TANARA PLETSCH DALLA COSTA

PLASTOMES OF THE SUBFAMILY CACTOIDEAE (CACTACEAE): GENOMIC, STRUCTURAL AND EVOLUTIONARY ASPECTS

Dissertation submitted to the Plant Physiology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

Adviser: Marcelo Rogalski

VIÇOSA - MINAS GERAIS
2022

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

T
D144p
2022

Dalla-Costa, Tanara Pletsch, 1996-
Plastomes of the subfamily Cactoideae (Cactaceae): genomic, structural and evolutionary aspects / Tanara Pletsch Dalla-Costa. – Viçosa, MG, 2022.
1 dissertação eletrônica (118 f.): il. (algumas color.).

Texto em inglês.
Orientador: Marcelo Rogalski.
Dissertação (mestrado) - Universidade Federal de Viçosa, Departamento de Biologia Vegetal, 2022.
Inclui bibliografia.
DOI: https://doi.org/10.47328/ufvbbt.2023.003
Modo de acesso: World Wide Web.

1. Plantas suculentas - Evolução. 2. Marcadores genéticos. 3. Plastídios. 4. Ácido ribonucléico - Síntese. I. Rogalski, Marcelo, 1975-. II. Universidade Federal de Viçosa. Departamento de Biologia Vegetal. Programa de Pós-Graduação em Fisiologia Vegetal. III. Título.

CDD 22. ed. 581.754

Bibliotecário(a) responsável: Euzébio Luiz Pinto CRB-6/3317

TANARA PLETSCH DALLA COSTA

PLASTOMES OF THE SUBFAMILY CACTOIDEAE (CACTACEAE): GENOMIC, STRUCTURAL AND EVOLUTIONARY ASPECTS

Dissertation submitted to the Plant Physiology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

APPROVED: October14, 2022.

Assent:

Documento assinado digitalmente
gov.br
TANARA PLETSCH DALLA COSTA
Data: 04/01/2023 23:02:15-0300
Verifique em https://verificador.iti.br

Tanara Plestch Dalla Costa
Author

Documento assinado digitalmente
gov.br
MARCELO ROGALSKI
Data: 09/01/2023 18:38:46-0300
Verifique em https://verificador.iti.br

Marcelo Rogalski
Adviser

ACKNOWLEDGMENTS

The biggest and most important thanks to God, Lord of my life, for the countless blessings He grants me, putting such good people in my path, and for allowing me to fulfill bigger dreams than I ever imagined.

I am very grateful to my parents Evair and Clari, my brothers Thiago and Thomas, my sister-in-law Thaís, and my beautiful niece Cecília, for being my base, motivation, and my source of love and encouragement. I am very blessed by the family that I have.

I thank my advisor, Professor Marcelo Rogalski, for all the conversations, remote meetings, teachings, and the trust built over these two years of my master's studies.

I thank all my colleagues of the Plant Molecular Physiology Laboratory, Amanda, Túlio, Daniel, Gleyson, Clara, João, and Ian, for their help with the analysis and the coffee moments. I especially thank Maria Carolina, for all the help during the development of this work, for advice, outbursts, happy hours, exchange of recipes, and, mainly, for the friendship.

Special thanks to my boyfriend Rhaí, for his love and companionship, for supporting me on the most difficult days, for sharing such happy moments, and for making everything lighter.

I thank my friends and master's colleagues Antônio and Ueliton, for sharing with me the joys and difficulties of the master's studies. Thank you to all the friends who stayed with me throughout this journey.

I thank the Universidade Federal de Viçosa, the Plant Physiology Graduate Program, and CNPq for this opportunity.

And to all who contributed directly and indirectly to the execution of this work, thank you very much!

“There is no limiting belief that resists a concrete result”.
(Kenia Gama)

ABSTRACT

DALLA COSTA, Tanara Pletsch, M.Sc., Universidade Federal de Viçosa, October, 2022. **Plastomes of the subfamily Cactoideae (Cactaceae): genomic, structural and evolutionary aspects**. Adviser: Marcelo Rogalski.

The family Cactaceae represents the greatest lineage of succulent plants, which contains a set of anatomical, morphological, and physiological modifications allowing adaptation and diversification in arid and semi-arid regions. Cactoideae is the richest subfamily in species diversity within Cactaceae. Representative species of this subfamily, such as *Melocactus glaucescens*, *Lepismium cruciforme*, and *Schlumbergera truncata* demonstrate interesting evolutionary aspects. They have different lifestyles and inhabit places with specific environmental conditions. Their distribution varies from a high degree of endemism and risk of extinction, to wide occurrence and status of extensive ornamental cultivation. Therefore, they are notable species to be used in studies related to plastid genomics. Plastid genomics allows analyzing and understanding phylogenetic and evolutionary relationships at the level of family/subfamily, as well as providing molecular markers that can help in conservation of natural resources. Although few cactus species with complete plastid sequences are available, many unusual features have been identified in the plastomes of the family. Thus, to expand the information concerning plastid evolution in Cactaceae, this study aimed at sequencing, assembling and analyzing the plastomes of *M. glaucescens*, *L. cruciforme*, and *S. truncata*. In addition, to provide complete plastome sequences, enabling analysis of plastome structures, frequency of codon usage, phylogeny, nucleotide diversity hotspots, gene divergence, signatures of positive selection, RNA editing sites and mapping of repetitive sequences. According to the data obtained, the plastomes showed several losses of genes and/or pseudogenization, including essential genes. Pseudogenization of the *trnT-GGU* gene in *L. cruciforme* was the first loss of gene functionality reported in the family. Moreover, *M. glaucescens* lost the *trnV-GAC*, *trnV-UAC* (common loss to the subfamily Cactoideae), and *trnA-UGC* genes, the last two being essential for plastid translation and cell survival, which strongly suggest the importation of tRNAs from the cytosol to the plastids in Cactoideae. Concerning plastomic structures, *M. glaucescens* showed unique rearrangements that culminated in the expansion of the inverted repeat (IRs) regions, the largest in the subfamily, and the duplication of unusual genes. On the other hand, the analysis of the plastomes of *L.*

cruciforme and *S. truncata* revealed a conserved structure among the tribe Rhipsalideae, with tribe-specific rearrangements and some variations in the size and content of the IRs. Molecular analyses in plastomes of the family Cactaceae was essential to identify high genetic divergence, numerous signatures of positive selection, and polymorphism of RNA editing sites. Furthermore, hundreds of plastid molecular markers were mapped for the three species, which will be useful for accessing the genetic information of natural populations and conservation of species, especially those threatened. Therefore, this study provides new insights into plastid evolution in Cactaceae, which is an exceptional lineage adapted to extreme environmental conditions and a notorious example of atypical plastome evolution.

Keywords: Cactaceae. Plastid evolution. Rearrangements. Gene divergence. RNA editing. Plastid molecular markers.

RESUMO

DALLA COSTA, Tanara Pletsch, M.Sc., Universidade Federal de Viçosa, outubro de 2022. **Plastomas da subfamília Cactoideae (Cactaceae): aspectos genômicos, estruturais e evolutivos**. Orientador: Marcelo Rogalski.

A família Cactaceae representa a maior linhagem de plantas suculentas, as quais possuem uma série de características anatômicas, morfológicas e fisiológicas que permitiram sua adaptação e diversificação em regiões áridas e semiáridas. A subfamília Cactoideae é a mais rica em diversidade de espécies. Representantes desta subfamília, como *Melocactus glaucescens*, *Lepismium cruciforme* e *Schlumbergera truncata*, contêm aspectos evolutivos interessantes. Possuem diferentes estilos de vida e habitam locais com condições ambientais específicas, a distribuição varia desde um alto grau de endemismo e risco de extinção até a ampla ocorrência e status de extenso cultivo ornamental. Assim, são espécies notáveis para estudos em genômica de plastídios, a qual permite analisar e compreender as relações filogenéticas e evolutivas da família, além de fornecer marcadores moleculares plastidiais que auxiliam no acesso e conservação dos recursos naturais. Embora ainda existam poucas espécies de cactos com sequências plastidiais completas disponíveis, muitas características incomuns foram identificadas nos plastomas da família. Dessa forma, visando ampliar as informações sobre a evolução do genoma plastidial de Cactaceae, este estudo teve como objetivo sequenciar, montar e analisar detalhadamente os plastomas de *M. glaucescens*, *L. cruciforme* e *S. truncata*. Além de fornecer as sequências plastidiais completas, foram realizadas análises comparativas das estruturas plastômicas, frequência de uso de códons, inferências filogenéticas, identificação de hotspots de diversidade de nucleotídeos, divergência gênica, predição de sítios de assinatura de seleção positiva e sítios de edição de RNA e mapeamento de sequências repetitivas. De acordo com os dados obtidos, os plastomas apresentaram extensa perda de genes e pseudogenizações, incluindo genes essenciais. A pseudogenização do gene *trnT-GGU* em *L. cruciforme* foi a primeira perda de funcionalidade do gene relatada na família. Além disso, *M. glaucescens* perdeu os genes *trnV-GAC*, *trnV-UAC* (perda comum para a subfamília Cactoideae) e *trnA-UGC*, sendo os dois últimos indispensáveis para a tradução plastidial e sobrevivência celular, sugerindo fortemente a importação de tRNAs do citosol aos plastídios em Cactoideae. Em relação às estruturas plastômicas, *M. glaucescens* apresentou rearranjos únicos que culminaram na expansão das IRs, a maior da subfamília, e na duplicação de genes incomuns. Por outro lado,

a análise dos plastomas de *L. cruciforme* e *S. truncata* revelou uma estrutura conservada entre a tribo Rhipsalideae, com rearranjos tribo-específicos e algumas variações no tamanho e conteúdo das regiões invertidas repetidas (IRs). Quanto à análise da evolução molecular na família Cactaceae, foi identificada alta divergência gênica, inúmeras assinaturas de seleção positiva e polimorfismo de sítios de edição de RNA. Além disso, foram mapeados centenas de marcadores moleculares plastidiais para as três espécies, que serão úteis para acessar informações genéticas de populações naturais e na conservação de espécies, principalmente aquelas ameaçadas. Portanto, este estudo fornece novos *insights* sobre a evolução plastidial em Cactaceae, que é uma linhagem excepcionalmente adaptada a condições ambientais extremas e um notório exemplo de evolução atípica dos plastomas.

Palavras-chave: Cactaceae. Evolução plastidial. Rearranjos. Divergência gênica. Edição de RNA. Marcadores moleculares plastidiais.

SUMMARY

1. GENERAL INTRODUCTION

Plants are responsible for one of the most important processes to maintain the life on Planet Earth: the photosynthesis. In this process, solar energy is converted into chemical energy to synthesize organic compounds for the plant, releasing oxygen into the atmosphere (Stirbet et al. 2020). Photosynthesis occurs in chloroplasts (i.e. differentiated plastids specialized in photosynthesis), organelles responsible for the synthesis of several essential compounds such as fatty acids, amino acids, pigments, hormones, nitrogen assimilation, among other functions (Daniell et al. 2016; Dobrogojski et al. 2020).

Chloroplasts are semiautonomous organelles since they import thousands of nuclear proteins, but they have their own DNA, called plastid genome or plastome (Dobrogojski et al. 2020). According to the endosymbiotic theory, a primitive eukaryotic cell would have engulfed an ancestral photosynthetic cyanobacterium, which originated the plastid (Allen 2015). However, plastomes have undergone an intense reorganization of structure and gene content throughout evolution, leading to a significant reduction in size (i.e. massive gene transfer to the nucleus; Allen 2015; Rogalski et al. 2015). The plastome is generally a circular molecule, containing a typical quadripartite structure, composed of two single-copy regions, one *large single copy* (LSC), and one *small single copy* (SSC), interspersed by two repeated and inverted sequences, the *inverted repeats* (IRA and IRB) (Lopes et al. 2019; Pacheco et al. 2020; Silva et al. 2021). The plastomes of land plants vary in size between 120-220 kb, which encode around 130 genes (Rogalski et al. 2015).

Plastid genes are mainly involved in two functional classes: photosynthetic and gene expression machineries. The first group is formed by: components of photosystems I and II, cytochrome *b6f* complex and ATP synthase (*psa*, *psb*, *pet* and *atp* genes, respectively); photosystem assembly factors (*ycf3* and *ycf4* genes); NAD(P)H-plastoquinone oxidoreductase complex (*ndh* genes); and the large subunit of the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO – *rbcL* gene). The second group is composed of RNA polymerase subunits, ribosomal proteins, as well as ribosomal and transport RNAs (*rpo*, *rps*, *rpl*, *rrn* and *trn* genes), translation initiation factor (*infA*); and intron maturase (*matK*). Other genes act in fatty acid biosynthesis (*accD*), cytochrome biogenesis (*ccsA*), inner envelope protein that acts on CO_2 uptake (*cemA*), protein homeostasis (*clpP*), and protein import to the plastids (*ycf1* and *ycf2*) (Bock 2007; Wicke et al. 2011; Daniell et al. 2016).

Plastid genomics offers useful tools for different studies (Rogalski et al. 2015). For example, comparative analyses among plastomes can elucidate relationships with environmental adaptation and speciation (Lopes et al. 2018; Lopes et al. 2021). Spacer regions and molecular markers are used for population genetics and phylogeography studies (Stefenon et al. 2019). Phylogenetic inferences and plant evolution can be determined through coding regions and genome rearrangements (Lopes et al. 2019; Pacheco et al. 2020; Silva et al. 2021). Furthermore, it is possible to genetically transform chloroplasts for studies in functional genetics and biotechnological approaches (Rogalski and Carrer 2011; Bock 2015).

Although the plastid genome of angiosperms has conserved structure and gene content (Dobrogojski et al. 2020), recent studies have revealed atypical events in several lineages. These events include IR expansion (Chumley et al. 2006) and contraction (Guisinger et al. 2011) in Geraniaceae, IR loss in Fabaceae (Moghaddam and Kazempour-Osaloo 2020), gene transfer to the nuclear genome in Ranunculaceae (Park et al. 2015) and Euphorbiaceae (Alqahtani and Jansen 2021), and numerous gene losses and pseudogenization in all the lineages mentioned above, as well as in Passifloraceae (Pacheco et al. 2020a; Pacheco et al. 2020b), Amaryllidaceae (Scobeyeva et al. 2021), Gentianaceae (Fu et al. 2021), among others.

The family Cactaceae, for example, exhibits several unusual characteristics. Among them, the reduction of plastomes (*M. zephyrantoides* with 107,343 kb in size; Solórzano et al. 2019), loss of the IR region in six species (*Carnegiea gigantea*, Sanderson et al. 2015; *Lophocereus schottii*; *Echinocactus grusonii*; *Ferocactus latispinus*; *Ferocactus setispinus*; *Leuchtenbergia principis*; information available in the NCBI database), IR expansion and contraction events, as well as highly rearranged genomes (Silva et al. 2021; Dalla Costa et al. 2022). In addition, many gene losses have been reported (i.e., genes *rpl20*, *rpl23*, *rpl33*, *rpl36*, *rps16*, *rps18*, *ycf1*, *ycf2* and *ycf4* and some tRNAs), including degeneration of the *ndh* complex in the subfamily Cactoideae (Sanderson et al. 2015; Solórzano et al. 2019; Silva et al. 2021; Oulo et al. 2020; Almeida et al. 2022; Dalla Costa et al. 2022). High gene divergence, intron loss, RNA editing sites, and positive selection signatures have also been described (Silva et al. 2021; Dalla Costa et al. 2022).

The family Cactaceae, belonging to the Caryophyllales order, has more than 100 genera and about 1400 recognized species. It is the largest lineage of succulent plants, adapted to arid and semi-arid environments, whose main centers of diversification are the Andes region, Central Mexico, and northeastern Brazil (Guerreiro et al. 2019). Although they grow slowly and under conditions unfavorable to survival, cacti diverged and diversified rapidly throughout

evolution (Arakaki et al. 2011), due to the anatomical, morphological, and physiological modifications in cacti. The main adaptations are: wax deposition and thick cuticle, a large proportion of parenchymatic tissue, leaves reduced to spines, photosynthetically active stem, superficial roots, and the acid metabolism of crassulacean (CAM) (Griffiths and Males 2017; Guerreiro et al. 2019).

Cactaceae is composed of four subfamilies, namely Cactoideae, Maihuenioideae, Opuntioideae and "Pereskioideae". The Cactoideae subfamily is the richest in species diversity, which represents about 85% of the family and is distributed within nine tribes (Anderson 2001; Guerreiro et al. 2019). These species are highly variable in growth habit and morphology, including tree-like, shrub, cespitose, vine or epiphytic (Anderson 2001).

The tribe Cereeae comprises most genera and about three-quarters of the species in the Cactaceae family. In Brazil, it is the most representative tribe, having the Brazilian Northeast as its center of diversity (Taylor 1997). One of the genera that make up the tribe Cereeae is *Melocactus* Link & Otto, which houses 37 species, of which 21 are endemic to northeastern Brazil, with Bahia being the main center of diversity for this genus (Taylor 2000). Species of the genus *Melocactus* are rich in bioactive compounds and have antimicrobial, antiparasitic, and antioxidant activity, among others (Brandão et al. 2017; Aquino-Martins et al. 2019). For this reason, they are widely used in folk medicine (Silva et al. 2015).

However, several species of the genus *Melocactus* are endangered, due to several factors, such as slow growth, high degree of endemism, small populations, degradation of the natural habitat, associated with predatory collection with the removal of entire individuals from nature, either to their use or for the trade of ornamental plants (Machado 2009; Goettsch et al. 2015). One of the species on the IUCN Red List of Endangered Species with a high risk of extinction is *Melocactus glaucescens* Buining & Brederoo. Endemic to a small area of 1,742.10 km² located in Morro do Chapéu-BA, *M. glaucescens* occurs in rupestrian fields, with only four small recognized populations (Taylor and Zappi 2004). *M. glaucescens* has low genetic variability, a high level of inbreeding, and evidence of hybridization with *M. ernestii*, characteristics that compromise the reproductive success of the species and reinforce the risk of extinction (Taylor and Zappi 2004).

In turn, Rhipsalideae represents the main tribe of epiphytic cacti of the Cactaceae family (Anderson 2001; Calvente et al. 2011a). Its distribution is restricted to South America, except for *Rhipsalis baccifera*, inhabiting tropical and subtropical forests. The Brazilian Atlantic Forest is considered the center of diversity of the Rhipsalideae tribe, which ranges from coastal

environments to altitudes higher than 3,000 meters and, therefore, has rare, endemic and restricted species (Hunt et al. 2006; Calvente et al. 2011b). The tribe is formed by the genera *Rhipsalis*, *Hatiora*, *Lepismium* and *Schlumbergera* (Anderson 2001; Calvente et al. 2011a). *Lepismium cruciforme* (Vell.) Miq. is widely distributed in the South and Southeast regions of Brazil and occurs in most forest formations (Anderson 2001). *Schlumbergera truncata* (Haw.) Moran, popularly known as "flor-de-maio" or "Christmas cactus", is endemic to the state of Rio de Janeiro. Both species are grown in pots for ornamental purposes, much appreciated for the beauty of their flowers (Leal et al. 2007; Chornobrov and Bilous 2021).

Cacti are key species for the aridest regions, as, beyond their important ecological role, they are used for food, fodder, civil construction, fuel, folk medicine, cosmetics, as well as for ornamental purposes (Silva et al. 2015). However, the Cactaceae family is among the five most endangered taxonomic groups, with more than 30% of the species threatened, mainly by human action, associated with intrinsic factors of the species, such as slow growth, high degree of endemism, and reduced populations (Goettsch et al. 2015). In addition, there is still little information published on the plastid genome of the Cactaceae family, only 21 complete plastomes are available in databases such as the NCBI (National Center for Biotechnology Information; https://www.ncbi.nlm.nih.gov/).

In this context, this is a very interesting taxon to expand studies in plastid genomics, to better understand the evolutionary relationships of the Cactaceae family and how the different lifestyles and specific environmental conditions of each habitat in which these species live are inserted imply the evolution of plastomes. In addition, plastomes are rich sources of molecular markers, such as simple sequence repeats (SSRs), which allow the identification of specific markers at different taxonomic levels. These markers are extremely useful for screening and detection of conservation strategies of natural populations and endangered species (Lopes et al. 2019; Lopes et al. 2021).

Therefore, we extensively sequenced and analyzed the plastomes of *M. glaucescens*, *L. cruciforme* and *S. truncata*. In addition to the sequenced plastomes, we provided data regarding phylogenetic inferences, polymorphism hotspots, structural analysis, codon usage, gene divergence, positive selection, prediction of RNA editing sites, and mapping of molecular markers. Taken together, these data allow new insights into the evolution of the family Cactaceae, as well as help in conservation approaches and risk mitigation of endangered species.

REFERENCES

Allen JF (2015) Why chloroplasts and mitochondria retain their own genomes and genetic systems: Colocation for redox regulation of gene expression. PNAS 112(33):10231-10238. doi: 10.1073/pnas.1500012112

Almeida EM, Sader MA, Rodriguez PE, Loeuille B, Felix LP, Pedrosa-Harand A (2021) Assembling the puzzle: Complete chloroplast genome sequences of *Discocactus bahiensis* Britton & Rose and *Melocactus ernestii* Vaupel (Cactaceae) and their evolutionary significance. Braz. J. Bot 44:877–888. doi: 10.1007/s40415-021-00772-2

Alqahtani AA, Jansen RK (2021) The evolutionary fate of *rpl32* and *rps16* losses in the *Euphorbia schimperi* (Euphorbiaceae) plastome. Sci Rep 11**:**7466. doi: 10.1038/s41598-021-86820-z

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. PNAS 108: 8379-8384. doi: 10.1073/pnas.1100628108

Aquino-Martins VGQ, Melo LFM, Silva LMP, Targino de Lima TR, Fernandes Queiroz M, Viana RLS, Zucolotto SM, Andrade VS, Rocha HAO, Scortecci KC (2019) *In vitro* antioxidant, anti-biofilm, and solar protection activities of *Melocactus zehntneri* (Britton & Rose) pulp extract. Antioxidants 8(10):439. doi: 10.3390/antiox8100439

Bock R (2007) Structure, function, and inheritance of plastid genomes. In: Bock R (ed) Cell and molecular biology of plastids. Springer, Berlin, pp 29-63.

Bock R (2015) Engineering plastid genomes: methods, tools, and applications in basic research and biotechnology. Annu Rev Plant Biol 66:211–241. doi: 10.1146/annurevarplant-050213-040212

Brandão GHA, Rigo G, Roque AA, Souza ACD, Scopel M, Nascimento CAO, Tasca T, Pereira CG, Giordani RB (2017) Extraction of bioactive alkaloids from *Melocactus zehntneri* using supercritical fluid. J Supercrit Fluids 129:28-35. doi: 10.1016/j.supflu.2016.12.012

Calvente A, Zappi DC, Forest F, Lohmanna LG (2011a) Molecular phylogeny of tribe Rhipsalideae (Cactaceae) and taxonomic implications for *Schlumbergera* and *Hatiora.* Mol Phylogenet Evol 58(3):456-468. doi: 10.1016/j.ympev.2011.01.001

Calvente A, Zappi DC, Forest F, Lohmanna LG (2011b) Molecular phylogeny, evolution, and biogeography of South American epiphytic cacti. Int J Plant Sci 172(7):902–914. doi: 10.1086/660881

Chornobrov O, Bilous S (2021) In vitro plant regeneration of Christmas cactus (*Schlumbergera truncata* (Haw.) Moran) by indirect morphogenesis. Folia For Pol Ser A For 63(1):68-73. doi: 10.2478/ffp-2021-0007

Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK (2006) The complete chloroplast genome sequence of *Pelargonium* × *hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol 23(11):2175-2190. doi: 10.1093/molbev/msl089

Daniell H, Lin CS, Yu M, Chang WJ (2016) Chloroplast genomes: diversity, evolution and applications in genetic engineering. Genome Biol 17:2–29.

Dalla Costa TP, Silva MC, Lopes AS, Pacheco TG, Oliveira JD, Baura VA, Balsanelli E, Souza EM, Pedrosa FO, Rogalski M (2022) The plastome of *Melocactus glaucescens* Buining & Brederoo reveals unique evolutionary features and loss of essential tRNA genes. Planta 255:57. doi: 10.1007/s00425-022-03841-2

Dobrogojski J, Adamiec M, Luciński R (2020) The chloroplast genome: a review. Acta Physiol Plant 42:98. doi: 10.1007/s11738-020-03089-x

Fu PC, Sun SS, Twyford AD, Li BB, Zhou RQ, Chen SL, Gao QB, Favre A (2021) Lineage-specific plastid degradation in subtribe Gentianinae (Gentianaceae). Ecol Evol 11(7):3286-3299. doi: 10.1002/ece3.7281

Goettsch B, Hilton-Taylor C, Cruz-Piñón G et al (2015) High proportion of cactus species threatened with extinction. Nature Plants 1:15142. doi: 10.1038/nplants.2015.142

Griffiths H, Males J (2017) Succulent plants. Curr Biol 27:890–896. doi: 10.1016/j.cub.2017.03.021

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(1):583-600. doi: 10.1093/molbev/msq229

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. doi: 10.1093/jhered/esy064

Hunt D, Taylor NP, Charles G (2006) The New Cactus Lexicon. 2 vols. DH Books, Milborne Port. 900p.

Leal L, Biondi D, Nunes JRS (2007) Propagação por sementes de *Schlumbergera* (Haw.) Moran (flor-de-maio) em diferentes substratos. Acta Sci Biol Sci 29(3):277-280. doi: 10.4025/actascibiolsci.v29i3.480

Lopes AS, Pacheco TG, Nimz T, Vieira LN, Guerra MP, Nodari RO, Souza EM, Pedrosa FO, Rogalski M (2018) 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. doi: 10.1007/s0042 5-018-2841-x

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. doi: 10.1007/s00425-019-03217-z

Lopes AS, Pacheco TG, Silva ON, Vieira LN, Guerra MP, Matta EPL, Baura VA, Balsanelli E, Souza EM, Pedrosa FO, Rogalski M (2021) Plastid genome evolution in Amazonian açaí palm (*Euterpe oleracea* Mart.) and Atlantic forest açaí palm (*Euterpe edulis* Mart.). Plant Mol Biol 105(4):559-574. doi: 10.1007/s11103-020-01109-5

Machado MC (2009) The genus *Melocactus* in eastern Brazil: part I - An introduction to *Melocactus*. British Cactus & Succulent Journal, 27:1-16.

Moghaddam M, Kazempour-Osaloo S (2020) Extensive survey of the *ycf*4 plastid gene throughout the IRLC legumes: Robust evidence of its locus and lineage specific accelerated rate of evolution, pseudogenization and gene loss in the tribe Fabeae. PLOS ONE 15(3): e0229846. doi: 10.1371/journal.pone.0229846

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. doi: 10.3390/plants9080979

Pacheco TG, Lopes AS, Oliveira JD, Otoni WC, Balsanelli E, Pedrosa FO, Souza EM, Rogalski M (2020a) The complete plastome of *Passiflora cirrhiflora* A. Juss.: structural features, RNA editing sites, hotspots of nucleotide diversity and molecular markers within the subgenus Deidamioides. Braz J Bot 43:839-853. doi: 10.1007/s40415-020-00655-y

Pacheco TG, Lopes AS, Welter JF, Yotoko KSC, Otoni WC, Vieira LN, Guerra MP, Nodari RO, Balsanelli E, Pedrosa FO, Souza EM, Rogalski M (2020b) Plastome sequences of the subgenus *Passiflora* reveal highly divergent genes and specific evolutionary features. Plant Mol Biol 104:21-37. doi: 10.1007/s11103-020-01020-z

Park S, Jansen RK, Park S (2015) Complete plastome sequence of *Thalictrum coreanum* (Ranunculaceae) and transfer of the *rpl32* gene to the nucleus in the ancestor of the subfamily Thalictroideae. BMC Plant Biol 15:40. doi: 10.1186/s12870-015-04326

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. doi: 10.1111/j.1467-7652.2011.00621.x

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. doi: 10.3389/fpls.2015.00586

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. doi: 10.3732/ajb.1500184

Scobeyeva VA, Artyushin IV, Krinitsina AA, Nikitin PA, Antipin MI, Kuptsov SV, Belenikin MS, Omelchenko DO, Logacheva MD, Konorov EA, Samoilov AE, Speranskaya AS (2021) Gene loss, pseudogenization in plastomes of genus *Allium* (Amaryllidaceae), and putative selection for adaptation to environmental conditions. Front Genet 12:674783. doi: 10.3389/fgene.2021.674783

Silva GM, Lopes AS, Pacheco TG, Machado KLG, Silva MC, Oliveira JD, Baura VA, Balsanelli E, 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. doi: 10.1007/s40415-020-00689-2

Silva VA (2015) Diversidade de uso das cactáceas no nordeste do Brasil: uma revisão. Gaia scientia 9:175-182.

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. doi: 10.3390/plants8100392

Stefenon VM, Klabunde G, Lemos RPM, Rogalski M, Nodari RO (2019) Phylogeography of plastid DNA sequences suggests postglacial southward demographic expansion and the existence of several glacial refugia for *Araucaria angustifolia*. Sci Rep 9:2752. doi: 10.1038/s41598-019-39308-w

Stirbet A, Dušan Lazár, Ya Guo, Govindjee Govindjee (2020) Photosynthesis: basics, history and modelling. Ann Bot 126(4):511–537. doi: 10.1093/aob/mcz171

Taylor NP (1997) Cactaceae. In: Olfield (Ed.), Cactus and Succulent Plants: Status Survey and Conservation Action Plan. IUCN/SSC. Cactus and Succulent Specialist Group, Gland, Switzerland and Cambridge, 17-20.

Taylor NP (2000) Taxonomy and phytogeography of the Cactaceae of eastern Brazil. PhD Thesis, The Open University and Royal Botanic Gardens, Kew.

Taylor NP & Zappi DC (2004) Cacti of Eastern Brazil. Royal Botanic Gardens, Kew, London. 511p.

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. doi: 10.1007/s11103-011-9762-4

2. CHAPTER I

The plastome of *Melocactus glaucescens* Buining & Brederoo reveals unique evolutionary features and loss of essential tRNA genes

Tanara P. Dalla Costa [1], Maria C. Silva [1], Amanda de Santana Lopes[1], Túlio Gomes Pacheco[1], José D. de Oliveira[1], Valter A. de Baura[2], Eduardo Balsanelli[2], Emanuel Maltempi de Souza[2], Fábio de Oliveira Pedrosa[2], Marcelo Rogalski[1*]

[1] Laboratório de Fisiologia Molecular de Plantas, Departamento de Biologia Vegetal, Universidade Federal de Viçosa, Viçosa-MG, Brasil.
[2] Núcleo de Fixação Biológica de Nitrogênio, Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Paraná, Curitiba-PR, Brasil.

*Corresponding author
E-mail address: rogalski@ufv.br

Published in:
Planta (2022) 255(57):1 – 19
DOI: 10.1007/s00425-022-03841-2

Planta (2022) 255:57
https://doi.org/10.1007/s00425-022-03841-2

ORIGINAL ARTICLE

The plastome of *Melocactus glaucescens* Buining & Brederoo reveals unique evolutionary features and loss of essential tRNA genes

Tanara P. Dalla Costa[1] · Maria C. Silva[1] · Amanda de Santana Lopes[1] · Túlio Gomes Pacheco[1] · José D. de Oliveira[1] · Valter A. de Baura[2] · Eduardo Balsanelli[2] · Emanuel Maltempi de Souza[2] · Fábio de Oliveira Pedrosa[2] · Marcelo Rogalski[1]

Received: 12 October 2021 / Accepted: 24 January 2022 / Published online: 3 February 2022

Abstract

Main conclusion **The plastome of *Melocactus glaucescens* shows unique rearrangements, IR expansion, and unprecedented gene losses in Cactaceae. Our data indicate tRNA import from the cytosol to the plastids in this species.**

Abstract Cactaceae represents one of the richest families in keystone species of arid and semiarid biomes. This family shows various specific features comprehending morphology, anatomy, and metabolism, which allow them to grow under unfavorable environmental conditions. The subfamily Cactoideae contains the most divergence of species, which are highly variable in growth habit and morphology. This subfamily includes the endangered species *Melocactus glaucescens* (tribe Cereeae), which is a cactus endemic to the biome Caatinga in Brazil. Aiming to analyze the plastid evolution and develop molecular markers, we sequenced and analyzed in detail the plastome of *M. glaucescens*. Our analyses revealed that the *M. glaucescens* plastome is the most divergent among the species of the family Cactaceae sequenced so far. We characterized here unique rearrangements, expanded IRs containing an unusual set of genes, and several gene losses. Some genes related to the ndh complex were lost during the plastome evolution, while others have lost their functionality. Additionally, the loss of three tRNA genes (*trnA-UGC*, *trnV-UAC*, and *trnV-GAC*) suggests tRNA import from the cytosol to the plastids in *M. glaucescens*. Moreover, we identified high gene divergence, several putative positive signatures, and possible unique RNA-editing sites. Furthermore, we mapped 169 SSRs in the plastome of *M. glaucescens*, which are helpful to access the genetic diversity of natural populations and conservation strategies. Finally, our data provide new insights into the evolution of plastids in Cactaceae, which is an outstanding lineage adapted to extreme environmental conditions and a notorious example of the atypical evolution of plastomes.

Keywords Caatinga · Cactaceae · Plastid evolution · Molecular markers · tRNA import · RNA editing

Introduction

The family Cactaceae represents one of the biggest lineages of succulent plants adapted to arid and semiarid regions, whose main centers of diversification are the Andes, Central Mexico, and northeastern Brazil (i.e., biome Caatinga; Arakaki et al. 2011; Guerreiro et al. 2019). The capacity to grow and develop under adverse environments is related to a set of evolving features such as anatomy, morphology, physiological modifications, and different lifestyles acquired during the evolution (Griffiths and Males 2017). Cactaceae comprises four subfamilies: Cactoideae, Maihuenioideae, Opuntioideae, and Pereskioideae. Cactoideae is the richest subfamily in the diversity of species, including species

Communicated by Anastasios Melis.

✉ Marcelo Rogalski
rogalski@ufv.br

[1] Laboratório de Fisiologia Molecular de Plantas, Departamento de Biologia Vegetal, Universidade Federal de Viçosa, Viçosa, MG, Brazil

[2] Núcleo de Fixação Biológica de Nitrogênio, Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Paraná, Curitiba, PR, Brazil

highly diversified in growth habit and morphology (i.e., tree-like, shrubby, caespitose, climbing, or epiphytes; Guerreiro et al. 2019).

Although they are species of great importance to human populations in arid regions, being used for food, animal forage, traditional medicine, or other uses (Shetty et al. 2012; Osuna-Martínez et al. 2014; Mayer and Cushman 2019), cacti are found among the five taxonomic groups most threatened with extinction (Goettsch et al. 2015). *Melocactus glaucescens* Buining & Brederoo is one of the species listed as endangered in the IUCN Red List of Threatened Species (IUCN 2020).

The genus *Melocactus* (Tribe Cereeae) harbors 37 species, of which 21 are endemic to northeastern Brazil. The state of Bahia is the main diversity center of this genus (Taylor 2000). Some species of the genus *Melocactus* are rich in bioactive compounds, such as essential oils, flavonoids, steroids, terpenoids, and alkaloids. In the same way, they also contain compounds with antimicrobial, antiparasitic, and antioxidant activities (Brandão et al. 2017; Aquino-Martins et al. 2019); therefore, they are widely used in folk medicine (Albuquerque et al. 2007). Given the importance of the genus, approaches related to in vitro propagation and conservation are being carried out to attenuate the risk of extinction (Torres-Silva et al. 2018, 2020, 2021a, b). In this context, plastid genomics can also assist in conservation strategies given that simple sequence repeats (SSRs) can provide plastid molecular markers useful for population genetic studies (Wheeler et al. 2014; Rogalski et al. 2015; Lopes et al. 2018a, b, c; Pacheco et al. 2020a, b, c). Furthermore, plastid genomics allows us to determine the evolutionary history of a plant group via structural rearrangements and coding regions of plastomes (Rogalski et al. 2015; Lopes et al. 2021). Recently, some reports have highlighted the importance of plastid genomics for studies related to plastid-cytoplasm incompatibility and plant speciation (Greiner et al. 2008a, b; 2011; Greiner and Köhl 2014; Pacheco et al. 2020b; Zupok et al. 2021).

Angiosperm plastomes encode a conserved set of approximately 120 genes. These genes are mainly involved in photosynthesis (i.e., photosystems I and II, cytochrome b_6f complex, and ATP synthase) and plastid expression (RNA polymerase subunits, ribosomal proteins, rRNA, and tRNA) (Bock 2007; Daniell et al. 2016). The plastid genes are responsible for about 80 proteins, 30 tRNAs, and 4 rRNAs (Hu et al. 2015). Most plastid proteins make up large chimeric complexes, whose subunits are encoded by plastid and nuclear genes. Therefore, the functionality of these complexes is dependent on well-established genetic interactions and co-evolution of both genomes (Greiner et al. 2013). On the other hand, plastids normally contain a minimal set of tRNA genes sufficient to carry out plastid translation via wobble and superwobbling rules, whereby a single tRNA can read up to four synonymous codons (Crick 1966; Rogalski et al. 2008a; Alkatib et al. 2012a, b).

Although the structure and gene content of most angiosperm plastomes are conserved, recent studies have revealed atypical events in Cactaceae, including plastome size reduction, new IR configurations, loss and pseudogenization of several genes, and family-specific rearrangements (Sanderson et al. 2015; Solórzano et al. 2019; Köhler et al. 2020; Oulo et al. 2020; Silva et al. 2021). However, only 13 species of cacti have their complete plastome available to date in databases such as the NCBI (National Center for Biotechnology Information). Therefore, this is a very interesting taxon to expand studies in plastid genomics to understand specific evolutionary relationships within this intriguing family.

Here, we report the complete plastome of *M. glaucescens* with extensive analyses concerning its structure and genetic content. *M. glaucescens* has several unique characteristics within Cactaceae, including rearrangements, gene losses, and unusual gene duplication. Additionally, we have identified high gene divergence, genes under putative positive selection, and possible new RNA-editing sites, which along with phylogenetic inference provide new insights into the evolution of plastome within Cactaceae. Moreover, we map the repetitive sequences of the *M. glaucescens* plastome, which are helpful to manage natural populations aiming for conservation strategies. In addition to the loss of *ndh* complex, we highlight the loss of three tRNA genes (*trnA-UGC, trnV-GAC*, and *trnV-UAC*). The loss of the three tRNA genes indicates the occurrence of tRNA import from the cytosol to the plastids given that two tRNAs are essential for plastid translation and cell survival.

Materials and methods

Plant material, chloroplast isolation, and plastid DNA extraction

Fresh and young cladodes of the *Melocactus glaucescens* were collected from plants maintained in the greenhouse and obtained as described by Torres-Silva et al. (2020). Before chloroplast isolation, the plants were kept in the dark at a temperature of 4 °C, for 5 days to reduce starch levels. Subsequently, chloroplast isolation and plastid DNA extraction were carried out as described before by Vieira et al. (2014).

Sequencing, assembly, annotation, and data archiving statement

To prepare the sequencing libraries, approximately 1 ng of chloroplast DNA was used with Nextera XT DNA Sample Prep Kit (Illumina Inc., San Diego, CA, USA) according to the manufacturer's instructions, and sequenced using

the Illumina MiSeq platform (Illumina Inc., San Diego, CA, USA). The sequencing of *M. glaucescens* resulted in 538,422 reads (average length of 194.4 bp). Four contigs were used to assemble the *M. glaucescens* plastome, which ranged from 490.02 to 222.87 of average coverage. The reads were trimmed under the threshold with a probability of error < 0.05. The trimmed reads (537,878) were de novo assembled in contigs using the CLC genomics Workbench 8.0.2 software (CLC Bio, Aarthus, Denmark). After that, a gap of 219 bp was identified. To close the gap and confirm the plastome sequence, we performed a second sequencing, which resulted in 546,506 reads (average length of 192.17). The gap of 219 bp was completely closed with the data from the second sequencing using a contig of 119.16 average coverage. In addition, we used the software NOVOPlasty 4.0 and a pre-assembled genome to check the sequence (Dierckxsens et al. 2016). The genes were first annotated by the Annotation of Organellar Genomes (GeSeq) (Tillich et al. 2017) and BLAST programs. Subsequently, initiation, termination, and intron-codon positions were determined based on alignment with plastid genomes of related species available in the GenBank database. Gene losses were verified through a search along the plastome of *M. glaucescens*, using the coding sequences of the respective functional genes of the species *Nicotiana tabacum*, *Spinacia oleracea*, and *Portulaca oleracea*. The tRNA sequences were analyzed using the tRNAscan-SE webserver (Lowe and Chan 2016). The physical genome of the plastome was drawn using the software Organellar Genome DRAW (OGDRAW) (Greiner et al. 2019). The complete nucleotide sequence of *M. glaucescens* plastome was deposited in the GenBank database under accession number OK298499.

Comparative analyses of plastome structure

The structural features of the *M. glaucescens* plastome were determined by multiple alignment analysis using the Mauve Genome Alignment v.2.4.0 (MAUVE) software (Darling et al. 2004). For this analysis, we used as a reference the LSC region of *P. oleracea*, which extends from the *trnH-GUG* gene to the *rps19* gene. Posteriorly, we compared with the corresponding regions of the plastomes of *M. glaucescens*, *Rhipsalis teres*, and *Carnegiea gigantea* (subfamily Cactoideae). The linear maps of the plastid genes were drawn by OGDRAW (Greiner et al. 2019).

Codon usage

Codon usage frequency was determined using the Codon Usage webserver (http://www.geneinfinity.org/sms/sms_codonusage.html). Thus, we extracted the sequences of all functional protein-coding genes of nine species of the family Cactaceae, *P. oleracea*, and *S. oleracea* listed in Supplementary Table S1.

Phylogenetic inference

To infer the phylogenetic position of *M. glaucescens* within the suborder Cactineae, 26 taxons belonging to the suborder Cactineae, containing complete plastome sequences available in GenBank. The sampling included representative species of the families Cactaceae, Basellaceae, Montiaceae, Halophytaceae, Portulacaceae, and Talinaceae. To root the tree, *S. oleracea* (Chenopodiaceae: Caryophyllales) was used as an external species. All plastomes sampled and analyzed here are listed in Supplementary Table S2. The phylogenetic inference was performed using the maximum-likelihood (ML) method based on 54 concatenated plastid genes. For this purpose, plastid genes were first extracted from the GenBank database and individually aligned using the MUSCLE software (Edgar 2004) implemented in MEGA 7.0.14 (Tamura et al. 2013). The gene sequences were concatenated using DnaSP v.6.12.03 software (Rozas et al. 2017). Subsequently, a maximum-likelihood tree was constructed using IQTREE v.1.6.12 (Nguyen et al. 2015). Five hundred non-parametric bootstrap replications were used to evaluate the branch supports, which generated five partitions, grouping the genes in the best replacement evolutionary models: TVM + F + I + G4 (*atpA*, *atpB*, *atpE*, *atpF*, *ccsA*, *cemA*, *matK*, *petA*, *psaC*, *psbH*, *psbJ*, *psbK*, *rbcL*, *rpl14*, *rpl16*, *rpoA*, *rpoB*, *rpoC1*, *rpoC2*, *rps4*, *rps8*, *rps14*, and *ycf3*); GTR + F + I + G4 (*atpH*, *atpI*, *petB*, *petD*, *petG*, *petN*, *psaA*, *psaB*, *psaI*, *psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbL*, *psbN*, *psbT*, and *psbZ*); TIM3 + F + G4 (*clpP*); TVM + F + G4 (*infA*, *psbM*, *rpl2*, *rps2*, *rps3*, *rps7*, *rps11*, *rps12*, *rps15*, *rps19*); TVM + F + G4 (*rpl22*). Finally, the consensus tree generated in this analysis was visualized using the FigTree v.1.4.4 software (http://tree.bio.ed.ac.uk/software/figtree/).

Gene divergence analysis in plastomes of Cactaceae

We aligned 59 protein-coding genes found in all 14 plastomes of the family Cactaceae, including *M. glaucescens*, and the other eight genes absent in one or more of these species, using the software Muscle implemented in Mega 7.0 (Edgar 2004; Tamura et al. 2013). The 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 above-mentioned for the phylogenetic inference. The gene divergence was estimated by the sum of total branch lengths that link the operational taxonomic units to the common ancestor of the species sampled here.

To investigate the presence of putative positive signatures (positive selection), 67 protein-coding genes were

aligned with 26 other related species as described before. The aligned codon sequences were used as input data for the server Selecton (http://selecton.tau.ac.il/index.html; Stern et al. 2007). The synonymous rate (Ks) and the non-synonymous rate (Ka) were estimated under the M8 evolution model (Yang et al. 2000). We consider in our analysis only sites where the positive selection was reliable (lower bound > 1).

Prediction of RNA-editing sites and SSRs

Potential RNA-editing sites in plastid protein-coding genes of the *M. glaucescens*, *Rhipsalis baccifera*, and *Selenicereus undatus* were predicted by the program Predictive RNA Editor for Plants (PREP) (Mower 2009). The sites found in our analysis were compared with sites previously identified in other species of the family Cactaceae by Silva et al. (2021). We analyzed 27 genes, with a cut-off value defined as 0.8. Absent genes (*rpl23*, *ndhA*, *ndhF*, and *ndhG*), pseudogenes (*ndhB*, and *ndhD*), and highly divergent genes (*accD*, and *clpP*) were excluded from this predictive analysis.

Simple sequence repeats (SSRs) were mapped using the webserver MIcroSAtellite (MISA) Identification (https://webblast.ipk-gatersleben.de/misa/), 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.

Results

General features of *Melocactus glaucescens* plastome

The *M. glaucescens* plastome is a circular molecule of 129,806 bp in length. The typical quadripartite structure contains two single-copy regions, a large single copy (LSC) of 51.465 bp and a small single copy (SSC) of 19.097 bp, interspersed between two inverted repeats, IRA and IRB (IRs), of 29.622 bp (Fig. 1). This plastome contains 98 unique genes, of which 67 are protein-coding genes, 27 are tRNAs, and 4 are rRNAs (Table 1). A total of 11 genes harbor one intron, four of them are tRNAs, and seven are protein-coding genes. The *ycf3* gene contains two introns. Two genes have lost one intron, *rpl2* and *rpoC1*, and the *clpP* gene has lost the two introns normally conserved in the sequence.

