GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Usefulness of cpDNA markers for phylogenetic and phylogeographic analyses of closely related cactus species

I.A.S. Bonatelli[1], D.C. Zappi[2,3], N.P. Taylor[2,4] and E.M. Moraes[1]

[1]Laboratório de Diversidade Genética e Evolução, Departamento de Biologia, Universidade Federal de São Carlos, Sorocaba, SP, Brasil
[2]Royal Botanic Gardens, Kew, Richmond, Surrey, United Kingdom
[3]Gardens by the Bay and Singapore Botanic Gardens, Singapore
[4]Singapore Botanic Gardens, National Parks Board, Singapore

Corresponding author: I.A.S. Bonatelli
E-mail: belbonatelli@gmail.com

Genet. Mol. Res. 12 (4): 4579-4585 (2013)
Received July 30, 2012
Accepted November 22, 2012
Published February 28, 2013
DOI http://dx.doi.org/10.4238/2013.February.28.27

ABSTRACT. Although plastid DNA has been widely explored as a marker of choice for phylogeny and phylogeography studies, little is known about its utility for examining relationships between closely related species. The slow evolutionary rates inherent to chloroplast (cp) DNA make it difficult to perform lower level taxonomic analyses, particularly at the population level. We characterized the nucleotide variation and investigated the utility of eight noncoding cpDNA regions in four closely related species of the *Pilosocereus aurisetus* group (Cactaceae), an endemic taxon of eastern South America. The plastid intergenic spacers *5'-trnS-trnG*, *3'-trnS-trnG* and *trnT-trnL* were the most variable regions and were the most useful for lower level taxonomic comparisons, especially when used together. We conclude that an adequate combination of regions alongside indels as an additional character improves the usefulness of cpDNA for phylogenetic studies.

Key words: cpDNA; Phylogeny; Phylogeography; *Pilosocereus*; Cactaceae

INTRODUCTION

Intra- and interspecific phylogenetic studies present an inherent challenge: the choice of appropriate molecular markers for the relevant taxonomic level. In plants, noncoding chloroplast DNA (cpDNA) regions have been explored for lower level taxonomic studies because they are under less selective pressure than coding regions and show an amount of polymorphic sites useful for phylogenetic studies (Pleines et al., 2009). However, the phylogenetic signal of different noncoding cpDNA regions can vary extensively among taxonomic groups (Shaw et al., 2005) and the evolutionary rates of different regions are poorly understood.

cpDNA has a maternal inheritance in most species, biparental in some, and paternal in others (Zhang and Sodmergen, 2010). The maternally inherited cpDNA has an interesting role because this allows comparisons between seed gene flow and pollen gene flow as well as the identification of hybridization events when compared to nuclear, biparentally inherited DNA. When it comes to intramolecular rearrangement, cpDNA shows a much more stable structure than does plant mitochondrial DNA (mtDNA). However, the substitution rate of the plastid genome is three to four times higher than that of plant mtDNA (Avise, 2009). Much of the variation observed in plastid noncoding regions is related to insertion-deletion mutations (indels) and should be treated carefully, since these are prone to homoplasy.

Pilosocereus is a genus of Neotropical columnar cacti subdivided in taxonomic groups of morphologically similar species (Zappi, 1994), including the *Pilosocereus aurisetus* group. The species belonging to this group exhibit substantial morphological variation across their range, leading to an unstable taxonomic history, with several species being placed in and removed from synonymy (Zappi, 1994; Taylor and Zappi, 2004; Hunt et al., 2006). Currently, this group consists of eight columnar species: *P. aureispinus* (Buining and Brederoo) Ritter; *P. aurisetus* (Werdermann) Byles and Rowley; *P. bohlei* Hofacker; *P. jauruensis* (Buining and Brederoo) Braun; *P. machrisii* (Dawson) Backeberg; *P. parvus* (Diers and Esteves Pereira) Braun; *P. pusillibaccatus* Braun and Esteves Pereira; and *P. vilaboensis* (Diers and Esteves Pereira) Braun. These species show a disjunct distribution pattern occurring in enclaves of dry vegetation, mostly on rock outcrops within the Cerrado domain in eastern and central Brazil (Zappi, 1994). Our aim was to characterize the variability of noncoding plastid regions and investigate their utility for phylogenetic and phylogeographic studies in closely related species of the *P. aurisetus* group. Our intention was to indicate promising candidate cpDNA regions to guide future phylogenetic and population studies in cactus species.

