MARIA CAROLINA SILVA

EVOLUTIONARY ASPECTS OF THE PLASTOMES OF SUBFAMILY OPUNTIOIDEAE (CACTACEAE)

Dissertation presented to the Universidade Federal de Viçosa, as part of the requirement of the Plant Physiology Graduate Program for obtention the title of *Magister Scientiae*

Advisor: Marcelo Rogalski

Co-advisor: Amanda de Santana Lopes

VIÇOSA - MINAS GERAIS
2021

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
S586e
2021

Silva, Maria Carolina, 1995-
Evolutionary aspects of the plastomes of subfamily Opuntioideae (Cactaceae) / Maria Carolina Silva. – Viçosa, MG, 2021.
1 dissertação eletrônica (120 f.): il. (algumas color.).

Orientador: Marcelo Rogalski.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Inclui bibliografia.
DOI: https://doi.org/10.47328/ufvbbt.2021.150
Modo de acesso: World Wide Web.

1. Biologia molecular. 2. Sequenciamento de nucleotídeo. 3. Plastídios. 4. Marcadores moleculares. 5. Edição de RNA. 6. Rearranjo Gênico. 7. Genômica. 8. Mapeamento genômico. 9. Filogenia. I. Universidade Federal de Viçosa. Departamento de Biologia Vegetal. Programa de Pós-Graduação em Fisiologia Vegetal. II. Título.

CDD 22. ed. 572.8633

Bibliotecário(a) responsável: Renata de Fátima Alves CRB6/2578

MARIA CAROLINA SILVA

EVOLUTIONARY ASPECTS OF THE PLASTOMES OF SUBFAMILY OPUNTIOIDEAE (CACTACEAE)

Dissertation presented to the Universidade Federal de Viçosa, as part of the requirement of the Plant Physiology Graduate Program for obtention the title of *Magister Scientiae*.

APPROVED: July 26t^{h} , 2021.

Assent:

Maria Carolina Silva
Author

Marcelo Rogalski
Adviser

DEDICATION

To my parents Maria Izabel and Agmar Rafael
for supporting me, trusting and example

With love,
I dedicate.

ACKNOWLEDGMENTS

I am very grateful to my family, my parents Maria Izabel and Agmar Rafael who provide me home, and good education, and my sisters Gabriela and Crisley, for being great professionals and inspiring me every day. I am very grateful to these three women with whom I share blood ties, for always showing me that women can do anything. And thank you very much Exú for the daily companionship.

I am very grateful to my advisor, Professor Dr. Marcelo Rogalski, for all the dedication to teaching me in these 2 years and 5 months, for respecting my time, and for always being open for both personal and professional conversations. I am very grateful.

I am very grateful to my co-advisor Dra. Amanda Lopes, for all your attention and availability and patience. I am very grateful for the friendship built here. Thanks for always not letting me give up.

I am extremely grateful for the friendships that accompany me over time, Luci, Marina, Lara, Victor, and Nalbert, thanky you. I am very grateful for the friendships that I built during this time in Viçosa, especially to my friends Bruno, Rita, Alexia, Hellen, and Tatiane, thank you, fellowship. I extend my thanks to my friends and colleagues from the Plant Physiology Graduate Program, especially from the Plant Molecular Physiology Lab for friendship and coffee. I especially thank Gleyson for taking the time to teach me molecular practices, and Túlio for teaching me almost everything I know about bioinformatics. I'm also grateful to the personal evolution group that was support mainly during the disciplines.

I am very grateful to Professors Valter Antônio de Baura (UFPR), Eduardo Balsanelli (UFPR), Emanuel Maltempi de Souza (UFPR), and Fábio de Oliveira Pedrosa (UFPR), for supporting the conduction of the study presented here.

Thank you very much to the teachers from the Plant Physiology Graduate Program and from the Biology Undergraduate Program at UFV for knowledge, quality education, and for encouraging me to walk the path of science and always learn more. I also thank Dra Amanda

de Santana Lopes, Dr. Dimas Mendes Ribeiro and Dr. Marcelo Gomes Marçal Vieira Vaz who accepted to participle in my Magister Scientiae committee.

Thanky you to CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for providing resources to carry out this work and FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais) for granting the scholarship.

"If education alone does not transform society, without it, neither does society change."

(Paulo Freire)

ABSTRACT

SILVA, Maria Carolina, M.Sc., Universidade Federal de Viçosa, July, 2021. **Evolutionary aspects of Opuntioideae subfamily (Cactaceae) plastomes.** Advisor: Marcelo Rogalski. Co-adviser: Amanda de Santana Lopes.

The plastid emerged through endosymbiosis. This organelle has a circular genome (plastome) of 100 to 220 kb with a quadripartite structure and harboring around 130 genes. The plastid genes are mainly involved in two processes, photosynthesis and gene expression. The non-propagation of plastid genes trough pollen, results at reduced evolutionary variation of these genes, favoring their use as molecular markers in phylogenetic studies at higher taxonomic ranks. On the other hand, there is a high variation in intergenic spaces and introns, which can be used in phylogenetic at lower taxonomic ranks and population genetic studies. Plastid genomics allow us to visualize the evolution of species through plastome rearrangements, gene divergence and degeneration, positive selection, and RNA editing sites. Therefore, the aim of this study was the sequencing of three plastomes of the subfamily Opuntioideae: *Brasiliopuntia brasiliensis*, *Opuntia monacantha*, and *Opuntia ficus-indica*. These species are widely cultivated and used for food, animal fodder, and ornamental purposes. Besides, they have pharmacological applications, as they present interesting bioactive compounds. Here, we characterize in detail the plastome structures, evolution of plastid genes, mapping of molecular markers, and phylogenetic relationships based on plastid sequences. Structurally, Opuntioideae shows unique rearrangements within Cactaceae, such as expansion of the inverted regions. Molecular evolution analyses of the protein-coding genes show a high divergence of genes involved in essential plastid functions. Furthermore, more than half of these plastid genes bear signatures of positive selection. Several shared RNA editing sites were predicted to occur in the plastid genes of the three species in study. Besides, hundreds of molecular markers were mapped in the plastomes sequenced here, which are informative sequences useful to improve taxonomic classifications and conservation strategies of these species. Moreover, the phylogeny based on concatenated genes resulted in a well-supported tree; with Opuntioideae forming a monophyletic clade that is a sister group of Cactoideae. Finally, this study shades light on the evolutionary patterns of the subfamily Opuntioideae (Cactaceae), and contributes with plenty of useful data to be explored and applied to access and understand the genetic diversity of these Opuntioideae species and trace adequate strategies of rational use and conservation.

Keywords: Plastome evolution. Gene divergence. Rearrangements. Positive selection. RNA editing. Molecular markers.

RESUMO

SILVA, Maria Carolina, M.Sc., Universidade Federal de Viçosa, julho de 2021. **Evolutionary aspects of Opuntioideae subfamily (Cactaceae) plastomes.** Orientador: Marcelo Rogalski. Coorientadora: Amanda de Santana Lopes.

O plastídeo emergiu por endossimbiose. Essa organela possui um genoma circular (plastoma) de 100 a 220 kb com estrutura quadripartida e abrigando cerca de 130 genes. Os genes plastidiais estão principalmente envolvidos em dois processos, fotossíntese e expressão gênica. A não propagação de genes de plastídios através do pólen, resulta na redução da variação evolutiva desses genes, favorecendo seu uso como marcadores moleculares em estudos filogenéticos em classificações taxonômicos superiores. Por outro lado, existe uma grande variação nos espaços intergênicos e íntrons, que podem ser usados na filogenética em classificações taxonômicas mais baixas e estudos de genética populacional. A genômica dos plastídios nos permite visualizar a evolução das espécies por meio de rearranjos nos plastomas, divergência e degeneração gênica, seleção positiva e locais de edição de RNA. Portanto, o objetivo deste estudo foi o sequenciamento de três plastomas da subfamília Opuntioideae: *Brasiliopuntia brasiliensis*, *Opuntia monacantha* e *Opuntia ficus-indica*. Essas espécies são amplamente cultivadas e usadas para alimentação, forragem animal e fins ornamentais. Além disso, apresentam aplicações farmacológicas, pois possuem compostos bioativos interessantes. Aqui, caracterizamos em detalhes as estruturas dos plastídios, a evolução dos genes plastidiais, o mapeamento de marcadores moleculares e as relações filogenéticas com base nas sequências dos plastídios. Estruturalmente, Opuntioideae mostra rearranjos únicos dentro de Cactaceae, como expansão das regiões invertidas. As análises de evolução molecular dos genes codificadores de proteínas mostram uma alta divergência de genes envolvidos nas funções essenciais dos plastídios. Além disso, mais da metade desses genes de plastídios carregam assinaturas de seleção positiva. Vários sítios compartilhados de edição de RNA foram previstos para ocorrer nos genes plastidiais das três espécies em estudo. Além disso, centenas de marcadores moleculares foram mapeados nos plastomas aqui sequenciados, que são sequências informativas úteis para melhorar as classificações taxonômicas e estratégias de conservação dessas espécies. Além disso, a filogenia baseada em genes concatenados resultou em uma árvore bem sustentada; com Opuntioideae formando um clado monofilético que é um grupo irmão de Cactoideae. Por fim, este estudo elucida os padrões evolutivos da subfamília Opuntioideae (Cactaceae), e contribui com muitos dados úteis a serem explorados e aplicados

para acessar e compreender a diversidade genética dessas espécies de Opuntioideae e traçar estratégias adequadas de uso racional e conservação.

Palavras-chave: Evolução do plastoma. Divergência gênica. Rearranjos. Seleção positiva. Edição de RNA. Marcadores moleculares.

SUMMARY

GENERAL INTRODUCTION

Plastids, as well as mitochondria, emerged through endosymbiosis. Specifically, plastids emerged from the incorporation of a cyanobacterium by ancestral free-living eubacteria billions of years ago (Sagan 1967; Gray 2017). From this and other endosymbiotic processes, arose the compartmentalized plant cell where it is possible to separately access the nuclear genome, mitochondrial genome, and plastid genome (plastome). The plastome has gone through an intense re-organization which leads to loss of shared genetic information, transference of genes to the nucleus, and importation of products. Ultimately, this process culminated in a complex interaction between nucleus and plastome (Timmis et al. 2004; Bock and Timmis 2008; Zimorski et al. 2014; Rogalski et al. 2015). This interaction is so well adjusted that in superior plants the nucleus supplies the plastome with almost 3000 imported proteins (Allen 2015). Currently, 130 functional genes are identified within the 100kb to 220kb that compose the plastome, while a cyanobacterium presents around 5000 genes (Allen 2015; Rogalski et al. 2015). Hence, the massive loss of plastome genes is a trend in plant evolution. The circular genome of the plastid is a quadripartite structure with two inverted regions (IRs) that can be divided into two regions of simple copies, the larger one (LSC) and the smaller one (SSC) (Bock 2007; Daniell et al. 2016).

The majority of terrestrial plants present primarily two groups of genes. The (i) first group is linked to the photosynthetic apparatus (photosystem I and II), transport of secondary electrons, ATP synthase, and major subunit of Rubisco, while the (ii) second group is related to gene expression in the ribosome subunit, tRNAs, maturase, RNA polymerase, fatty acids biosynthesis, protein degradation (protease *clpP*) and import of proteins (Wicke et al. 2011; Allen 2015; Book 2007).

Most angiosperms have a maternal inheritance of the plastid genome. Hence, they do not propagate the plastid genes through the pollen grain (Bock 2007). This characteristic is associated with the reduced evolutionary variation of these genes, favoring their use as molecular markers in phylogenetic studies of groups above order (Jansen et al. 2007; Rogalski et al. 2015; Yao et al. 2019). On the other hand, intergenic spaces, characterized by being non-coding regions (introns), present high variation and can be used in phylogenetic studies at lower taxonomic levels (Shaw et al. 2005, 2007; Rogalski et al. 2015). To analyze population-level variations, whether intra- or interspecific, simple repeated sequences present in the plastome can be useful (Rogalski et al. 2015). In addition, plastid genome sequencing allows visualization of the species evolution through rearrangements, deletions, gene loss, and recombinations that

can occur in the plastome at different taxonomic levels (Lopes et al. 2017, 2018abc, 2019; Pacheco et al. 2018 and 2020).

Members of the family Cactaceae have highly rearranged plastomes (Solórzano et al. 2019; Morais da Silva et al. 2021; Sanderson et al. 2015; Oulo et al. 2020; Köhler et al. 2020). These reported rearrangements include contraction and loss of IRs, deletion, and inversion of genes in LSC, and massive gene loss, probably due to the rearrangements. Most genes lost in cacti are related to photosynthetic machinery, *ycf4* more precisely related to photosystem I, genes of the NDH complex, which act in electrochemical gradient dissipation, and genes related to expression in the plastid, *rps* and *rpl* genes (Krech et al. 2012; Strand et al. 2019; Allen 2015).

Cactaceae appeared approximately 35 million years ago, it belongs to the order Caryophyllales with about 11600 species. This family is a monophyletic group, and is composed of four subfamilies: Cactoideae, Maihuenioideae, Opuntioideae, and Pereskioideae, comprising approximately 124 genera and around 1,400 species. (Arakaki et al. 2011; Anderson 2001; Guerrero et al. 2019). Morphologically and anatomically, they present cladodes with leaves reduced to thorns, adaptations that give the species better resistance to drought, and metabolically exhibit CAM metabolism or predominantly C3 in well-watered (Anderson 2001; Ogburn and Edwards 2010; Winter et al. 2011).

Opuntioideae is the subfamily within Cactaceae that occurs in the greatest diversity of habitats and is composed of 16 genera that are distributed in 3 tribes (Cylindropuntieae, Opuntieae, and Tephrocacteae) (Guerrero et al. 2019). The phylogenetic delimitations in Opuntioideae, are still poorly understood. In part, due to the high rate of hybridism between species, resulting in varied morphological characteristics in the hybrids. Phylogenies based on phenotypic aspects circumscribed the delimitation of several taxa within the subfamily. Conflict is seen when there is a construction of molecular phylogenies using nuclear ribosomal regions. However, phenotypic phylogenies are sustained when using plastid intergenic regions, such as the *trnK*/*matK* regions (Griffith and Porter 2009; Ritz et al. 2012). Phylogenetic complexity is amplified by speciation via polyploidy after hybridization, generating meta-species complexes due to the high degree of polyphilia (Anderson 2001; Majure and Puente 2014). Furthermore, the gap of sampling of individuals and the poverty of genetic information make more complete and conclusive studies difficult.

The genus *Opuntia* is popularly known as the prickly pear, belonging to the tribe Opuntieae, it is the most widespread, diverse genus, besides having largest number of species of Opuntioideae (Guerrero et al. 2019). Native to the Western Hemisphere and comprising

approximately 200 species of cacti, its distribution is concentrated in Mexico (Anderson 2001; Bauer and Waechter 2006; Lim 2012; Realini et al. 2015). The genus has high economic importance for many reasons, such as it is well distributed; widely cultivated for feeding purposes in arid areas since it serves as food for cattle, sheep, and goats; also, because the fruits are edible, being consumed both *in natura* and processed; widely cultivated for ornamental purposes (Lim 2012; Majure et al. 2012; Anderson 2001; Realini et al. 2015; Feugang et al. 2006). Furthermore, the intensification of global desertification aggravates the limitation of water resources making *Opuntia* spp. even more interesting as a food resource (Stintzing and Carle 2005).

Opuntia ficus-indica (L.) Mill.

The palm is a cactus with a shrubby habit that can reach up to 5 meters in height. It is probably native to Mexico but naturalized also in South Africa, Australia, and the Mediterranean (Anderson 2001). It features photosynthetic flattened phylloclades (Griffiths and Males 2017). The Indian fig tree has been cultivated for a long time, both for ornamental and feeding purposes (Anderson 2001; Grünwaldt et al. 2015). In the semiarid region of Brazil, cultivated *O. ficus-indica* clones stood out with a greater increase in dry mass compared to *Nopalea cochenillifera* (Silva et al. 2015), proving to be a very efficient alternative for animal forage production.

The fruits, that can be consumed both fresh and cooked, have a high content of antioxidants and vitamins, adding significantly to the nutrition of animals and humans. It is estimated that only the production in urban areas of Mexico City of the fig tree fruit (nopal, as it is known in the locality) moves a large market (Dieleman 2017).

In addition to the previously mentioned uses, the species has high ecological importance in the regions where it occurs, since it is home to several transitional animal and plant species (Oliveira et al. 1999). *O. ficus-indica* also showed potential for bioremediation, as mucilage was used to remove lead and cadmium from contaminated water (Miretzky et al. 2008).

Opuntia monacantha Haw

They are herbaceous plants that reach up to 2m in height, distributed in Argentina, Paraguay, Uruguay, and the coastal region of Brazil, where it is found along the south and southeast coast (Bauer and Waechter 2006). Therefore, occurring in the Atlantic Forest Biome,

one of which is more fragmented, degraded, and threatened by anthropic actions, due to intense tourist activity and agricultural expansion (Zappi et al. 2011).

O. monacantha has been reported for presenting bioactive compounds, antioxidants with high efficiency in fighting free radicals. These flavonoids, when they undergo acid hydrolysis, exhibit antioxidant and antidiabetic activities (Valente et al. 2010; Yang et al. 2008).

Antitumor and trypanocidal activities were also reported for the species using in vitro DNA repair bioassay in *Saccharomyces cerevisiae* and trypomastigote forms of *Trypanosoma cruzi*. The extracts of *O. monacantha* proved to be efficient for both species tested (Valente et al.2007). In addition, cladodes can be used to treat indigestion, asthma, and burns (Lim 2012).

Brasiliopuntia brasiliensis (Willd.) A. Berger

Brasiliopuntia, previously included in the genus *Opuntia*, is currently one among several other groups that present themselves in distinct genera within Opuntioideae. In this new classification, this *Brasiliopuntia* genus is composed of only two species, namely, *B.brasiliensis* and *Opuntia schikendantzii* (Majure et al. 2012). This species presents differentiated growth through dimorphic shoots, a ring of staminodes similar to trichomes in the flowers, and the exclusivity of having single pollen, with the pulporate grains with spinulose perforations (or bullate perforations) supporting the particularity of this genus (Anderson 2001; Majure and Puente 2014).

Brasiliopuntia brasiliensis, a specie native to Brazil, whose main habitat is humid forests, also occurring in the Cerrado and Caatinga and can still be found in other South American countries such as Peru, Bolivia, and Paraguay, it has, therefore, a wide distribution (Anderson 2001; Zappi et al. 2011; Majure and Puente 2014; Zappi and Taylor 2020).

The occurrence in these three biomes characterizes one of the important points of this study for obtaining genetic information on the specie in question, as they are extremely threatened environments and because they present a high number of endemic species. Emphasizing this characteristic of *B. brasiliensis* occupying such heterogeneous environments, from the semiarid in the Caatinga, a region characterized by low rainfall to Cerrado, where they occupy more rocky places and humid forests (Giulietti et al. 2004; Zappi et al. 2011).

It is an arboreal species that can reach up to 20 m in height, with a lignified trunk and lateral disposition of its cladodes (Anderson 2001; Mauseth and Mauseth 2006). *B. brasiliensis* has extrafloral nectaries that resemble common thorns, favoring their pollination and dispersal,

as they are also attractants that increase the number of nutrients in the soil around them, further reducing the presence of harmful insects for the species (Mauseth et al. 2016).

This work presents the complete plastomes of three species belonging to the subfamily Opuntioideae, characterizing their structures and evolutionary aspects. Besides, the sequence data and analyses turned available through this study aim to diminish the genetic information gap within this group.

REFERENCES

Allen JF (2015) Why chloroplasts and mitochondria retain their own genomes and genetic systems: Colocation for redox regulation of gene expression. *Proc Natl Acad Sci USA* **112**: 10231–10238

Anderson EF (2001) The cactus family. Cambridge Press 777, Portland, Oregon. ISBN 0-88192-498-9

Arakaki M, Christin PA, Nyffeler R, Lendel A, Eggli U, Ogburn RM, Spriggs E, Moore MJ, Edwards EJ (2011) Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proc Natl Acad Sci USA* **108**: 8379–8384

Bauer D, Waechter JL (2006) Sinopse taxonômica de Cactaceae epifíticas no Rio Grande do Sul, Brasil. *Acta Bot Brasilica* **20**: 225–239.

Bock R (2007) Structure, function, and inheritance of plastid genomes. *Top Curr Genet Spinger* **19**: 29–63, Berlin, Heidelberg.

Bock R, Timmis JN (2008) Reconstructing evolution: Gene transfer from plastids to the nucleus. *BioEssays* **30**: 556–566.

Daniell H, Lin CS, Yu M, Chang WJ (2016) Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol* **17**: 1–29

Dieleman H (2017) Urban agriculture in Mexico City; balancing between ecological, economic, social and symbolic value. *J Clean Prod* **163**: S156–S163

Edwards EJ, Ogburn RM (2012) Angiosperm responses to a low-CO2 world: CAM and C4 photosynthesis as parallel evolutionary trajectories. *International journal of plant sciences*, **173**(6): 724-733.

Feugang JM, Konarski P, Zou D, Stintzing FC, Zou C (2006) Nutritional and medicinal use of Cactus pear (Opuntia spp.) cladodes and fruits. *Front Biosci* **11**:2574–2589

Giulietti, AM, Neta ALB, Castro AAJF, Gamarra-Rojas CFL, Sampaio EVSB, Virgínio, JF, Queiroz LP, Figueiredo MA, Rodal MJN, Barbosa MRV, Harley RM (2004) Diagnóstico da vegetação nativa do bioma Caatinga. Ministério do Meio Ambiente 49-89, Brasília, DF. ISBN: 85-87166-47-6 *in Portuguese*

Guerrero PC, Majure LC, Cornejo-Romero A, Hernández-Hernández T (2019) Phylogenetic relationships and evolutionary trends in the cactus family. *J Hered* **110**: 4–21

Gray MW (2017) Lynn Margulis and the endosymbiont hypothesis: 50 years later. *Mol Biol Cell* **28**: 1285–1287

Griffith MP, Porter JM (2009) Phylogeny of Opuntioideae (Cactaceae). *Int J Plant Sci* **170**: 107–116

Griffiths H, Males J (2017) Succulent plants. *Current Biology Primer* **27**: 17 PR890-R896

Grünwaldt JM, Guevara JC, Grünwaldt EG, Carretero EM (2015) Cactus (Opuntia sps.) como forraje en las tierras secas de Argentina. *Rev la Fac Ciencias Agrar* **47** :263–282. ISSN: 03704661

Jansen RK, Cai Z, Raubeson LA, Daniell H, dePamphilis CW, Leebens-Mack J, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc Natl Acad Sci U S A* **104**: 19369–19374

Köhler M, Reginato M, Souza-Chies TT, Majure LC (2020) Insights Into Chloroplast Genome Evolution Across Opuntioideae (Cactaceae) Reveals Robust Yet Sometimes Conflicting Phylogenetic Topologies. *Frontiers in plant science*, **11**, 729

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL (2004) Versatile and open software for comparing large genomes. *Genome Biol* **5**: R12

Lim TK (2012) Edible medicinal and non-medicinal plants. Springer **2**: 683–686

Lopes AS, Pacheco TG, Nimz T, Vieira LN, Guerra MP, Nodari RO, de Souza EM, Pedrosa FO, Rogalski M (2018a) The complete plastome of macaw palm [*Acrocomia aculeata* (Jacq.) Lodd. ex Mart.] and extensive molecular analyses of the evolution of plastid genes in Arecaceae. *Planta* **247**: 1011–1030

Lopes AS, Pacheco TG, Santos KGD, Vieira LN, Guerra MP, Nodari RO, de Souza EM, Pedrosa FO, Rogalski M (2018b) The *Linum usitatissimum* L. plastome reveals atypical structural evolution, new editing sites, and the phylogenetic position of Linaceae within Malpighiales. *Plant Cell Rep* **37**: 307–328

Lopes AS, Pacheco TG, Vieira LN, Guerra MP, Nodari RO, de Souza EM, Pedrosa FO, Rogalski M (2018c) The *Crambe abyssinica* plastome: Brassicaceae phylogenomic analysis, evolution of RNA editing sites, hotspot and microsatellite characterization of the tribe Brassiceae. *Gene* **671**:36–49

Lopes AS, Pacheco TG, Silva ON, Cruz LM, Balsanelli E, Souza EM, Pedrosa FO, Rogalski M (2019) The plastomes of *Astrocaryum aculeatum* G. Mey. and *A. murumuru* Mart. show a flip-flop recombination between two short inverted repeats. *Planta* **250**: 1229–1246

Majure LC, Puente R (2014) Phylogenetic relationships and morphological evolution in Opuntia s. str. and closely related members of tribe Opuntieae. *Succul Plant Res* **8**: 9–30

Majure LC, Puente R, Griffith MP, Judd WS, Soltis PS, Soltis DE (2012) Phylogeny of Opuntia s.s. (Cactaceae): Clade delineation, geographic origins, reticulate evolution. *Am J Bot* **99**: 847–864

Mauseth JD, Mauseth JD (2006) Wood in the cactus subfamily opuntioideae has extremely diverse structure. *Bio One* **24**: 93–106

Mauseth JD, Rebmann JP, Machado SR (2016) Extrafloral Nectaries in Cacti. *Cactus Succul J* **88**: 156–171

Miretzky P, Muñoz C, Carrillo-Chávez A (2008) Experimental binding of lead to a low cost on biosorbent: Nopal (Opuntia streptacantha). *Bioresour Technol* **99**: 1211–1217

Morais da Silva G, Lopes AS, Pacheco TG, Machado KLG, Silva MC, Oliveira JD, de Baura VA, Balsanelli E, de Souza EM, Pedrosa FO, Rogalski M (2021) Genetic and evolutionary analyses of plastomes of the subfamily Cactoideae (Cactaceae) indicate relaxed protein biosynthesis and tRNA import from cytosol. *Braz. J. Bot* **44:** 97–116

Oliveira PS, Rico-Gray V, Díaz-Castelazo C, Castillo-Guevara C (1999) Interaction between ants, extrafloral nectaries and insect herbivores in Neotropical coastal sand dunes: Herbivore deterrence by visiting ants increases fruit set in Opuntia stricta (Cactaceae). *Funct Ecol* **13**: 623–631

Oulo M A, Yang J X, Dong X, Wanga V O, Mkala E M, Munyao J N, Onjolo VO, Rono PC, Hu GW, Wang Q F (2020) Complete Chloroplast Genome of Rhipsalis baccifera, the only Cactus with Natural Distribution in the Old World: Genome Rearrangement, Intron Gain and Loss, and Implications for Phylogenetic Studies. *Plants*, **9(8)**: 979

Pacheco TG, Lopes AS, Viana GD, Silva ON, Silva GM, Vieira LN, Guerra MP, Nodari RO, Souza EM, Pedrosa FO, Otoni WC, Rogalski M (2018) Genetic, evolutionary and phylogenetic aspects of the plastome of annatto (*Bixa orellana* L.), the Amazonian commercial species of natural dyes. *Planta* **249**: 563–582

Pacheco TG, Lopes AS, Welter JF, Yotoko KSC, Otoni WC, Vieira LN, Guerra MP, Nodari RO, Balsanelli E, Pedrosa FO, Souza EM, Rogalski M (2020) Plastome sequences of the subgenus *Passiflora* reveal highly divergent genes and specific evolutionary features. *Plant Mol Biol* **104**: 21–37

Realini MF, González GE, Font F, Picca PI, Poggio L, Gottlieb AM (2015) Phylogenetic relationships in Opuntia (Cactaceae, Opuntioideae) from southern South America. *Plant Syst Evol* **301**: 1123–1134

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

Rogalski M, Vieira LDN, Fraga HP, Guerra MP (2015) Plastid genomics in horticultural species: Importance and applications for plant population genetics, evolution, and biotechnology. *Front Plant Sci* **6**: 586

Sagan L (1967) On the origin of mitosing cells. *J Theor Biol* **14**: 225-IN6

Sanderson MJ, Copetti D, Búrquez A, Bustamante E, Charboneau JL, Eguiarte LE, Kumar S, Lee HO, Lee J, McMahon M, Steele K (2015) Exceptional reduction of the plastid genome of saguaro cactus (Carnegiea gigantea): Loss of the ndh gene suite and inverted repeat. *Am J Bot* **102**: 1115–1127

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL (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, 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

Silva TG, Primo JT, de Morais JE, da Silva Diniz WJ, de Souza CA, da Conceição Silva M (2015) Crescimento e produtividade de clones de palma forrageira no semiárido e relações com variáveis meteorológicas. *Rev Caatinga* **28**: 10–18

Solórzano S, Chincoya DA, Sanchez-Flores A, Estrada K, Díaz-Velásquez CE, González-Rodríguez A, Vaca-Paniagua F, Dávila P, Arias S (2019) De Novo Assembly Discovered Novel Structures in Genome of Plastids and Revealed Divergent Inverted Repeats in Mammillaria (Cactaceae, Caryophyllales). *Plants* **8**: 392

Stintzing FC, Carle R (2005) Cactus stems (Opuntia spp.): A review on their chemistry, technology, and uses. *Mol Nutr Food Res* **49**: 175–194

Strand DD, D'Andrea L, Bock R (2019) The plastid NAD(P)H dehydrogenase-like complex: structure, function and evolutionary dynamics. *Biochem J* **476(19)**: 2743-2756

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). *Theor Appl Genet* **106**: 411–422

Timmis JN, Ayliff MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet* **5**: 123–135

Valente LM, Scheinvar LA, da Silva GC, Antunes AP, Dos Santos FA, Oliveira TF, Tappin MR, Neto FR, Pereira AS, Carvalhaes SF, Siani AC (2007) Research Article Evaluation of the antitumor and trypanocidal activities and alkaloid profile in species of Brazilian Cactaceae. *Pharmacognosy Magazine* **3**: 167-172

Valente LM, da Paixão D, Do Nascimento AC, dos Santos PF, Scheinvar LA, Moura MR, Tinoco LW, Gomes LN, da Silva JF (2010) Antiradical activity, nutritional potential and flavonoids of the cladodes of Opuntia monacantha (Cactaceae). *Food Chemistry* **123(4)**: 1127-1131

Wicke S, Schneeweiss GM, Depamphilis CW, Müller KF, Quandt D (2011) The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol. Biol* **76**: 273–297

Winter K., Garcia M., Holtum J. A (2011). Drought-stress-induced up-regulation of CAM in seedlings of a tropical cactus, Opuntia elatior, operating predominantly in the C3 mode. *Journal of Experimental Botany* ***62*****(11)**: 4037-4042

Yang N, Zhao M, Zhu B, Yang B, Chen C, Cui C, Jiang Y (2008) Anti-diabetic effects of polysaccharides from Opuntia monacantha cladode in normal and streptozotocin-induced diabetic rats. *Innov Food Sci Emerg Technol* **9(4)**: 570-574

Yao G, Jin JJ, Li HT, Yang JB, Mandala VS, Croley M, Mostow R, Douglas NA, Chase MW, Christenhusz MJ, Soltis DE (2019) Plastid phylogenomic insights into the evolution of Caryophyllales. *Mol Phylogenet Evol* **134**: 74–86

Zappi D, Nigel T, Silva SR, Machado M, Moraes EM, Cruz ACB, Correia D, Larocca J, Assis JG, Aona L, de Menezes MOT, Meiado M, Marchi MN (2011) Plano de ação nacional para a conservação das cactáceas. Série Espécies Ameaçadas. Instituto Chico Mendes de Conservação da Biodiversidade 112, Brasília. ISBN 978-85-61842-00-0

Zappi D, Taylor N (2021) Cactaceae in Flora do Brasil 2020 under construction. In: Jard. Botânico do Rio Janeiro. http://reflora.jbrj.gov.br/reflora/floradobrasil/FB1430. Accessed July 9, 2020

Zimorski V, Ku C, Martin WF, Gould SB (2014) Endosymbiotic theory for organelle origins. *Curr Opin Microbiol* **22**: 38-48

CHAPTER I: The *Brasiliopuntia brasiliensis* (Willd.) A. Berger plastome reveals unique gene loss within Cactaceae and strong evidence of cytosolic tRNA valine import

ABSTRACT

The suborder Cactineae is characterized by having species associated with xeric habitats. These environmental conditions influenced these species anatomically and metabolically. The Cactaceae family occurs from Canada to the south of South America, in the most varied sizes and shapes. The genus *Brasiliopuntia* belongs to the family Cactaceae and comprises only two species, *Brasiliopuntia brasiliensis* and *Opuntia schikendantzii.* Here, we sequenced *B. brasiliensis* plastome and characterized it structurally and evolutionarily. Several analyzes were performed on plastid protein-coding genes, in order to understand how they have evolved within Cactaceae. Structurally, we identified several rearrangements in *B. brasiliensis* involving all regions of the plastome (LSC, SSC, and IR). We detect the presence of four genes that have degenerated to pseudogenes (*ycf1*, *ycf2*, *rp120,* and *trnV-UAC*). The *trnV-UAC* gene, here identified as a pseudogene due to mutations in its sequence that affected its structure, has been showed to be an essential gene by superwobbling theory. As our codon usage analysis reveal no difference in the frequency of valine codons between *B. brasiliensis* and other model species, we have hypothesized the import of the tRNA-Val from the cytosol. Interestingly, our analyses of synonymous and non-synonymous substitutions indicate that genes of the NDH complex missing in the subfamily Cactoideae are under negative selection in Opuntioideae. We have mapped 236 simple sequence repeats (SSR) along the *B. brasiliensis* plastome, providing informative sequences for further genetic studies. Lastly, our phylogenetic inference based on plastid sequences shows *B. brasiliensis* grouped with *Opuntia quimilo* and both forming a sister group with the other cacti belonging to the subfamily Cactoideae. Overall, the data generated and analyses discussed here shade light on the understanding of evolutionary patterns in Cactaceae.

Keywords: Opuntioideae, Superwobbling, gene divergence, rearrangements, molecular markers

1. INTRODUCTION

The Cactaceae family belongs to the suborder Cactineae, which is characterized by having species associated with xeric habitats, which reflects in morphological and anatomical characteristics, such as succulent tissues for water storage and CAM metabolism (Ocampo and Columbus 2010; Marinho et al. 2018). Cacti family occur only in the New World, from Canada

to the south of South America, inhabiting places with the most varied conditions and characteristics, therefore having the most varied sizes and shapes (Anderson 2001). The suborder Cactineae encompass about 130 genera and eight families, designated ACTP clade, well supported (Marinho et al. 2019).

The genus *Brasiliopuntia* belongs to the Cactaceae family and comprises two species, *Brasiliopuntia brasiliensis* and *Opuntia schikendantzii* (Majure et al. 2012). *Brasiliopuntia brasiliensis* has a wide distribution. It is a Brazilian native species, whose main habitat is humid forests but it is also spread into Cerrado and Caatinga biomes. Furthermore, this species can be found in other South American countries such as Peru, Bolivia, and Paraguay (Anderson 2001; Zappi et al. 2011; Majure and Puente 2014). It is an arboreal species that can reach up to 20 m in height, with a lignified trunk and lateral disposition of its cladodes (Mauseth and Mauseth 2006).

Plastid DNA is mostly maternally inherited and is, therefore, a molecule with an inheritance pattern without recombination (Rogalski et al. 2015; Vieira et al. 2016; Lopes et al. 2018a). Its set of genes that is responsible for coding essential plant functions is separated into two groups, genes related to the photosynthetic apparatus and the expression machinery. Generally, angiosperms present these groups of conserved genes, which includes 79 protein-coding genes involved with both processes mentioned above, 30 tRNAs, and 4 rRNAs (Daniell et al. 2016; Allen 2015; Rogalski et al. 2015). Hence, changes in essential genes can result in changes in plant phenotype and metabolism (Drescher et al. 2000; Kode et al. 2005; Rogalski et al. 2006, 2008a; Agrawal et al. 2020; Alkatib et al. 2012b).

Complete plastome sequences are a powerful source of molecular markers, particularly spread in regions with higher mutation rate such as introns and intergenic spacers. These sequences can be used in genetic studies as a molecular tool for species identification, phylogenetic inference, to understand the evolution of groups of plants (Rogalski et al. 2015; Pacheco et al. 2020b; Nyffeler 2002; Lopes et al. 2019).

A total of 18 plastomes belonging to the Cactaceae family have been completely sequenced to date (Sanderson et al. 2015; Solórzano et al. 2019; Oulo et. al 2020, Morais da Silva et al. 2021; Köhler et al. 2020, Liu et al. 2020; Qin et al. 2021), but, there is still little genetic income available from such a diverse family. The most species sequenced to date belong subfamily Cactoideae, and different from what is found in most angiosperm plastomes, this family interestingly revealed several rearrangements and losses of genes. More genetic information on Cactaceae is important for understanding the evolutionary patterns, in addition

to the complete plastome sequences being used to better support phylogenetic relationships, through molecular markers, study populations, and better identification between subfamilies.

Here, we reported the complete plastome sequence of *B. brasiliensis*, which was molecularly characterized in detail. The plastome of *B. brasiliensis* is a circular DNA molecule of 162,211 bp, the largest known plastome sequence within Cactaceae species to date. The *B. brasiliensis* plastome shows IRB-SSC expansion involving the *ycf1* gene. We detected 109 unique genes and 3 pseudogenes (*ycf1*, *ycf2* and *trnV-UAC*). The *trnV-UAC* gene is essential and was identified as a pseudogene, through mutations along its sequence, probably making its secondary structure non-functional, but our analysis of the use of codons indicates that this gene may be being imported from the cytosol. Also, we analyzed the divergence of plastid protein-coding genes, the synonymous and non-synonymous substitution rates of *ndh* genes and mapping of simple sequence repeats (SSRs). Finally, the phylogenetic inference of *B. brasiliensis* in the suborder Cactineae was performed.

2. MATERIAL AND METHODS

Plant material and plastid DNA extraction

Fresh and young *Brasiliopuntia brasiliensis'* cladodes were kept on dark for one week at 4 °C to decrease the starch content. Afterwards, chloroplast isolation and plastid DNA extraction were performed according to Vieira et al. (2014).

Plastid genome sequencing, assembling and annotation

The sequencing library was prepared with approximately 1 ng of plastid DNA using the sample preparation kit NexteraXT (Illumina Inc., San Diego, CA), according to the manufacturer's instructions. The obtained library was sequenced using MiSeq Reagent Kit v3 (600 cycles) on Illumina MiSeq Sequencer (Illumina Inc., San Diego, California, USA) at the Federal University of Paraná, Brazil. We performed four sequencing, with the better sequencing reaching a total of 636,811paired readings. They were trimmed below the limit with probability of error < 0.05. The trimmed readings (639,138) were *de novo* assembled in contigs using the CLC genomics Workbench 8.0.2 software (CLC Bio, Aarthus, Denmark). The contigs used for the assembly of the *B.brasiliensis* plastome ranged from 1,195.43 to 748.46 average coverage.

The gap ranged from 1 to 88 bp, which were closed using data from the other 3 sequencing, using SPAdes 3.13.0 software and pre-assembled genome (Bankevich et al. 2012).

Initial annotation of the plastome was carried out using Annotation of Organellar Genomes (Tillich et al. 2017), Dual Organellar GenoMe Annotator (DOGMA) (Wyman et al. 2004) and BLAST searches. From the initial annotation, putative start codons, stop codons, and introns were determined based on comparisons to homologous genes in other plastid genomes found in the GenBank database. All tRNA genes were manually verified by using tRNAscan-SE (Lowe and Chan 2016). A physical map of the plastid circular genome was drawn using Organellar Genome DRAW (OGDRAW) version 1.3.1 (Greiner et al. 2019). The complete nucleotide sequence of *B. brasiliensis* plastome was deposited in the GenBank database under accession number OK448351.

Plastome structure analysis

The LSC regions of *B.brasiliensis* and *Portulaca oleracea*, species available in the database containing a typical plastome organization, were compared based on strategy of multiple alignments using the software Mauve Genome Alignment v2.3.1 (MAUVE) (Darling et al. 2004). The MUMmer nucleotide script (NUCmer) Perl from the MUMmer 3.23 software was used to compare the complete plastomes of both species. Similarly, it was also used to compare the plastomes of *B.brasiliensis* and *Opuntia quimilo* (Kurtz et al. 2004). The linear map of genes was drawn by using OGDRAW aiming identification of possible rearrangements (Greiner et al. 2019).

Phylogenetic inference

The inference of the phylogenetic position of *B. brasiliensis* within the suborder Cactineae was realized using a phylogenetic approach, which was based on 54 plastid genes. The plastome of *Spinacia oleracea* (family Chenopodiaceae) was used as an outgroup species. The GenBank accession number of each taxon used here is shown in the **Supplementary Table S1**. The plastid genes were individually extracted from GenBank and aligned by the software MUSCLE (Edgar 2004) implemented in MEGAX (Kumar et al. 2018). The best-fit evolutionary models: TVM+F+I+G4 (*atpA*, *atpB*, *atpE*, *atpF*, *ccsA*, *cemA*, *matK*, *petA*, *psaC*, *psaI*, *psbH*, *psbK*, *psbT*, *rbcL*, *rpl14*, *rpl16*, *rpoB*, *rpoC1*, *rpoC2*, *rps2*, *rps4*, *rps14* and *ycf3*);

GTR+F+I+G4 (*atpH*, atpI, *petB*, *petD*, *petG*, *petN*, *psaA*, *psaB*, *psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbJ*, *psbL*, *psbN* and *psbZ*); TIM3+F+G4 (*clpP*); TVM+F+G4 (infA, *psbM*, *rpl2*, *rpoA*, *rps3*, *rps7*, *rps8*, *rps11*, *rps12*, *rps15* and *rps19*); TVM+F+G4 (*rpl22*), were selected after the maximum likelihood (ML) estimation be conducted using IQTREE v 1.6.10 (Nguyen et al. 2015). Posteriorly, 500 nonparametric bootstrap replications were used to assess branch support. Lastly, the consensus tree was visualized using FigTree 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Gene divergence analysis

For the gene divergence analysis, we selected the 56 protein-coding genes conserved in all 12 Cactaceae plastomes sequenced and published so far, including *B. brasiliensis* plastome. Additionally, other seven genes absent in one or more cacti species but present in B. brasiliensis were also analyzed. *Portulaca oleracea* was used as external group. The sequences were extracted from GenBank database and aligned using the software Muscle (Edgar 2004) implemented in Mega X (Kumar et al. 2018). The phylogenetic reconstruction of each gene was performed to estimate the gene divergence. The phylogenies were inferred based on the ML method, following the same steps used for the phylogenetic inference aforementioned. The gene divergence was estimated as the sum of total branch lengths linking the operational taxonomical units to the common ancestor of Cactaceae species sampled here.

Codon usage

All functional protein-coding genes were extracted from *B. brasiliensis* plastome and from other species available in the GenBank (**Supplementary Table S3**). The genes were individually set and submitted to the software Sequence Manipulation Suite: Codon Usage (https://www.bioinformatics.org/sms2/codon_usage.html).

Analysis of synonymous (dS) and non-synonymous (dN) substitution rates

Pairwise dS and dN substitution rates between *B. brasiliensis* and other representatives of Cactineae (shown in the **Supplementary Table S2**) were calculated based on 11 protein-coding genes of the NDH complex (*ndhA*, *B*, *C*, *D*, *E*, *F*, *G*, *H*, *I*, *J*, *K*). Some genes of this

complex were lost or are pseudogenes in the other cacti available in the database therefore we avoid the use of them in the divergence analysis. Firstly, the sequences of each gene were individually aligned by MUSCLE (Edgar 2004) implemented in MEGA X (Kumar et al. 2018), with pairwise deletion set to gaps and missing data. After, we calculated the dS and dN values using MEGA X under the Kumar model for each alignment (Kimura 2-parameter).

