\tau_2$ ) and theta in the ancient population ( $\theta$ ) in all models; we estimated contractions in the contemporary populations ( $\theta_r$ ; calculated as a ratio of the  $\theta$  in the ancient population) in the refuge model; we also sampled values for the ancient ( $M_a$ ) and current ( $M_c$ ) bidirectional migration rates between population pairs for the soft vicariance model; and the ratio of the founded and the current population sizes ( $\theta_r F-C$ ), which measures how much the population increased since the founder event, in the long-dispersal models.

The obtained values were used to simulate genealogies under each model (Figure 2). Prior distributions consisted of 200 000 simulated data sets for each demographic scenario in each species. The shape of the parameter distributions were based on a broad range of values, defined to cover biologically conceivable values for this species system (Supplementary Methods), considering a mutation rate of  $8.87 \times 10^{-4}$  per generation (Marriage *et al.*, 2009) for the SSR markers and a generation time of 15 years (a mean generation time observed in plants of different *Pilosocereus* species grown in greenhouse; Gerardus Olsthoorn pers. comm.). All parameters drawn from each model are depicted in Figure 2.

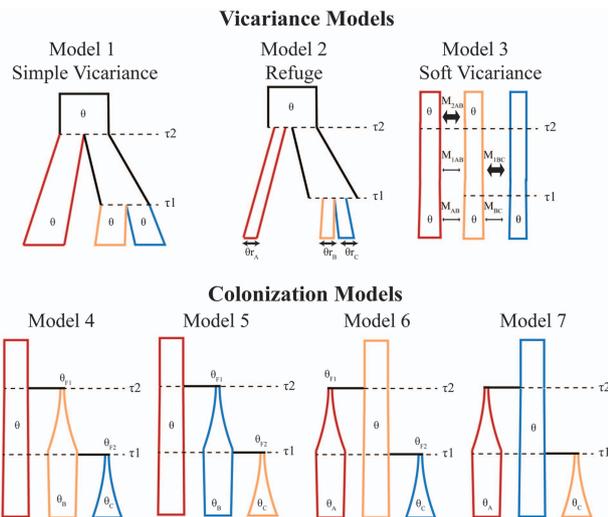
Summary statistics (SuSt) from simulated cpDNA sequence data (proportion of polymorphic sites— $\pi$ , number of segregating sites— $S$ , Tajima's  $D$ , Fay and Wu's  $\theta_H$  and  $\pi-H$ , proportion of polymorphic sites within each population— $\pi_w$ , and between populations— $\pi_B$ ) were calculated using a custom PERL script written by N Takebayashi (available at: <http://raven.iab.alaska.edu/~ntakebay/teaching/programming/coalsim/scripts/msSS.pl>). The same SuSt were estimated from the empirical cpDNA data (Supplementary Table S1). To calculate SuSt for SSR markers, the simulated data were converted to alleles differing in size using the software 'microsat' (Cox, 2011), and then converted to the Arlequin format using a Python script (microsat2arp.py; available at: <http://dx.doi.org/10.5061/dryad.8h6k1>). The software 'arlsuostat' (Excoffier and Lischer, 2010) was used to calculate the number of alleles ( $A$ ), expected heterozygosity ( $H_e$ ) and the modified version of Garza & Williamson's  $M$  (2001; Excoffier *et al.*, 2005) for each population ( $mM$ ), as well as the total and pairwise  $F_{ST}$  between pairs of populations for

both the simulated and the empirical data (Supplementary Table S1). This SuSt data set was selected with basis on its informativeness in relation the priors used, verified with linear regression analyzes (data not shown).

### ABC in the *pilosocereus* data set

To select the more appropriate strategy to perform ABC on our data set, we compared the performance of the ABC analysis using the original data and this same data transformed by Principal Component Analysis, an approach similar to the suggested by Bazin *et al.* (2010). We also tested different algorithms for the model selection step (Supplementary Methods). After choosing the best method to perform the ABC analysis, the empirical SuSt were used to perform the ABC model selection in both species, using the R package abc version 1.4 (Csilléry *et al.*, 2012, <http://cran.r-project.org/web/packages/abc/index.html>) with a threshold level of 0.005, resulting in 7000 simulations retained in the posterior. We considered both the posterior probabilities and the relative Bayes Factor estimates to define the best models for each species.

We also performed a parameter estimation step, conditional on posterior probabilities across all models, for parameters that were represented in all models ( $\tau_1$ ,  $\tau_2$  and  $\theta$ ). Estimation of parameters specific to the model with the highest support was performed using only the simulations from that model. For the parameter estimation step we used only the neural networks approach, as this method outperformed the other methods available in the abc package in a simulation study (Blum and François, 2010). The parameters were logit transformed using the prior boundaries to assure that the estimates are inside these intervals. We used posterior predictive checks to assess the performance of this step (Supplementary Methods).