Concerning the gene content, 15 genes are absent from *M. glaucescens* plastome. Ten of them were lost, including three tRNA genes (*trnA-UGC*, *trnV-GAC*, and *trnV-UAC*), *rpl23* gene, and six genes of the *ndh* complex (A, E, F, G, I, and K subunits). The other five genes of the *ndh* complex (B, C, D, H, and J subunits) are putative pseudogenes due to the presence of premature stop codons in its coding sequences or the presence of incomplete sequences, lacking initiation and termination codons. On the other hand, the *M. glaucescens* plastome exhibits two *clpP* gene segments, without introns, of which only one is functional, located between the *trnS-GCU* and *trnG-UCC* genes. The second segment of the *clpP* gene, located between the *rpl20* and *psbB* genes, is not functional and contains a copy of the first exon of the *rps12* gene.

M. glaucescens plastome is the second largest one of the subfamily Cactoideae among the cacti sequenced to date (Table 2). Furthermore, it exhibits the largest IRs of the family Cactaceae, which is possibly due to the expansion of the IRA/LSC border, resulting in the duplication of 26 genes. Of them, 22 are commonly located in the LSC region of angiosperms, such as *accD*, *rbcL*, *rps4*, *rps16*, *psbA*, *psbM*, *psaI*, *petN*, *atpB*, *atpE*, *matK*, *ycf4*, and nine tRNAs (*trnS-GGA* , *trnL-UAA*, *trnF -GAA*, *trnT-UGU*, *trnC-GCA*, *trnE-UUC*, *trnY-GUA*, *trnM-CAU*, *trnK-UUU*, *trnH-GUG*).

The region harboring the genes between the *trnH-GUG* and the *rps19* gene (i.e., part of IRA and all LSC region) of the *M. glaucescens* plastome was compared by multiple alignments by MAUVE with *P. oleracea* as a reference and other cacti (subfamily Cactoideae; Supplementary Fig. S1). The local collinear blocks (LCB) represent conserved segments among the species. The analysis revealed unique rearrangements in the *M. glaucescens* plastome, which include events of translocations, inversions, gene loss, expansion of the IRs, and contraction of the LSC region (Supplementary Fig. S1; Fig. 2).

To improve the visualization of the rearrangements, we performed a comparative analysis between the complete plastomes of *M. glaucescens* and *P. oleracea*, using linear genetic maps of the plastomes (Fig. 2). Thus, we identified translocation and inversion events involving a large segment in the LSC region, covering the sequence from the *trnQ-UUG* gene to *ycf4* gene (LCB B, C, D, H, I, and J—event 1; Fig. 2). This event resulted in a new configuration of the order and direction of the genes in the plastome of *M. glaucescens* (LCB J, H, I, D, C, E, and B). Subsequently, the LSC region underwent secondary inversions in two small segments of the plastome (event 2; Fig. 2), which reestablished the direction of the genes, if we compare with the plastome of *P. oleracea*. One of them corresponds to the gene block from *trnY-GUA* to *trnD-GUG* (LCB C and D), and the other to the gene block from *trnM-CAU* to *rbcL* (LCB H and I). Three blocks of the LSC region (LCB B, E, and J; Fig. 2) remained inverted in the plastome of *M. glaucescens*.

The blocks highlighted in red (i.e., Losses 1–7; Fig. 2) represent the gene losses of *M. glaucescens* plastome. Three genes were lost from the LSC region (*ndhK*, *trnV-UAC*, and *rpl23*), two genes from the IRs (*trnA-UGC* and *trnV-GAC*), and five genes from the SSC (*ndhA*, *ndhE*, *ndhF*, *ndhG*,

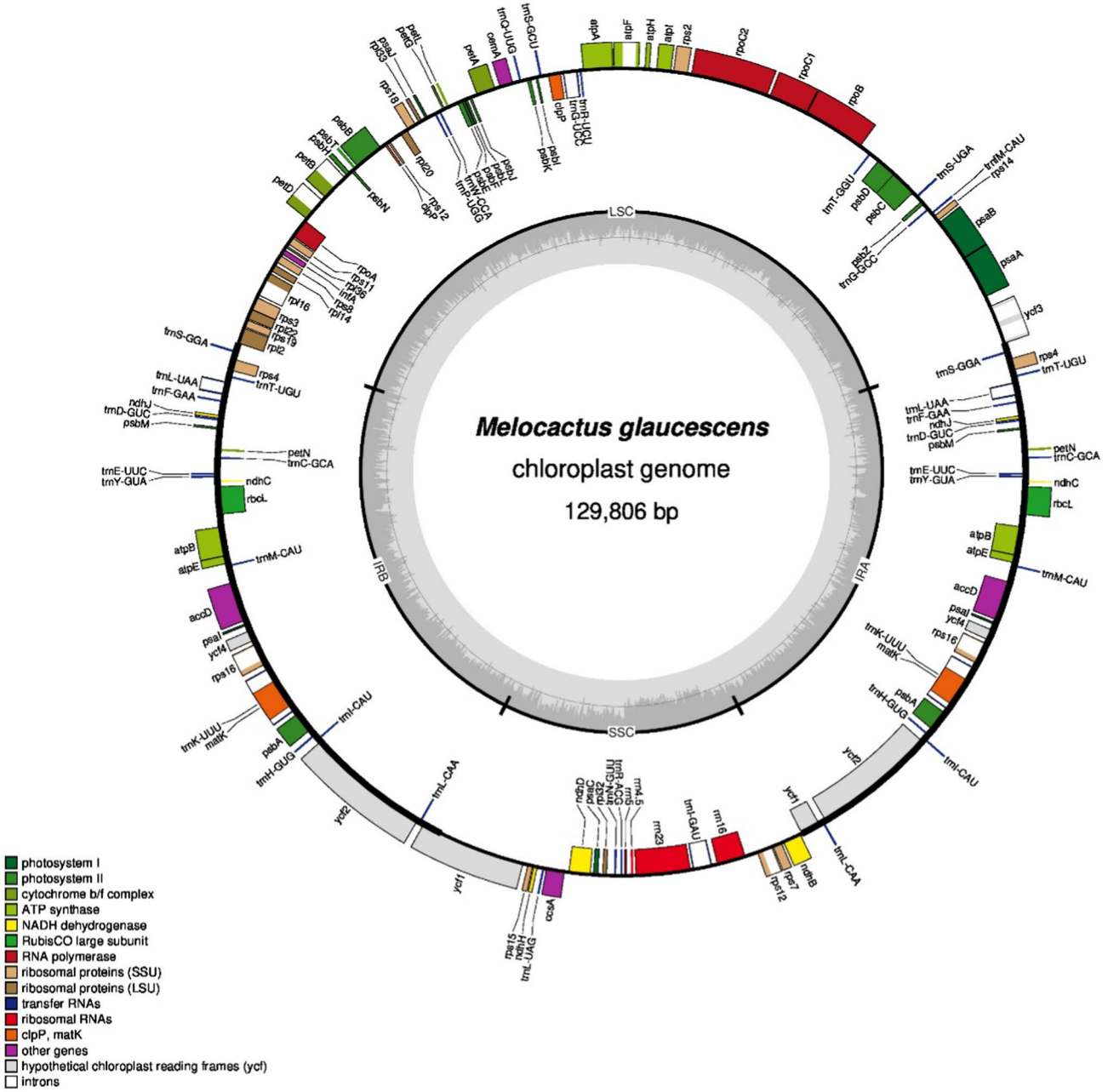


Fig. 1 Gene map of *M. glaucescens* plastome. Two inverted repeats, IR_A and IR_B, divide the circular DNA molecule into large (LSC) and small (SSC) single-copy regions. Genes drawn inside the circle are transcribed clockwise and genes drawn outside are expressed counter-clockwise. Genes belonging to different functional groups are color-coded. The darker gray in the inner circle corresponds to GC content, while the lighter gray corresponds to AT content. The dotted circle corresponds to 50% of AT/GC content

and *ndhI*) in comparison with the plastome of *P. oleracea*. In addition to gene losses, a reorganization of the SSC region in the plastome of *M. glaucescens* is also observed. The expansion of the SSC/IRA border culminated in the transfer of the four ribosomal genes (rRNA) from the IRs to the SSC region (i.e., Event 3; Fig. 2). Finally, the expansion of the IRA/LSC border resulted in IR expansion and LSC contraction in the plastome of *M. glaucescens* (i.e., Event 4; Fig. 2).

Codon usage

The *M. glaucescens* plastome lost three tRNA genes: *trnA-UGC, trnV-GAC*, and *trnV-UAC*, which encode three tRNAs, $tRNA^{Ala}$(UGC), $tRNA^{Val}$(GAC), and $tRNA^{Val}$(UAC), respectively. Given the complete absence of these genes in the plastome of *M. glaucescens*, we analyzed the codon usage frequency of the protein-coding genes in this cactus by

Table 1 List of genes identified in the plastome of *Melocactus glaucescens*

Group of gene	Name of gene
Gene expression machinery	
Ribosomal RNA genes	*rrn16; rrn23*; *rrn5; rrn4.5*
Transfer RNA genes	*trnC-GCA*[b]*; trnD-GUC*[b]*; trnE-UUC*[b]*; trnF-GAA*[b]*; trnfM-CAU; trnG-UCC*[a]*; trnG-GCC; trnH-GUG*[b]*; trnI-CAU*[b]*; trnI-GAU*[a]*; trnK-UUU*[ab]*; trnL-CAA*[b]*; trnL-UAA*[ab]*; trnL-UAG; trnM-CAU*[b]*; trnN-GUU; trnP-UGG; trnQ-UUG; trnR-ACG; trnR-UCU; trnS-GCU; trnS-UGA; trnS-GGA*[b]*; trnT-UGU*[b]*; trnT-GGU; trnW-CCA; trnY-GUA*[b]
Small subunit of ribosome	*rps2; rps3; rps4*[b]; *rps7; rps8; rps11; rps12*[a]; *rps14; rps15*; *rps16*[ab]; *rps18; rps19*
Large subunit of ribosome	*rpl2*; *rpl14; rpl16*[a]; *rpl20; rpl22; rpl32; rpl33; rpl36*
DNA-dependent RNA polymerase	*rpoA; rpoB; rpoC1*; *rpoC2*
Genes for photosynthesis	
Subunits of photosystem I (PSI)	*psaA; psaB; psaC; psaI*[b]*; psaJ; ycf3*[a]; *ycf4*[b]
Subunits of photosystem II (PSII)	*psbA*[b]; *psbB; psbC; psbD; psbE; psbF; psbH; psbI; psbJ; psbK; psbL; psbM*[b]*; psbN; psbT; psbZ*
Subunits of cytochrome b_6f	*petA; petB*[a]*; petD*[a]*; petG; petL; petN*[b]
Subunits of ATP synthase	*atpA; atpB*[b]*; atpE*[b]*; atpF*[a]; *atpH; atpI*
Large subunit of Rubisco	*rbcL*[b]
Other genes	
Maturase	*matK*[b]
Envelope membrane protein	*cemA*
Subunit of acetyl-CoA carboxylase	*accD*[b]
C-type cytochrome synthesis gene	*ccsA*
Clp Protease	*clpP*[c]
Component of TIC complex	*ycf1*[c]
Component of 2-MD heteromeric ppAAA-ATPase complex	*ycf2*[b]
Translation initiation factor IF-1	*infA*
Pseudogenes	*ndhB; ndhC; ndhD; ndhH; ndhJ*
Absent	*trnA-UGC; trnV-GAC; trnV-UAC; rpl23; ndhA; ndhE; ndhF; ndhG; ndhI; ndhK*

[a]Genes containing introns

[b]Duplicated gene

[c]Partially duplicated genes

Table 2 General features of plastomes within Cactaceae

Species	Subfamily	Size (bp)	LSC (bp)	SSC (bp)	IR (bp)	Number of genes	GC (%)	GenBank
M. glaucescens	Cactoideae	129,806	51,465	19,097	29,622	98	36.5	OK298499
Rhipsalis teres	Cactoideae	122,389	81,397	24,016	8,488	99	36.7	MT387452
Rhipsalis baccifera	Cactoideae	122,333	81,459	23,531	8530	101	36.7	MT821847
Selenicereus undatus	Cactoideae	133,326	68,256	21,716	21,677	99	36.4	NC_053698
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
M. pectinifera	Cactoideae	108,561	72,273	29,744	772	95	36.4	MN519716
M. crucigera	Cactoideae	115,505	71,565	29,418	7,261	96	36.3	MN517613
M. huitzilopochtli	Cactoideae	115,886	71,997	29,401	7,244	97	36.3	MN517612
M. solisioides	Cactoideae	115,356	71,690	29,238	7,214	95	36.4	MN518341
M. supertexta	Cactoideae	116,175	72,240	29,445	7,245	97	36.4	MN508963
M. zephyranthoides	Cactoideae	107,343	71,811	7,281	14,126	95	38.5	MN517611
Opuntia quimilo	Opuntioideae	150,347	101,475	4,115	22,392	109	36.6	MN114084

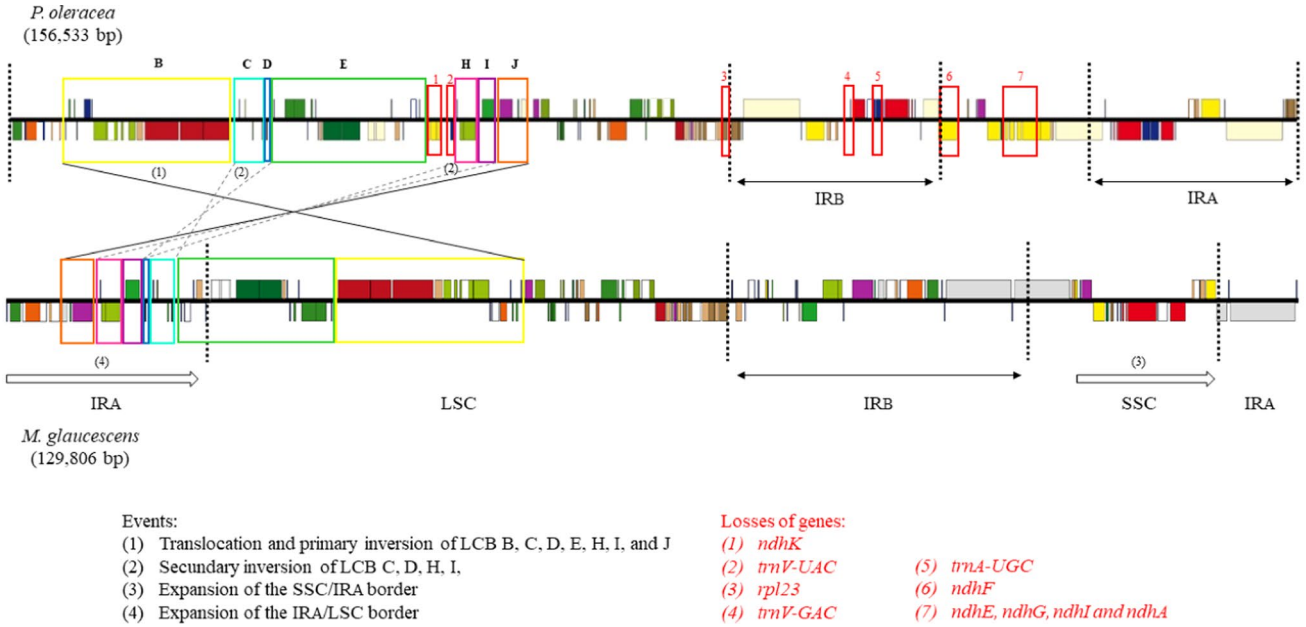


Fig. 2 Comparison of gene content and order between the plastomes of *M. glaucescens* and *P. oleracea* that presents the general structure of plastomes found in most angiosperms. The colored squares correspond to the LCB obtained by multiple alignments performed by MAUVE (Supplementary Fig. S1). Gene losses are pointed out by red squares in the plastome of *P. oleracea*. Solid cross lines mean translocation and inversion of the gene order. Dotted cross lines signify the only inversion of the gene order. Arrows indicate events of SSC/IRA and IRA/LSC border expansion. Rearrangement events are described according to the indicated numbering. Linear gene maps were drawn using OGDRAW (Greiner et al. 2019). LSC, large single-copy regions; IRA/B, inverted repeats A/B; SSC, small single-copy region

comparing it with other species of the family Cactaceae (Fig. 3). Both alanine and valine amino acids are encoded by a codon box (GC and GT, respectively). The alanine amino acid shows a high frequency of GCT codon (0.44), followed by GCA (0.29), GCC (0.16), and GCG (0.11) codons. These values are similar to the frequency of other cacti, whose averages were 0.43, 0.29, 0.16, and 0.10, respectively. In the same way, the frequencies of the valine codons observed in *M. glaucescens* and the average observed in other cacti are very similar. The most frequent codons are GTA (0.38 and 0.38) and GTT (0.36 and 0.37), and the less frequent ones are the GTG (0.14 and 0.14) and GTC codons (0.12 and 0.11), for *M. glaucescens* and the average of other species, respectively. These results indicate that despite the absence of tRNAs in the *M. glaucescens* plastome, the frequency of codon usage for the amino acids, alanine and valine, follows the same proportions observed in other species of the subfamily Cactoideae.

Phylogenetic reconstruction of the suborder Cactineae

The phylogenetic reconstruction of the suborder Cactineae by the maximum-likelihood (ML) method, based on concatenated plastid genes, generated a consensus tree with a log-likelihood (lnL) of -143,668,505 (Fig. 4). The family Montiaceae clade formed a sister group with the other families of the suborder Cactineae (100% of ML-BS). Basellaceae and Halophytaceae formed a sister group (50% of ML-BS), being Halophytaceae more divergent. Finally, the family Talinaceae was more closely related to the families Portulacaceae and Cactaceae (100% of ML-BS), which formed a well-supported sister group (91% of ML-BS). Within Cactaceae, *Opuntia quimilo* constituted the early branching of the family Cactaceae and formed a sister group with the subfamily Cactoideae clade (100% of ML-BS). The genus *Mammillaria* formed a monophyletic group highly supported and a sister group with the clade composed of the other species of the subfamily Cactoideae (100% of ML-BS). This group shows two clades: the first, a sister group among *S. undatus* and tribe Echinocereeae (*C. gigantea* and *L. schottii*); and the second, a sister group among the genera *Melocactus* (*M. glaucenscens*) and *Rhipsalis* (*R. baccifera* and *R. teres*), with 100% of ML-BS for all branches. *M. glaucescens* shows the longest branch length within Cactaceae, which means that it is the most divergent species in the family among the species analyzed here.

Fig. 3 Frequency and standard deviation (SD) of the codon usage analysis of *M. glaucescens* plastid protein-coding genes, compared to species of the family Cactaceae, *P. oleracea*, and *S. oleracea*. The red squares highlight the codons for the amino acids alanine and valine, whose respective tRNA genes (*trnA-UGC*, *tnrV-GAC*, *and trnV-UAC*) were lost from the *M. glaucescens* plastome

Gene divergence analysis and identification of putative positive signatures in plastid protein-coding genes of *Melocactus glaucescens*

From 67 plastid protein-coding genes of Cactaceae analyzed here, 35 demonstrated a low substitution rate, with branch length < 0.05 for all species (Supplementary Fig. S2). The other 32 genes exhibited branch lengths containing values higher than 0.05 in at least one species. The genes *accD*, *ycf1*, *clpP*, *rpl22*, *rpl32*, and *rps19* were the most divergent genes, demonstrating values of branch length > 0.2 (Fig. 5a). The *rpl32* gene showed the highest variation of branch lengths among the species. The highest substitution rate was observed in *M. glaucescens* (0.87) and lower rates for the genus *Mammillaria* (from 0.15 to 0.17). The *accD* is the most divergent gene for the family Cactaceae, with

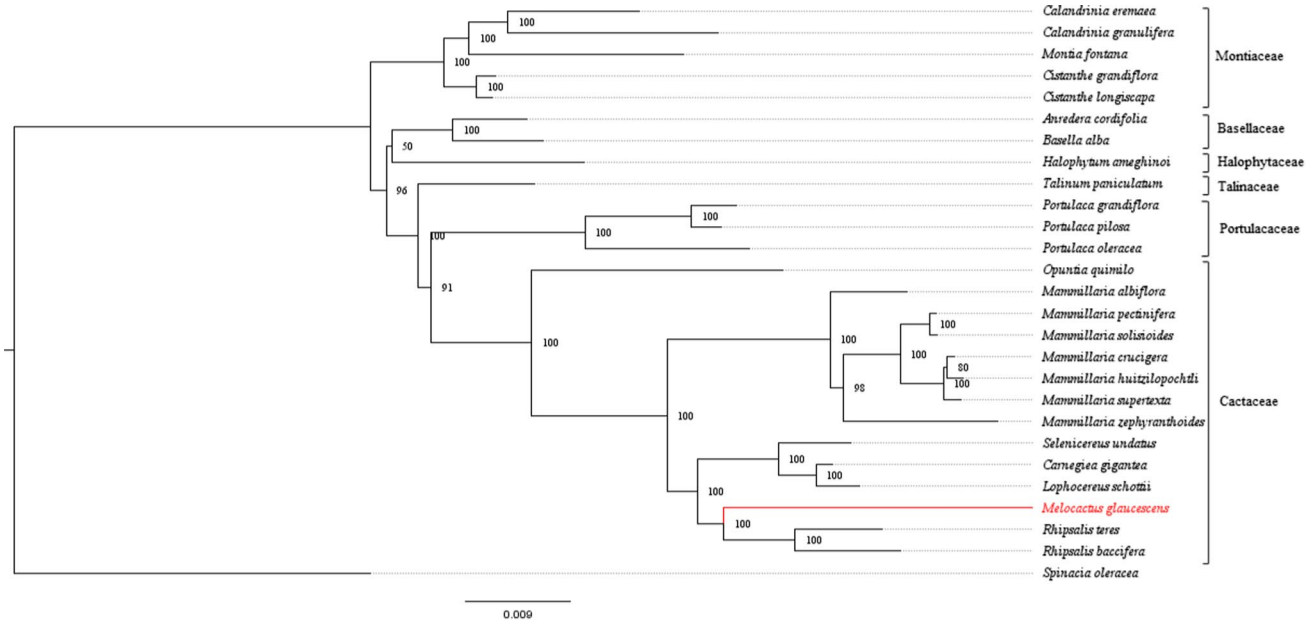


Fig. 4 Phylogenetic tree of 26 species of the suborder Cactineae and *Spinacia oleracea* as an outgroup, based on 54 protein-coding plastid genes using the maximum-likelihood (ML) method. Numbers (%) associated with branches are ML bootstrap support (BS) values. The branch length is proportional to the inferred divergence level. The scale bar indicates the number of inferred nucleic acid substitutions per site. The position of *M. glaucescens* is highlighted in red

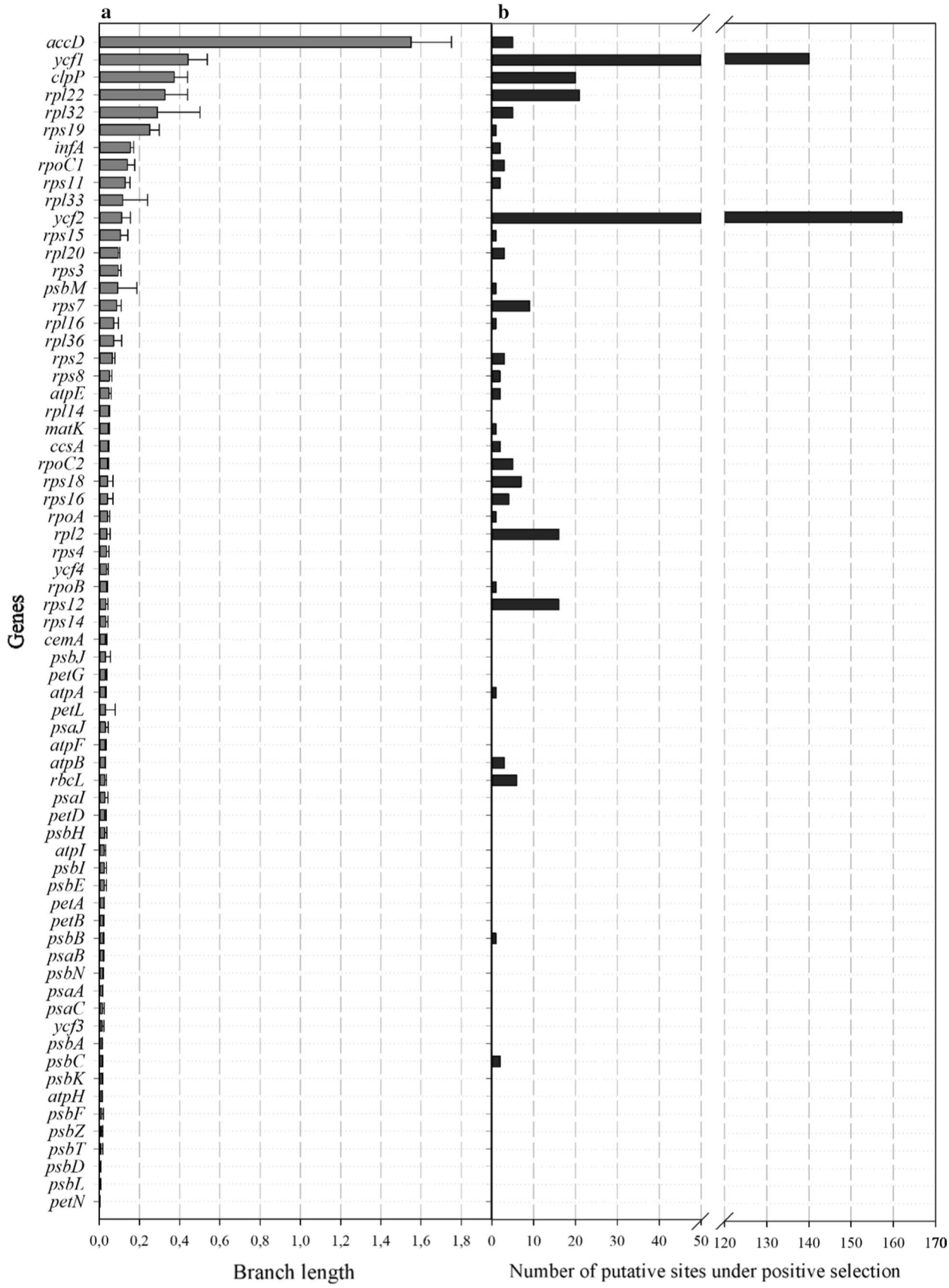
branch length values varying from 1.29 (*M. huitzilopochtli*) to 1.95 (*O. quimilo*). The *ycf1* gene, absent in *M. zephyranthoides*, showed branch lengths ranging from 0.34 (*C. gigantea*) to 0.71 (*O. quimilo*). The *rpl22* gene presented the lower substitution rate in *O. quimilo* (0.20), followed by the genus *Mammilaria* (0.24–0.31) and the higher substitution rate in *M. glaucescens* (0.58). On the other hand, the *clpP* gene exhibited branch lengths varying from 0.26 (*C. gigantea*) to 0.49 (*M. glaucescens*). Finally, in the gene, *rps19* was observed a variation of branch lengths from 0.16 (*O. quimilo*) to 0.31 (*C. gigantea*, *M. glaucescens*, and *M. solisioides*). Our analysis revealed that *M. glaucescens* showed higher substitution rates in 21 out of 67 genes analyzed here (31.3%); most of them (15 genes) involved in plastid gene expression machinery.

To investigate whether any plastid gene of Cactaceae underwent positive selection, we utilized the Selecton program. A total of 450 putative positive signatures distributed in 32 different genes were identified (Fig. 5b). More than half of genes under putative positive selection (18) are related to plastid gene expression [i.e., including subunits of RNA polymerase (*rpo* genes), RNA maturation (*matK* gene), and ribosomal proteins (*rpl* and rps genes)]. The other genes (14) are involved in functions such as photosynthesis [i.e., subunits of PSII (*psb* genes), ATP synthase (*atp* genes)], translation initiation (*infA* gene), cytochrome biosynthesis (*ccsA* gene), fatty acid biosynthesis (*accD* gene), protein import from the cytosol (*ycf1* and *ycf2* genes), and protein degradation (*clpP* gene). The sites under putative positive selection were plotted against the phylogeny based on concatenated plastid genes of the suborder Cactineae (Fig. 6; Supplementary Figs. S3–S24). The sites of putative positive selection are related to gene divergence, since most divergent genes also presented a high number of positive signatures. The high gene divergence naturally found in the family Cactaceae (for example in the *ycf1* and *ycf2* genes) is directly related to signatures of putative positive selection. Although the program Selecton interprets them as positive selection sites, we should interpret them with caution due to the high gene divergence found in plastomes of this family.

RNA-editing sites in plastid protein-coding genes of *Melocactus glaucescens* and other species of the subfamily Cactoideae

Here, we identified a total of 29 putative RNA-editing sites in 15 genes of *M. glaucescens* (Supplementary Table S3). All modifications are cytidine (C) to a uridine (U), at the first (27,6%) or second (72,4%) codon positions. Most editions changed the amino acid polarity (22 of 29) from polar to apolar (62,1%) or from apolar to polar (13,8%). Seven RNA-editing sites did not change amino acid polarity (three polar-polar and four apolar-apolar).

From a total of 29 RNA-editing sites, 11 are shared with all species of the subfamily Cactoideae analyzed here [i.e., *atpA* (424), *atpB* (138), *petB* (140), *psbL* (1), *rpoB* (158),


a
b
accD
ycf1
clpP
rpl22
rpl32
rps19
infA
rpoC1
rps11
rpl33
ycf2
rps15
rpl20
rps3
psbM
rps7
rpl16
rpl36
rps2
rps8
atpE
rpl14
matK
ccsA
rpoC2
rps18
rps16
rpoA
rpl2
rps4
ycf4
rpoB
rps12
rps14
cemA
psbJ
petG
atpA
petL
psaJ
atpF
atpB
rbcL
psaI
petD
psbH
atpI
psbI
psbE
petA
petB
psbB
psaB
psbN
psaA
psaC
ycf3
psbA
psbC
psbK
atpH
psbF
psbZ
psbT
psbD
psbL
petN
Genes
0,0 0,2 0,4 0,6 0,8 1,0 1,2 1,4 1,6 1,8
Branch length
0 10 20 30 40 50 120 130 140 150 160 170
Number of putative sites under positive selection

◂**Fig. 5** Molecular evolution analyses of plastid genes within the family Cactaceae. **a** Divergence of protein-coding genes. The gene divergence was estimated by the sum of total branch lengths in each gene tree inferred (mean ± SD). **b** Number of putative sites under positive selection

rpoB (189), *rpoC1* (11), *rpoC1* (209), *rpoC2* (488), *rps14* (27), and *rps14* (50)]. Three sites *atpA* (305), *rpl20* (87), and *rps2* (83) were found in almost all species of the subfamily Cactoideae, except in one taxon. Aiming at understanding the evolution of polymorphic sites, we traced RNA-editing sites of *M. glaucescens* with the phylogeny of the subfamily Cactoideae (Fig. 7). Two sites, *psbF* (26) and *rpoA* (28), are shared among closely related species of the Core Cactoideae. The site *rpoB* (184) is shared among Core Cactoideae, *M. albiflora*, and *M. zephyranthoides*. The site *rpoA* (295) is only shared among *M. glaucescens*, *M. albiflora*, and *M. zephyranthoides*. On the other hand, *M. glaucescens* does not contain two common sites found in species of the family Cactoideae species: *rpl20* (24) and *rpl20* (91). These exceptions are due to the mutations in plastid DNA, which fix a T, dismissing the need for RNA editing.

We identified here 11 possible specific sites in seven genes of *M. glaucescens*, *atpF*(31), *rpl20*(100), *rpoB*(358), *rpoC1*(390), *rpoC1*(420), *rpoC2*(725), *rpoC2*(739), *rpoC2*(1142), *rpoC2*(1311), *rps8*(48), and *ycf3*(26); Fig. 7 and Supplementary Table S3). In addition, other four possible specific RNA-editing sites were identified in the subfamily Cactoideae, being two in *R. baccifera* [*rpoA* (9) and *rpoA* (188)], and two in *S. undatus*, [*rpoC1* (321) and *rpoC2* (163)]. Another two sites are specific to the genus *Rhipsalis*: *rpoC1* (157) and *rpoC2* (606). Finally, the site *rpoC2* (770) is shared among some species of Core Cactoideae I (*S. undatus*, *C. gigantea*, and *L. schottii*) (Supplementary Table S3).

Identification of SSRs in *Melocactus glaucescens* plastome

The plastome of *M. glaucescens* contains 169 SSR loci, of which 82 are localized in intergenic spacers (IGSs), 59 in coding sequences (CDS), and 28 in introns. The SSRs localized in CDS were found with high frequency in the following genes: *ycf1* (11), *rpoC2* (6), *ycf2* (5), *rpoB* (3), and *matK* (3). The 28 SSRs located in introns are distributed on nine genes [e.g., *trnL-UAA* (6), *atpF* (5), and *ycf3* (4) genes]. Concerning the IGS, *trnS-UGA*/*psbZ*, *trnQ-UUG* /*cemA*, *rps4*/*trnT-UGU*, and *psaC*/*rpl32* showed three SSRs each. The most frequent type of SSRs are mono- (117) and dinucleotides (41), the others are tri- (2), tetra- (6), penta- (2), and hexapolymers (1). The bases A/T compose most of mono- and dinucleotide SSRs, with 97,4% and 63,4%, respectively (Supplementary Table S4). The size, sequence, and location of the SSRs identified in the *M. glaucescens* plastome are shown in Supplementary Table S5.

Discussion

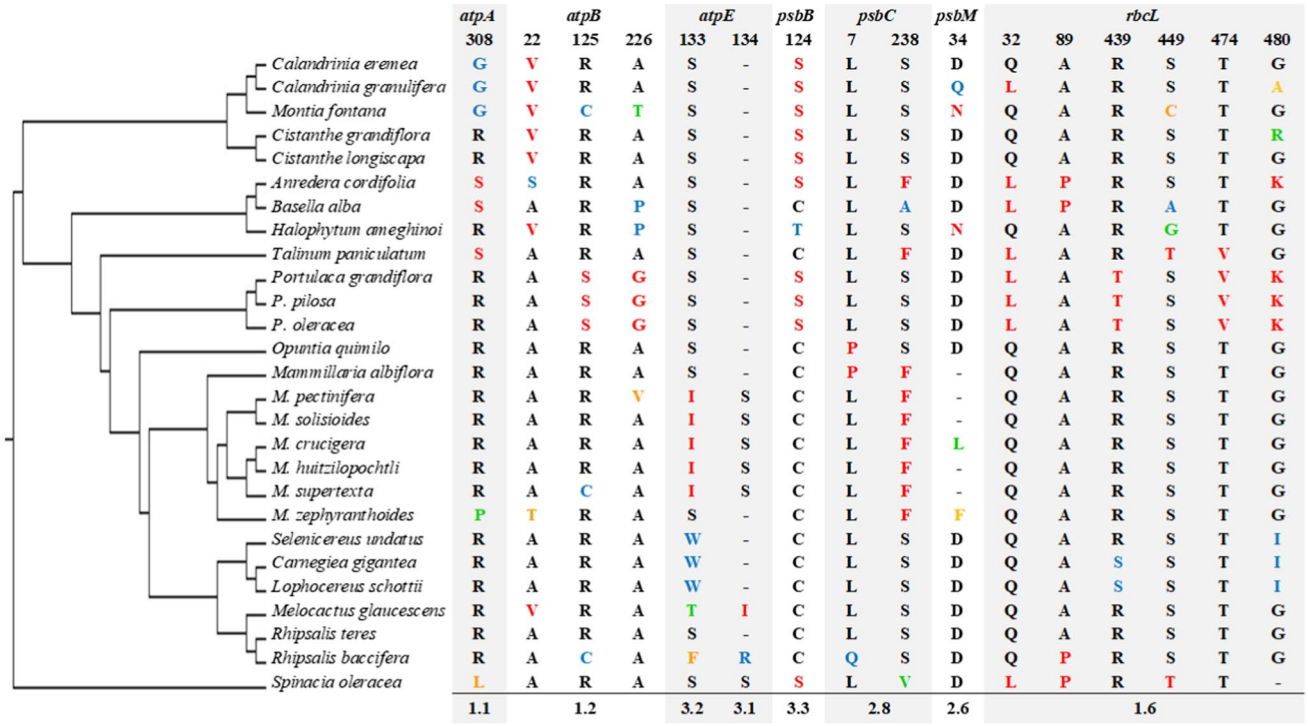
Unusual rearrangements, IR expansion, and massive gene losses characterize the *M. glaucescens* plastome

Usually, the plastome of angiosperms is a circular molecule containing a conserved structure and gene content (Bock et al. 2007). However, recent studies have revealed several plant lineages with atypical features (i.e., reduction of plastome size, unusual rearrangements, expansions and contractions of IRs, and losses and/or pseudogenization of several genes) (Chumley et al. 2006; Guisinger et al. 2011; Lee et al. 2017; Lopes et al. 2018b, 2019; Pacheco et al. 2020a, b), which include the family Cactaceae (Sanderson et al. 2015; Solórzano et al. 2019; Oulo et al. 2020; Silva et al. 2021).

Several plastome configurations have been frequently observed in plastomes of the subfamily Cactoideae, which are related to the SSC region and IRs (Sanderson et al. 2015; Solórzano et al. 2019; Silva et al. 2021). An outstanding example is related to the ribosomal RNA genes, which are commonly located in the IRs of plastomes of angiosperms (Lopes et al. 2018b,c, 2021; Pacheco et al. 2019, 2020a), including the cactus *Opuntia quimilo* of the subfamily Opuntioideae (Köhler et al. 2020). However, in *M. glaucescens*, *S. undatus*, and other cacti of the genera *Mammilaria* and *Rhipsalis*, they are located in the SSC, which appears to be a common feature found in species of the subfamily Cactoideae (Solórzano et al. 2019; Oulo et al. 2020; Silva et al. 2021).

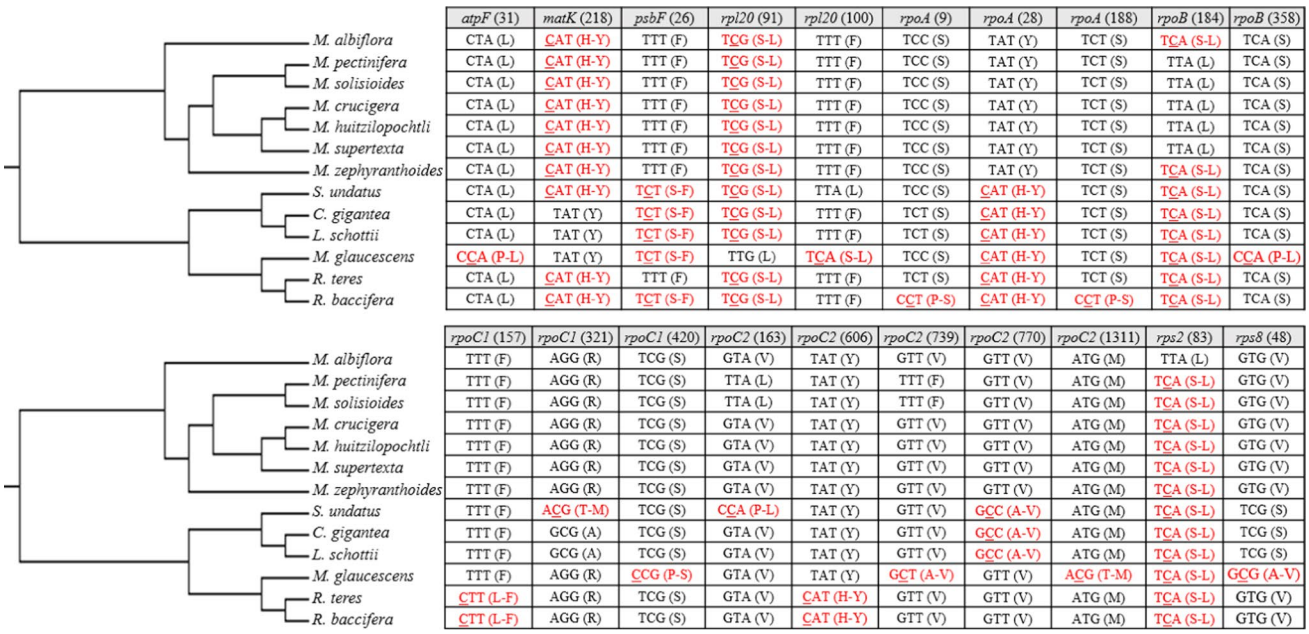
Nevertheless, the *M. glaucescens* plastome underwent intense and unprecedented structural reorganization, which does not follow either of the two rearrangement patterns (Tribe Cacteae and Core Cactoideae) previously observed in the subfamily Cactoideae (Silva et al. 2021). The translocation and inversion of a large segment of the LSC region and other subsequent events (e.g., the expansion of the IRs) resulted in the duplication of genes commonly located in the LSC. Thus, the plastome of *M. glaucescens* evolved under selective pressure to maintain these genes in the IRs. However, the reasons for such plastome rearrangement are still unclear.