Repeat sequence analysis

Simple sequence repeats (SSRs) in the *B. brasiliensis* plastome were detected using the MIcroSAtellite (MISA) Perl script (Thiel et al. 2003), with thresholds of eight repeat units for mononucleotide SSRs, four repeat units for di- and trinucleotide SSRs, and three repeat units for tetra-, penta- and hexanucleotide SSRs.

3. RESULTS

Overview of the Brasiliopuntia brasiliensis plastome

The *B. Brasiliensis* plastome is a circular DNA molecule of 162,211 bp in length and contains the general quadripartite structure, found in most flowering plants. It comprises two split inverted regions (IRs) of 35,316 bp separated by a large and a small regions of simple copy (LSC and SSC) with 87,186 and 4,393 bp, respectively (**Figure 1**). *B. brasiliensis* plastome is the largest one found in the family Cactaceae so far. It is characterized by containing expanded IRs and a reduced SSC. In Cactaceae, several structural plastome variations have been observed in different species, but a reduction of IRs and an increase of SSC are common features (**Table 1**). The overall GC content determined for *B. brasiliensis* is 36.8%, which is similar to all species of this family (**Table 1**).

The plastome contains 109 unique genes, of which 18 are completely duplicated and three (*ndhG*, *rps12* and *rps19*) are partially duplicated in the IRs (**Table 2**). The set of genes found in *B. brasiliensis* accounts for 76 unique protein-coding genes (11 are completely and three are partially duplicated), 29 unique tRNA genes (seven duplicated) and four unique rRNA genes (all of them duplicated). Among them, 14 genes harbor one intron (five tRNA genes and nine protein-coding genes) and one harbors two introns (*ycf3*). Four genes (*rpl20*, *ycf1*, *ycf2* and *trnV*-UAC) were identified as pseudogenes. The three protein-coding genes contain internal

stop codons in the coding sequence. The tRNA has some modifications in the sequence, which avoid the formation of a functional conformation. The two introns of the *clpP* gene and one intron of the *rpl2* gene were lost.

The LSC region (from *trnH-GUG* gene to *rps19* gene) of *B. brasiliensis* plastome was compared with the LSC of *P. oleracea* by multiple genome alignment by MAUVE (**Suplementary Figure S1**). Here, we hypothesize two possible explanations for the rearrangements founded. The first, a large inversion of block E to B was observed in *B. brasiliensis* in comparison with *P. oleracea*, species with a typical plastome in angiosperms. The analysis also demonstrated that gene position and gene are altered within the LSC in *B. brasiliensis*. Additionally, a second inversion occurred possibly later, because the LCB and D have the same orientation and order of genes as the ancestor. This second inversion probably reverted the primary inversion in the LCB B and LCB D blocks, which correspond from *petL* to *rps12*, and the pseudogene *trnV-UAC* only in *B. brasiliensis* to *rbcL*, leaving these blocks without rearrangements if we compare with the ancestor. It is clearly observed in the linear map (**Figure 3**). Other hypothesis would be the inverse of the first, that is, inversions occurred in blocks E and C first. The second inversion encompassed the entire region involved in these rearrangements (E to B), consequently it involved the blocks E and C, who have already suffered inversion.

Moreover, the complete plastome was analyzed by MUMmer (**Figure 2)** and some differences between *B. brasiliensis* and *P. oleracea* were also identified. The aforementioned rearrangements localized in the LSC region are indicated by the B to E letters in **Figure 2**. Besides, other rearrangmentes are pointed out in the IR borders (numbers 1 to 3 in **Figure 2**). Firstly, the *ycf1* gene was deleted from the boundary IRB-SSC (1). Secondly, an inversion occurred in the region corresponding to the SSC in *P. oleracea*, from *ycf1* gene to *trnL-UAG* gene (2). The last event observed in the plastome of *B. brasiliensis* indicates an expansion of the IRs. This feature occurs in the SSC/IRB boundary (3), which is demonstrated by the presence of a set of genes within the IRs usually found at the end of the SSC in *P. oleracea*. We also detailed these events in a linear map (**Figure 3**). Moreover, we also compared *B. brasiliensis* with *Opuntia quimilo* (**Figure 3** and **Supplementary Figure S2**), both species belonging the same subfamily. *O. quimilo* shows the same rearrangements, except for the contraction event from the IRs to the IRB-LSC boundary [**Figure 3** (4)]. This is the opposite of what occurred in *B. brasiliensis* and also compared with the ancestral *P. oleracea.*

Phylogenetic inference

The phylogenetic position of *B. brasiliensis* in the suborder Cactineae was inferred based on the sequence of 54 plastid protein-coding genes. This analysis included cacti of the subfamily Opuntioideae, represented by *Opuntia quimilo* and *B. brasiliensis*; subfamily Cactoideae, represented by *Carnegia gigantea*, *Lophocereus schotti*, *Rhipsalis teres*, and seven species of genus *Mammillaria*; and other 12 species of the suborder Cactineae (families Basellaceae, Montiaceae, Halophytaceae, Portulacaceae and Talinaceae). The species *Spinacea oleracea* (Chenopodiaceae) was used as outgroup (**Suplemmentary Table S1**). The maximum likelihood (ML) analysis produced a consensus tree (**Figure 4**) with a log-likelihood of -131303.593952.

The tree topology within Cactaceae, *B. brasiliensis* formed a sister group with *O. quimilo*, both species of the subfamily Opuntioideae, and a sister group with other cacti of the subfamily Cactoideae. Within the family Cactaceae, all nodes showed bootstrap support (BS) values of 100%, except for two nodes within the genus *Mammillaria*, including one with 99% (*M. zephrranthoides*) and another with 73% (*M. crucigera* and *M. hitzilopochtli*). The genus *Mammillaria* formed a monophyletic group inside this tree. Curiously, *Rhipsalis teres* formed a sister-group with *Carnegiea gigantea* and *Lophocereus schottii*, given that *R. teres* has epiphytic habit and the other two species are known to have terrestrial habitat and are giant cacti. Portulacaceae forms a sister-group with Cactaceae (BS of 88%) and both families formed a group-sister to Talinaceae (BS of 100%). Montiaceae diverged before the other families within Cactineae. Concerning the three genera of Montiaceae, *Calandrinia* is closely related to *Montia*, which form a sister group with *Cistanthe* (BS of 100%). Halophytaceae diverged first from Cactaceae, Portulacaceae, Talinaceae and Basellaceae (BS of 96%). The relationships among Cactaceae, Talinaceae and Baselaceae were low supported (BS of 50%).

Gene divergence analysis in Cactaceae plastomes

To analyze the evolution of plastidial genes within Cactaceae, we estimated the gene divergence of 63 protein-coding genes based on phylogenetic trees. We found that 56 out of 63 genes have low substitution rates (> 0.1) in all sampled taxa, suggesting their strong conservation within Cactaceae. The remain seven genes (*rpl22*, *psbM*, *rps19*, *petL*, *infA*, *clpP* and *accD*), on the other hand, show higher substitution rates (> 0.1) for some or all sampled

taxa **(Figure 5)**. The *psbM* and *rpl22* genes showed high divergence in some species within the subfamily Cactoideae, whilst they remain quite conserved among Opuntioideae species (substitution rates around 0.01). Similarly, the *infA* gene is slight more divergent in most Cactaceae species, varying from 0.1 to 0.19, while in *B. brasiliensis* the branch length is below 0.1. On the contrary, the *petL* gene showed a branch length of 0.13 in Opuntioideae, while its sequence is highly conserved in Cactoideae. The *clpP* and *accD* genes have a remarkably high substitution rates among all Cactaceae species sampled, ranging from 0.23 to 0.39 and from 0.72 to 1.8, respectively. In *B. brasiliensis*, the branch lengths are 0.32 and 1.73 for *clpP* and *accD*, respectively. Noteworthy, the *accD* gene, which shows the greatest variation of branch lengths, has the highest substitution rates within Opuntioideae.

Structural analysis trnV-UAC gene and codon usage

We compare here the sequence of *trnV-UAC* gene from *B. brasiliensis* with other 16 species from different taxonomic groups (i.e., angiosperms, gymnosperms, pteridophyte, bryophyte and hepatophyte). The analysis revealed several modified bases in the sequence, occurring transversions in positions 12, 27, 49 and 50 (highlighted by red arrow) and transitions in positions 25 and 42 (highlighted by blue arrow), modifying its primary structure (**Supplementary Figure S3**). The modified bases, transversions (12, 49 and 50) and transition (42), resulted in no correct pairing in the arms (D and C) and anticodon, if compared with the other species in the same region of the secondary structure (**Figure 8**, **Supplementary Figure S4**).

We then evaluated the use of valine codons in the plastid proteins of *B. brasiliensis* by comparing with other species (**Figure 6**, black circle). The analysis showed that the use of codons follow the same pattern found in other species (i.e. the rate of use of each codon is similar to that observed in plastomes containing both valine tRNAs).

Analysis of synonymous (dS) and non-synonymous (dN) substitution rates of ndh genes

The genes related to the NAD (P) H dehydrogenase complex (ndh genes), were lost or pseudogenes in the subfamily Cactoideae. Relationship of the *ndh* complex in the Cactaceae family, highlighted in the red blocks missing genes, pseudogene in the yellow block and functional gene in green (**Supplementary Figure S6**).

Therefore, in order to understand the behavior of the plastid genes related to the *ndh* genes within Cactaceae, substitutions synonyms (dS), non-synonyms (dN), and dN/dS values were estimated for the 11 *ndh* genes present in *B. brasiliensis* plastome and in closer relatives of the suborder Cactineae that have also maintained the whole set of *ndh* genes (species listed in **Supplementary Table S2**).

In general, dS values were low and ranged from 0.01 to 0.13 for *B. brasiliensis* and other species (**Supplementary Figure S5a**), although the dS values of most genes were slight lower in *B. brasiliensis* compared to the other species. The *ndhE* gene showed the highest dS value, whilst *ndhB* gene presented the lowest value. Similarly, we observed low dN values (below 0.05) in *B. brasiliensis ndh* genes, as well as in the other species sampled (**Supplementary Figure S5b**). Concerning the dN/dS ratio, all *ndh* genes from *B. brasiliensis* showed values very below (ranging from 0.1 to 0.4) (**Supplementary Figure S5c**). Altogether, these data indicates that the whole set of *ndh* genes found in *B. brasiliensis* are surprisingly under negative selection.

Repeat sequence analysis

The identification, type and distribution of simple sequence repeats (SSRs) in the plastome of *B. brasiliensis* were mapped, revealing a total of 236 SSR loci (**Supplementary Table S3**). Of them, the majority were classified as monopolymers (172) and dipolymers (43). In contrast, tripolymers (3), tetrapolymers (13), pentapolymers (3) and compound SSRs (1) were identified in lower frequency. The majority of the SSRs (88.13%) detected here are composed of A and T. Of the total mapped in the *B. brasiliensis* plastome, 172 are located in the LSC, 13 in the SSC and 51 in the IRs. In addition, 145 occur in intergenic spacers (IGS), 54 in CDS and 42 in introns. The SSRs located in CDSs comprise 29 genes, being the higher frequency identified in the *rpoc2* (7), *ycf1* (7), *ycf2* (4) and *accD* (4) genes. Most of the SSRs mapped to the introns occurred in the *ndhA* (6), *rpl16* (5), *trnK-UUU* (5), *trnL-UAA* (4), *atpF* (4), and *ycf3* (4) genes.

4. DISCUSSION

B. brasiliensis shows the largest plastome within Cactaceae, expanded IRs and a conserved set of genes

To date, *B. brasiliensis* has the largest plastome among cacti already sequenced. In addition, the *B. brasiliensis* plastome maintains a conserved set of genes and the typical quadripartite structure found in most flowering plants (Wicke et al. 2011; Ruhlman and Jansen 2014; Rogalski et al. 2015). The inversions and translocations of blocks of genes found in the LSC of *B. brasiliensis*, if compared with *P. oleracea* plastome (typical gene order), occurred in the same way in *O. quimilo*. Regarding the IR dynamics, however, some differences are reported beteween both Opuntioideae, in which *B. brasiliensis* bears a specific expansion of the IRB-SSC boundary while *O. quimilo* shows a contraction of the IRB-LSC boundary (**Figures 3**).

In the subfamily Cactoideae, complete plastome sequences of several species revealed intriguing rearrangements, gene losses and pseudogenes (Sanderson et al. 2015; Solórzano et al. 2019; Morais da Silva et al. 2021), while we identified that most of the genes reported as pseudogenes in Cactoideae are functional in *B. brasiliensis*. Besides, a tendency towards contracted IRs has been observed in Cactoideae, whereas in *B. brasiliensis* and *O. quimilo* we notice expanded IRs. However, *ycf1* and *ycf2* gene have degenerated into pseudogenes in *B. brasiliensis* as observed in other species of Cactaceae (*M. zephranthoides, O. quimilo* and *R. baccifera*) and other angiosperms (Solórzano et al. 2019; Köhler et al. 2020; Oulo et al. 2020; Guisinger et al. 2011; Fajardo et al. 2013; Rabah et al. 2019; Pacheco et al. 2020). A peculiar feature of *B. brasiliensis* plastome is the loss of the *rpl20* gene. This gene encodes the L20 ribosomal protein (50S subunit), which is essential for ribosome biogenesis and cell viability (Rogalski et al. 2008b). Replacement of *rpl20* gene by a nuclear copy was characterized in *Passiflora* (Sherestha et al. 2020), and a similar mechanism could be also occurring in *B. brasiliensis* to supply the lack of this essential gene.

All plastomes of the subfamily Cactoideae sequenced so far have lost the *ndh* genes, or they are present, but as pseudogenes. This is not the case of the subfamily Opuntioideae, in which *B. brasiliensis* and *O. quimilo* contain the whole set of 11 plastid *ndh* genes encoding subunits for the NDH complex, suggesting that this complex is fully functional in Opuntioideae. To investigate the putative functionality of *ndh* genes, we analyze the dN/dS ratio aiming to infer the evolution of these genes. The analysis surprisingly demonstrated that they are under negative selection in *B. brasiliensis*, which means that they are under pressure to be maintained functional in the subfamily Opuntioideae. The NDH complex works by dissipating electrochemical gradient and generating protomotor force in the thylakoid membrane (Shikanai

2016). Several reports demonstrated that plant mutants containing an inactivated NDH complex are weakly affect negatively in photosynthesis and growth (Burrows et al. 1998; Horváth et al. 2000), but the importance of the NDH complex in Cactaceae is still unclear and remains to be investigated in different species and environmental conditions. Several groups of plants growing under diverse environments have degenerated or lost the plastid *ndh* genes (Martín and Sabater 2010; Blazier et al. 2011; Lin et al. 2015, 2017; Ni et al. 2017; Strand et al. 2019). Nonetheless, the selection pressure acting to maintain or loss these genes is still unclear (Strand et al. 2019).

Furthermore, our analysis of gene divergence pointed out some highly divergent genes, particularly *clpP* and *accD* genes. The divergence found in the *petL* gene in Opuntioideae, but not in Cactoideae (0.01), may be a specific feature of this subfamily. The *clpP* gene is essential because compose a proteolytic subunit of the ATP-dependent Clp protease related to plastid protein homeostasis (Kuroda and Maliga 2003), and this gene has lost its introns in all Cactaceae sequenced so far, a trait also shared by other families with highly divergent *clpP* such as Passifloraceae, Linaceae (Lopes et al. 2018; Morais da Silva et al 2020; Köhler et al. 2020; Oulo et al. 2020; Pacheco et al. 2020). The *accD* gene is essential for fatty acid biosynthesis in plastids and cellular viability (Salie and Thelen 2016; Kode et al. 2005; Rogalski and Carrer 2011). The *accD* is the most divergent gene within Cactaceae, being considered as a pseudogene in Cactoideae by previous reports (Sanderson et al. 2015; Solorzano et al. 2019; Köhler et al. 2020). However, Cactaceae *accD* sequences do not contain any internal stop codon that would produce a truncated version of its product and, additionally, its 3'-end portion is fairly conserved (C-terminal domain that holds the β-carboxyl transferase activity), suggesting that the gene is also functional in Cactoideae (Morais da Silva et al. 2021).

Valine transporter RNA import signs

Plastomes usually contain genes (*trnV-UAC* and *trnV-GAC)* for isoacceptors of valine tRNAs $tRNA^{Val}$(UAC) or $tRNA^{Val}$(GAC), which are responsible for reading of the four codons (GTC, GTT, GTA, GTG) of the mRNA (Crick 1961). According to the conventional wobble rules, the $tRNA^{Val}$(UAC) encoded by the *trnV*-UAC gene would decode the two valine codons with a purine in the third codon position (GTA and GTG), whereas the $tRNA^{Val}$(GAC) encoded by the *trnV*-GAC gene would read the two valine codons with a pyrimidine in the third codon position (GTC and GTU) (Crick 1961). Afterwards, the mechanism of superwobbling was

demonstrated to improve the reading capacity of tRNAs containing a uridine (U) in the wobble position to four codons in box families (Rogalski et al. 2008a; Alkatib et al 2012a). Curiously, the *trnV-UAC* gene was lost in the subfamily Cactoideae (Morais da Silva et al. 2021) and the $tRNA^{Val}$(GAC), for stearic reasons, can not read the codons GTA and GTG. Consequently, the mutations in the *trnV-UAC* gene make the $tRNA^{Val}$(UAC) no functional in *B. brasiliensis*. Although the $tRNA^{Val}$(UAC) was considered functional in *O. quimilo* (Köhler et al. 2020), the gene contains the same mutations found in *B. brasiliensis*, and consequently, it is also nonfunctional. In other cacti of the subfamily Cactoideae, the *trnV-UAC* gene is absent (Sanderson et al. 2015; Solórzano et al. 2019; Oulo et al. 2020). The extremely rearranged region from the *trnM-CAU* to *rbcL* genes could be related to the loss of the tRNA in Cactoideae but an essential gene can not simply be lost without a replacing mechanism for the same essential function (Morais da Silva et al. 2021).

Our analysis of codon usage suggests a possible import mechanism for the $tRNA^{Val}$(UAC) given that it is not possible, for stearic reasons, for *trnV-GAC* to decode the four codons present in the mRNA (Rogalski et al. 2008a; Alkatib et al. 2012b). Heretofore, sequenced cactus plastomes contain only the *trnV-GAC* gene, however, the deletion of this gene in tobacco plants revealed that it is not essential for cell survival (Corneille et al. 2001; Alkatib et al. 2012b). The literature is rich in examples descripting tRNA import from the cytosol to the mitochondria and the mechanisms involved in that in different taxonomic groups, from Marchantiophyta to flowering plants (Akashi et al. 1998; Salinas-Giegé et al. 2015; Murcha et al. 2016; Warren and Sloan 2020; Reinbothe et al. 2021), whereas in plastids there is no experimental evidence of tRNA import so far (Legen et al. 2007; Rogalski et al. 2008a; Alkatib et al. 2012a, b; Agrawal et al. 2020).

Cactaceae phylogeny based on plastid sequences

Here we used concatenated plastid genes to construct the Cactineae phylogeny. Our phylogeny affirms the monophyly of Cactaceae, as the sampled Opuntioideae and Cactoideae subfamilies has formed a clade sister to Portulacaceae family. As expected, the species sequenced in this study, *B. brasiliensis*, was grouped with *Opuntia quimilo*, since both species belong to the subfamily Opuntioideae (Griffith and Porter 2009; Majure et al. 2012, Majure and Puente 2014). Regarding the subfamily Cactoideae, we sampled ten taxa comprising 4 genera . The genus *Mammillaria* (seven species) formed a monophyletic group, which is sister to a clade

encompassing *Lophocereus schottii*, *Carnegiea gigantea* and *Rhipsalis teres* species. These data agree with the current classification within this subfamily in which *Carnegiea* and *Lophocereus* belong to the core Pachycereeae (Core Cactoideae I) that is sister to the Core Cactoideae II, which includes the tribe Rhipsalidae, while *Mammilaria* is ranked into the tribe Cacteae (Copettii et al. 2017; Guerrero et al. 2019; Morais da Silva et al. 2021). The phylogenetic inferences coherently corroborate with the evolutionary and structural features observed in the plastomes of these species (Sanderson et al. 2015; Solórzano et al. 2019; Morais da Silva et al. 2021).

Several phylogenetic reconstructions based on IGS, ITS and transcriptomes formed a sister group among Cactaceae and Portulacaceae, and with Anacampserotaceae. Here, it is partially corroborated because no complete plastome of species of family Anacampserotaceae is available (Ocampo and Columbs 2010; Walker et al. 2018; Wang et al. 2019; Moore et al. 2018).

Importance of plastid molecular markers for Opuntioideae

We mapped in the plastome of *B. brasiliensis* 236 SSRs, which are simple sequence repeats normally characterized in plastid regions with a high mutation rate. They are interesting molecular markers with broad applications, such as in genetic studies of natural populations or germplasm collection (Rogalski et al. 2015). For more than 30 years, plastid markers have been used in analyses related to phylogenetic and genetic relationships among plants at different taxonomical levels (Gitzendanner et al. 2018). Cactaceae family has diversified from its ancestor approximately 35 million years ago (Mya) and the diversification of Opuntioideae subfamily is estimated around 6-8 Mya (Arakaki et al. 2011). The diversification of species under adverse environmental condition (Arakaki et al. 2011) may have contributed for rearrangements, gene losses and high mutation rates in plastomas of Cactaceae. Some regions mapped here containing many repetitions and are already widely used in phylogenies based on intergenic plastid regions, such as *trnK*-UUU /*matK* and *trnL*-UAA (Majure et al. 2012; Nyffeler 2002; Ritz et al. 2012). This kind of molecular markers can contribute to the classification of species, given that it is reported as monophyletic clade, but several hybrid plants were identified within the subfamily Opuntioideae (Griffith and Porter 2009; Majure et al. 2012).

5. CONCLUSIONS

The plastome of *B. brasiliensis* (162,211 bp) reported here is the largest among the cacti so far. This specific feature of the plastome can be determined by events of expansion of the IRB-SSC junction, with the fixation of *ycf1* sequence (pseudogene) in the IRs, and several inversions in the LSC. Unlike Cactoideae subfamily, gene losses and degeneration into pseudogenes is less frequent in Opuntioideae, a markable antagonistic feature between both subfamilies. Besides, we discussed here the strong evidence of *trnV-UAC* import from cytosol within Cactaceae. This essential plastid tRNA was lost in Cactoideae and we have showed here evidence of its degeneration in Opuntioideae. Notwithstanding, the codon usage of plastidial transcripts has surprisingly remained conserved in Cactaceae according our data, prompting us to hypothesed the import of cytosolic tRNA-Val(UAC) to circunvent the lack of a functional *trnV*-UAC gene. Finally, we have presented here the Cactaceae phylogeny based on concatenated-plastid genes and the complete mapping of SSRs loci in the *B. brasiliensis* plastome, turning available many useful and informative sequences for future genetic studies to acess the diversity of this Brasilian native species belonging to the outstanding family of cacti.

6. REFERENCES

Agrawal S, Karcher D, Ruf S, Bock R (2020) The functions of chloroplast Glutamyl-tRNA in translation and tetrapyrrole biosynthesis. *Plant Physiology* **183**: 263-276.

Allen JF (2015) Why chloroplasts and mitochondria retain their own genomes and genetic systems: Colocation for redox regulation of gene expression. *Proc Natl Acad Sci USA* **112**: 10231–10238

Alkatib S, Fleischmann T, Scharff LB, Bock R (2012a) Evolutionary constraints on the plastid tRNA set decoding methionine and isoleucine. *Nucleic Acids Res* **40**: 6713–6724

Alkatib S, Scharff LB, Rogalski M, Fleischmann TT, Matthes A, Seeger S, Schöttler MA, Ruf S, Bock R (2012b) The contributions of wobbling and superwobbling to the reading of the genetic code. *PLoS Genet* **8**: e1003076

Anderson EF (2001) The cactus family. Timber Press, Portland, Oregon

Arakaki M, Christin PA, Nyffeler R, Lendel A, Eggli U, Ogburn RM, Spriggs E, Moore MJ, Edwards EJ (2011) Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proc Natl Acad Sci USA* **108**: 8379–8384

Akashi K, Takenaka M, Yamaoka S, Suyama Y, Fukuzawa H, Ohyama K (1998) Coexistence of nuclear DNA-encoded tRNAVal (AAC) and mitochondrial DNA-encoded tRNAVal (UAC) in mitochondria of a liverwort Marchantia polymorpha. *Nucleic acids research*, ***26*****(9)**: 2168-2172

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19(5)**: 455-77

Blazier JC, Guisinger MM, Jansen RK (2011) Recent loss of plastid-encoded ndh genes within Erodium (Geraniaceae). *Plant Mol Biol* **76**: 263–272

Britton NL, Rose JN (1963) *The Cactaceae: descriptions and illustrations of plants of the cactus family* (Vol. 3) *Courier Corporation* pp 164-166

Burrows PA, Sazanov LA, Svab Z, Maliga P, Nixon PJ (1998) Identification of a functional respiratory complex in chloroplasts through analysis of tobacco mutants containing disrupted plastid ndh genes. *The EMBO Journal* **17(4)**: 868-876

Copetti D, Búrquez A, Bustamante E, Charboneau JL, Childs KL, Eguiarte LE, Sanderson MJ (2017) Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. *Proceedings of the National Academy of Sciences* **114(45)**: 12003 12008. DOI: 10.1073/pnas.1706367114

Corneille S, Lutz K, Svab Z. and Maliga P (2001) Efficient elimination of selectable marker genes from the plastid genome by the CRE-lox site-specific recombination system. *Plant J* **27**: 171–178

Crick, F.H.C (1966) Codon-anticodon pairing: the wobble hypothesis. *J. Mol. Biol* **19**: 548–555

Daniell H, Lin CS, Yu M, Chang WJ (2016) Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol* **17**: 1–29

Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* **14**: 1394–1403

Drescher A, Ruf S, Calsa T Jr, Carrer H, Bock R (2000) The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *Plant J*. **22**: 97–104

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

Fajardo D, Senalik D, Ames M, Zhu H, Steffan SA, Harbut R, Polas- hock J, Vorsa N, Gillespie E, Kron K, Zalapa JE (2013) Complete plastid genome sequence of Vaccinium macrocarpon: structure, gene content, and rearrangements revealed by next genera- tion sequencing. *Tree Genet Genomes* **9**: 489–498

Gitzendanner MA., Soltis PS, Yi TS, Li DZ, Soltis DE (2018). Plastome phylogenetics: 30 years of inferences into plant evolution. *Advances in botanical research* **85**: 293-313.

Greiner S, Lehwark P, Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res* **47**: W59–W64

Griffith MP, Porter JM. (2009). Phylogeny of Opuntioideae (Cactaceae). *International Journal of Plant Sciences* **170(1)**: 107-116

Guerrero PC, Majure LC, Cornejo-Romero A, Hernández-Hernández T (2019) Phylogenetic relationships and evolutionary trends in the cactus family. *J Hered* **110**: 4–21

Guisinger MM, Kuehl JV, Boore JL, Jansen RK (2011) Extreme reconfiguration of plastid genomes in the angiosperm family geraniaceae: rearrangements, repeats, and codon usage. *Mol Biol Evol* **28**: 583–600

Horváth EM, Peter SO, Joët T, Rumeau D, Cournac L, Horváth GV, Kavanagh TA, Schäfer C, Peltier G, Medgyesy P (2000) Targeted inactivation of the plastid ndhB gene in tobacco results in an enhanced sensitivity of photosynthesis to moderate stomatal closure. *Plant Physiol* **123**: 1337–1350

Kode V, Mudd EA, Iamtham S, Day A (2005) The tobacco plastid *accD* gene is essential and is required for leaf development. *Plant J* **44**: 237–244

Köhler M, Reginato M, Souza-Chies TT, & Majure LC (2020). Insights Into Chloroplast Genome Evolution Across Opuntioideae (Cactaceae) Reveals Robust Yet Sometimes Conflicting Phylogenetic Topologies. *Frontiers in plant science* **11**: 729

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution* ***35*****(6)**: 1547-1549.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Anto-nescu C, Salzberg SL (2004) Versatile and open software for comparing large genomes. *Genome Biol* **5(2)**: R12

Kuroda H, Maliga P (2003) The plastid clpP1 protease gene is essential for plant development. *Nature* **425**: 86–89

Legen J, Wanner G, Herrmann RG, Small I, Schmitz-Linneweber C (2007) Plastid tRNA Genes trnC-GCA and trnN-GUU are essential for plant cell development. *Plant J* **51**: 751–762

Lin CS, Chen JJ, Huang YT, Chan MT, Daniell H, Chang WJ, Hsu CT, Liao DC, Wu FH, Lin SY, Liao CF, Deyholos MK, Wong GKS, Albert VA, Chou ML, Chen CY, Shih MC (2015) The location and translocation of ndh genes of chloroplast origin in the Orchidaceae family *Sci Rep* **5**: 9040

Lin CS, Chen JJW, Chiu CC, Hsiao HC, Yang CJ, Jin XH, Shih MC (2017) Concomitant loss of NDH complex-related genes within chloroplast and nuclear genomes in some orchids. *Plant J* **90**: 994–1006

Liu J, Liu ZY, Zheng C, Niu YF (2021) Complete chloroplast genome sequence and phylogenetic analysis of dragon fruit (Selenicereus undatus (Haw.) DR Hunt) *Mitochondrial DNA Part B* **6(3)**: 1154-1156

Lopes AS, Pacheco TG, Vieira LN, Guerra MP, Nodari RO, Souza EM, Pedrosa FO, Rogalski M (2018a) The Crambe abyssi- nica plastome: Brassicaceae phylogenomic analysis, evolution of RNA editing sites, hotspot and microsatellite characteri- zation of the tribe Brassiceae. *Gene* **671**: 36–

Lopes AS, Pacheco TG, Santos KGD, Vieira LN, Guerra MP, Nodari RO, de Souza EM, Pedrosa FO, Rogalski M (2018b) The *Linum usitatissimum* L. plastome reveals atypical structural evolution, new editing sites, and the phylogenetic position of *Linaceae* within *Malpighiales*. *Plant Cell Rep* **37**: 307–328

Lopes AS, Pacheco TG, Silva ON, Cruz LM, Balsanelli E, Souza EM, Pedrosa FO, Rogalski M (2019) The plastomes of Astrocaryum aculeatum G. Mey. and A. murumuru Mart. show a flipflop recombination between two short inverted repeats. *Planta* **250**: 1229–1246

Lowe TM, Chan PP (2016) tRNAscan-SE On-line: integrating Search and context for analysis of transfer RNA genes. *Nucleic Acids Res* **44**: W54–W57

Majure LC, Puente R, Griffith MP, Judd WS, Soltis PS , Soltis DE (2012) Phylogeny of Opuntia ss (Cactaceae): clade delineation, geographic origins, and reticulate evolution. *American journal of botany* ***99*****(5)**: 847-864

Majure LC, Puente R (2014) Phylogenetic relationships and morphological evolution in Opuntia s. str. and closely related members of tribe Opuntieae. *Succulent Plant Research* **8**: 9-30

Martín M, Sabater B (2010) Plastid ndh genes in plant evolution. *Plant Physiol Biochem* **48**: 636–645

Marinho M A O, Souza G, Felix L P, De Carvalho R (2019) Comparative cytogenetics of the ACPT clade (Anacampserotaceae, Cactaceae, Portulacaceae, and Talinaceae): a very diverse group of the suborder Cactineae, Caryophyllales. *Protoplasma* **256**(3): 805-814

Mauseth JD, Mauseth JD (2006) Wood in the cactus subfamily opuntioideae has extremely diverse structure. **24**: 93–106

Moore AJ, Vos JM, Hancock LP, Goolsby E, Edwards EJ (2018) Targeted enrichment of large gene families for phylogenetic inference: phylogeny and molecular evolution of photosynthesis genes in the Portullugo Clade (Caryophyllales). *Syst Biol* **67**: 367–383

Morais da Silva G, Lopes AS, Pacheco TG, Machado KLG, Silva MC, Oliveira JD, de Baura VA, Balsanelli E, de Souza EM, Pedrosa FO, Rogalski M (2021) Genetic and evolutionary analyses of plastomes of the subfamily Cactoideae (Cactaceae) indicate relaxed protein biosynthesis and tRNA import from cytosol *Braz J Bot* **44**: 97–116

Murcha MW, Kubiszewski-Jakubiak S, Teixeira PF, Gügel IL, Kmiec B, Narsai R, Ivanova A, Megel C, Schock A, Kraus S, Berkowitz O, Glaser E, Philippar K, Maréchal-

Drouard L, Soll J, Whelan J (2016) Plant-specific preprotein and amino acid transporter proteins are required for tRNA import into mitochondria. *Plant Physiol* **172**: 2471–2490

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximumlikelihood phylogenies. *Mol Biol Evol* **32**: 268–274

Ni Z, Ye Y, Bai T, Xu M, Xu LA (2017) Complete chloroplast genome of Pinus massoniana (Pinaceae): gene rearrangements, loss of ndh Genes, and short inverted repeats contraction. *Expansion Mol* **22**: 1528

Nyffeler R (2002) Phylogenetic relationships in the cactus Family (*Cactaceae*) based on evidence from trnK/matK and *trnL-trnF* sequences. *Am J Bot* **89**: 312–326

Ocampo G, Columbus JT (2010) Molecular phylogenetics of suborder Cactineae (Caryophyllales), including insights into photo synthetic diversification and historical biogeography. *Am J Bot* **97**: 1827–1847

Oulo MA, Yang JX, Dong X, Wanga, VO, Mkala, EM, Munyao JN, Onjolo VO, Rono PC, Hu GW, Wang QF (2020) Complete Chloroplast Genome of Rhipsalis baccifera, the only Cactus with Natural Distribution in the Old World: Genome Rearrangement, Intron Gain and Loss, and Implications for Phylogenetic Studies. *Plants* **9**(8): 979

Pacheco TG, Lopes AS, Welter JF, Yotoko KSC, Otoni WC, Vieira LN, Guerra MP, Nodari RO, Balsanelli E, Pedrosa FO, Souza EM, Rogalski M (2020) Plastome sequences of the subgenus *Passiflora* reveal highly divergent genes and specific evolutionary features. *Plant Mol Biol* **104**: 21–37

Pacheco TG, da Morais SG, Lopes AS, Oliveira JD, Rogalski JM, Balsanelli E, Souza EM, Pedrosa FO, Rogalski M (2020b) Phylogenetic and evolutionary features of the plastome of Tropaeolum pentaphyllum Lam (Tropaeolaceae). *Planta* **252**: 17

Qin Q, Li J, Zeng S, Xu Y, Han F, Yu J (2021) The complete plastomes of red fleshed pitaya (Selenicereus monacanthus) and three related Selenicereus species: insights into gene losses, IR expansions and phylogenomic implications. *Research Square*

Rabah SO, Shrestha B, Hajrah NH, Sabir Mumdooh J, Alharby HF, Sabir Mernan J, Alhebshi AM, Sabir JSM, Gilbert LE, Ruhlman TA, Jansen RK (2019) Passiflora plastome sequencing reveals widespread genomic rearrangements: Passiflora plastome evolu- tion. *J Syst Evol* **57**: 1–14

Reinbothe S, Rossig C, Gray J, Rustgi S, von Wettstein D, Reinbothe C, Rassow J (2021). tRNA-Dependent Import of a Transit Sequence-Less Aminoacyl-tRNA Synthetase (LeuRS2) into the Mitochondria of Arabidopsis. *International journal of molecular sciences* **22**(8): 3808

Rogalski M, Ruf S, Bock R (2006) Tobacco plastid ribosomal protein S18 is essential for cell survival. *Nucleic Acids Res* **34**: 4537– 4545

Rogalski M, Carrer H (2011) Engineering plastid fatty acid biosynthesis to improve food quality and biofuel production in higher plants. *Plant Biotechnol J* **9**: 554–564

Rogalski M, Karcher D, Bock R (2008a) Superwobbling facilitates translation with reduced tRNA sets. *Nat Struct Mol Biol* **15**: 192–198

Rogalski M, Schottler MA, Thiele W, Schulze WX, Bock R (2008b) Rpl33, a nonessential plastid-encoded ribosomal protein in tobacco, is required under cold stress conditions. *Plant Cell* **20**: 2221–2237

Rogalski M, Nascimento VL, Fraga HP, Guerra MP (2015) Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution, and biotechnology. *Front Plant Sci* **6**: 586

Ruhlman TA, Jansen RK (2014) The plastid genomes of flowering plants. *Methods Mol Biol* **1132**: 3–38

Salie MJ, Thelen JJ (2016) Regulation and structure of the heteromeric acetyl-CoA carboxylase. *Biochim Biophys Acta* **1861**: 1207–1213

Salinas-Giegé T, Giegé R, Giegé P (2015) tRNA biology in mitochondria. *Int J Mol Sci* **16**: 4518–4559

Sanderson MJ, Copetti D, Búrquez A, Bustamante E, Charboneau JLM, Eguiarte LE, Kumar S, Lee HO, Lee J, McMahon M, Steele K, Wing R, Yang TJ, Zwickl D, Wojciechowski MF (2015) Exceptional reduction of the plastid genome of saguaro cactus (Carnegiea gigantea): loss of the ndh gene suite and inverted repeat. *Am J Bot* **102**:1115–1127

Shrestha B, Gilbert LE, Ruhlman TA, Jansen RK (2020) Rampant nuclear transfer and substitutions of plastid genes in *Passiflora*. *Genome Biol Evol* **12**(8): 1313-1329

Shikanai T (2016) Chloroplast NDH: a different enzyme with a structure similar to that of respiratory NADH dehydrogenase. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1857**(7): 1015-1022.

Solórzano S, Chincoya DA, Sanchez-Flores A, Estrada K, Díaz- Velásquez CE, González Rodríguez A, Vaca-Paniagua F, Dávila P, Arias S (2019) De novo assembly discovered novel structures in genome of plastids and revealed divergent inverted repeats in Mammillaria (Cactaceae, Caryophyllales). *Plants* **8**: 392

Strand DD, D'Andrea L, Bock R (2019). The plastid NAD (P) H dehydrogenase-like complex: structure, function and evolutionary dynamics. *Biochemical Journal* **476**(19): 2743-2756.

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* **106**: 411–422

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S (2017) GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res* **45**: W6–W11

Vieira LN, Faoro H, Fraga HPF, Rogalski M, de Souza EM, de Oliveira PF, Nodari RO, Guerra MP (2014) An improved protocol forintact chloroplasts and cpDNA isolation in conifers. *PLoS One* **9**(1): e84792

Vieira LN, Rogalski M, Faoro H, Fraga HP, Anjos KG, Picchi GFA, Nodari RO, Pedrosa FO, Souza EM, Guerra MP (2016) The plastome sequence of the endemic Amazonian conifer, Retrophyllum piresii (Silba) C.N.Page, reveals different recombination events and plastome isoforms. *Tree Genet Genomes* **12**: 10

Walker JF, Yang Y, Feng T, Timoneda A, Mikenas J, Hutchison V, Edwards C, Wang N, Ahluwalia S, Olivieri J, Walker-Hale N, Majure LC, Puente R, Kadereit G, Lauterbach M, Eggli U, Flores-Olvera H, Ochoterena H, Brockington SF, Moore MJ, Smith SA (2018) From cacti to carnivores: Improved phylotranscriptomic sampling and hierarchical homology inference provide further insight into the evolution of Caryophyllales. *Am J Bot* **105**: 446–462

Wang N, Yang Y, Moore MJ, Brockington SF, Walker JF, Brown JW, Liang B, Feng T, Edwards C, Mikenas J, Olivieri J, Hutchison V, Timoneda A, Stoughton T, Puente R, Majure LC, Eggli U, Smith SA (2019) Evolution of Portulacineae marked by gene tree conflict and gene family expansion associated with adaptation to harsh environments. *Mol Biol Evol* **36**: 112–126

Warren JM, Sloan DB (2020) Interchangeable parts: The evolutionarily dynamic tRNA population in plant mitochondria. *Mitochondrion* **52**: 144-156

Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol* **76**: 273–297

Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **20**: 3252–3255

Zappi D, Nigel T, Silva SR, Machado M, Moraes EM, Cruz ACB, Correia D, Larocca J, Assis JG, Aona L, de Menezes MOT, Meiado M, Marchi MN (2011) Plano de ação nacional para a conservação das cactáceas. Série Espécies Ameaçadas. Instituto Chico Mendes de Conservação da Biodiversidade 112, Brasília. ISBN 978-85-61842-00-0

7. FIGURES

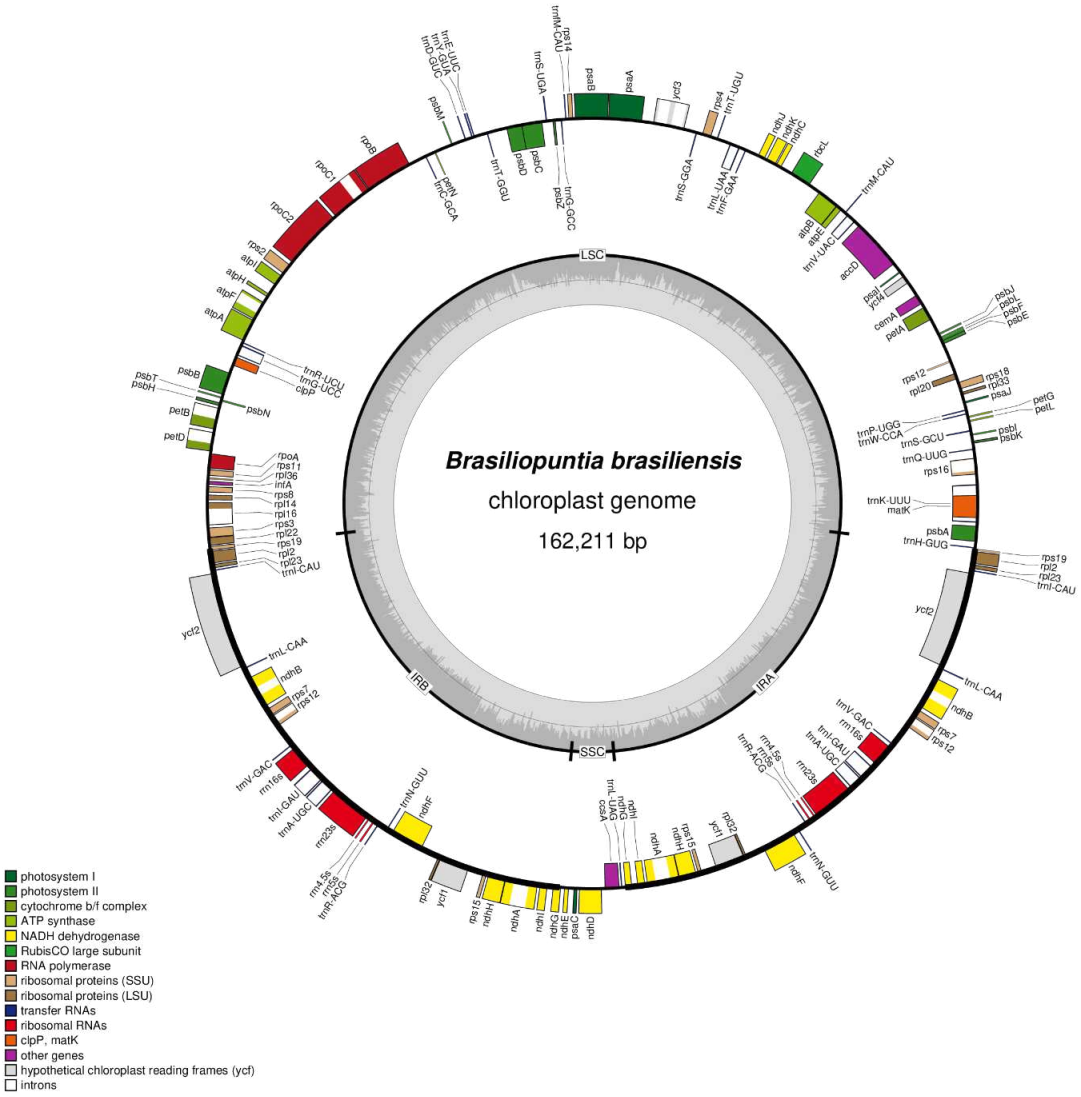


Figure 1. Gene map and genome organization of *B. brasiliensis.* Two inverted repeat regions IR_A and IR_B divide the circular DNA molecule into large (LSC) and small (SSC) single copy regions. Genes drawn inside the circle are transcribed clockwise, and genes drawn outside are expressed counterclockwise. Genes belonging to different functional groups are color-coded. The dark gray in the inner circle corresponds to GC content, while the light gray corresponds to AT content. Dotted circle corresponds to 50% of AT/GC content.