**Figure 2** Simulated scenarios of *P. machrisii* and *P. aurisetus* species divergence. Model 1—a simple vicariance with populations retaining their sizes after the separation; model 2—vicariance within refuges, with size reductions that accompanied the range fragmentation; model 3—soft vicariance, characterized by drastic reduction of gene flow. Models 4–7—all combinations of long-distance dispersal events, with one population serving as source and serial founder effects in the other populations. Symbols represent the estimated parameters of divergence times ( $\tau_1$  and  $\tau_2$ ) and theta in the ancient population ( $\theta$ ) from all models. Parameters specific to each model are contractions in the contemporary populations, calculated as a ratio of the  $\theta$  in the ancient population ( $\theta_r$ ), represented in the refuge model. Bidirectional ancient ( $M_a$ ) and current ( $M_c$ ) migration rates between population pairs, in the soft vicariance model. The ratio of the founded and the current population sizes ( $\theta_{rF-C}$ ), which measures how much the population increased since the founder event, in the long-dispersal models.

## RESULTS

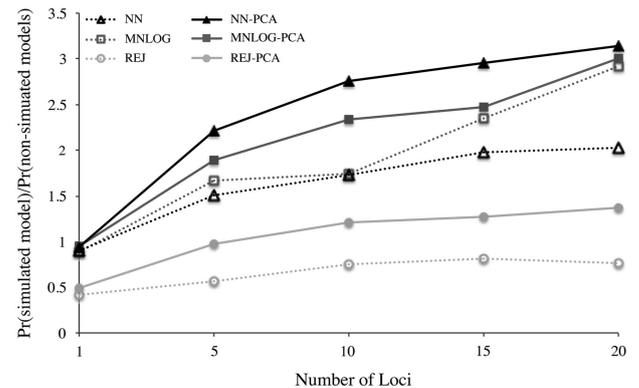
### Population structure

Results obtained in STRUCTURE suggested a main population structure in three groups ( $K=3$ ) for both species (Figure 1). In *P. aurisetus* the three recovered clusters consisted in the locality GMO, two localities in the center (*PaC*) and three in the south (*PaS*) of the species distribution. *P. machrisii* clusters were formed by four localities in the south (*PmS*), three in northeast (*PmNE*) and four in the northwest (*PmNW*) of the distribution of this species.

### ABC in the *pilosocereus* data set

The results obtained from the ABC simulation testing (Supplementary Methods) suggested that the regular rejection method showed the worst performance in choosing the simulated model, with higher probabilities of choosing non-simulated models (values lower than 1) in all tests, even when 20 loci were used. We observed that transforming the SuSt with Principal Component Analysis resulted in more accurate results compared with the original data set for the three methods, and that the neural networks method with the Principal Component Analysis correction had the best performance with the number of loci used here (10 for *P. aurisetus* and 11 for *P. machrisii*; Figure 3), showing rates higher than 2 when more than one locus was used; these values indicate a probability of choosing the generating model at least two times higher than the probability of choosing a non-simulated model. Therefore, we used this method for the remaining analyses.

The most likely scenario for *P. aurisetus* indicates that all populations experienced vicariant events with reduction in population size (Model 2) with a high posterior probability (PP=0.9993; Table 1). This result suggests that fragmentation events of a widespread ancestral population were much more likely than long-distance dispersal in this species. Vicariance was also supported in *P. machrisii*, but in this case the soft vicariance scenario (model 3), in which the three populations were connected by high levels of gene flow that decreases drastically in times  $\tau_2$  and  $\tau_1$ , was recovered with higher probability (PP=0.8; Table 1). Bayes Factor estimates confirm the inference power of the most likely scenarios described above, with the



**Figure 3** Cross-validation test to compare the performance of using different rejection algorithms and number of loci for the model selection step. Each point represent the average ratio of choosing the simulated model over choosing a non-simulated model for all the seven models tested. Therefore, values higher than 1.0 represent that the simulated model was recovered more often than non-simulated models. Abbreviations refer to each rejection model, as follows: REJ—regular rejection method, MNLOG—multinomial logistic regression, NN—neural networks. Dashed lines represent tests with the raw data, and solid lines refer to the data set transformed by PCA.