Another feature of *M. glaucescens* plastome is a massive loss of genes, which totalizes 15 genes that were lost or are not functional (i.e., pseudogenes). Out of this total, 11 genes of the *ndh* complex were lost, indicating a complete loss of its functionality in *M. glaucescens*. The *ndh* complex acts in the cyclical electron transport and production of ATP around Photosystem I (PSI) under stress conditions (Rumeau et al.



	atpA	*atpB*			*atpE*		*psbB*	*psbC*		*psbM*	*rbcL*					
	308	22	125	226	133	134	124	7	238	34	32	89	439	449	474	480
Calandrinia eremea	G	V	R	A	S	-	S	L	S	D	Q	A	R	S	T	G
Calandrinia granulifera	G	V	R	A	S	-	S	L	S	Q	L	A	R	S	T	A
Montia fontana	G	V	C	T	S	-	S	L	S	N	Q	A	R	C	T	G
Cistanthe grandiflora	R	V	R	A	S	-	S	L	S	D	Q	A	R	S	T	R
Cistanthe longiscapa	R	V	R	A	S	-	S	L	S	D	Q	A	R	S	T	G
Anredera cordifolia	S	S	R	A	S	-	S	L	F	D	L	P	R	S	T	K
Basella alba	S	A	R	P	S	-	C	L	A	D	L	P	R	A	T	G
Halophytum ameghinoi	R	V	R	P	S	-	T	L	S	N	Q	A	R	G	T	G
Talinum paniculatum	S	A	R	A	S	-	C	L	F	D	L	A	R	T	V	G
Portulaca grandiflora	R	A	S	G	S	-	S	L	S	D	L	A	T	S	V	K
P. pilosa	R	A	S	G	S	-	S	L	S	D	L	A	T	S	V	K
P. oleracea	R	A	S	G	S	-	S	L	S	D	L	A	T	S	V	K
Opuntia quimilo	R	A	R	A	S	-	C	P	S	D	Q	A	R	S	T	G
Mammillaria albiflora	R	A	R	A	S	-	C	P	F	-	Q	A	R	S	T	G
M. pectinifera	R	A	R	V	I	S	C	L	F	-	Q	A	R	S	T	G
M. solisioides	R	A	R	A	I	S	C	L	F	-	Q	A	R	S	T	G
M. crucigera	R	A	R	A	I	S	C	L	F	L	Q	A	R	S	T	G
M. huitzilopochtli	R	A	R	A	I	S	C	L	F	-	Q	A	R	S	T	G
M. supertexta	R	A	C	A	I	S	C	L	F	-	Q	A	R	S	T	G
M. zephyranthoides	P	T	R	A	S	-	C	L	F	F	Q	A	R	S	T	G
Selenicereus undatus	R	A	R	A	W	-	C	L	S	D	Q	A	R	S	T	I
Carnegiea gigantea	R	A	R	A	W	-	C	L	S	D	Q	A	S	S	T	I
Lophocereus schottii	R	A	R	A	W	-	C	L	S	D	Q	A	S	S	T	I
Melocactus glaucescens	R	V	R	A	T	I	C	L	S	D	Q	A	R	S	T	G
Rhipsalis teres	R	A	R	A	S	-	C	L	S	D	Q	A	R	S	T	G
Rhipsalis baccifera	R	A	C	A	F	R	C	Q	S	D	Q	P	R	S	T	G
Spinacia oleracea	L	A	R	A	S	S	S	L	V	D	L	P	R	T	T	-
	1.1	1.2			3.2	3.1	3.3	2.8		2.6	1.6					

Fig. 6 Non-synonymous substitutions in the *atpA*, *atpB, atpE*, *psbB*, *psbC*, *psbM*, and *rbcL* genes. The amino acids are plotted across the Cactineae phylogeny inferred based on plastid genes (Fig. 4). Different amino acid types identified at the same position are highlighted in distinct colors. The amino acid positions are relative to *M. glaucescens* plastid genes. The Ka/Ks scores for each codon position are indicated in the last line



	atpF (31)	*matK* (218)	*psbF* (26)	*rpl20* (91)	*rpl20* (100)	*rpoA* (9)	*rpoA* (28)	*rpoA* (188)	*rpoB* (184)	*rpoB* (358)
M. albiflora	CTA (L)	CAT (H-Y)	TTT (F)	TCG (S-L)	TTT (F)	TCC (S)	TAT (Y)	TCT (S)	TCA (S-L)	TCA (S)
M. pectinifera	CTA (L)	CAT (H-Y)	TTT (F)	TCG (S-L)	TTT (F)	TCC (S)	TAT (Y)	TCT (S)	TTA (L)	TCA (S)
M. solisioides	CTA (L)	CAT (H-Y)	TTT (F)	TCG (S-L)	TTT (F)	TCC (S)	TAT (Y)	TCT (S)	TTA (L)	TCA (S)
M. crucigera	CTA (L)	CAT (H-Y)	TTT (F)	TCG (S-L)	TTT (F)	TCC (S)	TAT (Y)	TCT (S)	TTA (L)	TCA (S)
M. huitzilopochtli	CTA (L)	CAT (H-Y)	TTT (F)	TCG (S-L)	TTT (F)	TCC (S)	TAT (Y)	TCT (S)	TTA (L)	TCA (S)
M. supertexta	CTA (L)	CAT (H-Y)	TTT (F)	TCG (S-L)	TTT (F)	TCC (S)	TAT (Y)	TCT (S)	TTA (L)	TCA (S)
M. zephyranthoides	CTA (L)	CAT (H-Y)	TTT (F)	TCG (S-L)	TTT (F)	TCC (S)	TAT (Y)	TCT (S)	TCA (S-L)	TCA (S)
S. undatus	CTA (L)	CAT (H-Y)	TCT (S-F)	TCG (S-L)	TTA (L)	TCC (S)	CAT (H-Y)	TCT (S)	TCA (S-L)	TCA (S)
C. gigantea	CTA (L)	TAT (Y)	TCT (S-F)	TCG (S-L)	TTT (F)	TCT (S)	CAT (H-Y)	TCT (S)	TCA (S-L)	TCA (S)
L. schottii	CTA (L)	TAT (Y)	TCT (S-F)	TCG (S-L)	TTT (F)	TCT (S)	CAT (H-Y)	TCT (S)	TCA (S-L)	TCA (S)
M. glaucescens	CCA (P-L)	TAT (Y)	TCT (S-F)	TTG (L)	TCA (S-L)	TCC (S)	CAT (H-Y)	TCT (S)	TCA (S-L)	CCA (P-L)
R. teres	CTA (L)	CAT (H-Y)	TTT (F)	TCG (S-L)	TTT (F)	TCT (S)	CAT (H-Y)	TCT (S)	TCA (S-L)	TCA (S)
R. baccifera	CTA (L)	CAT (H-Y)	TCT (S-F)	TCG (S-L)	TTT (F)	CCT (P-S)	CAT (H-Y)	CCT (P-S)	TCA (S-L)	TCA (S)

	rpoC1 (157)	*rpoC1* (321)	*rpoC1* (420)	*rpoC2* (163)	*rpoC2* (606)	*rpoC2* (739)	*rpoC2* (770)	*rpoC2* (1311)	*rps2* (83)	*rps8* (48)
M. albiflora	TTT (F)	AGG (R)	TCG (S)	GTA (V)	TAT (Y)	GTT (V)	GTT (V)	ATG (M)	TTA (L)	GTG (V)
M. pectinifera	TTT (F)	AGG (R)	TCG (S)	TTA (L)	TAT (Y)	TTT (F)	GTT (V)	ATG (M)	TCA (S-L)	GTG (V)
M. solisioides	TTT (F)	AGG (R)	TCG (S)	TTA (L)	TAT (Y)	TTT (F)	GTT (V)	ATG (M)	TCA (S-L)	GTG (V)
M. crucigera	TTT (F)	AGG (R)	TCG (S)	GTA (V)	TAT (Y)	GTT (V)	GTT (V)	ATG (M)	TCA (S-L)	GTG (V)
M. huitzilopochtli	TTT (F)	AGG (R)	TCG (S)	GTA (V)	TAT (Y)	GTT (V)	GTT (V)	ATG (M)	TCA (S-L)	GTG (V)
M. supertexta	TTT (F)	AGG (R)	TCG (S)	GTA (V)	TAT (Y)	GTT (V)	GTT (V)	ATG (M)	TCA (S-L)	GTG (V)
M. zephyranthoides	TTT (F)	AGG (R)	TCG (S)	GTA (V)	TAT (Y)	GTT (V)	GTT (V)	ATG (M)	TCA (S-L)	GTG (V)
S. undatus	TTT (F)	ACG (T-M)	TCG (S)	CCA (P-L)	TAT (Y)	GTT (V)	GCC (A-V)	ATG (M)	TCA (S-L)	TCG (S)
C. gigantea	TTT (F)	GCG (A)	TCG (S)	GTA (V)	TAT (Y)	GTT (V)	GCC (A-V)	ATG (M)	TCA (S-L)	TCG (S)
L. schottii	TTT (F)	GCG (A)	TCG (S)	GTA (V)	TAT (Y)	GTT (V)	GCC (A-V)	ATG (M)	TCA (S-L)	TCG (S)
M. glaucescens	TTT (F)	AGG (R)	CCG (P-S)	GTA (V)	TAT (Y)	GCT (A-V)	GTT (V)	ACG (T-M)	TCA (S-L)	GCG (A-V)
R. teres	CTT (L-F)	AGG (R)	TCG (S)	GTA (V)	CAT (H-Y)	GTT (V)	GTT (V)	ATG (M)	TCA (S-L)	GTG (V)
R. baccifera	CTT (L-F)	AGG (R)	TCG (S)	GTA (V)	CAT (H-Y)	GTT (V)	GTT (V)	ATG (M)	TCA (S-L)	GTG (V)

Fig. 7 Distribution of putative RNA-editing sites across the family Cactaceae phylogeny based on concatenated plastid genes. The codons highlighted in red have an editing site (underlined nucleotide)

2007; Strand et al. 2019). Several plant lineages have lost their *ndh* genes, including gymnosperms and angiosperms (Braukmann and Stefanović 2009; Blazier et al. 2011; Kim e Chase 2017; Lin et al. 2017; Fu et al. 2021). Although the reasons for the loss of *ndh* genes are unknown, it seems to be a common event in the subfamily Cactoideae (Silva et al. 2021). Likewise, the *rpl23* gene was lost from the *M. glaucescens* plastome, while in the other species of the subfamily Cactoideae, it is a pseudogene (Silva et al. 2021). The loss or pseudogenization of *rpl23* appears to be recurrent in Caryophyllales (Ramam and Park 2015). This gene encodes the L23 protein, an essential component of the 50S ribosome subunit (Fleischmann et al. 2011). Therefore, it is reasonable to assume that a similar L23 protein is nucleus-encoded and it is imported by the chloroplast from the cytosol as observed in other lineages (Lopes et al. 2018b).

On the other hand, the *rpl33* gene is functional in *M. glaucescens*, but it was lost or is a pseudogene in the other species of the subfamily Cactoideae (i.e., genus *Mammilaria*, genus *Rhipsalis*, *C. gigantea*, *L. schottii*, and *S. undatus*; Sanderson et al. 2015; Solórzano et al. 2019; Oulo et al. 2020; Silva et al. 2021). The *rpl33* gene encodes the L33 protein of the ribosome 50S subunit, which is required for proteins synthesis during the early stages of plant growth and development, especially during the formation of thylakoid membranes (Rogalski et al. 2008b). *M. glaucescens* is a slow-growing species, so the retention of this gene indicates that it is important for the maintenance of fitness and adaptation to its geographical occurrence and specific environmental condition.

Loss of essential tRNAs from the *M. glaucescens* plastome indicates tRNA import from cytosol

According to the wobbling rules proposed by Crick (1966), 32 tRNAs are needed to read all existing codons. However, plastomes of angiosperms harbor normally 30 tRNAs, which by nucleoside modification in the wobbling position and supperwobbling suffice for the reading of all codons (Rogalski et al. 2008a; Agris et al. 2007; Weixlbaumer et al. 2007; Kurata et al. 2008; Alkatib et al. 2012a, b; Mandal et al. 2014). Although it is an uncommon feature, some lineages of higher plants show less than 30 tRNAs in plastomes (Chumley et al. 2006; Guisinger et al. 2011; Lee et al. 2017). However, the loss of an essential tRNA (i.e., containing uridine in the wobbling position; *trnV*-UAC) has been reported recently in species of the subfamily Cactoidae (Silva et al. 2021). This kind of loss would require another mechanism to replace the essential tRNA for translation and cell viability, since plastid translation is essential for cell viability in most angiosperms (Rogalski et al. 2006, 2008a; b; Fleischmann et al. 2011; Alkatib et al. 2012b). A mechanism that would replace the loss of an essential tRNA is the tRNA import from the cytosol to the plastids, but this mechanism was still not demonstrated experimentally in plastids or even evidenced via genetic reverse in plastids (Legen et al. 2007; Rogalski et al. 2008a; Alkatib et al. 2012a, b; Agrawal et al. 2020). This is known that this mechanism is used in mitochondria to replace the absence of various tRNAs (Akashi et al. 1998; Dietrich et al. 1996; Alfonzo and Söll 2009; Salinas-Giegé et al. 2015; Murcha et al. 2016).

Here, we reported the loss of three tRNA genes (*trnA-UGC*, *trnV-GAC*, and *trnV-UAC*) in the plastome of *M. glaucescens*. As previously reported by Silva et al. (2021), the loss of the *trnV-UAC* gene appears to be a common feature of the subfamily Cactoideae. Surprisingly, *M. glaucescens* lost the two tRNA genes, *trnV-GAC*, and *trnV-UAC*, for the amino acid valine, both tRNAs are normally responsible to decode the four valine codons (i.e., GUA, GUC, GUG, and GUU) using the wobbling rules (Crick 1966). It means that plastid translation must operate with tRNAs presumably imported from the cytosol given that they are not encoded by plastids. The reverse genetics in plastids demonstrated that the *trnV-GAC* gene is dispensable, whereas the *trnV-UAC* is essential for cell viability (Alkatib et al. 2012b). More surprisingly, the *trnA-UGC* is also absent from the plastome of *M. glaucescens*. The $tRNA^{Ala}$(UGC) is responsible for the decoding of four alanine codons (i.e., GCA, GCC, GCG, and GCU), which can be done with high efficiency by uridine, containing an unknown modification, in the wobbling position (Pfitzinger et al. 1990; Alkatib et al. 2012b).

Given the total absence of valine and alanine tRNAs in the plastome of *M. glaucescens*, we analyzed codon usage in protein-coding genes of *M. glaucescens*, other cacti, and related species (i.e., spinach as a species of order Caryophyllales and model species for plastid studies). The codon usage analysis demonstrated that codon frequencies for the amino acids alanine and valine are very similar to frequencies found in the other species. Therefore, our data and results strongly suggest the import of tRNAs, at least $tRNA^{Ala}$(UGC) and $tRNA^{Val}$(UAC) according to decoding rules, from the cytosol to the plastids in *M. glaucescens*.

Suborder Cactineae phylogenetic reconstruction based on plastid genes

Our phylogenetic tree based on plastid protein-coding genes was similar to that reported by Silva et al. (2021). The same pattern of distribution of species within the families Montiaceae, Basellaceae, Portulacaceae, and subfamily Cactoideae was observed in both phylogenetic trees. However, some inconsistencies related to the families Halophytaceae and Talinaceae were verified. Our results demonstrated, with low support, that *H. ameghinoi* forms a sister group to the family Basellaceae, which differs from the results obtained by Silva et al. (2021). On the other hand, *Talinum paniculatam*

forms a sister group with the family Portulacaceae in the phylogenetic tree proposed by Silva et al. (2021), although this relationship is poorly supported. In our tree, the clade Portulacaceae is highly supported to be a sister group to the family Cactaceae. Complete plastome sequences from other species of families Halophytaceae and Talinaceae will resolve such inconsistencies.

Opuntia quimilo (Opuntioideae) constituted the early branching of the family Cactaceae and formed a sister group to the subfamily Cactoideae. Two clades compose the subfamily Cactoideae, one of genus *Mammillaria* and the other formed by the Core Cactoideae, corroborating with the phylogeny proposed by Guerrero et al. (2019). In our phylogenetic analysis, all species of the genus *Mammillaria* were grouped as previously reported by Solórzano et al. (2019) and Silva et al. (2021). However, a more comprehensive phylogeny of the clade Mammilloid was recently performed by Breslin et al. (2021), based on the LSC region of plastomes, which revealed that *Mammillaria* is no longer considered monophyletic. Probably, the limited number of plastomes of this genus available in the database and the proximity of the *Mammillaria* species analyzed here resulted in the formation of the clade Mammillaria. Finally, two groups form the Core Cactoideae clade. The first is composed of *C. gigantea* and *L. schottii* (a sister group of the tribe Echinocereeae) and *S. undatus* (tribe Hylocereeae), representing North American columnar cacti. The second group is formed by *M. glaucescens* (Cereeae), and a sister group composed of *R. teres* and *R. baccifera* (Rhipsalideae). These clades follow the phylogenetic inference reported by Guerrero et al. (2019), which are called Core Cactoideae I and Core Cactoideae II, respectively.

Molecular evolution of plastid protein-coding genes of *M. glaucescens*

The family Cactaceae shows a high divergence in several protein-coding genes. The genes *accD*, *ycf1*, *clpP*, *rpl22*, *rpl32*, and *rps19* exhibited mean values for branch length above 0.2. As previously reported by Silva et al. (2021), the *accD* is the most divergent gene from the family Cactaceae. This gene encodes the β-carboxyl-transferase subunit of the acetyl-CoA carboxylase, which is related to plastid fatty acid biosynthesis (Rogalski and Carrer 2011; Salie and Thelen 2016). Although it is divergent, the C-terminus, where is located the carboxyl-transferase domain, remains conserved (Greiner et al. 2008b; Silva et al. 2021). The *ycf1* and *ycf2* genes are involved in the protein import machinery in plastids, being essential for cell survival (Drescher et al. 2000; Kikuchi et al. 2013, 2018). However, the high divergence of both genes has been reported in several plastomes of angiosperms, as well as events of pseudogenization and gene loss (Vries et al. 2015; Lopes et al. 2018b, 2021), including cacti (Solárzano et al. 2019; Köhler et al. 2020). The *clpP* gene is also an essential gene for cell viability (Shikanai et al. 2001; Kuroda and Maliga 2003), whereas some plant lineages show highly divergent sequences (Erixan and Oxelman 2008; Rockenbach et al. 2016; Pacheco et al. 2020a, b).

To obtain more information about the evolution of the plastid genes in Cactaceae, we next analyzed the presence of positive selection in each protein-coding gene. We identified a total of 450 putative positive signatures in almost half of the genes analyzed (32 out of 67), including genes related to gene expression, photosynthesis, fatty acid biosynthesis, cytochrome, and others. According to our analyses, most divergent genes also contain a high number of putative positive signatures. The presence of positive signatures was also characterized in other plant lineages, but the reason for such signatures and their effects are still poorly understood (Lopes et al. 2018a, 2021). Curiously, we observed in our analyses a correlation between gene divergence and positive signatures.

The *ycf1* and *ycf2* genes harbor more than half of all putative positive signatures identified in our analysis (303 signatures), which are distributed throughout the protein. In the *accD* gene, we identified five putative positive signatures, none of them located in the C-terminal region. Although the *accD* gene has an outstanding importance for the plant cell (Kode et al. 2005), accelerated rates of evolution and even functional transfer to the nucleus have been previously reported (Rousseau-Gueutin et al. 2013; Rockenbach et al. 2016). The sequence of *clpP* gene revealed 20 putative sites of positive signatures in our analysis. According to the Conserved Domains Database (https://www.ncbi.nlm.nih.gov/cdd), only one of them, *clpP* (153), is located at the protein–protein interaction interface and represents a wide change of amino acid among the species of Cactaceae. The conserved amino acid (lysine) was changed to six different amino acids among the species. Oligomerization is essential to build a functional protein complex, wherefore it is likely that the nuclear genes (i.e., other subunits) and plastid *clpP* gene are undergoing a process of co-evolution in Cactaceae (Greiner and Bock 2013; Zhang et al. 2015).

The genes involved in plastid translation also acquired several sites of putative positive signatures in the family Cactaceae. As demonstrated in the previous studies, accelerated evolutionary rates and positive selection frequently occur in plastid genes involved in gene expression (Krawczyk and Sawicki 2013; Hu et al. 2015; Xu et al. 2015; Weng et al. 2016). Four genes (*rpl20*, *rps2*, *rps7*, and *rps12*) encoding plastid ribosomal proteins have signatures at sites of interaction with ribosomal proteins and ribosomal RNA. The L20 (i.e., amino acid position 100) is a binding site interacting with protein L21 in the 50S ribosomal subunit. In our analysis, this site is under putative positive selection and contains six different amino acids considering the different species

of Cactaceae. Three sites that interact directly with 16S rRNA, located in the ribosomal proteins S2 (39), S7(94), and S12(31), are under putative positive selection (Wimberly et al. 2000; Schuwirth et al. 2005; Krawczyk and Sawicki 2013; Hu et al. 2015; Xu et al. 2015; Weng et al. 2016). Our analysis revealed that genes from several functional classes are under putative positive selection. The consequences of these signatures on protein function remain unclear and have to be investigated functionally. However, they suggest that natural selection acts on different sites, possibly bringing adaptation in each species to specific environmental conditions (Greiner and Bock 2013; Hu et al. 2015; Lopes et al. 2021).

The *M. glaucescens* plastome shows possible unique RNA-editing sites

The RNA editing is a post-translational mechanism of nucleotide correction, which changes the bases from C to U or more rarely from U to C. Plastid RNA editing occurs in most land plants to regulate gene expression under different stress conditions or adaptation needs (Takenaka et al. 2013). In this study, we identified 29 putative RNA-editing sites in 15 genes of *M. glaucescens* plastome using the PREP program.

Here, all RNA-editing sites were identified at the first or second positions of the codons and C-to-U conversions as observed in other studies (Lopes et al. 2018a,c; Silva et al. 2021). Of the total of 29 RNA-editing sites, 18 of them were also identified in other species of subfamily Cactoideae (Silva et al. 2021). However, we found new sites for this subfamily, being two in *S. undatus* [*rpoC1* (321) and *rpoC2* (163)], two in *R. baccifera* [*rpoA*(9) and *rpoA*(188)] and 11 in *M. glaucescens.*

M. glaucescens revealed 11 possible specific RNA-editing sites distributed among the genes *atpF*(31), *rpl20*(100), *rpoB*(358), *rpoC1*(390, 420), *rpoC2*(725, 739, 1142, and 1311), *rps8* (48), and *ycf3* (26). This is the first species of the tribe Cereeae to have the complete plastome widely analyzed. The sequencing and publication of other plastomes will enable us to infer whether these putative sites are unique to *M. glaucescens* or occur at the level of genus or tribe.

The high number of SSRs identified in the *M. glaucescens* plastome represents an important source of molecular markers aiming at conservation strategies

A total of 169 SSR loci were identified in the plastome of *M. glaucescens* (Supplementary Table S4 and S5), being dominant occurrence in non-coding regions (IGS and introns). These regions evolve faster than coding regions and thereby are more interesting to study the genetic variability of populations (Rogalski et al. 2015). The plastid molecular markers show several interesting features for various genetic studies such as high polymorphism, non-recombinant, and uniparental inheritance in most angiosperms (Wheeler et al. 2014). More than 93% of SSRs found here are characterized as mono and dinucleotides, constituted by A and T. Only three SSRs mononucleotides G/C were observed in this plastome. The location and the type of SSRs identified are congruent with other studies (Pacheco et al. 2019, 2020c; Lopes et al. 2021; Silva et al. 2021).

The family Cactaceae is one of the most threatened taxonomic groups, with more than 30% of species at risk of extinction (Goettsch et al. 2015). *M. glaucescens* is among the endangered species of cacti, due to a high degree of endemism, reduced population, slow growth, and anthropogenic action, either by the indiscriminate collection of specimens or by habitat destruction (Taylor and Zappy 2004; Machado et al. 2009). Therefore, the full plastome sequence and SSRs characterized in this study are useful information for investigation in natural populations of *M. glaucescens* and assist in conservation strategies for this threatened species.

Conclusions

Here, we report the complete plastome of *M. glaucescens*, an endemic cactus to a small region of the State of Bahia (Brazil) and highly endangered. The plastome of *M. glaucescens* contains unique structural features within the subfamily Cactoideae. Several rearrangement events culminated in the expansion of IRA/LSC border, the largest IRs found in Cactoidae, and the duplication of 22 genes commonly located in the LSC. Concerning gene divergence, we identified high divergent genes, a high number of putative positive signatures in the plastid protein-coding genes, and possible unique RNA-editing sites. Additionally, all *ndh* genes were lost or became pseudogenes. Moreover, the *trnA-UGC*, *trnV-UAC* , and *trnV-GAC* genes are absent from the *M. glaucescens* plastome. The absence of both tRNAs suggests the mechanism of tRNA import from the cytosol to the plastids in this plant lineage, since they are essential genes for plastid translation and cell viability. Furthermore, we characterize in the plastome of *M. glaucescens* plastome several SSRs, which represent an extremely useful tool for conservation strategies in natural populations of endangered species. Finally, we also emphasize the importance of environmental conservation since different and specific habitats represent hotspots of endemic species with risk of extinction and importance for understanding the evolution of plastomes, plant cells, and species.

Author contribution statement TPDC, MCS, ASL, and MR conceived and designed the research. TPDC, MCS, ASL, TGP, JDO, EB, VAB, EB, and MR conducted experiments

and analyzed the data. ASL, TGP, EMS, FOP, and MR contributed with reagents and materials. TPDC and MR wrote the manuscript. All authors read and approved the manuscript.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00425-022-03841-2.

Acknowledgements This research was supported by the National Council for Scientific and Technological Development, Brazil (CNPq—Grants 459698/2014-1, 310654/2018-1, and 436407/2018-3). We are grateful to INCT-FBN and for the scholarships granted by the CNPq to TPDC, JDO, FOP, EMS, and MR. The research fellowship granted by Minas Gerais Research Foundation (FAPEMIG) to MCS is also gratefully acknowledged. We also thank Prof. Wagner Campos Otoni (UFV) and Dr. Gabriela Torres da Silva (UFV) for providing the plants (*M. glaucescens*) used in this study.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Agrawal S, Karcher D, Ruf S, Bock R (2020) The functions of chloroplast Glutamyl-tRNA in translation and tetrapyrrole biosynthesis. Plant Physiol 183:263–276. https://doi.org/10.1104/pp.20.00009

Agris PF, Vendeix FA, Graham WD (2007) tRNA's wobble decoding of the genome: 40 years of modification. J Mol Biol 366(1):1–13. https://doi.org/10.1016/j.jmb.2006.11.046

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 Res 26(9):2168–2172. https://doi.org/10.1093/nar/26.9.2168

Albuquerque UP, Medeiros PM, Almeida ALS, Monteiro JM, Lins Neto EMF, Melo JG, Santos JP (2007) Medicinal plants of the caatinga (semi-arid) vegetation of NE Brazil: a quantitative approach. J Ethnopharmacol 114:325–354. https://doi.org/10.1016/j.jep.2007.08.017

Alfonzo JD, Söll D (2009) Mitochondrial tRNA import—the challenge to understand has just begun. Biol Chem 390:717–722. https://doi.org/10.1515/BC.2009.101

Alkatib S, Fleischmann TT, Scharff LB, Bock R (2012a) Evolutionary constraints on the plastid tRNA set decoding methionine and isoleucine. Nucleic Acids Res 40(14):6713–6724. https://doi.org/10.1093/nar/gks350

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(11):e1003076. https://doi.org/10.1371/journal.pgen.1003076

Aquino-Martins VGdQ, Melo LFMd, Silva LMP, Targino de Lima TR, Fernandes Queiroz M, Viana RLS, Zucolotto SM, Andrade VS, Rocha HAO, Scortecci KC (2019) In vitro antioxidant, antibiofilm, and solar protection activities of *Melocactus zehntneri* (Britton & Rose) pulp extract. Antioxidants 8(10):439. https://doi.org/10.3390/antiox8100439

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. PNAS 108:8379–8384. https://doi.org/10.1073/pnas.1100628108

Blazier CJ, Guisinger MM, Jansen RK (2011) Recent loss of plastid-encoded ndh genes within *Erodium* (Geraniaceae). Plant Mol Biol 76:263–272. https://doi.org/10.1007/s11103-011-9753-5

Bock R (2007) Structure, function, and inheritance of plastid genomes. In: Bock R (ed) Cell and molecular biology of plastids. Springer, Berlin, pp 29–63

Brandão GHA, Rigo G, Roque AA, Souza ACD, Scopel M, Nascimento CAO, Tasca T, Pereira CG, Giordani RB (2017) Extraction of bioactive alkaloids from *Melocactus zehntneri* using supercritical fluid. J Supercrit 129:28–35. https://doi.org/10.1016/j.supflu.2016.12.012

Braukmann TWA, Kuzmina M, Stefanović S (2009) Loss of all plastid *ndh* genes in *Gnetales* and conifers: extent and evolutionary significance for the seed plant phylogeny. Curr Genet 55:323–337. https://doi.org/10.1007/s00294-009-0249-7

Breslin PB, Wojciechowski MF, Majure LC (2021) Molecular phylogeny of the Mammilloid clade (Cactaceae) resolves the monophyly of Mammillaria. Taxon 70(2):308–323. https://doi.org/10.1002/tax.12451

Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK (2006) The complete chloroplast genome sequence of *Pelargonium* x *hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol 23(11):2175–2190. https://doi.org/10.1093/molbev/msl089

Crick FHC (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 17:134. https://doi.org/10.1186/s13059-016-1004-2

Darling ACE, Mau B, Blattner FR, Perna NT (2004) Mauve: Multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403. https://doi.org/10.1101/gr.2289704

Dierckxsens N, Mardulyn P, Smits G (2016) NOVOPlasty: De novo assembly of organelle genomes from whole genome data. Nucleic Acids Res 45:e18. https://doi.org/10.1093/nar/gkw955

Dietrich A, Small I, Cosset A, Weil JH, Maréchal-Drouard L (1996) Editing and import: strategies for providing plant mitochondria with a complete set of functional transfer RNAs. Biochimie 78:518–529. https://doi.org/10.1016/0300-9084(96)84758-4

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. https://doi.org/10.1046/j.1365-313x.2000.00722.x

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340

Erixon P, Oxelman B (2008) Whole-gene positive selection, elevated synonymous substitution rates, duplication, and indel evolution of the chloroplast *clpP1* gene. PLoS ONE 3(1):e1386. https://doi.org/10.1371/journal.pone.0001386

Fleischmann TT, Scharff LB, Alkatib S, Hasdorf S, Schöttler MA, Bock R (2011) Nonessential plastid-encoded ribosomal proteins in tobacco: a developmental role for plastid translation and implications for reductive genome evolution. Plant Cell 23(9):3137–3155. https://doi.org/10.1105/tpc.111.088906

Fu P-C, Sun S-S, Twyford AD, Li B-B, Zhou R-Q, Chen S-L, Gao Q-B, Favre A (2021) Lineage-specific plastid degradation in subtribe *Gentianinae* (Gentianaceae). Ecol Evol 11:3286–3299. https://doi.org/10.1002/ece3.7281

Goettsch B, Hilton-Taylor C, Cruz-Piñón G et al (2015) High proportion of cactus species threatened with extinction. Nat Plants 1:15142. https://doi.org/10.1038/nplants.2015.142

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. https://doi.org/10.1002/bies.201200137

Greiner S, Köhl K (2014) Growing evening primroses (*Oenothera*). Front Plant Sci 5:38. https://doi.org/10.3389/fpls.2014.00038

Greiner S, Wang X, Herrmann RG, Rauwolf U, Mayer K, Haberer G, Meurer J (2008a) The complete nucleotide sequences of the 5 genetically distinct plastid genomes of *Oenothera*, subsection Oenothera: II. A microevolutionary view using bioinformatics and formal genetic data. Mol Biol Evol 25:2019–2030. https://doi.org/10.1093/molbev/msn149

Greiner S, Wang X, Rauwolf U, Silber MV, Mayer K, Meurer J, Haberer G, Herrmann RG (2008b) The complete nucleotide sequences of the five genetically distinct plastid genomes of Oenothera, subsection Oenothera: I. Sequence evaluation and plastome evolution. Nucleic Acids Res 36(7):2366–2378. https://doi.org/10.1093/nar/gkn081

Greiner S, Rauwolf U, Meurer J, Herrmann RG (2011) The role of plastids in plant speciation. Mol Ecol 20:671–691. https://doi.org/10.1111/j.1365-294X.2010.04984.x

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. https://doi.org/10.1093/nar/gkz238

Griffiths H, Males J (2017) Succulent plants. Curr Biol 27:890–896. https://doi.org/10.1016/j.cub.2017.03.021

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. https://doi.org/10.1093/jhered/esy064

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(1):583–600. https://doi.org/10.1093/molbev/msq229

Hu S, Sablok G, Wang B, Qu D, Barbaro E, Viola R, Li M, Varotto C (2015) Plastome organization and evolution of chloroplast genes in *Cardamine* species adapted to contrasting habitats. BMC Genom 16:306. https://doi.org/10.1186/s12864-015-1498-0

IUCN (2020) IUCN Red list of threatened species. version 2021.1. https://www.iucnredlist.org/. Accessed 15 Aug 2021.

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. https://doi.org/10.1126/science.1229262

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 AAAAT-Pase complex is required for chloroplast protein import. Plant Cell 30:2677–2703. https://doi.org/10.1105/tpc.18.00357

Kim HT, Chase MW (2017) Independent degradation in genes of the plastid *ndh* gene family in species of the orchid genus *Cymbidium* (Orchidaceae; Epidendroideae). PLoS ONE 12(11):e0187318. https://doi.org/10.1371/journal.pone.0187318

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. https://doi.org/10.1111/j.1365-313X.2005.02533.x

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. Front Plant Sci 19(11):729. https://doi.org/10.3389/fpls.2020.00729

Krawczyk K, Sawicki J (2013) The uneven rate of the molecular evolution of gene sequences of DNA-dependent RNA polymerase I of the genus *Lamium* L. Int J Mol Sci 14:11376–11391. https://doi.org/10.3390/ijms140611376

Kurata S, Weixlbaumer A, Ohtsuki T, Shimazaki T, Wada T, Kirino Y, Takai K, Watanabe K, Ramakrishnan V, Suzuki T (2008) Modified uridines with C5-methylene substituents at the first position of the tRNA anticodon stabilize U.G wobble pairing during decoding. J Biol Chem 283(27):18801–18811. https://doi.org/10.1074/jbc.M800233200

Kuroda H, Maliga P (2003) The plastid *clpP1* protease gene is essential for plant development. Nature 425:86–89. https://doi.org/10.1038/nature01909

Lee D-H, Cho W-B, Choi B-H, Lee J-H (2017) Characterization of two complete chloroplast genomes in the tribe *Gnaphalieae* (Asteraceae): gene loss or pseudogenization of *T-GGU* and implications for phylogenetic relationships. Hortic Sci Technol 35(6):769–783. https://doi.org/10.12972/kjhst.20170081

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. https://doi.org/10.1111/j.1365-313X.2007.03177.x

Lin C-S, Chen JJW, Chiu C-C, Hsiao HCW, Yang C-J, Jin X-H, Leebens-Mack J, de Pamphilis CW, Huang Y-T, Yang L-H, Chang W-J, Kui L, Wong GK-S, Hu J-M, Wang W, Shih M-C (2017) Concomitant loss of NDH complex-related genes within chloroplast and nuclear genomes in some orchids. Plant J 90:994–1006. https://doi.org/10.1111/tpj.13525

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. https://doi.org/10.1007/s00425-018-2841-x

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. https://doi.org/10.1007/s00299-017-2231-z

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. https://doi.org/10.1016/j.gene.2018.05.088

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. https://doi.org/10.1007/s00425-019-03217-z

Lopes AS, Pacheco TG, Silva ON, Vieira LN, Guerra MP, Matta EPL, Baura VA, Balsanelli E, Souza EM, Pedrosa FO, Rogalski M (2021) Plastid genome evolution in Amazonian açaí palm (*Euterpe oleracea* Mart.) and Atlantic forest açaí palm (*Euterpe edulis* Mart.). Plant Mol Biol 105(4):559–574. https://doi.org/10.1007/s11103-020-01109-5

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. https://doi.org/10.1093/nar/gkw413

Machado MC (2009) The genus *Melocactus* in eastern Brazil: part I—an introduction to *Melocactus*. Br Cactus Succul J 27:1–16

Mandal D, Köhrer C, Su D, Babu IR, Chan CT, Liu Y, Söll D, Blum P, Kuwahara M, Dedon PC, Rajbhandary UL (2014) Identification and codon reading properties of 5-cyanomethyl uridine,

a new modified nucleoside found in the anticodon wobble position of mutant haloarchaeal isoleucine tRNAs. RNA 20(2):177–188. https://doi.org/10.1261/rna.042358.113

Mayer JA, Cushman JC (2019) Nutritional and mineral content of prickly pear cactus: a highly water-use efficient forage, fodder and food species. J Agro Crop Sci 205:625–634. https://doi.org/10.1111/jac.12353

Mower JP (2009) The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. Nucleic Acids Res 37:W253–W259. https://doi.org/10.1093/nar/gkp337

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(4):2471–2490. https://doi.org/10.1104/pp.16.01519

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300

Osuna-Martínez U, Reyes-Esparza J, Rodríguez-Fragoso L (2014) Cactus (*Opuntia ficus-indica*): a review on its antioxidants properties and potential pharmacological use in chronic diseases. Nat Prod Chem Res 2:6. https://doi.org/10.4172/2329-6836.1000153

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. https://doi.org/10.3390/plants9080979

Pacheco TG, Lopes AS, Viana GD, 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. https://doi.org/10.1007/s00425-018-3023-6

Pacheco TG, Lopes AS, Oliveira JD, Otoni WC, Balsanelli E, Pedrosa FO, Souza EM, Rogalski M (2020a) The complete plastome of *Passiflora cirrhiflora* A. Juss.: structural features, RNA editing sites, hotspots of nucleotide diversity and molecular markers within the subgenus *Deidamioides*. Braz J Bot 43:839–853. https://doi.org/10.1007/s40415-020-00655-y

Pacheco TG, Lopes AS, Welter JF, Yotoko KSC, Otoni WC, Vieira LN, Guerra MP, Nodari RO, Balsanelli E, Pedrosa FO, Souza EM, Rogalski M (2020b) Plastome sequences of the subgenus *Passiflora* reveal highly divergent genes and specific evolutionary features. Plant Mol Biol 104:21–37. https://doi.org/10.1007/s11103-020-01020-z

Pacheco TG, Morais da Silva G, Lopes AS, Oliveira JD, Rogalski JM, Balsanelli E, Souza EM, Pedrosa FO, Rogalski M (2020c) Phylogenetic and evolutionary features of the plastome of *Tropaeolum pentaphyllum* Lam. (Tropaeolaceae). Planta 252:17. https://doi.org/10.1007/s00425-020-03427-w

Pfitzinger H, Weil JH, Pillay DTN, Guillemaut P (1990) Codon recognition mechanisms in plant chloroplasts. Plant Mol Biol 14:805–814. https://doi.org/10.1007/BF00016513

Raman G, Park S (2015) Analysis of the complete chloroplast genome of a medicinal plant, *Dianthus superbus* var. *longicalycinus*, from a comparative genomics perspective. PLoS ONE 10(10):e0141329. https://doi.org/10.1371/journal.pone.0141329

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. https://doi.org/10.1534/genetics.116.188268

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. https://doi.org/10.1111/j.1467-7652.2011.00621.x

Rogalski M, Ruf S, Bock R (2006) Tobacco plastid ribosomal protein S18 is essential for cell survival. Nucleic Acids Res 34:4537–4545. https://doi.org/10.1093/nar/gkl634

Rogalski M, Karcher D, Bock R (2008a) Superwobbling facilitates translation with reduced tRNA sets. Nat Struct Mol Biol 15:192–198. https://doi.org/10.1038/nsmb.1370

Rogalski M, Schöttler 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. https://doi.org/10.1105/tpc.108.060392

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. https://doi.org/10.3389/fpls.2015.00586

Rousseau-Gueutin M, Huang X, Higginson E, Ayliffe M, Day A, Timmis JN (2013) Potential functional replacement of the plastidic Acetyl-CoA carboxylase subunit (*accD*) gene by recent transfers to the nucleus in some angiosperm lineages. Plant Physiol 161(4):1918–1929. https://doi.org/10.1104/pp.113.214528

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol 34:3299–3302. https://doi.org/10.1093/molbev/msx248

Rumeau D, Peltier G, Cournac L (2007) Chlororespiration and cyclic electron flow around PSI during photosynthesis and plant stress response. Plant Cell Environ 30:1041–1051. https://doi.org/10.1111/j.1365-3040.2007.01675.x

Salie MJ, Thelen JJ (2016) Regulation and structure of the heteromeric acetyl-CoA carboxylase. Biochim Biophys Acta 1861:1207–1213. https://doi.org/10.1016/j.bbalip.2016.04.004

Salinas-Giegé T, Giegé R, Giegé P (2015) tRNA biology in mitochondria. Int J Mol Sci 16:4518–4559. https://doi.org/10.3390/ijms16034518

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. https://doi.org/10.3732/ajb.1500184

Schuwirth BS, Borovinskaya MA, Hau CW, Zhang W, Vila-Sanjurjo A, Holton JM, Cate JH (2005) Structures of the bacterial ribosome at 3.5 A resolution. Science 310(5749):827–834. https://doi.org/10.1126/science.1117230

Shetty AA, Rana MK, Preetham SP (2012) Cactus: a medicinal food. J Food Sci Technol 49:530–536. https://doi.org/10.1007/s13197-011-0462-5

Shikanai T, Shimizu K, Ueda K, Nishimura Y, Kuroiwa T, Hashimoto T (2001) The chloroplast *clpP* gene, encoding a proteolytic subunit of ATP-dependent protease, is indispensable for chloroplast development in tobacco. Plant Cell Physiol 42:264–273. https://doi.org/10.1093/pcp/pce031

Silva GM, Lopes AS, Pacheco TG, Machado KLG, Silva MC, Oliveira JD, Baura VA, Balsanelli E, 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. Rev Bras Bot 44:97–116. https://doi.org/10.1007/s40415-020-00689-2

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(10):392. https://doi.org/10.3390/plants8100392

Stern A, Doron-Faigenboim A, Erez E, Martz E, Bacharach E, Pupko T (2007) Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. Nucleic Acids Res 35:W506–W511. https://doi.org/10.1093/nar/gkm382

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. https://doi.org/10.1042/BCJ20190365

Takenaka M, Zehrmann A, Verbitskiy D, Härtel B, Brennicke A (2013) RNA editing in plants and its evolution. Annu Rev Genet 47:335–352. https://doi.org/10.1146/annurev-genet-111212-133519

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729. https://doi.org/10.1093/molbev/mst197

Taylor NP (2000) Taxonomy and phytogeography of the Cactaceae of eastern Brazil. PhD Thesis. The Open University e Royal Botanic Gardens, Kew. p 414. https://doi.org/10.21954/ou.ro.0000d49b

Taylor NP, Zappi DC (2004) Cacti of Eastern Brazil. Royal Botanic Gardens, Kew, London, p 511

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. https://doi.org/10.1093/nar/gkx391

Torres-Silva G, Resende SV, Lima-Brito A, Bezerra HB, de Santana JRF, Schnadelbach AS (2018) In vitro shoot production, morphological alterations and genetic instability of *Melocactus glaucescens* (Cactaceae), an endangered species endemic to eastern Brazil. S Afr J Bot 115:100–107. https://doi.org/10.1016/j.sajb.2018.01.001

Torres-Silva G, Matos EM, Correia LF, Fortini EA, Soares WS, Batista DS, Otoni CG, Azevedo AA, Viccini LF, Koelhler AD, Resende SV, Secht CD, Otoni WC (2020) Anatomy, flow cytometry, and X-ray tomography reveal tissue organization and ploidy distribution in long-term in vitro cultures of *Melocactus* species. Front Plant Sci 11:1314. https://doi.org/10.3389/fpls.2020.01314

Torres-Silva G, Correia LNF, Koehler AD, Batista DS, Faria DV, Resende SV, Strickler SR, Fouracre J, Romanel E, Specht CD, Otoni WC (2021a) Expression of *Melocactus glaucescens* SERK1 sheds new light on the mechanism of areolar activation in cacti. Plant Cell Tiss Organ Cult. https://doi.org/10.1007/s11240-021-02137-9(Inpress)

Torres-Silva G, Schnadelbach AS, Bezerra HB, Lima-Brito A, Resende SV (2021b) In vitro conservation and genetic diversity of threatened species of *Melocactus* (Cactaceae). Biodivers Conserv 30:1067–1080. https://doi.org/10.1007/s10531-021-02132-8

Vieira LN, Faoro H, Fraga HPF, Rogalski M, de Souza EM, de Oliveira PF, Nodari RO, Guerra MP (2014) An improved protocol for intact chloroplasts and cpDNA isolation in conifers. PLoS ONE 9:e84792. https://doi.org/10.1371/journal.pone.0084792

Vries J, Sousa FL, Bölter B, Soll J, Goulda SB (2015) YCF1: a green TIC? Plant Cell 27:1827–1833. https://doi.org/10.1105/tpc.114.135541

Weixlbaumer A, Murphy FV, Dziergowska A, Malkiewicz A, Vendeix FAP, Agris PF, Ramakrishnan V (2007) Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines. Nat Struct Mol Biol 14:498–502. https://doi.org/10.1038/nsmb1242

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. https://doi.org/10.1093/gbe/evw115

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(12):1400. https://doi.org/10.3732/apps.1400059

Wimberly BT, Brodersen DE, Clemons WM Jr, Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V (2000) Structure of the 30S ribosomal subunit. Nature 407:327–339. https://doi.org/10.1038/35030006

Xu JH, Liu Q, Hu W, Wang T, Xue Q, Messing J (2015) Dynamics of chloroplast genomes in green plants. Genomics 106:221–231. https://doi.org/10.1016/j.ygeno.2015.07.004

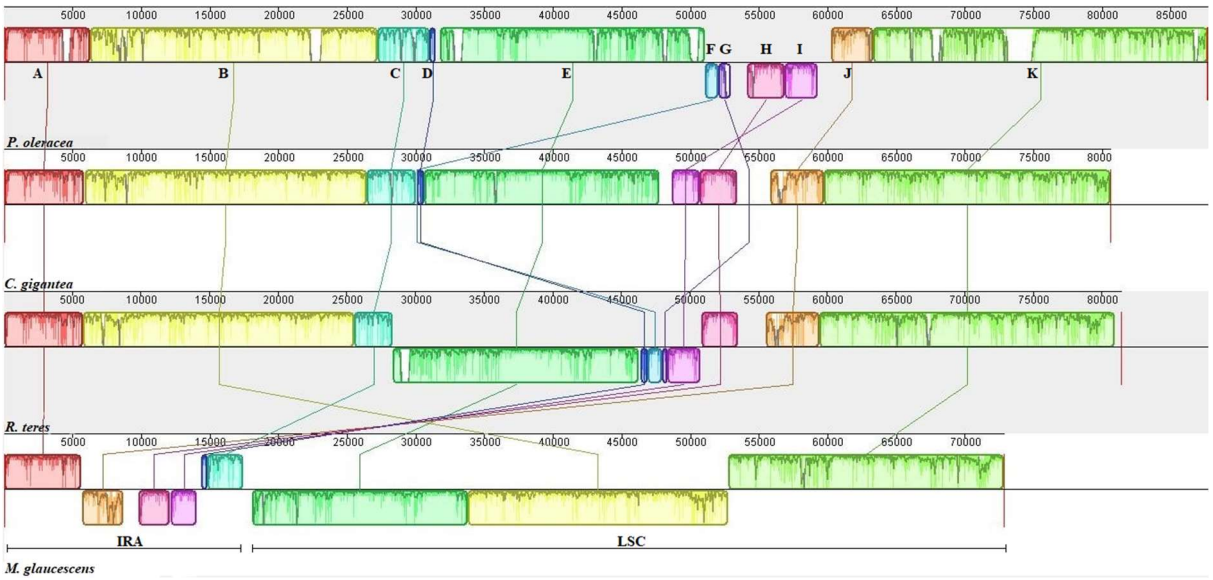
Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155(1):431–449. https://doi.org/10.1093/genetics/155.1.431

Zhang J, Ruhlman TA, Sabir J, Blazier JC, Jansen RK (2015) Coordinated rates of evolution between interacting plastid and nuclear genes in Geraniaceae. Plant Cell 27:563–573. https://doi.org/10.1105/tpc.114.134353

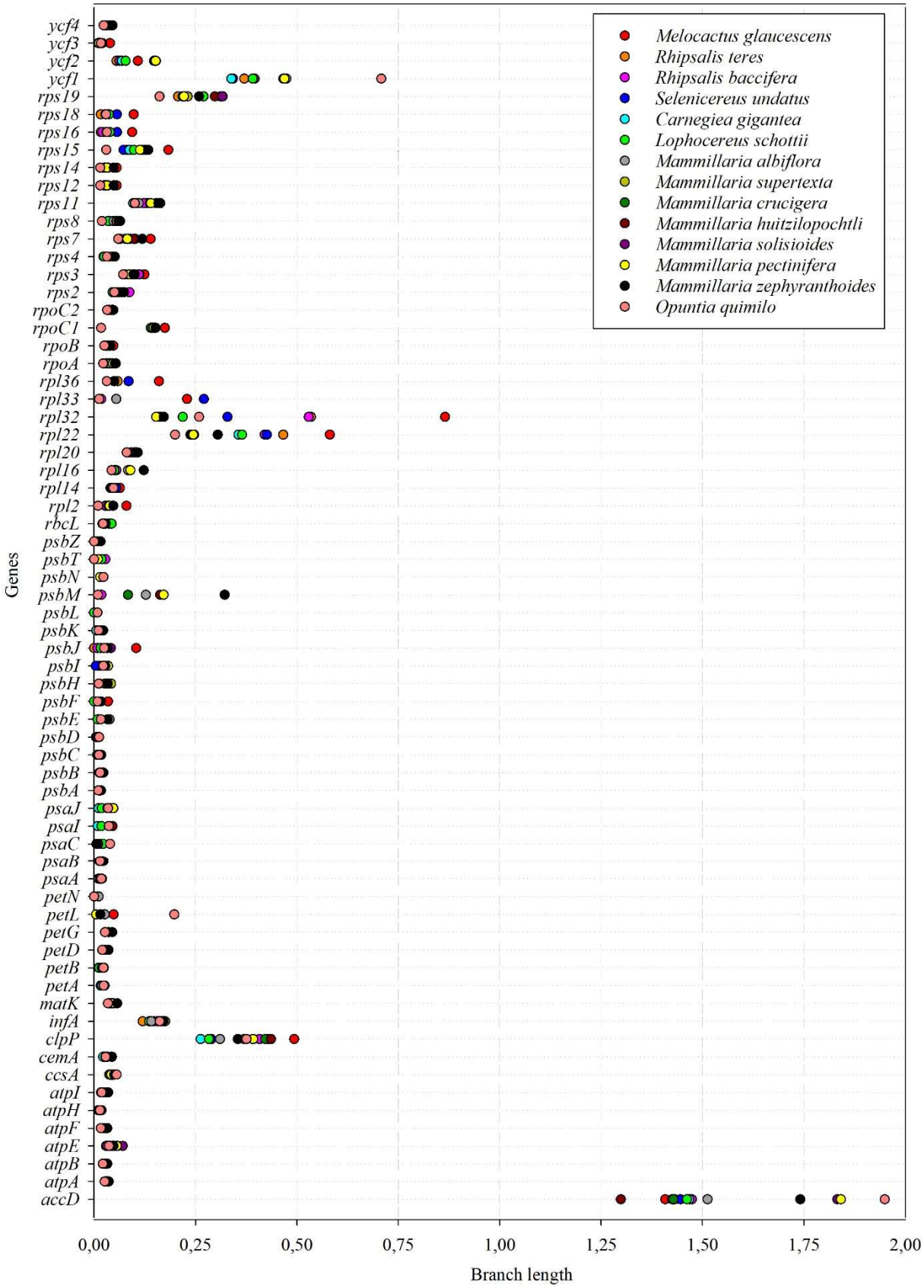
Zupok A, Kozul D, Schöttler MA, Niehörster J, Garbsch F, Liere K, Fischer A, Zoschke R, Malinova I, Bock R, Greiner S (2021) A photosynthesis operon in the chloroplast genome drives speciation in evening primroses. Plant Cell 33:2583–2601. https://doi.org/10.1093/plcell/koab155

