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).