**Table 1** List of tested models in *P. machrisii* and *P. aurisetus* and their respective PP and BF compared with the best alternative model according to Jeffreys (1961)

Model	<i>P. aurisetus</i>			<i>P. machrisii</i>		
	Description	PP	BF	Description	PP	BF
1	Simple Vicariance	0	Negative	Simple Vicariance	0.0537	Negative
2	Refuge	<b>0.9997</b>	<b>Very strong</b>	Refuge	0.002	Negative
3	Soft Vicariance	0	Negative	Soft Vicariance	<b>0.8019</b>	<b>Strong</b>
4	Colonization from GMO to PaC	0	Negative	Colonization from PmS to PmNW	0.0604	Negative
5	Colonization from PaGMO to PaS	0	Negative	Colonization from PmS to PmNE	0.001	Negative
6	Colonization from PaS	0	Negative	Colonization from PmNE	0.0218	Negative
7	Colonization from PaC	0.0003	Negative	Colonization from PmNW	0.0592	Negative

Abbreviations: BF, Bayes Factor; PP, posterior probabilities. Values in bold indicates the best model for each species. Abbreviations refer to the STRUCTURE groups in Figure 1, and representation of the models is showed in Figure 2.

preferred models showing very high values compared with the other models (Supplementary Table S2). No model related to colonization was recovered with a posterior probability higher than 0.06 (Table 1) for either species. Furthermore, the sum of the posterior probabilities of all models involving colonization did not exceed 0.14 in both cases, showing that long-dispersal events might not have an important role in the evolutionary history of these species.

Parameter estimations of effective population size for the current populations are several times smaller than those of the ancestral population in *P. aurisetus*, principally in GMO with the most severe bottleneck estimates (current size is 0.03–0.05 of the original population). Divergence times for the most recent lineages ( $\tau_1$ ) were centered in the early to middle Pleistocene for *P. aurisetus* (Median = 0.739 Mya, 95 %HPD = 0.079–1.056) and for *P. machrisii* (Median = 0.619 Mya, 95% HPD = 0.093–1.893). The most ancient separation between lineages ( $\tau_2$ ) within both species was estimated to take place mostly in the early Pleistocene (*P. aurisetus* Median: 0.938 Mya, 95% HPD = 0.370–1.7187; *P. machrisii* Median: 1.184 Mya, 95% HPD 0.373–2.539). The migration parameters estimated for *P. machrisii* showed that current migration rates between PmS and PmNE were very low (Median: 0.0062, 95% HPD = 0.0005–0.0626) and much smaller than between PmS and PmNW, and between PmNE and PmNW (Table 2). The ancient migration estimates ( $M_a$ ) showed low informativeness, as the posterior distribution of the parameters was almost the same as the 95% HPD (Table 2). Plotting the simulations recovered in the posterior of the model selection step revealed that the observed data set is clearly surrounded by the simulated data sets when the SuSt were transformed by Principal Component Analysis, indicating that this strategy outperformed the use of untransformed SuSt with our data (Figures 3 and 4). The results from the posterior predictive checks also suggested a good fit from our simulations to the empirical data, as almost all SuSt rendered simulated data sets containing the empirical values (Supplementary Table S1). The only exceptions were the  $F_{ST}$ s in *P. aurisetus*, and  $\theta_H$ , number of alleles in PmNE and PmNW, total  $F_{ST}$  and mM (Garza and Williamson, 2001) in all populations for *P. machrisii*.

## DISCUSSION

Independent evidences (for example, Prado and Gibbs, 1993; Ledru *et al.*, 1996; Pennington *et al.*, 2004; Collevatti *et al.*, 2009; 2012) have been considered to explain the shifts in vegetation dynamics and species diversification due to quaternary climate oscillations in eastern South America. Although there is no consensus about the extent to which the Pleistocenic climatic fluctuations affected the species