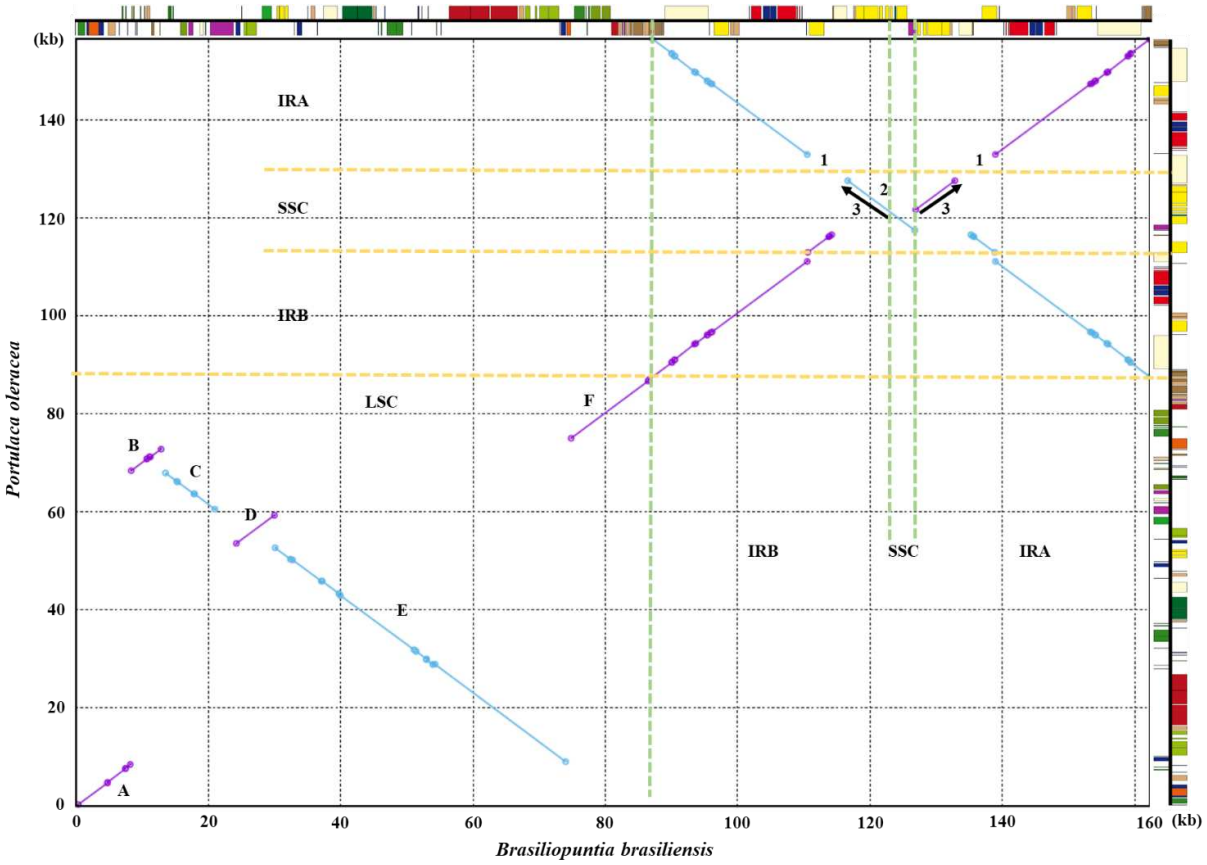


Figure 2. Dot-plot analyses comparing the plastomes of *Brasiliopuntia brasilienses* (x-axis) against *Portulaca oleracea* (y-axis). A positive slope denotes that the pair of sequences compared is in the same orienta- tion. A negative slope denotes that the pair of sequences compared can be aligned, but their orientation is opposite. Sequences in the same direction are purple and inversions are blue. The black arrows indicate contratrion and expansions in the region IRB-SSC. Vertical green lines indicate the regions SSC, LSC and IR's in *B. brasiliensis*, while in *P. oleracea* is indicates by yellow horizontal lines the same regions. More details about regions highlighted with letters and numbers **Figure 2**.

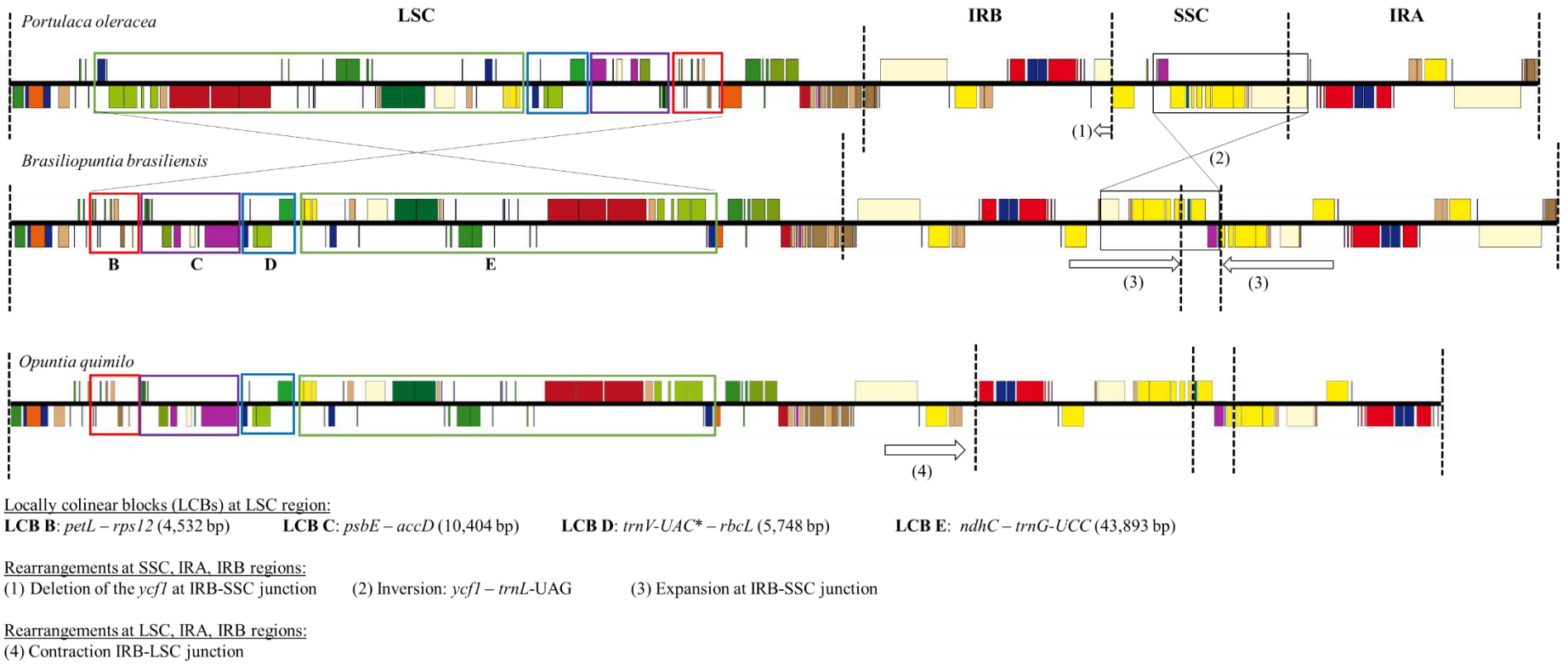


Figure 3. Comparison of the gene content and order between the *Brasiliopuntia brasiliensis* with *Portulaca oleracea* plastomes, which presents the general structure found in most angiosperms plastomes, and also between Opuntioideas. *B.brasiliensis* is also compared with *Opuntia quimilo*, both Opuntioideae. The genes involved in inversions are highlighted by crossed and straight lines gray arrows, while a secondary inversion (ie, occurred in a region that suffered inversion before) is highlighted by red (B) and blue (D) square. The white arrows indicates contraction and expansion. The (*) indicates pseudogene. Dotted black lines horizontally separate as regions of the plastoma, LSC and SSC, large and small single-copy regions; IRA / B, inverted repeat region. The linear gene maps were drawn by using OGDRAW.

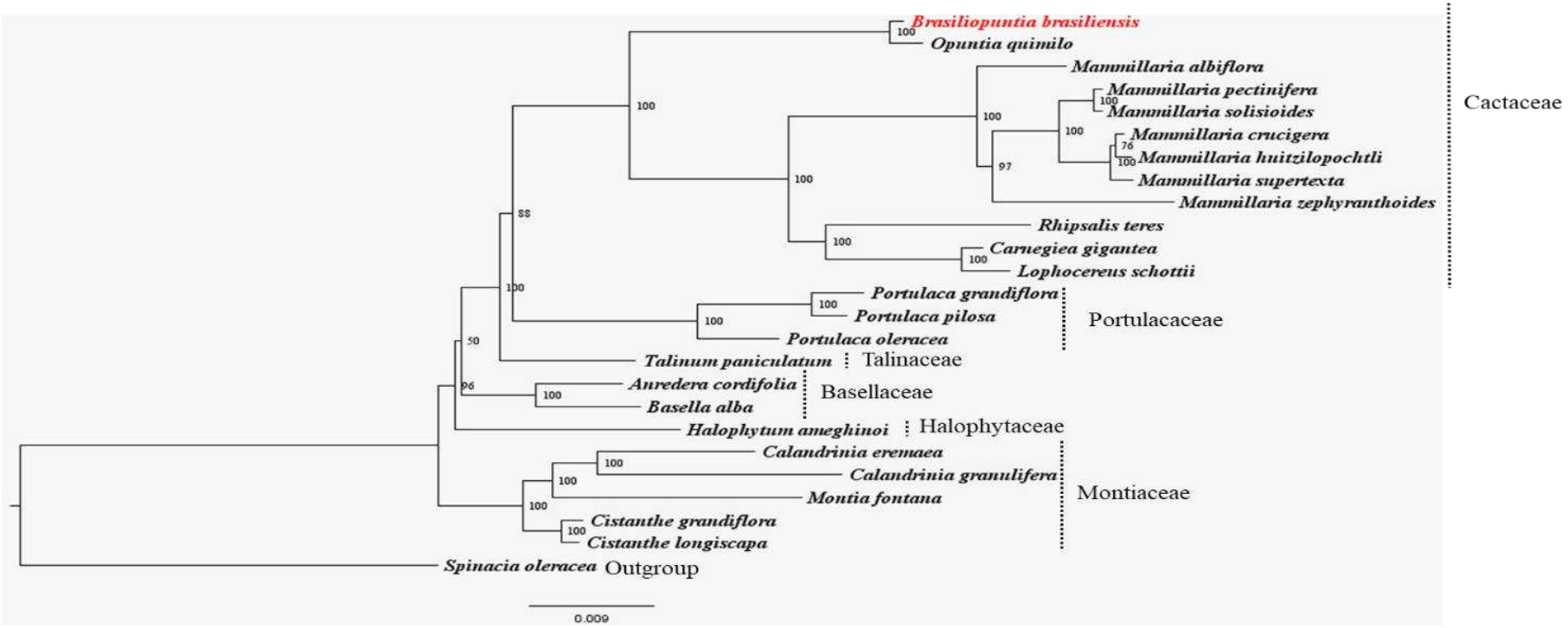


Figure 4. Cactineae phylogenetic tree of 25 taxa (24 species of suborder Cactineae and one outgroup) based on 54 protein-coding plastid genes using maximum likelihood (ML) method. Numbers (%) associated with branches are ML bootstrap support (BS) values. The branch length is proportional to the inferred divergence level. The scale bar indicates the number of inferred nucleic acid substitutions per site. The position of *Brasiliopuntia brasiliensis* is highlighted in red. *Spinacia oleracea* was used to root the tree. Tree reconstruction was performed using IQTREE software, with TVM+F+I+G4, GTR+F+I+G4, TIM3+F+G4 and TVM+F+G4, models.

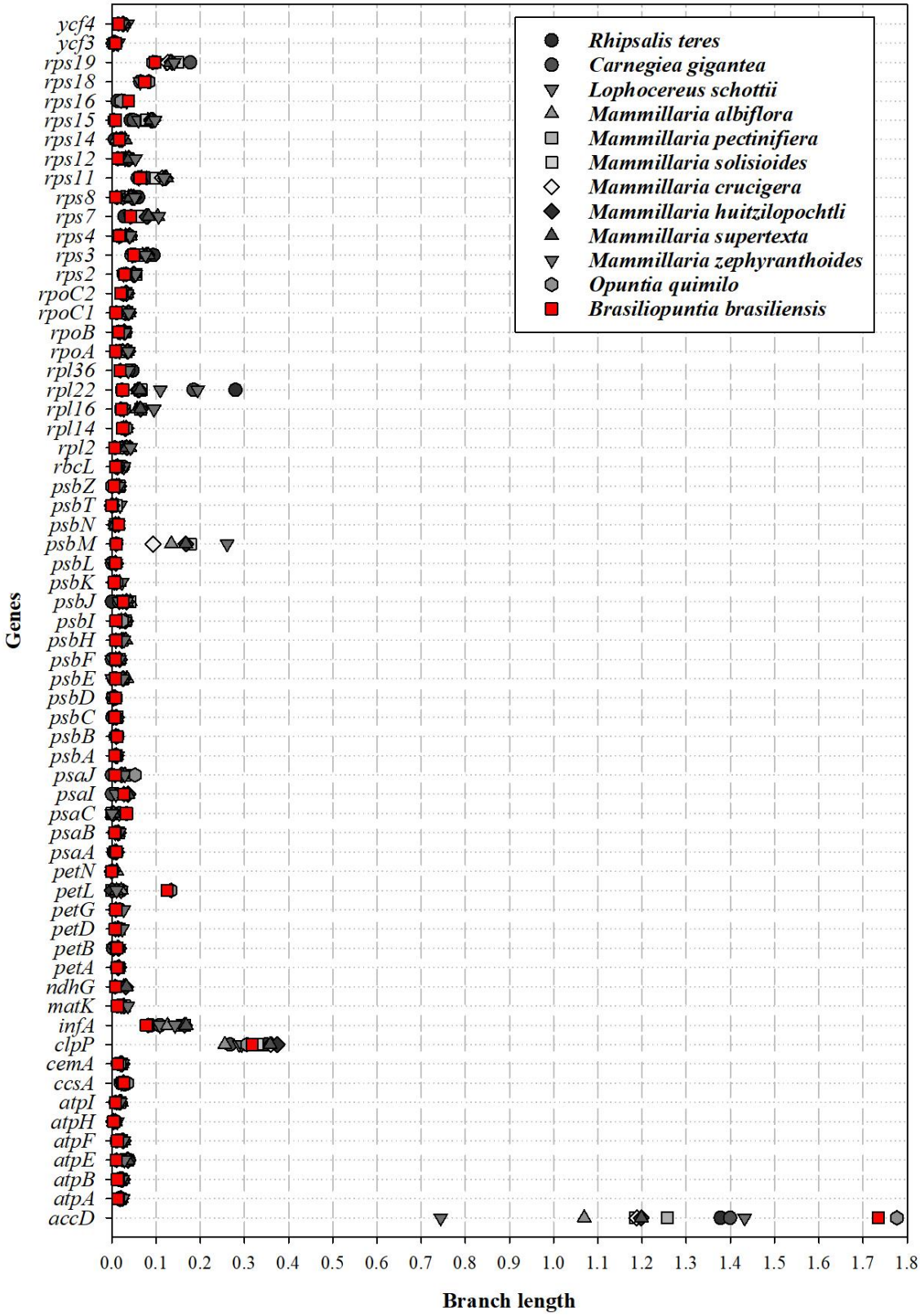


Figure 5. Divergence of plastid protein-coding genes among species of the family Cactaceae. The gene divergence was estimated by the sum of total branch lengths in each gene tree infererd.

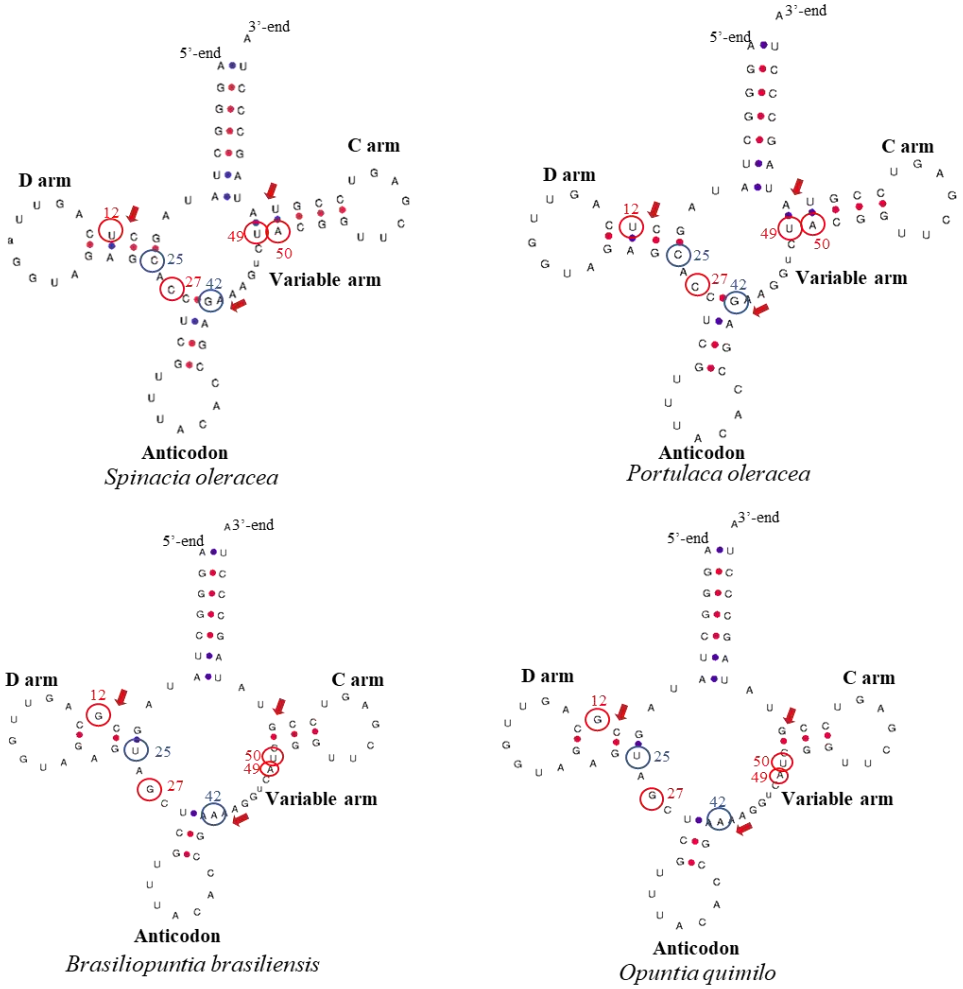


Figure 6. Secondary structure of the trnV-UAC consists of 3 arms, 2 of which are non-variable (C-arm and D-arm), and arm where the anticondon is located. A variable portion (Variable arm). The reading pattern is also indicated 5'end (start) and 3'end stop. The red arrows indicate the non-pairing of 1 base pair in the arms of *B. brasiliensis* and *O. quimilo* when compared with *S. oleracea* and *P. oleracea*. Change of bases bases (highlighted by red circle, base transverse comparing *B. brasiliensis* and *O.quimilo* in compare P.oleracea and Spinacia oleracea and blue circle base transition. The numbers indicates position variable base in *B.brasiliensis* and *O.quimilo* compared with other species alignment in (**Supplementary Figure S3**).

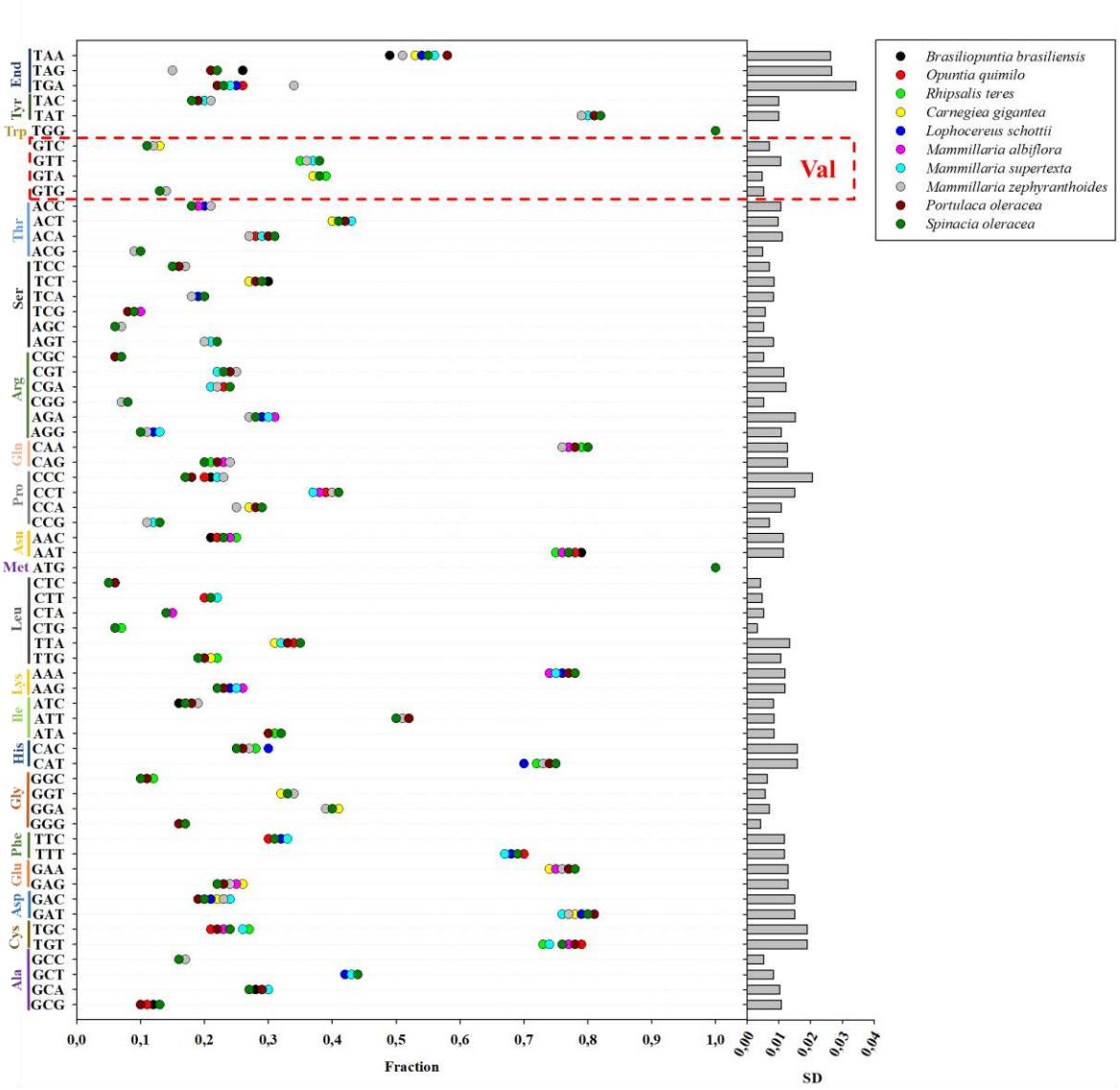


Figure 7. Comparison of the Codon usage between 10 species represented by balls with their respective colors (legend) in the graph. The x-axis represents the fractions (0 to 1) and the left most standard deviation (SD) in gray bars. The y-axis is representing the codons and a vertical line and the respective names of the amino acids are highlighted beside it.

8. TABLES

Table 1. Comparison of plastome size and GC content between *B. brasiliensis,* subfamily Opuntioideae and other species of Cactaceae.

Species	Subfamily	Size (bp)	LSC (bp)	SSC (bp)	IR (bp)	Number of genes	GC (%)	GenBank
Brasiliopuntia brasiliensis	Opuntioideae	162,211	87,186	4,393	35,316	109	36.8	OK448351
Opuntia quimilo	Opuntioideae	150,347	101,475	4,115	22,392	109	36.6	MN114084
Carnegiea gigantea	Cactoideae	113,064	-	-	-	99	36.7	NC_027618
Lophocereus schottii	Cactoideae	113,204	-	-	-	99	36.5	NC_041727
Mammillaria albiflora	Cactoideae	110,789	78,380	31,061	674	96	36.4	MN517610
Mammillaria pectinifera	Cactoideae	108,561	72,273	29,744	772	95	36.4	MN519716
Mammillaria crucigera	Cactoideae	115,505	71,565	29,418	7,261	96	36.3	MN517613
Mammillaria huitzilopochtli	Cactoideae	115,886	71,997	29,401	7,244	97	36.3	MN517612
Mammillaria solisioides	Cactoideae	115,356	71,690	29,238	7,214	95	36.4	MN518341
Mammillaria supertexta	Cactoideae	116,175	72,240	29,445	7,245	97	36.4	MN508963
Mammillaria zephyranthoides	Cactoideae	107,343	71,811	7,281	14,126	95	38.5	MN517611
Rhipsalis teres	Cactoideae	122,389	81,397	24,016	8,488	99	36.7	MT387452

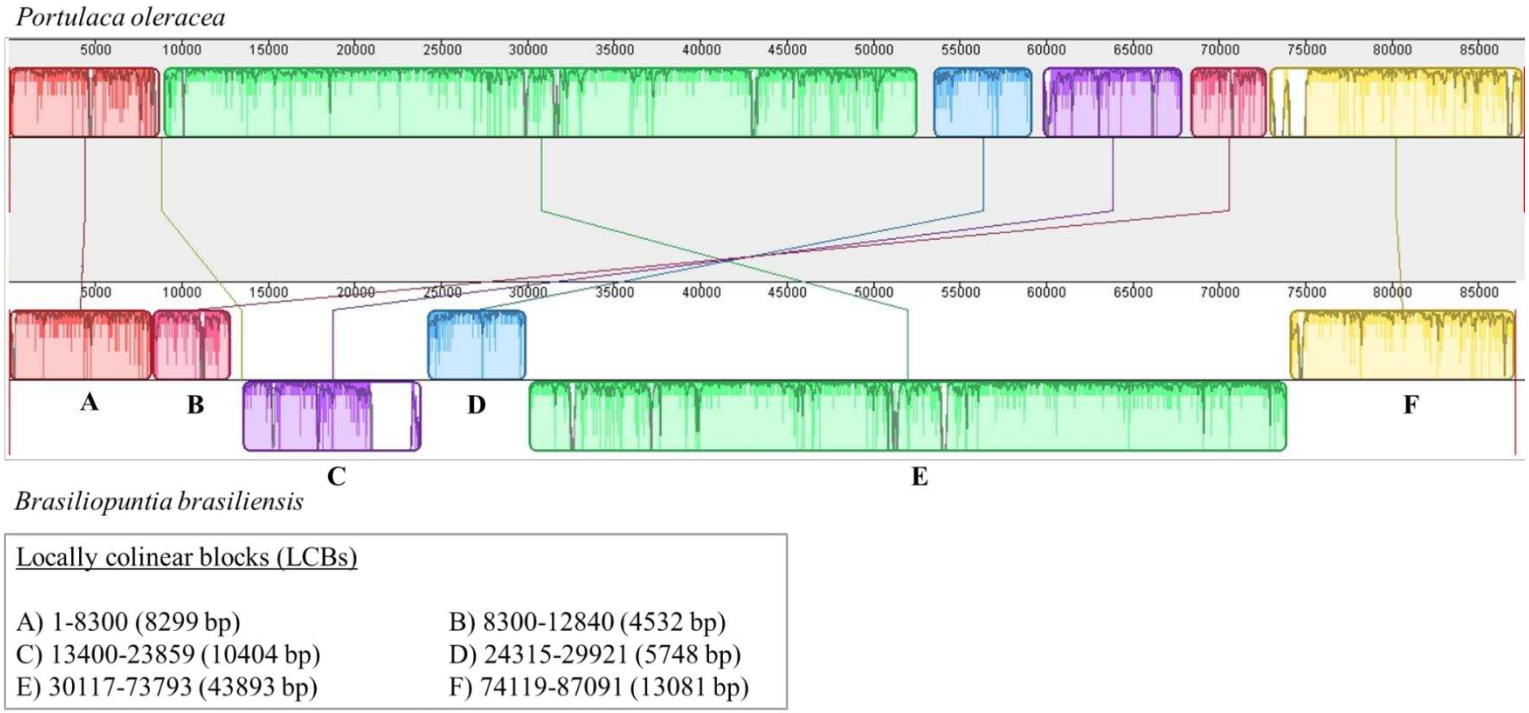
Table 2. List of genes identified in the plastome of *Brasiliopuntia brasiliensis*.

Group of gene	Name of gene
Gene expression machinery	
Ribosomal RNA genes	*rrn16*[b]*; rrn23*[b]; *rrn5*[b]*; rrn4.5*[b]
Transfer RNA genes	*trnA* –UGC[ab]; *trnC* –GCA; *trnD* –GUC; *trnE* –UUC; *trnF* –GAA; *trnfM* –CAU; *trnG* –UCC[a]; *trnG* –GCC; *trnH* –GUG; *trnI* –CAU[b]; *trnI* –GAU[ab]; *trnK* –UUU[a]; *trnL* –CAA[b]; *trnL* –UAA[a]; *trnL* –UAG; *trnM* –CAU; *trnN* –GUU[b]; *trnP* –UGG; *trnQ* –UUG; *trnR* –ACG[b]; *trnR* –UCU; *trnS* –GCU; *trnS* –UGA; *trnS* –GGA; *trnT* –UGU; *trnT* –GGU; *trnV* –GAC[b]; *trnW* –CCA; *trnY* –GUA
Small subunit of ribosome	*rps2; rps3; rps4*; *rps7*[b]*; rps8; rps11; rps12*[ac]; *rps14; rps15*[b]; *rps16*[a]; *rps18; rps19*[c]
Large subunit of ribosome	*rpl2*; *rpl14; rpl16*[a]; *rpl22; rpl23*[b]*; rpl32*[b]*; rpl33; rpl36*
DNA-dependent RNA polymerase	*rpoA; rpoB; rpoC1*[a]; *rpoC2*
Translational initiation factor	*infA*
Genes for photosynthesis	
Subunits of photosystem I (PSI)	*psaA; psaB; psaC; psaI; psaJ; ycf3*[a]; *ycf4*
Subunits of photosystem II (PSII)	*psbA*; *psbB; psbC; psbD; psbE; psbF; psbH; psbI; psbJ; psbK; psbL; psbM; psbN; psbT; psbZ*
Subunits of cytochrome b_6f complex	*petA; petB*[a]*; petD*[a]*; petG; petL; petN*
Subunits of ATP synthase	*atpA; atpB; atpE; atpF*[a]; *atpH; atpI*
Subunits of NADH dehydrogenase	*ndhA*[ab]; *ndhB*[ab]; *ndhC; ndhD; ndhE; ndhF*[b]*; ndhG*[c]*; ndhH*[b]; *ndhI*[b]*; ndhJ; ndhK*
Large subunit of Rubisco	*rbcL*
Others genes	
Maturase	*matK*
Envelope membrane protein	*cemA*
Subunit of acetyl-CoA carboxylase	*accD*
C-type cytochrome synthesis gene	*ccsA*
ATP-dependent Protease	*clpP*
Pseudogenes	*trnV* –UAC[a]; *rpl20; ycf1*[b]*; ycf2*[b]

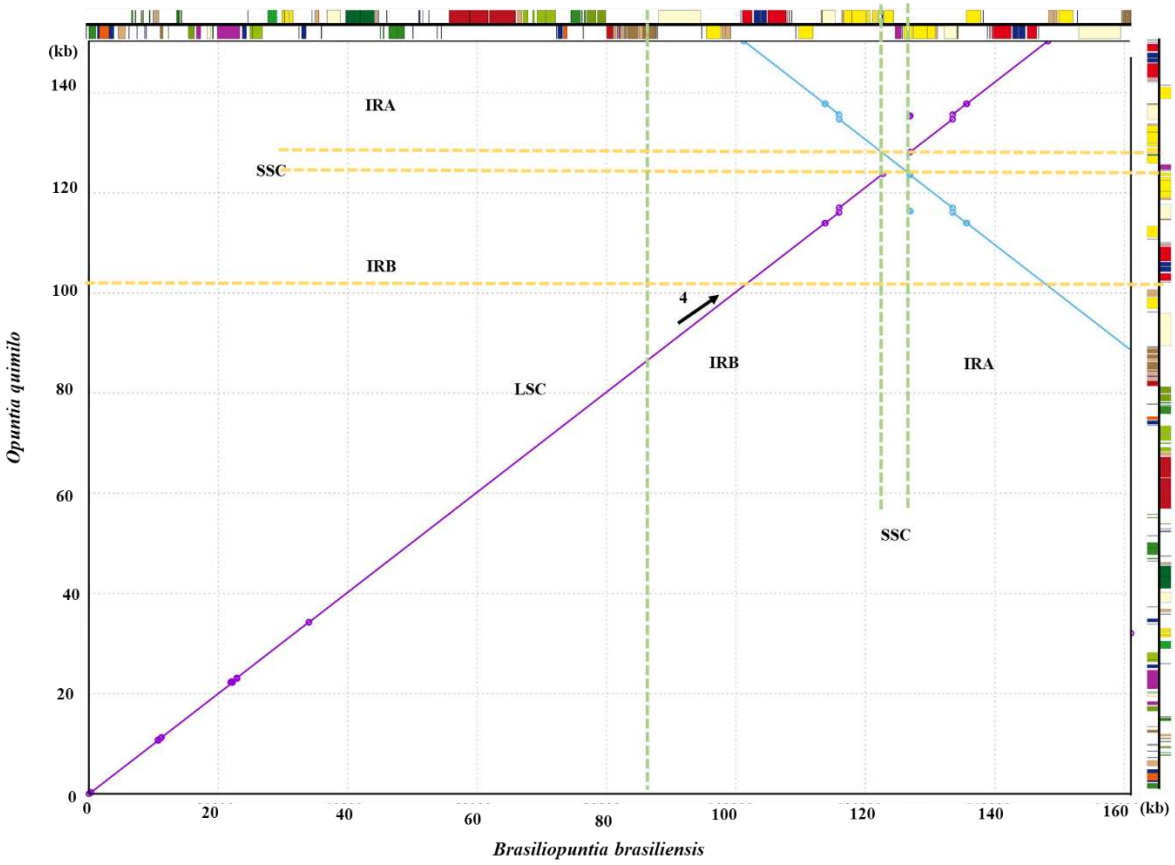
[a] Genes containing introns; [b] Duplicated gene, [c] Partially duplicated genes

9. SUPPLEMENTARY MATERIAL

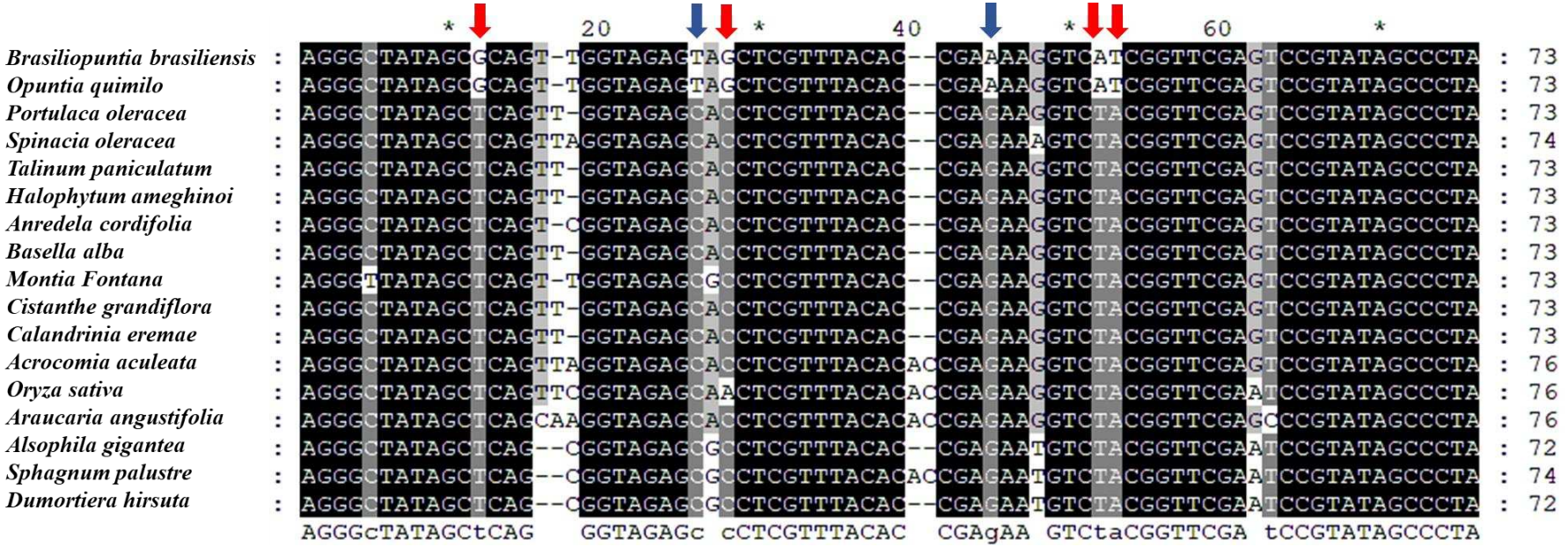
SUPPLEMENTARY FIGURES



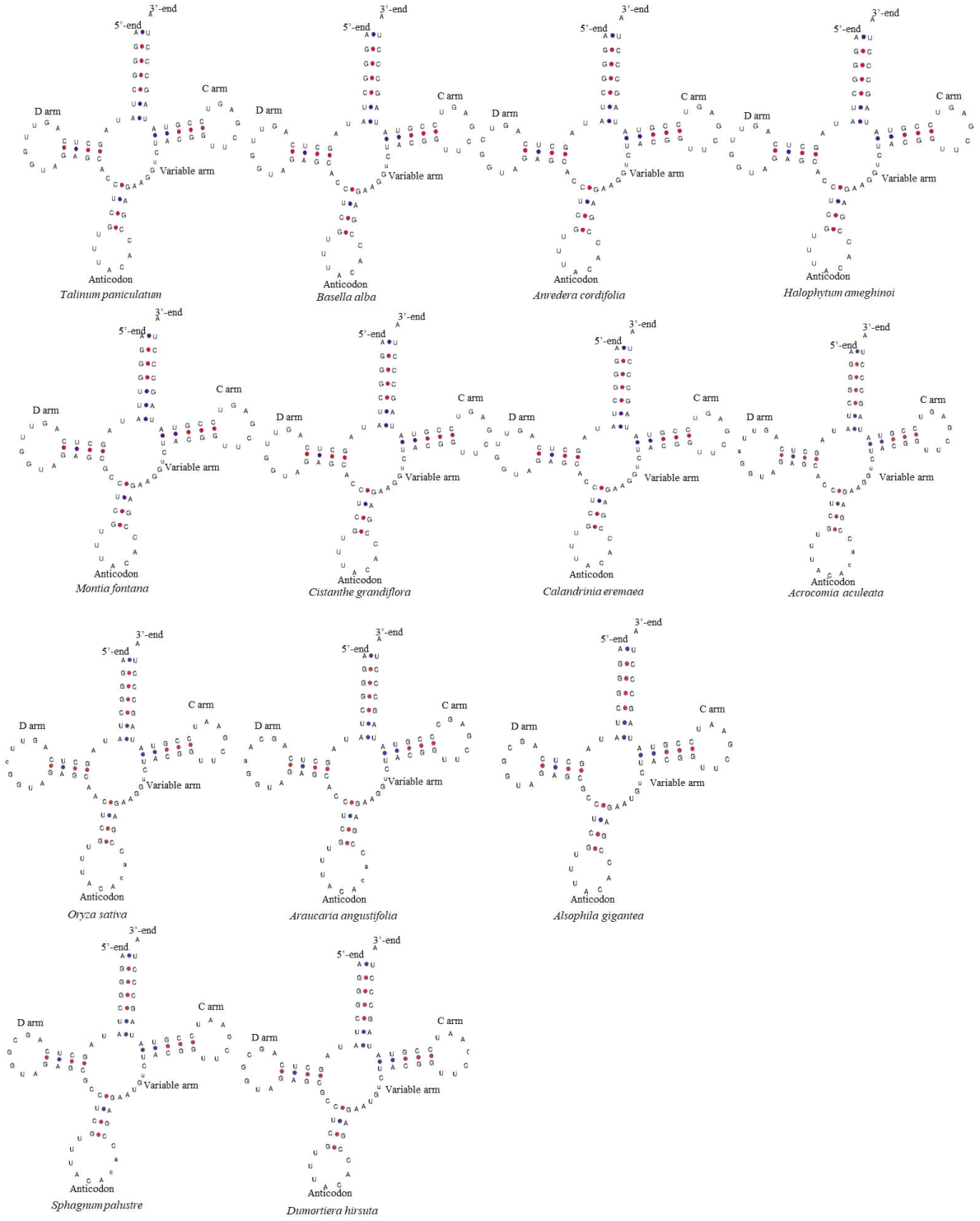
Supplementary Figure S1. Multiple alignment of *Brasiliopuntia brasiliensis* plastomes and *Portulaca oleracea* as reference. The region aligned corresponds region LSC of *P. oleracea* plastome, from the trnH-GUG to rps19 gene. Locally collinear blocks (LCBs) are color-coded. The LCBs related to inverted segments are drawn down. More details about the gene content of the LCBs involved in rearrangements are present in **Figure 3**



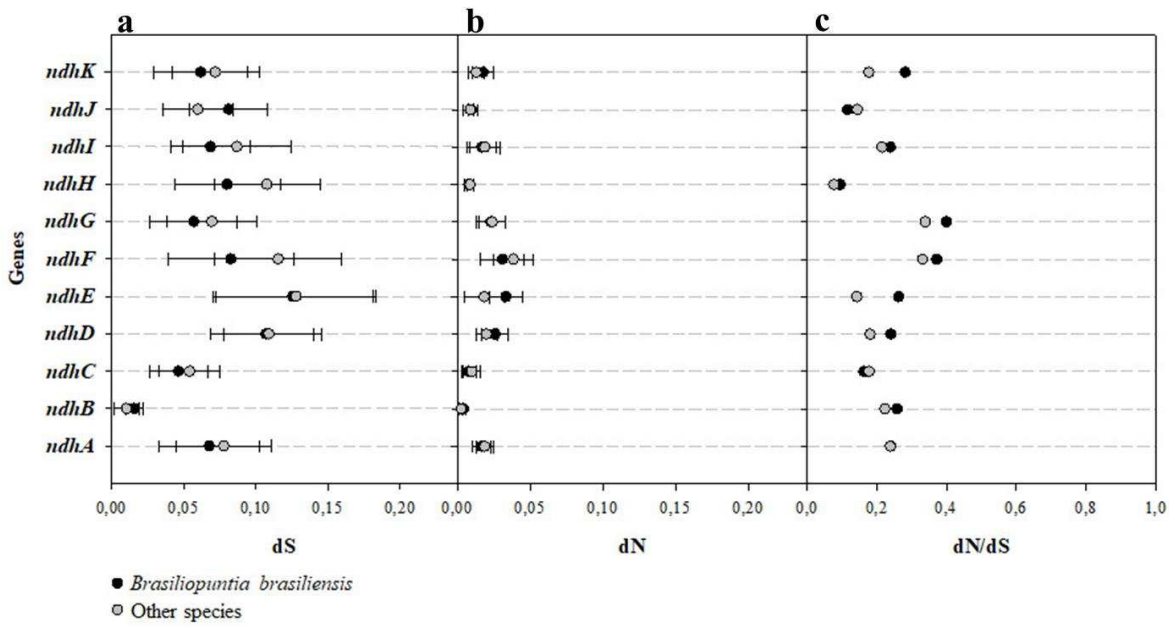
Supplementary Figure S2. Dot plot analyses comparing the plastomes of *Brasiliopuntia brasilienses* (x-axis) against *Opuntia quimilo* (y-axis). A positive slope denotes that the pair of sequences compared is in the same orientation. A negative slope denotes that the pair of sequences compared can be aligned, but their orientation is opposite. Sequences in the same direction are purple and inversions are blue. The black arrow (4) indicates contratrion in the region IRB-LSC. Vertical green lines indicates the regions SSC, LSC and IR's in *B. brasiliensis*, while in *O. quimilo* is indicates by yellow horizontal lines the same regions. More details about regions **Figure 2.**



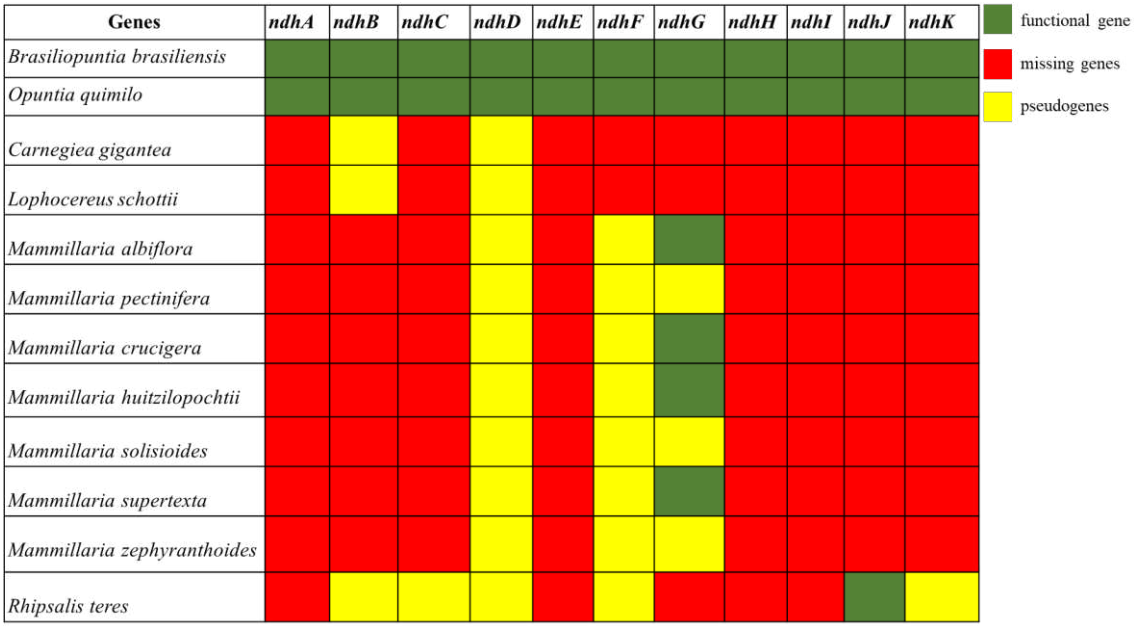
Supplementary Figure S3. Clustal W alignment of the primary structure of *trnV-UAC*, showing the change of bases (highlighted by red arrows, base transverse comparing *B. brasiliensis* and *O.quimilo* with other species and blue arrows base transition). Black shadowing indicates conserved sequence among all taxa in the alignment, dark gray indicates conserved sequence among all except for five taxon, light gray indicate conserved sequence among for eleven taxa, and white indicate non conserved nucleotide. The numbers above of line indicate the nucleotide position in the alignment. The numbers in the end of line indicates the total length of *trnV-UAC* in each species.