SUPPLEMENTARY MATERIAL

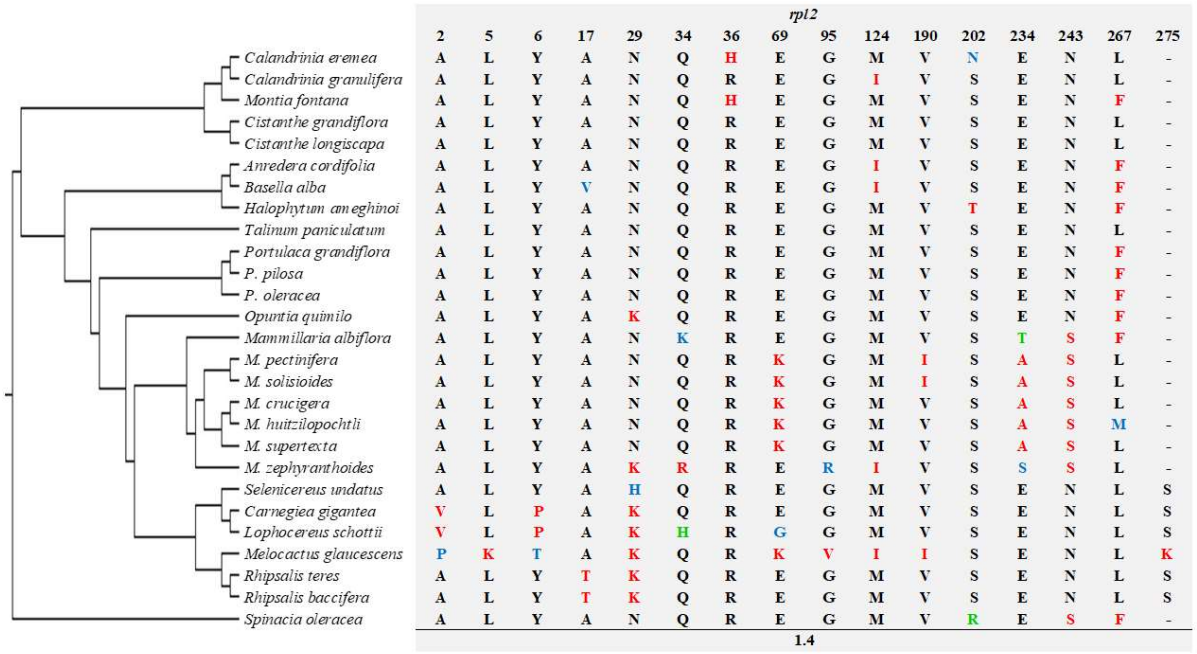
Supplementary Figures



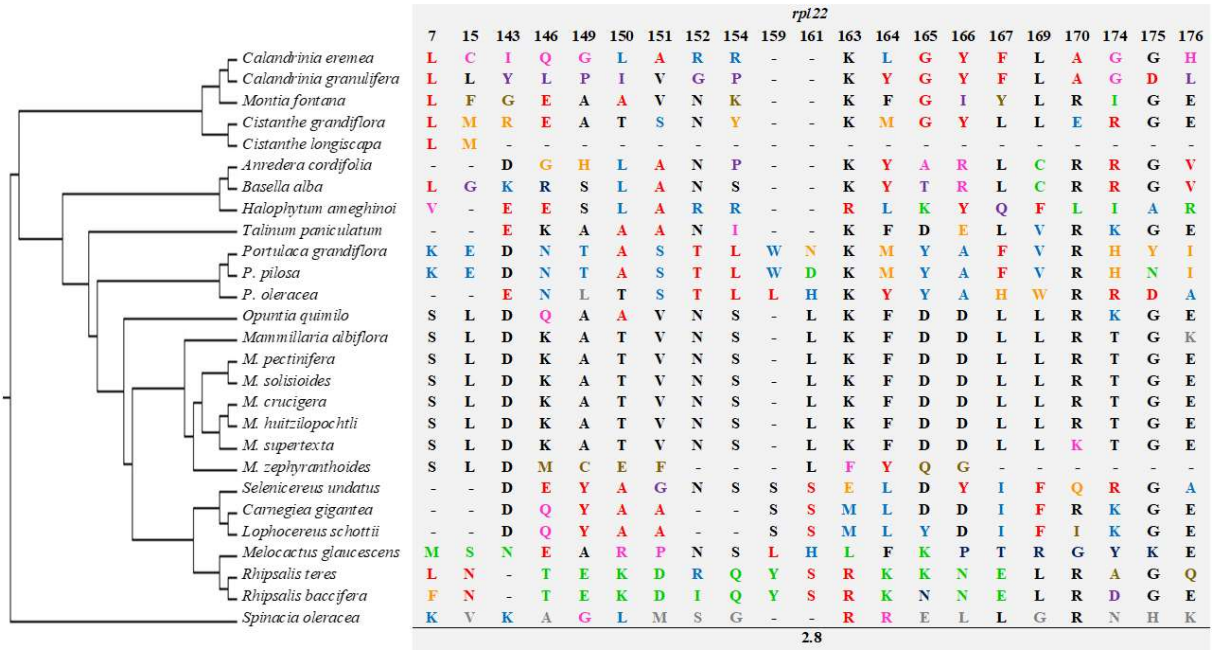
Supplementary Fig. S1. Multiple alignments of *M. glaucescens*, *R. teres*, *C. gigantea* plastomes (subfamily Cactoideae), and *P. oleracea* as reference species. The region aligned corresponds to the LSC of *P. oleracea* plastome, from the *trnH-GUG* to the *rps19* genes. Locally collinear blocks (LCB) are color-coded. More details about the gene content of the LCBs involved in rearrangements are present in Figure 2



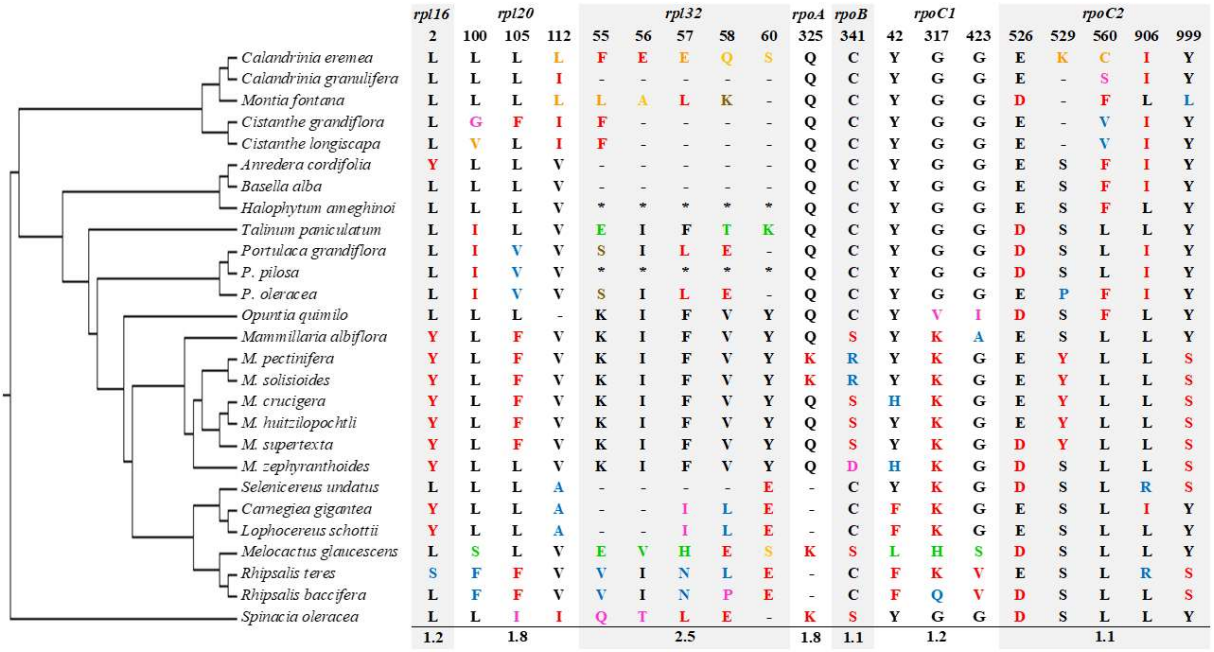
Supplementary Fig. S2. Divergence of protein-coding genes in the plastomes of family Cactaceae based on phylogenetic reconstruction. The gene divergence was estimated by the sum of total branch lengths until the common ancestor node in each gene tree inferred



Supplementary Fig. S3. Non-synonymous substitutions in the *rpl2* gene. The amino acids are plotted across the Cactineae phylogeny inferred based on plastid genes (Figure 4). Different amino acid types identified at the same position are highlighted in distinct colors. The amino acid positions are relative to *M. glaucescens* plastid genes. The Ka/Ks scores for each codon position are indicated in the last line

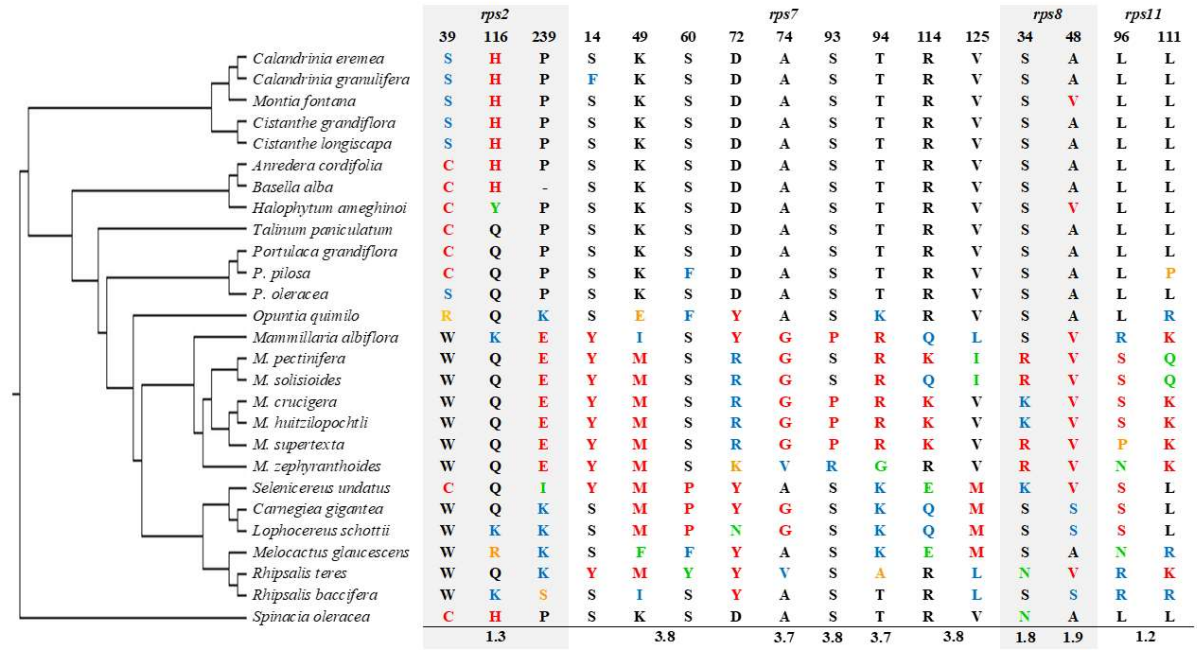


Supplementary Fig. S4. Non-synonymous substitutions in the *rpl22* gene. The amino acids are plotted across the Cactineae phylogeny inferred based on plastid genes (Figure 4). Different amino acid types identified at the same position are highlighted in distinct colors. The amino acid positions are relative to *M. glaucescens* plastid genes. The Ka/Ks scores for each codon position are indicated in the last line



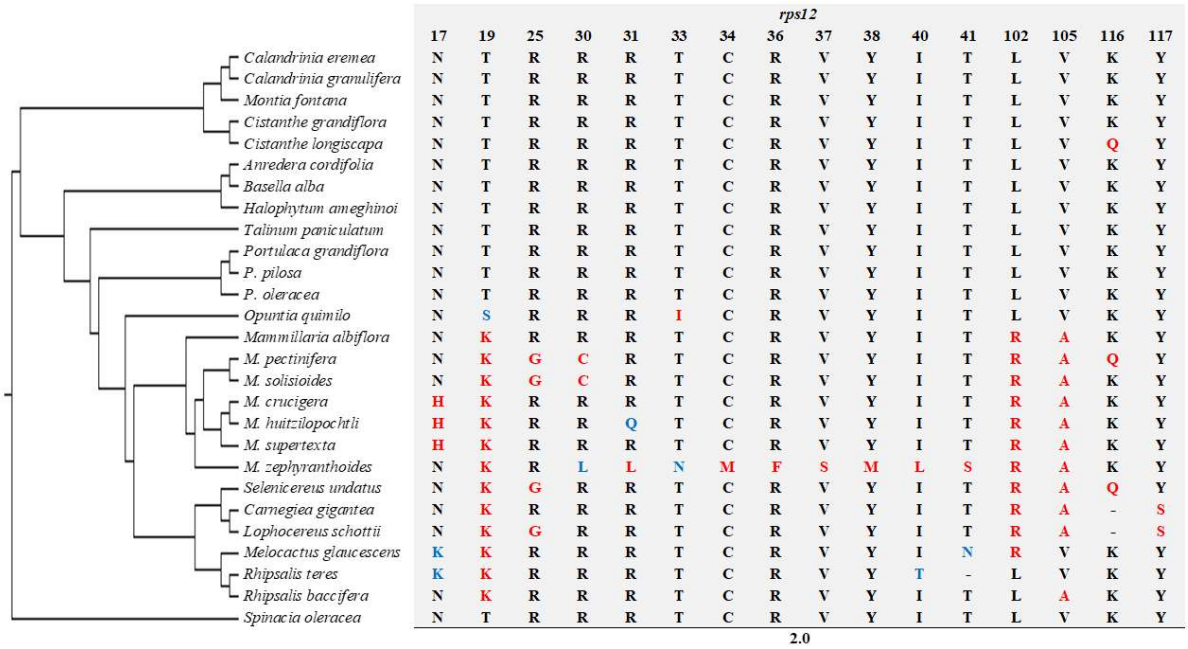
	rpl16	*rpl20*			*rpl32*					*rpoA*	*rpoB*	*rpoC1*			*rpoC2*				
	2	100	105	112	55	56	57	58	60	325	341	42	317	423	526	529	560	906	999
Calandrinia eremea	L	L	L	L	F	E	E	Q	S	Q	C	Y	G	G	E	K	C	I	Y
Calandrinia granulifera	L	L	L	I	-	-	-	-	-	Q	C	Y	G	G	E	-	S	I	Y
Montia fontana	L	L	L	L	L	A	L	K	-	Q	C	Y	G	G	D	-	F	L	L
Cistanthe grandiflora	L	G	F	I	F	-	-	-	-	Q	C	Y	G	G	E	-	V	I	Y
Cistanthe longiscapa	L	V	L	I	F	-	-	-	-	Q	C	Y	G	G	E	-	V	I	Y
Anredera cordifolia	Y	L	L	V	-	-	-	-	-	Q	C	Y	G	G	E	S	F	I	Y
Basella alba	L	L	L	V	-	-	-	-	-	Q	C	Y	G	G	E	S	F	I	Y
Halophytum ameghinoi	L	L	L	V	*	*	*	*	*	Q	C	Y	G	G	E	S	F	L	Y
Talinum paniculatum	L	I	L	V	E	I	F	T	K	Q	C	Y	G	G	D	S	L	L	Y
Portulaca grandiflora	L	I	V	V	S	I	L	E	-	Q	C	Y	G	G	D	S	L	I	Y
P. pilosa	L	I	V	V	*	*	*	*	*	Q	C	Y	G	G	D	S	L	I	Y
P. oleracea	L	I	V	V	S	I	L	E	-	Q	C	Y	G	G	E	P	F	I	Y
Opuntia quimilo	L	L	L	-	K	I	F	V	Y	Q	C	Y	V	I	D	S	F	L	Y
Mammillaria albiflora	Y	L	F	V	K	I	F	V	Y	Q	S	Y	K	A	E	S	L	L	Y
M. pectinifera	Y	L	F	V	K	I	F	V	Y	K	R	Y	K	G	E	Y	L	L	S
M. solisioides	Y	L	F	V	K	I	F	V	Y	K	R	Y	K	G	E	Y	L	L	S
M. crucigera	Y	L	F	V	K	I	F	V	Y	Q	S	H	K	G	E	Y	L	L	S
M. huitzilopochtli	Y	L	F	V	K	I	F	V	Y	Q	S	Y	K	G	E	Y	L	L	S
M. supertexta	Y	L	F	V	K	I	F	V	Y	Q	S	Y	K	G	D	Y	L	L	S
M. zephyranthoides	Y	L	L	V	K	I	F	V	Y	Q	D	H	K	G	D	S	L	L	S
Selenicereus undatus	L	L	L	A	-	-	-	-	E	-	C	Y	K	G	D	S	L	R	S
Carnegiea gigantea	Y	L	L	A	-	-	I	L	E	-	C	F	K	G	E	S	L	I	Y
Lophocereus schottii	Y	L	L	A	-	-	I	L	E	-	C	F	K	G	E	S	L	L	Y
Melocactus glaucescens	L	S	L	V	E	V	H	E	S	K	S	L	H	S	D	S	L	L	Y
Rhipsalis teres	S	F	F	V	V	I	N	L	E	-	C	F	K	V	E	S	L	R	S
Rhipsalis baccifera	L	F	F	V	V	I	N	P	E	-	C	F	Q	V	D	S	L	L	S
Spinacia oleracea	L	L	I	I	Q	T	L	E	-	K	S	Y	G	G	D	S	L	L	Y
	1.2	1.8			2.5					1.8	1.1	1.2			1.1				

Supplementary Fig. S5. Non-synonymous substitutions in the *rpl16*, *rpl20*, *rpl32*, *rpoA*, *rpoB*, *rpoC1* and *rpoC2* genes. The amino acids are plotted across the Cactineae phylogeny inferred based on plastid genes (Figure 4). Different amino acid types identified at the same position are highlighted in distinct colors. The amino acid positions are relative to *M. glaucescens* plastid genes. The Ka/Ks scores for each codon position are indicated in the last line. Absent genes are indicated by an asterisk

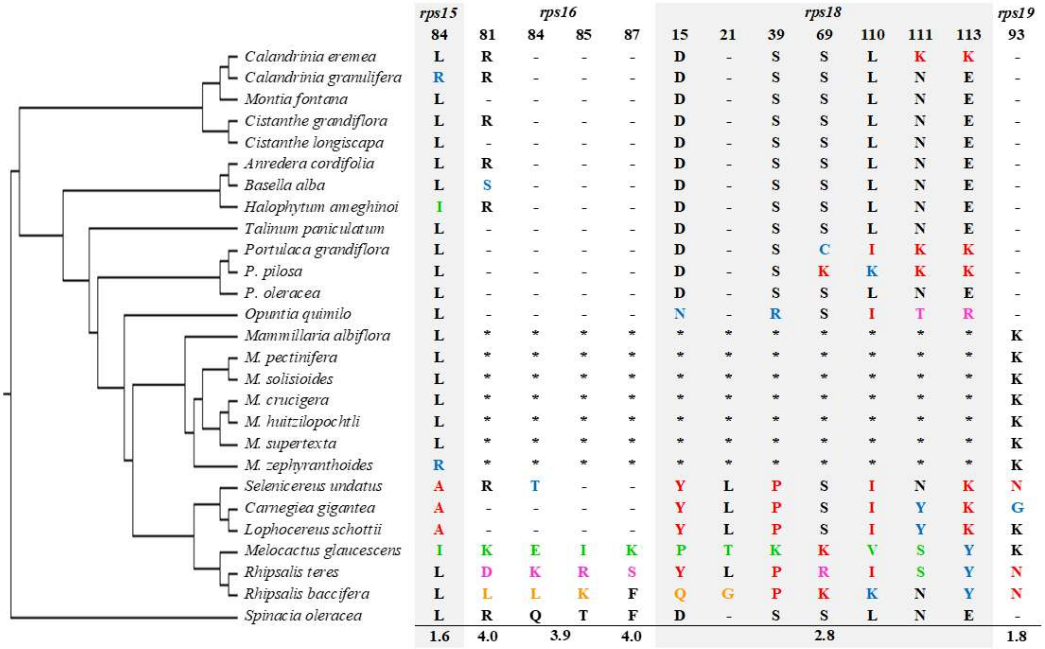


	rps2			*rps7*									*rps8*		*rps11*	
	39	116	239	14	49	60	72	74	93	94	114	125	34	48	96	111
Calandrinia eremea	S	H	P	S	K	S	D	A	S	T	R	V	S	A	L	L
Calandrinia granulifera	S	H	P	F	K	S	D	A	S	T	R	V	S	A	L	L
Montia fontana	S	H	P	S	K	S	D	A	S	T	R	V	S	V	L	L
Cistanthe grandiflora	S	H	P	S	K	S	D	A	S	T	R	V	S	A	L	L
Cistanthe longiscapa	S	H	P	S	K	S	D	A	S	T	R	V	S	A	L	L
Anredera cordifolia	C	H	P	S	K	S	D	A	S	T	R	V	S	A	L	L
Basella alba	C	H	-	S	K	S	D	A	S	T	R	V	S	A	L	L
Halophytum ameghinoi	C	Y	P	S	K	S	D	A	S	T	R	V	S	V	L	L
Talinum paniculatum	C	Q	P	S	K	S	D	A	S	T	R	V	S	A	L	L
Portulaca grandiflora	C	Q	P	S	K	S	D	A	S	T	R	V	S	A	L	L
P. pilosa	C	Q	P	S	K	F	D	A	S	T	R	V	S	A	L	P
P. oleracea	S	Q	P	S	K	S	D	A	S	T	R	V	S	A	L	L
Opuntia quimilo	R	Q	K	S	E	F	Y	A	S	K	R	V	S	A	L	R
Mammillaria albiflora	W	K	E	Y	I	S	Y	G	P	R	Q	L	S	V	R	K
M. pectinifera	W	Q	E	Y	M	S	R	G	S	R	K	I	R	V	S	Q
M. solisioides	W	Q	E	Y	M	S	R	G	S	R	Q	I	R	V	S	Q
M. crucigera	W	Q	E	Y	M	S	R	G	P	R	K	V	K	V	S	K
M. huitzilopochtli	W	Q	E	Y	M	S	R	G	P	R	K	V	K	V	S	K
M. supertexta	W	Q	E	Y	M	S	R	G	P	R	K	V	R	V	P	K
M. zephyranthoides	W	Q	E	Y	M	S	K	V	R	G	R	V	R	V	N	K
Selenicereus undatus	C	Q	I	Y	M	P	Y	A	S	K	E	M	K	V	S	L
Carnegiea gigantea	W	Q	K	S	M	P	Y	G	S	K	Q	M	S	S	S	L
Lophocereus schottii	W	K	K	S	M	P	N	G	S	K	Q	M	S	S	S	L
Melocactus glaucescens	W	R	K	S	F	F	Y	A	S	K	E	M	S	A	N	R
Rhipsalis teres	W	Q	K	Y	M	Y	Y	V	S	A	R	L	N	V	R	K
Rhipsalis baccifera	W	K	S	S	I	S	Y	A	S	T	R	L	S	S	R	R
Spinacia oleracea	C	H	P	S	K	S	D	A	S	T	R	V	N	A	L	L
	1.3			3.8				3.7	3.8	3.7	3.8		1.8	1.9	1.2	

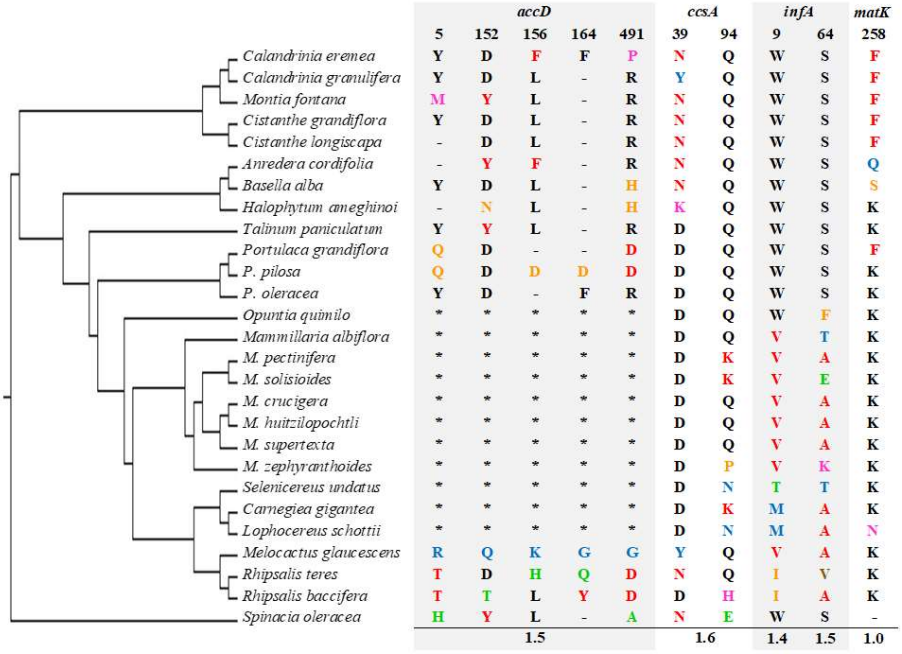
Supplementary Fig. S6. Non-synonymous substitutions in the *rps2*, *rps7*, *rps8*, and *rps11* genes. The amino acids are plotted across the Cactineae phylogeny inferred based on plastid genes (Figure 4). Different amino acid types identified at the same position are highlighted in distinct colors. The amino acid positions are relative to *M. glaucescens* plastid genes. The Ka/Ks scores for each codon position are indicated in the last line



Supplementary Fig. S7. Non-synonymous substitutions in the *rps12* gene. The amino acids are plotted across the Cactineae phylogeny inferred based on plastid genes (Figure 4). Different amino acid types identified at the same position are highlighted in distinct colors. The amino acid positions are relative to *M. glaucescens* plastid genes. The Ka/Ks scores for each codon position are indicated in the last line

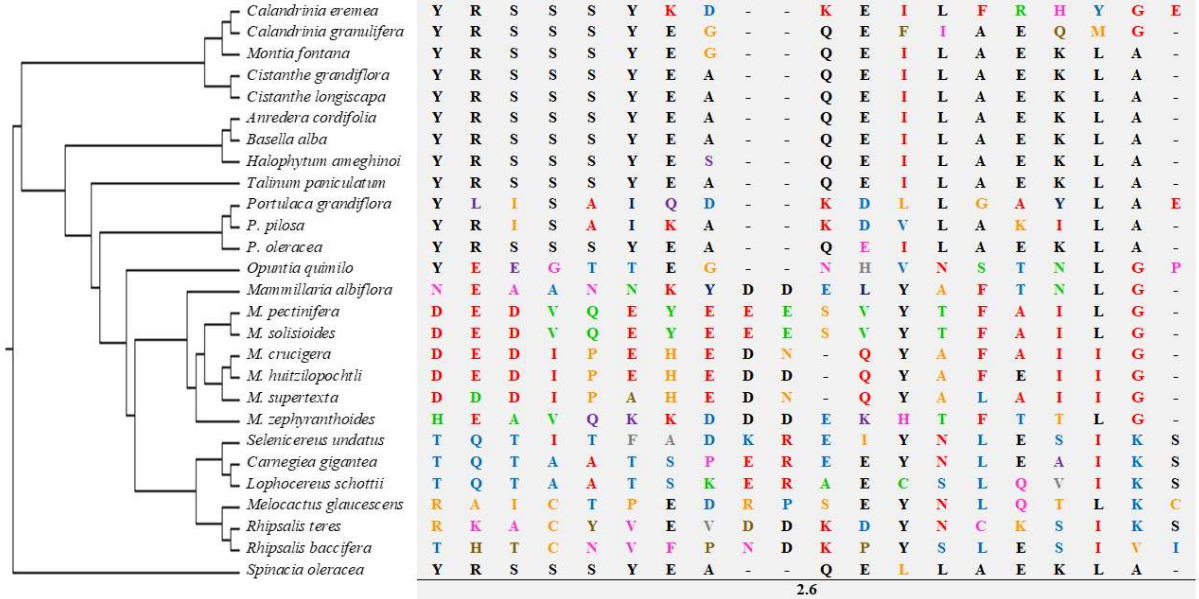


Supplementary Fig. S8. Non-synonymous substitutions in the *rps15*, *rps16*, *rps18*, and *rps19* genes. The amino acids are plotted across the Cactineae phylogeny inferred based on plastid genes (Figure 4). Different amino acid types identified at the same position are highlighted in distinct colors. The amino acid positions are relative to *M. glaucescens* plastid genes. The Ka/Ks scores for each codon position are indicated in the last line. Absent genes are indicated by an asterisk



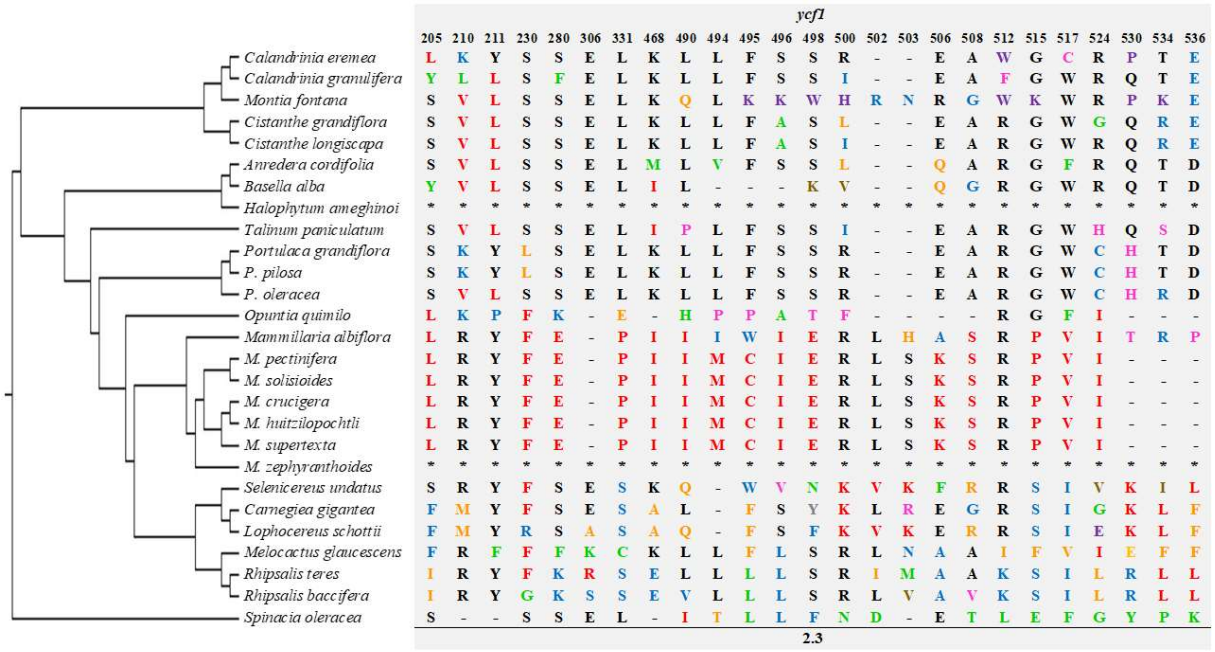
	accD					*ccsA*		*infA*		*matK*
	5	152	156	164	491	39	94	9	64	258
Calandrinia eremea	Y	D	F	F	P	N	Q	W	S	F
Calandrinia granulifera	Y	D	L	-	R	Y	Q	W	S	F
Montia fontana	M	Y	L	-	R	N	Q	W	S	F
Cistanthe grandiflora	Y	D	L	-	R	N	Q	W	S	F
Cistanthe longiscapa	-	D	L	-	R	N	Q	W	S	F
Anredera cordifolia	-	Y	F	-	R	N	Q	W	S	Q
Basella alba	Y	D	L	-	H	N	Q	W	S	S
Halophytum ameghinoi	-	N	L	-	H	K	Q	W	S	K
Talinum paniculatum	Y	Y	L	-	R	D	Q	W	S	K
Portulaca grandiflora	Q	D	-	-	D	D	Q	W	S	F
P. pilosa	Q	D	D	D	D	D	Q	W	S	K
P. oleracea	Y	D	-	F	R	D	Q	W	S	K
Opuntia quimilo	*	*	*	*	*	D	Q	W	F	K
Mammillaria albiflora	*	*	*	*	*	D	Q	V	T	K
M. pectinifera	*	*	*	*	*	D	K	V	A	K
M. solisioides	*	*	*	*	*	D	K	V	E	K
M. crucigera	*	*	*	*	*	D	Q	V	A	K
M. huitzilopochtli	*	*	*	*	*	D	Q	V	A	K
M. supertexta	*	*	*	*	*	D	Q	V	A	K
M. zephyranthoides	*	*	*	*	*	D	P	V	K	K
Selenicereus undatus	*	*	*	*	*	D	N	T	T	K
Carnegiea gigantea	*	*	*	*	*	D	K	M	A	K
Lophocereus schottii	*	*	*	*	*	D	N	M	A	N
Melocactus glaucescens	R	Q	K	G	G	Y	Q	V	A	K
Rhipsalis teres	T	D	H	Q	D	N	Q	I	V	K
Rhipsalis baccifera	T	T	L	Y	D	D	H	I	A	K
Spinacia oleracea	H	Y	L	-	A	N	E	W	S	-
			1.5			1.6		1.4	1.5	1.0

Supplementary Fig. S9. Non-synonymous substitutions in the *accD*, *ccsA*, *infA,* and *matK* genes. The amino acids are plotted across the Cactineae phylogeny inferred based on plastid genes (Figure 4). Different amino acids identified at the same position are highlighted in distinct colors. Amino acid positions are relative to *M. glaucescens* plastid genes. The Ka/Ks scores for each codon position are indicated in the last line. Absent genes are indicated by an asterisk

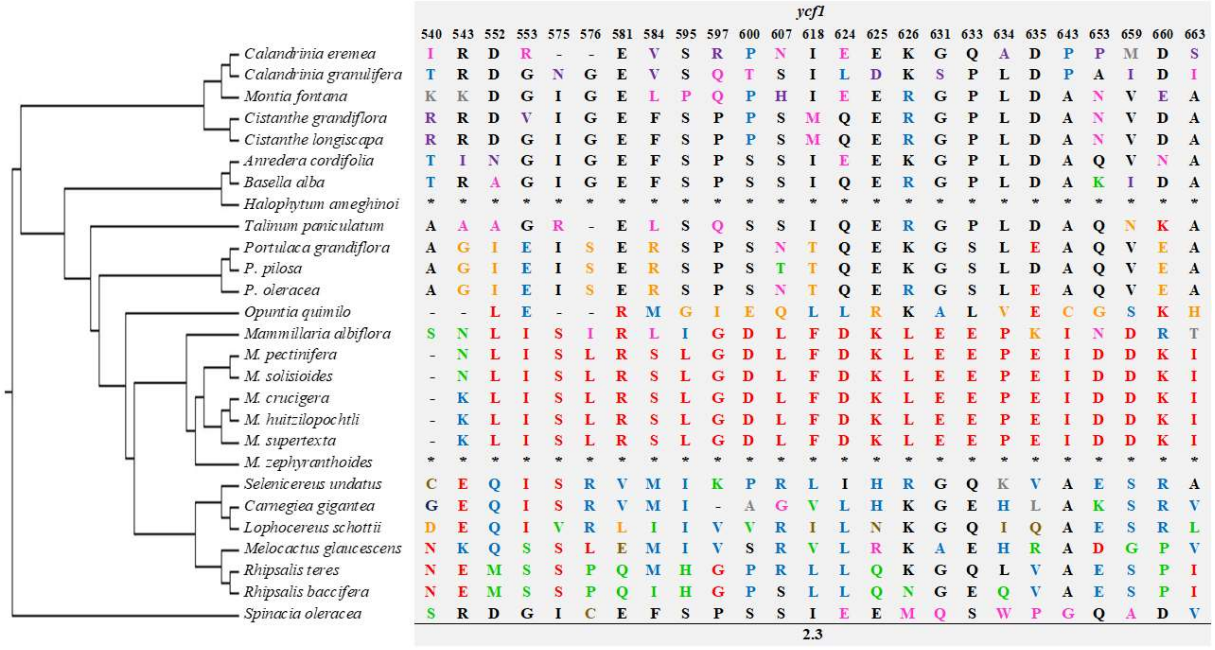


	clpP																			
	24	26	41	128	129	133	134	135	137	139	140	143	145	146	148	149	153	154	198	201
Calandrinia eremea	Y	R	S	S	S	Y	K	D	-	-	K	E	I	L	F	R	H	Y	G	E
Calandrinia granulifera	Y	R	S	S	S	Y	E	G	-	-	Q	E	F	I	A	E	Q	M	G	-
Montia fontana	Y	R	S	S	S	Y	E	G	-	-	Q	E	I	L	A	E	K	L	A	-
Cistanthe grandiflora	Y	R	S	S	S	Y	E	A	-	-	Q	E	I	L	A	E	K	L	A	-
Cistanthe longiscapa	Y	R	S	S	S	Y	E	A	-	-	Q	E	I	L	A	E	K	L	A	-
Anredera cordifolia	Y	R	S	S	S	Y	E	A	-	-	Q	E	I	L	A	E	K	L	A	-
Basella alba	Y	R	S	S	S	Y	E	A	-	-	Q	E	I	L	A	E	K	L	A	-
Halophytum ameghinoi	Y	R	S	S	S	Y	E	S	-	-	Q	E	I	L	A	E	K	L	A	-
Talinum paniculatum	Y	R	S	S	S	Y	E	A	-	-	Q	E	I	L	A	E	K	L	A	-
Portulaca grandiflora	Y	L	I	S	A	I	Q	D	-	-	K	D	L	L	G	A	Y	L	A	E
P. pilosa	Y	R	I	S	A	I	K	A	-	-	K	D	V	L	A	K	I	L	A	-
P. oleracea	Y	R	S	S	S	Y	E	A	-	-	Q	E	I	L	A	E	K	L	A	-
Opuntia quimilo	Y	E	E	G	T	T	E	G	-	-	N	H	V	N	S	T	N	L	G	P
Mammillaria albiflora	N	E	A	A	N	N	K	Y	D	D	E	L	Y	A	F	T	N	L	G	-
M. pectinifera	D	E	D	V	Q	E	Y	E	E	E	S	V	Y	T	F	A	I	L	G	-
M. solisioides	D	E	D	V	Q	E	Y	E	E	E	S	V	Y	T	F	A	I	L	G	-
M. crucigera	D	E	D	I	P	E	H	E	D	N	-	Q	Y	A	F	A	I	I	G	-
M. huitzilopochtli	D	E	D	I	P	E	H	E	D	D	-	Q	Y	A	F	E	I	I	G	-
M. supertexta	D	D	D	I	P	A	H	E	D	N	-	Q	Y	A	L	A	I	I	G	-
M. zephyranthoides	H	E	A	V	Q	K	K	D	D	D	E	K	H	T	F	T	T	L	G	-
Selenicereus undatus	T	Q	T	I	T	F	A	D	K	R	E	I	Y	N	L	E	S	I	K	S
Carnegiea gigantea	T	Q	T	A	A	T	S	P	E	R	E	E	Y	N	L	E	A	I	K	S
Lophocereus schottii	T	Q	T	A	A	T	S	K	E	R	A	E	C	S	L	Q	V	I	K	S
Melocactus glaucescens	R	A	I	C	T	P	E	D	R	P	S	E	Y	N	L	Q	T	L	K	C
Rhipsalis teres	R	K	A	C	Y	V	E	V	D	D	K	D	Y	N	C	K	S	I	K	S
Rhipsalis baccifera	T	H	T	C	N	V	F	P	N	D	K	P	Y	S	L	E	S	I	V	I
Spinacia oleracea	Y	R	S	S	S	Y	E	A	-	-	Q	E	L	L	A	E	K	L	A	-
											2.6									

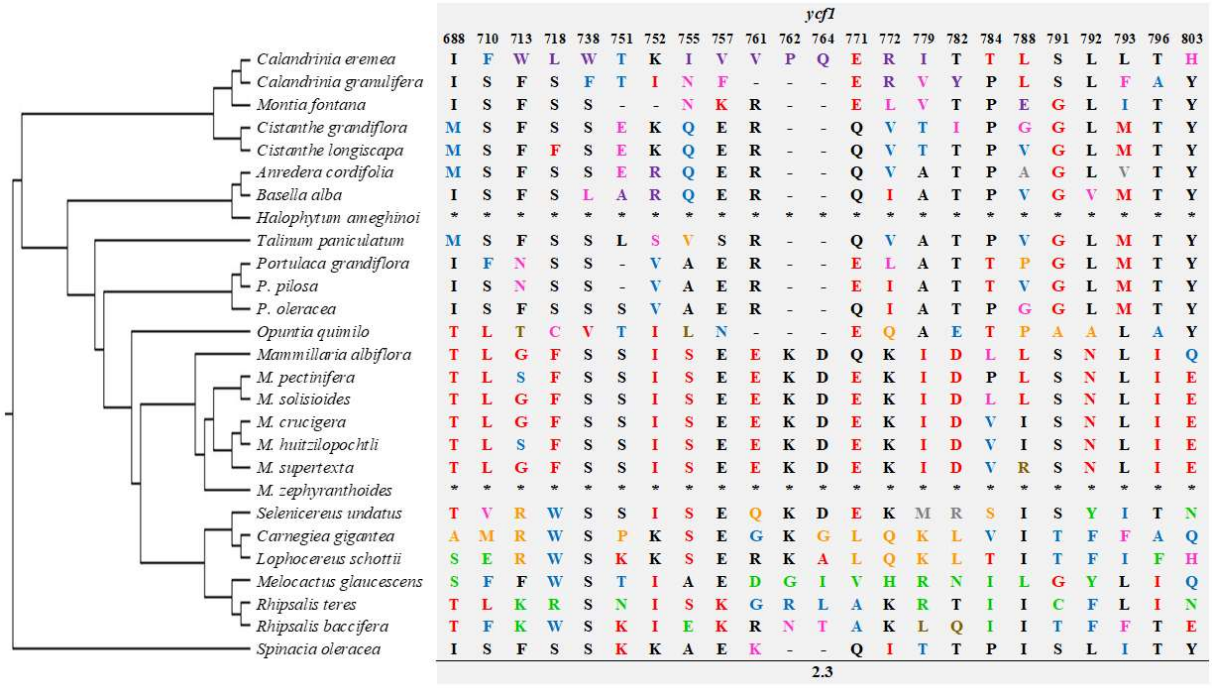
Supplementary Fig. S10. Non-synonymous substitutions in the *clpP* gene. The amino acids are plotted across the Cactineae phylogeny inferred based on plastid genes (Figure 4). Different amino acids identified at the same position are highlighted in distinct colors. Amino acid positions are relative to *M. glaucescens* plastid genes. The Ka/Ks scores for each codon position are indicated in the last line. Absent genes are indicated by an asterisk



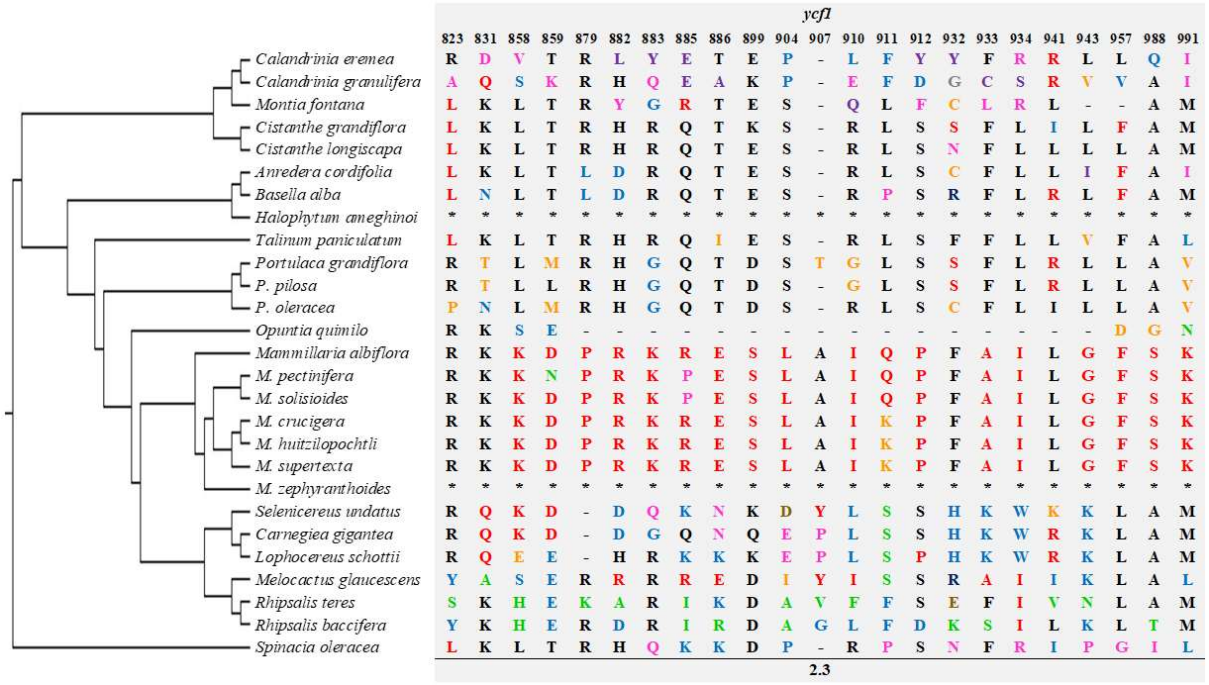
Supplementary Fig. S11. Non-synonymous substitutions in the *ycf1* gene. The amino acids are plotted across the Cactineae phylogeny inferred based on plastid genes (Figure 4). Different amino acids identified at the same position are highlighted in distinct colors. Amino acid positions are relative to *M. glaucescens* plastid genes. The Ka/Ks scores for each codon position are indicated in the last line. Absent genes are indicated by an asterisk