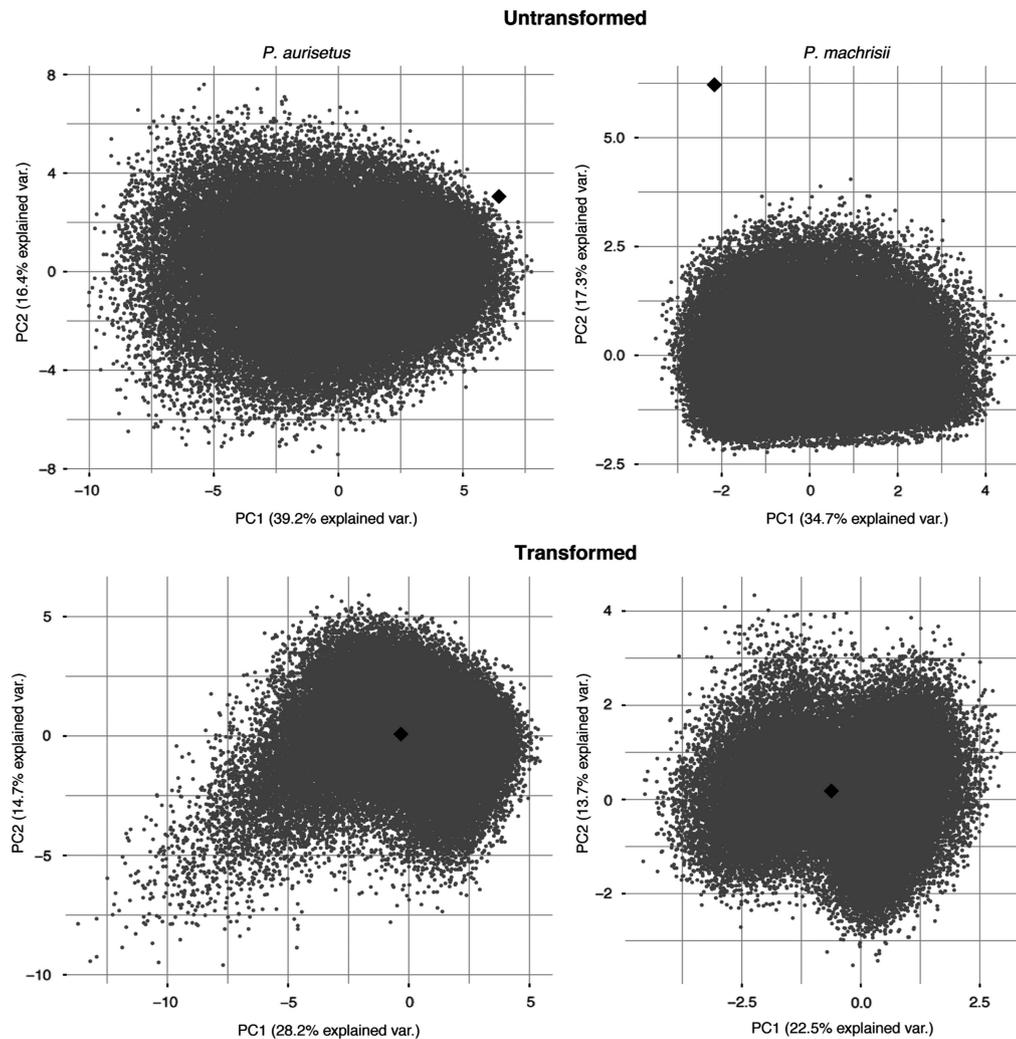
**Table 2** Demographic parameters estimated under neural network regression for *P. machrisii* and *P. aurisetus*

Species	Parameters	Median	95% CI
<i>P. aurisetus</i>	<b><math>N_e</math></b>	3171.93	1848.78–3768.00
	<b><math>\tau_1</math></b>	0.7304	0.0792–1.0599
	<b><math>\tau_2</math></b>	0.9377	0.3697–1.7187
	$\theta_{rPaGMO}$	0.0148	0.0036–0.0475
	$\theta_{rPaS}$	0.1817	0.0525–0.4388
	$\theta_{rPaC}$	0.4356	0.1767–0.7116
<i>P. machrisii</i>	<b><math>N_e</math></b>	2800.90	1663.15–3773.09
	<b><math>\tau_1</math></b>	0.6189	0.0927–1.8929
	<b><math>\tau_2</math></b>	1.1839	0.3727–2.5389
	$M_c PmS <-> PmNE$	0.0062	0.0005–0.0626
	$M_c PmS <-> PmNW$	0.0759	0.0068–0.5189
	$M_c PmNE <-> PmNW$	0.0490	0.0073–0.3495
	$M_a PmS <-> PmNE$	49.38	3.55–96.98
	$M_a PmS <-> PmNW$	49.13	4.26–96.53
$M_a PmNE <-> PmNW$	48.43	3.28–97.59	

Abbreviation: CI, confidence interval.  $N_e$  in units of individuals, divergence times ( $\tau_1$  and  $\tau_2$ ) in Mya,  $M_c$  and  $M_a$  (current and ancestral migration rates) in number of alleles per generation, and  $\theta_r$  is the ratio of the ancestral population size in the current population. Parameters estimated from simulations from all models are shown in bold. Abbreviations refer to the STRUCTURE groups showed in Figure 1.

distribution and evolution of different groups of organisms (Hewitt, 2004; Werneck *et al.*, 2012; Turchetto-Zolet *et al.*, 2013), it seems clear that the response of each species to climatic fluctuations is related to the dispersal abilities and ecological niches of the particular species (Bonatelli *et al.*, 2014). In the *P. aurisetus* species group, the climate oscillations appear to have promoted successive events of distributional fragmentation and isolation leading to allopatric differentiation, which could be related, with the formation of quaternary interglacial microrefugia in the species from the *P. aurisetus* complex (Bonatelli *et al.*, 2014). The disjunct distribution of plant species in rocky environments within Cerrado was also investigated by Collevatti *et al.*, 2009, 2012 using the species *Lychnophora ericoides* and *Tibouchina papyrus*. The extant phylogeographic pattern of both species is likely shaped by vicariant processes that took place during warmer and moister conditions of the interglacial periods, whereas in the Pleistocene glacial periods the species have occupied a broader distribution.