Supplementary Figure S4. Secondary structure of the *trnV-UAC* of thirteen species, of these 7 dicots, 2 monocots, 1 gymnosperm, 1 pteridophyte, 1 bryophyte and 1 Marchantiophyta. The structure consists of 3 arms, 2 of which are non-variable (C-arm and D-arm), and arm where the anticondon is located. A variable portion (Variable arm). The reading pattern is also indicated 5'end (start) and 3'end stop.



Supplementary Figure S5. Synonymous (dS), nonsynononymous (dN), and dN/dS values of 11 plastid- protein-coding genes of *ndh* complex. *B. brasiliensis* is represented by black circles and the mean the other species (see **Supplementary Table S2**) is represented by light gray circles.



Supplementary Figure S6. Relationship of the *ndh* complex in the Cactaceae family, highlighted in the red blocks missing genes, pseudogene in the yellow block and functional gene in green.

SUPPLEMENTARY TABLE

Supplementary Table S1. List of the taxa sampled in the phylogenetic analysis. (*) Outgroup.

Species	**Family**	**GenBank**
Brasiliopuntia brasiliensis	Cactaceae	
Opuntia quimilo	Cactaceae	MN114084
Rhipsalis teres	Cactaceae	MT387452
Carnegiea gigantea	Cactaceae	NC_027618.1
Lophocereus schottii	Cactaceae	NC_041727.1
Mammillaria albiflora	Cactaceae	MN517610
Mammillaria pectinifera	Cactaceae	MN519716
Mammillaria crucigera	Cactaceae	MN517613
Mammillaria huitzilopochtli	Cactaceae	MN517612
Mammillaria solisioides	Cactaceae	MN518341
Mammillaria supertexta	Cactaceae	MN508963
Mammillaria zephyranthoides	Cactaceae	MN517611
Anredera cordifolia	Basellaceae	NC_041274.1
Basella alba	Basellaceae	NC_041293.1
Calandrinia eremaea	Montiaceae	NC_041259.1
Calandrinia granulifera	Montiaceae	NC_041260.1
Cistanthe grandiflora	Montiaceae	NC_041295.1
Cistanthe longiscapa	Montiaceae	NC_035140.1
Montia fontana	Montiaceae	NC_041269.1
Halophytum ameghinoi	Halophytaceae	NC_040949.1
Portulaca grandiflora	Portulacaceae	NC_041299.1
Portulaca oleracea	Portulacaceae	NC_036236.1
Portulaca pilosa	Portulacaceae	NC_041264.1
Talinum paniculatum	Talinaceae	NC_037748.1
*Spinacia oleracea**	Chenopodiaceae	NC_002202.1

Supplementary Table S2. List of the taxa sampled in the non-synonymous analysis.

Species	Family	GenBank
Brasiliopuntia brasiliensis	Cactaceae	
Opuntia quimilo	Cactaceae	MN114084
Anredera cordifolia	Basellaceae	NC_041274.1
Basella alba	Basellaceae	NC_041293.1
Calandrinia eremaea	Montiaceae	NC_041259.1
Calandrinia granulifera	Montiaceae	NC_041260.1
Cistanthe grandiflora	Montiaceae	NC_041295.1
Montia fontana	Montiaceae	NC_041269.1
Portulaca grandiflora	Portulacaceae	NC_041299.1
Portulaca oleracea	Portulacaceae	NC_036236.1
Portulaca pilosa	Portulacaceae	NC_041264.1
Talinum paniculatum	Talinaceae	NC_037748.1

Supplementary Table S3. List of the taxa sampled in the codon usage analysis

Species	Family	Quantity of coding genes used
Brasiliopuntia brasiliensis	Cactaceae	76
Opuntia quimilo	Cactaceae	76
Rhipsalis teres	Cactaceae	64
Carnegiea gigantea	Cactaceae	63
Lophocereus schottii	Cactaceae	63
Mammillaria albiflora	Cactaceae	62
Mammillaria supertexta	Cactaceae	63
Mammillaria zephyranthoides	Cactaceae	58
Portulaca oleracea	Portulacaceae	78
Spinacia oleracea	Chenopodiaceae	78

Supplementary Table S4. List of simple sequence repeats (SSRs) identified in the plastome of *B. brasiliensis*.

SSR sequence	**Number of repetitions**															**Total**
	3	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	
A/T	-	-	-	-	-	79	39	20	10	8	4	3	-	3	2	168
C/G	-	-	-	-	-	1	2	1	-	1	-	-	-	-	-	5
AG/CT	-	10	-	-	-	-	-	-	-	-	-	-	-	-	-	10
AT/AT	-	22	7	2	-	-	-	1	1	-	-	-	-	-	-	33
AAT/ATT	-	3	1	-	-	-	-	-	-	-	-	-	-	-	-	4
AAAC/GTTT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AAAG/CTTT	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5
AAAT/ATTT	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
AATC/ATTG	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
ACAT/ATGT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ACCT/AGGT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ATCC/ATGG	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AAAAC/GTTTT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AAAGT/ACTTT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AATAT/ATATT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Total																**236**

Supplementary Table S5. Distribution of SSR loci in the *B. brasiliensis* plastome.

SSRtype	SSR	Size	Start	End	Location
mono	(A)13	13	234	246	*trnH-GUG*/*psbA* (IGS)
mono	(A)16	16	350	365	*trnH-GUG*/psbA (IGS)
mono	(T)9	9	2001	2009	*trnK-UUU* (intron)
mono	(A)8	8	2586	2593	trnK-UUU (intron)/*matK* (CDS)
mono	(T)9	9	3237	3245	trnK-UUU (intron)/*matK* (CDS)
mono	(A)8	8	3621	3628	trnK-UUU (intron)/*rps16* (IGS)
mono	(T)8	8	3974	3981	*trnK-UUU* (intron)/*rps16* (IGS)
mono	(T)14	14	4544	4557	*trnK-UUU*/*rps16* (IGS)
mono	(A)8	8	4607	4614	trnK-UUU/*rps16* (IGS)
tri	(TAA)4	12	4765	4776	*trnK-UUU*/rps16 (IGS)
mono	(A)9	9	4850	4858	*trnK-UUU*/*rps16* (IGS)
mono	(T)10	10	4957	4966	*trnK-UUU*/*rps16* (IGS)
mono	(A)9	9	5426	5434	*rps16* (intron)
di	(AT)4	8	5727	5734	*rps16* (intron)
di	(AT)4	8	6207	6214	*rps16*/*trnQ-UUG* (IGS)
mono	(T)13	13	6229	6241	*rps16*/*trnQ-UUG* (IGS)
mono	(A)8	8	6256	6263	*rps16*/*trnQ-UUG* (IGS)
mono	(A)12	12	6967	6978	trnQ-UUG/psbK (IGS)
mono	(A)8	8	7475	7482	psbK/psbI (IGS)
compound	(A)10(AAT)5*	23	7574	7596	psbK/psbI (IGS)
mono	(A)8	8	7779	7786	*psbI*/trnS-GCU (IGS)
di	(TA)4	8	7817	7824	*psbI*/trnS-GCU (IGS)
di	(AT)4	8	8084	8091	trnS-GCU/petL (IGS)
di	(AT)5	10	8239	8248	trnS-GCU/petL (IGS)
mono	(T)9	9	8316	8324	trnS-GCU/petL (IGS)
mono	(A)11	11	8509	8519	trnS-GCU/petL (IGS)
mono	(T)12	12	8802	8813	petL/petG (IGS)
mono	(A)8	8	9249	9256	trnW-CAA/trnP-UGG(IGS)
di	(AT)4	8	9546	9553	trnW-CAA/trnP-UGG(IGS)
mono	(A)9	9	9628	9636	trnW-CAA/trnP-UGG(IGS)
mono	(A)8	8	9793	9800	trnW-CAA/trnP-UGG(IGS)
mono	(T)10	10	10748	10757	rpl33/rps18 (IGS)
mono	(T)9	9	11456	11464	rps18/rpl20 (IGS)
mono	(T)8	8	11480	11487	rps18/rpl20 (IGS)
mono	(T)8	8	11900	11907	rpl20 (CDS)
mono	(T)8	8	12114	12121	rpl20/rps12 (IGS)
di	(AT)5	10	12870	12879	rps12 (intron)
di	(TA)4	8	13268	13275	rps12/psbE (IGS)
tetra	(ATCA)3	12	13379	13390	rps12/psbE (IGS)
mono	(A)8	8	13539	13546	rps12/psbE (IGS)
mono	(T)9	9	15197	15205	psbJ/psbA (IGS)
mono	(A)16	16	15322	15337	psbJ/psbA (IGS)
mono	(C)9	9	15478	15486	psbJ/psbA (IGS)
mono	(A)12	12	15662	15673	psbJ/psbA (IGS)
mono	(A)8	8	15676	15683	psbJ/psbA (IGS)
mono	(T)8	8	16417	16424	petA (CDS)
di	(TA)4	8	16776	16783	petA (CDS)

mono	(T)12	12	17808	17819	cemA (CDS)/ycf4 (IGS)
mono	(T)8	8	17939	17946	cemA/ycf4 (IGS)
penta	(CAAAA)3	15	18275	18289	cemA/ycf4 (IGS)
mono	(T)9	9	18519	18527	cemA/ycf4 (IGS)
mono	(A)10	10	18615	18624	cemA/ycf4 (IGS)
mono	(A)8	8	19162	19169	ycf4 (CDS)
mono	(G)8	8	19179	19186	ycf4 (CDS)
mono	(T)11	11	19799	19809	psaI/accD (IGS)
di	(TA)4	8	19883	19890	psaI/accD (IGS)
mono	(A)8	8	20003	20010	psaI/accD (IGS)
mono	(T)8	8	20121	20128	psaI/accD (IGS)
mono	(A)8	8	20151	20158	psaI/accD (IGS)
mono	(T)8	8	20174	20181	psaI/accD (IGS)
di	(AT)4	8	20230	20237	psaI/accD (IGS)
di	(AT)4	8	20656	20663	accD (CDS)
mono	(T)8	8	20669	20676	accD (CDS)
mono	(A)8	8	22562	22569	accD (CDS)
mono	(A)8	8	23076	23083	accD (CDS)
mono	(T)11	11	24336	24346	trnV-UAC (intron)
mono	(T)8	8	24790	24797	trnV-UAC (intron)
mono	(T)8	8	24930	24937	trnV-UAC/trnM-CAU (IGS)
di	(TA)4	8	25225	25232	trnM-CAU/ atpE (IGS)
mono	(T)9	9	25288	25296	trnM-CAU/ atpE (IGS)
mono	(T)10	10	27284	27293	atpB (CDS)
mono	(A)9	9	27604	27612	atpB/rbcL (IGS)
di	(AT)4	8	27964	27971	atpB/rbcL (IGS)
di	(GA)4	8	28643	28650	rbcL(CDS)
mono	(T)9	9	30121	30129	rbcL/ndhC(IGS)
mono	(T)8	8	31522	31529	ndhK/ndhJ(IGS)
mono	(A)9	9	31573	31581	ndhK/ndhJ(IGS)
tetra	(AAAT)3	12	31619	31630	ndhK/ndhJ(IGS)
mono	(G)9	9	32418	32426	ndhJ/trnF-GAA (IGS)
mono	(C)10	10	32480	32489	ndhJ/trnF-GAA (IGS)
mono	(C)12	12	32593	32604	ndhJ/trnF-GAA (IGS)
mono	(A)10	10	32803	32812	ndhJ/trnF-GAA (IGS)
di	(AT)4	8	33749	33756	trnL-UAA(intron)
di	(TA)5	10	33799	33808	trnL-UAA(intron)
mono	(T)9	9	33974	33982	trnL-UAA(intron)
mono	(T)11	11	34013	34023	trnL-UAA(intron)
mono	(T)12	12	34171	34182	trnL-UAA/trnT-UGU(IGS)
di	(AT)4	8	34431	34438	trnL-UAA/trnT-UGU(IGS)
mono	(T)10	10	34606	34615	trnL-UAA/trnT-UGU(IGS)
mono	(A)8	8	34897	34904	trnL-UAA/trnT-UGU(IGS)
mono	(T)9	9	34938	34946	trnL-UAA/trnT-UGU(IGS)
mono	(T)13	13	35127	35139	trnT-UGU/rps4(IGS)
di	(AT)4	8	36371	36378	rps4/trnS-GAA(IGS)
tetra	(TGAT)3	12	36728	36739	trnS-GAA/ycf3(IGS)
mono	(A)9	9	37231	37239	trnS-GAA/ycf3(IGS)
mono	(T)17	17	37545	37561	ycf3(intron)
mono	(A)8	8	38641	38648	ycf3(intron)

mono	(A)8	8	38776	38783	ycf3(intron)
di	(TA)4	8	38872	38879	ycf3(intron)
di	(TA)4	8	39723	39730	ycf3/psaA(IGS)
mono	(T)8	8	39757	39764	ycf3/psaA(IGS)
di	(AT)4	8	39891	39898	ycf3/psaA(IGS)
mono	(A)8	8	40156	40163	ycf3/psaA(IGS)
tetra	(TTTC)3	12	44793	44804	psaB(CDS)/rps14(IGS)
di	(AT)6	12	45761	45772	trnG-GCC/psbZ(IGS)
di	(CT)4	8	45946	45953	trnG-GCC/psbZ(IGS)
di	(TA)4	8	46489	46496	psbZ/trnS-UGA(IGS)
di	(CT)4	8	46719	46726	trnS-UGA(CDS)
mono	(A)8	8	46827	46834	trnS-UGA/psbC(IGS)
mono	(A)11	11	46866	46876	trnS-UGA/psbC(IGS)
mono	(T)9	9	46896	46904	trnS-UGA/psbC(IGS)
mono	(A)9	9	49553	49561	psbD/trnT-GGU(IGS)
tetra	(AAGA)3	12	49908	49919	psbD/trnT-GGU(IGS)
mono	(A)10	10	50345	50354	psbD/trnT-GGU(IGS)
tetra	(TACA)3	12	50847	50858	trnT-GGU/trnE-UUC(IGS)
di	(AT)6	12	51096	51107	trnT-GGU/trnE-UUC(IGS)
di	(AT)5	10	51110	51119	trnT-GGU/trnE-UUC(IGS)
di	(TA)4	8	51139	51146	trnT-GGU/trnE-UUC(IGS)
tri	(TAT)4	12	51369	51380	trnT-GGU/trnE-UUC(IGS)
tri	(ATA)4	12	51399	51410	trnT-GGU/trnE-UUC(IGS)
di	(AT)5	10	52310	52319	trnD-GUC/psbM(IGS)
mono	(T)8	8	52667	52674	trnD-GUC/psbM(IGS)
mono	(T)16	16	52941	52956	trnD-GUC/psbM(IGS)
mono	(T)8	8	53461	53468	psbM/petN(IGS)
mono	(A)8	8	53770	53777	psbM/petN(IGS)
mono	(T)8	8	54747	54754	petN/trnC-GCA(IGS)
mono	(T)11	11	55101	55111	petN/trnC-GCA(IGS)
mono	(T)8	8	55244	55251	trnC-GCA/rpoB(IGS)
di	(AT)4	8	55343	55350	trnC-GCA/rpoB(IGS)
mono	(A)9	9	55871	55879	trnC-GCA/rpoB(IGS)
mono	(T)10	10	56086	56095	trnC-GCA/rpoB(IGS)
mono	(A)8	8	57094	57101	rpoB(CDS)
mono	(T)9	9	60256	60264	rpoC1(intron)
mono	(T)8	8	60478	60485	rpoC1(intron)
mono	(T)8	8	60914	60921	rpoC1(CDS)
mono	(T)9	9	62844	62852	rpoC2(CDS)
di	(AT)5	10	63362	63371	rpoC2(CDS)
mono	(A)8	8	64195	64202	rpoC2(CDS)
mono	(A)9	9	64674	64682	rpoC2(CDS)
mono	(A)8	8	64846	64853	rpoC2(CDS)
mono	(T)8	8	65496	65503	rpoC2(CDS)
mono	(A)9	9	66561	66569	rpoC2(CDS)
mono	(T)11	11	69274	69284	atpI/atpH(IGS)
mono	(T)8	8	70167	70174	atpF(intron)
mono	(T)8	8	70485	70492	atpF(intron)
mono	(A)14	14	70582	70595	atpF(intron)
mono	(T)14	14	70665	70678	atpF(intron)

mono	(T)10	10	71309	71318	atpF/atpA(IGS)
mono	(A)9	9	73708	73716	trnG-UCC(CDS)
mono	(T)9	9	74052	74060	trnG-UCC/clpP(IGS)
mono	(T)8	8	74746	74753	clpP/psbB(IGS)
mono	(T)8	8	74768	74775	clpP/psbB(IGS)
mono	(A)11	11	74928	74938	clpP/psbB(IGS)
mono	(T)8	8	76769	76776	psbB/psbT(IGS)
mono	(T)8	8	76833	76840	psbB/psbT(IGS)
mono	(T)8	8	77997	78004	petB(intron)
mono	(A)8	8	78182	78189	petB(intron)
mono	(A)11	11	79340	79350	petB/petD(IGS)
di	(AG)4	8	79481	79488	petD(intron)
mono	(T)10	10	79625	79634	petD(intron)
di	(AT)11	22	79779	79800	petD(intron)
di	(GA)4	8	81680	81687	rpoA(CDS)
mono	(T)9	9	82345	82353	rps11/rpl36(IGS)
mono	(T)9	9	83462	83470	rps8/rpl14(IGS)
mono	(A)9	9	83478	83486	rps8/rpl14(IGS)
mono	(T)8	8	83597	83604	rpl14(CDS)
tetra	(CAAA)3	12	83954	83965	rpl14/rpl16(IGS)
di	(TA)5	10	84804	84813	rpl16(intron)
tetra	(TTTC)3	12	85309	85320	rpl16(intron)
mono	(T)17	17	85366	85382	rpl16(intron)
mono	(A)9	9	85419	85427	rpl16(intron)
mono	(T)8	8	85558	85565	rpl16(intron)
mono	(T)9	9	87320	87328	rps19/rpl2(IGS)
tetra	(AATA)3	12	88765	88776	trnI-CAU/ycf2(IGS)
mono	(A)12	12	88789	88800	trnI-CAU/ycf2(IGS)
di	(GA)4	8	88898	88905	ycf2(CDS)
di	(GA)4	8	89852	89859	ycf2(CDS)
mono	(A)9	9	91776	91784	ycf2(CDS)
mono	(T)8	8	93650	93657	ycf2(CDS)
mono	(T)9	9	95985	95993	trnL-CAA/ndhB(IGS)
di	(AG)4	8	96617	96624	ndhB(CDS)
mono	(T)9	9	99038	99046	rps7(CDS)
mono	(A)8	8	99505	99512	rps12(intron)
mono	(T)9	9	101230	101238	rps12/trnV-GAC(IGS)
mono	(T)8	8	104121	104128	trnI-GAU(intron)
mono	(T)8	8	105452	105459	trnA-UGC(intron)
tetra	(AGGT)3	12	107468	107479	rrn23s(CDS)
di	(CT)4	8	107506	107513	rrn23s(CDS)
di	(GA)4	8	107830	107837	rrn23s(CDS)
mono	(A)10	10	109258	109267	rrn5s/trnR-ACG(IGS)
mono	(T)8	8	109659	109666	trnR-ACG/trnN-GUU(IGS)
penta	(TATAA)3	15	110481	110495	trnN-GUU/ndhF(IGS)
mono	(A)13	13	110532	110544	trnN-GUU/ndhF(IGS)
mono	(A)8	8	111280	111287	ndhF(CDS)
mono	(T)8	8	111403	111410	ndhF(CDS)
mono	(A)8	8	112898	112905	ndhF/rpl32(IGS)
di	(AT)4	8	113157	113164	ndhF/rpl32(IGS)

mono	(T)12	12	113388	113399	ndhF/rpl32(IGS)
penta	(TTACT)3	15	113449	113463	ndhF/rpl32(IGS)
mono	(A)9	9	113484	113492	ndhF/rpl32(IGS)
mono	(A)10	10	113603	113612	ndhF/rpl32(IGS)
di	(AT)10	20	113874	113893	ndhF/rpl32(IGS)
mono	(A)8	8	114128	114135	rpl32(CDS)
mono	(A)10	10	114210	114219	rpl32(CDS)
mono	(A)8	8	114221	114228	rpl32(CDS)
mono	(A)8	8	114315	114322	ycf1(CDS)
mono	(T)8	8	114786	114793	ycf1(CDS)
mono	(A)8	8	115316	115323	ycf1(CDS)
mono	(A)9	9	115537	115545	ycf1(CDS)
mono	(A)9	9	115560	115568	ycf1(CDS)
mono	(A)9	9	116238	116246	ycf1(CDS)
mono	(A)10	10	116248	116257	ycf1(CDS)/rps15(IGS)
mono	(A)8	8	116262	116269	ycf1/rps15(IGS)
mono	(A)10	10	117227	117236	ycf1/rps15(IGS)
tetra	(CCAT)3	12	117692	117703	ndH(CDS)
tetra	(TCTT)3	12	119457	119468	ndhA(intron)
tetra	(CTTT)3	12	119959	119970	ndhA(intron)
mono	(A)8	8	120027	120034	ndhA(intron)
mono	(T)9	9	120107	120115	ndhA(intron)
mono	(A)8	8	120124	120131	ndhA(intron)
mono	(T)8	8	120166	120173	ndhA(intron)
mono	(A)8	8	121018	121025	ndhA/ndhI(IGS)
mono	(T)8	8	122687	122694	ndhG/ndhE(IGS)
mono	(T)12	12	123111	123122	ndhE/psaC(IGS)
mono	(T)11	11	123272	123282	ndhE/psaC(IGS)
mono	(A)10	10	123417	123426	ndhE/psaC(IGS)
mono	(T)8	8	123724	123731	psaC/ndhD(IGS)
mono	(A)8	8	124707	124714	ndhD(CDS)
mono	(T)8	8	125057	125064	ndhD(CDS)
mono	(T)8	8	125292	125299	ndhD(CDS)
mono	(A)8	8	125335	125342	ndhD/ccsA(IGS)
mono	(A)10	10	125345	125354	ndhD/ccsA(IGS)
mono	(T)8	8	126218	126225	ccsA(CDS)
mono	(A)10	10	126508	126517	ccsA/trnL-UAG(IGS)
mono	(T)10	10	126769	126778	trnL-UAG/ndhG(IGS)

CDS, coding sequences; IGS, intergenic spacers

CHAPTER II: The plastomes of two *Opuntia*, *Opuntia monacantha* Haw and *Opuntia ficus-indica* (L.) Mill. reveal little gene loss and unique rearrangements within Cactaceae

ABSTRACT

Opuntia arose ca.6-8 million years ago (MA) and comprises around 220-350 species, with the center of origin is South America. *Opuntia* is cultivated worldwide due to its economic importance, being used in food, animal fodder, ornamental purposes, and medicinal purposes. We sequenced and characterized in detail the plastomes of *Opuntia monacantha* and *Opuntia ficus-indica*. Aiming to understand the evolutionary patterns in Cactaceae and in the suborder Cactineae, we conducted extensive analyzes with the plastid genes. The analyzes showed genes with high divergence (*accD*, *clpP*, *infA*, *rpl22*, *rps18*) and we also found that 49 of 79) plastid genes under positive selection. A total of 1758 nucleotides are under positive selection and the high divergence genes mentioned above with are responsible for more than half of these nucleotide sites. Moreover, we predicted 68 RNA editing sites in Opuntioideae, and mapped 240 and 253 simple sequence repeats (SSR) to *O. monacantha* and *O. ficus-indica*, respectively. Furthermore, we performed a phylogeny analysis based on 54 concatenated genes of 30 taxa, representing the suborder Cactineae, in which *Opuntia* formed a monophyletic group. The data presented here is important for genetic studies in Cactaceae and provide an elucidation of the evolution on plastid genes.

Keywords**:** Prickly pear, Opuntioideae, Positive selection, Plastome evolution, Expansion LSC, Expansion IRs

1. INTRODUCTION

The Cactaceae family contains approximately 1400 species (Guerrero et al. 2019). *Opuntia* comprises about 220-350 species (Britton and Rose 1919; Anderson 2001), and has South America as its center of origin, having diversified thought the Miocene a 6-8 million years ago (Majure et al. 2012; Arakaki et al.2011). *Opuntia* commonly known prickly pear, belong to subfamily Opuntioideae, and is cultivated worldwide due to its economic importance. It is used in food, animal fodder, for ornamental and medicinally purposes, (Stintzing and Carle 2005; Grünwaldt et al. 2015; Santiago et al. 2018; Feugang et al. 2006; Yang et al. 2008).

In the semiarid region of Brazil and in Mexico, *Opuntia ficus-indica* is popularly known as *palma forrageira* and Nopal (*figueira da india*), respectively. In those locations, harvest takes place biannually (Silva et al. 2015; Dieleman et al. 2017). The fruits have a high content of antioxidants and vitamins, adding significantly to the nutrition of animals and humans (Lucena et al. 2013; Dieleman et al. 2017). Also, the species has high ecological importance, since it is home to several transitional animal and plant species (Oliveira et al. 1999).

In Brazil, *Opuntia monacantha* occurs along the south and southeast coast, in the Atlantic Forest biome, which is among the biome most fragmented, degraded, and threatened by human activities, such as tourist activity and agricultural expansion. This species has been widely studied because its bioactive compounds and antioxidants, wich have high efficiency in fighting free radicals (Bauer and Waechter 2006; Zappi et al. 2011; Valente et al. 2010).

Generally, superior plants have a plastome with a typical quadripartite structure, with conserved genes, divided into two groups. The first group is linked to the photosynthetic apparatus (photosystem I and II, ATP synthase, and complex cytochrome b6f). The second group is related to gene expression in the ribosomes subunit, tRNAs, rRNAs, and RNA polymerase (Daniell et al. 2016; Allen 2015; Rogalski et al. 2015).

The structure, order, and gene content are generally conserved in angiosperms, but there are reports of groups that present highly rearranged plastomes, gene losses, wich involve genes considered essential (e.g., *ycf1*, *ycf2*, *trnV-UAC*) (Kikuchi et al. 2013 and 2018; Alkatib et al 2012). Some of these groups are Fabaceae (Schwarz et al. 2015), Campunalaceae (Haberle et al. 2008), Geraniaceae (Weng et al. 2014), Passifloraceae (Rabah et al. 2019; Pacheco et al. 2020), and Cactaceae (Sanderson et al. 2015; Solórzano et al. 2019; Oulo et. al 2020, Morais da Silva et al. 2021; Köhler et al. 2020), a family that is the object of study in this work.

Obtaining the complete plastomes of the species, we can perform several analyses, including the study of population genetics through molecular markers mapped and listed for the group, and phylogenetic inferences at various taxonomic levels (Rogalski et al. 2015; Quiao et al. 2016). Furthermore, understanding the evolution of plastid genes through the prediction of RNA editing sites, positive selection, gene divergence, evaluating functionality, degeneration, and/or gene loss (Lopes et al. 2018ab; Pacheco et al. 2019 and 2020).

Here, we report the complete plastomes of *Opuntia monacantha* and *Opuntia ficus-indica*, we characterize in detail the structures identifying unique rearrangements and translocations in Opuntioideae (expansion of regions like IR and LSC). With the complete plastome sequenced we map the simple sequence repeat (SSR) and use the protein-encoding

genes we predict possible RNA editing sites. We inferred the gene divergence and positive selection in addition to the phylogenetic position of these species in the suborder Cactineae.

2. MATERIAL AND METHODS

Plant material and plastid DNA extraction, sequencing, assembling and annotation

Opuntia monacantha and *Opuntia ficus-indica* fresh and young cladodes were collected at Universidade Federal de Viçosa, Viçosa-MG, Brazil and kept on dark for 1 week at 4 °C to decrease the starch content. After that, the chloroplast isolation and plastid DNA extraction were performed according to Vieira et al. (2014).

Plastome sequencing, assembling, and annotation

Around 1 ng of plastid DNA was used to prepare sequencing libraries with Nextera XT DNA Sample Prep Kit (Illumina Inc., San Diego, CA), according to the manufacturer's instructions. Obtaining the library was sequencing using Illumina MiSeq platform (Illumina Inc., San Diego, California, USA) at the Universidade Federal do Paraná, Brazil. The reads obtained were trimmed (threshold with a probability of error <0.05) and de novo assembled in contigs using CLC genomics Workbench 8.0.2 software (CLC Bio, Aarthus, Denmark). The sequencing of *O. monacantha* and *O. ficus-indica* resulted, respectively, in 2,283 reads of average coverage 1,152.56 and 3,845 reads of average coverage 451.45.

The contigs used for assembling the plastomes varied from 631.39 to 1,156.59 for *O. monacantha* and from 926.88 to 240.02 for *O. ficus-indica*, average coverage. Additionally, a gap of 247 bp in *O.ficus-indica* was closed with a contig of 118299 size and 51.27 coverage, using the SPAdes 3.13.0 software (Bankevich et al. 2012), based on the pre-assembled genome and the aid of another sequencing.

For annotation of genes, we use Annotation of Organellar Genomes (Tillich et al. 2017) , Dual Organellar GenoMe Annotator (DOGMA) (Wyman et al. 2004) and BLAST searches. From this initial annotation, putative start codons, stop codons, and intron positions were determined based on comparisons to homologous genes in other plastid genomes at the GenBank database. All tRNA genes were further verified by using tRNAscan-SE (Lowe and Chan 2016). The physical map of the plastid circular genome was drawn using Organellar

Genome DRAW (OGDRAW) version 1.3.1 (Greiner et al. 2019) . The complete nucleotide sequences and gene features of the plastomes of *O. monacantha* and *O. ficus-indica* were deposited in the GenBank database under accession number MZ579523 and OK448352, respectively.

Comparative analysis of plastome structure Opuntia

The LSC region of *O. monacantha*, *O. ficus-indica*, *O. quimilo*, and *Portulaca oleracea*, specie available in the database had a typical plastome organization, were compared through multiple alignment using the software Mauve Genome Alignment v2.3.1 (MAUVE) (Darling et al. 2004). After that, the MUMmer nucleotide script (NUCmer) Perl from the MUMmer 3.23 software was used to compare the complete plastome of *O. monacantha* and *O. ficus-indica* with that of *P. oleracea* (Kurtz et al. 2004). The linear maps of genes were drawn by using OGDRAW for identification of possible rearrangements (Greiner et al. 2019) .

Phylogenetic inference

The inference of phylogenetic position of *O. monacantha* and *O. ficus-indica* inside the suborder Cactineae was realized based on 54 plastid genes. *Spinacia oleracea* (Chenopodiaceae) was used as an outgroup (**Supplementary Table S1**). The plastid genes were extracted from GenBank individually and aligned the by MUSCLE (Edgar 2004) implemented in MEGAX (Kumar et al. 2018) The best- fit evolutionary models: TVM+F+I+G4: (*atpA*, *atpB*, *atpE*, *atpF*, *ccsA*, *cemA*, *matK*, *petA*, *psaC*, *psbH*, *psbK*, *psbT*, *rbcL*, *rpl14*, *rpl16*, *rpoA*, *rpoB*, *rpoC1*, *rpoC2*, *rps4*, *rps14*, *ycf3*); GTR+F+I+G4: (*atpH*, *atpI*, *petB*, *petD*, *petG*, *petN*, *psaA*, *psaB*, *psaI*, *psbA*, *psbB*, *psbC*,*psbD*, *psbE*, *psbF*, *psbJ*, *psbL*, *psbN*, *psbZ*); TIM3+F+G4: (*clpP)*; TVM+F+G4: (*infA*, *psbM*, *rpl2*, *rps2*, *rps3*, *rps7*, *rps8*, *rps11*, *rps12*, *rps15*, *rps19*); TVM+F+G4: (*rpl22*), were selected after the a maximum likelihood (ML) estimation was conducted using IQTREE v 1.6.10 (Nguyen et al. 2015). Posteriorly 500 non-parametric bootstrap replications were used to asses branch support. Finally, the consensus tree was visualized through FigTree 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

SSR identication and RNA editing sites prediction

Simple sequence repeats (SSRs) in the *O. monacantha* and *O. ficus-indica* plastome were detected using the MIcroSAtellite (MISA) Perl script (Thiel et al. 2003), with thresholds of eight repeat units for mononucleotide SSRs, four repeat units for di- and trinucleotide SSRs, and three repeat units for tetra-, penta- and hexanucleotide SSRs.
Potential RNA editing sites where predicted 35 plastid protein-coding genes (*accD*, *atpA*, *atpB*, *atpF*, *atpI*, *ccsA*, *clpP*, *matK*, *ndhA*, *ndhB*, *ndhD*, *ndhF*, *ndhG*, *petB*, *petD*, *petG*, *petL*, *psaB*, *psaI*, *psbB*, *psbE*, *psbF*, *psbL*, *rpl2*, *rpl20*, *rpl23*, *rpoA*, *rpoB*, *rpoc1*, *rpoc2*, *rps2*, *rps8*, *rps14*, *rps16*, and *ycf3*) of the species *O. monacantha*, *O. ficus-indica*, *O.quimilo* and *Brasiliensis brasiliensis*, using the software Predictive RNA Editor for Plants (PREP) suite (Mower 2009). The cutoff value was set to 0.8.

Gene divergence analysis in cacti and identification of positive signatures in plastid protein-coding genes within Cactineae

The 57 protein-coding genes present in all 17 Cactaceae plastomes already sequenced, including *O. monacantha* and *O. ficus indica*, other 7 genes absent in some species and *Portulaca oleracea* (as external group) were extracted from GenBank and aligned using the software Muscle (Edgar 2004) implemented in Mega X (Kumar et al. 2018) . Phylogenetic reconstruction of each gene was performed to assess the gene divergence. The phylogenies were inferred based on the ML method following the same steps mentioned above for the phylogenetic inference using plastid genes. The gene divergence was estimated by the sum of total branch lengths that link the operational taxonomical units to the common ancestor of Cactaceae species sampled here.

We also investigated for the presence of positive signatures (positive selection) of 77 functional genes in *O. monacantha* and *O. ficus indica*, that were aligned as described before. The positive signatures were analyzed using Selecton version 2.4 (http://selecton.tau.ac.il//; Stern et al.2007), performing the comparison M8 (allows positive selection) against M8a (null model) with one degree of freedom and cutoff (ε) of 0.1. We consider in our analysis only sites where reliable positive selection was inferred (lower bound >1 and test with probability < 0.01.

3. RESULTS

General features of the plastomes Opuntia

The plastomes of *O. monacantha* (**Figure 1**) and *O. ficus-indica* (**Figure 2**) has a circular DNA molecule with the typical quadripartite structure, containing two inverted repeat regions (IRA and IRB) among two single copy regions (LSC and SSC), the larger and the smaller, respectively. The SSC region was the only one that practically did not vary among these species, 4109 bp in *O. monacantha* to 4110bp in *O. ficus-indica*. On the other hand, LSC varied significantly, 101,384 bp in *O. monacantha* to 87.248 bp in *O. ficus-indica* and IR 21,791bp to 30,515 bp, respectively (**Table 1**). The overall GC content determined to *O. monacantha* is 36.6%, and *O.ficus-indica* 36.7%, similarlly to all species in this family (**Table 1**).

The gene content of both *Opuntia* plastomes includes 110 unique genes, of which 77 are protein-coding genes, 29 are tRNAs genes, and four are rRNA genes (**Table 2** and **3**). *O. monacantha* has 17 completely duplicated genes, including 8 protein-coding genes, four tRNAs genes, and four rRNA genes. Three pseudogenes (*ycf1*, *ycf2*, and *trnV-UAC*) were identified by the presence of stop codons in the two protein-coding genes. O. *ficus-indica* on the other hand has 25 completely duplicated genes, and *rps19* gene partially duplicated in the IRs boundaries. The set of completely duplicated genes include, 14 of protein-coding, seven tRNA, and four rRNA. Two pseudogenes, *ycf1* and *trnV-UAC*, and a lost gene *ycf2*. Both species have 14 genes harbor one intron (five tRNA genes and nine protein-coding genes) and one harbor two introns (*ycf3* gene). The two introns of the *clpP* gene and one intron of the rpl2 gene were lost in both species.

To facilitate interpretation, we used multiple alignment analyses to compare only the LSC region (from *trnH-GUG* to *rps19* genes) (i) *P. oleracea* with *O. monacantha* and *O. ficus-indica* (ii) *O. monacantha*, *O. ficus-indica* and *O. quimilo* (**Supplementary Figure S1**). Here, we hypothesize two possible explanations for the rearrangements founded. The first hypothesis is that a large inversion of block E to B has occurred, explaining why *Opuntia* has a different position and gene order inside LSC. Posteriorly, a second inversion has occurred, reversing the effects of the first, in LCB E the region *petL* to *rps12* and in LCB C the region *trnV-UAC* pseudogene in *Opuntia* to *rbcL*, explaining why these blocks have the same orientation and order of genes of their ancestor.

Other hypothesis would be the inverse of the first, that is, inversions occurred in blocks E and C first. The second inversion encompassed the entire region involved in these

rearrangements (E to B), consequently it involved the blocks E and C, who have already suffered inversion, (see **Figure 3**).

When comparing the complete plastomes of *O. monacantha* and *O. ficus-indica* with *P. oleracea*, by Mummer (**Supplementary Figure S2** and **S3**), respectively. The rearrangements and inversions found here in the SSC-IRs region of *Opuntia*, in addition to the LSC region mentioned earlier (**Supplementary Figure S1**, **Figure 3** (**E** and **C**), are shared with *Brasiliopuntia brasiliensis* (Chapter 1).

The differences found between the two *Opuntia* consist in the fact *O.mocantha* shares IRB/LSC contraction with *O.quimilo*, both having a more expanded LSC when compared to *O.ficus - indica* and *P.oleracea* (**Figure 3**).

In addition to the rearrangements that have been identified in Opuntioideae, only *O.ficus-indica* shows the loss of the *ycf2* gene (**Supplementary Figure S2** (number 1 highlight in red)), which in theory would be in the IR, due to the SSC/IRB contraction event (**Figure 3**(**3**)). *O. monacantha* and *O. quimilo*, present the *ycf2* in the LSC region, but as a pseudogene (**Figure 3**).