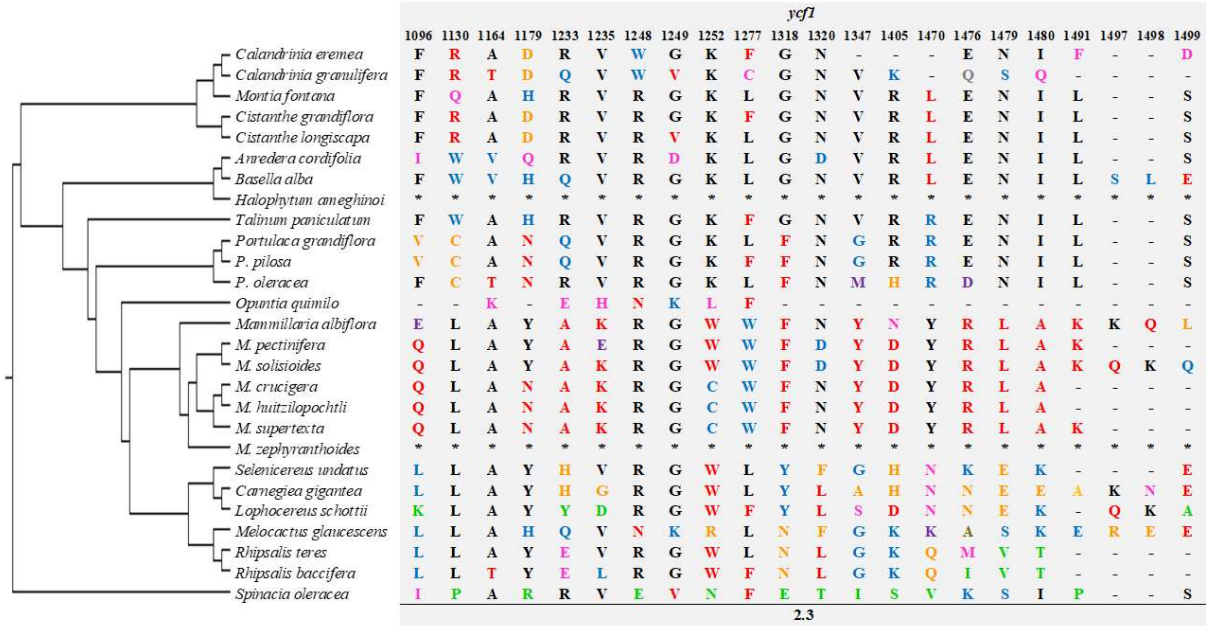
Supplementary Fig. S12. Non-synonymous substitutions in the *ycf1* gene. The amino acids are plotted across the Cactineae phylogeny inferred based on plastid genes (Figure 4). Different amino acids identified at the same position are highlighted in distinct colors. Amino acid positions are relative to *M. glaucescens* plastid genes. The Ka/Ks scores for each codon position are indicated in the last line. Absent genes are indicated by an asterisk



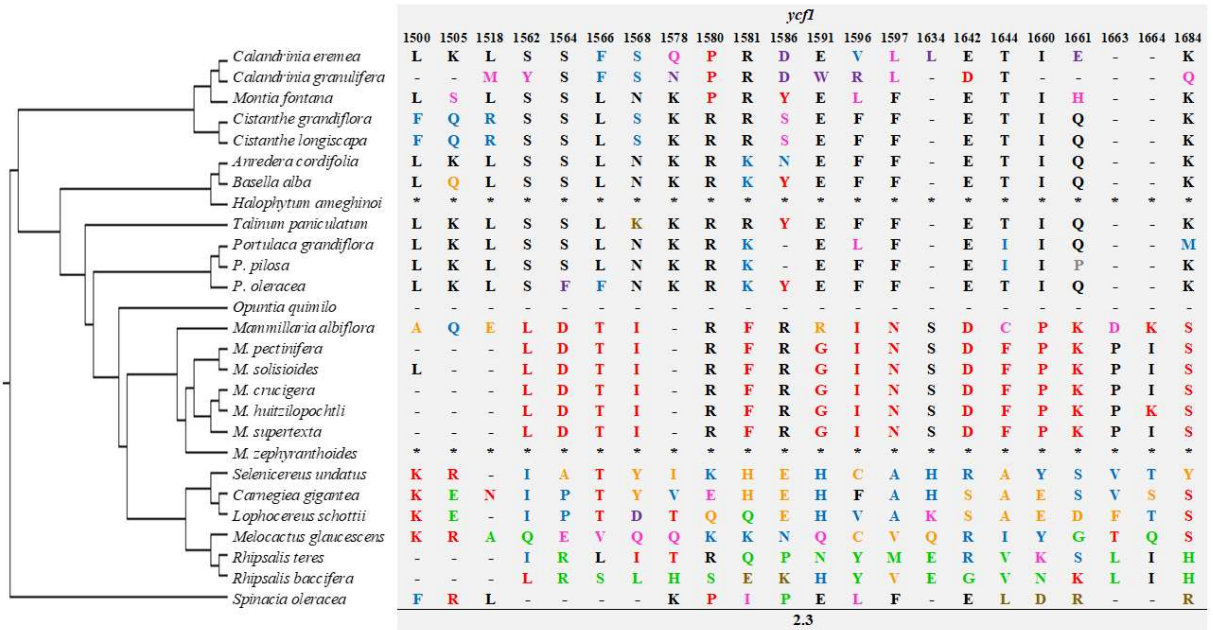
Supplementary Fig. S13. Non-synonymous substitutions in the *ycf1* gene. The amino acids are plotted across the Cactineae phylogeny inferred based on plastid genes (Figure 4). Different amino acids identified at the same position are highlighted in distinct colors. Amino acid positions are relative to *M. glaucescens* plastid genes. The Ka/Ks scores for each codon position are indicated in the last line. Absent genes are indicated by an asterisk



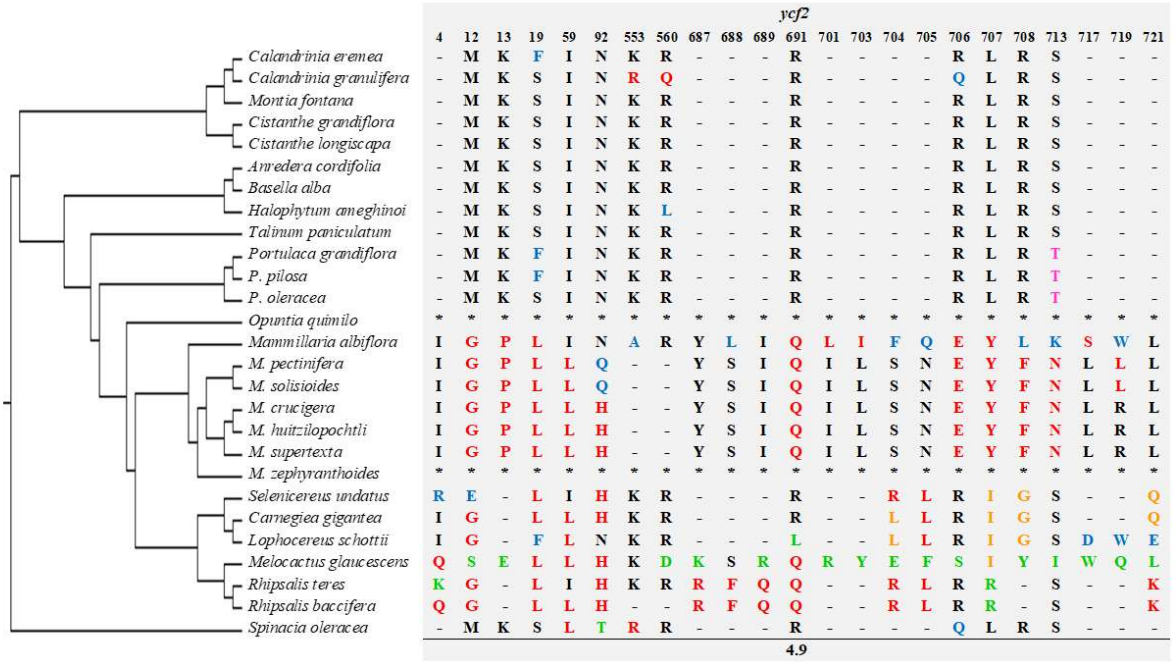
Supplementary Fig. S14. Non-synonymous substitutions in the *ycf1* gene. The amino acids are plotted across the Cactineae phylogeny inferred based on plastid genes (Figure 4). Different amino acids identified at the same position are highlighted in distinct colors. Amino acid positions are relative to *M. glaucescens* plastid genes. The Ka/Ks scores for each codon position are indicated in the last line. Absent genes are indicated by an asterisk



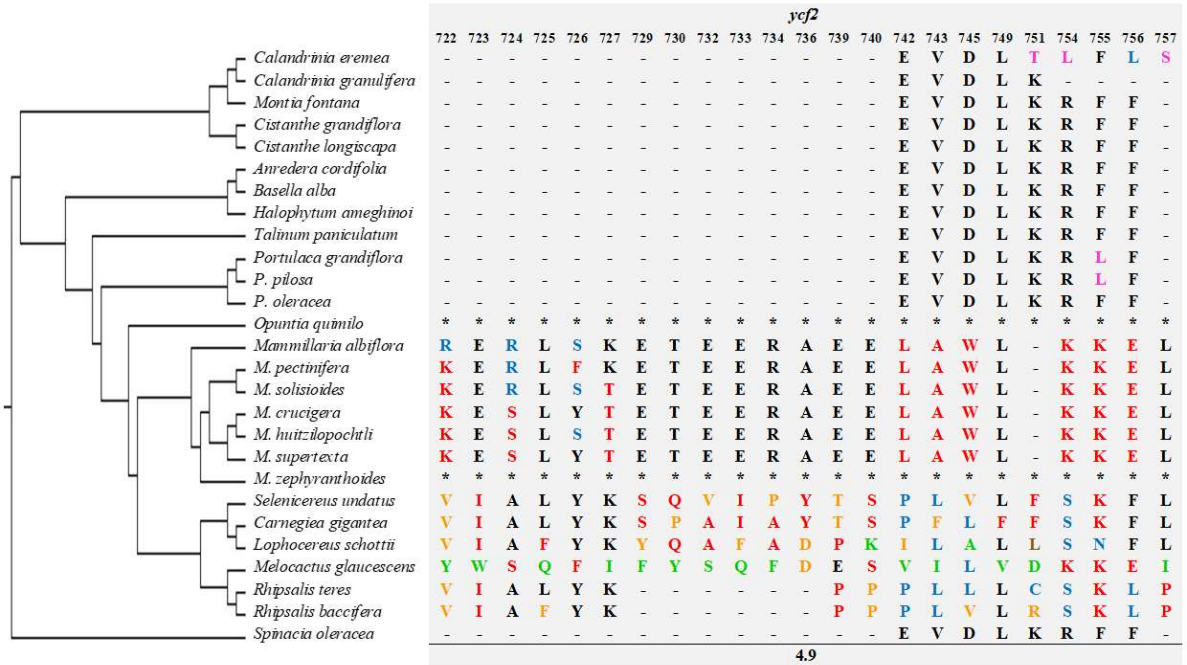
Supplementary Fig. S15. Non-synonymous substitutions in the *ycf1* gene. The amino acids are plotted across the Cactineae phylogeny inferred based on plastid genes (Figure 4). Different amino acids identified at the same position are highlighted in distinct colors. Amino acid positions are relative to *M. glaucescens* plastid genes. The Ka/Ks scores for each codon position are indicated in the last line. Absent genes are indicated by an asterisk



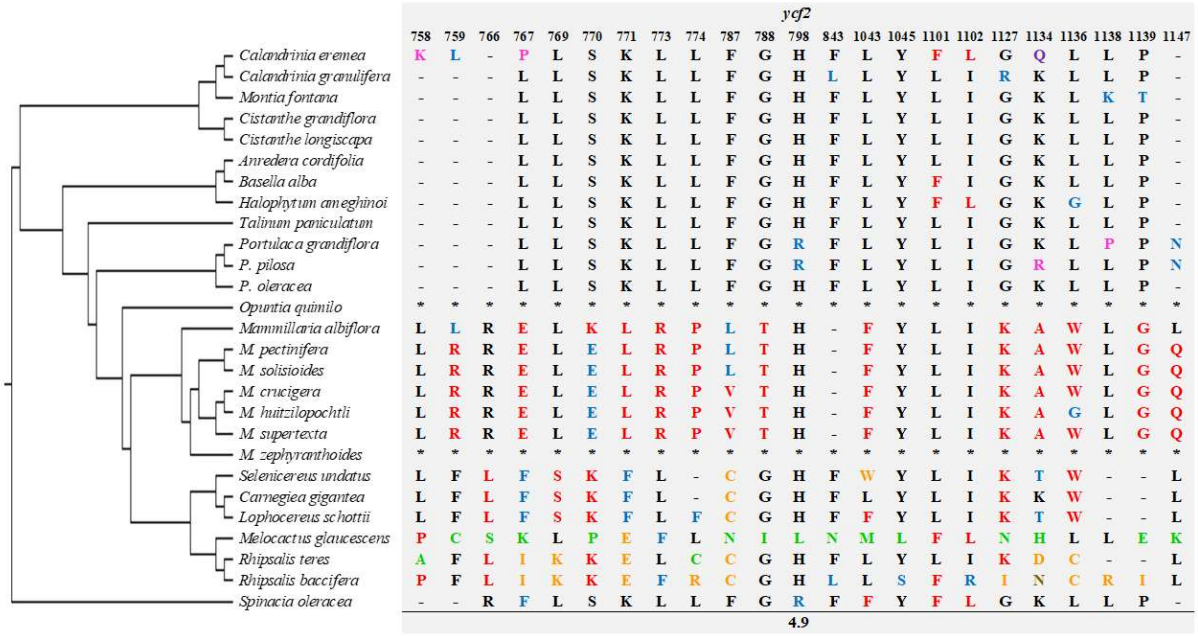
Supplementary Fig. S16. Non-synonymous substitutions in the *ycf1* gene. The amino acids are plotted across the Cactineae phylogeny inferred based on plastid genes (Figure 4). Different amino acids identified at the same position are highlighted in distinct colors. Amino acid positions are relative to *M. glaucescens* plastid genes. The Ka/Ks scores for each codon position are indicated in the last line. Absent genes are indicated by an asterisk



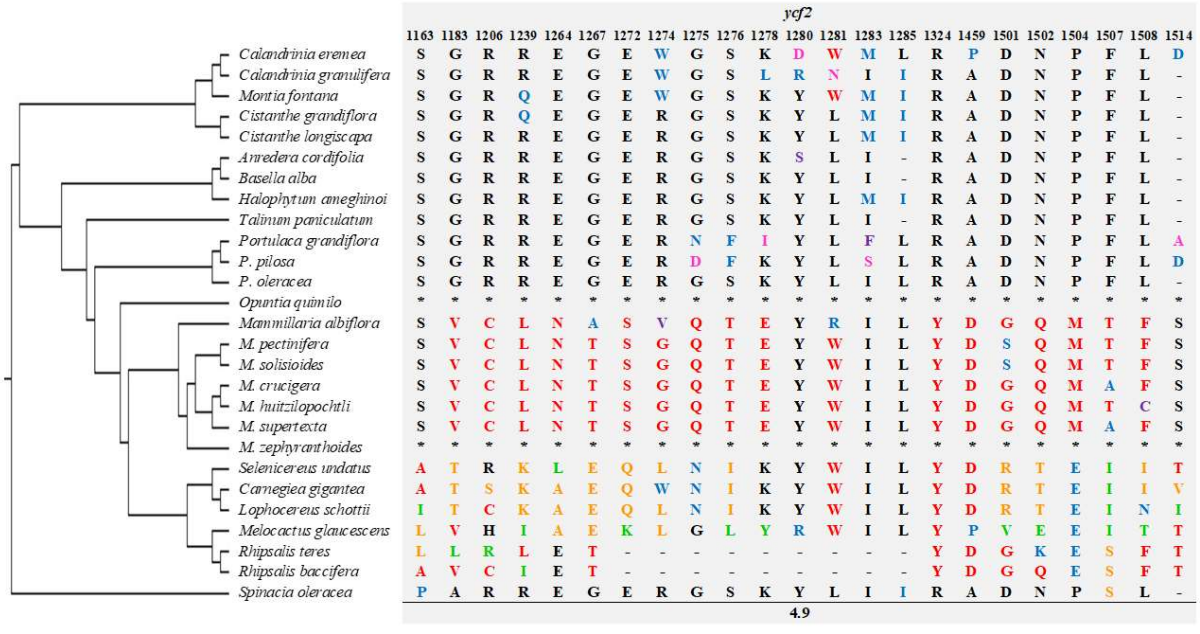
Supplementary Fig. S17. Non-synonymous substitutions in the *ycf2* gene. The amino acids are plotted across the Cactineae phylogeny inferred based on plastid genes (Figure 4). Different amino acids identified at the same position are highlighted in distinct colors. Amino acid positions are relative to *M. glaucescens* plastid genes. The Ka/Ks scores for each codon position are indicated in the last line. Absent genes are indicated by an asterisk



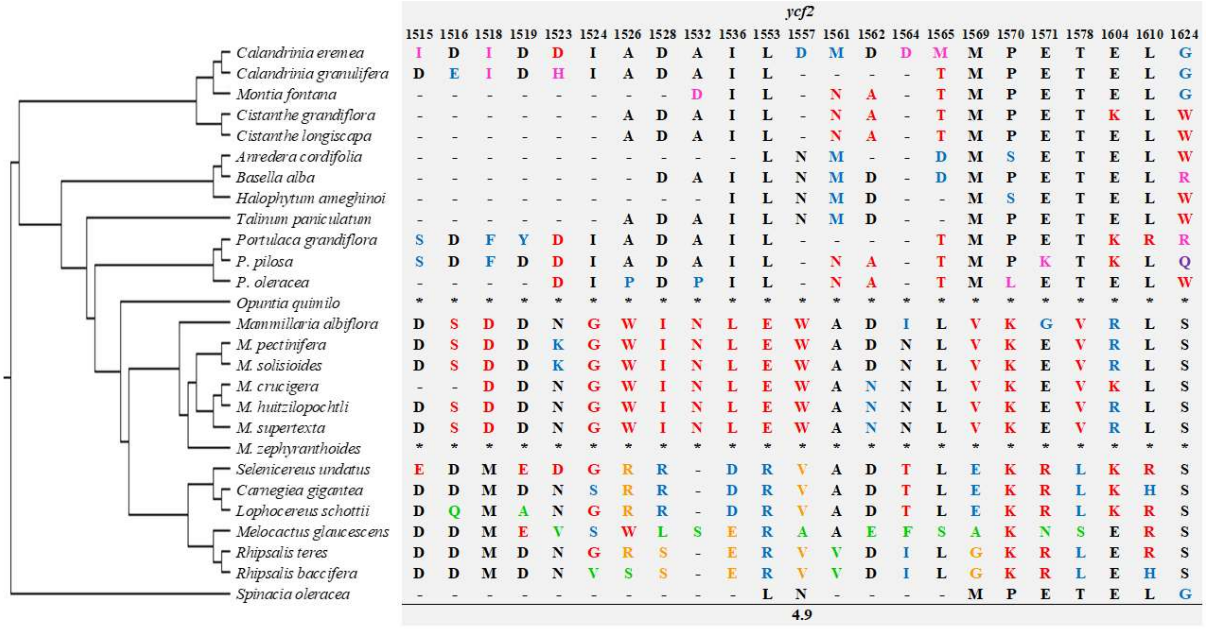
Supplementary Fig. S18. Non-synonymous substitutions in the *ycf2* gene. The amino acids are plotted across the Cactineae phylogeny inferred based on plastid genes (Figure 4). Different amino acids identified at the same position are highlighted in distinct colors. Amino acid positions are relative to *M. glaucescens* plastid genes. The Ka/Ks scores for each codon position are indicated in the last line. Absent genes are indicated by an asterisk



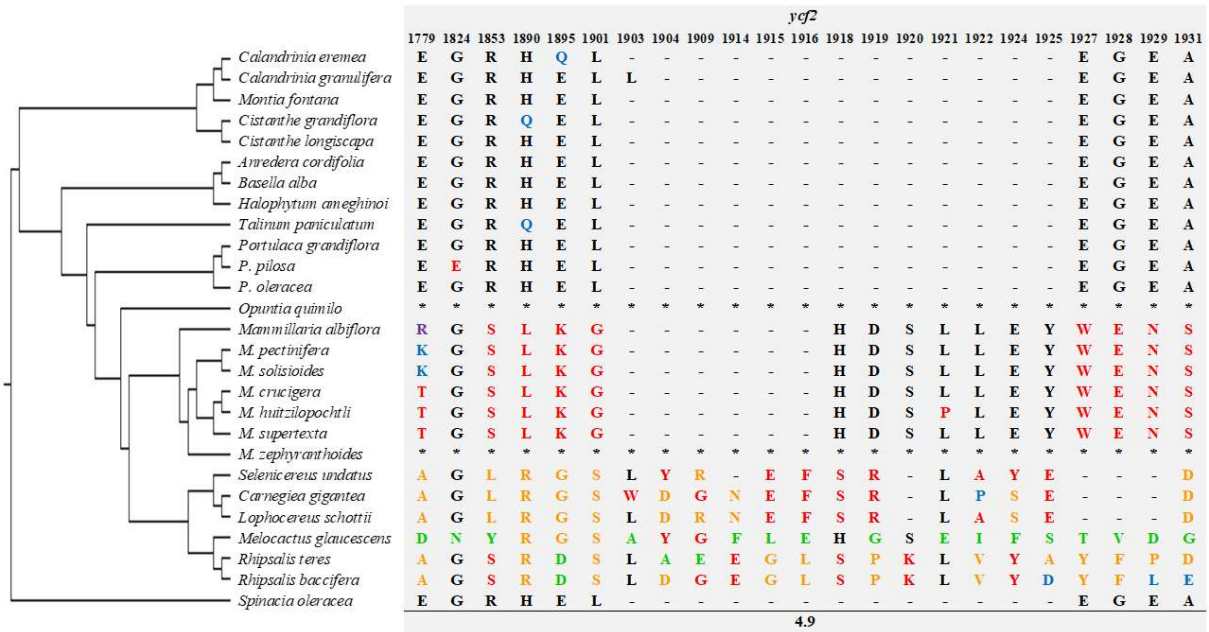
Supplementary Fig. S19. Non-synonymous substitutions in the *ycf2* gene. The amino acids are plotted across the Cactineae phylogeny inferred based on plastid genes (Figure 4). Different amino acids identified at the same position are highlighted in distinct colors. Amino acid positions are relative to *M. glaucescens* plastid genes. The Ka/Ks scores for each codon position are indicated in the last line. Absent genes are indicated by an asterisk



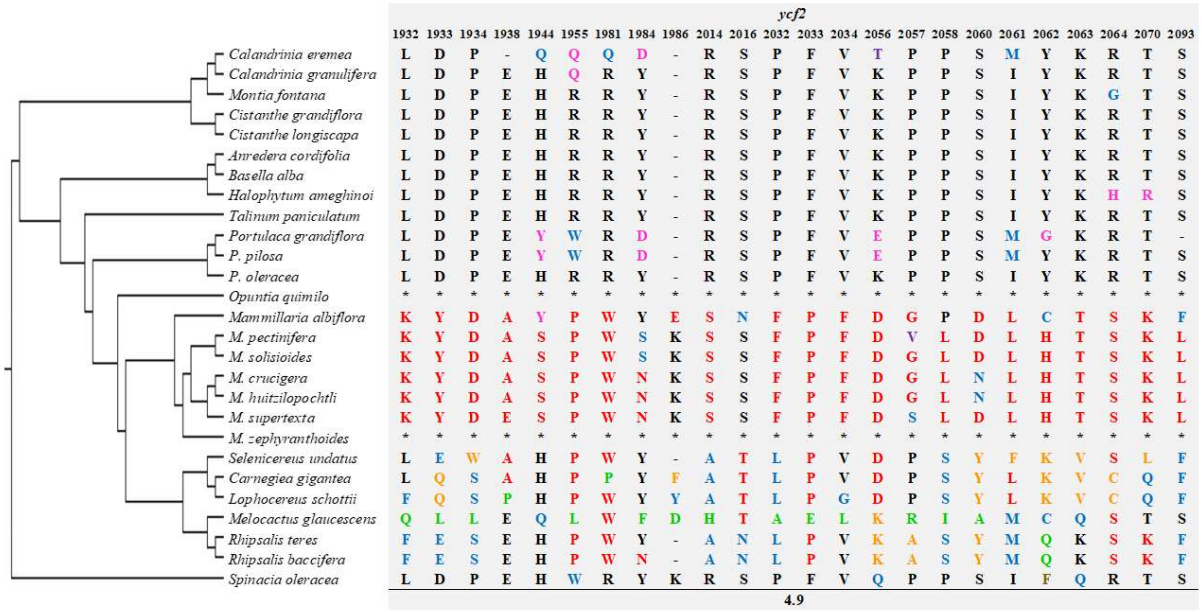
Supplementary Fig. S20. Non-synonymous substitutions in the *ycf2* gene. The amino acids are plotted across the Cactineae phylogeny inferred based on plastid genes (Figure 4). Different amino acids identified at the same position are highlighted in distinct colors. Amino acid positions are relative to *M. glaucescens* plastid genes. The Ka/Ks scores for each codon position are indicated in the last line. Absent genes are indicated by an asterisk



Supplementary Fig. S21. Non-synonymous substitutions in the *ycf2* gene. The amino acids are plotted across the Cactineae phylogeny inferred based on plastid genes (Figure 4). Different amino acids identified at the same position are highlighted in distinct colors. Amino acid positions are relative to *M. glaucescens* plastid genes. The Ka/Ks scores for each codon position are indicated in the last line. Absent genes are indicated by an asterisk



Supplementary Fig. S22. Non-synonymous substitutions in the *ycf2* gene. The amino acids are plotted across the Cactineae phylogeny inferred based on plastid genes (Figure 4). Different amino acids identified at the same position are highlighted in distinct colors. Amino acid positions are relative to *M. glaucescens* plastid genes. The Ka/Ks scores for each codon position are indicated in the last line. Absent genes are indicated by an asterisk



Supplementary Fig. S23. Non-synonymous substitutions in the *ycf2* gene. The amino acids are plotted across the Cactineae phylogeny inferred based on plastid genes (Figure 4). Different amino acids identified at the same position are highlighted in distinct colors. Amino acid positions are relative to *M. glaucescens* plastid genes. The Ka/Ks scores for each codon position are indicated in the last line. Absent genes are indicated by an asterisk

Supplementary Tables

Supplementary Table S1. List of species included in the codon usage analysis

Species	Family	GenBank
Melocactus glaucescens	Cactaceae	OK298499
Rhipsalis teres	Cactaceae	MT387452
Rhipsalis baccifera	Cactaceae	MT821847
Selenicereus undatus	Cactaceae	NC_053698
Carnegiea gigantea	Cactaceae	NC_027618
Lophocereus schottii	Cactaceae	NC_041727
Mammillaria albiflora	Cactaceae	MN517610
Mammillaria supertexta	Cactaceae	MN508963
Mammillaria zephyranthoides	Cactaceae	MN517611
Opuntia quimilo	Cactaceae	MN114084
Portulaca oleracea	Portulacaceae	NC_041264.1
Spinacia oleracea	Chenopodiaceae	NC_002202.1

Supplementary Table S2. List of species included in the analysis of Cactineae phylogeny

Species	Family	GenBank
Melocactus glaucescens	Cactaceae	OK298499
Rhipsalis teres	Cactaceae	MT387452
Rhipsalis baccifera	Cactaceae	MT821847
Selenicereus undatus	Cactaceae	NC_053698
Carnegiea gigantea	Cactaceae	NC_027618
Lophocereus schottii	Cactaceae	NC_041727
Mammillaria albiflora	Cactaceae	MN517610
M. pectinifera	Cactaceae	MN519716
M. crucigera	Cactaceae	MN517613
M. huitzilopochtli	Cactaceae	MN517612
M. solisioides	Cactaceae	MN518341
M. supertexta	Cactaceae	MN508963
M. zephyranthoides	Cactaceae	MN517611
Opuntia quimilo	Cactaceae	MN114084
Anredera cordifolia	Basellaceae	NC_041274.1
Basella alba	Basellaceae	NC_041293.1
Calandrinia eremea	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 pilosa	Portulacaceae	NC_036236.1
Portulaca oleracea	Portulacaceae	NC_041264.1
Talinum paniculatum	Talinaceae	NC_037748.1
Spinacia oleracea *	Chenopodiaceae	NC_002202.1

* Outgroup

Supplementary Table S3. List of RNA editing sites predicted by PREP program. The species analyzed belong to the subfamily Cactoideae (family Cactaceae). The abbreviation "aa pos" meaning the amino acid position of the RNA editing site in the aligned sequence

Gene	aa pos	*M. glaucescens*	*R. baccifera*	*S. undatus*	Observation based on Silva et al. (2021) Predictive analyses
atpA	305	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	Editing site conserved in subf. Cactoideae, except in *M. albiflora* (T fixed)
	424	CCC (P) => TCC (S)	CCC (P) => TCC (S)	CCC (P) => TCC (S)	Editing site conserved in subf. Cactoideae
atpB	138	ACC (T) => ATC (I)	ACC (T) => ATC (I)	ACC (T) => ATC (I)	Editing site conserved in subf. Cactoideae
atpF	31	CCA (P) => CTA (L)	CTA (L)	CTA (L)	T fixed in other species of the subf. Cactoideae
matK	217	TAT (Y)	CAT (H) => TAT (Y)	CAT (H) => TAT (Y)	Editing site conserved in subf. Cactoideae, except in *C. gigantea* and *L. schottii* (T fixed)
petB	140	CGG (R) => TGG (W)	CGG (R) => TGG (W)	CGG (R) => TGG (W)	Editing site conserved in subf. Cactoideae
psbF	26	TCT (S) => TTT (F)	TCT (S) => TTT (F)	TCT (S) => TTT (F)	Editing site in *C. gigantea* and *L. schottii*; the other species has a T fixed
psbL	1	ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	Editing site conserved in subf. Cactoideae
rpl20	20	TTC (F)	TCC (S) => TTC (F)	TCC (S) => TTC (F)	Editing site conserved in subf. Cactoideae
	83	CAC (H) => TAC (Y)	CAC (H) => TAC (Y)	CAC (H) => TAC (Y)	Editing site conserved in subf. Cactoideae, except in *M. zephyranthoides* (T fixed)
	87	TTG (L)	TCG (S) => TTG (L)	TCG (S) => TTG (L)	Editing site conserved in subf. Cactoideae
	100	TCA (S) => TTA (L)	TTT (F)	TTA (L)	T fixed in other species of the subf. Cactoideae
rpoA	9	TCC (S)	CCT (P) => TCT (S)	TCC (S)	T fixed in other species of the subf. Cactoideae
	28	CAT (H) => TAT (Y)	CAT (H) => TAT (Y)	CAT (H) => TAT (Y)	Editing site conserved in subf. Cactoideae, except in *Mammillaria* genus (T fixed)
	291	TCG (S) => TTG (L)	TTG (L)	TTG (L)	Editing site in *M. albiflora* and *M. zephyranthoides*
	188	TCT (S)	CCT (P) => TCT (S)	TCT (S)	T fixed in other species of the subf. Cactoideae
rpoB	158	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	Editing site conserved in subf. Cactoideae
	184	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	Editing site in *R. teres*, *C. gigantea*, *L. schottii*, *M. albiflora* and *M. zephyranthoides*
	189	TCG (S) => TTG (L)	TCG (S) => TTG (L)	TCG (S) => TTG (L)	Editing site conserved in subf. Cactoideae
	358	CCA (P) => TCA (S)	TCA (S)	TCA (S)	T fixed in other species of the subf. Cactoideae
rpoC1	11	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	Editing site conserved in subf. Cactoideae

	157	TTT (F)	CTT (L) => TTT (F)	TTT (F)	Editing site in *R. teres*; the other species has a T fixed
	209	ACT (T) => ATT (I)	ACT (T) => ATT (I)	ACT (T) => ATT (I)	Editing site conserved in subf. Cactoideae
	321	AGG (R)	AGG (R)	ACG (T) => ATG (M)	T fixed in other species of the subf. Cactoideae
	390	CCA (P) => TCA (S)	TCA (S)	TCA (S)	T fixed in other species of the subf. Cactoideae
	420	CCG (P) => TCG (S)	TCG (S)	TCG (S)	T fixed in other species of the subf. Cactoideae
rpoC2	163	GTA (V)	GTA (V)	CCA (P) => CTA (L)	T fixed in other species of the subf. Cactoideae
	498	ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	Editing site conserved in subf. Cactoideae
	606	TAT (Y)	CAT (H) => TAT (Y)	TAT (Y)	Editing site in *R. teres*; the other species has a T fixed
	725	CAT (H) => TAT (Y)	TAT (Y)	TAT (Y)	T fixed in other species of the subf. Cactoideae
	739	GCT (A) => GTT (V)	GTT (V)	GTT (V)	T fixed in other species of the subf. Cactoideae
	747	GTC (V)	GTC (V)	GCC (A) => GTC (V)	Editing site in *C. gigantea* and *L. schottii*; the other species has a T fixed
	1142	TCT (S) => TTT (F)	TTC (F)	TTC (F)	T fixed in other species of the subf. Cactoideae
	1311	ACG (T) => ATG (M)	ATG (M)	ATG (M)	T fixed in other species of the subf. Cactoideae
rps2	83	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	Editing site conserved in subf. Cactoideae, except in *M. albiflora* (T fixed)
rps8	48	GCG (A) => GTG (V)	GTG (V)	TCG (S)	T fixed in other species of the subf. Cactoideae
rps14	27	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	Editing site conserved in subf. Cactoideae
	50	CCA (P) => CTA (L)	CCA (P) => CTA (L)	CCA (P) => CTA (L)	Editing site conserved in subf. Cactoideae
ycf3	26	ACT (T) => ATT (I)	ATT (I)	ATT (I)	T fixed in other species of the subf. Cactoideae

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

SSR sequence	3	4	5	6	7	8	9	10	11	12	13	Total
A/T	-	-	-	-	-	58	31	13	9	-	3	114
C/G	-	-	-	-	-	3	-	-	-	-	-	3
AC/GT	-	1	-	-	-	-	-	-	-	-	-	1
AG/CT	-	14	-	-	-	-	-	-	-	-	-	14
AT/AT	-	15	4	1	5	-	1	-	-	-	-	26
AAT/ATT	-	2	-	-	-	-	-	-	-	-	-	2
AAAG/CTTT	2	-	-	-	-	-	-	-	-	-	-	2
AAAT/ATTT	2	-	1	-	-	-	-	-	-	-	-	3
AATT/AATT	1	-	-	-	-	-	-	-	-	-	-	1
AAAGT/ACTTT	1	-	-	-	-	-	-	-	-	-	-	1
AATAG/ATTCT	1	-	-	-	-	-	-	-	-	-	-	1
AAAGTT/AACTTT	1	-	-	-	-	-	-	-	-	-	-	1
Total												169

Supplementary Table S5. Distribution of SSR loci in the plastome of *M. glaucescens*

SSR type	SSR	Size	Start	End	Location
mono	(A)8	8	498	505	*ycf3* (íntron)
mono	(A)10	10	1151	1160	*ycf3* (íntron)
di	(TA)4	8	1385	1392	*ycf3* (íntron)
mono	(A)8	8	1764	1771	*ycf3* (íntron)
mono	(T)11	11	2168	2178	*ycf3/psaA* (IGS)
di	(AG)4	8	2436	2443	*ycf3/psaA* (IGS)
di	(AG)4	8	7176	7183	*rps14* (CDS)
mono	(T)9	9	7841	7849	*trnfM-CAU/trnG-GCC* (IGS)
di	(AT)7	14	7995	8008	*trnG-GCC/psbZ* (IGS)
di	(CT)4	8	8822	8829	*trnS-UGA* (CDS)
mono	(A)8	8	8923	8930	*trnS-UGA/psbC* (IGS)
mono	(A)11	11	8962	8972	*trnS-UGA/psbC* (IGS)
mono	(T)9	9	8987	8995	*trnS-UGA/psbC* (IGS)
mono	(T)8	8	11564	11571	*psbD/trnT-GGU* (IGS)
mono	(A)8	8	11781	11788	*psbD/trnT-GGU* (IGS)
mono	(A)9	9	12415	12423	*trnT-GGU/rpoB* (IGS)
mono	(T)9	9	12473	12481	*trnT-GGU/rpoB* (IGS)
mono	(A)10	10	13318	13327	*rpoB* (CDS)
mono	(A)8	8	13424	13431	*rpoB* (CDS)
di	(GA)4	8	15055	15062	*rpoB* (CDS)
mono	(T)8	8	16455	16462	*rpoC1* (CDS)
di	(AT)5	10	18889	18898	*rpoC2* (CDS)
mono	(A)8	8	19722	19729	*rpoC2* (CDS)
mono	(A)11	11	20163	20173	*rpoC2* (CDS)
di	(AT)4	8	20200	20207	*rpoC2* (CDS)
mono	(T)8	8	20984	20991	*rpoC2* (CDS)
mono	(A)9	9	22037	22045	*rpoC2* (CDS)
mono	(A)8	8	22850	22857	*rps2* (CDS)
mono	(T)11	11	24271	24281	*atpI/atpH* (IGS)
mono	(T)9	9	25040	25048	*atpF* (íntron)
mono	(A)8	8	25257	25264	*atpF* (íntron)
mono	(T)8	8	25344	25351	*atpF* (íntron)
mono	(A)10	10	25426	25435	*atpF* (íntron)
mono	(T)8	8	25505	25512	*atpF* (íntron)
mono	(T)9	9	26141	26149	*atpF/atpA* (IGS)
mono	(A)9	9	28424	28432	*trnG-UCC* (íntron)
mono	(A)10	10	28501	28510	*trnG-UCC* (íntron)
mono	(T)10	10	28820	28829	*trnG-UCC/clpP* (IGS)
mono	(A)8	8	29311	29318	*clpP* (CDS)
di	(AT)4	8	29766	29773	*clpP/trnS-GCU* (IGS)
di	(TA)4	8	29805	29812	*clpP/trnS-GCU* (IGS)

di	(AT)7	14	29991	30004	*trnS-GCU/psbI* (IGS)
mono	(A)10	10	31224	31233	*trnQ-UUG/cemA* (IGS)
mono	(T)8	8	31287	31294	*trnQ-UUG/cemA* (IGS)
tri	(AAT)4	12	31346	31357	*trnQ-UUG/cemA* (IGS)
mono	(A)11	11	31359	31369	*cemA* (CDS)
di	(AT)4	8	32379	32386	*petA* (CDS)
mono	(A)8	8	32739	32746	*petA* (CDS)
mono	(T)11	11	33314	33324	*petA/psbJ* (IGS)
mono	(T)8	8	33335	33342	*petA/psbJ* (IGS)
mono	(T)8	8	33803	33810	*psbJ/psbL* (IGS)
di	(AT)5	10	35783	35792	*trnP-UGG/psaJ* (IGS)
di	(TA)4	8	35794	35801	*trnP-UGG/psaJ* (IGS)
mono	(T)8	8	36299	36306	*psaJ/rpl33* (IGS)
mono	(T)8	8	36661	36668	*rpl33/rps18* (IGS)
mono	(A)8	8	36883	36890	*rps18* (CDS)
mono	(A)10	10	37248	37257	*rps18/rpl20* (IGS)
mono	(T)8	8	37628	37635	*rpl20* (CDS)
mono	(T)9	9	38309	38317	*rpl20/rps12* (IGS)
di	(AT)7	14	38463	38476	*rps12/clpP* (IGS)
mono	(T)8	8	40455	40462	*psbB/psbT* (IGS)
mono	(T)8	8	41599	41606	*petB* (íntron)
mono	(A)9	9	41801	41809	*petB* (íntron)
penta	(TTCTA)3	15	41971	41985	*petB* (íntron)
mono	(A)10	10	42928	42937	*petB/petD* (IGS)
di	(AG)4	8	43063	43070	*petD* (íntron)
mono	(T)8	8	43185	43192	*petD* (íntron)
di	(TA)4	8	43354	43361	*petD* (íntron)
mono	(T)9	9	44522	44530	*rpoA* (CDS)
di	(GA)4	8	45202	45209	*rpoA* (CDS)
mono	(T)8	8	45332	45339	*rpoA/rps11* (IGS)
mono	(T)8	8	45990	45997	*rpl36* (CDS)
mono	(T)8	8	46360	46367	*infA* (CDS)
mono	(T)11	11	46949	46959	*rps8/rpl14* (IGS)
tetra	(TTTC)3	12	48546	48557	*rpl16* (íntron)
mono	(T)9	9	49458	49466	*rps3* (CDS)
mono	(T)9	9	49726	49734	*rpl22* (CDS)
mono	(A)9	9	50195	50203	*rpl22/rps19* (IGS)
mono	(T)8	8	50312	50319	*rps19* (CDS)
mono	(T)8	8	50606	50613	*rps19/rpl2* (IGS)
di	(TA)5	10	52087	52096	*trnS-GGA/rps4* (IGS)
mono	(T)8	8	53084	53091	*rps4/trnT-UGU* (IGS)
mono	(A)8	8	53174	53181	*rps4/trnT-UGU* (IGS)
mono	(A)8	8	53194	53201	*rps4/trnT-UGU* (IGS)

mono	(A)9	9	53678	53686	*trnL-UAA* (íntron)
mono	(A)9	9	53717	53725	*trnL-UAA* (íntron)
mono	(A)8	8	53813	53820	*trnL-UAA* (íntron)
di	(TA)7	14	53881	53894	*trnL-UAA* (íntron)
di	(AT)4	8	53900	53907	*trnL-UAA* (íntron)
di	(AT)4	8	53951	53958	*trnL-UAA* (íntron)
mono	(T)8	8	54285	54292	*trnL-UAA/trnF-GAA* (IGS)
mono	(C)8	8	54863	54870	*trnF-GAA/ndhJ* (IGS)
mono	(A)8	8	55443	55450	*ndhJ/trnD-GUC* (IGS)
di	(TA)9	18	55672	55689	*trnD-GUC/psbM* (IGS)
mono	(T)8	8	56302	56309	*psbM/petN* (IGS)
mono	(A)10	10	57178	57187	*petN/trnC-GCA* (IGS)
mono	(T)8	8	57402	57409	*petN/trnC-GCA* (IGS)
di	(AT)7	14	57613	57626	*trnC-GCA/trnE-UUC* (IGS)
mono	(A)9	9	58024	58032	*trnC-GCA/trnE-UUC* (IGS)
mono	(T)8	8	58541	58548	*trnY-GUA/ndhC* (IGS)
mono	(A)8	8	58580	58587	*trnY-GUA/ndhC* (IGS)
di	(AT)4	8	58784	58791	*ndhC/rbcL* (IGS)
di	(TC)4	8	59886	59893	*rbcL* (CDS)
di	(AT)4	8	60561	60568	*rbcL/atpB* (IGS)
di	(AT)4	8	60668	60675	*rbcL/atpB* (IGS)
mono	(A)10	10	61038	61047	*atpB* (CDS)
tri	(ATA)4	12	62952	62963	*atpE/trnM-CAU* (IGS)
mono	(A)8	8	64829	64836	*accD* (CDS)
di	(TA)4	8	65688	65695	*accD* (CDS)
tetra	(ATAA)3	12	65962	65973	*accD/psaI* (IGS)
mono	(A)8	8	66053	66060	*accD/psaI* (IGS)
mono	(G)8	8	66357	66364	*psaI/ycf4* (IGS)
mono	(C)8	8	66676	66683	*ycf4* (CDS)
mono	(T)11	11	66690	66700	*ycf4* (CDS)
mono	(T)8	8	67130	67137	*ycf4/rps16* (IGS)
di	(AT)4	8	67181	67188	*ycf4/rps16* (IGS)
di	(TA)5	10	67646	67655	*rps16* (íntron)
mono	(A)8	8	67771	67778	*rps16* (íntron)
mono	(A)8	8	67894	67901	*rps16* (íntron)
mono	(A)11	11	68384	68394	*rps16* (CDS)
mono	(T)8	8	68487	68494	*rps16/trnK-UUU* (IGS)
di	(TA)4	8	68603	68610	*rps16/trnK-UUU* (IGS)
mono	(A)9	9	68999	69007	*trnK-UUU/matK* (IGS)
mono	(T)8	8	69337	69344	*matK* (CDS)
mono	(A)9	9	69720	69728	*matK* (CDS)
mono	(T)8	8	70366	70373	*matK* (CDS)
mono	(A)9	9	70950	70958	*matK/trnK-UUU* (IGS)

mono	(T)10	10	72592	72601	*psbA/trnH-GUG* (IGS)
mono	(T)10	10	72676	72685	*psbA/trnH-GUG* (IGS)
mono	(T)9	9	72921	72929	*trnH-GUG/trnI-CAU* (IGS)
tetra	(ATAA)5	20	73206	73225	*trnI-CAU/ycf2* (IGS)
di	(GA)4	8	73344	73351	*ycf2* (CDS)
di	(GA)4	8	73356	73363	*ycf2* (CDS)
di	(GA)4	8	74265	74272	*ycf2* (CDS)
mono	(A)9	9	76132	76140	*ycf2* (CDS)
di	(GA)4	8	76153	76160	*ycf2* (CDS)
mono	(A)9	9	80049	80057	*trnL-CAA/ycf1* (IGS)
mono	(A)10	10	80879	80888	*ycf1* (CDS)
mono	(A)13	13	81467	81479	*ycf1* (CDS)
mono	(A)8	8	81596	81603	*ycf1* (CDS)
mono	(A)9	9	81680	81688	*ycf1* (CDS)
mono	(T)8	8	82873	82880	*ycf1* (CDS)
mono	(A)8	8	83789	83796	*ycf1* (CDS)
mono	(A)9	9	83814	83822	*ycf1* (CDS)
tetra	(AATT)3	12	83845	83856	*ycf1* (CDS)
mono	(A)8	8	83900	83907	*ycf1* (CDS)
mono	(A)8	8	84482	84489	*ycf1* (CDS)
mono	(A)13	13	84814	84826	*ycf1* (CDS)
mono	(A)9	9	85805	85813	*rps15* (CDS)
mono	(T)9	9	86490	86498	*ndhH/trnL-UAG* (IGS)
mono	(T)9	9	87899	87907	*ccsA/ndhD* (IGS)
tetra	(TTTC)3	12	88570	88581	*ndhD* (CDS)
mono	(A)9	9	89133	89141	*ndhD/psaC* (IGS)
mono	(T)8	8	89443	89450	*psaC/rpl32* (IGS)
penta	(AGTAA)3	15	89478	89492	*psaC/rpl32* (IGS)
tetra	(TAAA)3	12	89565	89576	*psaC/rpl32* (IGS)
mono	(T)13	13	89676	89688	*rpl32* (CDS)
mono	(T)8	8	89815	89822	*rpl32* (CDS)
di	(AT)6	12	90046	90057	*rpl32/trnN-GUU* (IGS)
mono	(T)9	9	90232	90240	*rpl32/trnN-GUU* (IGS)
di	(CA)4	8	90519	90526	*trnN-GUU/trnR-ACG* (IGS)
mono	(A)8	8	90720	90727	*trnR-ACG/rrn5* (IGS)
di	(GA)4	8	91036	91043	*rrn5/rrn4.5* (IGS)
di	(AG)4	8	92478	92485	*rrn23* (CDS)
hexa	(CTTTAA)3	18	94200	94217	*rrn23/trnI-GAU* (IGS)
mono	(A)9	9	94831	94839	*trnI-GAU* (intron)
mono	(A)8	8	98598	98605	*rps7* (CDS)
mono	(T)8	8	99148	99155	*rps7/ndhB* (IGS)
di	(TC)4	8	100127	100134	*ndhB/ycf1* (IGS)

CDS, coding sequences; IGS, intergenic spacers

3. CHAPTER II

The plastomes of *Lepismium cruciforme* (Vell.) Miq and *Schlumbergera truncata* (Haw.) Moran reveals tribe-specific rearrangements and the first loss of *trnT-GGU* gene in the family Cactaceae

Tanara P. Dalla Costa [1], Maria C. Silva [1], Amanda de Santana Lopes[1], Túlio Gomes Pacheco[1], Gleyson Morais da Silva[1], José D. de Oliveira[1], Valter A. de Baura[2], Eduardo Balsanelli[2], Emanuel Maltempi de Souza[2], Fábio de Oliveira Pedrosa[2], Marcelo Rogalski[1*]

[1] Laboratório de Fisiologia Molecular de Plantas, Departamento de Biologia Vegetal, Universidade Federal de Viçosa, Viçosa-MG, Brasil.
[2] Núcleo de Fixação Biológica de Nitrogênio, Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Paraná, Curitiba-PR, Brasil.