The results of our ABC analysis indicate that models containing vicariant fragmentation have a higher support than those involving



**Figure 4** Plot of the first two Principal Component axes for the simulations recovered in the posterior of the model selection step. Each simulation is represented by a gray point, with the empirical data represented as a black diamond. The two plots above represent the untransformed data set for *P. aurisetus* and *P. machrisii*, and the two plots below represent the data sets transformed by PCA.

dispersal in the diversification history of two species of the group. This result is consistent with previous genetic analyses, as well as with the results of paleoecological niche modeling, which show that suitable habitat was patchy during glacial maxima (Bonatelli *et al.*, 2014). Therefore, diversification in the patchily distributed *P. machrisii* and *P. aurisetus* complex seem to be largely driven by long-term isolation and genetic drift, rather than dispersal/colonization dynamics. The low ability to seed dispersal could also have increased the level of isolation among populations, allowing the appearance of new species and differentiated lineages.

The diversification time estimates that result from the analysis of the microsatellite data indicate a very recent diversification, an expected feature in Cactaceae (Arakaki *et al.*, 2011; Hernández-Hernández *et al.*, 2011). Importantly, these estimates were similar to those obtained from two cpDNA and one nuclear gene (Bonatelli *et al.*, 2014), supporting that the splitting times for the most ancient lineages in the two species took place primarily during the early Pleistocene whereas the derived lineages diversified in a period within the early and middle Pleistocene. These estimates fell within a period characterized by range shifts in dry vegetation (Prado and Gibbs, 1993; Pennington *et al.*,

2004) and also recognized as an age of speciation and intraspecific differentiation of other species associated with the open and xeric environments in eastern Brazil, such as the rodent species of the genus *Calomys* (Almeida *et al.*, 2007) and some cactophilic drosophilids of the *D. buzatti* cluster (Moraes *et al.*, 2009; Franco and Manfrin, 2013).

The estimates of effective population sizes in both species and the ancestral and current migration rates in *P. machrisii* exhibit 95% confidence intervals that are similar to the prior distributions, which suggests that the data lack informativeness to estimate these parameters. Conversely, the bottleneck parameter in *P. aurisetus* indicate that the population reduction was more severe in the northern range of *P. aurisetus*, with the current GMO population size up to 4% of the ancestral population size, according to the confidence interval 95% (0.0036–0.0475). The populations in the center of *P. aurisetus* distribution had a moderate size reduction (confidence interval 95% 0.0525–0.4388) and the populations in the south of the distribution had the less severe reduction (confidence interval 95% 0.1767–0.7116), a result in agreement with the fact that the southern part of the distribution of this taxon occurs in a more continuous habitat, in the core of the Espinhaço Mountain Range.

As demonstrated here, model-based methods can enhance traditional phylogeographic inferences by allowing researchers to quantify the statistical support for scenarios implied by SuSt or parameter estimates (Beaumont *et al.*, 2010; Csilléry *et al.*, 2010). We presented results obtained using an ABC framework applied with several methods to validate the results obtained in each step, as suggested by Sunnåker *et al.*, 2013. Specifically, we implemented a cross-validation method to select the best approach to perform model selection with our data set. We also compared the performance of each approach with different number of loci, and observed that the number of loci presented here showed a high accuracy in determining the model with high support (both the MNLOG and NN methods showed a ratio of choosing the simulated model over choosing a non-simulated model ('false model') higher than 1), and also that adding more loci than what we showed here does not present a great improvement in the performance (Figure 3). Although the amount of genetic information and the loci type are more important in increasing the resolution of genetic analysis than the number of loci itself, especially in analysis based on coalescence (Corl and Ellegren, 2013), when we increase the number of loci in our analysis we were also increasing the genetic variability, once all simulated loci had the same genetic information content. Another result that supports our results was that we were able to recover the empirical value for 20 over 24 SuSt in *P. aurisetus* and for 17 over 24 SuSt for *P. machrisii* in a Posterior Predictive Check analysis. It is noteworthy that some of the SuSt that were not well recovered in this step were not directly related to the model used to simulate the data sets for the PPC analysis. For example, the SuSt not recovered in *P. aurisetus* were all calculated as  $F_{ST}$ s that are usually related to migration rates, a parameter that was absent in the model used to simulate data sets for this species. In addition, three of the SuSt not recovered for *P. machrisii* (mM for each population) are informative for size fluctuation in the populations, a parameter that was also absent in the model used to simulate data in this species.