Cactineae phylogenetic reconstruction based on plastid genes

We carried out phylogenetic analysis, based 54 protein-coding plastid genes aiming to infer the position of *Opuntia monacantha* and *Opuntia ficus-indica* within Cactaceae and suborder Cactineae. This analysis is an update of the phylogeny of the suborder Cactineae (Chapter 1; Morais da Silva et al. 2020). In this update were included cacti of the subfamily Opuntioideae, *O. monacantha* and *O. ficus-indica*, sequenced in this work, one basal specie, represented by *Pereskia aculeata*, and from the subfamily Cactoideae, were included *Selenicereus undatus* and *Rhipsalis baccifer*a (**Supplementary Table S1**). The maximum likelihood (ML) analysis produced a consensus tree (**Figure4**) with a log-likelihood of -137906.927813. In this tree topology, the relationships among the families are in accordance to previously works (Chapter 1; Morais da Silva et al. 2020).

The separation of the two subfamilies (Opuntioideae and Cactoideae) is noticeable and has high branch support, most with 100% of ML-BS. In Cactoideae, this phylogeny groups *Rhipsalis teres* with *R. baccifera*, within the *Carnegiea* and *Lophocereus* clade. *C. gigantea* and *L. schotti* forming a sister group with *Seleniceus undatus*.

Within Opuntioideae, *O. ficus-indica* diverged a little earlier, but it stays grouped with *O. monacantha* and *O. quimilo*, forming a sister group, which, in turn, all formed a sister group with *Brasiliopuntia brasiliensis*.

Sequence repeats content and identification of putative RNA editing sites in plastid protein-coding genes of three species of genus Opuntia

We identified 240 and 253 SSR loci in the plastomes of *O. monacantha* and *O. ficus-indica* (**Supplementary Table S2** and **S4**), respectively. Most of them, approximately 70%, comprises A/T mononucleotide repeats (**Supplementary Table S3** and **S5)**. Of the total simple sequence repeat mapped in *O. monacantha* plastome, 188 are located in the LSC region, 13 in the SSC and 37 in the IRs. Already in *O. ficus- indica*, 176, 15 and 59, respectively. Additionally, in *O. monacantha* the total SSR mapped, 150 are present in intergenic spacers (IGS), 38 in introns and 52 in CDS. In *O. ficus-indica* similar numbers were found for the regions, 154 (IGS), 46 (intron) and 53 (CDS). The distribution of SSR loci is well conserved among species, 227 occurrences are shared among them.

The predictive analysis of the RNA editing sites in plastid protein-coding genes of the subfamily Opuntioideae was carried using the webserver PREP. We identified 68 sites for *O. ficus-indica*, *O. monacantha and Brasiliopuntia brasiliensis,* distributed in 22 genes for species genus *Opuntia* and in 20 genes for *B.brasiliensis*. *O. quimilo*, in turn, showed 69 sites, distributed in 21 genes (**Supplementary Table S6**). All RNA editions predicted are cytidine (C) to uridine (U) conversions, at the first (24.6%) or second (75.4%) positions of the codons.

Most predicted RNA sites (63,23%) changed the encoded amino acid from polar to apolar. When a change from apolar to polar occurred, only 6 of 68 sites of change were predicted, representing 8,83% of sites. There were also 17 editions that did not change the polarity of the amino acid (sixteen changes of apolar to apolar and three polar to polar), filling 27.34% of the changes in general.

Of the 68 sites predicted for Opuntioideae, 64 of these are shared by all 4 species analyzed. Some RNA editing sites are unique in the Opuntioideae subfamily, except from *accd* gene, which is a very divergent gene and presents distinct sites between species. The *psbB* gene (e.g., CTT(L)=>TTT(F)) in *O. ficus-indica*, *rpoc2* gene (e.g. TCC(S)=> TTC(F)) in *O. quimilo*, and *matK* gene (e.g. CCG(P)) =>CTG(L)) *B. brasiliensis*.

Molecular evolution analysis of protein-coding genes in Opuntia plastomes

This analysis of gene divergence in these cacti indicates that 50 of those genes are evolving under low substitution rates, around 0.01, in all species (**Figure 5**).

In general, the gene divergence analysis indicates that the plastid protein-coding genes in Cactaceae are very divergent. The *rps* genes varied significantly in branch length, being close to 0.1 for all species. Genes like *accD* followed by *clpP* and *infA*, show extensive variation Mainly *accD* in species of subfamily Opuntioideae, reaching 0.80 branch length. Other genes, such as *psbM*, and *rpl22*, are more divergent in subfamily Cactoideae, on the other hand, *petL* is divergent only in Opuntioideae.

We also investigate if any plastid gene in Cactineae undergoes positive selection. We identified a total of 189 putative positive signatures distributed in just over half of plastid genes (35 of 79 protein-coding genes). This positive selection analysis corroborates what was identified in the genetic divergence analysis. The more divergent the gene, the greater the number of positive signatures.

Positive signatures were identified and evolutionary patterns were identified in Cactineae. We plotted the sites under positive selection against phylogenetic inference in several genes involved in different functions, such as photosynthesis and others functions [(**Figure 6**) ATP synthase (*atpA*, *B* and *F*), RNA maturation (*matk*), cythocrhome biosynthesis (*ccsA* gene), envelope membrane protein (*cem A*), and translation initiation factor (*infA*), (**Figure 7**) (*psb* genes, *rbcL*), and ndh complex (*ndh* genes) (**Figure 8**)]. The genes related to express machinery including subunits of RNA polymerase (*rpo* genes), and ribossomal proteins (*rpl* and *rps* genes), (**Supplementary Figure S4, S5, S6, and S7**).

Some evolutionary types are notorious when we compare the selection signatures to the phylogeny (relationship between the suborder Cactineae). In the case of the *rpl* and *rps* genes, for example, the sites are quite heterogeneous and position does not always occur in all species (**Supplementary Figure S5, S6 and S7)**. However, most of the times the positive selection site in Opuntioideae is shared. Genes related to photosynthesis, usually more conserved genes, the sites identified are shared in different families (**Figure 6** and **7**). The genes related to the *ndh* complex, absent in the Cactoideae subfamily, but present in Opuntioideae, showed positive selection sites, identical to the other species in this topology, not closely related (**Figure S8)**.

4. DISCUSSION

Opuntia have few genetic losses and/or pseudogenization, and have the same IRB/SSC inversions and rearrangements

The plastomes of Opuntioideae are conserved in gene content (**Table 1**; Chapter 1) and the rearrangements are just the expansion in the IRB-SSC junction. The LSC and IRs are the regions in which these plastomes have wide variation while SSC is very similar between species. *Opuntia* has the smaller plastomes compared to *B. brasiliensis*, but Opuntioideae has the largest plastomes within Cactaceae (**Table 1**).

The inversions and translocations in the LSC region that occurred in Opuntioideae, occur in the same way for all species this group, but the rearrangements we have exceptions, in addition, to rearrangements in IRB-SSC that occurred in *O. ficus-indica* and *B. brasiliensis*, *O. monacantha* shared contraction IRB-LSC with *O. quimilo*, therefore both having the largest LSC, ~20 kb more than other species in Opuntioideae. In general, Cactaceae with present lot of rearrangements in their plastomes, from species with total IR loss (Sanderson et al. 2015) to a very reduced IR with SSC expansion (Solórzano et al. 2019; Oulo et al. 2020; Morais da Silva et al. 2021). However, until now, the expansion of IRs and LSC was reported only in Opuntioideae (Chapter 1; Köhler et al. 2020).

Opuntia has a very conserved gene content, except for the pseudogenes *ycf1*, *ycf2*, *trnV-UAC*. The *ycf2* gene was lost in *O. ficus-indica*. Degeneration of *ycf1* and *ycf2* is not recurrent in Cactaceae, only *Mammillaria zephyranthoides*, *Rhipsalis baccifera*, and *Opuntia quimilo*, present these genes as pseudogenes (Solórzano et. al 2019; Oulo et al. 2020; Köhler et al. 2020). However, in angiosperms, complete loss or degeneration of these genes is common (Guisinger et al. 2011; Fajardo et al. 2013; Rabah et al. 2019; Pacheco et al. 2020). Although both genes are considered essential, their functions have not been discovered and there are no reports of alternative mechanisms to make up for the lack of both (Drescher et al. 2000).

The *trnV-UAC* is an essential transporter RNA, according to the Superwobling theory (Alkatib et al 2012; Rogalski et al. 2008). This gene in Cactoideae as missing, and the loss is justified because it occurs in a region of the LSC where a lot of rearrangements has been reported for a long time (i.e., **LCB C Figure 3**) (Sanderson et al. 2015; Solórzano et al. 2019; Oulo et al. 2020; Morais da Silva et al. 2021; Wallace 1995). However, identified *trnV-UAC* as a pseudogene in *B. brasiliensis* (Chapter 1), through analysis of the primary and secondary structure of this gene, in addition the structural analysis, through analysis codon usage, verified that it is possibly used of that RNA transporter valine, indicating import. Here we identified this

trnV-UAC as a pseudogene in *Opuntia* and based on the work of Silva and collaborators, we continue to hypothesize the import due to your essentiality.

All species of the subfamily Cactoideae available to the database, have revealed in their plastomes a vast gene loss and/or degeneration to pseudogenes. The NDH complex, related to NAD(P)H dehydrogenase, acting mainly in the cyclic transfer of electrons under stressful conditions (Li et al. 2014; Yamori and Shikanai 2016), has been the most reported in these cases, followed by genes related to the machinery of expression, more precisely the 50s (*rpl23*, *rpl33*, *rpl36*) and 30s (*rps16* and *rps18*) subunits of the ribosome, in addition to the photosynthesis-related ycf4 (Krech et al. 2012).

In *Opuntia*, all these genes degenerated to pseudogenes and/or lost in Cactoideae are functional, most of the *ndh* genes are present in the region of the IRs, and we could even hypothesize that they were being maintained because they were undergoing gene conversion (mutation corrections) (Khakhlova and Bock 2006). Although, we believe that this is not the case due to the fact that 5 subunits (*ndhC*, *D*, *E*, *J*, and *K*) occur in the LSC and SSC region, sites that do not have gene coverage, as they do not have duplicate copies in the plastome.

Interestingly, these genes are directly involved with processes that impact plant growth (protein biosynthesis, especially in the early stages of development and photosynthesis), although most are not considered essential. Cactoideae has a slow development compared to Opuntioideae, which are cacti of long, and flat cladodes, it is likely that the absence of these genes may be impairing development.

SSR mapping and prediction

To date, plastid molecular markers have not been reported in *Opuntia* populations. Likewise, the Cactaceae family in general is poorly studied in several aspects related to genetic diversity of populations. Here we detect, 240 SSR in the plastome of *O. monacantha* and 253 in *O. ficus-indica*, which represents a contribution of genetic data to the genus *Opuntia* that can be used in studies of conservation and population genetics, in addition to helping to identify the population structure (Rogalski et al. 2015). Most of the simple sequence repeats (microsatellites) mapped here in *Opuntia* occurred mainly in intergenic spaces (IGS), and are monopolymers of A and T (Morais da Silva et al. 2021; George et al. 2015; Wheeler et al. 2014).

In *Brasiliopuntia brasiliensis*, 236 microsatellites were mapped, of which 218 are shared with *Opuntia*, signaling that these potential molecular markers can be used at the subfamily level (Chapter 1).

Morais da Silva and collaborators (2021) signaled, that genes with a high degree of divergence (e.g., *accD* and *clpP*) would be good molecular markers for the Cactoideae subfamily. Here, we complement the use of these two genes already mentioned, with the genes *petL*, for Opuntioideae, *psbM*, *rps18* and *rpl22*, for all cacti.

RNA editing sites are highly conserved in Opuntioideae

Plastid RNA editing is a post-transcriptional modification mechanism in RNA, in flowering plants (C-U), and in ferns and mosses (U-C). In parallel with the number of RNA edits being decreasing with the evolution of plants (Takenaka et al. 2013), it was identified that there is a tendency for these edits to occur more in non-essential genes and that if they occur in essential genes, it is in portions that the change does not interfere with the gene's functionality (Fiebig et al. 2014). In addition, RNA editing can act to correction mutations, indirectly repairing (restoring amino acids, start and stop codons) (Takenaka et al. 2013; Ichinose and Sugita 2017).

We identified a total of 68 RNA editing sites, of which 64 were shared by all Opuntioideae analyzed (**Supplementary Table S6)**, demonstrating a high conservation of the edited sites. Edits occurred in the first and second positions of the codons, which has been reported for angiosperms (Takenaka et al. 2013). More than half of the predicted sites caused the amino acid change from polar to apolar, increasing the number of hydrophobic proteins, which can impact the interaction of protein complexes and transmembrane domains (He et al. 2016; Lopes et al. 2019).

Comparing to the RNA editing sites identified in the subfamily Cactoideae by Morais da silva and collaborators (2021), we can consider as gain the site in the *psbB* gene (193), which was identified only in *O. ficus-indica*, matK (54) in *B. brasiliensis*. Already (610) rpoc2 in *O. quimilo*, it was identified in all Cactoideae, but it is unique among Opuntioideae species.

Also comparing with that found in Cactoideae, the sites in the *ccsA*, *ndh* complex and *rps8* genes identified here occur only in Opuntioideae. On the other hand, sites identified by Morais da Silva and collaborators (2021) in the *rpl2* and *rpoA* genes only occur in species of the Opuntioideae subfamily.

It is noteworthy that the genes of the *ndh* complex are responsible for 30 out of 68 of the editing sites and that in Cactoideae there was degeneration of the complex. Several angiosperms have an accumulation of RNA editing sites in genes of this complex (Tsudzuki et al. 2001).

We present here results of the first RNA editing prediction analysis performed for Opuntioideae, but due to the lack of taxa it is not possible inferring the evolutionary pattern for the subfamily.

Molecular evolution of plastid protein-coding genes with Opuntioideae

We performed analyzes to identify gene divergence and also investigate positive selection in protein-coding genes, in order to elucidate the evolution of these genes in Cactineae. In our gene divergence analysis, we identified some genes with wide variations in branch length (e.g., *accD*, *clpP*, *psbM*, *rps19*, *infA*, and *petL* genes).

All of these genes, with the exception of the *petL* gene, were reported as highly divergent in Cactoideae (Morais da Silva et al. 2021). The *accD* gene had been considered a pseudogene in Cactoideae (Sanderson et al. 2015; Solórzano et al. 2019; Köhler et al. 2020), but it is only responsible for all fatty acid biosynthesis Rogalski and Carrer 2011; Salie and Thelen 2016), here we take into account the finding by Morais da Silva and collaborators (2021) that the C-terminal portion houses the domains that guarantee the functionality of the gene is conserved, therefore considering the gene to be functional even with a high degree of divergence. This gene is also reported to be well divergent in *Passiflora* and *Brasiliopuntia brasiliensis* (Rabah et al. 2019; Pacheco et al. 2020; Chapter 1).

The loss of introns and the fact that it is very divergent has been reported for *clpP* in several cacti and other species (Morais da Silva et al 2020; Köhler et al. 2020; Oulo et al. 2020; Pacheco et al.2020; Lopes et al. 2018; Chapter 1).

There are reports in the *Silene* genus that the high divergence of *accD* and *clpP* genes coincide with rapid nuclear evolution and also rapid evolution of these plastid genes coinciding with high mitochondrial divergence, since, these genes encode multi-subunits and are formed by nuclear and plastid gene products (Rockenbach et al. 2016; Sloan et al. 2012; Greiner and Bock 2013)

Furthermore, we investigated if any Cactineae plastid gene were under positive selection. We expected that very divergent genes might be under positive selection, we detected 189 nucleotides under selection, distributed across 35 protein-coding genes.

We inquired whether the nucleotides under positive selection occurred in a protein domain interaction site. According to Conserved Domains Database (https://www.ncbi.nlm.nih.gov/cdd), where out of a total of 35 genes, 19 genes were available for consultation, and of these, only *rps2* and *rps7* genes had interactions identified.

In the *rps2* gene occurred indication of interaction between the protein domains, in position (39) rRNA interaction site. For the *rps7* gene at position 93, rRNA binding site. The plastid genes involved in gene expression, especially the *rpl* and *rps* genes, is reported evolving in an accelerated way (Weng et al. 2016).

5. CONCLUSIONS

In this study we reported the complete plastomes of two species of the genus *Opuntia*, *O. monacantha* and *O. ficus-indica*. Our comparative analyzes revealed well rearranged plastomes, both compared within Opuntioideae, and only within the genus *Opuntia*. The plastome of *Opuntia monacantha* revealed a contraction at the IRB-LSC junction that is shared only with *Opuntia quimilo*, both forming a sister group phylogenetically, while *Opuntia ficus-indica* showed all shared rearrangements within *Opuntia* except for the loss of the *ycf2* gene in the region of IRs, in other closely related species, this gene is located in the LSC region, but is a pseudogene. Both species present the *ycf1* and *trnV-UAC* gene as pseudogenes, genes reported in the literature as essential. Regarding the evolution analysis of plastid genes, we found high genetic divergence for the genes *accD*, *clpP*, *rps19*, *infA*, *rpl22*, and *petL*. We also identified 189 putative positive selection sites. Moreover, we identified 68 RNA editing sites, distributed in 22 genes and also mapped 240 and 253 microsatellites to *Opuntia monacantha* and *O. ficus-indica*, respectively. Finally, the complete *Opuntia* plastomes, sequenced here, helped to understand the evolution of the Cactaceae family and the Opuntioideae subfamily, in addition to the molecular markers listed here are useful for population genetic studies, as these species in specific inhabit well degraded biomes.

6. REFERENCES

Allen JF (2015) Why chloroplasts and mitochondria retain their own genomes and genetic systems: Colocation for redox regulation of gene expression. *Proc Natl Acad Sci USA* **112**: 10231–10238

Alkatib S, Scharff LB, Rogalski M, Fleischmann TT, Matthes A, Seeger S, Schöttler MA, Ruf S, Bock R (2012) The contributions of wobbling and superwobbling to the reading of the genetic code. *PLoS Genet* **8**: e1003076

Anderson EF (2001) The cactus family. Timber Press, Portland, Oregon

Arakaki M, Christin PA, Nyffeler R, Lendel A, Eggli U, Ogburn RM, Spriggs E, Moore MJ, Edwards EJ (2011) Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proc Natl Acad Sci USA* **108**: 8379–8384

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**(5): 455-77

Bauer D, Waechter JL (2006) Sinopse taxonômica de Cactaceae epifíticas no Rio Grande do Sul, Brasil. *Acta Bot Brasilica* **20**: 225–239

Britton NL, Rose JN (1919) The Cactaceae, vol 1. Carnegie Institution of Washington, Washington

Daniell H, Lin CS, Yu M, Chang WJ (2016) Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol* **17**: 1–29

Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* **14**: 1394–1403

Dieleman H (2017) Urban agriculture in Mexico City; balancing between ecological, economic, social and symbolic value. *J Clean Prod* **163**: S156–S163

De Santiago E, Domínguez-Fernández M, Cid C, De Peña MP (2018). Impact of cooking process on nutritional composition and antioxidants of cactus cladodes (Opuntia ficus-indica). *Food chemistry* **240**: 1055-1062.

Drescher A, Ruf S, Calsa T Jr, Carrer H, Bock R (2000) The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *Plant J* **22**: 97–104

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

Fajardo D, Senalik D, Ames M, Zhu H, Steffan SA, Harbut R, Polas- hock J, Vorsa N, Gillespie E, Kron K, Zalapa JE (2013) Complete plastid genome sequence of Vaccinium macrocarpon: structure, gene content, and rearrangements revealed by next genera- tion sequencing. *Tree Genet Genomes* **9**: 489–498

Feugang JM, Konarski P, Zou D, Stintzing FC, Zou C (2006). Nutritional and medicinal use of Cactus pear (Opuntia spp.) cladodes and fruits. *Front Biosci* **11**(1): 2574-2589

Fiebig A, Stegemann S, Bock R (2004) Rapid evolution of RNA editing sites in a small non-essential plastid gene. *Nucleic Acids Res* **32**: 3615–3622

George B, Bhatt BS, Awasthi M, George B, Singh AK (2015) Comparative analysis of microsatellites in chloroplast genomes of lower and higher plants. *Curr Genet* **61**: 665–677

Greiner S, Bock R (2013) Tuning a ménage à trois: co-evolution and co-adaptation of nuclear and organellar genomes in plants. *BioEssays* **35**: 354–365

Greiner S, Lehwark P, Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res* **47**: W59–W64

Grünwaldt JM, Guevara JC, Grünwaldt EG, Carretero EM (2015). Cacti (Opuntia sps.) as forage in Argentina dry lands. *Revista de la Facultad de Ciencias Agrarias* **47**(1): 263-282

Guerrero PC, Majure LC, Cornejo-Romero A, Hernández-Hernández T (2019) Phylogenetic relationships and evolutionary trends in the cactus family. *J Hered* **110**: 4–21

Guisinger MM, Kuehl JV, Boore JL, Jansen RK (2011) Extreme reconfiguration of plastid genomes in the angiosperm family geraniaceae: rearrangements, repeats, and codon usage. *Mol Biol Evol* **28**: 583–600

Haberle RC, Fourcade HM, Boore JL, Jansen RK (2008) Extensive rearrangements in the chloroplast genome of Trachelium caer- uleum are associated with repeats and tRNA genes. *J Mol Evol* **66**: 350–361

He P, Huang S, Xiao G, Zhang Y, Yu J (2016) Abundant RNA editing sites of chloroplast protein-coding genes in Ginkgo biloba and an evolutionary pattern analysis. *BMC Plant Biol* **16**: 257

Ichinose M, Sugita M (2017) RNA editing and its molecular mechanism in plant organelles. *Genes* **8**:5

Kikuchi S, Bédard J, Hirano M, Hirabayashi Y, Oishi M, Imai M, Takase M, Ide T, Nakai M (2013) Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science* **339**: 571–574

Kikuchi S, Asakura Y, Imai M, Nakahira Y, Kotani Y, Hashiguchi Y, Nakai Y, Takafuji K, Bédard J, Hirabayashi-Ishioka Y, Mori H, Shiina T, Nakai M (2018) A Ycf2-FtsHi Heteromeric AAA- ATPase complex is required for chloroplast protein import. *Plant Cell* **30**: 2677–2703

Khakhlova O, Bock R (2006). Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J* **46**(1):85-94

Köhler M, Reginato M, Souza-Chies TT, & Majure LC (2020). Insights Into Chloroplast Genome Evolution Across Opuntioideae (Cactaceae) Reveals Robust Yet Sometimes Conflicting Phylogenetic Topologies. *Frontiers in plant science* **11**: 729

Krech K, Ruf S, Masduki FF, Thiele W, Bednarczyk D, Albus CA, Tiller N, Hasse C, Schöttler MA, Bock R (2012) The plastid genome-encoded Ycf4 protein functions as a nonessential assembly factor for photosystem I in higher plants. *Plant Physiol*. **159**(2): 579-91

Kumar S, Stecher G, Li M, Knyaz C, & Tamura K (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution* ***35***(6): 1547-1549

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Anto- nescu C, Salzberg SL (2004) Versatile and open software for comparing large genomes. *Genome Biol* **5**(2): R12

Li XG, Duan W, Meng QW, Zou Q, Zhao SJ (2004) The function of chloroplastic NAD(P)H dehydrogenase in tobacco during chilling stress under low irradiance. *Plant Cell Physiol* **45**(1): 103-8

Lopes AS, Pacheco TG, Santos KGD, Vieira LN, Guerra MP, Nodari RO, de Souza EM, Pedrosa FO, Rogalski M (2018a) The *Linum usitatissimum* L. plastome reveals atypical structural evolution, new editing sites, and the phylogenetic position of *Linaceae* within *Malpighiales*. *Plant Cell Rep* **37**: 307–328

Lopes AS, Pacheco TG, Nimz T, Vieira LN, Guerra MP, Nodari RO, Souza EM, Pedrosa FO, Rogalski M (2018b) The complete plastome of macaw palm [Acrocomia aculeata (Jacq.) Lodd. ex Mart.] and extensive molecular analyses of the evolution of plastid genes in Arecaceae. *Planta* **247**: 1011–1030

Lowe TM, Chan PP (2016) tRNAscan-SE On-line: integrating Search and context for analysis of transfer RNA genes. *Nucleic Acids Res* **44**: W54–W57

Lucena CM, Lucena RF, Costa GM, Carvalho TK, Costa GG, Alves RR, Pereira DD, Ribeiro JE, Alves CA, Quirino ZG, Nunes EN (2013) Use and knowledge of Cactaceae in Northeastern Brazil. *J Ethnobiol Ethnomed* **28**;9(1): 62

Majure LC, Puente R, Griffith MP, Judd WS, Soltis PS, Soltis DE (2012) Phylogeny of Opuntia ss (Cactaceae): clade delineation, geographic origins, and reticulate evolution. *American journal of botany* **99**(5): 847-864.

Morais da Silva G, Lopes AS, Pacheco TG, Machado KLG, Silva MC, Oliveira JD, de Baura VA, Balsanelli E, de Souza EM, Pedrosa FO, Rogalski M (2021) Genetic and evolutionary analyses of plastomes of the subfamily Cactoideae (Cactaceae) indicate relaxed protein biosynthesis and tRNA import from cytosol. *Braz. J. Bot* **44:** 97–116

Mower JP (2009) The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Res* **37**:253–259

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximumlikelihood phylogenies. *Mol Biol Evol* **32**: 268–274

Oulo MA, Yang JX, Dong X, Wanga, VO, Mkala, EM, Munyao JN, Onjolo VO, Rono PC, Hu GW, Wang QF (2020) Complete Chloroplast Genome of Rhipsalis baccifera, the only Cactus with Natural Distribution in the Old World: Genome Rearrangement, Intron Gain and Loss, and Implications for Phylogenetic Studies. *Plants* **9**(8): 979

Oliveira PS, Rico-Gray V, Díaz-Castelazo C, Castillo-Guevara C (1999) Interaction between ants, extrafloral nectaries and insect herbivores in Neotropical coastal sand dunes: Herbivore deterrence by visiting ants increases fruit set in Opuntia stricta (Cactaceae). *Funct Ecol* **13**: 623–631

Pacheco TG, Lopes AS, Viana GDM, Silva ON, Silva GM, Vieira LN, Guerra MP, Nodari RO, Souza EM, Pedrosa FO, Otoni WC, Rogalski M (2019) Genetic, evolutionary and phylogenetic aspects of the plastome of annatto (Bixa orellana L.), the Amazonian commercial species of natural dyes. *Planta* **249**: 563–582

Pacheco TG, Lopes AS, Welter JF, Yotoko KSC, Otoni WC, Vieira LN, Guerra MP, Nodari RO, Balsanelli E, Pedrosa FO, Souza EM, Rogalski M (2020) Plastome sequences of the subgenus *Passiflora* reveal highly divergent genes and specific evolutionary features. *Plant Mol Biol* **104**: 21–37

Qiao J, Cai M, Yan G, Wang N, Li F, Chen B, Gao G, Xu K, Li J, Wu X (2016) High-throughput multiplex cpDNA resequencing clarifies the genetic diversity and genetic relationships among Brassica napus, Brassica rapa and Brassica oleracea. *Plant Biotechnol J* **14**: 409–418

Rabah SO, Shrestha B, Hajrah NH, Sabir MJ, Alharby HF, Sabir MJ, Alhebshi AM, Sabir JSM, Gilbert LE, Ruhlman TA, Jansen RK (2019) Passiflora plastome sequencing reveals widespread genomic rearrangements. *J Syst Evol* **57**:1–14

Rogalski M, Karcher D, Bock R (2008) Superwobbling facilitates translation with reduced tRNA sets. *Nat Struct Mol Biol* **15**: 192–198

Rogalski M, Carrer H (2011) Engineering plastid fatty acid biosyn- thesis to improve food quality and biofuel production in higher plants. *Plant Biotechnol J* **9**: 554–564

Rogalski M, Vieira LN, Fraga HP, Guerra MP (2015) Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution, and biotechnology. *Front Plant Sci* **6**: 586

Rockenbach K, Havird JC, Monroe JG, Triant DA, Taylor DR, Sloan DB (2016) Positive Selection in Rapidly Evolving Plastid-Nuclear Enzyme Complexes. *Genetics* **204**(4): 1507-1522

Takenaka M, Zehrmann A, Salie MJ, Thelen JJ (2016) Regulation and structure of the heteromeric acetyl-CoA carboxylase. *Biochim Biophys Acta* **1861**: 1207–1213

Sanderson MJ, Copetti D, Búrquez A, Bustamante E, Charboneau JLM, Eguiarte LE, Kumar S, Lee HO, Lee J, McMahon M, Steele K, Wing R, Yang TJ, Zwickl D, Wojciechowski MF (2015) Exceptional reduction of the plastid genome of saguaro cactus (Carnegiea gigantea): loss of the ndh gene suite and inverted repeat. *Am J Bot* **102**: 1115–1127

Schwarz EN, Ruhlman TA, Sabir JSM, Hajrah NH, Alharbi NS, Al-Malki AL, Bailey CD, Jansen RK (2015) Plastid genome sequences of legumes reveal parallel inversions and multiple losses of rps16 in papilionoids. *J Syst Evol* **53**: 458–468

Silva TG, Primo JT, de Morais JE, da Silva Diniz WJ, de Souza CA, da Conceição Silva M (2015) Crescimento e produtividade de clones de palma forrageira no semiárido e relações com variáveis meteorológicas. *Rev Caatinga* **28**: 10–18. ISSN: 1983-2125 *in portuguese*

Shikanai T, Shimizu K, Ueda K, Nishimura Y, Kuroiwa T, Hashimoto T (2001) The chloroplast clpP gene, encoding a proteolytic subu- nit of ATP-dependent protease, is indispensable for chloroplast development in tobacco. *Plant Cell Physiol* **42**: 264–273

Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR (2012) Recent Acceleration of Plastid Sequence and Structural Evolution Coincides with Extreme Mitochondrial Divergence in the Angiosperm Genus Silene. *Genome Biology and Evolution* **4**(3): 294–306

Solórzano S, Chincoya DA, Sanchez-Flores A, Estrada K, DíazVelásquez CE, González-Rodríguez A, Vaca-Paniagua F, Dávila P, Arias S (2019) De novo assembly discovered novel structures in genome of plastids and revealed divergent inverted repeats in Mammillaria (Cactaceae, Caryophyllales). *Plants* **8**: 392

Stintzing FC, Carle, R (2005). Cactus stems (Opuntia spp.): A review on their chemistry, technology, and uses. *Molecular nutrition & food research* **49**(2): 175-194

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* **106**: 411–422

Tsudzuki T, Wakasugi T, Sugiura M (2001) Comparative analysis of RNA editing sites in higher plant chloroplasts. *J Mol Evol* **53**: 327–332

Valente LM, da Paixão D, Do Nascimento AC, dos Santos PF, Scheinvar LA, Moura MR, Tinoco LW, Gomes LN, da Silva JF (2010) Antiradical activity, nutritional potential and flavonoids of the cladodes of Opuntia monacantha (Cactaceae). *Food Chemistry* **123**(4): 1127-1131

Verbitskiy D, Härtel B, Brennicke A (2013) RNA editing in plants and its evolution. *Annu Rev Genet* **47**: 335–352

Vieira LN, Faoro H, Fraga HPF, Rogalski M, Souza EM, Pedrosa FB, Nodari RO, Guerra MP (2014) An improved protocol for intact chloroplasts and cpDNA isolation in Conifers. *PLoS ONE* **9**: e84792

Wallace RS (1995) Molecular systematic study of the Cactaceae: Using chloroplast DNA variation to elucidate Cactus phylogeny. *Bradleya* (13): 1-12

Weng ML, Blazier JC, Govindu M, Jansen RK (2014) Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correla- tion between genome rearrangements, repeats, and nucleotide sub- stitution rates. *Mol Biol Evol* **31**: 645–659

Weng ML, Ruhlman TA, Jansen RK (2016) Plastid-nuclear interaction and accelerated coevolution in plastid ribosomal genes in Geraniaceae. *Genome Biol Evol* **8**: 1824–1838

Wheeler GL, Dorman HE, Buchanan A, Challagundla L, Wallace LE (2014) A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology. *Appl Plant Sci* **2**: 1400059

Yamori W, Shikanai T (2016) Physiological Functions of Cyclic Electron Transport Around Photosystem I in Sustaining Photosynthesis and Plant Growth. *Annu Rev Plant Biol* 29;**67**: 81-106

Yang N, Zhao M, Zhu B, Yang B, Chen C, Cui C, Jiang Y (2008) Anti-diabetic effects of polysaccharides from Opuntia monacantha cladode in normal and streptozotocin-induced diabetic rats. *Innovative food science & emerging technologies* **9**(4): 570-574

Zappi D, Nigel T, Silva SR, Machado M, Moraes EM, Cruz ACB, Correia D, Larocca J, Assis JG, Aona L, de Menezes MOT, Meiado M, Marchi MN (2011) Plano de ação nacional para a conservação das cactáceas. Série Espécies Ameaçadas. Instituto Chico Mendes de Conservação da Biodiversidade 112, Brasília. ISBN 978-85-61842-00-0

7. FIGURES

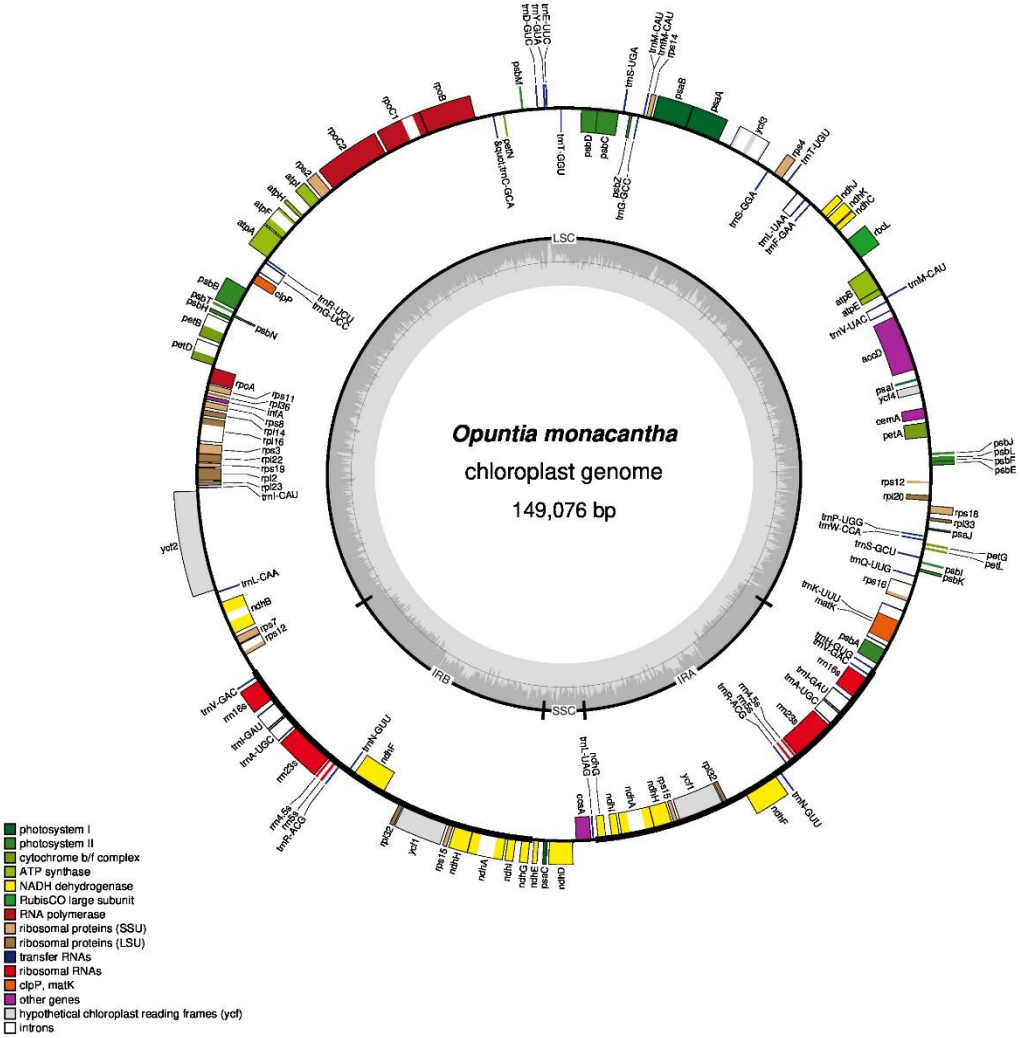


Figure 1. Gene map and genome organization of *O. monacantha.* Two inverted repeat regions IR_A and IR_B divide the circular DNA molecule into large (LSC) and small (SSC) single copy regions. Genes drawn inside the circle are transcribed clockwise, and genes drawn outside are expressed counterclockwise. Genes belonging to different functional groups are color-coded. The dark gray in the inner circle corresponds to GC content, while the light gray corresponds to AT content. The dotted circle corresponds to 50% of AT/GC content.

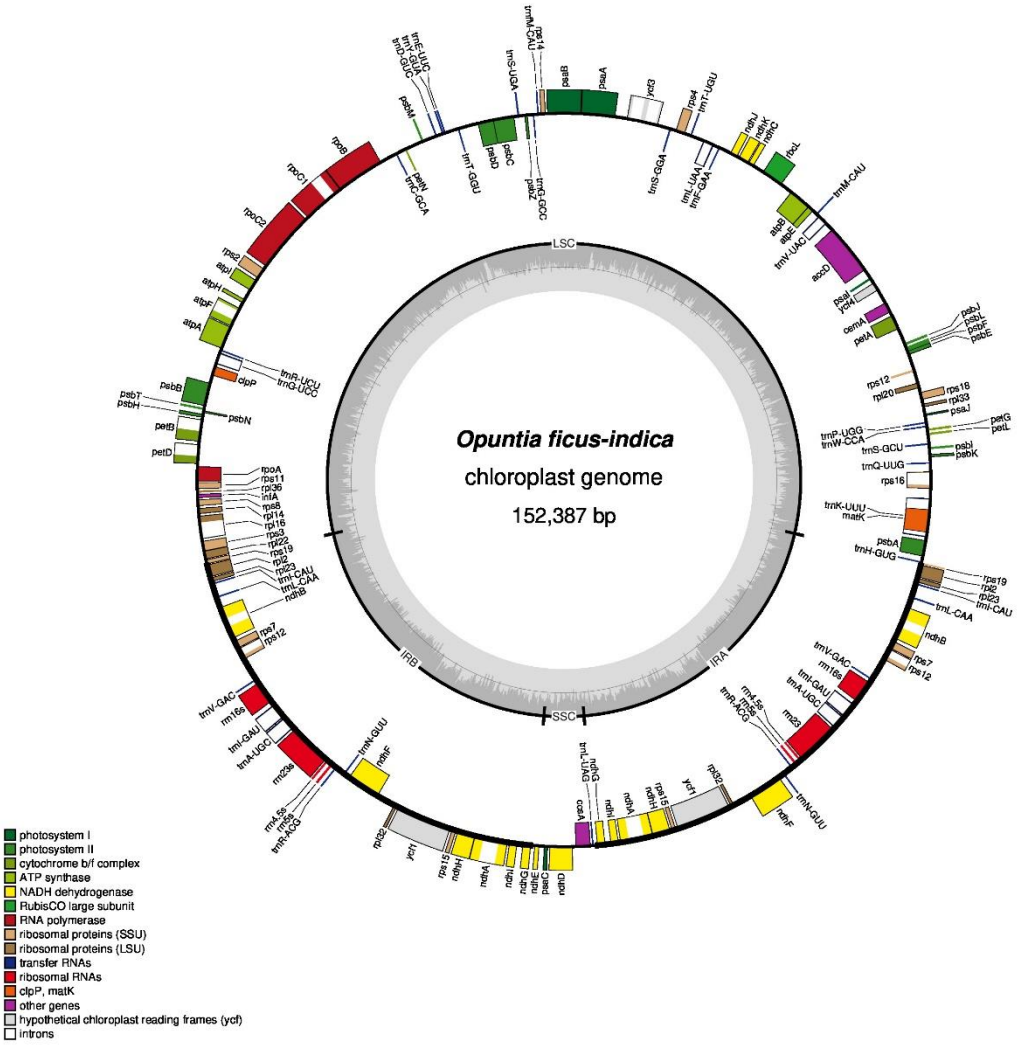


Figure 2. Gene map and genome organization of *O. ficus-indica.* Two inverted repeat regions IR_A and IR_B divide the circular DNA molecule into large (LSC) and small (SSC) single copy regions. Genes drawn inside the circle are transcribed clockwise, and genes drawn outside are expressed counterclockwise. Genes belonging to different functional groups are color-coded. The dark gray in the inner circle corresponds to GC content, while the light gray corresponds to AT content. The dotted circle corresponds to 50% of AT/GC content.