MATERIAL AND METHODS

Seven populations, comprising 2 to 10 individuals per population, were analyzed across the distribution of the *P. aurisetus* group, in eastern and central Brazil: *P. machrisii* from Altinópolis, SP (FOR), Delfinópolis, SP (DEL), Alto Paraíso de Goiás, GO (APA), and Cristalina, GO (CRI) localities; *P. aurisetus* from Cardeal Mota, MG (CMO); *P. vilaboensis* from Pirenópolis, GO (PIR); and *P. aureispinus* from Ibotirama, BA (IBO).

Genomic DNA was extracted from root tissue using the Dneasy Plant Mini kit (Qiagen, Hilden, Germany). Eight regions of cpDNA were selected based on the availability

of polymerase chain reaction (PCR) primers and previous reports on phylogenetic utility, including the intron *trnL* and the intergenic spacers *trnS-trnG*, *trnH-psbA*, *trnT-trnL*, *trnL-trnF*, *3'-rps16-5'-trnK* and *atpI-atpH*. All these regions are located in the cpDNA large single copy region.

trnS-trnG was amplified in two halves overlapping in 104 bp using previously reported primers in combination with primers designed in this study (primers *SGRev2* and *SGFwd2*). The region *5'-trnS-trnG* was amplified using primers $trnS^{GCU}$ (Shaw et al., 2005) and *SGRev2* (5'-TCC GCT CAT TAG CTC TCC TC-3'), while *3'-trnS-trnG* was amplified using primers *5'-trnG2S* (Shaw et al., 2005) and *SGFwd2* (5'-CAC CCA TGG TTC CCA TTA GA-3'). PCR was performed in 25 µL containing 1X reaction buffer (100 mM Tris-HCl, pH 8.8, 500 mM KCl, 0.8% Nonidet P40), 1.5 mM $MgCl_2$, 200 µM of each dNTP, 0.1 µM of each primer, 1 U Taq DNA polymerase (Fermentas, Burlington, Ontario, Canada), and 1 µL 5-40 ng template DNA. The *trnH-psbA* region was amplified using primers developed by Hamilton (1999), 2 mM $MgCl_2$, 0.4 µM of each primer, and 0.5 U Taq DNA polymerase. The intron *trnL* and the intergenic spacers *trnT-trnL* and *trnL-trnF* were amplified with primers by Taberlet et al. (1991). A final concentration of 3 mM $MgCl_2$ and 1.25 U Taq polymerase were used for *trnT-trnL* and 2 mM $MgCl_2$ for *trnL* and *trnL-trnF*. The regions *3'-rps16-5'-trnK* and *atpI-atpH* were amplified using primers of Shaw et al. (2007), 3 mM $MgCl_2$ and 0.8 U Taq DNA polymerase. All amplifications were performed using an Eppendorf Mastercycler® gradient (Eppendorf AG, Hamburg, Germany) according to the PCR conditions described by Shaw et al. (2005, 2007) with the following annealing temperatures: 62°C for *3'-trnS-trnG*, *5'-trnG2S*, and *trnL*, 53°C for *trnH-psbA*, 50°C for *trnT-trnL*, 60°C for *trnL-trnF*, 48°C for *3'-rps16-5'-trnK*, and 56°C for *atpI-atpH*.

The PCR products were run on 1.5% agarose gels stained with ethidium bromide and purified by ExoSAP - IT or Illustra GFX PCR DNA and Gel Band Purification kit (GE Healthcare, Piscataway, NJ, USA). Both strands were sequenced using the same primers as for the amplification. The samples were sequenced with an ABI PRISM 3730 DNA Analyzer (Applied Biosystems, Foster City, CA, USA) and prepared using the Big Dye terminator version 3.1 cycle sequencing kit (Applied Biosystems).

Forward and reverse DNA sequences were combined in the Chromas 1.5 software (Technelysium Pty Ltd., Tewantin, Australia), and sequences were aligned using ClustalW (Thompson et al., 1994). Indels were manually coded as a single mutation when they shared identical boundaries and length.

Nucleotide analyses such as the number of substitutions (S), indel, parsimoniously informative substitutions and indels, and haplotypes (h) were carried out in DNAsp (Rozas et al., 2003). The proportion of variable sites (% variability) was calculated as (S + indels / L) x 100, where *S* is the number of substitution sites, indel is the number of indel sites, and *L* is the total sequence length. Based on our results, the three most variable regions (*5'-trnS-trnG*, *3'-trnS-trnG* and *trnT-trnL*) were used separately as well as in combination to construct parsimony haplotype networks implemented in TCS 1.21 (Clement et al., 2000).