By using the methods with best performance in the cross-validation step, we calculated the relative posterior probability and compared the Bayes Factors of several demographic models that were derived to test if the present-day fragmented distribution of two cactus species was achieved by vicariance or long-distance dispersal. Essentially, this analysis provides a quantitative evaluation of competing historical scenarios for these species, and thus an effective extension of the qualitative analysis conducted earlier in the *P. aurisetus* complex (Bonatelli *et al.*, 2014).

Although rocky savanna habitats in central and eastern Brazil covers <1% of the Brazil's continental surface, they harbor a remarkable floristic diversity and endemism, with ~14% of the entire Brazilian vascular flora (Silveira *et al.*, 2015). Phylogeographical studies on plant and animal species associated to these habitats have pointed to the Pleistocene climatic fluctuations and their demographic consequences as an important diversification mechanism accounting to the high diversity and endemism of rocky savanna habitats in eastern South America (Moraes *et al.*, 2009; Barbosa *et al.*, 2012; Franco and Manfrin, 2013; Collevatti *et al.*, 2009, 2013; Bonatelli *et al.*, 2014; Machado *et al.*, 2014). These studies have provided growing support to the hypothesis that rocky savannas in Eastern South America are interglacial refuge for xerophytic species. However, this hypothesis has not been thoroughly tested in a model-based statistical framework as ABC. Thus, initiatives such as the one presented here might be very useful to compare competing models of diversification in dry vegetation communities of eastern South America.

## DATA ARCHIVING

DNA sequences were deposited in GenBank under accession nos: JN035381-JN035385; JN035388-JN035402, JN035403-JN035405; JN035420-JN035430; JN035437; JN035441-JN035445; JN035449-JN035451; JN035570-JN035574; JN035576-JN035580; JN035585; JN035590-JN035594; JN035600-JN035604; KC621129-KC621152; KC621159-KC621163; KC621184-KC621208; KC621213-KC621217; KC621228-KC621237; KC621242; KC621243; KC621245-KC621260; KC779260-KC779263; KC779314-KC779335; KC779341-KC779345; KC779366-KC779390; KC779395-KC779399; KC779410-KC779437. Microsatellite data and scripts available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.8h6k1>.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

This work was supported through grants from the São Paulo Research Foundation (2010/02665-1 to MFP, 2010/02186-6 to IASB and 2005/55200-8 to EMM), CNPq (471963/2007-0 and 500903/2009-3 to EMM) and CAPES (5638/13-5 to MFP and 5639/13-1 to IASB). We thank Heidi Utsunomiya for laboratory assistance and Gerardus Olsthoorn for share the valuable information on generation time of *Pilosocereus* species. We also thank Michael Hickerson and an anonymous referee for valuable comments on the manuscript.