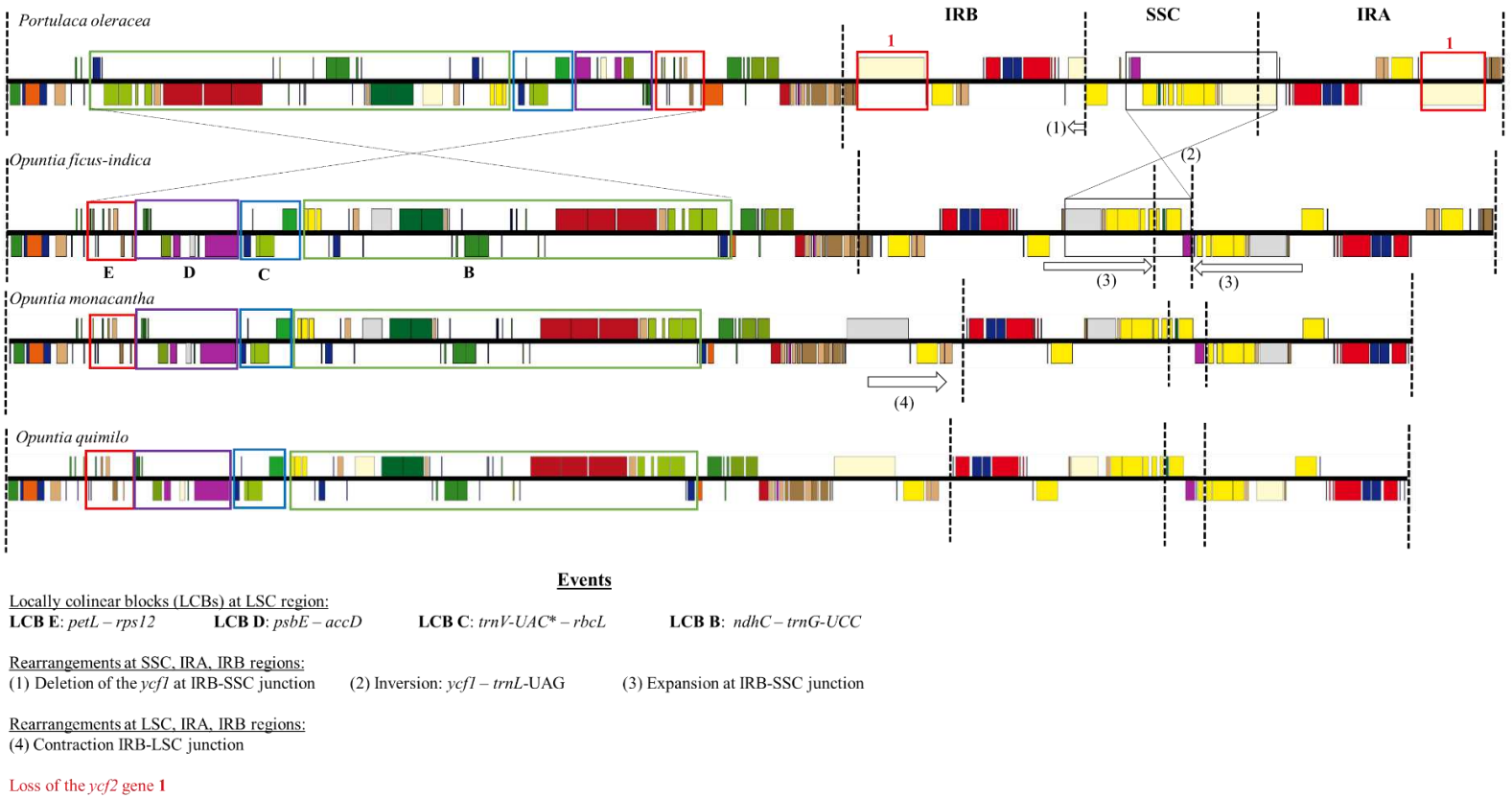


Figure 3. Comparison of the gene content and order between the *Opuntia ficus-indica* and *Opuntia monacantha* with *Portulaca oleracea* plastomes, which presents the general structure found in most angiosperms plastomes, and also between Opuntioideas. *O.ficus-indica* and *O.monacantha* is also compared with *Opuntia quimilo*, all Opuntioideae. The genes involved in inversions are highlighted by crossed and straight lines gray arrows. The white arrows indicate contraction and expansion. The (*) indicates pseudogene. Gene loss is pointed out by a red square in the plastome of *P.oleracea*. Dotted black lines horizontally separate as regions of the plastome, LSC and SSC, large and small single-copy regions; IRA /B, inverted repeat region. The linear gene maps were drawn by using OGDRAW**.**

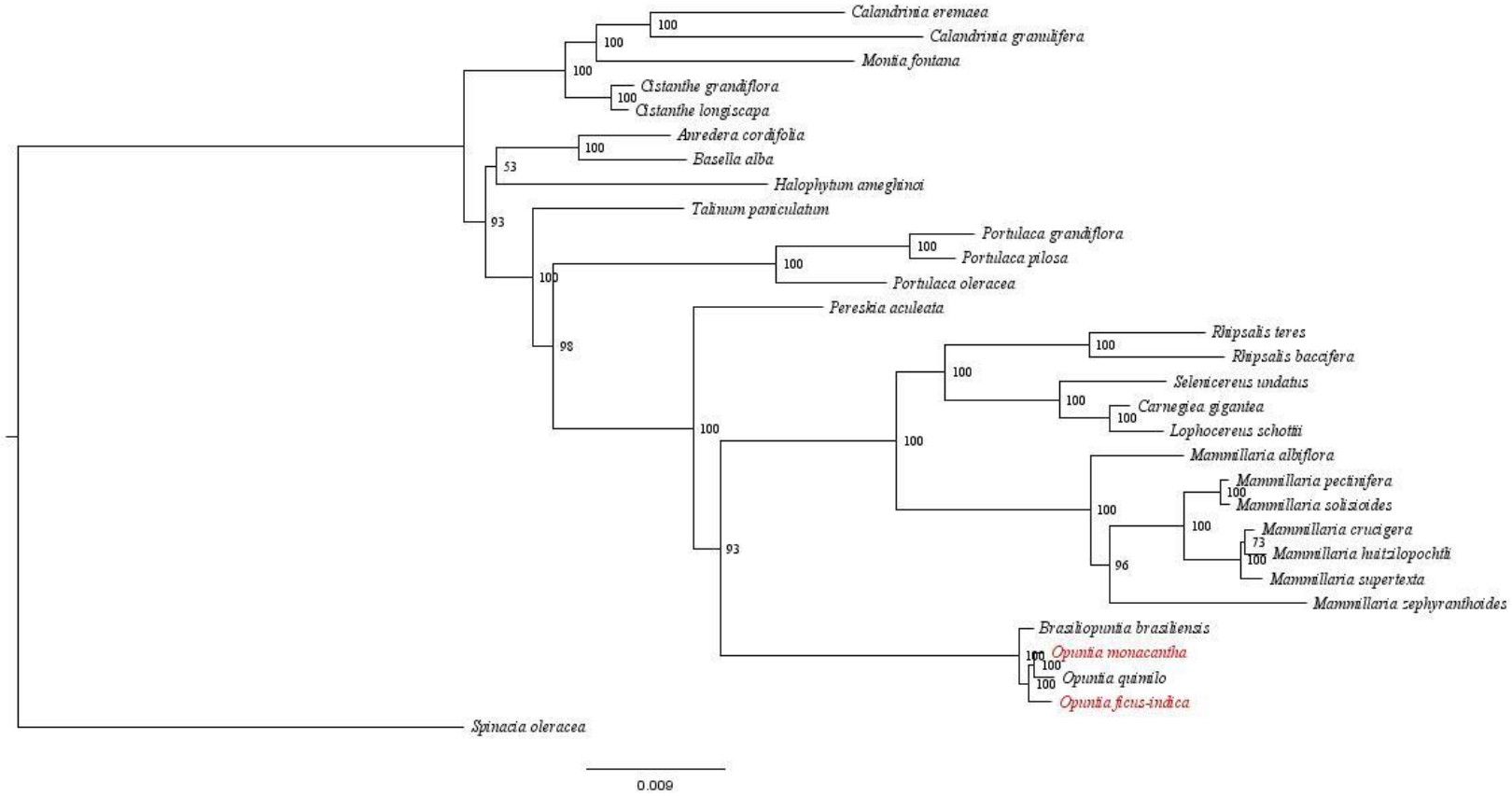


Figure 4. Cactineae phylogenetic tree of 30 taxa (29 species of suborder Cactineae and one outgroup) based on 54 protein-coding plastid genes using maximum likelihood (ML) method. Numbers (%) associated with branches are ML bootstrap support (BS) values. The branch length is proportional to the inferred divergence level. The scale bar indicates the number of inferred nucleic acid substitutions per site. The position of *Opuntia monacantha* and *Opuntia ficus-indica* is highlighted in red. *Spinacia oleracea* was used to root the tree. Tree reconstruction was performed using IQTREE software, with TVM+F+I+G4, GTR+F+I+G4, TIM3+F+G4 and TVM+F+G4, models.

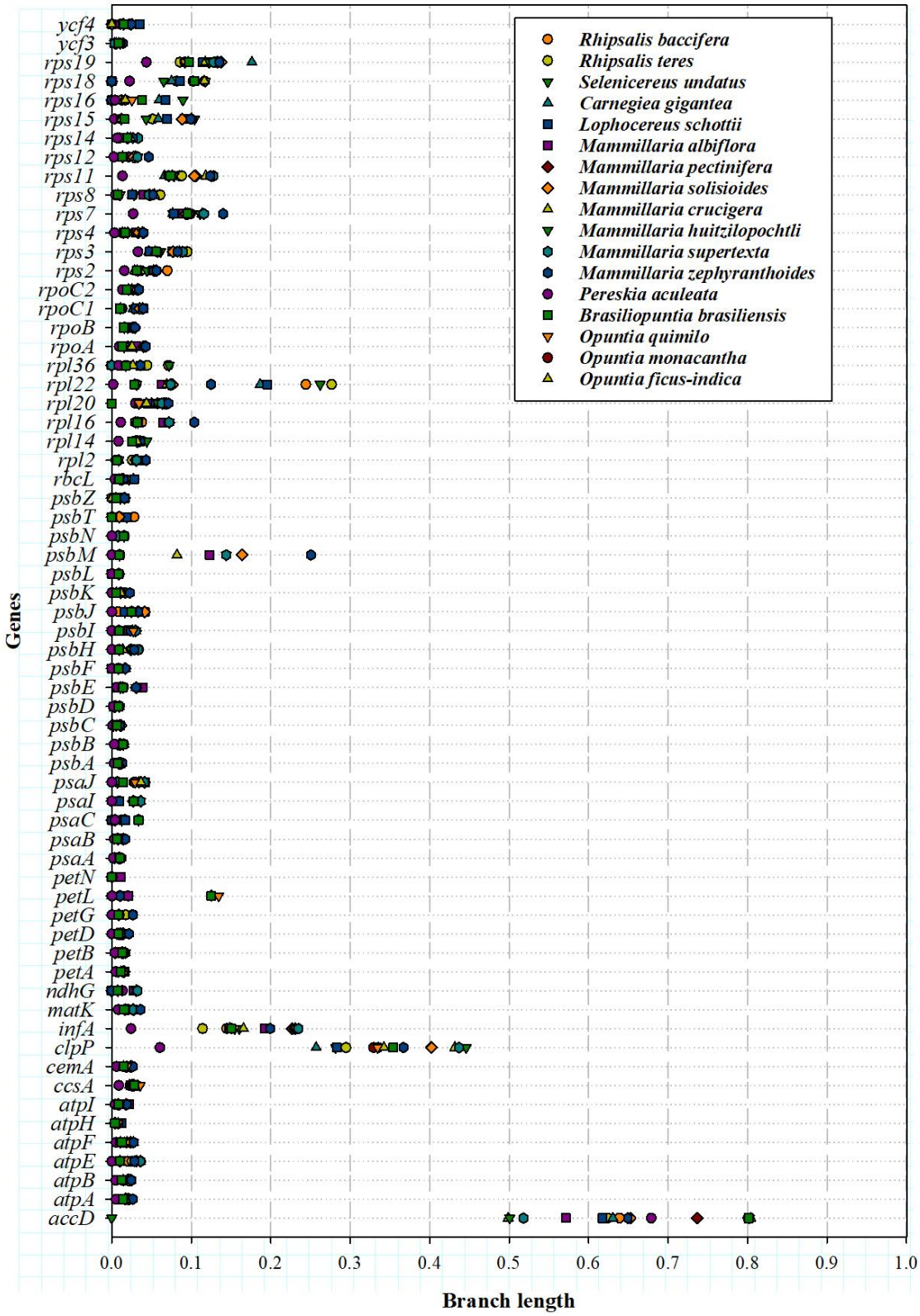
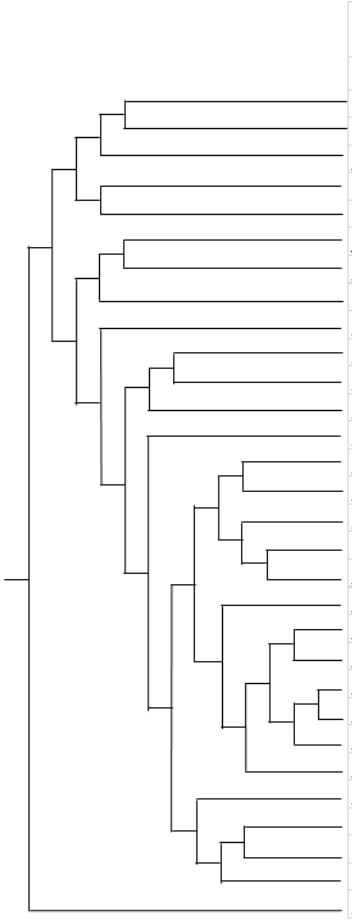
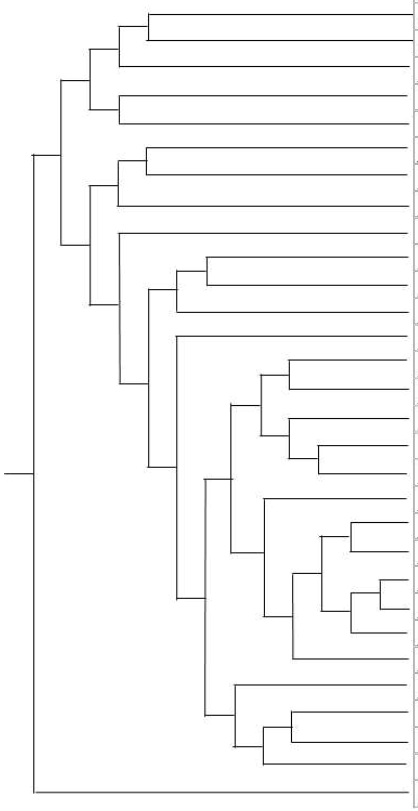


Figure 5. Divergence of plastid protein-coding genes among species of the family Cactaceae. The gene divergence was estimated by the sum of total branch lengths in each gene tree infererd.



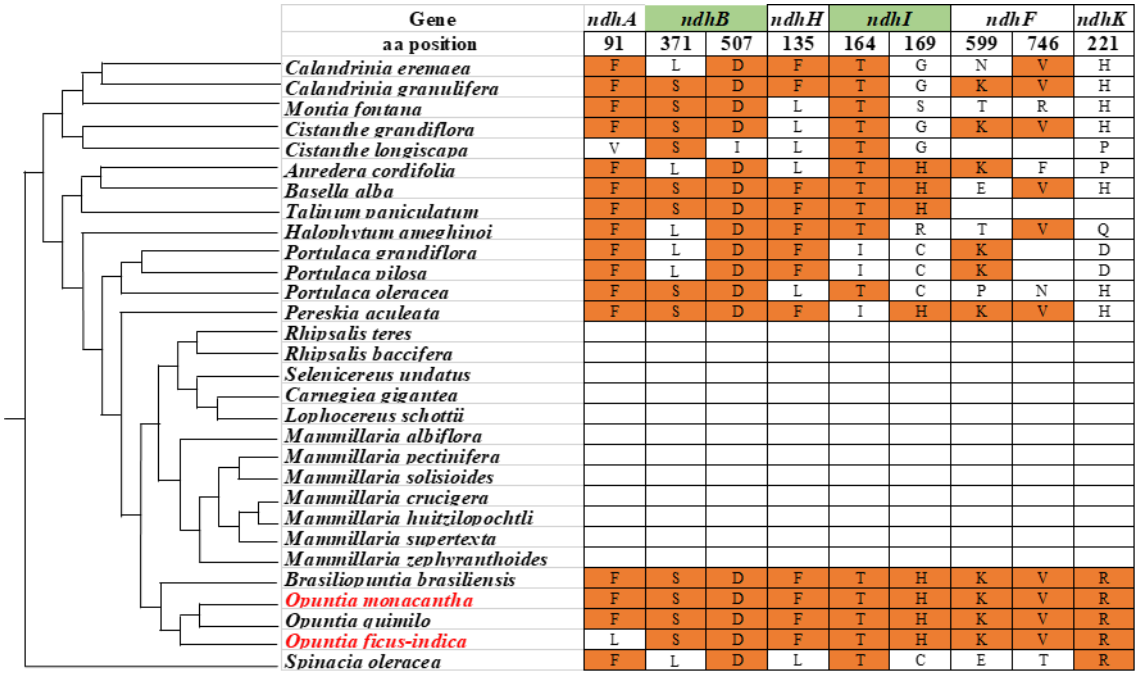
Gene	*atpA*						*atpB*					*atpF*	ccsA			cemA	infA		*matK*
aa position	79	308	427	456	459	476	3	127	228	387	452	90	93	174	177	67	5	59	258
Calandrinia eremaea	D	G	A	K	V	I	I	R	A	T	T	M	Q	R	D	H	W	S	K
Calandrinia granulifera	D	G	A	K	V	L	I	R	A	I	A	M	Q	R	D	H	W	S	K
Montia fontana	N	G	A	K	V	L	I	C	T	I	K	M	Q	R	Y	H	W	S	K
Cistanthe grandiflora	D	R	A	K	V	I	I	R	A	T	A	M	Q	S	Y	H	W	S	K
Cistanthe longiscapa	D	S	A	K	V	I	I	R	A	T	A	M	Q	R	Y	H	W	S	K
Anredera cordifolia	D	R	A	K	V	I	I	R	A	T	A	I	Q	R	Y	C	W	S	K
Basella alba	D	S	A	K	V	I	I	R	P	T	A	M	Q	R	D	C	W	S	K
Talinum paniculatum	D	S	A	K	V	I	T	R	A	T	A	M	Q	R	D	C	W	S	K
Halophytum ameghinoi	D	R	A	K	V	L	I	R	P	I	V	M	Q	R		S	W	S	K
Portulaca grandiflora	D	R	A	K	G	I	T	S	G	T	A	L	Q	R	Y	C	W	S	K
Portulaca pilosa	D	R	A	K	G	I	T	S	G	T	A	L	Q	R	Y	C	W	S	K
Portulaca oleracea	D	R	A	K	V	L	N	S	G	T	A	L	Q	R	Y	C	W	S	K
Pereskia aculeata	A	R	D	K	V	I	T	R	A	T	A	M	K	R	Y	C	S	S	K
Rhipsalis teres	S	R	E	D	E	L	A	R	A	T	A	M	Q	R	D	R	I	V	K
Rhipsalis baccifera	S	R	E	D	D	L	A	C	A	T	A	M	H	R	D	R	I	A	K
Selenicereus undatus	G	R	E	G	D	L	A	R	A	T	A	M	N	R	D	R	T	T	K
Carnegiea gigantea	G	R	K	Y	D	L	A	R	A	T	A	M	K	R	D	R	M	A	K
Lophocereus schottii	G	R	K	D	D	L	A	R	A	T	A	M	N	R	D	R	M	A	K
Mammillaria albiflora	G	R	E	D	V	L	V	R	A	T	A	I	Q	R	D	R	V	T	K
Mammillaria pectinifera	G	R	E	D	V	L	A	R	V	T	A	I	K	R	D	R	V	A	K
Mammillaria solisioides	G	R	E	D	V	L	A	R	A	T	A	I	K	R	D	R	V	E	K
Mammillaria crucigera	G	R	E	D	V	L	A	R	A	T	A	I	Q	R	Y	H	V	A	K
Mammillaria huitzilopochtli	G	R	E	D	V	L	A	R	A	T	A	I	Q	R	Y	H	V	A	K
Mammillaria supertexta	G	R	E	D	V	L	A	C	A	T	A	I	Q	R	Y	H	V	A	K
Mammillaria zephyranthoides	G	P	E	D	V	L	A	R	A	T	A	I	P	I	D	R	V	K	N
Brasiliopuntia brasiliensis	D	R	T	K	G	I	N	R	A	T	A	K	Q	R	D	L	W	L	K
Opuntia monacantha	D	R	T	K	G	I	N	R	A	T	A	K	Q	R	D	L	W	L	K
Opuntia quimilo	D	R	T	K	G	I	N	R	A	T	A	K	Q	S	D	L	W	F	K
Opuntia ficus-indica	D	R	T	K	G	I	N	R	A	T	A	K	Q	R	D	L	F	L	K
Spinacia oleracea	D	L	A	K	V	I	I	R	A	I	A	M	E	Q	D	R	W	S	K

Figure 6. Sites under positive selection in plastid genes involved in the photosynthetic process and others fumctions. The amino acids are plotted across Cactineae phylogeny, based on plastid genes concatenated. Colored amino acids in orange mean the nucleotide is under positive selection.The amino acids positions are relative to *O. monacantha* plastid genes.



Gene	*psbB*		*psbC*		*psbK*		*psbM*	*psbZ*	*rbcL*						
aa position	124	296	7	242	18	43	34	39	10	32	358	359	449	474	480
Calandrinia eremaea	S	Q	L	S	G	L	D	F	S	Q	A	D	S	T	G
Calandrinia granulifera	S	Q	L	S	S	L	D	F	S	L	A	D	S	T	A
Montia fontana	S	K	L	S	S	L	N	F	G	Q	A	D	C	T	G
Cistanthe grandiflora	S	Q	L	S	S	L	D	L	S	Q	A	D	S	T	R
Cistanthe longiscapa	S	Q	L	S	S	L	D	F	S	Q	A	D	S	T	G
Anredera cordifolia	S	Q	L	F	R	L	D	F	S	L	E	D	S	T	N
Basella alba	C	Q	L	A	S	V	D	F	F	L	D	D	A	T	G
Halophytum ameghinoi	T	Q	L	S	S	L	N	F	S	Q	A	D	G	T	
Talinum paniculatum	C	Q	L	F	S	V	D	F	F	L	A	S	T	V	G
Portulaca grandiflora	S	Q	L	S	N	V	D	F	S	L	A	D	S	V	K
Portulaca pilosa	S	Q	L	S	N	V	D	F	S	L	A	D	S	V	K
Portulaca oleracea	S	Q	L	S	S	V	D	F	S	L	A	E	S	V	N
Pereskia aculeata	C	Q	L	S	S	V	D	F	S	Q	A	N	S	T	G
Rhipsalis teres	C	K	L	S	S	V	D	F	S	Q	A	N	S	T	G
Rhipsalis baccifera	C	Q	Q	S	S	V	D	F	S	Q	A	N	S	T	G
Selenicereus undatus	C	K	L	S	S	V	D	F	S	Q	A	N	S	T	G
Carnegiea gigantea	C	K	L	S	S	V	Q	F	S	Q	A	N	S	T	G
Lophocereus schottii	C	K	L	S	S	V	D	F	S	Q	A	N	S	T	G
Mammillaria albiflora	C	K	P	F	S	V	N	F	S	Q	A	N	S	T	G
Mammillaria pectinifera	C	Q	L	F	S	V	G	L	S	Q	A	N	S	T	G
Mammillaria solisioides	C	Q	L	F	S	V	G	L	S	Q	A	N	S	T	G
Mammillaria crucigera	C	Q	L	F	S	V	L	F	S	Q	A	N	S	T	G
Mammillaria	C	Q	L	F	S	V	G	F	S	Q	A	N	S	T	G
Mammillaria supertexta	C	Q	L	F	S	V	G	F	S	Q	A	N	S	T	G
Mammillaria	C	Q	L	F	R	V	F	F	S	Q	A	N	S	T	G
Brasiliopuntia	C	Q	P	S	S	V	D	F	S	Q	A	N	S	T	G
Opuntia monacantha	C	Q	P	S	S	V	D	F	S	Q	A	N	S	T	G
Opuntia quimilo	C	Q	P	S	S	V	D	F	S	Q	A	N	S	T	G
Opuntia ficus-indica	C	Q	P	S	S	V	D	F	S	Q	A	N	S	T	G
Spinacia oleracea	S	Q	L	V	S	L	D	I	S	L	R	S	T	T	V

Figure 7. Sites under positive selection in plastid genes involved in the photosynthetic process. The amino acids are plotted across Cactineae phylogeny, based on plastid genes concatenated. Colored amino acids in orange mean the nucleotide is under positive selection.The amino acids positions are relative to *O. monacantha* plastid genes.



Gene	*ndhA*	*ndhB*		*ndhH*	*ndhI*		*ndhF*		*ndhK*
aa position	91	371	507	135	164	169	599	746	221
Calandrinia eremaea	F	L	D	F	T	G	N	V	H
Calandrinia granulifera	F	S	D	F	T	G	K	V	H
Montia fontana	F	S	D	L	T	S	T	R	H
Cistanthe grandiflora	F	S	D	L	T	G	K	V	H
Cistanthe longiscapa	V	S	I	L	T	G			P
Anredera cordifolia	F	L	D	L	T	H	K	F	P
Basella alba	F	S	D	F	T	H	E	V	H
Talinum paniculatum	F	S	D	F	T	H			
Halophytum ameghinoi	F	L	D	F	T	R	T	V	Q
Portulaca grandiflora	F	L	D	F	I	C	K		D
Portulaca pilosa	F	L	D	F	I	C	K		D
Portulaca oleracea	F	S	D	L	T	C	P	N	H
Pereskia aculeata	F	S	D	F	I	H	K	V	H
Rhipsalis teres									
Rhipsalis baccifera									
Selenicereus undatus									
Carnegiea gigantea									
Lophocereus schottii									
Mammillaria albiflora									
Mammillaria pectinifera									
Mammillaria solisioides									
Mammillaria crucigera									
Mammillaria huitzilopochtli									
Mammillaria supertexta									
Mammillaria zephyranthoides									
Brasiliopuntia brasiliensis	F	S	D	F	T	H	K	V	R
Opuntia monacantha	F	S	D	F	T	H	K	V	R
Opuntia quimilo	F	S	D	F	T	H	K	V	R
Opuntia ficus-indica	L	S	D	F	T	H	K	V	R
Spinacia oleracea	F	L	D	L	T	C	E	T	R

Figure 8. Sites under positive selection in plastid genes involved in the photosynthetic process. The amino acids are plotted across Cactineae phylogeny, based on plastid genes concatenated. Colored amino acids in orange mean the nucleotide is under positive selection.The amino acids positions are relative to *O. monacantha* plastid genes.

8. TABLES

Table 1. Comparison of plastome size and GC content between *O.monacantha* and *O. ficus-indica*, both subfamily Opuntioideae and other species of Cactaceae.

Species	Subfamily	Size (bp)	LSC (bp)	SSC (bp)	IR (bp)	Num-ber of genes	GC (%)	GenBank
Opuntia monacantha	Opuntioideae	149,076	101,384	4110	21791	110	36.6	MZ579523
Opuntia fícus-indica	Opuntioideae	152,387	87,248	4109	30,515	110	36.7	OK448352
Opuntia quimilo	Opuntioideae	150,347	101,475	4,115	22,392	109	36.6	MN114084
Brasiliopuntia brasiliensis	Opuntioideae	162,211	87,186	4,393	35,316	109	36.8	OK448351
Carnegiea gigantea	Cactoideae	113,064	-	-	-	99	36.7	NC_027618
Lophocereus schottii	Cactoideae	113,204	-	-	-	99	36.5	NC_041727
Mammillaria albiflora	Cactoideae	110,789	78,380	31,061	674	96	36.4	MN517610
Mammillaria pectinifera	Cactoideae	108,561	72,273	29,744	772	95	36.4	MN519716
Mammillaria crucigera	Cactoideae	115,505	71,565	29,418	7,261	96	36.3	MN517613
Mammillaria huitzilopochtli	Cactoideae	115,886	71,997	29,401	7,244	97	36.3	MN517612
Mammillaria solisioides	Cactoideae	115,356	71,690	29,238	7,214	95	36.4	MN518341
Mammillaria supertexta	Cactoideae	116,175	72,240	29,445	7,245	97	36.4	MN508963
Mammillaria zephyranthoides	Cactoideae	107,343	71,811	7,281	14,126	95	38.5	MN517611
Pereskia aculeata	Cactoideae	140,922	88,221	34,207	9,247	113	35.8	OK448353
Rhipsalis teres	Cactoideae	122,389	81,397	24,016	8,488	99	36.7	MT387452
Rhipsalis baccifera	Cactoideae	122,333	81,459	25,531	8,530	103	36.7	NC_05355
Selenicereus undatus	Cactoideae	133,326	68,256	21,715	21,676	100	36.4	NC_053681

Table 2. List of genes identified in the plastome of *Opuntia monacantha*.

Group of gene	Name of gene
Gene expression machinery	
Ribosomal RNA genes	*rrn16*[b]*; rrn23*[b]; *rrn5*[b]*; rrn4.5*[b]
Transfer RNA genes	*trnA* –UGC[ab]; *trnC* –GCA; *trnD* –GUC; *trnE* –UUC; *trnF* –GAA; *trnfM* –CAU; *trnG* –UCC[a]; *trnG* –GCC; *trnH* –GUG; *trnI* –CAU; *trnI* –GAU[ab]; *trnK* –UUU[a]; *trnL* –CAA; *trnL* –UAA[a]; *trnL* –UAG; *trnM* –CAU; *trnN* –GUU[b]; *trnP* –UGG; *trnQ* –UUG; *trnR* –ACG[b]; *trnR* –UCU; *trnS* –GCU; *trnS* –UGA; *trnS* –GGA; *trnT* –UGU; *trnT* –GGU; *trnV* –GAC[b]; *trnW* –CCA; *trnY* –GUA
Small subunit of ribosome	*rps2; rps3; rps4*; *rps7; rps8; rps11; rps12*[ac]; *rps14; rps15*[b]; *rps16*[a]; *rps18; rps19*
Large subunit of ribosome	*rpl2*; *rpl14; rpl16*[a]; *rpl20; rpl22; rpl23; rpl32*[b]*; rpl33; rpl36*
DNA-dependent RNA polymerase	*rpoA; rpoB; rpoC1*[a]; *rpoC2*
Translational initiation factor	*infA*
Genes for photosynthesis	
Subunits of photosystem I (PSI)	*psaA; psaB; psaC; psaI; psaJ; ycf3*[a]; *ycf4*
Subunits of photosystem II (PSII)	*psbA*; *psbB; psbC; psbD; psbE; psbF; psbH; psbI; psbJ; psbK; psbL; psbM; psbN; psbT; psbZ*
Subunits of cytochrome b_6f complex	*petA; petB*[a]*; petD*[a]*; petG; petL; petN*
Subunits of ATP synthase	*atpA; atpB; atpE; atpF*[a]; *atpH; atpI*
Subunits of NADH dehydrogenase	*ndhA*[ab]; *ndhB*[a]; *ndhC; ndhD; ndhE; ndhF*[b]*; ndhG*[b]*; ndhH*[b]; *ndhI*[b]*; ndhJ; ndhK*
Large subunit of Rubisco	*rbcL*
Others genes	
Maturase	*matK*
Envelope membrane protein	*cemA*
Subunit of acetyl-CoA carboxylase	*accD*
C-type cytochrome synthesis gene	*ccsA*
ATP-dependent Protease	*clpP*
Pseudogenes	*trnV* –UAC[a]*; ycf1*[b]*; ycf2*

[a] Genes containing introns; [b] Duplicated gene, [c] Partially duplicated genes

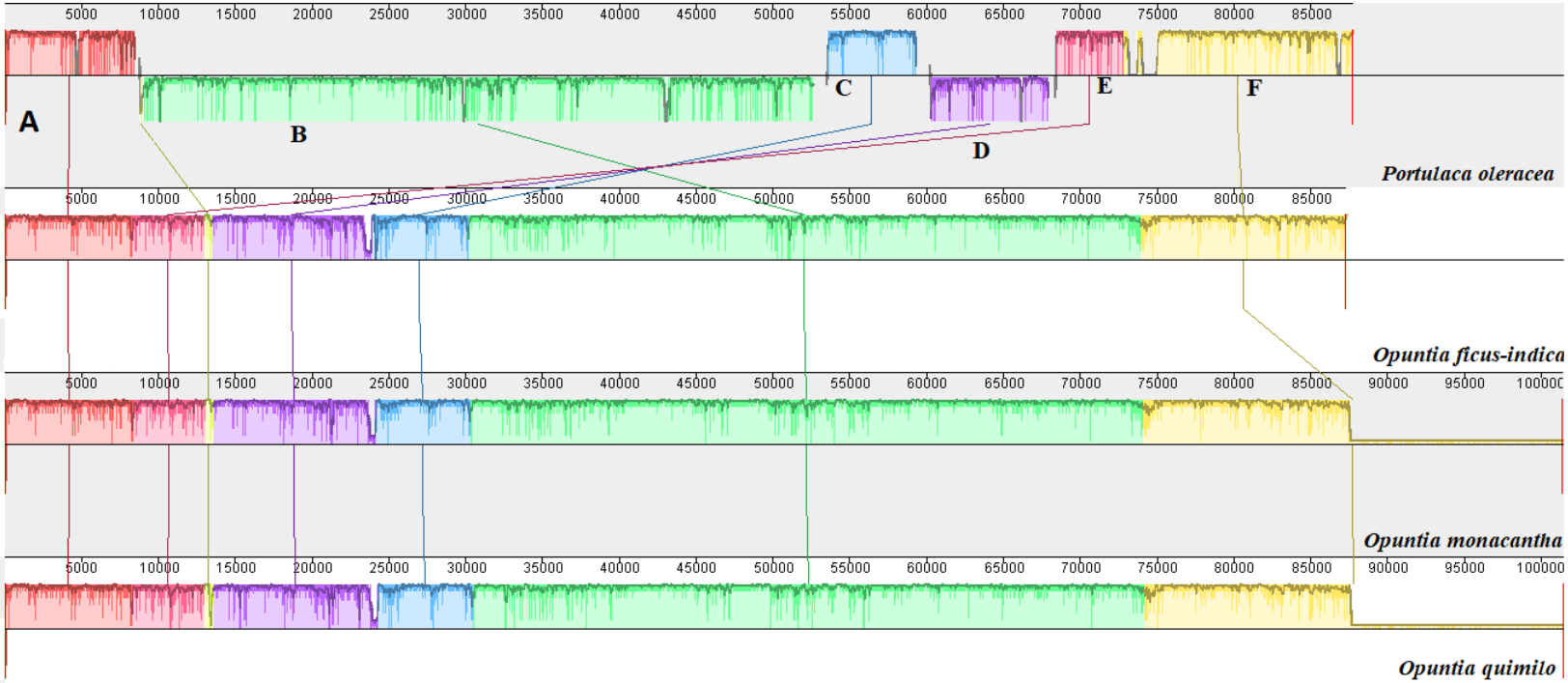
Table 2. List of genes identified in the plastome of *Opuntia ficus-indica*.

Group of gene	Name of gene
Gene expression machinery	
Ribosomal RNA genes	*rrn16*[b]*; rrn23*[b]; *rrn5*[b]*; rrn4.5*[b]
Transfer RNA genes	*trnA* –UGC[ab]; *trnC* –GCA; *trnD* –GUC; *trnE* –UUC; *trnF* –GAA; *trnfM* –CAU; *trnG* –UCC[a]; *trnG* –GCC; *trnH* –GUG; *trnI* –CAU[b]; *trnI* –GAU[ab]; *trnK* –UUU[a]; *trnL* –CAA[b]; *trnL* –UAA[a]; *trnL* –UAG; *trnM* –CAU; *trnN* –GUU[b]; *trnP* –UGG; *trnQ* –UUG; *trnR* –ACG[b]; *trnR* –UCU; *trnS* –GCU; *trnS* –UGA; *trnS* –GGA; *trnT* –UGU; *trnT* –GGU; *trnV* –GAC[b]; *trnW* –CCA; *trnY* –GUA
Small subunit of ribosome	*rps2; rps3; rps4*; *rps7*[b]*; rps8; rps11; rps12*[ac]; *rps14; rps15*[b]; *rps16*[a]; *rps18; rps19*[c]
Large subunit of ribosome	*rpl2* [b]; *rpl14; rpl16*[a]; *rpl20*; *rpl22; rpl23*[b]*; rpl32*[b]*; rpl33; rpl36*
DNA-dependent RNA polymerase	*rpoA; rpoB; rpoC1*[a]; *rpoC2*
Translational initiation factor	*infA*
Genes for photosynthesis	
Subunits of photosystem I (PSI)	*psaA; psaB; psaC; psaI; psaJ; ycf3*[a]; *ycf4*
Subunits of photosystem II (PSII)	*psbA*; *psbB; psbC; psbD; psbE; psbF; psbH; psbI; psbJ; psbK; psbL; psbM; psbN; psbT; psbZ*
Subunits of cytochrome b_6f complex	*petA; petB*[a]*; petD*[a]*; petG; petL; petN*
Subunits of ATP synthase	*atpA; atpB; atpE; atpF*[a]; *atpH; atpI*
Subunits of NADH dehydrogenase	*ndhA*[ab]; *ndhB*[ab]; *ndhC; ndhD; ndhE; ndhF*[b]*; ndhG*[b]*; ndhH*[b]; *ndhI*[b]*; ndhJ; ndhK*
Large subunit of Rubisco	*rbcL*
Others genes	
Maturase	*matK*
Envelope membrane protein	*cemA*
Subunit of acetyl-CoA carboxylase	*accD*
C-type cytochrome synthesis gene	*ccsA*
ATP-dependent Protease	*clpP*
Pseudogenes	*trnV* –UAC[a]*; ycf1*[b]
Absent	*ycf2*

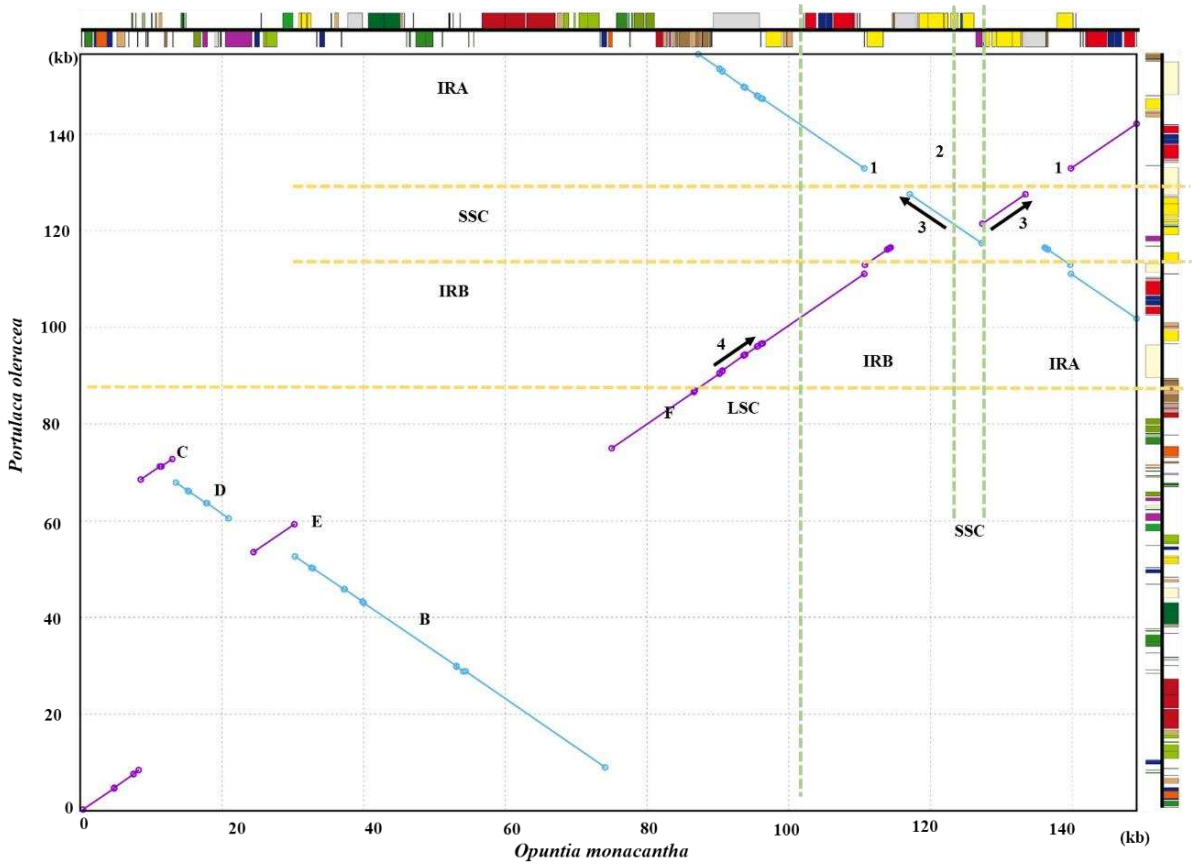
[a] Genes containing introns; [b] Duplicated gene, [c] Partially duplicated genes

9. SUPPLEMENTARY MATERIAL

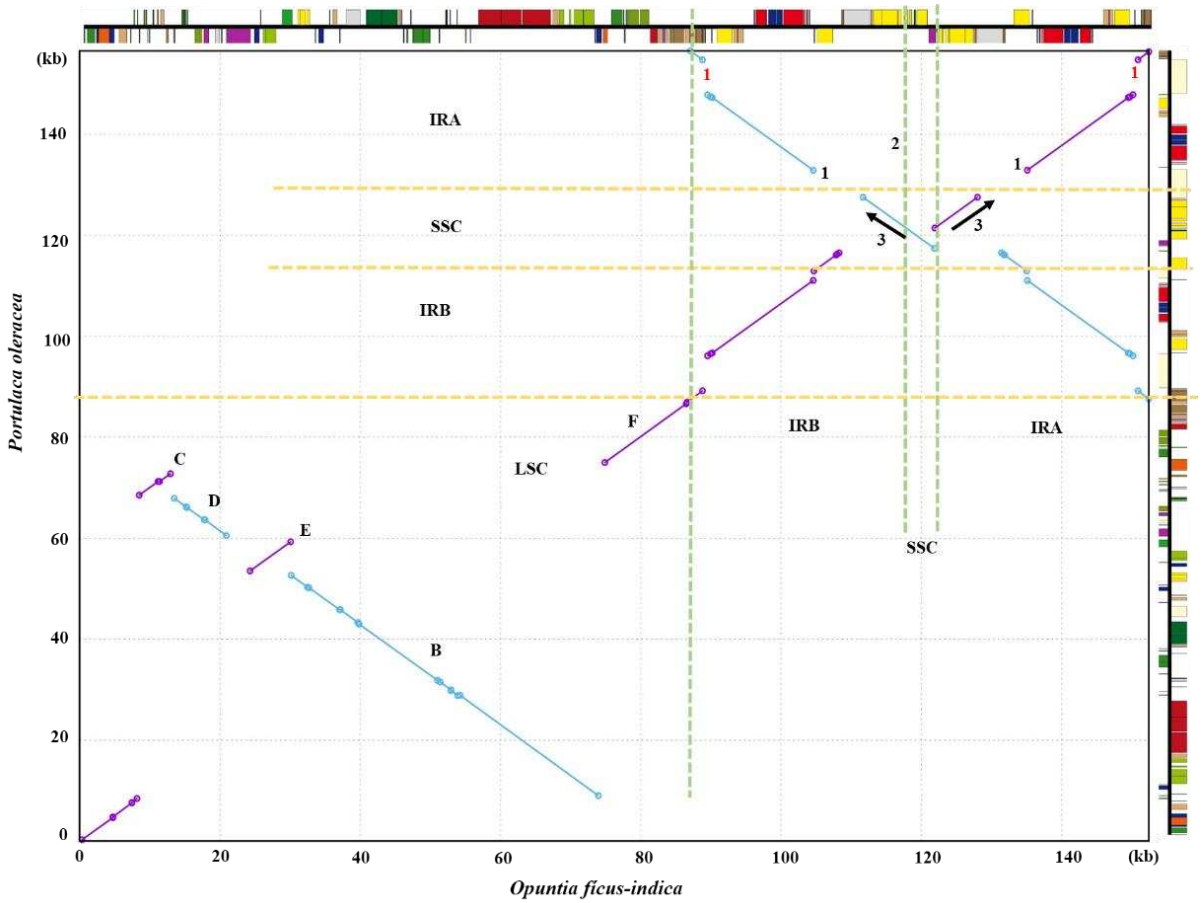
SUPPLEMENTARY FIGURES



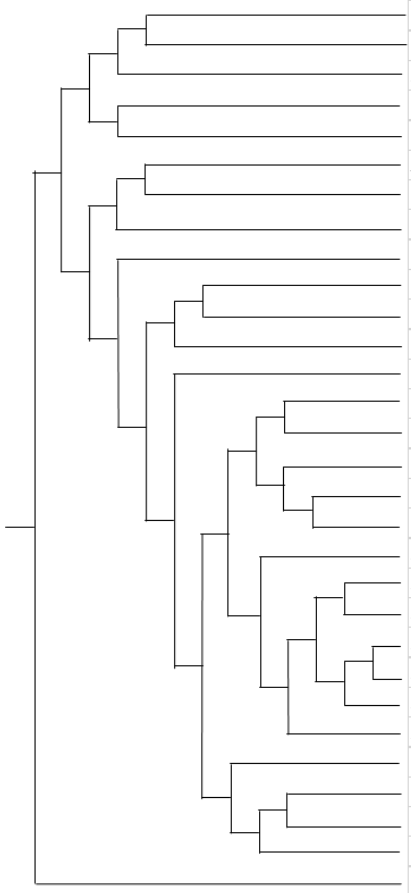
Supplementary Figure S1. Multiple alignment of *Opuntia ficus-indica*, *Opuntia monacantha*, *Opuntia quimilo* plastomes and *Portulaca oleracea* as reference. The region aligned corresponds region LSC of *P. oleracea* plastome, from the trnH-GUG to rps19 gene. Locally collinear blocks (LCBs) are color-coded. The LCBs related to inverted segments are drawn down. More details about the gene content of the LCBs involved in rearrangements are present in **Figure 3**.