RESULTS

We analyzed a total of 3719 bp obtained from 270 sequences (Table 1), including 40 sequences of the *5'-trnS-trnG* region (GenBank accession Nos. JN035420-JN035459), 38

of *3'-trnS-trnG* (JN035381-JN035418), 48 of *trnH-psbA* (JN035466-JN035513), 44 of *trnT-trnL* (JN035570-JN035613), 24 of *trnL* (JN035515-JN035538), 31 of *trnL-trnF* (JN035539-JN035569), 17 of *3'-rps16-5'-trnK* (JN035368-JN035372, JN035374-JN035379, JQ323547-JQ323552), and 28 of *atpI- atpH* (JN035462, JN035464, JQ323521-JQ323546).

Table 1. Characterization of nucleotide variation of the plastid regions analyzed for the major sample of *Pilosocereus aurisetus* group.

	5'-trnS-trnG	*3'-trnS-trnG*	*trnH-psbA*	*trnT-trnL*	*trnL*	*trnL-trnF*	*3'-rps16-5'-trnK*	*atpI-atpH*
Length (bp)	555	649	296	293	618	378	439	595
N/h	40/4	38/5	48/3	44/5	24/3	31/2	17/3	28/5
% variability	1.26	1.85	0.68	1.7	0.32	0.26	0.68	1.0
S/S informative	4/3	12/11	0	3/2	1/1	0	1/1	1/1
Id/Id informative	3/3	0	2/2	2/1	1/1	1/1	2/2	5/3
% GC	30.8	40.2	26.1	32.0	27.6	35.2	23.8	32.9
Taxonomic level variation	Sp/Pw	S/Pw/Pa	Sp/Pa	Sp/Pa	Sp	Sp	Sp	Sp/Pa

N = number of sequences analyzed; h = haplotypes; S = substitution sites; S informative = parsimoniously informative substitution sites; Id = indel; Id informative = parsimoniously informative indels; Sp = interspecies variation; Pw = within-population variation; Pa = among-population variation.

The percentage of variability found in each region analyzed ranged from 0.26% in the *trnL-trnF* region to 1.85% in the *3'-trnS-trnG* region. The main source of variability was nucleotide substitutions, despite the fact that the analyzed regions also showed indels, with the exception of *3'-trnS-trnG* (Table 1). We found indels ranging from 1 bp (*5'-trnS-trnG* and *trnH-psbA*) to 356 bp (*atpI-atpH*).

Five regions (*5'-trnS-trnG*, *3'-trnS-trnG*, *trnH-psbA*, *trnT-trnL*, and *atpI-atpH*) showed variation at both the population and species level, while 3 regions (*trnL*, *trnL-trnF* and *3'-rps16-5'-trnK*) were variable only at the interspecies level (Table 1). Within-population variation was only observed for *5'-trnS-trnG* and *3'-trnS-trnG* in *P. aureispinus*, while among-population variation was found for *3'-trnS-trnG*, *trnH-psbA*, *trnT-trnL*, and *atpI-atpH* in *P. machrisii*.

We selected the most variable regions (*5'-trnS-trnG*, *3'-trnS-trnG*, and *trnT-trnL*) to construct the haplotype networks (Figure 1). When the regions were used separately they provided different topologies, and no region was able to show phylogenetic distinctiveness for all species. The phylogenetic resolution of the networks could be increased by including indels in the analysis. The most informative topologies were provided by *3'-trnS-trnG* and *trnT-trnL* (Figure 1) regions. The *3'-trnS-trnG* haplotypes distinguished the central-western (APA and CRI) from the southernmost (FOR and DEL) populations of *P. machrisii*, while *trnT-trnL* haplotypes distinguished two close *P. machrisii* populations in the central-western region (APA and CRI) (Figure 1). However, the topology of *3'-trnS-trnG* network provided a non-monophyletic grouping for *P. machrisii* haplotypes, since the same haplotype was shared by *P. aurisetus* and the southernmost populations of *P. machrisii* (FOR and DEL).

A higher informative topology was obtained by concatenating the three most variable regions (*5'-trnS-trnG*/*3'-trnS-trnG/trnT-trnL*). In this analysis, we found seven haplotypes that resolved the species into distinct branches (Figure 1). *P. machrisii* showed three haplotypes of which only one was shared by two close populations (FOR and DEL).