- Almeida FC, Bonvicino CR, Cordeiro-Estrela P (2007). Phylogeny and temporal diversification of Calomys (Rodentia, Sigmodontinae): implications for the biogeography of an endemic genus of the open/dry biomes of South America. *Mol Phylog Evol* **42**: 449–466.
- Arakaki M, Christin PA, Nyffeler R, Lendel A, Eggli U, Ogburn RM *et al.* (2011). Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proc Natl Acad Sci USA* **108**: 8379–8384.
- Avise JC, Walker D, Johns GC (1998). Species durations and pleistocene effects on vertebrate phylogeography. *Proc R Soc London B* **265**: 1707–1712.
- Barbosa AR, Fiorini CF, Silva-Pereira V, Mello-Silva R, Borba EL (2012). Geographical genetic structuring and phenotypic variation in the *Vellozia hirsuta* (Velloziaceae) ochlospecies complex. *Am J Bot* **99**: 1477–1488.
- Bazin E, Dawson KJ, Beaumont MA (2010). Likelihood-free inference of population structure and local adaptation in a Bayesian hierarchical model. *Genetics* **185**: 587–602.
- Beaumont MA (2010). Approximate Bayesian computation in evolution and ecology. *Annu Rev Ecol Syst* **41**: 379–406.
- Becquet C, Przeworski M (2007). A new approach to estimate parameters of speciation models with application to apes. *Genome Res* **17**: 1505–1519.
- Beheregaray LB (2008). Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Mol Ecol* **17**: 3754–3774.
- Blum MGB, François O (2010). Non-linear regression models for Approximate Bayesian Computation. *Stat Comput* **20**: 63–73.
- Bonatelli IAS, Perez MF, Townsend-Peterson A, Zappi DC, Taylor NP, Machado MC *et al.* (2014). Interglacial microrefugia and diversification of a cactus species complex: phylogeography and paleodistributional reconstructions for *Pilosocereus aurisetus* and allies. *Mol Ecol* **23**: 3044–3063.
- Carstens BC, Stoute HN, Reid NM (2009). An information theoretical approach to phylogeography. *Mol Ecol* **18**: 4270–4282.
- Collevatti RG, Castro TG, Lima JS, Telles MPC (2012). Phylogeography of *Tibouchina papyrus* (Pohl) Toledo (Melastomataceae), an endangered tree species from rocky savannas, suggests bidirectional expansion due to climate cooling in the Pleistocene. *Ecol Evol* **2**: 1024–1035.
- Collevatti RG, Rabelo SG, Vieira RF (2009). Phylogeography and disjunct distribution in *Lychnophora eicoides* (Asteraceae), an endangered cerrado shrub species. *Ann Bot* **104**: 655–664.
- Collevatti RG, Telles MPC, Nabout JC, Chaves LJ, Soares TN (2013). Demographic history and the low genetic diversity in *Dipteryx alata* (Fabaceae) from Brazilian Neotropical savannas. *Heredity* **111**: 97–105.
- Corl A, Ellegren H (2013). Sampling strategies for species trees: the effects on phylogenetic inference of the number of genes, number of individuals, and whether loci are mitochondrial, sex-linked, or autosomal. *Mol Phylog Evol* **69**: 358–366.
- Cox M (2011). Microsat. Available at <http://massey.genomicus.com/software.html>.
- Csilléry K, Blum MGB, Gaggiotti OE, François O (2010). Approximate Bayesian Computation (ABC) in practice. *Trends Ecol Evol* **25**: 410–418.
- Csilléry K, François O, Blum MGB (2012). abc: an R package for approximate Bayesian computation (ABC). *Method Ecol Evol* **3**: 475–479.