*Corresponding author
E-mail address: rogalski@ufv.br

Manuscript in preparation to be submitted to **Planta**

Abstract
Recent studies have revealed atypical features in the plastomes of the family Cactaceae, the largest lineage of succulent species adapted to arid and semi-arid regions. However, few plastomes are available to date, and many genera do not have any sequenced species, including *Lepismium* and *Schlumbergera*. Aiming to contribute with new insights into the plastid evolution of Cactaceae, we sequenced and analyzed the plastomes of *Lepismium cruciforme* and *Schlumbergera truncata*, two South American epiphytic cacti (Rhipsalideae tribe). Our data revealed many gene losses in both plastomes and the first loss of functionality of the *trnT-GGU* gene in the family Cactaceae. The *trnT-GGU* gene was pseudogenized in *L. cruciforme* plastome and appears to be degenerating in the tribe Rhipsalideae. Although the plastomic structure is conserved among the species of the tribe Rhipsalideae, with tribe-specific rearrangements, we identified nine nucleotide polymorphism hotspots, useful to improve the phylogenetic resolutions of Rhipsalideae. Furthermore, our analysis indicated high gene divergence and rapid evolution of RNA editing sites in plastid protein-coding genes from Cactaceae. Finally, we mapped repetitive sequences from the plastomes of *L. cruciforme* and *S. truncata*, which can be used to access natural populations and assist in conservation strategies for both species.

Keywords: Atlantic Forest, Rhipsalideae, Plastome evolution, Gene divergence, Polymorphism hotspots, Plastid SSRs

Introduction

Chloroplasts are responsible, among other functions, for the process of photosynthesis, therefore, they are essential for the maintenance of life on Earth (Stirbet et al. 2020). These organelles of endosymbiotic origin have an independent genome called the plastid genome or plastome (Allen 2015). This genome usually is represented as a circular DNA molecule, with a quadripartite structure in two single copy regions (large single copy - LSC; and small single copy - SSC), separated by two inverted repeat regions (IRA and IRB) (Pacheco et al. 2020a; b; Lopes et al. 2021; Silva et al. 2021). Unique features such as relatively small molecules, multiple copies per organelle, mutation rate lower than the nuclear genome, uniparental inheritance, and non-recombinant nature make the plastome an excellent object for different genetic studies (Bock 2007; Wicke et al. 2011; Rogalski et al. 2015; Daniell et al. 2016).

Of the more than 400,000 recognized land plants (Lugadha et al. 2016), only 8,000 have their complete plastid genome available in NCBI database (National Center for Biotechnology Information; https://www.ncbi.nlm.nih.gov/). Therefore, a lot of new information can be discovered by plastid genomics. In angiosperms, for example, although it has been postulated that gene structure and content are generally conserved (Wicke et al. 2011; Daniell et al. 2016), atypical events have been revealed in several lineages (Chumley et al. 2006; Park et al. 2015;

Moghaddam and Kazempour-Osaloo 2020; Alqahtani and Jansen 2021; Fu et al. 2021; Scobeyeva et al. 2021). A lineage that is extremely unusual in terms of structure and plastome genomic content is the family Cactaceae. Intensely rearranged plastomes, different configurations of IR regions, intron losses, massive gene losses, unusual gene duplications, gene divergence, polymorphic RNA editing sites, and positive selection signatures are some characteristics previously reported (Sanderson et al. 2015; Solórzano et al. 2019; Oulo et al. 2020; Silva et al. 2021; Almeida et al. 2022; Dalla Costa et al. 2022).

All these discoveries allowed new insights into the evolution of the family Cactaceae, indicating that such characteristics imposed on the plastomes allowed them to adapt and diversify to the hostile environments they inhabit (Silva et al. 2021). This family represents the largest lineage of succulent plants, rich in species diversity, which are anatomically, morphologically, and physiologically adapted to arid and semi-arid regions (Anderson 2001; Arakaki et al. 2011; Griffiths and Males 2017; Guerrero et al. 2019). However, only 21 of the more than 1400 species (Guerrero et al. 2019) have the complete plastome sequences available. Many genera do not have any plastome sequenced yet, as is the case with *Lepismium* Pfeiff. and *Schlumbergera* Lem., for example. Both genera belong to the tribe Rhipsalideae, the largest group of epiphytic cacti, which occur in tropical and subtropical humid South American forests (Anderson 2001; Calvente et al. 2011a; b), with the Brazilian Atlantic Forest as their center of diversification (Hunt et al. 2006; Calvente et al. 2011).

In this work, we sequenced and analyzed the plastomes of *L. cruciforme* (Vell.) Miq. and *S. truncata* (Haw.) Moran, two epiphytic cacti of ornamental importance. Our data revealed conserved plastomic structure among species of the tribe Rhipsalideae and tribe-specific rearrangements, with some variations at the IR region boundaries. We report here the extensive loss of genes from these plastomes, as well as the first loss of functionality of the *trnT-GGU* gene in the family Cactaceae. Furthermore, we inferred the phylogenetic position of both species within the family Cactaceae, identified nucleotide polymorphism hotspots of the tribe, analyzed gene divergence, predicted new putative RNA editing sites, and mapped repetitive sequences in the plastomes of *L. cruciforme* and *S. truncata*.

Materials and methods

Plant material, chloroplast isolation, and plastid DNA extraction

The plant material was obtained from plants of *Lepismium cruciforme* and *Schlumbergera truncata* cultivated under greenhouse conditions. Fresh and young cladodes were collected and stored for five days in a dark environment at a temperature of 4 °C to reduce starch levels. Then, chloroplast isolation and plastid DNA extraction was carried out as previously described by Vieira et al. (2014).

Sequencing, assembly, annotation, and data archiving statement

The sequencing libraries were prepared with approximately 1 ng of chloroplast DNA using the sample preparation kit NexteraXT (Illumina Inc., San Diego, CA, USA), according to the manufacturer's instructions. The library was sequenced using the Illumina MiSeq platform (Illumina Inc., San Diego, CA, USA) at the Federal University of Paraná, Brazil. The readings were trimmed under the threshold with a probability of error < 0.05 and de novo assembled in contigs using the CLC genomics Workbench 8.0.2 software (CLC Bio, Aarthus, Denmark). We performed three sequencings for *L. cruciforme*, with a total of 334,196 to 521,463 paired reads, and average length between 182.7 and 198.7. The contigs used to assemble the *L. cruciforme* plastome ranged from 650.53 to 137.06 of average coverage. One gap from 358 bp was closed using NOVOPlasty 4.0 software (Dierckxsens et al. 2017) and a pre-assembled genome. To assemble the *S. truncata* plastome, we used one sequencing with a total of 499,582 paired reads and 189.28 average length. The contigs used for assembling the *S. truncata* plastome varied from 489.33 to 107.99 of average coverage. Three gaps of 5,984 bp, 4,130 bp, and 134 bp also were resolved through the NOVOPlasty 4.0 software, using a second sequencing and the pre-assembled genome.

To close one gap of 403 bp in the *S. truncata* plastome, we carried out a PCR amplification using the following pair of primers: 5′GTCTTTCATTAGTCAATTCTTGACTGA-3′ and 5′TTACACGATTCCATAGAGTTTTTGTTTAT-3′. In 25 µl reactions, 50 ng total genomic DNA was amplified in a reaction mixture containing 200 µM of each dNTP, 2 mM MgCl2, 10 pmol of each primer and 1U Platinum™ Taq DNA Polymerase (Invitrogen). The standard PCR program was 40 cycles of 30 s at 94 °C, 40 s at 61 °C and 45 s at 72 °C, with 5 min extension of the first cycle at 94 °C and 5 min final extension at 72 °C. PCR products were analyzed by electrophoresis in 1,5% agarose gels. For sequencing, PCR products were purified by electrophoresis on 1,5% agarose gels and subsequent extraction from gel slices using the Quick Gel Extraction & PCR Purification Combo Kit (Invitrogen). The samples were

sequenced by the Sanger method (sequencer ABI 3500xl) by GoGenetic Company (www.gogenetic.com.br).

The genes were annotated initially by the Annotation of Organellar Genomes (GeSeq) (Tillich et al. 2017) and BLAST programs. Subsequently, initiation, termination, and intron codon positions were determined based on alignment with plastid genomes of related species available in the GenBank database. The gene losses were verified through a search along the plastomes, using the coding sequences of the respective functional genes of the *Nicotiana tabacum*, *Spinacia oleracea*, and *Portulaca oleracea* species, whose plastomes are conversed. The tRNA sequences were analyzed using the tRNAscan-SE webserver (Lowe and Chan 2016). The physical maps of the plastomes were drawn using the online software Organellar Genome DRAW (OGDRAW) (Greiner et al. 2019). The complete nucleotide sequences of plastomes were deposited in the GenBank database under accession ON005622 for *L. cruciforme* and ON005623 for *S. truncata.*

Comparative analyses of plastome structure

To analyze the structural characteristics of the *L. cruciforme* and *S. truncata* plastomes, we performed multiple alignments using the Mauve Genome Alignment v.2.4.0 (MAUVE) software (Darling et al. 2004). For this, we used the complete sequence of plastomes, excluding only the IRa region. The comparative analysis included species *Rhipsalis baccifera*, *Rhipsalis teres* and *Pereskia aculeata* (family Cactaceae), and *Portulaca oleracea*, used as reference. Linear maps of plastid genes were drawn by OGDRAW (Greiner et al. 2019).

Comparative analyses of the secondary structure of *trnT*-GGU and Codon Usage

Pseudogenization of the *trnT*-GGU gene in the plastome of *L. cruciforme* was confirmed by annotation using tRNAscan-SE v.2.0.744. Structural variations within *trnT*-GGU were analyzed by alignment through the clustalW (Thompson et al. 1994) implemented in Mega7.0 (Kumar et al. 2016) and manually inspected. We compared the secondary structure of the plastid *trnT*-GGU of *L. cruciforme* with 24 species from different taxonomic groups. Among them, three other species of the tribe Rhipsalideae (*S. truncata*, *R. baccifera*, and *R. teres*), 11 species of cacti, dicots, monocots, gymnosperm, pteridophyte, and marcanthiophyta (Supplementary Table S1).

To analyze the effect of trnT-GGU pseudogenization on protein-coding gene sequences, we performed codon usage analysis. For that, we sampled sequences of all genes encoding

functional proteins from 21 species of the family Cactaceae, *P. oleracea* and *Spinacia oleracea*, listed in Supplementary Table S2. The frequencies were determined by the Codon Usage web server (http://www.geneinfinity.org/sms/smscodonusage.html).

Phylogenetic reconstruction

The *L. cruciforme* and *S. truncata* phylogenetic positions within the suborder Cactineae were inferred based on 54 concatenated plastid genes. Thirty nine taxons belonging to the suborder Cactineae were sampled, including representative species of Cactaceae, Basellaceae, Montiaceae, Halophytaceae, Portulacaceae, and Talinaceae families. *S. oleracea* (Chenopodiaceae: Caryophyllales) was used as outgroup species. All complete plastome sequences sampled are available in GenBank and listed in Supplementary Table S3. A total of 54 protein-coding genes were extracted from each species and individually aligned using Muscle software (Edgar 2004) implemented in Mega 7.0 (Kumar et al. 2016). Subsequently, the gene sequences were concatenated using DnaSP v.6.12.03 software (Rozas et al. 2017). The phylogenetic inference based on maximum likelihood (ML) method was performed using IQTREE v.1.6.12 (Nguyen et al. 2015). The branch supports were evaluated using 500 non-parametric bootstrap replications, which grouped the genes into six different partitions each partition with genes sharing the best replacement evolutionary models: TVM+F+I+G4 (*atpA*, *atpB*, *atpE*, *atpF*, *ccsA*, *cemA*, *petA*, *psaI*, *psbH*, *psbJ*, *psbK*, *psbT*, *psbZ*, *rbcL*, *rpl14*, *rpoA*, *rpoB*, *rpoC1*, *rps4*, *rps14*, *ycf3*); GTR+F+I+G4 (*atpH*, *atpI*, *petB*, *petD*, *petG*, *petN*, *psaA*, *psaB*, *psaC*, *psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbL*, *psbN*); GTR+F+G4 (*clpP*, *infA*, *matK*, *psbM*); TVM+F+I+G4 (*rpl2*, *rpoC2*, *rps2*, *rps3*, *rps7*, *rps8*, *rps11*, *rps12*, *rps15*, *rps19*); K3Pu+F+G4 (*rpl16*); TVM+F+G4 (*rpl22*). The same dataset was used for carried out Bayesian inference (BI), using Partition Finder version 1.1.1 (Lanfear et al. 2012) to estimate the best evolutionary model for each gene. The genes were grouped in two models, being then GTR+I+G (*atpA, atpB, atpE, atpF, atpH, atpI, ccsA, cemA, clpP, matK, petA, petB, petD, petG, petN, psaA, psaB, psaC, psaI, psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbJ, psbK, psbL, psbN, psbT, psbZ, rbcL, rpl14, rpl16, rpl2, rpoA, rpoB, rpoC1, rpoC2, rps2, rps3, rps4, rps7, rps8, rps11, rps12, rps14, rps15, ycf3*) and GTR+G (*infA*, *psbM*, *rps19*, *rpl22*). Posteriorly, the BI was constructed by the MrBayes version 3.2 software (Ronquist et al. 2012), with one million generations of two runs of four Markov Chains (three hot and one cold per run). Finally, the consensus trees were visualized using the software FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Gene divergence analysis in plastomes of Cactaceae

We aligned 59 protein-coding genes found in all 28 plastomes of the Cactaceae family, including *L. cruciforme* and *S. truncata*, and eight genes absent in one or more of these species, using the software Muscle implemented in Mega 7.0 (Edgar 2004; Kumar et al. 2016). The 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 above mentioned for the phylogenetic inference. The gene divergence was estimated by the sum of total branch lengths that link the operational taxonomical units to the common ancestor of the species sampled here.

Sliding window analysis

Nucleotide polymorphism hotspots from the tribe Rhipsalideae (i.e., *L. cruciforme*, *S. truncata*, *R. baccifera*, and *R. teres*) were inferred by sliding window analysis. Firstly, we aligned six different combinations of whole plastomes from the four species mentioned above, excluding the IRa region, through the MAFFT web server (https://mafft.cbrc.jp/alignment/server/; Katoh et al. 2019). Then, we performed the analysis of the sliding window, with window length adjustment set to 300 bp and the step sizes at 75 bp, using the DnaSP v.6.12.03 software (Rozas et al. 2017).

Prediction of RNA editing sites and SSRs

Putative RNA editing sites in genes encoding plastid proteins were identified using the Predictive RNA Editor for Plants (PREP) program (Mower 2009). In addition to *L. cruciforme* and *S. truncata*, we analyzed the potential RNA editing sites in eight other cactus species (*Ferocactus latispinus*, *F. setispinus*, *Echinocactus grusonii*, *Leuchtenbergia principis*, *Hylocereus polyrhizus*, *Selenicereus anthonyanus*, *S. grandiflorus*, and *S. validus*). The data obtained were compared with sites previously identified in other species of the Cactaceae family by Silva et al. (2021) and Dalla-Costa et al. (2022). A total of 27 genes were analyzed, with a cut-off value set at 0.8. Missing genes (*ndhA*, *ndhF*, and *ndhG*), pseudogenes (*rpl23*, *ndhB*, and *ndhD*), and highly divergent genes (*accD* and *clpP*) were excluded from this predictive analysis.

Simple sequence repeats (SSRs) were mapped using the webserver MIcroSAtellite (MISA) Identification (https://webblast.ipk-gatersleben.de/misa/), 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.

Results

Gene content and structure of *Lepismium cruciforme* and *Schlumbergera truncata* plastomes

The *L. cruciforme* and *S. truncata* plastomes are typical circular DNA molecules of 123,113 bp and 119,945 bp in length, respectively, with two single-copy regions (LSC and SSC) separated by two inverted repeats regions (IRs - IRa and IRb) (Fig. 1 and 2). Both species present 99 unique genes, of which 11 genes contain one intron, one gene harbor two introns (*ycf3*), and three genes lose their introns (*clpP, rpl2,* and *rpoC1*) (Table 1). The *L. cruciforme* plastome presents 67 protein-coding, 28 tRNAs, and four rRNAs genes. The *S. truncata* plastome exhibit 66 protein-coding, 29 tRNAs, and four rRNAs genes. The two species shared six gene losses (i.e., *trnV-UAC*, *ndhA*, *ndhE*, *ndhG*, *ndh,* and *ndhI*) and six pseudogenes (i.e., *ndhB, ndhC, ndhD, ndhF, ndhK,* and *rpl23*). In addition, *L. cruciforme* lost the *rpl33* gene and *trnT-GGU* is a pseudogene. In *S. truncata* plastome, the *ycf4* is a putative pseudogene due to the presence of a premature stop codon. The gene content and general features from both plastomes are shown in Table 1 and Supplementary Table S4.

To analyze the structure of the plastomes, we performed multiple alignments by MAUVE (Supplementary Fig. S1). Here we compared the whole plastomes, excluding the IRa region, of *Schlumbergera truncata*, *Lepismium cruciforme, Rhipsalis baccifera, Rhipsalis teres, Pereskia aculeata* (Cactaceae family), and *Portulaca oleracea* as reference. In this analysis, the plastomes were delimited in 10 locally collinear blocks (LCBs), which refers to conserved regions among species. Events of inversion, translocation, expansion and contraction of the IR boundaries were identified in *L. cruciforme* and *S. truncata* plastomes.

To visualize in detail the rearrangements, was carried out a comparative analysis with linear genetic maps of the species plastomes mentioned above (Fig. 3). Three initial events occurred in the *P. aculeata* plastome, a primitive species of the family Cactaceae. A primary inversion of the LCBs D and E in the LSC region, which harbors the gene segment from *trnV-UAC* to *rbcL* genes, changed the order and direction of genes (event 1; Fig. 3). In the SSC region, the block of genes from *ccsA* to *ycf1* (LCB J) suffered an inversion (event 2; Fig. 3), so that the *ycf1* gene fragment that was partially in the IR of *P. oleracea* was lost (red squared 1;

Fig. 3). In addition, *P. aculeata* plastome presented a contraction of the IRb/SSC border, culminating in the expansion of the SSC region in comparison with *P. oleracea* (event 3; Fig. 3).

Concerning the rearrangements of the tribe Rhipsalideae, (i.e., *L. cruciforme*, *S. truncata*, *R. teres,* and *R. baccifera*) inversion of two segments of the region LSC was observed. One inversion occurred in the LCB B, harboring the genes from *trnT-GGU* to *ndhJ* (event 4; Fig. 3). The other event refers to the secondary inversion of the LCB E, which restored the initial direction of the *rbcL* gene found in the *P. oleracea* plastome (event 5; Fig. 3). In the SSC region, there were inversion and translocation of the gene segment, that comprehends the genes from *ndhB* to *rpl32* (LCB G), changing the order and direction of genes (event 6; Fig. 3).

Despite the similarity among the plastomes of the tribe Rhipsalideae, there are some variations in the size and gene content of the IR region. The Rhipsalideae tribe presented an event of initial expansion of the IRb/SSC border, incorporating part of the *ycf1* gene and expanding the IR region, in comparison with *P. aculeata* (event 7; Fig. 3). This expansion provided *L. cruciforme* with the largest IR region of the tribe, with 9.675 bp in size, harboring *rps19*, *rpl2*, *rpl23*, *trnI-CAU*, *ycf2*, *trnL-CAA*, and *ycf1* (fragment) genes. On the other hand, an event of the expansion of the LSC/IRb border in the *S. truncata* plastome encompassed the *rpl2* and *rpl23* genes in the LSC region (event 8; Fig. 3), leading to the reduction of its IR region, which is the smallest IR of the tribe (8,380 bp), and contain the *trnI-CAU*, *ycf2*, *trnL-CAA* genes and fragments of the genes *ycf1* and *rps19*. Finally, both *Rhipsalis* species also underwent LSC/IRb border expansion, reducing their gene content to four genes (*trnI-CAU*, *ycf2*, *trnL-CAA*, and *ycf1* fragment). The gene losses and pseudogenes of *L. cruciforme* and *S. truncata* are highlighted by red squares and asterisks, respectively, and described according to the indicated numbering in Figure 3.

Comparative analyses of *trnT-GGU* gene and Codon Usage

The threonine amino acid is transported in plastids by two tRNAs, $tRNA^{Thr}$(UGU) and $tRNA^{Thr}$(GGU), encoded by *trnT-UGU* and *trnT-GGU* genes, respectively. However, the *trnT-GGU* is a pseudogene in the *L. cruciforme* plastome. By analyzing the *trnT-GGU* sequence, we identified six mutations, being three transitions (1, 30 and 32 positions), two transversions (2 and 28 positions), and one insertion (67 position) of nucleotides (Supplementary Fig. S2). Some of these mutations also occurred in the other three species of the tribe Rhipsalideae, with *R. baccifera* having five mutations and *R. teres* and *S. truncata* having three alterations each. The

transitions in positions 1, 30 and 32 are common to the tribe, except for position 1, which is conserved in *R. teres*. A transversion at position 36 is shared between the two *Rhipsalis* species, and another transversion at position 42 occurs only in *R. baccifera* (Supplementary Fig. S2).

A comparative analysis of the secondary structure of *trnT-GGU* was carried out among Rhipsalideae species and *Spinacia oleracea* (Fig. 4) and other 20 species of different taxonomic groups, including other cactus, dicots, monocots, marcanthiophyta, pteridophyte, and gymnosperm species (Supplementary Fig. S3). All the species analyzed here presented the perfect cloverleaf structure, except for the four species of the tribe Rhipsalideae. In these species, mutations occurred mainly in the acceptor stem and the anticodon loop, resulting in mismatched regions (Fig. 4). The secondary structures of Rhipsalideae, predicted by the webserver tRNAscan-SE, showed lower infernal values than *S. oleracea* (61.3), which ranged from 28.9 (*L. cruciforme*) to 44.0 (*R. teres*). The infernal value reflects the accuracy of the predicted tRNAs and their functional classification, compared to tRNAs available in databases. Therefore, the low infernal value of *trnT-GGU* in *L. cruciforme* reinforces its pseudogenization. On the other hand, the *trnT-GGU* gene appears to be functional in *S. truncata*, *R. baccifera* and *R. teres.*

To analyze the effect of *trnT-GGU* pseudogenization on protein-coding gene sequences of *L. cruciforme*, we compared the codon usage frequency of this species with 12 species of the family Cactaceae, *P. oleracea*, and *S. oleracea* (Fig. 5). The threonine amino acid is encoded by four codons (i.e., ACA, ACC, ACG, and ACT), which codon usage frequencies were very similar among *L. cruciforme* and other species analyzed here. The codon most frequent for threonine amino acid in the *L. cruciforme* plastome was the ACT (0.41), followed by ACA (0.30), while ACC and ACG codons were the less frequent (0.18 and 0.10, respectively). Similar frequencies were verified in the mean of the codon usage of cacti, being them 0.41 (ACT), 0.29 (ACA), 0.19 (ACC), and 0.10 (ACG). Therefore, the pseudogenization of the *trnT-GGU* gene did not change the codon usage frequency in *L. cruciforme* plastome.

Phylogeny of the suborder Cactineae based on plastid genes

To infer the position of *L. cruciforme* and *S. truncata* within the family Cactaceae, we performed a phylogenetic analysis based on concatenated plastid genes. For that, we updated our Cactaceae phylogeny previously published (Dalla-Costa et al. 2022) with the addition of 14 species of the family. A consensus tree was produced by the maximum likelihood (ML) method, with a log-likelihood (lnL) of - 140,997.1611 (Figure 6). The relationships among the families

and the species within subfamilies followed our previous results (Dalla-Costa et al. 2022), except in the subfamily Halophytaceae position. Within the family Cactaceae, *Pereskia aculeata* constituted the early branching, forming a sister group with Opuntioideae and Cactoideae subfamilies. The subfamily Opuntioideae configured a clade well-supported (100% of ML-BS), in which *Opuntia monacantha* exhibited the longest branch, being the species most divergent of the subfamily. Concerning the subfamily Cactoideae, two clades were formed (100% of ML-BS): the first formed by the tribe Cacteae (*Mammillaria*, *Echinocactus*, *Ferocactus*, and *Leuchtenbergia* genres), and the second by the Core Cactoideae. In the Core Cactoideae, the *Selenicereus* genus formed a sister group with *C. gigantia* and *L. schottii* (tribe Echinocereeae), and *M. glaucescens* formed a sister group with Rhipsalideae tribe, both with high support of ML-BS (100 and 99%, respectively). Curiously, within the tribe Rhipsalideae, the two species of *Rhipsalis* (*R. teres* and *R. baccifera*) did not group together. *R. baccifera* formed a sister group with *S. truncata*, and *R. teres* formed a sister group with *L. cruciforme*, which relationships were highly supported (100% of ML-BS). To certify this result, a Bayesian inference (BI) analysis was carried out with the same dataset. A Bayesian phylogenetic tree was generated, with a - lnL = 160,149.4898 (Supplementary Fig. S6). The BI presented a similar pattern of distribution of the families and species within the families, except for *H. ameghinoi* and *S. anthonyanus* positions. However, the same relationship within the tribe Rhipsalideae was observed, with 100% of the posterior probability.

Hotspots of nucleotide diversity in the plastomes of the tribe *Rhipsalideae*

Because the two species *Rhipsalis* were not grouped together in our phylogeny (Fig. 6), we decided to characterize the plastid regions of rapid evolution among the four species of the tribe Rhipsalideae. For this, we performed sliding window analysis for six different combinations of whole plastome alignments (i.e., *L. cruciforme* x *S. truncata*, *L. cruciforme* x *R. baccifera*, *L. cruciforme* x *R. teres*, *S. truncata* x *R. baccifera*, *S. truncata* x *R. teres*, and *R. baccifera* x *R. teres*) (Fig. 7). Therefore, we identified nine regions as nucleotide diversity hotspots of the tribe Rhipsalideae plastomes. The highest polymorphism peak encompasses the IGS regions of *trnM-CAU/accD* and CDS of the *accD* gene (pi>0.36), observed in all combinations analyzed here. Another point of high polymorphism among the four species is at the end of the SSC region, where the *ndhB* (pi>0.24) is located, which is a pseudogene in Rhipsalideae. The *clpP* gene also showed high nucleotide variability when comparing the plastomes of *L. cruciforme* x *S. truncata* and *S. truncata* x *R. teres* (pi>0.22) (Fig. 7).

The other polymorphic regions are specific for each pair of plastomes analyzed. For example, a comparison between the plastomes of *L. cruciforme* and *R. teres* revealed that the IGS *trnI-GAU/rrn16* region (pi>0.38) is even more polymorphic than the *accD* gene (pi>0.35). Likewise, the *rpoB/trnC-GCA* and *ycf4/cemA* intergenic regions presented high nucleotide diversity (pi>0.26) between the plastomes of *L. cruciforme* and *R. baccifera*. Between *R. teres* and *R. baccifera* plastomes, the intergenic region *petN/psbM* showed high polymorphism (pi>0.26). Finally, comparing the *S. truncata* and *R. baccifera* plastomes was identified high nucleotide diversity in the *ycf1* gene (pi>0.2) and the IGS *rpl32/ndhF* region (pi>0.25) (Fig. 7).

Gene divergence and RNA editing sites in protein-coding genes of *L. cruciforme* and *S. truncata*

To expand the study of plastid gene evolution within the family Cactoideae, we inferred the gene divergence of the 67 protein-coding genes based on phylogenetic trees. Of the 67 genes analyzed, only 32 of them showed a low substitution rate for all cacti species (<0.05) (Supplementary Fig. S2). On the other hand, eight genes were highly divergent, namely *accD*, *clpP*, *infA*, *rps18*, *ycf1*, *rpl32*, *rpl22*, and *rps19*, whose mean branch length values were higher than 0.2 (Supplementary Fig. S3). The greatest variations in branch lengths among species were observed in the *accD* and *rps18* genes. The *accD* gene is the most divergent in the Cactaceae family, whose branch length ranged from 0.82 (*L. principis*) to 1.84 (*O. ficus-indica*). The *rps18* gene showed the highest substitution rate in *F. setispinus* (1.12) and the lowest rate for *P. aculeata* (0.08). Although the average branch length of the *psbM* gene was low (0.09), this gene also presented a high divergence among species, whose branch lengths ranged from 0.01 to 0.66 (Supplementary Fig S2-S3).

Concerning the RNA editing sites in the protein-coding genes, a total of 37 putative RNA editing sites distributed in 17 genes among species were predicted (Supplementary Table S5). All editions are cytidine (C) to uridine (U), at the first or second codon position. *L. cruciforme* and *S. truncata* showed 22 and 24 putative RNA editing sites, respectively. Of these, 19 sites are shared between both species, and most of them are conserved within the subfamily Cactoideae [i.e., *atpA* (305), *atpA* (424), *atpB* (138), *matK* (217), *petB* (140), *psbL* (1), *rpl20* (20), *rpl20* (83), *rpl20* (87), *rpoB* (158), *rpoB* (184), *rpoB* (189), *rpoC1* (14), *rpoC1* (210), *rpoC2* (498), *rps2* (83), *rps14* (27), and *rps14* (50)]. The *rpoC1* (157) and *rpoC2* (605) sites are present only in *L. cruciforme*, *S. truncata*, *R. baccifera,* and *R. teres*, therefore, are specific to

the tribe Rhipsalideae. The *psbF* (26) site found in *L. cruciforme* is shared with other species of Core Cactoideae, but absent in *S. truncata* and *R. teres* (Supplementary Table S3). Here we identified three putative specific sites in two genes of *S. truncata*: *rpoB* (664), *rpoC2* (652), and *rpoC2* (769). Furthermore, four species also exhibited unique sites, being them *E. grusonii* [*atpI* (33)], *F. latispinus* [*rpl2* (217)], *F. setispinus* [*psbB* (193), *rpoC2* (69), and *rps2* (203)], and *S. grandiflorus* [*petB* (163)]. Two new sites [*rpoC2* (163) and *rpoC2* (747)] are shared among *H. polyrhizus, S. anthonyanus, S. grandifloras,* and *S. validus* (Supplementary Table S5).

Simple sequence repeats analysis

A total of 214 and 165 SRRs loci were identified in *L. cruciforme* and *S. truncata* plastomes, respectively. In both species, the most frequent of SSRs are mono and dinucleotides, with 158 and 127, and 39 and 33 repeats, respectively. The bases A/T compose most of the SSR mono (96%) e dinucleotides (41%) (Supplementary Table S6). The repeats are localized most frequently in intergenic spacers (IGSs), followed by coding sequences (CDS), and last in introns. The *L. cruciforme* plastome show 118 SSRs in IGS, being most abundant in *psbK*/*psbI* (6) and *rpoB*/*trnC-GCA* (6) regions. The 65 SSRs in CDS are distributed principally in *ycf1* (11)*, rpoC2* (9), and *ycf2* (7) genes. And the SSRs localized in introns are greater occurrences in *atpF*, *petD,* and *trnL-UAA*, with five SSRs each. In its turn, the *S. truncata* plastome presents 88 SSRs localized in IGS, 50 in CDS, and 27 in intron regions, respectively. The *rbcL/atpB*, *trnK-UUU/rps16*, *rps16/trnQ-UUG*, and *ycf3/psaA* IGS regions content more repeats, all with four SSRs. Again, the SSRs localized in CDS were most frequent in *ycf1* (8), *rpoC2* (8), and *ycf2* (6) genes. Concerning SSR in introns, the genes more abundant were *atpF* (6), *rpl16* (5), and *petD* (4). The size, sequence, and location of the SSRs identified in the *L. cruciforme* and *S. truncata* plastomes are shown in Supplementary Table S7-S8.

Discussion

General features of the *Lepismium cruciforme* and *Schlumbergera truncata* plastomes reveal conserved structure in the Rhipsalideae tribe

The plastid genomes of angiosperms commonly have structure and gene content conserved, and gene loss is extremely rare (Wicke et al. 2011). However, this conserved pattern does not occur in the family Cactaceae. In addition to presenting the greatest diversity of species in the family, the subfamily Cactoideae exhibits many unusual features in its plastomes. This

includes intense rearrangements, losses or expansions and contractions of the IR region, and extensive gene losses (Sanderson et al. 2015; Solórzano et al. 2019; Koler et al. 2020; Oulo et al. 2020; Almeida et al. 2021; Silva et al. 2021; Dalla Costa et al. 2022; Qin et al. 2022).

The tribe Rhipsalideae (i.e., *L. cruciforme*, *S. truncata*, *R. teres,* and *R. baccifera*) has a conserved plastome structure, with some variations of IRs boundaries. The four species exhibit the same pattern of rearrangements, of which two inversions in the LSC region (LCBs B and E, Fig. 3 and Supplementary Fig. S1) appear to be recent and tribe-specific events. On the other hand, the SSC region of species of the subfamily Cactoideae has undergone intense rearrangement and, currently, harbors genes generally located in the IRs of angiosperms, such as the ribosomal RNA genes (Pacheco et al. 2020a, b; Lopes et al. 2021). This configuration also occurs in *Pereskia aculeata* (unpublished data), a basal species of the family Cactaceae, indicating that this complex rearrangement occurred earlier than the changes in the IRs boundaries (Silva et al. 2021), which may have triggered the different configurations of cacti plastomes.

Regarding the gene content, *L. cruciforme* and *S. truncata* had extensive loss and pseudogenization. Shared losses between both species include the genes of the *ndh* complex, *rpl23*, and *trnV-UAC*. Furthermore, *ycf4* is a pseudogene in *S. truncata* plastome due to a premature stop codon in the sequence, and the *rpl33* gene was lost from the *L. cruciforme* plastome. All of them have been previously reported in the subfamily Cactoideae and are widely discussed by Silva et al. (2021) and Dalla Costa et al. (2022). However, the pseudogenization of *trnT-GGU* in *L. cruciforme*, discussed below, is a novelty for the family Cactaceae.

Is the *trnT-GGU* gene in the process of degeneration in the Rhipsalideae tribe?

The plastid genome contains two genes encoding transfer RNA for the aminoacid threonine: *trnT-UGU* and *trnT-GGU*, which encode $tRNA^{Thr}$(UGU), and $tRNA^{Thr}$(GGU), respectively. These transporters recognize four codons for threonine (i.e., ACA, ACC, ACG, and ACT), according to Crick's (1966) oscillation rules. However, the *trnT-GGU* gene appears to be degenerating in the tribe Rhipsalideae. All four species of the tribe analyzed here present at least three mutations in the gene sequence, reflected in the alteration of the secondary structure of the tRNAs, with some pairing incompatibilities and consequent pseudogenization in *L. cruciforme*.

Independent *trnT-GGU* pseudogenization has been previously reported in three families, namely Geraniaceae (Chumley et al. 2006), Cupressaceae (Hirao et al. 2008), and

Asteraceae (Lee et al. 2017). More recently, Abdullah et al. (2021) carried out a comprehensive investigation of the *trnT-GGU* gene in the Asteraceae family, analyzing species from 13 subfamilies, and found that this is a pseudogene in the core of Asteraceae. Two possible causes for the pseudogenization of the *trnT-GGU* gene have been proposed. One is related to an inversion event of a large gene segment, in which the *trnT-GGU* gene was located (Chumley et al. 2006). The other refers to an insertion event, which can increase mutation rates and cause pseudogenization or even gene loss (Abdullah et al. 2021). In Rhipsalideae, there was also an inversion in the gene segment that harbors the genes from *ndhJ* to *trnT-GGU*, but this did not lead to pseudogenization in the four species of the tribe. On the other hand, only *L. cruciforme* obtained an insertion in the gene sequence, which is the possible cause for the pseudogenization of *trnT-GGU*.

In a work based on reverse genetics with tobacco, Alkatib et al. (2012) reported the essentiality of the *trnT-UGU* gene, while *trnT-GGU* is dispensable for translation and cell viability in plastids. This can be explained by the superwobbling mechanism, in which a tRNA containing uridine with an unknown change in wobble position (*trnT-UGU*) can efficiently decode all codons (Pfitzinger et al. 1990; Rogalski et al. 2008; Alkatib et al. 2012; Abdullah et al. 2021). In a similar way, *L. cruciforme* plastome may be using the superwobbling mechanism to compensate the lack of functionality of *trnT-GGU*, since there was no change in the frequency of use of codons for the protein-coding sequences of *L. cruciforme*, but frequencies similar to species that keep the functionality of both threonine tRNAs.

Suborder Cactineae phylogeny and nucleotide divergence hotspots as a tool for phylogenetic analysis of the tribe Rhipsalideae

The phylogenetic tree based on 54 protein-coding plastid genes was very similar to the inferences previously published by Silva et al. (2021) and Dalla Costa et al. (2022). Comparing with these previous inferences, the phylogenetic positions of the families Halophytaceae and Talinaceae remain inconsistent, as both families have only one taxon with their complete plastomes available in databases (*Halophytum ameghinoi* and *Talinum paniculatam*, respectively). Within Cactaceae, *Pereskia aculeata* constituted the initial branch of the family, forming a sister group with the subfamilies Opuntioideae and Cactoideae, both of monophyletic origin, as previously reported (Hernández-Hernández et al. 2011; Guerrero et al. 2019a). The relationships found within the Cactoideae subfamily, referring to the Cacteae and Core Cactoideae clades (I and II), are congruent with recen resolved phylogenies (Guerrero et al.

2019a; Silva et al. 2021; Dalla Costa et al. 2022). However, a curious relationship in the tribe Rhipsalideae was inferred. The species *Rhipsalis baccifera* and *Rhipsalis teres* did not constitute a clade, but formed a sister group with *S. truncata* and *L. cruciforme*, respectively, in a highly supported relationship.

Korotkova et al. (2011) inferred the phylogeny of Rhipsalideae through seven different regions, being two introns (*trnK*, *rpl16*), three intergenic spacers (*psbA-trnH*, *trnQ-rps16*, and *rps3-rpl16*) and two coding regions (*rbcL* and *matK*), sampling all 52 recognized species of the tribe. In this study, the genera *Rhipsalis* and *Lepismium* were found to constitute a clade, and a monophyletic origin of both genera was corroborated(Korotkova et al. 2011). Other phylogenies based on intergenic regions *trnK-matK* (Bárcenas et al. 2011), *psbA-trnH*, *trnQ-rps16*, and *rpl32-trnL* (Calvente et al. 2011a), as well as with plastid and nuclear markers (Calvente et al. 2011b) also reported the *Rhipsalis* monophyly. Resolving phylogenetic trees for Cactaceae has been challenging, and many relationships remain unresolved, especially in South American lineages, namely Core Cactoideae II (Korothova et al. 2011; Guerrero et al. 2019a). Although the phylogenetic relationship of the tribe Rhipsalideae was strongly supported in our analyses, this will be better understood when more taxa with complete plastid sequences become available.

Hotspots of nucleotide polymorphisms in plastome sequences provide interesting information for genetic studies, being useful for improving phylogenetic resolutions at the tribe level, for example (Rogalski et al. 2015; Lopes et al. 2018). Our sliding window analysis revealed nine nucleotide diversity hotspots within the tribe Rhipsalideae. Of these nine regions, only two of them have been used in phylogenetic studies of Cactaceae previously, that is, the spacer region *rpl32*/*ndhF* (Majure et al. 2012) and the coding region *ycf1* (Cruz et al. 2016; Larridon et al. 2018; Guerrero et al. 2019b). However, none of these studies covered species of the tribe Rhipsalideae. We identified the largest nucleotide polymorphisms in *trnM-CAU*/*accD*, *petN/psbM*, *rpoB/trnC-GCA*, *trnI-GAU/rrn16*, and *rpl32/ndhF* IGS, as well as in *accD*, *ycf1*, *clpP*, and *ndhB* CDS. It is already known that *accD*, *ycf1*, and *clpP* genes are more divergent, including in Cactaceae (Silva et al. 2021; Dalla Costa et al. 2022). Due to this divergence, the *ycf1* gene is considered the most promising plastid DNA barcode of land plants (Dong et al. 2015). Likewise, the *accD* gene is suggested as a plastid marker for genetic studies and, in addition, accD and clpP can be used in phylogenies in lower taxa in Cactaceae (Silva et al. 2021). Therefore, these hotspots bring genetic information that can improve the phylogeny of the tribe Rhipsalideae.

Molecular evolution of plastid protein-coding genes and RNA editing sites in *L. cruciforme* and *S. truncata* plastomes

According to our data and previous publications, the protein-coding genes of the family Cactaceae are highly divergent (Silva et al. 2021; Dalla Costa et al. 2022). Our analysis indicated that the genes with the highest substitution rates are *accD*, *clpP*, *infA*, *rps18*, *ycf1*, *rpl32*, *rpl22*, and *rps19*, respectively. Six of these eight genes had already shown mean replacement rates above 0.2 in our previous analysis (Dalla Costa et al. 2022). As expected (Silva et al. 2021; Dalla Costa et al. 2022), the *accD* gene is the most divergent in the family Cactaceae. On the other hand, the analysis upload with the inclusion of 14 new cactus species revealed the high divergence of the *infA* and *rps18* genes. The *infA* gene encodes a translation initiation factor and, together with two other factors encoded by the nuclear genome, initiates plastid translation (Wicke et al. 2011; Daniell et al. 2016). The high divergence of *infA* in Cactaceae also was reported by Silva et al. (2021). In addition, many losses (Pacheco et al. 2020b; Scobeyeva et al. 2021) and pseudogenization (Pacheco et al. 2020a; Alqahtani and Jansen 2021), as well as the transfer of the *infA* gene to the nucleus (Park et al. 2015) have been described in several species. In turn, the *rps18* gene encodes the protein S18, an essential component of the 30S ribosome subunit (Rogalski et al. 2006). For this reason, few losses and pseudogenization have been reported (Cai et al. 2008; Dong et al. 2013). However, *rps18* is pseudogenized in the genus *Mammilaria* (Solórzano et al. 2019) and *Leuchtenbergia principis*. Furthermore, the most divergent sequences were identified in *Ferocactus setispinus* and *Echinocactus grusonii* species. All these species belong to the tribe Cacteae, so it is possible that the *rps18* gene is degenerating in the tribe and that a functional copy has been transferred to the nuclear genome (Silva et al. 2021).

All land plants have a post-transcriptional modification mechanism based on RNA editing, except for some *Marchantiales* species (Small et al. 2020). Through conversions from cytidine to uridine, or less frequently from U-to-C, start or stop codons can be created and amino acids can be changed, either for error correction and functional regulation or for adaptation and stress conditions (Takenaka et al. 2013). In this study, were predicted 37 putative RNA editing sites and all edits are C-to-U conversions, at the first or second codon position, as occurs in most angiosperms (Takenaka et al. 2013). Of these 37 sites, 26 had already been identified in the subfamily Cactoideae in previous studies (Silva et al. 2021; Dalla Costa et al. 2022). Only 12 sites (32.4%) are shared with all species of the subfamily. On the other hand, we identified 11 new putative RNA editing sites. *S. truncata* has three possibly unique sites,

namely *rpoB* (664), *rpoC2* (652), and *rpoC2* (769). The other eight new sites found are distributed among the species *E. grusonii*, *F. latispinus*, *F. setispinus*, *H. polyrhizus*, *S. anthonyanus*, *S. grandiflorus*, and *S. validus*. These results reveal the high polymorphism of editing sites in the subfamily Cactoideae, indicating the rapid evolution of this mechanism in the subfamily (Silva et al. 2021).