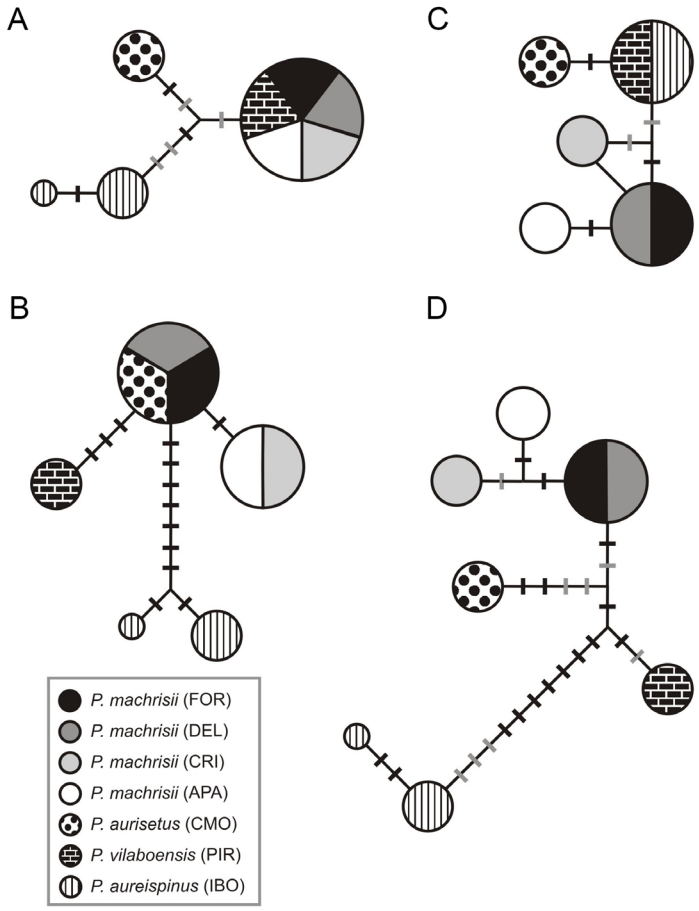

Figure 1. Statistical parsimony network of the *Pilosocereus aurisetus* species group based on noncoding cpDNA spacers *5'-trnS-trnG* (**A**), *3'-trnS-trnG* (**B**), *trnT-trnL* (**C**), and on the combination of these three regions (**D**). The circle sizes are proportional to the frequency of the haplotypes in the sample. Each species/populations was assigned according to the fill patterns and colors shown in the inset box. Black and gray crossed lines represent differences due to indels or substitutions, respectively.

DISCUSSION

In this study, we showed that some widely used plastid regions such as the intergenic spacers *trnH-psbA* and *trnL-trnF*, along with the *trnL* intron, can be very conservative in related species. As demonstrated in this study, the *trnH-psbA* region has shown high occurrence of indels even in closely related species (Aldrich et al., 1988; Shaw et al., 2005). *trnL* and *trnL-trnF* regions showed hardly any variation, while their adjacent region *trnT-trnL* seemed to be more informative as noted by previous studies (Shaw et al., 2005). We observed that the *trnS-trnG* intergenic spacer had the most variable sites in the 3'-region portion. For *atpI-atpH*, we found the occurrence of large indels, as also previously observed for the genus *Magnolia* (Magnoliaceae) (Shaw et al., 2007).

There is no consensus regarding the contribution of indels to cpDNA polymorphism, despite that recent studies have shown that such source of variation may occur more frequently compared to substitutions (Borsch and Quandt, 2009). Plastid genomes are generally AT-rich, which facilitates the occurrence of DNA secondary structure formation and the occurrence of indels, mainly the larger ones. Mononucleotide repeats of A or T also increase the chance of indels in these regions due to polymerase slippages in DNA replication. In our analysis, indels were important in resolving haplotype relationships, and since they are a common feature in cpDNA, we advocate their use in phylogenetic analyses.

In the *Pilosocereus* species studied here, most intraspecific variation observed was due to differences among populations. Such pattern of variation is commonly observed in phylogeographic studies of plant species with a patchy distribution in the Neotropics, such as bromeliads, cacti, and *Petunia* (Solanaceae) (Barbará et al., 2008; Helsen et al., 2009; Lorenz-Lemke et al., 2010). Probably, the lack of intrapopulation variation derives more from the distribution pattern and demographic history of these Neotropical species than from the molecular marker.

Our data reflect the difficulty of finding the appropriate molecular markers for intra- and interspecific studies in closely related plant species. No single region was able to provide a complete taxonomic resolution within the species analyzed when used separately, even the *trnH-psbA* region, which has been chosen as a barcode for plant species (Kress et al., 2005).