Supplementary Figure S2. Dot-plot analyses comparing the plastomes of *Opuntia monacantha* (x-axis) against *Portulaca oleracea* (y-axis). A positive slope denotes that the pair of sequences compared is in the same orientation. A negative slope denotes that the pair of sequences compared can be aligned, but their orientation is opposite. Sequences in the same direction are purple and inversions are blue. The black arrows indicate contraction and expansions in the region IRB-SSC. Vertical green lines indicate the regions SSC, LSC and IRs in *O. monacantha*, while in *P. oleracea* is indicates by yellow horizontal lines the same regions. More details about regions highlighted with letters and numbers **Figure 3**.

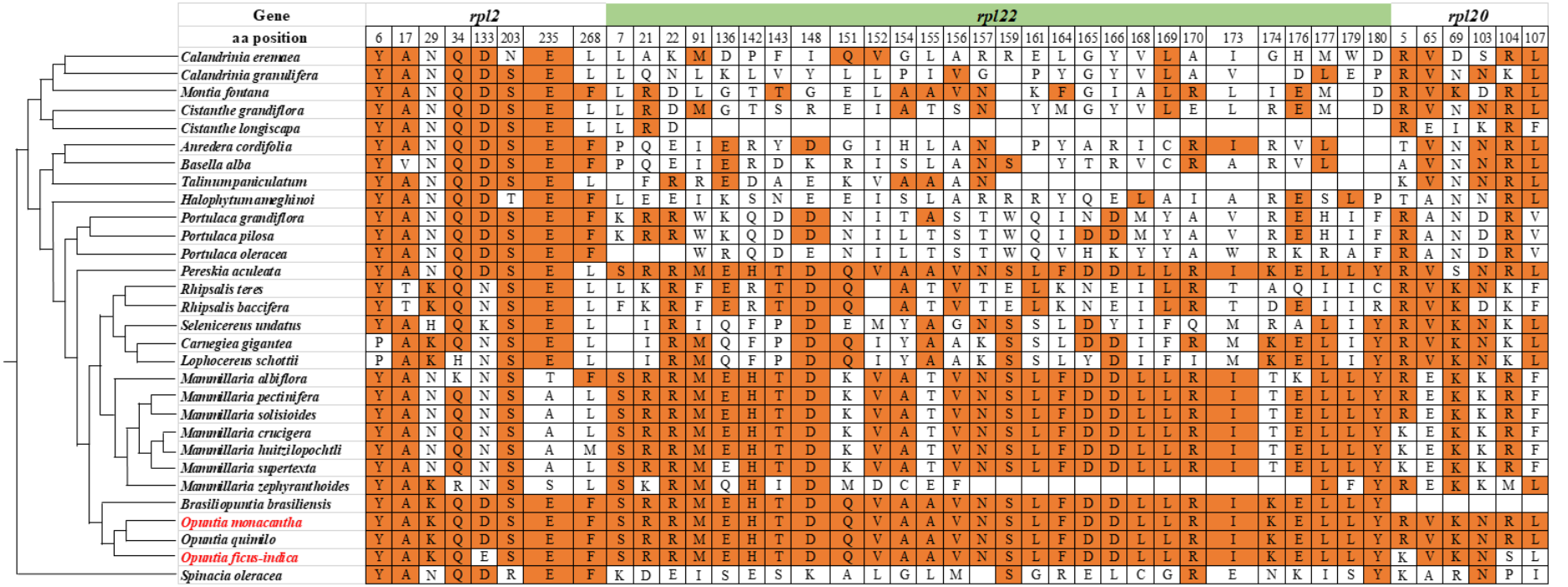


Supplementary Figure S3. Dot plot analyses comparing the plastomes of *Opuntia ficus-indica* (x-axis) against *Portulaca oleracea* (y-axis). A positive slope denotes that the pair of sequences compared is in the same orientation. A negative slope denotes that the pair of sequences compared can be aligned, but their orientation is opposite. Sequences in the same direction are purple and inversions are blue. The black arrows indicate contraction and expansions in the region IRB-SSC. Gene loss is pointed out by red number 1. Vertical green lines indicate the regions SSC, LSC and IRs in *O. ficus-indica*, while in *P. oleracea* is indicates by yellow horizontal lines the same regions. More details about regions highlighted with letters and numbers **Figure 3.**



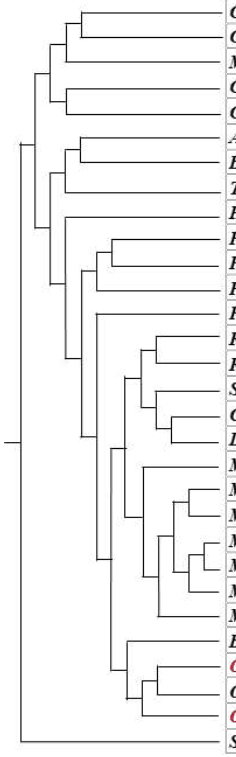
Gene	*rpoA*						*rpoB*	*rpoc2*									
aa position	261	266	294	319	334	335	340	163	240	405	522	556	866	915	996	1008	1055
Calandrinia eremaea	N	S	L	F	K	R	C	L	F	F	E	C	L	I	F	Y	H
Calandrinia granulifera	K	A	L	F	Q	N	C	L	F	F	E	S	L	I	F	Y	H
Montia fontana	K	A	L	F	E	K	C	L	F	F	D	F	L	L	L	L	Y
Cistanthe grandiflora	N	A	L	L	K	N	C	L	F	F	E	V	L	I	F	Y	Q
Cistanthe longiscapa	N	A	L	L	K	N	C	L	F	F	E	V	L	I	F	Y	Q
Anredera cordifolia	N	A	L	F	K	R	C	L	F	F	E	F	L	I	F	Y	H
Basella alba	N	A	L	F	K	N	C	L	F	F	E	F	L	I	L	Y	H
Talinum paniculatum	N	A	L	L	K	N	C	L	F	F	D	L	L	L	I	Y	H
Halophytum ameghinoi	N	A	S	L	K	N	C	L	F	F	E	F	L	L	F	Y	H
Portulaca grandiflora	K	R	L	L	Q	R	C	L	F	F	D	L	L	I	F	Y	H
Portulaca pilosa	K	R	L	L	Q	R	C	L	F	F	D	L	L	I	F	Y	H
Portulaca oleracea	N	T	L	L	Q	N	C	L	F	F	E	F	L	I	F	Y	H
Pereskia aculeata	N	P	L	L	E	K	W	L	F	I	D	F	L	L	L	Y	H
Rhipsalis teres	K	A	L	L			C	V	F	F	E	L	S	R	L	S	H
Rhipsalis baccifera	K	A	W	L			C	V	F	F	D	L	L	L	L	S	H
Selenicereus undatus	K	A	L	W			C	P	L	C	D	L	F	R	F	S	H
Carnegiea gigantea	K	A	L	L			C	V	F	C	E	L	F	I	F	Y	H
Lophocereus schottii	K	A	L	L			C	V	F	C	E	L	L	L	F	Y	H
Mammillaria albiflora	N	E	S	L	M	K	S	V	F	I	E	L	L	L	F	Y	H
Mammillaria pectinifera	K	E	W	L	K	K	R	L	L	L	E	L	L	L	F	S	H
Mammillaria solisioides	K	E	W	L	K	K	R	L	L	L	E	L	L	L	F	S	H
Mammillaria crucigera	K	E	W	L	M	K	S	V	L	L	E	L	L	L	F	S	H
Mammillaria huitzilopochtli	M	E	W	L	M	K	S	V	L	L	E	L	L	L	F	S	H
Mammillaria supertexta	K	E	W	L	K	N	S	V	L	L	D	L	L	L	F	S	H
Mammillaria zephyranthoides	N	E	S	L	M	K	D	L	W	L	D	L	F	L	F	S	
Brasiliopuntia brasiliensis	N	A	L	L	E	K	C	L	F	L	D	F	L	L	F	Y	H
Opuntia monacantha	N	A	L	L	E	K	C	L	F	L	D	F	L	L	F	Y	H
Opuntia quimilo	N	A	L	L			C	L	F	L	D	F	L	L	F	Y	H
Opuntia ficus-indica	N	E	L	Y	E	K	C	L	F	L	D	F	L	L	F	Y	R
Spinacia oleracea	N	A	L	F	R		S	L	L	F	D	L	S	L	F	Y	K

Supplementary Figure S4. Sites under positive selection in plastid genes involved in the subunits DNA- dependent RNA polymerase. The amino acids are plotted across Cactineae phylogeny, based on plastid genes concatenated. Colored amino acids in orange mean the nucleotide is under positive selection.The amino acids positions are relative to *O. monacantha* plastid genes.



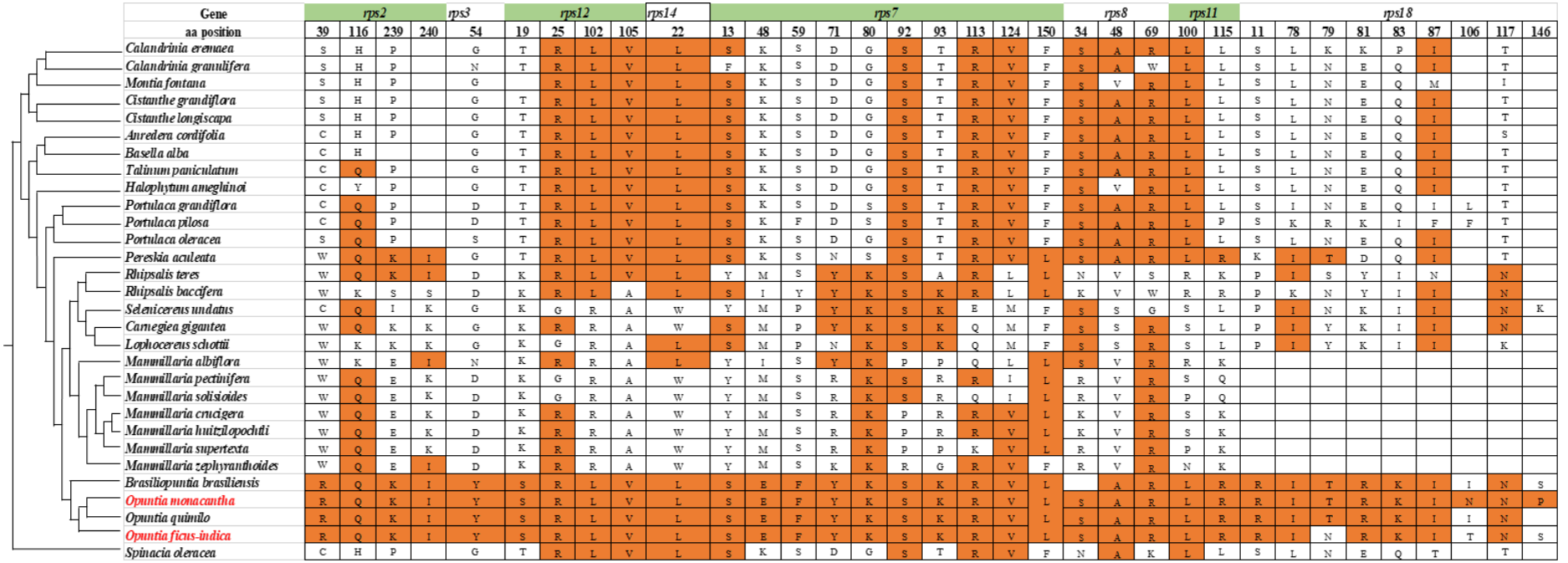
Gene	*rpl2*								*rpl22*																												*rpl20*					
aa position	6	17	29	34	133	203	235	268	7	21	22	91	136	142	143	148	151	152	154	155	156	157	159	161	164	165	166	168	169	170	173	174	176	177	179	180	5	65	69	103	104	107
Calandrinia eremaea	Y	A	N	Q	D	N	E	L	L	A	K	M	D	P	F	I	Q	V	G	L	A	R	R	E	L	G	Y	V	L	A	I	G	H	M	W	D	R	V	D	S	R	L
Calandrinia granulifera	Y	A	N	Q	D	S	E	L	L	Q	N	L	K	L	V	Y	L	L	P	I	V	G		P	Y	G	Y	V	L	A	V		D	L	E	P	R	V	N	N	K	L
Montia fontana	Y	A	N	Q	D	S	E	F	L	R	D	L	G	T	T	G	E	L	A	A	V	N		K	F	G	I	A	L	R	L	I	E	M		D	R	V	K	D	R	L
Cistanthe grandiflora	Y	A	N	Q	D	S	E	L	L	R	D	M	G	T	S	R	E	I	A	T	S	N		Y	M	G	Y	V	L	E	L	R	E	M		D	R	V	N	N	R	L
Cistanthe longiscapa	Y	A	N	Q	D	S	E	L	L	R	D																										R	E	I	K	R	F
Anredera cordifolia	Y	A	N	Q	D	S	E	F	P	Q	E	I	E	R	Y	D	G	I	H	L	A	N		P	Y	A	R	I	C	R	I	R	V	L			T	V	N	N	R	L
Basella alba	Y	V	N	Q	D	S	E	F	P	Q	E	I	E	R	D	K	R	I	S	L	A	N	S		Y	T	R	V	C	R	A	R	V	L			A	V	N	N	R	L
Talinum paniculatum	Y	A	N	Q	D	S	E	L		F	R	R	E	D	A	E	K	V	A	A	A	N																K	V	N	N	R	L
Halophytum ameghinoi	Y	A	N	Q	D	T	E	F	L	E	E	I	K	S	N	E	E	I	S	L	A	R	R	R	Y	Q	E	L	A	I	A	R	E	S	L	P	T	A	N	N	R	L
Portulaca grandiflora	Y	A	N	Q	D	S	E	F	K	R	R	W	K	Q	D	D	N	I	T	A	S	T	W	Q	I	N	D	M	Y	A	V	R	E	H	I	F	R	A	N	D	R	V
Portulaca pilosa	Y	A	N	Q	D	S	E	F	K	R	R	W	K	Q	D	D	N	I	L	T	S	T	W	Q	I	D	D	M	Y	A	V	R	E	H	I	F	R	A	N	D	R	V
Portulaca oleracea	Y	A	N	Q	D	S	E	F				W	R	Q	D	E	N	I	L	T	S	T	W	Q	V	H	K	Y	Y	A	W	R	K	R	A	F	R	A	N	D	R	V
Pereskia aculeata	Y	A	N	Q	D	S	E	L	S	R	R	M	E	H	T	D	Q	V	A	A	V	N	S	L	F	D	D	L	L	R	I	K	E	L	L	Y	R	V	S	N	R	L
Rhipsalis teres	Y	T	K	Q	N	S	E	L	L	K	R	F	E	R	T	D	Q		A	T	V	T	E	L	K	N	E	I	L	R	T	A	Q	I	I	C	R	V	K	N	K	F
Rhipsalis baccifera	Y	T	K	Q	N	S	E	L	F	K	R	F	E	R	T	D	Q		A	T	V	T	E	L	K	N	E	I	L	R	T	D	E	I	I	R	R	V	K	D	K	F
Selenicereus undatus	Y	A	H	Q	K	S	E	L		I	R	I	Q	F	P	D	E	M	Y	A	G	N	S	S	L	D	Y	I	F	Q	M	R	A	L	I	Y	R	V	K	N	K	L
Carnegiea gigantea	P	A	K	Q	N	S	E	L		I	R	M	Q	F	P	D	Q	I	Y	A	A	K	S	S	L	D	D	I	F	R	M	K	E	L	I	Y	R	V	K	N	K	L
Lophocereus schottii	P	A	K	H	N	S	E	L		I	R	M	Q	F	P	D	Q	I	Y	A	A	K	S	S	L	Y	D	I	F	I	M	K	E	L	I	Y	R	V	K	N	K	L
Mammillaria albiflora	Y	A	N	K	N	S	T	F	S	R	R	M	E	H	T	D	K	V	A	T	V	N	S	L	F	D	D	L	L	R	I	T	K	L	L	Y	R	E	K	K	R	F
Mammillaria pectinifera	Y	A	N	Q	N	S	A	L	S	R	R	M	E	H	T	D	K	V	A	T	V	N	S	L	F	D	D	L	L	R	I	T	E	L	L	Y	R	E	K	K	R	F
Mammillaria solisioides	Y	A	N	Q	N	S	A	L	S	R	R	M	E	H	T	D	K	V	A	T	V	N	S	L	F	D	D	L	L	R	I	T	E	L	L	Y	R	E	K	K	R	F
Mammillaria crucigera	Y	A	N	Q	N	S	A	L	S	R	R	M	E	H	T	D	K	V	A	T	V	N	S	L	F	D	D	L	L	R	I	T	E	L	L	Y	K	E	K	K	R	F
Mammillaria huitzilopochtli	Y	A	N	Q	N	S	A	M	S	R	R	M	E	H	T	D	K	V	A	T	V	N	S	L	F	D	D	L	L	R	I	T	E	L	L	Y	K	E	K	K	R	F
Mammillaria supertexta	Y	A	N	Q	N	S	A	L	S	R	R	M	E	H	T	D	K	V	A	T	V	N	S	L	F	D	D	L	L	R	I	T	E	L	L	Y	K	E	K	K	R	F
Mammillaria zephyranthoides	Y	A	K	R	N	S	S	L	S	K	R	M	Q	H	I	D	M	D	C	E	F														L	F	Y	R	E	K	K	M	L
Brasiliopuntia brasiliensis	Y	A	K	Q	D	S	E	F	S	R	R	M	E	H	T	D	Q	V	A	A	V	N	S	L	F	D	D	L	L	R	I	K	E	L	L	Y						
Opuntia monacantha	Y	A	K	Q	D	S	E	F	S	R	R	M	E	H	T	D	Q	V	A	A	V	N	S	L	F	D	D	L	L	R	I	K	E	L	L	Y	R	V	K	N	R	L
Opuntia quimilo	Y	A	K	Q	D	S	E	F	S	R	R	M	E	H	T	D	Q	V	A	A	V	N	S	L	F	D	D	L	L	R	I	K	E	L	L	Y	R	V	K	N	R	L
Opuntia ficus-indica	Y	A	K	Q	E	S	E	F	S	R	R	M	E	H	T	D	Q	V	A	A	V	N	S	L	F	D	D	L	L	R	I	K	E	L	L	Y	K	V	K	N	S	L
Spinacia oleracea	Y	A	N	Q	D	R	E	F	K	D	E	I	S	E	S	K	A	L	G	L	M		S	G	R	E	L	C	G	R	E	N	K	I	S	Y	K	A	R	N	P	I

Supplementary Figure S5. Sites under positive selection in plastid genes involved in expression machinery. The amino acids are plotted across Cactineae phylogeny, based on plastid genes concatenated. Colored amino acids in orange mean the nucleotide is under positive selection.The amino acids positions are relative to *O. monacantha* plastid genes.



Gene	*rpl23*														*rpl32*										
aa position	66	68	69	71	72	73	74	75	81	82	85	86	87	88	11	49	52	73	75	77	83	84	85	86	88	89
Calandrinia eremaea															T	E		K	F	F	E	E	Q	E	A	
Calandrinia granulifera															S	Q		K	N							
Montia fontana	M	T	Y	A	L	Q	T	Y	D	H	S	I	G	L	A	Q		K	N	L	A	L	K			
Cistanthe grandiflora	M	H	T	H	Y	R	R	M	S	G	I	P	P	L	T	L		K	N	F						
Cistanthe longiscapa	M	H	T	H	Y	R	R	M		G		P	P	L	A	Q		K	I	F						
Anredera cordifolia	M	H	T	H	Y	R	R	M	S	G	I	P	P	L	S	Q		K	N							
Basella alba	M	H	T	H	Y	R	R	M	S	G	I	P	P	L	S	Q		K	N							
Talinum paniculatum															S	K		T	K	E	I	F	T	Q	T	I
Halophytum ameghinoi	M	H	T	H	Y	R	R	M	S	G	I	S	P	L												
Portulaca grandiflora															S	E		K	N	S	I	L	E			
Portulaca pilosa				H	Y	R	R	M	S	G	I	P	P	L												
Portulaca oleracea	M	H	T	H	Y	R	R	M	S	G	I	P	P	L	S	E		K	N	S	I	L	E			
Pereskia aculeata	L	H	T	H	Y	R	R	M	S	G	I	L	P	H	S	K		K	F	E	Y	E	Q	S	S	A
Rhipsalis teres															S	S	G	R	I	L	L	E	I	Q	I	
Rhipsalis baccifera															S	S	G	R	I	p	L	E	I	Q	I	
Selenicereus undatus																										
Carnegiea gigantea															S			K	L		F	E	Q	S	L	A
Lophocereus schottii															S			K	L		F	E	Q	S	L	A
Mammillaria albiflora															S	K		K	F	E	Y	E	Q	S	S	A
Mammillaria pectinifera															S	K		K	F	E	Y	E	Q	S	S	A
Mammillaria solisioides															S	K		K	F	E	Y	E	Q	S	S	A
Mammillaria crucigera															S	K		K	F	E	Y	E	Q	S	S	A
Mammillaria huitzilopochtli															S	K		K	F	E	Y	E	Q	S	S	A
Mammillaria supertexta															S	K		K	F	E	Y	E	Q	S	S	A
Mammillaria zephyranthoides																										
Brasiliopuntia brasiliensis	L	H	T	H	Y	R	R	M	S	G	I	L	P	H	S	K	K	L	T							
Opuntia monacantha	L	H	T	H	Y	R	R	M	S	G	I	L	L	H	S	K	K	K	L	T	A	S	Y	S	L	K
Opuntia quimilo															S	K	K	K	F	E	Y	E	Q	S	Y	E
Opuntia ficus-indica	S	H	T	H	Y	R	R	M	S	G	I	L	L	H	S	K	K	K	F	E	Y	E	Q	S	Y	E
Spinacia oleracea															Y	R		K	S	Q	T	L	E			

Supplementary Figure S6. Sites under positive selection in plastid genes involved in expression machinery.The amino acids are plotted across Cactineae phylogeny, based on plastid genes concatenated. Colored amino acids in orange mean the nucleotide is under positive selection.The amino acids positions are relative to *O. monacantha* plastid genes.



Gene	*rps2*				*rps3*	*rps12*				*rps14*	*rps7*										*rps8*			*rps11*		*rps18*								
aa position	39	116	239	240	54	19	25	102	105	22	13	48	59	71	80	92	93	113	124	150	34	48	69	100	115	11	78	79	81	83	87	106	117	146
Calandrinia erenaea	S	H	P		G	T	R	L	V	L	S	K	S	D	G	S	T	R	V	F	S	A	R	L	L	S	L	K	K	P	I		T	
Calandrinia granulifera	S	H	P		N	T	R	L	V	L	F	K	S	D	G	S	T	R	V	F	S	A	W	L	L	S	L	N	E	Q	I		T	
Montia fontana	S	H	P		G		R	L	V	L	S	K	S	D	G	S	T	R	V	F	S	V	R	L	L	S	L	N	E	Q	M		I	
Cistanthe grandiflora	S	H	P		G	T	R	L	V	L	S	K	S	D	G	S	T	R	V	F	S	A	R	L	L	S	L	N	E	Q	I		T	
Cistanthe longiscapa	S	H	P		G	T	R	L	V	L	S	K	S	D	G	S	T	R	V	F	S	A	R	L	L	S	L	N	E	Q	I		T	
Anredera cordifolia	C	H	P		G	T	R	L	V	L	S	K	S	D	G	S	T	R	V	F	S	A	R	L	L	S	L	N	E	Q	I		S	
Basella alba	C	H			G	T	R	L	V	L	S	K	S	D	G	S	T	R	V	F	S	A	R	L	L	S	L	N	E	Q	I		T	
Talinum paniculatum	C	Q	P		G	T	R	L	V	L	S	K	S	D	G	S	T	R	V	F	S	A	R	L	L	S	L	N	E	Q	I		T	
Halophytum ameghinoi	C	Y	P		G	T	R	L	V	L	S	K	S	D	G	S	T	R	V	F	S	V	R	L	L	S	L	N	E	Q	I		T	
Portulaca grandiflora	C	Q	P		D	T	R	L	V	L	S	K	S	D	S	S	T	R	V	F	S	A	R	L	L	S	I	N	E	Q	I	L	T	
Portulaca pilosa	C	Q	P		D	T	R	L	V	L	S	K	F	D	S	S	T	R	V	F	S	A	R	L	P	S	K	R	K	I	F	F	T	
Portulaca oleracea	S	Q	P		S	T	R	L	V	L	S	K	S	D	G	S	T	R	V	F	S	A	R	L	L	S	L	N	E	Q	I		T	
Pereskia aculeata	W	Q	K	I	G	T	R	L	V	L	S	K	S	N	S	S	T	R	V	L	S	A	R	L	R	K	I	T	D	Q	I		T	
Rhipsalis teres	W	Q	K	I	D	K	R	L	V	L	Y	M	S	Y	K	S	A	R	L	L	N	V	S	R	K	P	I	S	Y	I	N		N	
Rhipsalis baccifera	W	K	S	S	D	K	R	L	A	L	S	I	Y	Y	K	S	K	R	L	L	K	V	W	R	R	P	K	N	Y	I	I		N	
Selenicereus undatus	C	Q	I	K	G	K	G	R	A	W	Y	M	P	Y	K	S	K	E	M	F	S	S	G	S	L	P	I	N	K	I	I		N	K
Carnegiea gigantea	W	Q	K	K	G	K	R	R	A	W	S	M	P	Y	K	S	K	Q	M	F	S	S	R	S	L	P	I	Y	K	I	I		N	
Lophocereus schottii	W	K	K	K	G	K	G	R	A	L	S	M	P	N	K	S	K	Q	M	F	S	S	R	S	L	P	I	Y	K	I	I		K	
Mammillaria albiflora	W	K	E	I	N	K	R	R	A	L	Y	I	S	Y	K	P	P	Q	L	L	S	V	R	R	K									
Mammillaria pectinifera	W	Q	E	K	D	K	G	R	A	W	Y	M	S	R	K	S	R	R	I	L	R	V	R	S	Q									
Mammillaria solisioides	W	Q	E	K	D	K	G	R	A	W	Y	M	S	R	K	S	R	Q	I	L	R	V	R	P	Q									
Mammillaria crucigera	W	Q	E	K	D	K	R	R	A	W	Y	M	S	R	K	P	R	R	V	L	K	V	R	S	K									
Mammillaria huitzilopochtli	W	Q	E	K	D	K	R	R	A	W	Y	M	S	R	K	P	R	R	V	L	K	V	R	S	K									
Mammillaria supertexta	W	Q	E	K	D	K	R	R	A	W	Y	M	S	R	K	P	P	K	V	L	R	V	R	P	K									
Mammillaria zephyranthoides	W	Q	E	I	D	K	R	R	A	W	Y	M	S	K	K	R	G	R	V	F	R	V	R	N	K									
Brasiliopuntia brasiliensis	R	Q	K	I	Y	S	R	L	V	L	S	E	F	Y	K	S	K	R	V	L		A	R	L	R	R	I	T	R	K	I	I	N	S
Opuntia monacantha	R	Q	K	I	Y	S	R	L	V	L	S	E	F	Y	K	S	K	R	V	L	S	A	R	L	R	R	I	T	R	K	I	N	N	P
Opuntia quimilo	R	Q	K	I	Y	S	R	L	V	L	S	E	F	Y	K	S	K	R	V	L	S	A	R	L	R	R	I	T	R	K	I	I	N	
Opuntia ficus-indica	R	Q	K	I	Y	S	R	L	V	L	S	E	F	Y	K	S	K	R	V	L	S	A	R	L	R	R	I	N	R	K	I	T	N	S
Spinacia oleracea	C	H	P		G	T	R	L	V	L	S	K	S	D	G	S	T	R	V	F	N	A	K	L	L	S	L	N	E	Q	T		T	

Supplementary Figure S7. Sites under positive selection in plastid genes involved in expression machinery.The amino acids are plotted across Cactineae phylogeny, based on plastid genes concatenated. Colored amino acids in orange mean the nucleotide is under positive selection.The amino acids positions are relative to *O. monacantha* plastid genes.

SUPPLEMENTARY TABLES

Supplementary Table S1. List of the taxa sampled in the phylogenetic analysis (*) Outgroup.

Species	**Family**	**GenBank**
Opuntia monacantha	Cactaceae	MZ579523
Opuntia ficus indica	Cactaceae	OK448352
Opuntia quimilo	Cactaceae	MN114084
Brasiliopuntia brasiliensis	Cactaceae	OK448351
Carnegiea gigantea	Cactaceae	NC_027618.1
Lophocereus schottii	Cactaceae	NC_041727.1
Mammillaria albiflora	Cactaceae	MN517610
Mammillaria pectinifera	Cactaceae	MN519716
Mammillaria crucigera	Cactaceae	MN517613
Mammillaria huitzilopochtli	Cactaceae	MN517612
Mammillaria solisioides	Cactaceae	MN518341
Mammillaria supertexta	Cactaceae	MN508963
Mammillaria zephyranthoides	Cactaceae	MN517611
Pereskia aculeata	Cactaceae	OK448353
Rhipsalis teres	Cactaceae	MT387452
Rhipsalis baccifera	Cactaceae	NC_05355
Selenicereus undatus	Cactaceae	NC_053681
Anredera cordifolia	Basellaceae	NC_041274.1
Basella alba	Basellaceae	NC_041293.1
Calandrinia eremaea	Montiaceae	NC_041259.1
Calandrinia granulifera	Montiaceae	NC_041260.1
Cistanthe grandiflora	Montiaceae	NC_041295.1
Cistanthe longiscapa	Montiaceae	NC_035140.1
Montia fontana	Montiaceae	NC_041269.1
Halophytum ameghinoi	Halophytaceae	NC_040949.1
Portulaca grandiflora	Portulacaceae	NC_041299.1
Portulaca oleracea	Portulacaceae	NC_036236.1
Portulaca pilosa	Portulacaceae	NC_041264.1
Talinum paniculatum	Talinaceae	NC_037748.1
*Spinacia oleracea**	Chenopodiaceae	NC_002202.1

Supplementary Table S2. List of simple sequence repeats (SSRs) identified in the plastome of *O. monacantha*

SSR sequence	Number of repetitions															Total
	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
A/T	-	-	-	-	-	80	42	23	4	8	3	3	3	1	1	168
C/G	-	-	-	-	-	1	1	1	-	-	-	-	-	-	-	3
AG/CT	-	10	-	-	-	-	-	-	-	-	-	-	-	-	-	10
AT/AT	-	24	6	1	-	-	1	1	-	-	-	-	-	-	-	33
AAG/CTT	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	2
AAT/ATT	-	4	-	-	-	-	-	-	-	-	-	-	-	-	-	4
AAAC/GTTT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AAAG/CTTT	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5
AAAT/ATTT	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4
AATC/ATTG	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
AATT/AATT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ACCT/AGGT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ATCC/ATGG	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AAAAC/GTTTT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AAAGT/ACTTT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AAATT/AATTT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ACCCC/GGGGT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ACTAG/AGTCT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
																240

Supplementary Table S3. Distribution of SSR loci in the plastome of *O. monacantha*

SSR type	SSR	Size	Start	End	Location
compound	(ATAA)2(A)8	19	258	276	*trnH-GUG*/*psbA* (IGS)
mono	(A)13	13	380	392	*trnH-GUG*/*psbA* (IGS)
mono	(T)9	9	2028	2036	*trnK-UUU* (intron)
mono	(A)8	8	2613	2620	*matK* (CDS)
mono	(T)9	9	3264	3272	*matK* (CDS)
mono	(A)10	10	3648	3657	*trnk-UUU*(intron)*)/rps16* (IGS)
mono	(T)8	8	4003	4010	*trnk-UUU*(intron)*)/rps16* (IGS)
mono	(T)8	8	4574	4581	*trnk-UUU/rps16* (IGS)
mono	(A)8	8	4631	4638	*trnk-UUU/rps16* (IGS)
tri	(TAA)4	12	4789	4800	*trnk-UUU/rps16* (IGS)
mono	(A)8	8	4874	4881	*trnk-UUU/rps16* (IGS)
mono	(T)12	12	4980	4991	*trnk-UUU/rps16* (IGS)
mono	(A)10	10	5451	5460	*rps16* (intron)
di	(AT)4	8	5753	5760	*rps16* (intron)
di	(AT)4	8	6233	6240	*rps16*/ *trnQ-UUG* (IGS)

mono	(T)12	12	6255	6266	*rps16*/ *trnQ-UUG* (IGS)
mono	(A)8	8	6281	6288	*rps16*/ *trnQ-UUG* (IGS)
mono	(A)17	17	6993	7009	*trnQ-UUG/psbK* (IGS)
mono	(A)8	8	7506	7513	*psbK*/*psbI* (IGS)
compound	(A)9(AAT)3	20	7605	7624	*psbK*/*psbI* (IGS)
mono	(A)9	9	7807	7815	*psbI*/*trnS-GCU* (IGS)
di	(TA)4	8	7846	7853	*psbI*/*trnS-GCU* (IGS)
di	(AT)4	8	8113	8120	*trnS-GCU*/*petL* (IGS)
di	(AT)5	10	8268	8277	*trnS-GCU*/*petL* (IGS)
mono	(T)8	8	8345	8352	*trnS-GCU*/*petL* (IGS)
mono	(A)13	13	8537	8549	*trnS-GCU*/*petL* (IGS)
mono	(T)9	9	8832	8840	*petL*/*petG* (IGS)
mono	(A)8	8	9276	9283	*trnW-CCA*/*trnP-UGG* (IGS)
di	(AT)4	8	9573	9580	*trnP-UGG*/psaJ (IGS)
mono	(A)12	12	9655	9666	*trnP-UGG*/psaJ (IGS)
mono	(A)8	8	9823	9830	*trnP-UGG*/psaJ (IGS)
di	(AT)5	10	10792	10801	*rpl33*/*rps18* (IGS)
mono	(T)10	10	10858	10867	*rpl33*/*rps18* (IGS)
mono	(T)9	9	11566	11574	*rps18*/*rpl20* (IGS)
mono	(T)8	8	11590	11597	*rps18*/*rpl20* (IGS)
mono	(T)8	8	12010	12017	*rpl20* (CDS)
mono	(T)8	8	12224	12231	*rpl20*/*rps12* (IGS)
di	(AT)4	8	12980	12987	*rps12*/*psbE* (IGS)
di	(TA)5	10	13361	13370	*rps12*/*psbE* (IGS)
mono	(A)8	8	13516	13523	*rps12*/*psbE* (IGS)
mono	(T)10	10	15162	15171	*psbJ*/*petA* (IGS)
mono	(A)14	14	15288	15301	*psbJ*/*petA* (IGS)
penta	(TAGTC)3	15	15501	15515	*psbJ*/*petA* (IGS)
mono	(A)10	10	15629	15638	*psbJ*/*petA* (IGS)
mono	(A)8	8	15641	15648	*psbJ*/*petA* (IGS)
mono	(T)8	8	16382	16389	*petA* (CDS)
di	(TA)4	8	16741	16748	*petA* (CDS)
mono	(T)8	8	17773	17780	*cemA*/*ycf4* (IGS)
mono	(T)8	8	17900	17907	*cemA*/*ycf4* (IGS)
tetra	(ATCA)3	12	18146	18157	*cemA*/*ycf4* (IGS)
penta	(CAAAA)3	15	18240	18254	*cemA*/*ycf4* (IGS)
mono	(T)11	11	18484	18494	*cemA*/*ycf4* (IGS)
mono	(A)10	10	18583	18592	*cemA*/*ycf4* (IGS)
mono	(A)8	8	19130	19137	*ycf4* (CDS)
mono	(G)8	8	19147	19154	*ycf4* (CDS)
mono	(T)10	10	19767	19776	*psaI*/*accD* (IGS)
di	(TA)4	8	19850	19857	*psaI*/*accD* (IGS)
mono	(A)8	8	19970	19977	*psaI*/*accD* (IGS)
mono	(T)8	8	20088	20095	*psaI*/*accD* (IGS)

mono	(A)8	8	20118	20125	*psaI*/*accD* (IGS)
mono	(T)8	8	20141	20148	*psaI*/*accD* (IGS)
di	(AT)4	8	20197	20204	*psaI*/*accD* (IGS)
di	(AT)4	8	20625	20632	*accD* (CDS)
mono	(T)8	8	20638	20645	*accD* (CDS)
tri	(TCT)5	15	20963	20977	*accD* (CDS)
mono	(A)8	8	22639	22646	*accD* (CDS)
mono	(A)9	9	24397	24405	*accD*/*trnV-UAC* (IGS)
mono	(T)9	9	24553	24561	*trnV-UAC* (intron)
mono	(T)8	8	25005	25012	*trnV-UAC* (intron)
mono	(T)9	9	25145	25153	*trnV-UAC*/*trnM-CAU* (IGS)
di	(TA)4	8	25441	25448	*trnM-CAU*/*atpE* (IGS)
mono	(T)10	10	25504	25513	*trnM-CAU*/*atpE* (IGS)
mono	(T)10	10	27501	27510	*atpB* (CDS)
mono	(T)8	8	27711	27718	*atpB*/*rbcL* (IGS)
mono	(A)12	12	27822	27833	*atpB*/*rbcL* (IGS)
di	(AT)4	8	28185	28192	*atpB*/*rbcL* (IGS)
di	(GA)4	8	28864	28871	*rbcL* (CDS)
mono	(T)9	9	30347	30355	*rbcL*/*ndhC* (IGS)
mono	(T)8	8	31748	31755	*ndhK*/*ndhJ* (IGS)
mono	(A)9	9	31799	31807	*ndhK*/*ndhJ* (IGS)
tetra	(AAAT)3	12	31845	31856	*ndhK*/*ndhJ* (IGS)
penta	(TGGGG)3	15	32643	32657	*ndhJ*/*trnF-GAA* (IGS)
mono	(C)10	10	32708	32717	*ndhJ*/*trnF-GAA* (IGS)
mono	(G)9	9	32804	32812	*ndhJ*/*trnF-GAA* (IGS)
mono	(A)14	14	32936	32949	*ndhJ*/*trnF-GAA* (IGS)
di	(AT)4	8	33881	33888	*trnL-UAA* (intron)
di	(TA)6	12	33931	33942	*trnL-UAA* (intron)
mono	(T)8	8	34010	34017	*trnL-UAA* (intron)
mono	(T)10	10	34109	34118	*trnL-UAA* (intron)
mono	(T)14	14	34149	34162	*trnL-UAA* (intron)
mono	(T)8	8	34310	34317	*trnL-UAA*/*trnT-UGU* (IGS)
di	(AT)4	8	34567	34574	*trnL-UAA*/*trnT-UGU* (IGS)
mono	(T)12	12	34742	34753	*trnL-UAA*/*trnT-UGU* (IGS)
mono	(A)8	8	35035	35042	*trnL-UAA*/*trnT-UGU* (IGS)
mono	(T)9	9	35076	35084	*trnL-UAA*/*trnT-UGU* (IGS)
mono	(T)15	15	35265	35279	*trnT-UGU*/*rps4* (IGS)
di	(AT)4	8	36498	36505	*rps4*/*trnS-GGA* (IGS)
tetra	(TGAT)3	12	36855	36866	*trnS-GGA*/*ycf3* (IGS)
mono	(A)8	8	37358	37365	*trnS-GGA*/*ycf3* (IGS)
mono	(T)15	15	37672	37686	*ycf3* (intron)
mono	(A)8	8	38734	38741	*ycf3* (intron)
mono	(A)8	8	38869	38876	*ycf3* (intron)
di	(TA)4	8	38965	38972	*ycf3* (intron)

mono	(T)9	9	39869	39877	*ycf3*/*psaA* (IGS)
di	(AT)4	8	40004	40011	*ycf3*/*psaA* (IGS)
mono	(A)8	8	40297	40304	*ycf3*/*psaA* (IGS)
tetra	(TTTC)3	12	44934	44945	psaB/rps14 (IGS)
di	(AT)4	8	45908	45915	*trnG-GCC*/*psbZ* (IGS)
di	(AT)5	10	45917	45926	*trnG-GCC*/*psbZ* (IGS)
tetra	(ATTT)3	12	46008	46019	*trnG-GCC*/*psbZ* (IGS)
di	(CT)4	8	46093	46100	*trnG-GCC*/*psbZ* (IGS)
di	(TA)4	8	46636	46643	*trnS-UGA* (CDS)
di	(CT)4	8	46866	46873	*trnS-UGA* (CDS)
mono	(A)8	8	46974	46981	*trnS-UGA*/*psbC* (IGS)
mono	(A)10	10	47013	47022	*trnS-UGA*/*psbC* (IGS)
mono	(T)10	10	47042	47051	*trnS-UGA*/*psbC* (IGS)
mono	(A)10	10	49700	49709	*psbD*/*trnT-GGU* (IGS)
tetra	(AAGA)3	12	50056	50067	*psbD*/*trnT-GGU* (IGS)
mono	(A)10	10	50493	50502	*psbD*/*trnT-GGU* (IGS)
di	(TA)4	8	51244	51251	*trnT-GGU*/*trnE-UUC* (IGS)
tri	(TAT)4	12	51518	51529	*trnT-GGU*/*trnE-UUC* (IGS)
tri	(ATA)4	12	51548	51559	*trnT-GGU*/*trnE-UUC* (IGS)
di	(AT)4	8	52459	52466	*trnD-GUC*/*psbM* (IGS)
mono	(T)8	8	52822	52829	*trnD-GUC*/*psbM* (IGS)
mono	(T)16	16	53096	53111	*trnD-GUC*/*psbM* (IGS)
mono	(A)8	8	53924	53931	*psbM*/*petN* (IGS)
mono	(T)8	8	54900	54907	*petN*/*trnC-GCA* (IGS)
mono	(T)11	11	55254	55264	*petN*/*trnC-GCA* (IGS)
mono	(T)9	9	55397	55405	*trnC-GCA*/*rpoB* (IGS)
di	(AT)4	8	55497	55504	*trnC-GCA*/*rpoB* (IGS)
mono	(A)9	9	55986	55994	*trnC-GCA*/*rpoB* (IGS)
mono	(T)9	9	56201	56209	*trnC-GCA*/*rpoB* (IGS)
mono	(A)8	8	57208	57215	*rpoB* (CDS)
mono	(T)8	8	60370	60377	*rpoc1* (intron)
mono	(T)8	8	60591	60598	*rpoc1* (intron)
mono	(T)8	8	61028	61035	*rpoc1* (CDS)
mono	(T)9	9	62954	62962	*rpoc2* (CDS)
di	(AT)5	10	63472	63481	*rpoc2* (CDS)
mono	(A)8	8	64305	64312	*rpoc2* (CDS)
mono	(A)9	9	64784	64792	*rpoc2* (CDS)
mono	(A)8	8	64956	64963	*rpoc2* (CDS)
mono	(T)8	8	65606	65613	*rpoc2* (CDS)
mono	(A)9	9	66671	66679	*rpoc2* (CDS)
mono	(T)10	10	69384	69393	*atpI*/*atpH* (IGS)
mono	(T)8	8	70276	70283	*atpF* (intron)
mono	(A)12	12	70686	70697	*atpF* (intron)
mono	(T)9	9	70767	70775	*atpF* (intron)