Simple sequence repeats in *L. cruciforme* and *S. truncata* plastomes

Plastid markers present a high level of polymorphism, uniparental inheritance, and non-recombinant nature in most angiosperms, and for this reason, they are widely used (Wheeler et al. 2014). Among them are the SSRs, useful for several studies, such as genetic diversity (Angioi et al. 2010), phylogenetic relationships (Liu et al. 2012), demographic history of species (Jaramillo-Correa et al. 2015), gene flow for herbicide resistance alleles (Busconi et al. 2012) identification of maternal lines and clones in vegetatively propagated cultures (Gardner et al. 2015), among others. Here we identified and mapped a high number of SSRs in *L. cruciforme* (214) and *S. truncata* (165) plastomes, predominantly in non-coding regions (IGS and intros). These regions are advantageous for studies related to the genetic variability of populations because they present faster evolution in comparison with the coding regions (Rogalski et al. 2015). The high number of SSRs is common among Rhipsalideae species, as the *R. baccifera* and *R. teres* species have 169 and 200 SSRs, respectively (Silva et al. 2021). In *M. glaucescens*, 169 SSRs were also previously identified (Dalla Costa et al. 2022). The Cactaceae family is one of the most endangered taxa, either by anthropic action or by intrinsic characteristics of cactus (Machado et al. 2009; Goettsch et al. 2015). Therefore, the SSRs provided here are useful for accessing the genetic diversity of natural populations and assisting in conservation strategies for *L. cruciforme* and *S. truncata*.

Conclusions

Here we report the *L. cruciforme* and *S. truncata* complete plastomes, two species of South American epiphytic cacti of the tribe Rhipsalideae. The plastomic structure is conserved among species of the tribe Rhipsalideae, with two recent tribe-specific inversion events in the LSC region and some variations in the boundaries of the IR region. On the other hand, these plastomes suffered a massive loss of genes and pseudogenization, including the *ndh* complex, *rpl23*, *rpl33*, *ycf4*, and *trnV-UAC* genes. Additionally, *trnT-GGU* is a pseudogene in the

plastome of *L. cruciforme* and appears to be degenerating in the tribe Rhipsalideae. In addition, the plastid protein-coding genes of the family Cactaceae show high gene divergence and rapid evolution of the RNA editing mechanism. Curiously, our phylogenetic inference based on concatenated plastid genes resulted in the separation of the species *Rhipsalis baccifera* and *Rhipsalis teres*, which grouped with *S. truncata* and *L. cruciforme*, respectively. Thus, we identified nine regions of high nucleotide polymorphism among Rhipsalideae species, which may help to clarify the phylogenetic relationships of the tribe. Finally, we mapped the SSRs in *L. cruciforme* and *S. truncata* plastomes, useful for genetic studies, access to natural populations, and conservation strategies for both species.

References

Allen JF (2015) Why chloroplasts and mitochondria retain their own genomes and genetic systems: Colocation for redox regulation of gene expression. PNAS 112(33):10231-10238. https://doi.org/10.1073/pnas.1500012112

Almeida EM, Sader MA, Rodriguez PE, Loeuille B, Felix LP, Pedrosa-Harand A (2021) Assembling the puzzle: Complete chloroplast genome sequences of *Discocactus bahiensis* Britton & Rose and *Melocactus ernestii* Vaupel (Cactaceae) and their evolutionary significance. Braz. J. Bot 44:877–888. https://doi.org/10.1007/s40415-021-00772-2

Abdullah, Mehmood F, Heidari P, Rahim A, Ahmed I, Poczai P (2021) Pseudogenization of the chloroplast threonine (*trnT-GGU*) gene in the sunflower family (Asteraceae). Sci Rep 11:21122. https://doi.org/10.1038/s41598-021-00510-4

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 Genetics 8(11): e1003076. https://doi.org/10.1371/journal.pgen.1003076

Alqahtani AA, Jansen RK (2021) The evolutionary fate of *rpl32* and *rps16* losses in the *Euphorbia schimperi* (Euphorbiaceae) plastome. Sci Rep 11**:**7466. https://doi.org/10.1038/s41598-021-86820-z

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

Angioi SA, Rau D, Attene G, Nanni L, Belluci E, Logozzo G, Negri V, Spagnoletti Zeuli PL, Papa R (2010) Beans in Europe: origin and structure of the European landraces of *Phaseolus vulgaris* L. Theor Appl Genet 121:829-843 (2010). https://doi.org/10.1007/s00122-010-1353-2

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. PNAS 108:8379–8384. https://doi.org/10.1073/pnas.1100628108

Bárcenas RT, Yesson C, Hawkins JA (2011) Molecular systematics of the Cactaceae. Cladistics 27(5):470-489. https://doi.org/10.1111/j.1096-0031.2011.00350.x

Bock R (2007) Structure, function, and inheritance of plastid genomes. In: Bock R (ed) Cell and molecular biology of plastids. Springer, Berlin, pp 29-63. https://doi.org/10.1007/4735_2007_0223

Busconi M, Rossi D, Lorenzoni C, Baldi G, Fogher C (2012) Spread of herbicide-resistant weedy rice (red rice, *Oryza sativa* L.) after 5 years of Clearfield rice cultivation in Italy. Plant Biol 14:751–759. https://doi.org/10.1111/j.1438-8677.2012.00570.x

Cai Z, Guisinger M, Kim HG, Ruck E, Blazier JC, McMurtry V, Kuehl JV, Boore J, Jansen RK (2008) Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. J Mol Evol 67:696704. https://doi.org/10.1007/s00239-008-9180-7

Calvente A, Zappi DC, Forest F, Lohmann LG (2011) Molecular phylogeny, evolution, and biogeography of South American epiphytic cacti. Int J Plant Sci 172(7):902-914. https://dx.doi.org/10.1086/660881

Calvente A, Zappi DC, Forest F, Lohmann LG (2011) Molecular phylogeny of tribe Rhipsalideae (Cactaceae) and taxonomic implications for *Schlumbergera* and *Hatiora*. Mol Phylogenet Evol 58:456-468. https://doi.org/10.1016/j.ympev.2011.01.001

Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK (2006) The complete chloroplast genome sequence of *Pelargonium* x *hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol 23(11):2175–2190. https://doi.org/10.1093/molbev/msl089

Crick FHC (1966) Codon-anticodon pairing: The wobble hypothesis. J Mol Biol 19:548–555.

Cruz MÁ, Arias S, Terrazas T (2016) Molecular phylogeny and taxonomy of the genus *Disocactus* (Cactaceae), based on the DNA sequences of six chloroplast markers. Willdenowia, 46(1):145-164. https://doi.org/10.3372/wi.46.46112

Dalla Costa TP, Silva MC, Lopes AS, Pacheco TG, Oliveira JD, Baura VA, Balsanelli E, Souza EM, Pedrosa FO, Rogalski M (2022) The plastome of *Melocactus glaucescens* Buining & Brederoo reveals unique evolutionary features and loss of essential tRNA genes. Planta 255:57. https://doi.org/10.1007/s00425-022-03841-2

Daniell H, Lin CS, Yu M, Chang WJ (2016) Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol 17:134. https://doi.org/10.1186/s13059-016-1004-2

Darling ACE, Mau B, Blattner FR, Perna NT (2004) Mauve: Multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403. https://doi.org/10.1101/gr.2289704

Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: de novo assembly of organelle genomes from whole genome data, Nucleic Acids Res 45(4):e18. https://doi.org/10.1093/nar/gkw955

Dong W, Xu C, Cheng T, Zhou S (2013) Complete chloroplast genome of *Sedum sarmentosum* and chloroplast genome evolution in Saxifragales. PloS one 8(10):e77965. https://doi.org/10.1371/journal.pone.0077965

Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, Cheng T, Guo J, Zhou S (2015) *ycf1*, the most promising plastid DNA barcode of land plants. Sci Rep 5:8348. https://doi.org/10.1038/srep08348

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340

Fu P-C, Sun S-S, Twyford AD, Li B-B, Zhou R-Q, Chen S-L, Gao Q-B, Favre A (2021) Lineage-specific plastid degradation in subtribe *Gentianinae* (Gentianaceae). Ecol Evol 11:3286– 3299. https://doi.org/10.1002/ece3.7281

Gardner EM, Laricchia KM, Murphy M, Ragone D, Scheffler BE, Simpson S, Williams EW, Zerega NJ (2015) Chloroplast microsatellite markers for *Artocarpus* (Moraceae) developed from transcriptome sequences. Appl Plant Sci 3(9):1500049. https://doi.org/10.3732/apps.1500049

Goettsch B, Hilton-Taylor C, Cruz-Piñón G et al. (2015) High proportion of cactus species threatened with extinction. Nat Plants 1:15142. https://doi.org/10.1038/nplants.2015.142

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. https://doi.org/10.1093/nar/gkz238

Griffiths H, Males J (2017) Succulent plants. Curr Biol 27:890–896. https://doi.org/10.1016/j.cub.2017.03.021

Guerrero PC, Majure LC, Cornejo-Romero A, Hernández-Hernández T (2019a) Phylogenetic relationships and evolutionary trends in the cactus family. J Hered 110:4–21. https://doi.org/10.1093/jhered/esy064

Guerrero PC, Walter HE, Arroyo MT, Peña CM, Tamburrino I, De Benedictis M, Larridon I (2019b) Molecular phylogeny of the large South American genus *Eriosyce* (Notocacteae, Cactaceae): generic delimitation and proposed changes in infrageneric and species ranks. Taxon 68(3):557-573. https://doi.org/10.1002/tax.12066

Hernández-Hernández T, Hernández HM, De-Nova JA, Puente R, Eguiarte LE, Magallón S (2011) Phylogenetic relationships and evolution of growth form in Cactaceae (Caryophyllales, Eudicotyledoneae). Am J Bot 98(1):44-61. doi: 10.3732/ajb.1000129. https://doi.org/10.3732/ajb.1000129

Hirao T, Watanabe A, Kurita M, Kondo T, Takata K (2008) Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: Diversified genomic structure of coniferous species. BMC Plant Biol 8:70. https://doi.org/10.1186/1471-2229-8-70

Hunt D, Taylor NP, Charles G (2006) The New Cactus Lexicon. 2 vols. DH Books, Milborne Port. 900p.

Jaramillo-Correa JP, Gérardi S, Beaulieu J, Ledig FT, Bousquet J (2015) Inferring and outlining past population declines with linked microsatellites: a case study in two spruce species. Tree Genet Genomes 11:9. https://doi.org/10.1007/s11295-015-0835-4

Katoh K, Rozewicki J, Yamada KD (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinformatics 20(4):1160–1166. https://doi.org/10.1093/bib/bbx108

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. Front Plant Sci 19(11):729. https://doi.org/10.3389/fpls.2020.00729

Korotkova N, Borsch T, Quandt D, Taylor NP, Müller KF, Barthlott W (2011) What does it take to resolve relationships and to identify species with molecular markers? An example from the epiphytic Rhipsalideae (Cactaceae). Am J Bot: 98(9):1549-1572. https://doi.org/10.3732/ajb.1000502

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33(7):1870–1874. https://doi.org/10.1093/molbev/msw054

Lanfear R, Calcott B, Ho SYW, Guindon S (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol 29:1695–1701. https://doi.org/10.1093/molbev/mss020

Larridon I, Walter HE, Rosas M, Vandomme V, Guerrero PC (2018) Evolutionary trends in the columnar cactus genus *Eulychnia* (Cactaceae) based on molecular phylogenetics, morphology, distribution, and habitat. Syst Biodiver 16(7):643-657. https://doi.org/10.1080/14772000.2018.1473898

Lee D-H, Cho W-B, Choi B-H, Lee J-H (2017) Characterization of two complete chloroplast genomes in the tribe *Gnaphalieae* (Asteraceae): gene loss or pseudogenization of *T-GGU* and implications for phylogenetic relationships. Hortic Sci Technol 35(6):769–783. https://doi.org/10.12972/kjhst.20170081

Liu Z-L, Yang X, Zhang H, Zhang J, Li J-F (2012) Systematic position of *Pinus henryi* (Pinaceae) as revealed by multiple evidence. Nord J Bot 30:671–679. https://doi.org/10.1111/j.1756-1051.2012.01554.x

Lopes AS, Pacheco TG, Vieira LN, Guerra MP, Nodari RO, de Souza EM, Pedrosa FO, Rogalski M (2018) The *Crambe abyssinica* plastome: Brassicaceae phylogenomic analysis,

evolution of RNA editing sites, hotspot and microsatellite characterization of the tribe Brassiceae. Gene 671:36-49. https://doi.org/10.1016/j.gene.2018.05.088

Lopes AS, Pacheco TG, Silva ON, Vieira LN, Guerra MP, Matta EPL, Baura VA, Balsanelli E, Souza EM, Pedrosa FO, Rogalski M (2021) Plastid genome evolution in Amazonian açaí palm (*Euterpe oleracea* Mart.) and Atlantic forest açaí palm (*Euterpe edulis* Mart.). Plant Mol Biol 105(4):559–574. https://doi.org/10.1007/s11103-020-01109-5

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. https://doi.org/10.1093/nar/gkw413

Lughadha EN, Govaerts R, Belyaeva I, Black N, Lindon H, Allkin R, Magill RE, Nicolson N (2016) Counting counts: revised estimates of numbers of accepted species of flowering plants, seed plants, vascular plants and land plants with a review of other recent estimates. Phytotaxa, 272(1):82-88. https://doi.org/10.11646/phytotaxa.272.1.5

Machado MC (2009) The genus *Melocactus* in eastern Brazil: part I - An introduction to *Melocactus*. British Cactus & Succulent Journal, 27:1–16.

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. Am J Bot 99(5):847-864. https://doi.org/10.3732/ajb.1100375

Moghaddam M, Kazempour-Osaloo S (2020) Extensive survey of the *ycf*4 plastid gene throughout the IRLC legumes: Robust evidence of its locus and lineage specific accelerated rate of evolution, pseudogenization and gene loss in the tribe Fabeae. PLOS ONE 15(3): e0229846. https://doi.org/10.1371/journal.pone.0229846

Mower JP (2009) The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. Nucleic Acids Res 37: W253–W259. https://doi.org/10.1093/nar/gkp337

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300

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. https://doi.org/10.3390/plants9080979

Pacheco TG, Lopes AS, Oliveira JD, Otoni WC, Balsanelli E, Pedrosa FO, Souza EM, Rogalski M (2020a) The complete plastome of *Passiflora cirrhiflora* A. Juss.: structural features, RNA editing sites, hotspots of nucleotide diversity and molecular markers within the subgenus Deidamioides. Braz J Bot 43:839-853. https://doi.org/10.1007/s40415-020-00655-y

Pacheco TG, Lopes AS, Welter JF, Yotoko KSC, Otoni WC, Vieira LN, Guerra MP, Nodari RO, Balsanelli E, Pedrosa FO, Souza EM, Rogalski M (2020b) Plastome sequences of the

subgenus *Passiflora* reveal highly divergent genes and specific evolutionary features. Plant Mol Biol 104:21-37. https://doi.org/10.1007/s11103-020-01020-z

Park S, Jansen RK, Park S (2015) Complete plastome sequence of *Thalictrum coreanum* (Ranunculaceae) and transfer of the *rpl32* gene to the nucleus in the ancestor of the subfamily Thalictroideae. BMC Plant Biol 15:40. https://doi.org/10.1186/s12870-015-0432-6

Pfitzinger H, Weil JH, Pillay DTN, Guillemaut P (1990) Codon recognition mechanisms in plant chloroplasts. Plant Mol Biol 14:805–814. https://doi.org/10.1007/BF00016513

Qin Q, Li J, Zeng S, Xu Y, Han F, Yu J (2022) The complete plastomes of red fleshed pitaya (*Selenicereus monacanthus*) and three related *Selenicereus* species: insights into gene losses, inverted repeat expansions and phylogenomic implications. Physiol Mol Biol Plants 28(1):123-137. https://doi.org/10.1007/s12298-021-01121-z

Rogalski M, Karcher D, Bock R (2008) Superwobbling facilitates translation with reduced tRNA sets. Nat Struct Mol Biol 15:192–198. https://doi.org/10.1038/nsmb.1370

Rogalski M, Ruf S, Bock R (2006) Tobacco plastid ribosomal protein S18 is essential for cell survival. Nucleic Acids Res 34:4537–4545. https://doi.org/10.1093/nar/gkl634

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. https://doi.org/10.3389/fpls.2015.00586

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542. https://doi.org/10.1093/sysbio/sys029

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol 34:3299–3302. https://doi.org/10.1093/molbev/msx248

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. https://doi.org/10.3732/ajb.1500184

Scobeyeva VA, Artyushin IV, Krinitsina AA, Nikitin PA, Antipin MI, Kuptsov SV, Belenikin MS, Omelchenko DO, Logacheva MD, Konorov EA, Samoilov AE, Speranskaya AS (2021) Gene loss, pseudogenization in plastomes of genus *Allium* (Amaryllidaceae), and putative selection for adaptation to environmental conditions. Front Genet 12:674783. https://doi.org/10.3389/fgene.2021.674783

Silva GM, Lopes AS, Pacheco TG, Machado KLG, Silva MC, Oliveira JD, Baura VA, Balsanelli E, 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. Rev Bras Bot 44:97–116. https://doi.org/10.1007/s40415-020-00689-2

Small ID, Schallenberg-Rüdinger M, Takenaka M, Mireau H, Ostersetzer-Biran O (2020) Plant organellar RNA editing: what 30 years of research has revealed. Plant J 101(5):1040-1056. https://doi.org/10.1111/tpj.14578

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(10):392. https://doi.org/10.3390/plants8100392

Stirbet A, Dušan Lazár, Ya Guo, Govindjee Govindjee (2020) Photosynthesis: basics, history and modelling. Ann Bot 126(4):511–537. https://doi.org/10.1093/aob/mcz171

Takenaka M, Zehrmann A, Verbitskiy D, Härtel B, Brennicke A (2013) RNA editing in plants and its evolution. Annu Rev Genet 47:335–352. https://doi.org/10.1146/annurev-genet-111212-133519

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22(22):4673-4680. https://doi.org/10.1093/nar/22.22.4673

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. https://doi.org/10.1093/nar/gkx391

Vieira LN, Faoro H, Fraga HPF, Rogalski M, de Souza EM, de Oliveira PF, Nodari RO, Guerra MP (2014) An improved protocol for intact chloroplasts and cpDNA isolation in conifers. PLoS ONE 9: e84792. https://doi.org/10.1371/journal.pone.0084792

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(12):1400. https://doi.org/10.3732/apps.1400059

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. https://doi.org/10.1007/s11103-011-9762-4

FIGURES

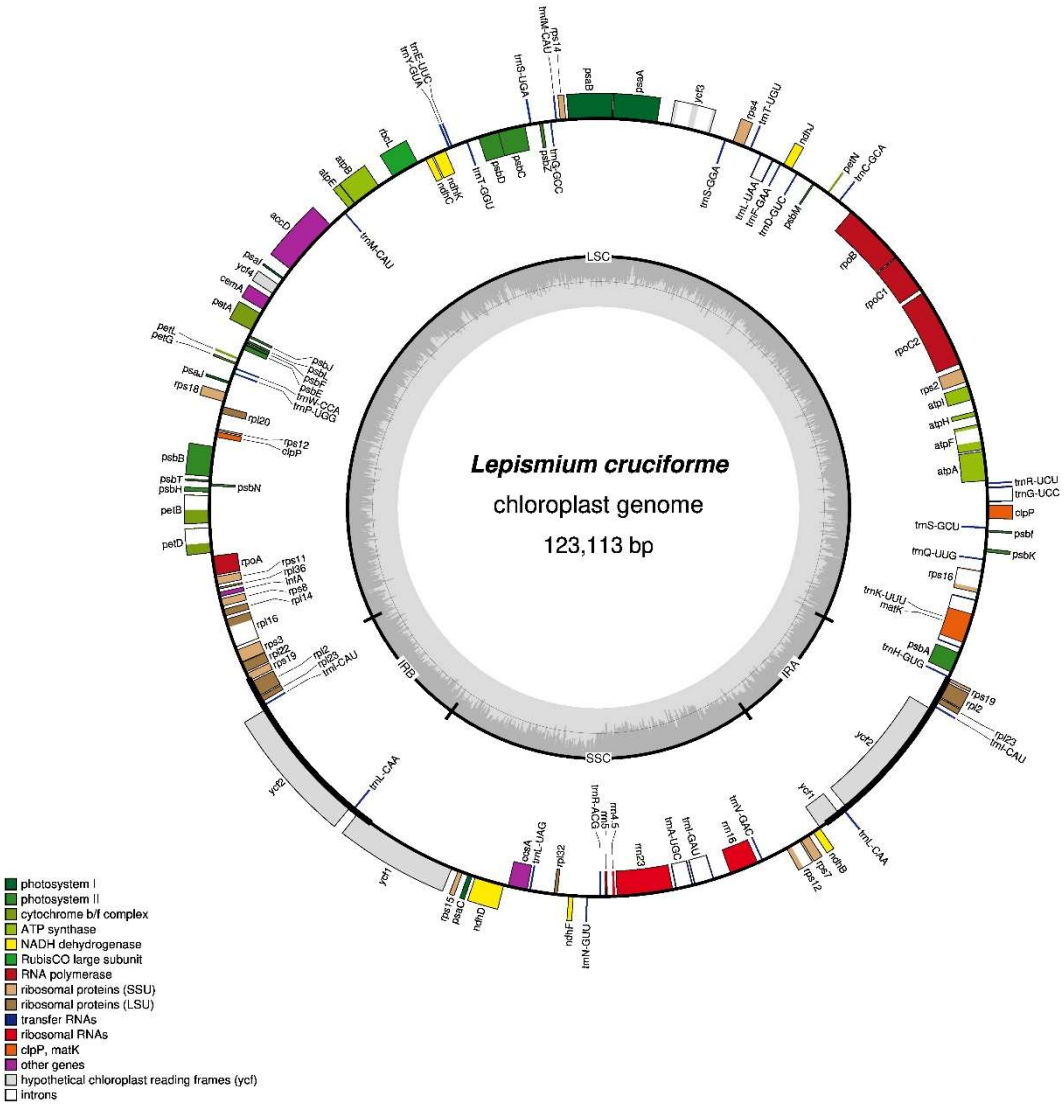


Figure 1. Gene map of *Lepismium cruciforme* plastome. Genes drawn inside the circle are transcribed clockwise and genes drawn outside are expressed counterclockwise direction. Different functional groups of genes are color-coded. The darker gray in the inner circle corresponds to GC content, while the lighter gray corresponds to AT content. *LSC* Large Single Copy, *SSC* Small Single Copy, *IRA*/*IRB* Inverted Repeats A and B.

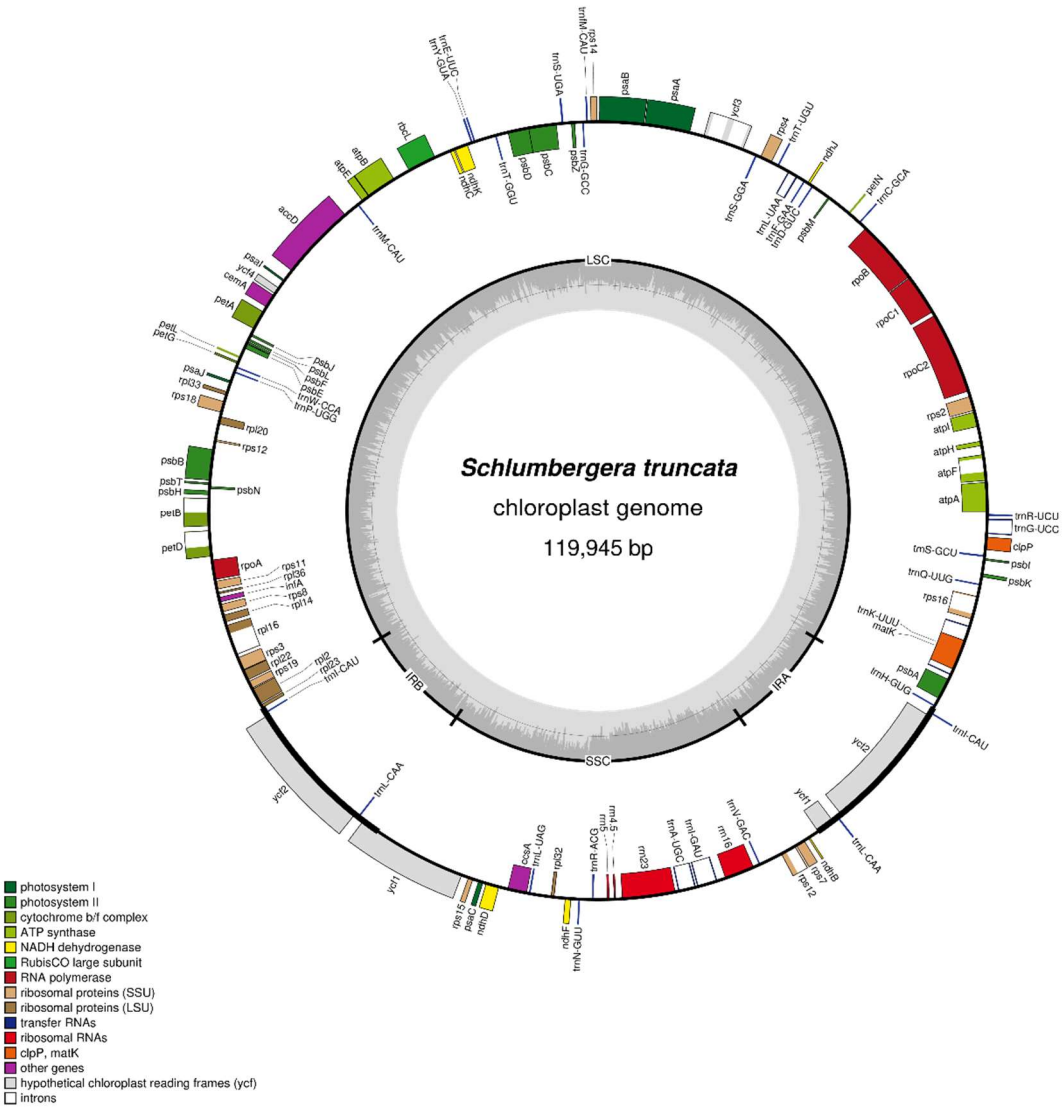


Figure 2. Gene map of *Schlumbergera truncata* plastome. Genes drawn inside the circle are transcribed clockwise and genes drawn outside are expressed counterclockwise direction. Different functional groups of genes are color-coded. The darker gray in the inner circle corresponds to GC content, while the lighter gray corresponds to AT content. *LSC* Large Single Copy, *SSC* Small Single Copy, *IRA*/*IRB* Inverted Repeats A and B.

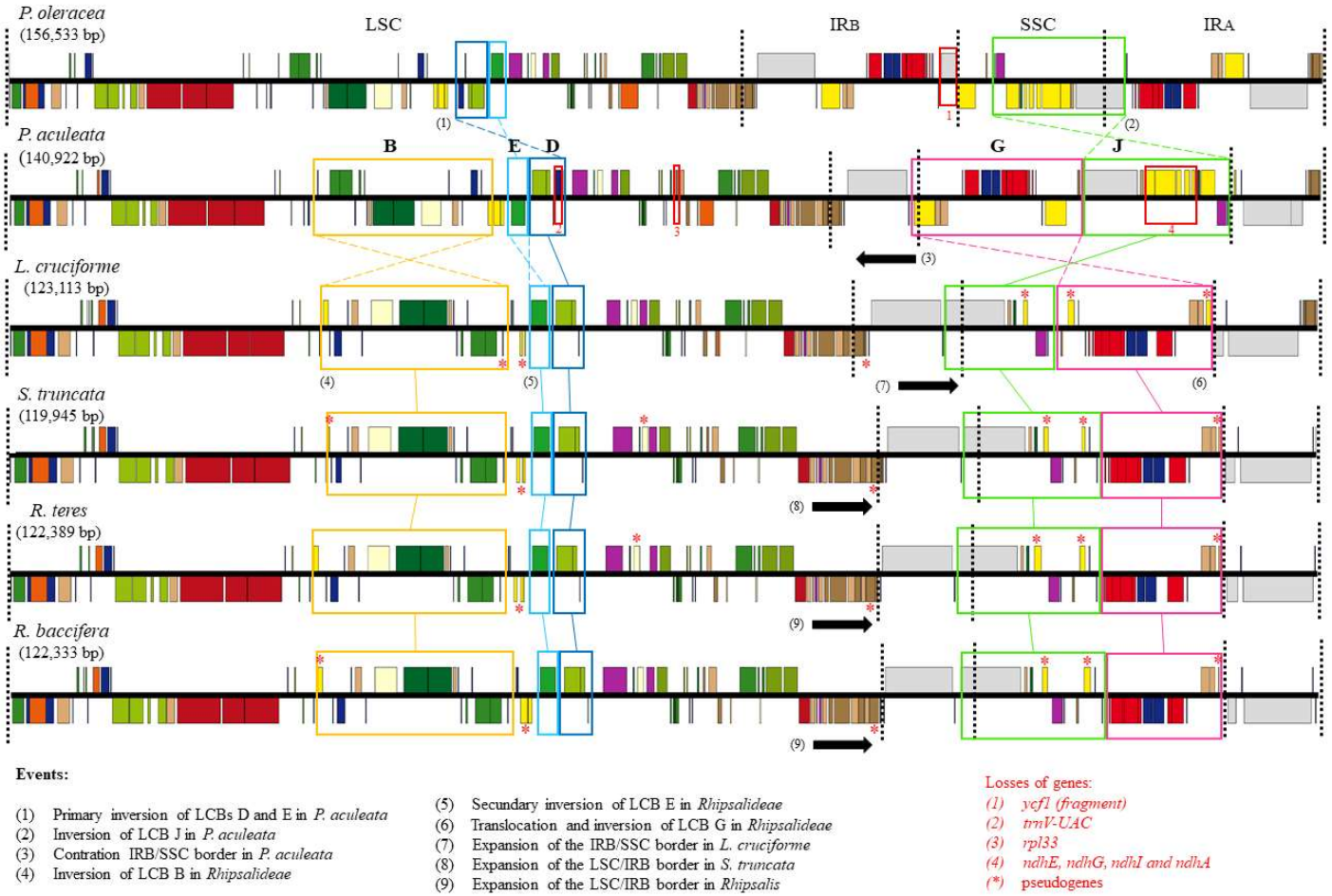


Figure 3. Gene content and order among the plastomes of *Lepismium cruciforme*, *Schlumbergera truncata*, *Rhipsalis teres, Rhipsalis baccifera* and *Pereskia aculeata* compared with the *Portulaca oleracea* plastome, which represents a plastome conserved of the angiosperms. The colored squares correspond to the LCB obtained by the multiple alignments performed by MAUVE (Supplementary Fig. S1). The red squares pointed out in the plastome of *P. oleracea*, and *P. aculeata* represent gene losses; red asterisks indicate the pseudogenes. Dotted cross lines indicate events of translocation and inversion of the gene order. Black arrows represent LSC/IRB, and IRB/SSC border expansion and contraction events. The rearrangement events are described according to the indicated numbering. Linear gene maps were drawn by using OGDRAW. *LSC* Large Single Copy, *SSC* Small Single Copy, *IRA*/*IRB* Inverted Repeats A and B.

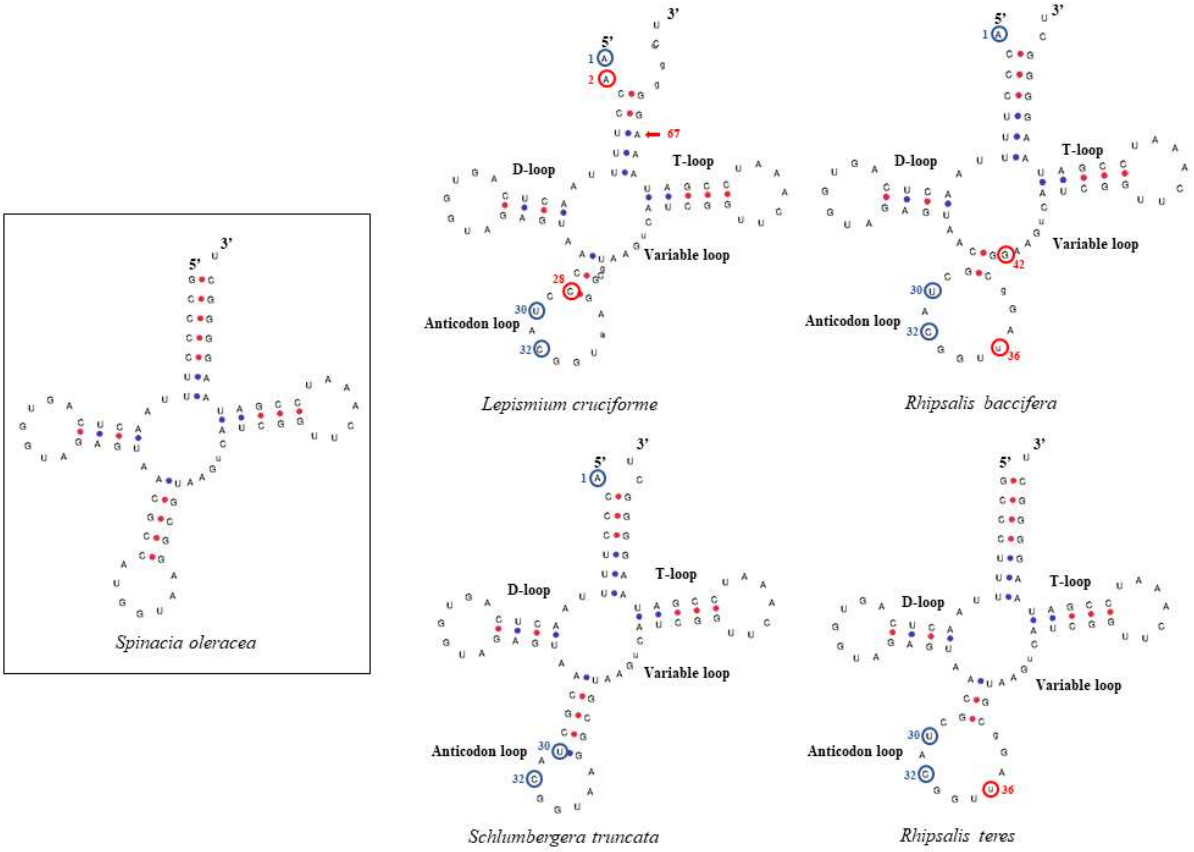


Figure 4. Putative secondary structure of *trnT*-GGU of four species of Rhipsalideae tribe compared with *Spinacia oleracea*, whose structure is a perfect cloverleaf. The circles indicate substitutions of nucleotides, transitions (blue circles) or transversions (red circles). The secondary structure was predicted by the tRNAscan-SE.

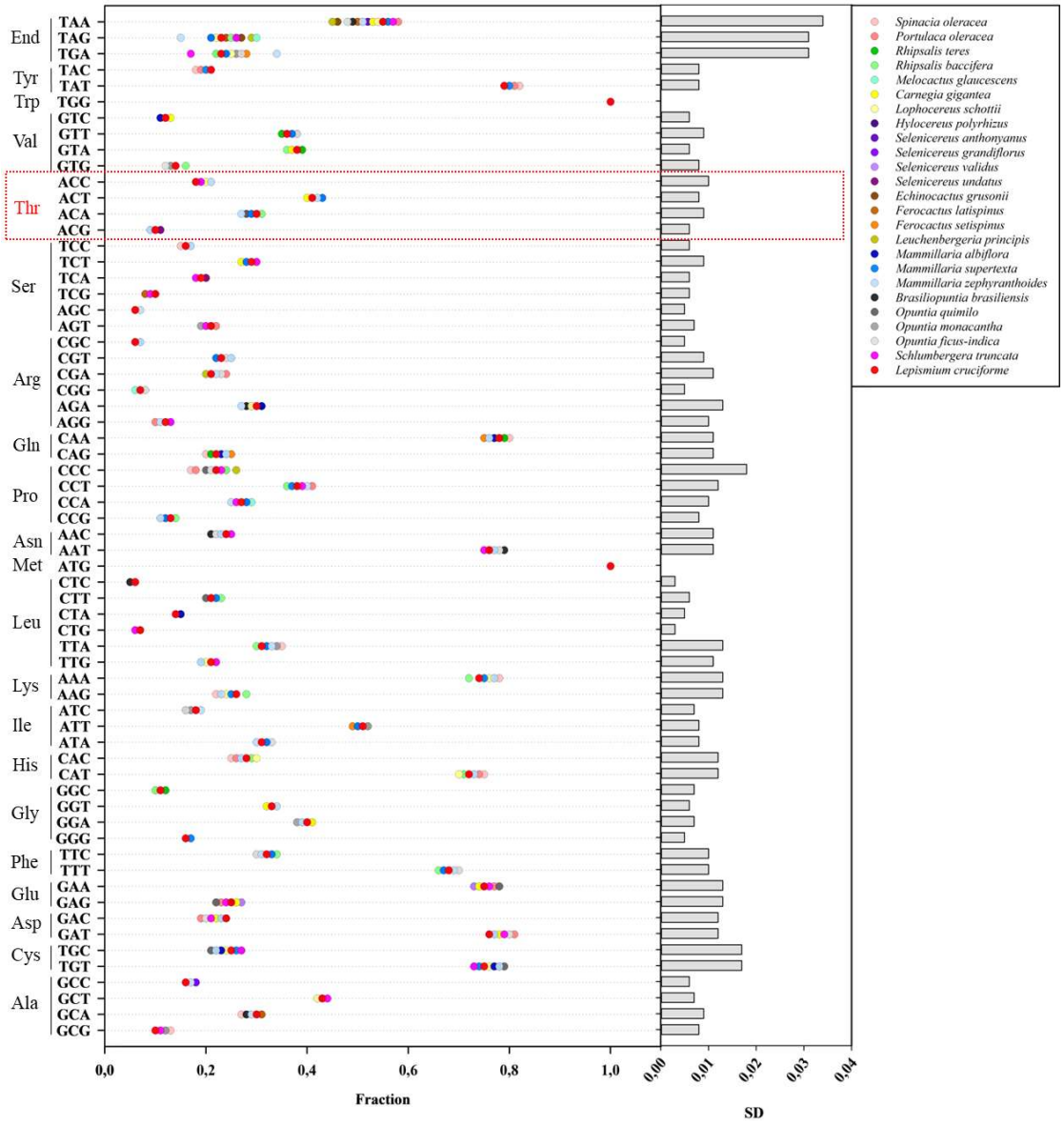


Figure 5. Frequency and standard deviation (SD) of the codon usage analysis of *Lepismium cruciforme* protein-coding plastid genes, compared to 22 species of the Cactaceae family, *P. oleracea* and *S. oleracea*. The red square highlights codons for the amino acid threonine, whose respective tRNA gene (*trnT*-GGU) is a pseudogene in the *L. cruciforme* plastome.

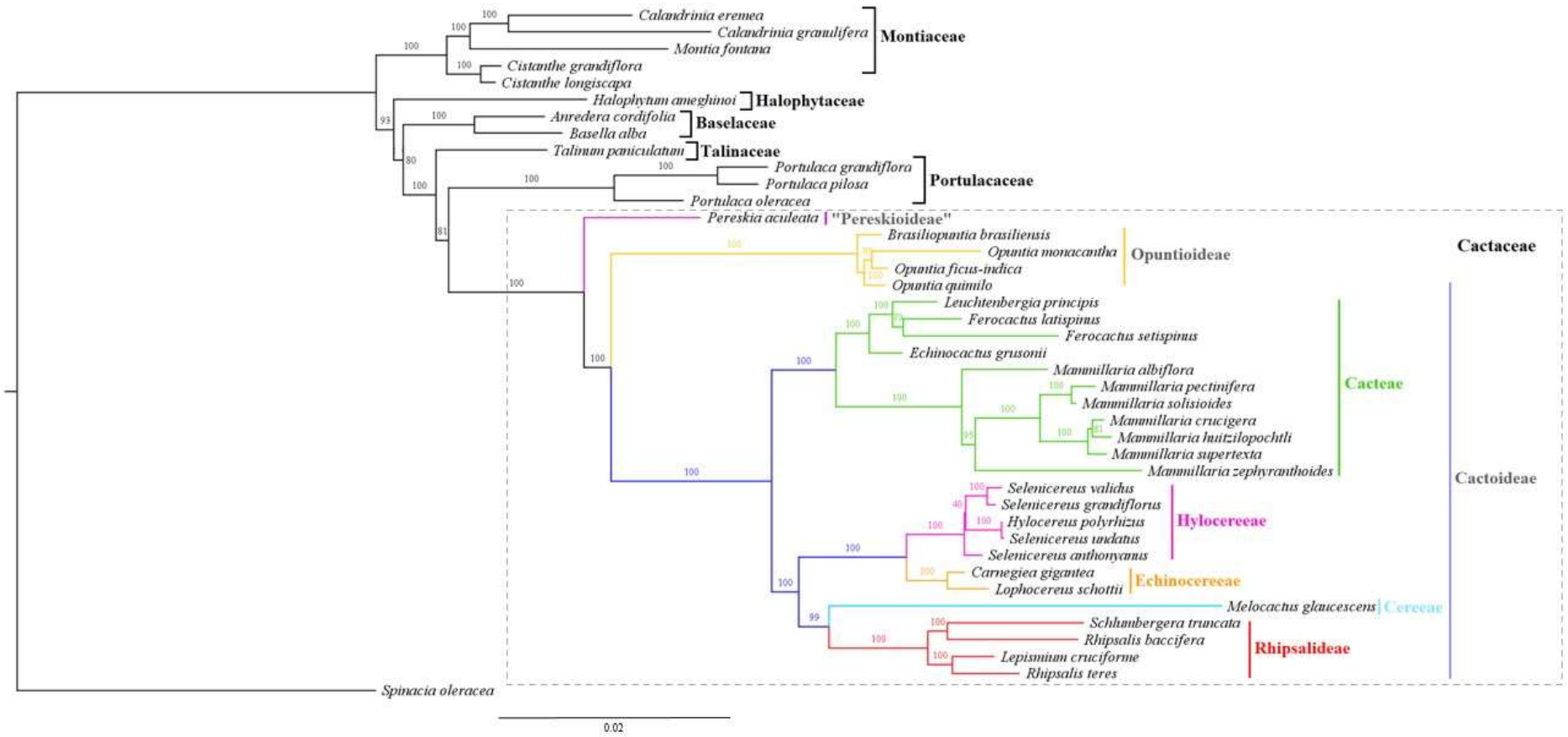


Figure 6. Maximum likelihood (ML) phylogenetic analysis based on 54 protein-coding plastid genes shows the position of *Lepismium cruciforme* and *Schlumbergera truncata* (highlighted in red) within the suborder Cactineae. The families are indicated to the right of the respective species. Numbers (%) associated with branches are ML bootstrap support (BS) values. The branch length is proportional to the inferred divergence level, and the scale bar indicates the number of inferred nucleic acid substitutions per site. *Spinacia oleracea* was used to root the tree.

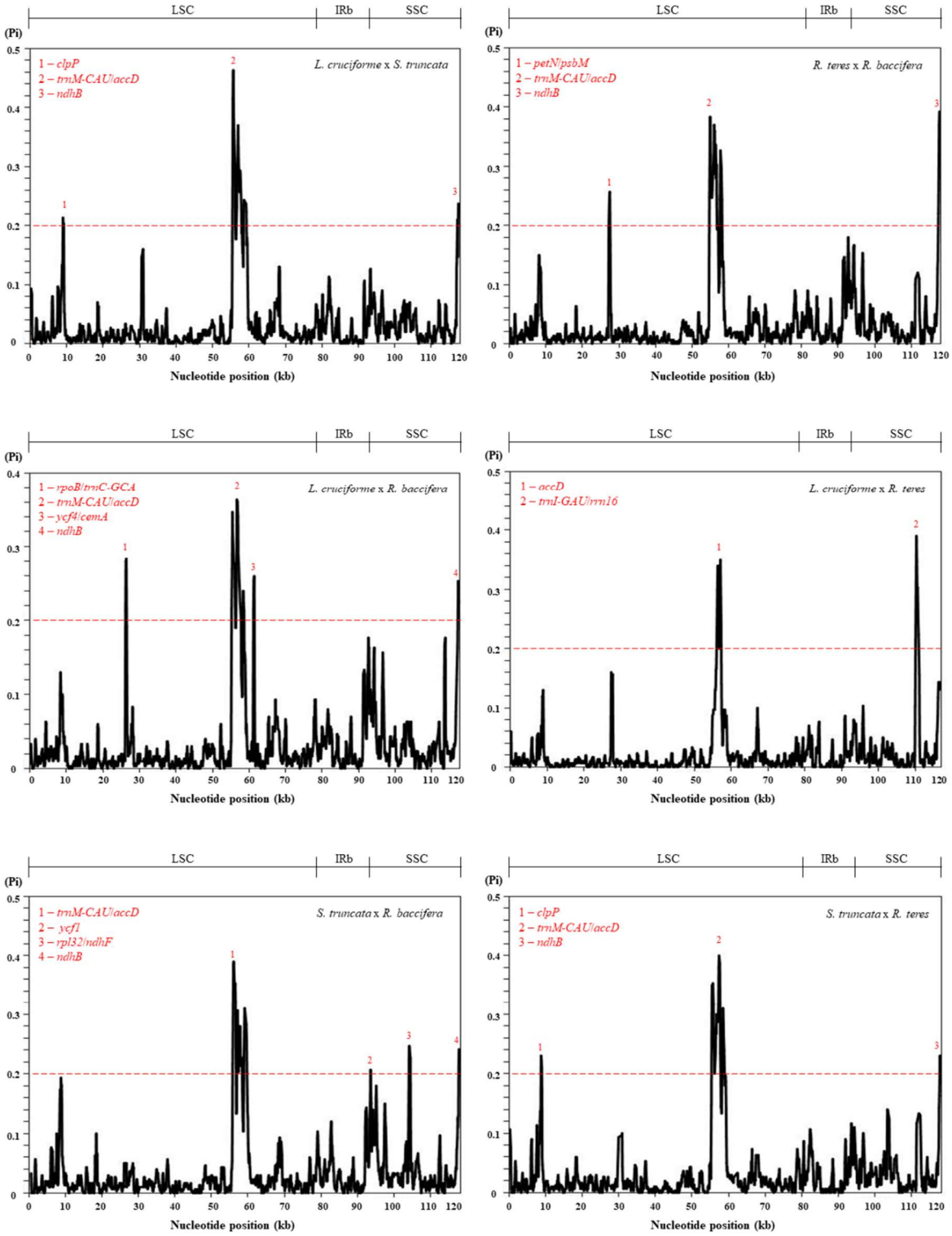


Figure 7. Sliding window analysis of the combination of whole plastome alignments of *Lepismium cruciforme*, *Schlumbergera truncata*, *Rhipsalis baccifera*, and *Rhipsalis teres*. The numbers indicate the regions with high nucleotide variability (Pi>0.2, above the red line). Pi (y-axis) represents the nucleotide diversity of each window, according to the nucleotide position (in kb) of the midpoint (x-axis). On the upper x-axis, the approximate intervals of the LSC, IRb, and SSC regions are indicated. Window length, 300pb. Step size, 75pb.