Previous studies have shown that combining different cpDNA regions is more useful for differentiating species as well as increasing the resolution and support for phylogenetic topologies (see Miller et al., 2009). In our study, the network topology provided by the three concatenated regions (*5'-trnS-trnG*/*3'-trnS-trnG*/*trnT-trnL*) agreed with morphological data that point to *P. aureispinus* as the more differentiated species within the *P. aurisetus* group (Zappi, 1994). Some divergent morphological characters found in *P. aureispinus* are stems with horizontal rings of bristles in the flowering zones (*vs* longitudinally organized pseudocephalia), narrow (*vs* infundibuliform) flower tube, relatively small fruits, and conical testa cells with coarse cuticular folds (*vs* flat testa cells with less coarse cuticular folds) (Zappi, 1994). Furthermore, the use of those three regions in combination revealed differences among *P. machrisii* populations more effectively. However, even when adopting such a procedure, we were unable to find within-population variation in most populations, an important source of genetic signals of historic demographic events for phylogeographic analyses.

Herein, we uncovered the potential of some cpDNA regions to help us resolve the relationships among closely related species and to perform population genetics studies on the Cactaceae. Thus, on the basis of the molecular data shown in this study, we suggest the use of the concatenated regions *5'-trnS-trnG*/*3'-trnS-trnG*/*trnT-trnL* in future studies of the *P. aurisetus* group, which certainly have an interesting evolutionary history to tell us.

ACKNOWLEDGMENTS

Research supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). We thank H. Utsunomiya, F. Madia, and M. Perez for laboratory assistance, M. Machado for assistance with the field sampling, and F.F. Franco and anonymous reviewers for helpful comments on the manuscript.

REFERENCES

Aldrich J, Cherney BW, Merlin E and Christopherson L (1988). The role of insertions/deletions in the evolution of the intergenic region between psbA and trnH in the chloroplast genome. *Curr. Genet.* 14: 137-146.

Avise JC (2009). Phylogeography: retrospect and prospect. *J. Biogeogr.* 36: 3-15.

Barbará T, Lexer C, Martinelli G, Mayo S, et al. (2008). Within-population spatial genetic structure in four naturally fragmented species of a neotropical inselberg radiation, *Alcantarea imperialis*, *A. geniculata*, *A. glaziouana* and *A. regina* (Bromeliaceae). *Heredity* 101: 285-296.

Borsch T and Quandt D (2009). Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA. *Plant Syst. Evol.* 282: 169-199.

Clement M, Posada D and Crandall KA (2000). TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9: 1657-1659.

Hamilton MB (1999). Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Mol. Ecol.* 8: 521-523.

Helsen P, Browne RA, Anderson DJ, Verdyck P, et al. (2009). Galápagos' *Opuntia* (prickly pear) cacti: extensive morphological diversity, low genetic variability. *Biol. J. Linn. Soc.* 96: 451-461.

Hunt D, Taylor NP and Charles G (2006). The New Cactus Lexicon. Atlas & Text. dh Books, Milborne Port.

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, et al. (2005). Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U. S. A.* 102: 8369-8374.

Lorenz-Lemke AP, Togni PD, Mader G, Kriedt RA, et al. (2010). Diversification of plant species in a subtropical region of eastern South American highlands: a phylogeographic perspective on native *Petunia* (Solanaceae). *Mol. Ecol.* 19: 5240-5251.

Miller JS, Kamath A and Levin RA (2009). Do multiple tortoises equal a hare? The utility of nine noncoding plastid regions for species-level phylogenetics in tribe Lycieae (Solanaceae). *Syst. Bot.* 34: 796-804.

Pleines T, Jakob SS and Blattner FR (2009). Application of non-coding DNA regions in intraspecific analyses. *Plant Syst. Evol.* 282: 281-294.

Rozas J, Sánchez-Delbarrio JC, Messeguer X and Rozas R (2003). DNASP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496-2497.

Shaw J, Lickey EB, Beck JT, Farmer SB, et al. (2005). The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* 92: 142-166.

Shaw J, Lickey EB, Schilling EE and Small RL (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am. J. Bot.* 94: 275-288.

Taberlet P, Gielly L, Pautou G and Bouvet J (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* 17: 1105-1109.

Taylor NP and Zappi DC (2004). Cacti of Eastern Brazil. Royal Botanic Gardens, Kew.

Thompson JD, Higgins DG and Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680.

Zappi DC (1994). *Pilosocereus* (Cactaceae) the Genus in Brazil. Royal Botanic Garden, Kew.

Zhang Q and Sodmergen (2010). Why does biparental plastid inheritance revive in angiosperms? *J. Plant Res.* 123: 201-206.