mono	(T)9	9	71406	71414	*atpF*/*atpA* (IGS)
mono	(A)10	10	73728	73737	*trnG-UCC* (intron)
mono	(A)8	8	73812	73819	*trnG-UCC* (intron)
mono	(T)9	9	74227	74235	*trnG-UCC*/*clpP* (IGS)
tri	(TCT)4	12	74799	74810	*clpP* (CDS)
mono	(A)9	9	74875	74883	*clpP*/*psbB* (IGS)
mono	(T)8	8	74948	74955	*clpP*/*psbB* (IGS)
mono	(A)8	8	75106	75113	*clpP*/*psbB* (IGS)
mono	(T)8	8	76944	76951	*psbB*/*psbT* (IGS)
mono	(T)8	8	77008	77015	*psbB*/*psbT* (IGS)
mono	(T)9	9	78172	78180	*petB* (intron)
mono	(A)8	8	78358	78365	*petB* (intron)
mono	(A)10	10	79516	79525	*petB*/*petD* (IGS)
di	(AG)4	8	79656	79663	*petD* (IGS)
mono	(T)10	10	79800	79809	*petD* (IGS)
di	(AT)9	18	79954	79971	*petD* (IGS)
di	(GA)4	8	81851	81858	*rpoA* (CDS)
mono	(T)9	9	82516	82524	*rps11*/*rpl36* (IGS)
mono	(T)8	8	83634	83641	*rps8*/*rpl14* (IGS)
mono	(A)8	8	83649	83656	*rps8*/*rpl14* (IGS)
mono	(A)9	9	83658	83666	*rps8*/*rpl14* (IGS)
mono	(T)8	8	83773	83780	*rpl14* (CDS)
tetra	(CAAA)3	12	84130	84141	*rpl14*/*rpl16* (IGS)
di	(TA)5	10	84980	84989	*rpl16* (intron)
tetra	(TTTC)3	12	85485	85496	*rpl16* (intron)
mono	(T)12	12	85542	85553	*rpl16* (intron)
mono	(A)9	9	85590	85598	*rpl16* (intron)
mono	(T)8	8	85729	85736	*rpl16* (intron)
mono	(T)10	10	87491	87500	*rps19*/*rpl2* (IGS)
mono	(T)8	8	88402	88409	*rpl2*/*rpl23* (IGS)
tetra	(AATA)3	12	88937	88948	*trnI-CAU*/*ycf2* (IGS)
mono	(A)9	9	88961	88969	*trnI-CAU*/*ycf2* (IGS)
di	(GA)4	8	89067	89074	*ycf2* (CDS)
di	(GA)4	8	90021	90028	*ycf2* (CDS)
mono	(A)9	9	91945	91953	*ycf2* (CDS)
mono	(T)8	8	93819	93826	*ycf2* (CDS)
mono	(T)9	9	96154	96162	*trnL-CAA*/*ndhB* (IGS)
di	(AG)4	8	96786	96793	*ndhB* (CDS)
mono	(T)9	9	99207	99215	*rps7* (intron)
mono	(A)8	8	99674	99681	*rps12* (intron)
mono	(T)9	9	101399	101407	*rps12* (CDS)
mono	(T)8	8	104290	104297	*trnI-GAU* (intron)
mono	(T)8	8	105621	105628	*trnA-UGC* (intron)
tetra	(AGGT)3	12	107633	107644	*rrn23s* (CDS)

di	(CT)4	8	107671	107678	*rrn23s* (CDS)
di	(GA)4	8	107995	108002	*rrn23s* (CDS)
mono	(A)9	9	109423	109431	*rrn5s*/*trnR-ACG* (IGS)
mono	(T)8	8	109823	109830	trnR-ACG/trnN-GUU (IGS)
mono	(A)13	13	110691	110703	*trnN-GUU*/*ndhF* (IGS)
mono	(A)8	8	111439	111446	*ndhF*(CDS)
mono	(T)8	8	111562	111569	*ndhF*(CDS)
mono	(A)8	8	113057	113064	*ndhF*/*rpl32* (IGS)
mono	(A)8	8	113283	113290	*ndhF*/*rpl32* (IGS)
di	(AT)4	8	113323	113330	*ndhF*/*rpl32* (IGS)
mono	(T)11	11	113554	113564	*ndhF*/*rpl32* (IGS)
penta	(TTACT)4	20	113614	113633	*ndhF*/*rpl32* (IGS)
mono	(A)10	10	113654	113663	*ndhF*/*rpl32* (IGS)
mono	(A)9	9	113774	113782	*ndhF*/*rpl32* (IGS)
di	(AT)10	20	114044	114063	*ndhF*/*rpl32* (IGS)
tetra	(TAAT)3	12	114154	114165	*ndhF*/*rpl32* (IGS)
mono	(A)8	8	114302	114309	*rpl32* (CDS)
mono	(A)10	10	114384	114393	*rpl32* (CDS)
mono	(A)8	8	114395	114402	*rpl32* (CDS)
mono	(T)8	8	115029	115036	*ycf1* (CDS)
mono	(A)8	8	115568	115575	*ycf1* (CDS)
mono	(A)9	9	115789	115797	*ycf1* (CDS)
di	(TA)4	8	116492	116499	*ycf1* (CDS)
mono	(A)9	9	116711	116719	*ycf1* (CDS)
mono	(A)9	9	116724	116732	*ycf1* (CDS)
mono	(A)12	12	117690	117701	*ycf1*/*rps15* (IGS)
tetra	(CCAT)3	12	118157	118168	*ndhH* (CDS)
tetra	(CTTT)3	12	120424	120435	*ndhA* (intron)
tetra	(TCTT)3	12	119922	119933	*ndhA* (intron)
mono	(A)8	8	120492	120499	*ndhA* (intron)
mono	(T)9	9	120572	120580	*ndhA* (intron)
mono	(A)9	9	120589	120597	*ndhA* (intron)
mono	(T)8	8	120632	120639	*ndhA* (intron)
penta	(TTAAA)3	15	123176	123190	*ndhG*/*ndhE* (IGS)
mono	(T)9	9	123593	123601	*ndhE*/*psaC* (IGS)
mono	(T)8	8	123751	123758	*ndhE*/*psaC* (IGS)
mono	(A)10	10	123893	123902	*ndhE*/*psaC* (IGS)
mono	(T)9	9	124200	124208	*psaC*/*ndhD* (IGS)
mono	(T)8	8	124250	124257	*psaC*/*ndhD* (IGS)
mono	(A)8	8	125185	125192	*ndhD* (CDS)
mono	(T)8	8	125535	125542	*ndhD* (CDS)
mono	(T)8	8	125770	125777	*ndhD* (CDS)
mono	(A)8	8	125813	125820	*ndhD*/*ccsA* (IGS)
mono	(A)9	9	125823	125831	*ndhD*/*ccsA* (IGS)

mono	(T)8	8	126695	126702	*ccsA* (CDS)
mono	(A)15	15	126985	126999	*ccsA*/*trnL-UAG* (IGS)
mono	(T)11	11	127251	127261	*trnL-UAG*/*ndhG* (IGS)

CDS, coding sequences; IGS, intergenic spacers

Supplementary Table S4. List of simple sequence repeats (SSRs) identified in the plastome of *O. ficus-indica*

SSR sequence	Number of repetitions														Total
	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
A/T	-	-	-	-	-	82	40	21	10	9	4	2	1	3	172
C/G	-	-	-	-	-	1	2	-	-	2	-	-	-	-	5
AG/CT	-	8	-	-	-	-	-	-	-	-	-	-	-	-	8
AT/AT	-	26	7	1	-	-	1	-	-	1	-	-	-	-	36
AAT/ATT	-	4	-	-	-	-	-	-	-	-	-	-	-	-	4
AAAC/GTTT	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1
AAAG/CTTT	-	5	-	-	-	-	-	-	-	-	-	-	-	-	5
AAAT/ATTT	-	5	-	-	-	-	-	-	-	-	-	-	-	-	5
AATC/ATTG	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1
AATT/AATT	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1
ACCT/AGGT	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1
ATCC/ATGG	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1
AAAAC/GTTTT	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1
AAAGT/ACTTT	-	2	-	-	-	-	-	-	-	-	-	-	-	-	2
AAGTG/ACTTC	-	1	2	-	-	-	1	-	-	-	-	1	-	-	5
AATAT/ATATT	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1
ACTAG/AGTCT	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1
AAAATT/AATTTT	-	-	-	1	-	-	-	-	-	-	-	-	-	-	1
AATAGT/ACTATT	-	-	1	-	-	-	-	-	-	-	-	-	-	-	1
AATATT/AATATT	-	-	1	-	-	-	-	-	-	-	-	-	-	-	1
															253

Supplementary Table S5. Distribution of SSR loci in the plastome of *O. ficus-indica*

SSR type	SSR	Size	Start	End	Location
tetra	(ATAA)3	12	207	218	*trnH-GUG*/*psbA* (IGS)
mono	(A)13	13	326	338	*trnH-GUG*/*psbA* (IGS)
mono	(T)9	9	1974	1982	*trnK-UUU* (intron)
mono	(A)8	8	2559	2566	*trnK-UUU* (intron)/matk (CDS)
mono	(T)9	9	3210	3218	*trnK-UUU* (intron)/matk (CDS)
mono	(A)9	9	3594	3602	*trnK-UUU (intron)*/*rps16* (IGS)
mono	(T)8	8	3948	3955	*trnK-UUU (intron)*/*rps16* (IGS)
mono	(T)9	9	4519	4527	*trnK-UUU* /*rps16* (IGS)

mono	(A)8	8	4577	4584	*trnK-UUU* /*rps16* (IGS)
tri	(TAA)4	12	4735	4746	*trnK-UUU* /*rps16* (IGS)
mono	(A)8	8	4820	4827	*trnK-UUU* /*rps16* (IGS)
mono	(T)11	11	4926	4936	*trnK-UUU* /*rps16* (IGS)
mono	(A)10	10	5396	5405	*rps16* (intron)
di	(AT)4	8	5698	5705	*rps16* (intron)
di	(AT)4	8	6178	6185	*rps16*/*trnQ-UUG* (IGS)
mono	(T)12	12	6200	6211	*rps16*/*trnQ-UUG* (IGS)
mono	(A)8	8	6226	6233	*rps16*/*trnQ-UUG* (IGS)
mono	(A)11	11	6938	6948	*trnQ-UUG*/*psbK* (IGS)
mono	(T)12	12	6968	6979	*trnQ-UUG*/*psbK* (IGS)
mono	(A)8	8	7445	7452	*psbK*/*psbI* (IGS)
compound	(A)10-(AAT)3	21	7544	7564	*psbK*/*psbI* (IGS)
mono	(A)9	9	7747	7755	psbI/trnS-GCU (IGS)
di	(TA)4	8	7786	7793	psbI/trnS-GCU (IGS)
di	(AT)4	8	8053	8060	*trnS-GCU*/*petL* (IGS)
di	(AT)5	10	8208	8217	*trnS-GCU*/*petL* (IGS)
mono	(T)9	9	8285	8293	*trnS-GCU*/*petL* (IGS)
mono	(A)14	14	8478	8491	*trnS-GCU*/*petL* (IGS)
mono	(T)12	12	8774	8785	*petL*/*petG* (IGS)
mono	(A)8	8	9221	9228	*trnW-CCA*/*trnP-UGG* (IGS)
di	(AT)4	8	9518	9525	*trnP-UGG*/*psaJ* (IGS)
mono	(A)8	8	9767	9774	*trnP-UGG*/*psaJ* (IGS)
di	(AT)4	8	10709	10716	*rpl33*/*rps18* (IGS)
di	(AT)5	10	10736	10745	*rpl33*/*rps18* (IGS)
mono	(T)9	9	10821	10829	*rpl33*/*rps18* (IGS)
mono	(T)11	11	11528	11538	*rps18*/*rpl20* (IGS)
mono	(T)8	8	11554	11561	*rps18*/*rpl20* (IGS)
mono	(T)8	8	11974	11981	*rpl20* (CDS)
mono	(T)8	8	12188	12195	*rpl20*/*rps12* (IGS)
di	(AT)4	8	12944	12951	*rps12*/*psbE* (IGS)
mono	(A)10	10	13164	13173	*rps12*/*psbE* (IGS)
di	(TA)5	10	13328	13337	*rps12*/*psbE* (IGS)
p6	(TTTAAT)6	36	13367	13402	*rps12*/*psbE* (IGS)
mono	(A)8	8	13481	13488	*rps12*/*psbE* (IGS)
mono	(T)11	11	15137	15147	*psbJ*/*petA* (IGS)
mono	(A)13	13	15264	15276	*psbJ*/*petA* (IGS)
penta	(TAGTC)3	15	15476	15490	*psbJ*/*petA* (IGS)
mono	(A)9	9	15604	15612	*psbJ*/*petA* (IGS)
mono	(A)8	8	15615	15622	*psbJ*/*petA* (IGS)
mono	(T)8	8	16356	16363	*petA* (CDS)
di	(TA)4	8	16715	16722	*petA* (CDS)
mono	(T)8	8	17747	17754	*cemA* (CDS)/*ycf4* (IGS)

mono	(T)8	8	17874	17881	*cemA*/*ycf4* (IGS)
penta	(CAAAA)3	15	18210	18224	*cemA*/*ycf4* (IGS)
mono	(T)8	8	18454	18461	*cemA*/*ycf4* (IGS)
mono	(A)10	10	18549	18558	*cemA*/*ycf4* (IGS)
mono	(A)8	8	19096	19103	*ycf4* (CDS)
mono	(G)8	8	19113	19120	*ycf4* (CDS)
mono	(T)9	9	19733	19741	*psaI*/*accD* (IGS)
di	(TA)4	8	19815	19822	*psaI*/*accD* (IGS)
mono	(A)11	11	19935	19945	*psaI*/*accD* (IGS)
mono	(T)8	8	20058	20065	*psaI*/*accD* (IGS)
mono	(A)8	8	20088	20095	*psaI*/*accD* (IGS)
mono	(T)8	8	20111	20118	*psaI*/*accD* (IGS)
di	(AT)4	8	20167	20174	*psaI*/*accD* (IGS)
di	(AT)4	8	20595	20602	*accD* (CDS)
mono	(T)8	8	23801	23808	*accD*/*trnV-UAC* (IGS)
mono	(T)10	10	24400	24409	*trnV-UAC* (intron)
mono	(T)12	12	24858	24869	*trnV-UAC* (intron)
mono	(T)9	9	25002	25010	*trnV-UAC*/*trnM-CAU* (IGS)
di	(TA)4	8	25298	25305	*trnM-CAU*/ *atpE* (IGS)
mono	(T)10	10	25361	25370	*trnM-CAU*/ *atpE* (IGS)
di	(TA)4	8	25450	25457	*trnM-CAU*/ *atpE* (IGS)
mono	(T)10	10	27359	27368	*atpB* (CDS)
penta	(TAAAG)3	15	27443	27457	*atpB*/*rbcL* (IGS)
mono	(T)8	8	27579	27586	*atpB*/*rbcL* (IGS)
mono	(A)10	10	27690	27699	*atpB*/*rbcL* (IGS)
di	(AT)4	8	28055	28062	*rbcL* (CDS)
di	(GA)4	8	28734	28741	*rbcL* (CDS)
mono	(T)8	8	30198	30205	*rbcL*/ *ndhC* (IGS)
mono	(T)8	8	31589	31596	*ndhK*/*ndhJ* (IGS)
mono	(A)15	15	31640	31654	*ndhK*/*ndhJ* (IGS)
tetra	(AAAT)3	12	31692	31703	*ndhK*/*ndhJ* (IGS)
mono	(G)12	12	32496	32507	*ndhJ*/*trnF-GAA* (IGS)
mono	(C)9	9	32555	32563	*ndhJ*/*trnF-GAA* (IGS)
mono	(C)12	12	32628	32639	*ndhJ*/*trnF-GAA* (IGS)
mono	(A)13	13	32781	32793	*ndhJ*/*trnF-GAA* (IGS)
di	(AT)4	8	33725	33732	*trnL-UAA* (intron)
di	(TA)6	12	33775	33786	*trnL-UAA* (intron)
mono	(T)10	10	33854	33863	*trnL-UAA* (intron)
mono	(T)12	12	33955	33966	*trnL-UAA* (intron)
mono	(T)11	11	33997	34007	*trnL-UAA* (intron)
mono	(T)12	12	34155	34166	*trnL-UAA*/ *trnT-UGU* (IGS)
di	(AT)4	8	34414	34421	*trnL-UAA*/ *trnT-UGU* (IGS)
mono	(T)8	8	34589	34596	*trnL-UAA*/ *trnT-UGU* (IGS)
mono	(A)8	8	34878	34885	*trnL-UAA*/ *trnT-UGU* (IGS)

mono	(T)9	9	34919	34927	*trnL-UAA*/ *trnT-UGU* (IGS)
mono	(T)16	16	35108	35123	*trnT-UGU*/*rps4* (IGS)
di	(AT)4	8	36337	36344	*rps4*/ *trnS-GGA* (IGS)
tetra	(TGAT)3	12	36694	36705	*trnS-GGA*/*ycf3* (IGS)
mono	(A)8	8	37197	37204	*trnS-GGA*/*ycf3* (IGS)
mono	(T)13	13	37511	37523	*ycf3* (intron)
mono	(A)8	8	38571	38578	*ycf3* (intron)
mono	(A)8	8	38705	38712	*ycf3* (intron)
di	(TA)4	8	38801	38808	*ycf3* (intron)
mono	(T)9	9	39686	39694	*ycf3*/*psaA* (IGS)
di	(AT)4	8	39821	39828	*ycf3*/*psaA* (IGS)
mono	(A)8	8	40114	40121	*ycf3*/*psaA* (IGS)
tetra	(TTTC)3	12	44751	44762	*psaB*/*rps14* (IGS)
di	(AT)5	10	45719	45728	*trnG-GCC*/*psbZ* (IGS)
di	(AT)5	10	45730	45739	*trnG-GCC*/*psbZ* (IGS)
tetra	(TATT)3	12	45848	45859	*trnG-GCC*/*psbZ* (IGS)
di	(CT)4	8	45906	45913	*trnG-GCC*/*psbZ* (IGS)
di	(TA)4	8	46449	46456	*psbZ*/*trnS-UGA* (IGS)
di	(CT)4	8	46679	46686	*trns-UGA* (CDS)
mono	(A)9	9	46787	46795	*trnS-UGA*/*psbC* (IGS)
mono	(A)9	9	46827	46835	*trnS-UGA*/*psbC* (IGS)
mono	(T)10	10	46855	46864	*trnS-UGA*/*psbC* (IGS)
mono	(A)9	9	49513	49521	*psbD*/*trnT-GGU* (IGS)
tetra	(AAGA)3	12	49868	49879	*psbD*/*trnT-GGU* (IGS)
mono	(A)14	14	50305	50318	*psbD*/*trnT-GGU* (IGS)
compound	(ATATTA)2 (ATAT<TA >)(TA)3	24	51095	51118	*trnT-GGU*/*trnE-UUC* (IGS)
tri	(TAT)4	12	51362	51373	*trnT-GGU*/*trnE-UUC* (IGS)
tri	(ATA)4	12	51392	51403	*trnT-GGU*/*trnE-UUC* (IGS)
di	(AT)4	8	52303	52310	*trnD-GUC*/*psbM* (IGS)
mono	(T)8	8	52658	52665	*trnD-GUC*/*psbM* (IGS)
mono	(T)16	16	52932	52947	*trnD-GUC*/*psbM* (IGS)
mono	(T)9	9	53452	53460	*psbM*/*petN* (IGS)
mono	(T)8	8	54731	54738	*petN*/*trnC-GCA* (IGS)
mono	(T)12	12	55085	55096	*petN*/*trnC-GCA* (IGS)
mono	(T)9	9	55229	55237	*trnC-GCA*/*rpoB* (IGS)
di	(AT)4	8	55329	55336	*trnC-GCA*/*rpoB* (IGS)
tetra	(ATAA)3	12	55673	55684	*trnC-GCA*/*rpoB* (IGS)
mono	(A)9	9	55821	55829	*trnC-GCA*/*rpoB* (IGS)
mono	(T)9	9	56036	56044	*trnC-GCA*/*rpoB* (IGS)
mono	(A)8	8	57043	57050	*rpoB* (CDS)
mono	(T)8	8	60205	60212	*rpoc1* (intron)
mono	(T)8	8	60426	60433	*rpoc1* (intron)

mono	(T)8	8	60863	60870	*rpoc1*(CDS)
mono	(T)9	9	62793	62801	*rpoc2*(CDS)
di	(AT)5	10	63311	63320	*rpoc2*(CDS)
mono	(A)8	8	64144	64151	*rpoc2*(CDS)
mono	(A)9	9	64623	64631	*rpoc2*(CDS)
mono	(A)8	8	64795	64802	*rpoc2*(CDS)
mono	(T)8	8	65445	65452	*rpoc2*(CDS)
mono	(A)9	9	66510	66518	*rpoc2*(CDS)
mono	(T)10	10	69223	69232	atpI/atpH (IGS)
mono	(T)8	8	70115	70122	*atpF* (intron)
mono	(A)16	16	70525	70540	*atpF* (intron)
mono	(T)10	10	70610	70619	*atpF* (intron)
mono	(T)9	9	71250	71258	*atpA* (CDS)
mono	(A)10	10	73572	73581	*trnG-UCC* (intron)
mono	(A)10	10	73656	73665	*trnG-UCC* (intron)
mono	(T)9	9	74073	74081	*trnG-UCC*/*clpP* (IGS)
mono	(T)8	8	74798	74805	*clpP*/*psbB* (IGS)
mono	(A)8	8	74956	74963	*clpP*/*psbB* (IGS)
mono	(T)8	8	76794	76801	*psbB*/*psbT*(IGS)
mono	(T)8	8	76858	76865	*psbB*/*psbT*(IGS)
mono	(A)8	8	77590	77597	*petB* (intron)
mono	(T)8	8	78022	78029	*petB* (intron)
mono	(A)8	8	78207	78214	*petB* (intron)
mono	(A)10	10	79365	79374	*petB*/*petD* (IGS)
di	(AG)4	8	79505	79512	*petD* (intron)
mono	(T)9	9	79649	79657	*petD* (intron)
di	(AT)9	18	79802	79819	*petD* (intron)
mono	(T)9	9	80833	80841	*rpoA* (CDS)
di	(GA)4	8	81703	81710	*rpoA* (CDS)
mono	(T)9	9	82368	82376	*rps11*/*rpl36* (intron)
mono	(T)8	8	83495	83502	*rps8*/*rpl14* (intron)
mono	(A)8	8	83510	83517	*rps8*/*rpl14* (intron)
mono	(T)8	8	83637	83644	rpl14 (CDS)
tetra	(CAAA)3	12	83994	84005	*rpl14*/*rpl16* (IGS)
di	(TA)5	10	84835	84844	*rpl16* (intron)
tetra	(TTTC)3	12	85340	85351	*rpl16* (intron)
mono	(T)8	8	85397	85404	*rpl16* (intron)
mono	(A)10	10	85441	85450	*rpl16* (intron)
mono	(T)8	8	85581	85588	*rpl16* (intron)
mono	(T)10	10	87343	87352	*rps19*/*rpl2* (IGS)
mono	(A)10	10	88789	88798	*trnI-CAU*/*trnL-CAA* (IGS)
mono	(A)8	8	88813	88820	*trnI-CAU*/*trnL-CAA* (IGS)
mono	(A)8	8	88835	88842	*trnI-CAU*/*trnL-CAA* (IGS)
penta	(AAGTG)8	40	88849	88888	*trnI-CAU*/*trnL-CAA* (IGS)

mono	(A)8	8	88892	88899	*trnI-CAU*/*trnL-CAA* (IGS)
penta	(AAGTG)13	65	88906	88970	*trnI-CAU*/*trnL-CAA* (IGS)
mono	(A)8	8	88974	88981	*trnI-CAU*/*trnL-CAA* (IGS)
penta	(AAGTG)4	20	89000	89019	*trnI-CAU*/*trnL-CAA* (IGS)
penta	(AAGTG)4	20	89049	89068	*trnI-CAU*/*trnL-CAA* (IGS)
penta	(AAGTG)3	15	89098	89112	*trnI-CAU*/*trnL-CAA* (IGS)
mono	(T)10	10	90044	90053	*trnL-CAA*/*ndhB* (IGS)
di	(AG)4	8	90677	90684	*ndhB* (CDS)
mono	(T)9	9	93098	93106	*rps7* (CDS)
mono	(A)8	8	93565	93572	*rps12* (CDS)
mono	(T)9	9	95290	95298	*rps12*/*trnV-GAC* (IGS)
mono	(G)9	9	95765	95773	*trnV-GAC*/*rrn16s* (IGS)
mono	(T)8	8	98181	98188	*trnI-GAU* (intron)
mono	(T)9	9	99512	99520	*trnA-UGC* (intron)
tetra	(AGGT)3	12	101526	101537	*rrn23s* (CDS)
di	(CT)4	8	101564	101571	*rrn23s* (CDS)
di	(GA)4	8	101888	101895	*rrn23s* (CDS)
mono	(A)9	9	103316	103324	*rrn5s*/*trnR-ACG* (IGS)
compound	(T)7(<T>ATTAC)(TATTAC)2	25	103716	103740	trnR-ACG/trnN-GUU (IGS)
penta	(TATAA)3	15	104556	104570	*trnN-GUU*/*ndhF* (IGS)
mono	(A)9	9	104607	104615	*trnN-GUU*/*ndhF* (IGS)
mono	(A)8	8	105351	105358	*ndhF* (CDS)
mono	(T)8	8	105474	105481	*ndhF* (CDS)
mono	(A)8	8	106969	106976	*ndhF*/*rpl32* (IGS)
mono	(A)8	8	107189	107196	*ndhF*/*rpl32* (IGS)
di	(AT)4	8	107229	107236	*ndhF*/*rpl32* (IGS)
mono	(A)8	8	107438	107445	*ndhF*/*rpl32* (IGS)
mono	(T)11	11	107461	107471	*ndhF*/*rpl32* (IGS)
mono	(A)8	8	107500	107507	*ndhF*/*rpl32* (IGS)
penta	(TTACT)3	15	107522	107536	*ndhF*/*rpl32* (IGS)
mono	(A)9	9	107557	107565	*ndhF*/*rpl32* (IGS)
mono	(A)11	11	107676	107686	*ndhF*/*rpl32* (IGS)
di	(AT)12	24	107948	107971	*ndhF*/*rpl32* (IGS)
mono	(A)8	8	108206	108213	*rpl32* (CDS)
mono	(A)12	12	108288	108299	*rpl32* (CDS)
mono	(A)8	8	108301	108308	*rpl32* (CDS)
mono	(A)8	8	108455	108462	*rpl32*/*ycf1* (IGS)
mono	(T)8	8	108926	108933	*ycf1* (CDS)
mono	(A)8	8	109438	109445	*ycf1* (CDS)
mono	(A)9	9	109659	109667	*ycf1* (CDS)
mono	(A)9	9	109682	109690	*ycf1* (CDS)
di	(TA)4	8	110676	110683	*ycf1* (CDS)

mono	(A)8	8	110874	110881	*ycf1* (CDS)
di	(TA)4	8	111081	111088	*ycf1* (CDS)
mono	(A)9	9	111300	111308	*ycf1* (CDS)
mono	(A)10	10	111313	111322	*ycf1* (CDS)
mono	(A)10	10	112280	112289	*ycf1*/*rps15* (IGS)
tetra	(CCAT)3	12	112745	112756	*ndhH* (CDS)
tetra	(TCTT)3	12	114510	114521	*ndhA* (intron)
tetra	(CTTT)3	12	115012	115023	*ndhA* (intron)
mono	(A)8	8	115080	115087	*ndhA* (intron)
mono	(T)9	9	115160	115168	*ndhA* (intron)
mono	(A)9	9	115177	115185	*ndhA* (intron)
mono	(T)8	8	115220	115227	*ndhA* (intron)
mono	(T)11	11	118176	118186	*ndhE*/*psaC* (IGS)
tetra	(AATA)3	12	118310	118321	*ndhE*/*psaC* (IGS)
mono	(T)9	9	118340	118348	*ndhE*/*psaC* (IGS)
mono	(A)10	10	118483	118492	*ndhE*/*psaC* (IGS)
mono	(T)8	8	118790	118797	*psaC*/*ndhD* (IGS)
mono	(T)8	8	118839	118846	*psaC*/*ndhD* (IGS)
mono	(A)8	8	119774	119781	*ndhD* (CDS)
mono	(T)8	8	120124	120131	*ndhD* (CDS)
mono	(T)8	8	120359	120366	*ndhD* (CDS)
mono	(A)8	8	120402	120409	*ndhD*/*ccsA* (IGS)
mono	(A)9	9	120412	120420	*ndhD*/*ccsA* (IGS)
tetra	(TTAA)3	12	120593	120604	*ccsA* (CDS)
mono	(T)8	8	121288	121295	*ccsA* (CDS)
mono	(A)12	12	121578	121589	*ccsA*/*trnL-UAG* (IGS)
mono	(T)8	8	121841	121848	*trnL-UAG*/*ndhG* (IGS)

CDS, coding sequences; IGS, intergenic spacers

Gene	aa pos	*Opuntia ficus-indica*	*aa pos*	*O. monacantha*	aa pos	*O. quimilo*	aa pos	*Brasiliopuntia brasiliensis*
accD	585	CCA (P) => TCA (S)	*566*	CCG (P) => TCG (S)	584	CCG (P) => TCG (S)	541	CCG (P) => TCG (S)
	677	CCT (P) => TCT (S)	*649*	GCG (A) => GTG (V)	727	CCC (P) => TCC (S)	580	TCT (S) => TTT (F)
	862	TCG (S) => TTG (L)	*658*	CCG (P) => TCG (S)	761	CCT (P) => TCT (S)	590	CAT (H) => TAT (Y)
	931	CCC (P) => TCC (S)	*676*	CTT (L) => TTT (F)	1018	TCC (S) => TTC (F)	629	CCA (P) => CTA (L)
	1123	GCT (A) => GTT (V)	*678*	CCT (P) => TCT (S)	1084	GCA (A) => GTA (V)	703	CCT (P) => TCT (S)
	1128	CTT (L) => TTT (F)	*1071*	GCA (A) => GTA (V)	1220	GCT (A) => GTT (V)	1146	GCT (A) => GTT (V)
			1212	CTT (L) => TTT (F)	1225	CTT (L) => TTT (F)	1151	CTT (L) => TTT (F)
atpA	305	TCA (S) => TTA (L)	*305*	TCA (S) => TTA (L)	914	TCA (S) => TTA (L)	305	TCA (S) => TTA (L)
	424	CCC (P) => TCC (S)	*424*	CCC (P) => TCC (S)	1270	CCC (P) => TCC (S)	424	CCC (P) => TCC (S)
atpB	140	ACC (T) => ATC (I)	*140*	ACC (T) => ATC (I)	140	ACC (T) => ATC (I)	140	ACC (T) => ATC (I)
ccsA	188	TCT (S) => TTT (F)	*188*	TCT (S) => TTT (F)	188	TCT (S) => TTT (F)	188	TCT (S) => TTT (F)
clpP	143	CTT (L) => TTT (F)	*143*	CTT (L) => TTT (F)	143	CTT (L) => TTT (F)	142	CTT (L) => TTT (F)
matK	217	CAT (H) => TAT (Y)	*217*	CAT (H) => TAT (Y)	217	CAT (H) => TAT (Y)	54	CCG (P) => CTG (L)
							217	CAT (H) => TAT (Y)
ndhA	42	ACA (T) => ATA (I)	*42*	ACA (T) => ATA (I)	42	ACA (T) => ATA (I)	42	ACA (T) => ATA (I)
	114	TCA (S) => TTA (L)	*114*	TCA (S) => TTA (L)	114	TCA (S) => TTA (L)	114	TCA (S) => TTA (L)
	182	ACA (T) => ATA (I)	*182*	ACA (T) => ATA (I)	182	ACA (T) => ATA (I)	182	ACA (T) => ATA (I)
	189	TCA (S) => TTA (L)	*189*	TCA (S) => TTA (L)	189	TCA (S) => TTA (L)	189	TCA (S) => TTA (L)
	358	TCC (S) => TTC (F)	*358*	TCC (S) => TTC (F)	358	TCC (S) => TTC (F)	358	TCC (S) => TTC (F)
ndhB	50	TCA (S) => TTA (L)	*50*	TCA (S) => TTA (L)	50	TCA (S) => TTA (L)	50	TCA (S) => TTA (L)
	156	CCA (P) => CTA (L)	*156*	CCA (P) => CTA (L)	156	CCA (P) => CTA (L)	156	CCA (P) => CTA (L)
	181	ACG (T) => ATG (M)	*181*	ACG (T) => ATG (M)	181	ACG (T) => ATG (M)	181	ACG (T) => ATG (M)
	196	CAT (H) => TAT (Y)	*196*	CAT (H) => TAT (Y)	196	CAT (H) => TAT (Y)	196	CAT (H) => TAT (Y)
	204	TCA (S) => TTA (L)	*204*	TCA (S) => TTA (L)	204	TCA (S) => TTA (L)	204	TCA (S) => TTA (L)
	246	CCA (P) => CTA (L)	*246*	CCA (P) => CTA (L)	246	CCA (P) => CTA (L)	246	CCA (P) => CTA (L)
	249	TCT (S) => TTT (F)	*249*	TCT (S) => TTT (F)	249	TCT (S) => TTT (F)	249	TCT (S) => TTT (F)
	277	TCA (S) => TTA (L)	*277*	TCA (S) => TTA (L)	277	TCA (S) => TTA (L)	277	TCA (S) => TTA (L)

	279	TCA (S) => TTA (L)	*279*	TCA (S) => TTA (L)	279	TCA (S) => TTA (L)	279	TCA (S) => TTA (L)
	371	TCA (S) => TTA (L)	*371*	TCA (S) => TTA (L)	371	TCA (S) => TTA (L)	371	TCA (S) => TTA (L)
	494	CCA (P) => CTA (L)	*494*	CCA (P) => CTA (L)	494	CCA (P) => CTA (L)	494	CCA (P) => CTA (L)
ndhD	9	ACA (T) => ATA (I)	*9*	ACA (T) => ATA (I)	9	ACA (T) => ATA (I)	9	ACA (T) => ATA (I)
	128	TCG (S) => TTG (L)	*128*	TCG (S) => TTG (L)	128	TCG (S) => TTG (L)	128	TCG (S) => TTG (L)
	182	GCT (A) => GTT (V)	*182*	GCT (A) => GTT (V)	182	GCT (A) => GTT (V)	182	GCT (A) => GTT (V)
	293	TCA (S) => TTA (L)	*293*	TCA (S) => TTA (L)	293	TCA (S) => TTA (L)	293	TCA (S) => TTA (L)
	433	TCA (S) => TTA (L)	*433*	TCA (S) => TTA (L)	433	TCA (S) => TTA (L)	433	TCA (S) => TTA (L)
	437	TCA (S) => TTA (L)	*437*	TCA (S) => TTA (L)	437	TCA (S) => TTA (L)	437	TCA (S) => TTA (L)
	466	ACT (T) => ATT (I)	*466*	ACT (T) => ATT (I)	466	ACT (T) => ATT (I)	466	ACT (T) => ATT (I)
	482	GCT (A) => GTT (V)	*482*	GCT (A) => GTT (V)	482	GCT (A) => GTT (V)	482	GCT (A) => GTT (V)
	490	GCG (A) => GTG (V)	*490*	GCG (A) => GTG (V)	490	GCG (A) => GTG (V)	490	GCG (A) => GTG (V)
ndhF	97	TCA (S) => TTA (L)	*97*	TCA (S) => TTA (L)	97	TCA (S) => TTA (L)	97	TCA (S) => TTA (L)
	291	ACG (T) => ATG (M)	*291*	ACG (T) => ATG (M)	291	ACG (T) => ATG (M)	291	ACG (T) => ATG (M)
ndhG	56	CAT (H) => TAT (Y)	*56*	CAT (H) => TAT (Y)	56	CAT (H) => TAT (Y)	56	CAT (H) => TAT (Y)
	84	CCA (P) => TCA (S)	*84*	CCA (P) => TCA (S)	84	CCA (P) => TCA (S)	84	CCA (P) => TCA (S)
	105	ACA (T) => ATA (I)	*105*	ACA (T) => ATA (I)	105	ACA (T) => ATA (I)	105	ACA (T) => ATA (I)
petB	140	CGG (R) => TGG (W)	*140*	CGG (R) => TGG (W)	140	CGG (R) => TGG (W)	140	CGG (R) => TGG (W)
psbB	193	CTT (L) => TTT (F)						
psbF	26	TCT (S) => TTT (F)	*26*	TCT (S) => TTT (F)	26	TCT (S) => TTT (F)	26	TCT (S) => TTT (F)
psbL	1	ACG (T) => ATG (M)	*1*	ACG (T) => ATG (M)	1	ACG (T) => ATG (M)	1	ACG (T) => ATG (M)
rpl20	89	TCG (S) => TTG (L)	*89*	TCG (S) => TTG (L)	149	TCG (S) => TTG (L)		
rpoB	158	TCA (S) => TTA (L)	*158*	TCA (S) => TTA (L)	164	TCA (S) => TTA (L)	158	TCA (S) => TTA (L)
	184	TCA (S) => TTA (L)	*184*	TCA (S) => TTA (L)	190	TCA (S) => TTA (L)	184	TCA (S) => TTA (L)
	189	TCG (S) => TTG (L)	*189*	TCG (S) => TTG (L)	195	TCG (S) => TTG (L)	189	TCG (S) => TTG (L)
	478	ACG (T) => ATG (M)	*478*	ACG (T) => ATG (M)	484	ACG (T) => ATG (M)	478	ACG (T) => ATG (M)
	505	TCC (S) => TTC (F)	*505*	TCC (S) => TTC (F)	511	TCC (S) => TTC (F)	505	TCC (S) => TTC (F)
	767	GCA (A) => GTA (V)	*767*	GCA (A) => GTA (V)	773	GCA (A) => GTA (V)	767	GCA (A) => GTA (V)

	809	TCA (S) => TTA (L)	*809*	TCA (S) => TTA (L)	815	TCA (S) => TTA (L)	809	TCA (S) => TTA (L)
rpoc1	14	TCA (S) => TTA (L)	*14*	TCA (S) => TTA (L)	14	TCA (S) => TTA (L)	14	TCA (S) => TTA (L)
	210	ACT (T) => ATT (I)	*210*	ACT (T) => ATT (I)	210	ACT (T) => ATT (I)	210	ACT (T) => ATT (I)
rpoc2	379	CTT (L) => TTT (F)	*379*	CTT (L) => TTT (F)	383	CTT (L) => TTT (F)	379	CTT (L) => TTT (F)
	498	ACG (T) => ATG (M)	*498*	ACG (T) => ATG (M)	502	ACG (T) => ATG (M)	498	ACG (T) => ATG (M)
	643	CTT (L) => TTT (F)	*643*	CTT (L) => TTT (F)	610	TCC (S) => TTC (F)	643	CTT (L) => TTT (F)
	741	CGG (R) => TGG (W)	*741*	CGG (R) => TGG (W)	647	CTT (L) => TTT (F)	741	CGG (R) => TGG (W)
	743	ACC (T) => ATC (I)	*743*	ACC (T) => ATC (I)	745	CGG (R) => TGG (W)	743	ACC (T) => ATC (I)
	905	TCG (S) => TTG (L)	*905*	TCG (S) => TTG (L)	747	ACC (T) => ATC (I)	905	TCG (S) => TTG (L)
	913	CCT (P) => TCT (S)	*913*	CCT (P) => TCT (S)	909	TCG (S) => TTG (L)	913	CCT (P) => TCT (S)
	1050	GCA (A) => GTA (V)	*1050*	GCA (A) => GTA (V)	917	CCT (P) => TCT (S)	1050	GCA (A) => GTA (V)
					1054	GCA (A) => GTA (V)		
rps2	83	TCA (S) => TTA (L)	*83*	TCA (S) => TTA (L)	83	TCA (S) => TTA (L)	83	TCA (S) => TTA (L)
rps8	48	GCG (A) => GTG (V)	*48*	GCG (A) => GTG (V)	48	GCG (A) => GTG (V)	48	GCG (A) => GTG (V)
rps14	27	TCA (S) => TTA (L)	*27*	TCA (S) => TTA (L)	27	TCA (S) => TTA (L)	27	TCA (S) => TTA (L)
	50	CCA (P) => CTA (L)	*50*	CCA (P) => CTA (L)	50	CCA (P) => CTA (L)	50	CCA (P) => CTA (L)

Supplementary Table S6. List of RNA editing sites by PREP program. The four species analyzed here belong to the subfamily Opuntioideae (Cactaceae). The abbreviation “aa pos” meaning the amino acid position of the RNA editing site in the aligned sequence. The codons highlighted in red edition gain in specie. The underlined under the amino acid refers to the edit position at the codon.

GENERAL CONCLUSIONS

In this work, we reported the complete plastomes of *Brasiliopuntia brasiliensis*, *Opuntia monacantha*, and *Opuntia ficus-indica*, all species belonging to the subfamily Opuntioideae, and *B. brasiliensis* being the biggest plastome at the moment in Cactaceae, 162,211 bp. We identified in the plastomes, a lot of rearrangements, inversions and translocation in the LSC region, expansions IRB-SSC boundaries, and only *O. manacantha* present IRB-LSC contraction.

Concerning the gene content, the plastomes of all species present the genes, *ycf1*, *ycf2* e *trnV-UAC* as a pseudogene, in addition, *B. brasiliensis* present *rpl20* gene like pseudogene, and O. *ficus indica* with gene loss *ycf2*. The gene *trnV-UAC* is absent in all Cactoideae, but is an essential gene, and how is it here a pseudogene, we lead analysis structural and also of the codon usage e the importation is hypothesis more possible.

More than half of genes identified under positive selection are the coding-protein plastid genes: *accD*, *clpP*, *rpl*, *rps* genes, with are the most divergence genes. These genes related to essential functions in the plastid (i.e., fatty acid biosynthesis, gene expression, and photosynthesis).

The NDH complex is degenerate in Cactoideae, here in Opuntioideae this complex is functional, therefore, we calculation substitutions synonymous (dS) and non-synonymous (dN) rates for these ndh genes, and all 11 genes compose this complex present low ratio substitution dN/dS in *B. brasiliensis*. In addition, these genes present low divergence in Opuntioideae, demonstrating to be conserved within this subfamily.

The features of plastid genes of the plastome sequenced here included 68 sites of RNA editing sites, of which 64 shared in Opuntioideae. We also map microsatellites (SSR), and 227 occurrences are shared among *Opuntia* and *B. brasiliensis*. These plastid molecular markers could provide information useful for genetic studies in Cactaceae, and conservation of natural resources.

The phylogeny presents here, based on concatenated plastid genes, result in well-supported. Opuntioideae and Cactoideae got separated, and Opuntioideae formed a monophyletic group.

Finally, the complete *Opuntia* and *Brasiliopuntia* plastomes, sequenced here, helped to understand the evolution of the Cactaceae family and the Opuntioideae subfamily.