- Espindola A, Carstens BC, Alvarez N (2014). Comparative phylogeography of mutualists and the effect of the host on the genetic structure of its partners. *Biol J Linn Soc* **113**: 1021–1035.
- Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol* **14**: 2611–2620.
- Excoffier L, Laval G, Schneider S (2005). Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol Bioinform Online* **1**: 47–50.
- Excoffier L, Lischer HEL (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* **10**: 564–567.
- Fagundes NJR, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL *et al*. (2007). Statistical evaluations of alternative models of human evolution. *Proc Natl Acad Sci USA* **104**: 17614–17619.
- Franco FF, Manfrin MH (2013). Recent demographic history of cactophilic *Drosophila* species can be related to quaternary palaeoclimatic changes in South America. *J Biogeogr* **40**: 142–154.
- Garza JC, Williamson EG (2001). Detection of reduction in population size using data from microsatellite loci. *Mol Ecol* **10**: 305–318.
- Garzón-Orduña JJ, Benetti-Longhini JE, Brower AVZ (2014). Timing the diversification of the Amazonian biota: butterfly divergences are consistent with Pleistocene refugia. *J Biogeogr* **41**: 1631–1638.
- Haffer J (1969). Speciation in Amazonian forest birds. *Science* **165**: 131–137.
- Hernández-Hernández T, Hernández HM, De-Nova JA, Puentes R, Eguiarte LE, Magallón S (2011). Phylogenetic relationships and evolution of growth form in Cactaceae (Caryophyllales, Eudicotyledoneae). *Am J Bot* **98**: 44–61.
- Hewitt GM (1996). Some genetic consequences of ices ages, and their role in divergence and speciation. *Biol J Linn Soc* **58**: 247–276.
- Hewitt GM (2004). Genetic consequences of climatic oscillations in the quaternary. *Phil Trans R Soc B* **359**: 183–195.
- Hickerson MJ, Meyer CP (2008). Testing comparative phylogeographic models of marine vicariance and dispersal using a hierarchical Bayesian approach. *BMC Evol Biol* **8**: 322.
- Hoorn C, Wesselingh FP, ter Steege H, Bermudez MA, Mora A, Sevink J *et al*. (2010). Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* **330**: 927–931.
- Hudson RR (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**: 337–338.
- Hughes CE, Pennington RT, Antonelli A (2013). Neotropical plant evolution: assembling the big picture. *Bot J Linn Soc* **171**: 1–18.
- Hunt D, Taylor NP, Charles G (2006). The new cactus lexicon: atlas & text. DH Books, Milborne Port, England.
- Jeffreys H (1961). *Theory of Probability*. 3rd edn, Oxford University Press: New York, NY, USA.
- Keppel G, Van Niel KP, Wardell-Johnson GW, Yates CJ, Byrne M, Mucina L *et al*. (2012). Refugia: identifying and understanding safe havens for biodiversity under climate change. *Global Ecol Biogeogr* **21**: 393–404.
- Knowles LL (2001). Did the Pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshoppers. *Mol Ecol* **10**: 691–701.
- Ledru MP, Braga PIS, Soubiès F, Fournier M, Martin L, Suguio K, Turcq B (1996). The last 50,000 years in the Neotropics (southern Brazil): evolution of vegetation and climate. *Palaeogeogr Palaeoclimatol Palaeoecol* **123**: 239–257.
- Machado T, Silva VX, Silva MJJ (2014). Phylogenetic relationships within *Bothrops neuwiedi* group (Serpentes, Squamata): Geographically highly-structured lineages, evidence of introgressive hybridization and Neogene/Quaternary diversification. *Mol Phyl Evol* **71**: 1–14.
- Marriage TN, Hudman S, Mort ME, Orive ME, Shaw RG, Kelly JK (2009). Direct estimation of the mutation rate at dinucleotide microsatellite loci in *Arabidopsis thaliana* (Brassicaceae). *Heredity* **103**: 310–317.
- Moraes EM, Yotoko KSC, Manfrin MH, Solferini VN, Sene FM (2009). Phylogeography of the cactophilic species *Drosophila gouveai*: demographic events and divergence timing in dry vegetation enclaves in eastern Brazil. *J Biol* **36**: 2136–2147.
- Nason JD, Hamrick JL, Fleming TH (2002). Historical vicariance and postglacial colonization effects on the evolution of genetic structure in *Lophocereus*, a sonoran desert columnar cactus. *Evolution* **56**: 2214–2226.
- Palero F, Lopes J, Abelló P, Macpherson E, Pascual M, Beaumont MA (2009). Rapid radiation in spiny lobsters (*Palinurus* spp) as revealed by classic and ABC methods using mtDNA and microsatellite data. *BMC Evol Biol* **9**: 263.
- 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. *Phil Trans R Soc B* **359**: 515–537.
- Pinheiro F, de Barros F, Palma-Silva C, Fay MF, Lexer C, Cozzolino S (2011). Phylogeography and genetic differentiation along the distributional range of the orchid *Epidendrum fulgens*: a Neotropical coastal species not restricted to glacial refugia. *J Biogeogr* **38**: 1923–1935.
- Prado DE, Gibbs PE (1993). Patterns of species distributions in the dry seasonal forests of South America. *Ann Missouri Bot Gard* **80**: 902–927.
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Riddle BR, Hafner DJ (2006). A step-wise approach to integrating phylogeographic and phylogenetic biogeographic perspectives on the history of a core North American warm deserts biota. *J Arid Environ* **66**: 435–461.
- Ronquist F (1997). Dispersal-variance analysis: a new approach to the quantification of historical biogeography. *Syst Biol* **46**: 195–203.
- Rosenberg N (2004). DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Not* **4**: 137–138.
- Rull V (2011). Neotropical biodiversity: timing and potential drivers. *Trends Ecol Evol* **26**: 508–513.
- Silveira FAO, Negreiros D, Barbosa NPU, Buisson E, Carmo FF, Carstensen DW *et al*. (2015). Ecology and evolution of plant diversity in the endangered campo rupestre: a neglected conservation priority. *Plant Soil*; doi:10.1007/s11104-015-2637-8.
- Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006). Comparative phylogeography of unglaciated eastern North America. *Mol Ecol* **15**: 4261–4293.
- Stewart JR, Lister AM, Barnes I, Dalen L (2010). Refugia revisited: individualistic responses of species in space and time. *Proc R Soc B* **277**: 661–671.
- Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C (2013). Approximate bayesian computation. *PLoS Comput Biol* **9**: e1002803.
- Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Mol Ecol* **7**: 453–464.
- Taylor NP, Zappi DC (2004). Cacti of eastern Brazil. Royal Botanic Gardens, Kew, London, UK.
- Turchetto-Zolet AC, Pinheiro F, Salgueiro F, Palma-Silva C (2013). Phylogeographic patterns shed light on evolutionary process in South America. *Mol Ecol* **22**: 1193–1213.
- Werneck FP, Gamble T, Colli GR, Rodrigues MT, Sites JW (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.
- Yoder AD, Nowak MD (2006). Has vicariance or dispersal been the predominant biogeographic force in Madagascar? Only time will tell. *Ann Rev Ecol Syst* **37**: 405–431.
- Zappi DC (1994). *Pilosocereus* (Cactaceae). *The genus in Brazil*. Royal Botanic Gardens, Kew.

Supplementary Information accompanies this paper on Heredity website (<http://www.nature.com/hdy>)