TABLES

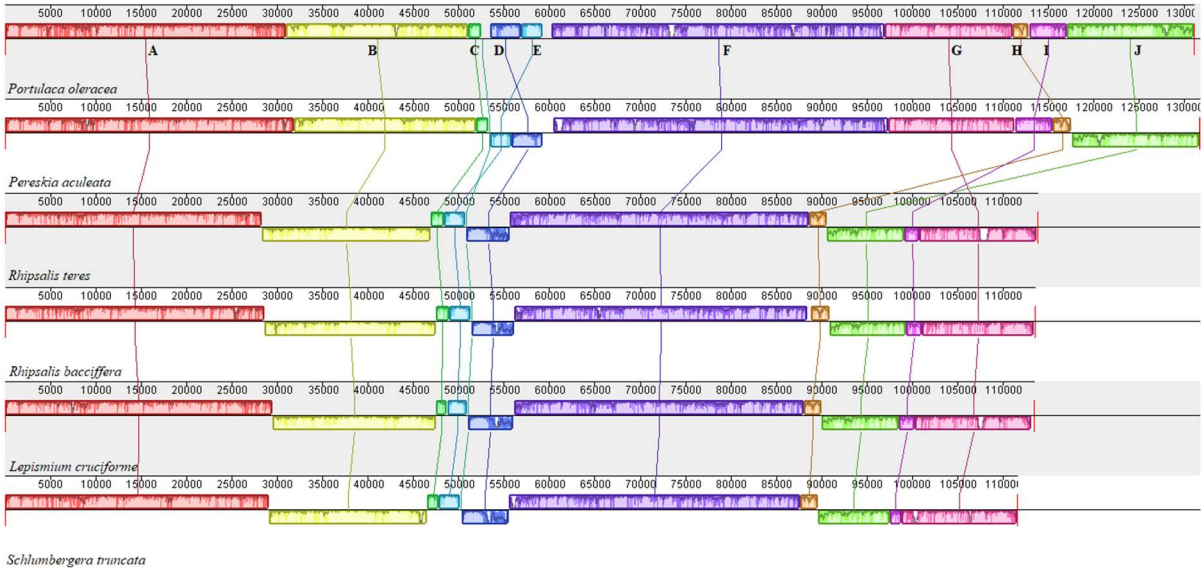
Table 1. List of genes identified in the plastomes of *Lepismium cruciforme* and *Schlumbergera truncata*

Group of gene	Name of gene
Gene expression machinery	
Ribosomal RNA genes	*rrn16; rrn23*; *rrn5; rrn4.5*
Transfer RNA genes	*trnA*-UGC*; trnC*-GCA*; trnD*-GUC*; trnE*-UUC*; trnF*-GAA*; trnfM*-CAU*; trnG*-UCC[a]*; trnG*-GCC*; trnH*-GUG*; trnI*-CAU[b]*; trnI*-GAU[a]*; trnK*-UUU[a]*; trnL*-CAA[b]*; trnL*-UAA[a]*; trnL*-UAG*; trnM*-CAU*; trnN*-GUU*; trnP*-UGG*; trnQ*-UUG*; trnR*-ACG*; trnR*-UCU*; trnS*-GCU*; trnS*-UGA*; trnS*-GGA*; trnT*-UGU*; trnT*-GGU[1]*; trnV*-GAC*; trnW*-CCA*; trnY*-GUA
Small subunit of ribosome	*rps2; rps3; rps4*; *rps7; rps8; rps11; rps12*[a]; *rps14; rps15*; *rps16*[a]; *rps18; rps19*[e]
Large subunit of ribosome	*rpl2*[d]; *rpl14; rpl16*[a]; *rpl20; rpl22; rpl32; rpl33*[*]*; rpl36*
DNA-dependent RNA polymerase	*rpoA; rpoB; rpoC1*; *rpoC2*
Genes for photosynthesis	
Subunits of photosystem I (PSI)	*psaA; psaB; psaC; psaI; psaJ; ycf3*[a]; *ycf4* [2]
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	*petA; petB*[a]*; petD*[a]*; petG; petL; petN*
Subunits of ATP synthase	*atpA; atpB; atpE; atpF*[a]; *atpH; atpI*
Subunits of NADH dehydrogenase	*ndhJ* [2]
Large subunit of Rubisco	*rbcL*
Others genes	
Maturase	*matK*[b]
Envelope membrane protein	*cemA*
Subunit of acetyl-CoA carboxylase	*accD*
C-type cytochrome synthesis gene	*ccsA*
Clp Protease	*clpP*[c]
Component of TIC complex	*ycf1*[c]
Component of 2-MD heteromeric ppAAA-ATPase complex	*ycf2*[b]
Translation initiation factor IF-1	*infA*
Pseudogenes	*ndhB; ndhC; ndhD; ndhF; ndhK; rpl23*[e]
Absent	*trnV-UAC; ndhA; ndhE; ndhG; ndhH; ndhI*

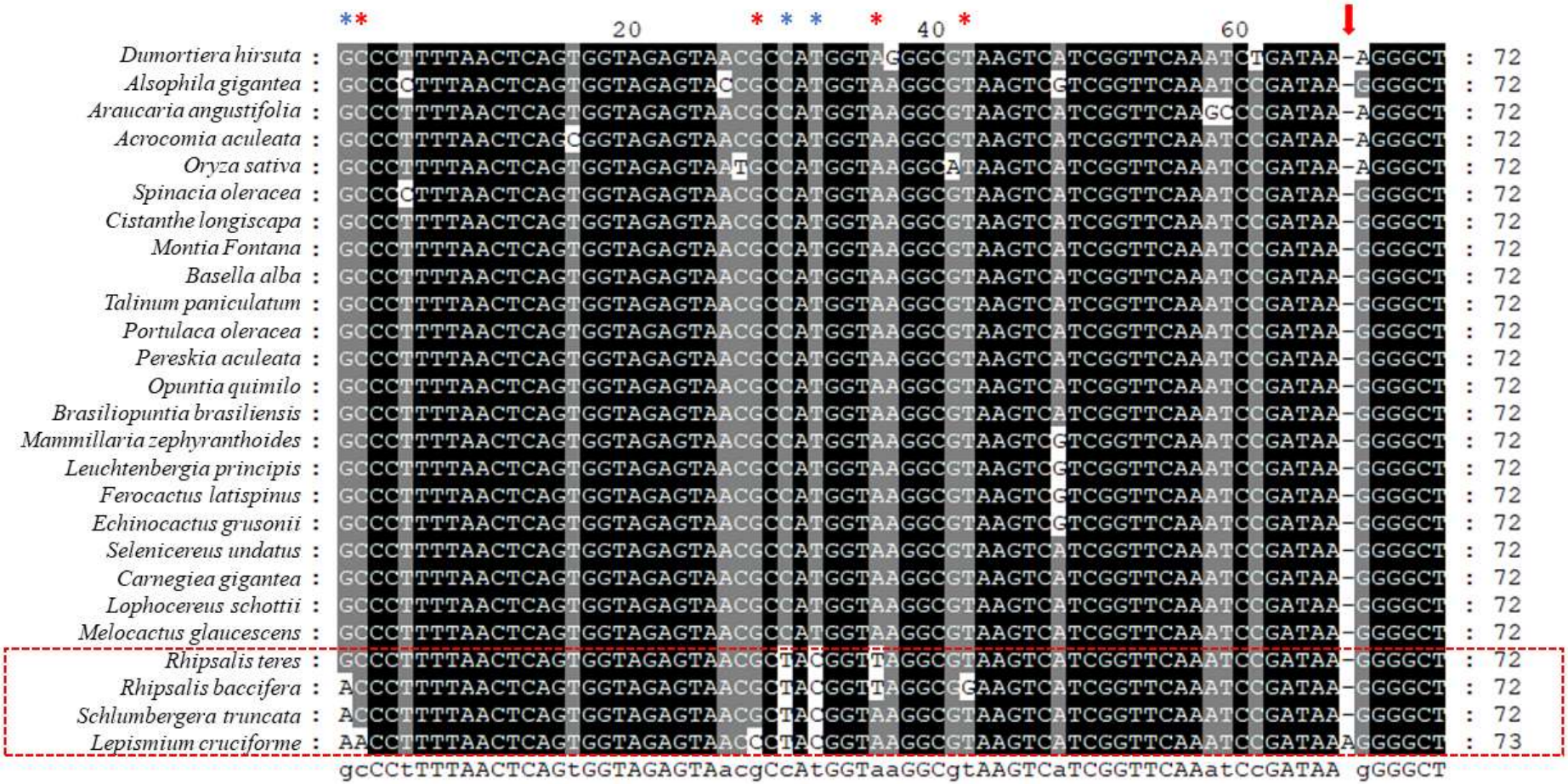
[a]Genes containing introns; [b]Duplicated gene in both species; [c]Partially duplicated genes in both species; [d]Duplicated gene in *L.cruciforme*; [e]Partially duplicated genes in *L. cruciforme*; [1]Pseudogene in *L. cruciforme* plastome; [2]Pseudogene in *S. truncata* plastome; [*]Gene loss in *L. cruciforme* plastome.

SUPPLEMENTARY MATERIALS

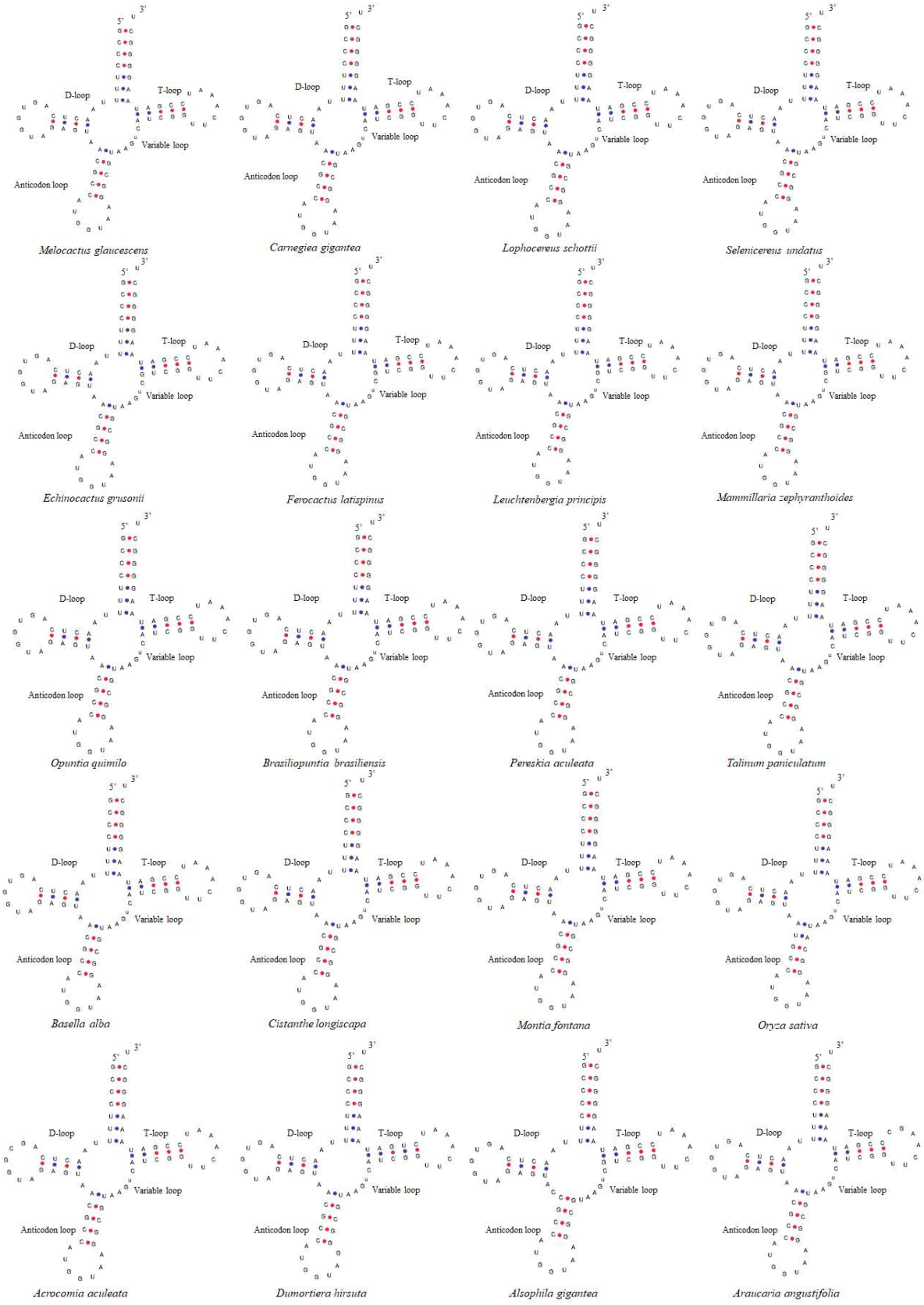
Supplementary Figures



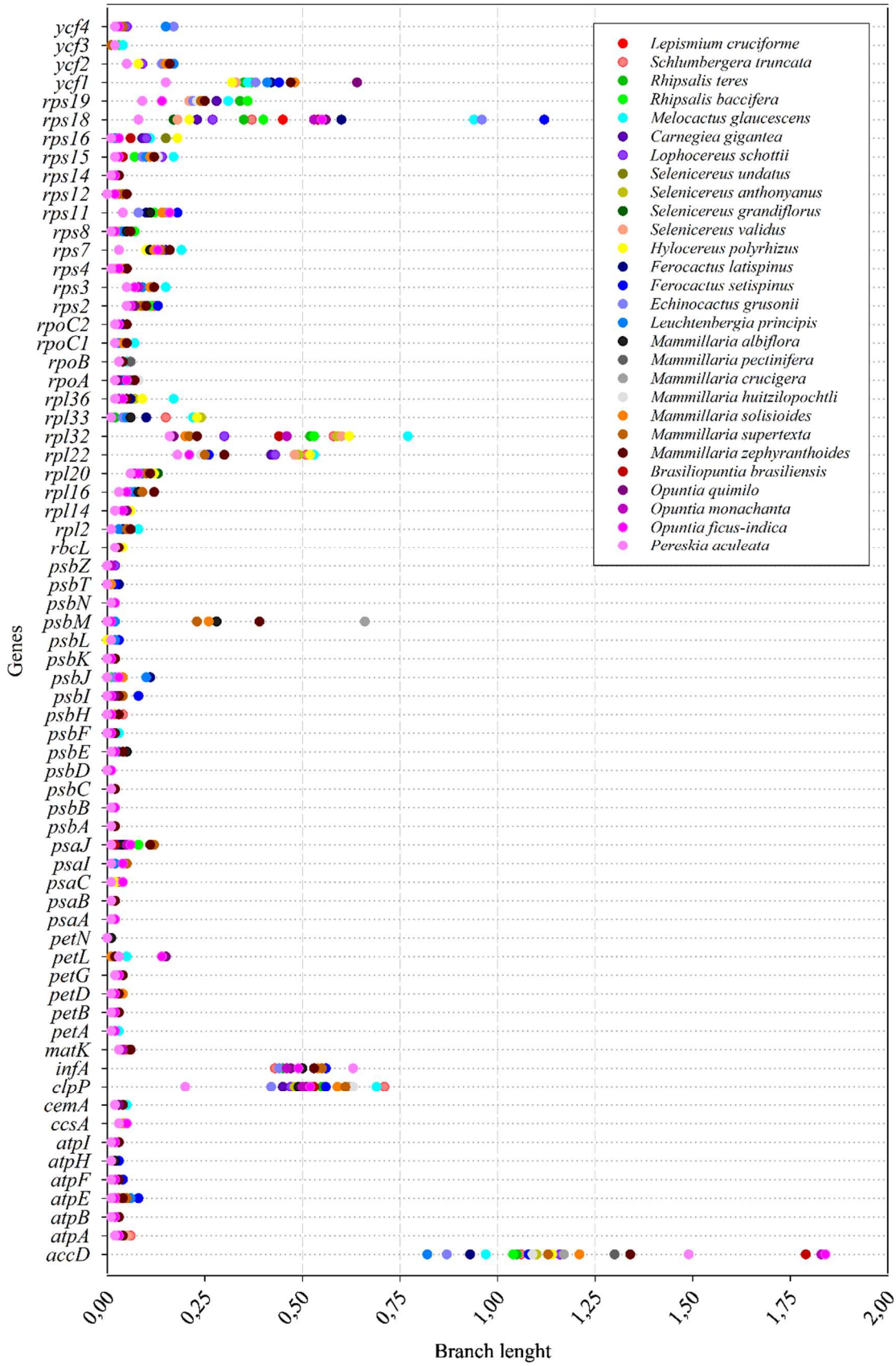
Supplementary Fig. S1. Multiple alignments of *Schlumbergera truncata*, *Lepismium cruciforme, Rhipsalis baccifera, Rhipsalis teres, Pereskia aculeata* plastomes (Cactaceae family) and *Portulaca oleracea* as reference. The region aligned corresponds to complete plastome sequence, except for removal of IRA. Locally collinear blocks (LCB) are color-coded. More details about the gene content of the LCBs involved in rearrangements are present in Figure 3.



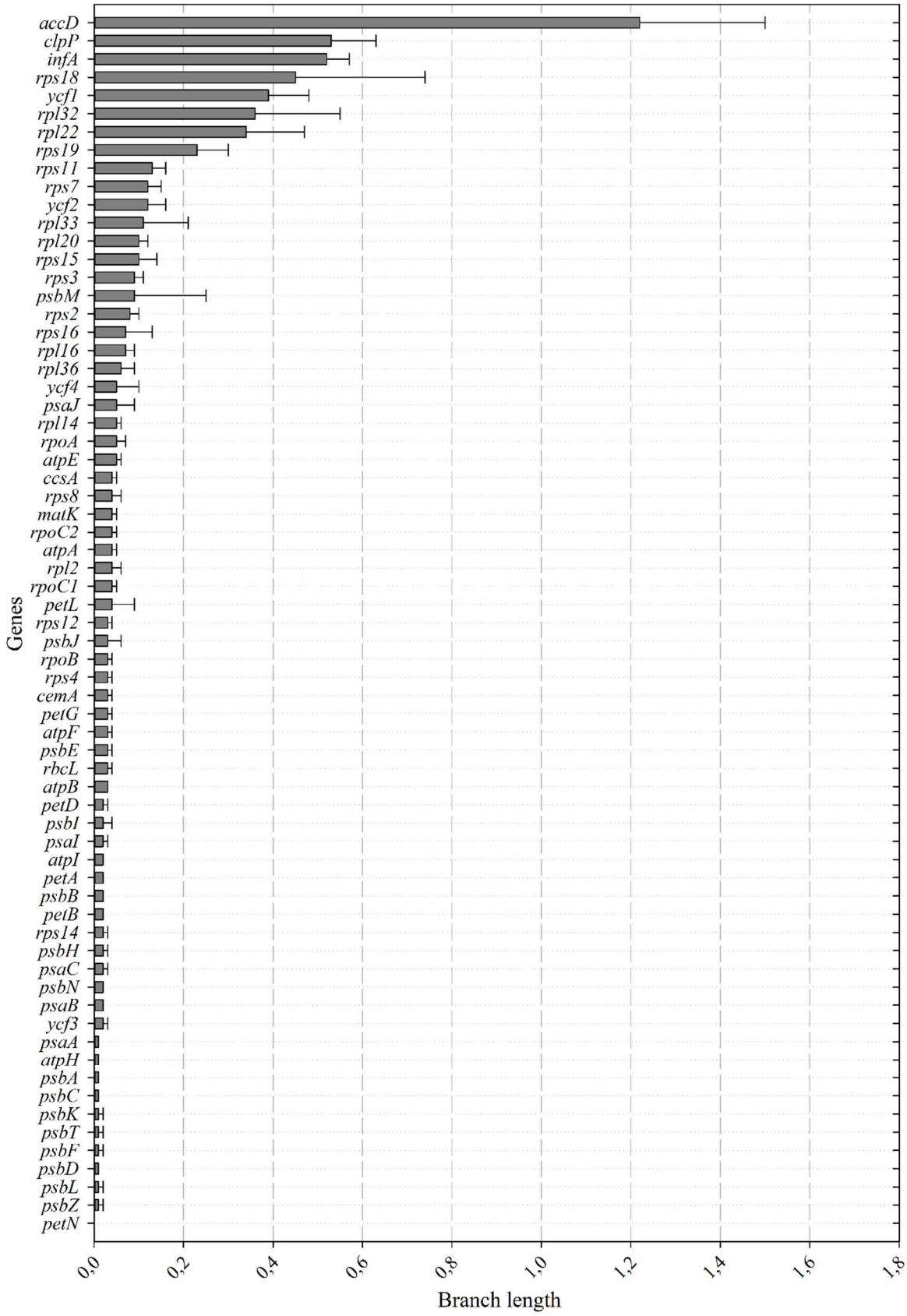
Supplementary Fig. S2. Multiple sequence alignment of the plastid threonine (trnT-GGU) gene. Asterisks represent mutation events in the sequences of species Rhipsalideae tribe (highlighted by the red square); blue asterisks signify transitions and red asterisks indicate transversions. The red arrow points to the insertion of a nucleotide in the *Lepismium cruciform* gene sequence.



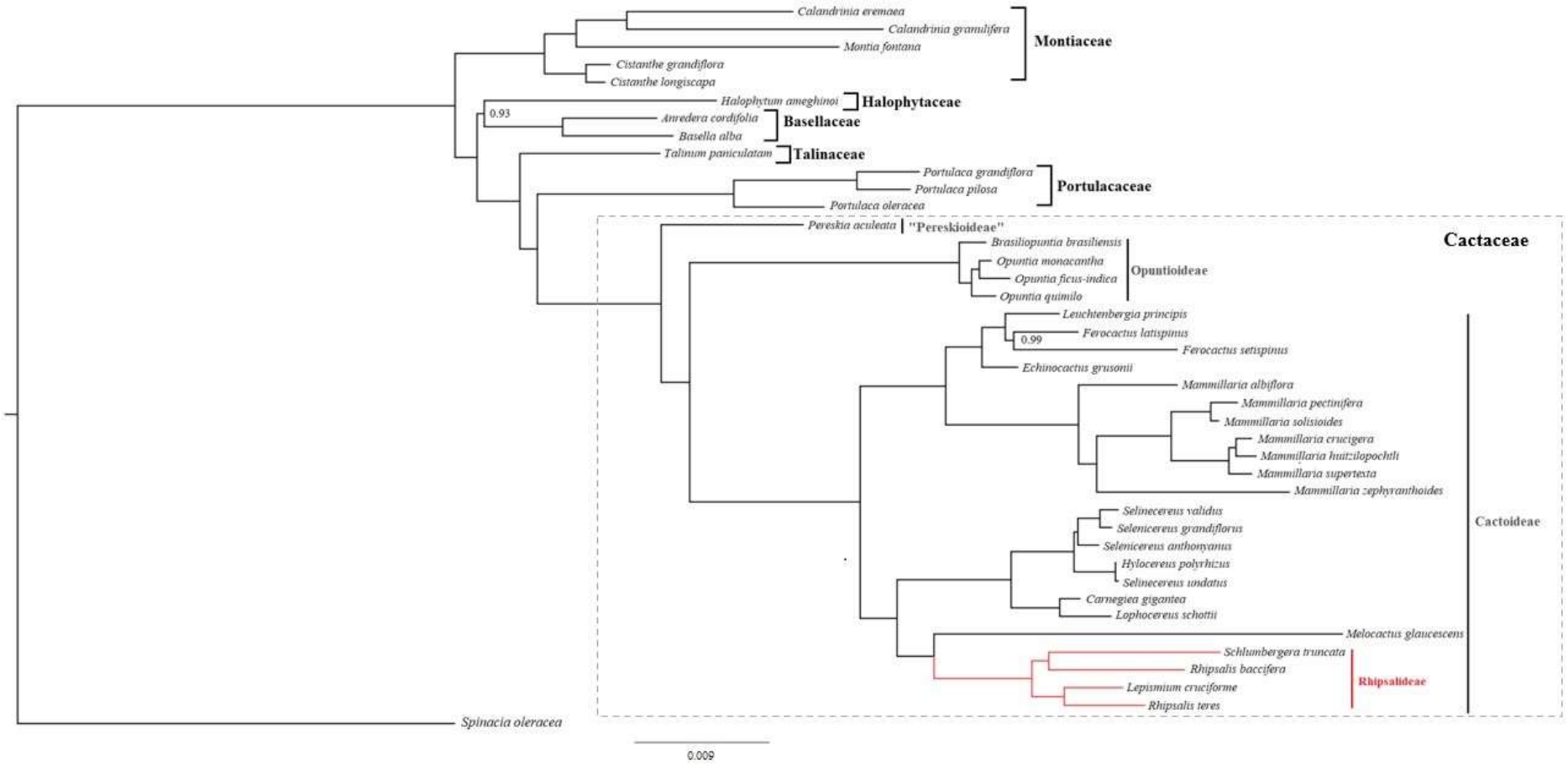
Supplementary Fig. S3. Putative secondary structure of *trnT*-GGU in different taxonomic groups of plants. Elevenspecies of the Cactaceae family, along with species of dicots, monocots, marcanthiophyta, pteridophyte, and gymnosperm were included. All species show the perfect cloverleaf structure, containing the D-loop, T-loop, variable loop, and anticodon loop. The secondary structure was predicted by the tRNAscan-SE.



Supplementary Fig. S4. Divergence of protein-coding genes in the plastomes of family Cactaceae based on phylogenetic reconstruction. The gene divergence was estimated by the sum of total branch lengths until the common ancestor node in each gene tree inferred.



Supplementary Fig. S5. Molecular evolution analyses of plastid genes within the family Cactaceae. The divergence of protein-coding genes is shown. The gene divergence was estimated by the sum of total branch lengths in each gene tree inferred (mean ± SD).



Supplementary Fig. S6. Bayesian inference (BI) analysis based on 54 protein-coding plastid genes shows the position of *Lepismium cruciforme* and *Schlumbergera truncata* (highlighted in red) within the suborder Cactineae. The families are indicated to the right of the species. Numbers in front of the nodes are values from posterior probabilities (PP) from BI; nodes with 100% of PP values were omitted. The branch length is proportional to the inferred divergence level, and the scale bar indicates the number of inferred nucleic acid substitutions per site. *Spinacia oleracea* was used to root the tree.

Supplementary Tables

Supplementary Table S1. List of species included in comparative analysis of *trnT-GGU* structure secondary

Species	Taxonomic Group / Family	GenBank
Lepismium cruciforme	Eudicotyledons/Cactaceae	ON005622
Schlumbergera truncata	Eudicotyledons/Cactaceae	ON005623
Rhipsalis teres	Eudicotyledons/Cactaceae	MT387452
Rhipsalis baccifera	Eudicotyledons/Cactaceae	MT821847
Melocactus glaucescens	Eudicotyledons/Cactaceae	OK298499
Carnegiea gigantea	Eudicotyledons/Cactaceae	NC_027618
Lophocereus schottii	Eudicotyledons/Cactaceae	NC_041727
Selenicereus undatus	Eudicotyledons/Cactaceae	NC_053698
Echinocactus grusonii	Eudicotyledons/Cactaceae	MW553048
Leuchtenbergia principis	Eudicotyledons/Cactaceae	MW553057
Ferocactus latispinus	Eudicotyledons/Cactaceae	MW553072
Mammillaria zephyranthoides	Eudicotyledons/Cactaceae	MN517611
Brasiliopuntia brasiliesis	Eudicotyledons/Cactaceae	OK448351
Opuntia quimilo	Eudicotyledons/Cactaceae	MN114084
Pereskia aculeata	Eudicotyledons/Cactaceae	OK448353
Talinum paniculatum	Eudicotyledons/Talinaceae	NC_037748.1
Basella alba	Eudicotyledons/Basellaceae	NC_041293.1
Cistanthe longiscapa	Eudicotyledons/Montiaceae	NC_035140.1
Montia fontana	Eudicotyledons/Montiaceae	NC_041269.1
Spinacia oleracea	Eudicotyledons/Chenopodiaceae	NC_002202.1
Oryza sativa	Monocotyledons/Poaceae	NC_031333.1
Acrocomia aculeata	Monocotyledons/Arecaceae	NC_037084.1
Dumortiera hirsuta	Marchantiophyta/Dumortieraceae	NC_039590.1
Alsophila gigantea	Pteridophyte/Cyatheaceae	NC_044079.1
Araucaria angustifolia	Gymnosperm/Araucariaceae	NC_039155.1

Supplementary Table S2. List of species included in the codon usage analysis

Species	Family	GenBank
Lepismium cruciforme	Cactaceae	ON005622
Schlumbergera truncata	Cactaceae	ON005623
Rhipsalis teres	Cactaceae	MT387452
Rhipsalis baccifera	Cactaceae	MT821847
Melocactus glaucescens	Cactaceae	OK298499
Carnegiea gigantea	Cactaceae	NC_027618
Lophocereus schottii	Cactaceae	NC_041727
Hylocereus polyrhizus	Cactaceae	MW553055
Selinecereus anthonyanus	Cactaceae	MW553068
Selenicereus grandiflorus	Cactaceae	MW553069
Selenicereus undatus	Cactaceae	NC_053698
Selenicereus validus	Cactaceae	MW553070
Echinocactus grusonii	Cactaceae	MW553048
Leuchtenbergia principis	Cactaceae	MW553057
Ferocactus latispinus	Cactaceae	MW553072
Ferocactus setispinus	Cactaceae	MW553071
Mammillaria albiflora	Cactaceae	MN517610
Mammillaria supertexta	Cactaceae	MN508963
Mammillaria zephyranthoides	Cactaceae	MN517611
Brasiliopuntia brasiliesis	Cactaceae	OK448351
Opuntia ficus-indica	Cactaceae	OK448352
Opuntia monacantha	Cactaceae	MZ579523
Opuntia quimilo	Cactaceae	MN114084
Portulaca oleracea	Portulacaceae	NC_041264.1
Spinacia oleracea	Chenopodiaceae	NC_002202.1

Supplementary Table S3. List of species included in the Cactineae phylogeny analysis

Species	Family	GenBank
Lepismium cruciforme	Cactaceae	ON005622
Schlumbergera truncata	Cactaceae	ON005623
Rhipsalis teres	Cactaceae	MT387452
Rhipsalis baccifera	Cactaceae	MT821847
Melocactus glaucescens	Cactaceae	OK298499
Carnegiea gigantea	Cactaceae	NC_027618
Lophocereus schottii	Cactaceae	NC_041727
Hylocereus polyrhizus	Cactaceae	MW553055
Selinecereus anthonyanus	Cactaceae	MW553068
Selenicereus grandiflorus	Cactaceae	MW553069
Selenicereus undatus	Cactaceae	NC_053698
Selenicereus validus	Cactaceae	MW553070
Echinocactus grusonii	Cactaceae	MW553048
Leuchtenbergia principis	Cactaceae	MW553057
Ferocactus latispinus	Cactaceae	MW553072
Ferocactus setispinus	Cactaceae	MW553071
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
Brasiliopuntia brasiliesis	Cactaceae	OK448351
Opuntia ficus-indica	Cactaceae	OK448352
Opuntia monacantha	Cactaceae	MZ579523
Opuntia quimilo	Cactaceae	MN114084
Pereskia aculeata	Cactaceae	OK448353
Anredera cordifolia	Basellaceae	NC_041274.1
Basella alba	Basellaceae	NC_041293.1
Calandrinia eremea	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 pilosa	Portulacaceae	NC_036236.1
Portulaca oleracea	Portulacaceae	NC_041264.1
Talinum paniculatum	Talinaceae	NC_037748.1
Spinacia oleracea *	Chenopodiaceae	NC_002202.1

* Outgroup.

Supplementary Table S4. General features of plastid genomes within the family Cactaceae.

Species	Subfamily	Size (bp)	LSC (bp)	SSC (bp)	IR (bp)	Number of genes	GC (%)	GenBank
Lepismium cruciforme	Cactoideae	123,113	79,309	24,454	9,675	99	36.5	ON005622
Schlumbergera truncata	Cactoideae	119,945	80,276	22,909	8,380	99	36.6	ON005623
M. glaucescens	Cactoideae	129,806	51,465	19,097	29,622	98	36.5	OK298499
Rhipsalis teres	Cactoideae	122,389	81,397	24,016	8,488	99	36.7	MT387452
Rhipsalis baccifera	Cactoideae	122,333	81,459	23,531	8,530	101	36.7	MT821847
Hylocereus polyrhizus	Cactoideae	133,146	68,076	21,716	21,677	99	36.4	MW553055
Selenicereus anthonyanus	Cactoideae	133,317	68,203	21,766	21,674	99	36.4	MW553068
S. grandiflorus	Cactoideae	134,211	68,839	22,014	21,679	99	36.3	MW553069
S. undatus	Cactoideae	133,326	68,256	21,716	21,677	99	36.4	NC_053698
S.validus	Cactoideae	134,211	68,839	22,014	21,679	99	36.3	MW553070
Carnegiea gigantea	Cactoideae	113,064	-	-	-	99	36.7	NC_027618
Lophocereus schottii	Cactoideae	113,204	-	-	-	99	36.5	NC_041727
Echinocactus grusonii	Cactoideae	119,201	-	-	-	101	36.1	MW553048
Ferocactus latispinus	Cactoideae	116,692	-	-	-	99	36.2	MW553072
Ferocactus setispinus	Cactoideae	110,524	-	-	-	97	36.2	MW553071
Leuchtenbergia principis	Cactoideae	116,863	-	-	-	99	36.2	MW553057
Mammillaria albiflora	Cactoideae	110,789	78,380	31,061	674	96	36.4	MN517610
M. pectinifera	Cactoideae	108,561	72,273	29,744	772	95	36.4	MN519716
M. crucigera	Cactoideae	115,505	71,565	29,418	7,261	96	36.3	MN517613
M. huitzilopochtli	Cactoideae	115,886	71,997	29,401	7,244	97	36.3	MN517612
M. solisioides	Cactoideae	115,356	71,690	29,238	7,214	95	36.4	MN518341
M. supertexta	Cactoideae	116,175	72,240	29,445	7,245	97	36.4	MN508963
M. zephyranthoides	Cactoideae	107,343	71,811	7,281	14,126	95	38.5	MN517611
Pereskia aculeata	“Pereskioideae”	140,922	88,158	34,268	9,248	113	35.8	OK448353
Brasiliopuntia brasiliensis	Opuntioideae	162,211	87,186	4,393	35,316	109	36.8	OK448351
Opuntia ficus-indica	Opuntioideae	152,387	87,248	4,109	30,515	110	36.7	OK448352
Opuntia monacantha	Opuntioideae	149,076	101,384	4,110	21,791	110	36.6	MZ579523
Opuntia quimilo	Opuntioideae	150,347	101,475	4,115	22,392	109	36.6	MN114084

Supplementary Table S5. List of RNA editing sites predicted by PREP program. The species analyzed belong to the subfamily Cactoideae (Cactaceae family). The abbreviation "aa pos" meaning the amino acid position of the RNA editing site in the aligned sequence

Gene	aa pos	*L. cruciforme*	*S. truncata*	*F. latispinus*	*F. setispinus*	*E. grusonii*	*L. principis*	*H. polyrhizus* and *S. anthonyanus*	*S. grandiflorus*	*S. validus*	Observation based on Silva et al. (2021) and Dalla Costa et al. (2021/2022) predictive analyses
atpA	305	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	Editing site conserved in subf. Cactoideae, except in *M. albiflora* (T fixed)
	424	CCC (P) => TCC (S)	CCC (P) => TCC (S)	CCC (P) => TCC (S)	CCC (P) => TCC (S)	CCC (P) => TCC (S)	CCC (P) => TCC (S)	CCC (P) => TCC (S)	CCC (P) => TCC (S)	CCC (P) => TCC (S)	Editing site conserved in subf. Cactoideae
atpB	138	ACC (T) => ATC (I)	ACC (T) => ATC (I)	ACC (T) => ATC (I)	ACC (T) => ATC (I)	ACC (T) => ATC (I)	ACC (T) => ATC (I)	ACC (T) => ATC (I)	ACC (T) => ATC (I)	ACC (T) => ATC (I)	Editing site conserved in subf. Cactoideae
atpI	33	TTC (F)	TTC (F)	TTC (F)	TTC (F)	CTC (L) => TTC (F)	TTC (F)	TTC (F)	TTC (F)	TTC (F)	T fixed in other species of the subf. Cactoideae
matK	217	CAT (H) => TAT (Y)	CAT (H) => TAT (Y)	TAT (Y)	TAT (Y)	CAT (H) => TAT (Y)	CAT (H) => TAT (Y)	CAT (H) => TAT (Y)	CAT (H) => TAT (Y)	CAT (H) => TAT (Y)	Editing site conserved in subf. Cactoideae, except in *C. gigantea*, *L. schottii* and *M. glaucescens* (T fixed)
petB	140	CGG (R) => TGG (W)	CGG (R) => TGG (W)	CGG (R) => TGG (W)	CGG (R) => TGG (W)	CGG (R) => TGG (W)	CGG (R) => TGG (W)	CGG (R) => TGG (W)	CGG (R) => TGG (W)	CGG (R) => TGG (W)	Editing site conserved in subf. Cactoideae
	163	TCA(S)	TCA(S)	TCA(S)	TCA(S)	TCA(S)	TCA(S)	TCA(S)	CCA (P) => TCA (S)	TCT (S)	T fixed in other species of the subf. Cactoideae
psbB	193	TTT (F)	TAT (Y)	TTT (F)	CTT (L) => TTT (F)	TTT (F)	TTT (F)	TTT (F)	TTT (F)	TTT (F)	T fixed in other species of the subf. Cactoideae
psbF	26	TCT (S) => TTT (F)	TTT (F)	TTT (F)	TTT (F)	TTT (F)	TTT (F)	TCT (S) => TTT (F)	TCT (S) => TTT (F)	TCT (S) => TTT (F)	Editing site in *C. gigantea*, *L. schottii*, *S. undatus*, *M. glaucescens* and *R. baccifera*; the other species has a T fixed
psbL	1	ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	Editing site conserved in subf. Cactoideae
rpl2	217	GTA (V)	GTA (V)	GCA (A) => GTA (V)	GTA (V)	GTA (V)	GTA (V)	GTA (V)	GTA (V)	GTA (V)	T fixed in other species of the subf. Cactoideae

rpl20	20	TCC (S) => TTC (F)	TCC (S) => TTC (F)	TCC (S) => TTC (F)	TCC (S) => TTC (F)	TCC (S) => TTC (F)	TCT (S) => TTT (F)	TCT (S) => TTT (F)	TTT (F)	TTT (F)	Editing site conserved in subf. Cactoideae
	83	CAC (H) => TAC (Y)	CAC (H) => TAC (Y)	CAC (H) => TAC (Y)	CAC (H) => TAC (Y)	CAC (H) => TAC (Y)	CAC (H) => TAC (Y)	CAC (H) => TAC (Y)	CAC (H) => TAC (Y)	CAC (H) => TAC (Y)	Editing site conserved in subf. Cactoideae, except in *M. zephyranthoides* (T fixed)
	87	TCG (S) => TTG (L)	TCG (S) => TTG (L)	TCG (S) => TTG (L)	TCG (S) => TTG (L)	TCG (S) => TTG (L)	TCG (S) => TTG (L)	TCG (S) => TTG (L)	TCG (S) => TTG (L)	TCG (S) => TTG (L)	Editing site conserved in subf. Cactoideae, except in *M. glaucescens* (T fixed)
rpoA	28	CAT (H) => TAT (Y)	CAT (H) => TAT (Y)	TAT (Y)	TAT (Y)	TAT (Y)	TAT (Y)	CAT (H) => TAT (Y)	CAT (H) => TAT (Y)	CAT (H) => TAT (Y)	Editing site in *C. gigantea, L. schottii, S. undatus, M. glaucescens, R. baccifera* and *R. teres*; the other species has a T fixed
	294	TTG (L)	TTG (L)	TTG (L)	TCG (S) => TTG (L)	TTG (L)	TTG (L)	TTG (L)	TTG (L)	TTG (L)	Editing site in *M. albiflora, M. zephyranthoides* and *M. glaucescens*; the other species has a T fixed
rpoB	158	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	Editing site conserved in subf. Cactoideae
	184	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	Editing site in *C. gigantea, L. schottii, S. undatus, M. glaucescens, R. baccifera, R. teres, M. albiflora* and *M. zephyranthoides*; the other species has a T fixed
	189	TCG (S) => TTG (L)	TCG (S) => TTG (L)	TCG (S) => TTG (L)	TCG (S) => TTG (L)	TCG (S) => TTG (L)	TCG (S) => TTG (L)	TCG (S) => TTG (L)	TCG (S) => TTG (L)	TCG (S) => TTG (L)	Editing site conserved in subf. Cactoideae
	664	TAT (Y)	CAT (H) => TAT (Y)	TAT (Y)	TAT (Y)	TAT (Y)	TAT (Y)	TAT (Y)	TAT (Y)	TAT (Y)	T fixed in other species of the subf. Cactoideae
rpoC1	14	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	Editing site conserved in subf. Cactoideae
	157	CTT (L) => TTT (F)	CTT (L) => TTT (F)	TTT (F)	TTT (F)	TTT (F)	TTT (F)	TTT (F)	TTT (F)	TTT (F)	Editing site in *R. baccifera* and *R. teres*; the other species has a T fixed
	210	ACT (T) => ATT (I)	ACT (T) => ATT (I)	ACT (T) => ATT (I)	ACT (T) => ATT (I)	ACT (T) => ATT (I)	ACT (T) => ATT (I)	ACT (T) => ATT (I)	ACT (T) => ATT (I)	ACT (T) => ATT (I)	Editing site conserved in subf. Cactoideae

	322	AGG (R)	TGG (W)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	Editing site in *S. undatus*
rpoC2	69	ATA (I)	ATA (I)	ATA (I)	ACA (T) => ATA (I)	ATA (I)	ATA (I)	ATA (I)	ATA (I)	ATA (I)	T fixed in other species of the subf. Cactoideae
	163	GTA (V)	GTA (V)	TTA (L)	TTA (L)	GTA (V)	TTA (L)	CCA (P) => CTA (L)	CCA (P) => CTA (L)	CCA (P) => CTA (L)	T fixed in other species of the subf. Cactoideae
	498	ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	ACG (T) => ATG (M)	Editing site conserved in subf. Cactoideae
	605	CAT (H) => TAT (Y)	CAT (H) => TAT (Y)	TAT (Y)	TAT (Y)	TAT (Y)	TAT (Y)	TAT (Y)	TAT (Y)	TAT (Y)	Editing site in *R. baccifera* and *R. teres*; the other species has a T fixed
	652	TTT (F)	TCT (S) => TTT (F)	TTT (F)	TTT (F)	TTT (F)	TTT (F)	TTT (F)	TTT (F)	TTT (F)	T fixed in other species of the subf. Cactoideae
	742	TGG (W)	TGG (W)	CGG (R) => TGG (W)	CGG (R) => TGG (W)	CGG (R) => TGG (W)	CGG (R) => TGG (W)	TGG (W)	TGG (W)	TGG (W)	Editing site in *Mammillaria* genus; the other species has a T fixed
	747	CCC (P)	CCC (P)	CCC (P)	CCC (P)	CCC (P)	CCC (P)	GCC (A) => GTC (V)	GCC (A) => GTC (V)	GCC (A) => GTC (V)	C fixed in other species of the subf. Cactoideae
	769	GTT (V)	GCT (A) => GTT (V)	GTT (V)	GTT (V)	GTT (V)	GTT (V)	GTT (V)	GTT (V)	GTT (V)	T fixed in other species of the subf. Cactoideae
rps2	83	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	Editing site conserved in subf. Cactoideae, except in *M. albiflora* (T fixed)
	203	TCG (S)	TCG (S)	TCG (S)	CCG (P) => TCG (S)	TCG (S)	TCG (S)	TCG (S)	TCG (S)	TCG (S)	T fixed in other species of the subf. Cactoideae
rps8	48	GTG (V)	GTG (V)	GCG (A) => GTG (V)	GCG (A) => GTG (V)	ACG (T)	GCA (A) => GTA (V)	TCG (S)	TCG (S)	TCG (S)	Editing site in *M. glaucescens*; T fixed in other species of the subf. Cactoideae
rps14	27	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	TCA (S) => TTA (L)	Editing site conserved in subf. Cactoideae
	50	CCA (P) => CTA (L)	CCA (P) => CTA (L)	CCA (P) => CTA (L)	CCA (P) => CTA (L)	CCA (P) => CTA (L)	CCA (P) => CTA (L)	CCA (P) => CTA (L)	CCA (P) => CTA (L)	CCA (P) => CTA (L)	Editing site conserved in subf. Cactoideae

Supplementary Table S6. List of simple sequence repeats (SSRs) identified in the *L. cruciforme* and *S. truncata* plastomes

***Lepismium cruciforme* plastome**

SSR sequence	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total
A/T	-	-	-	-	-	65	38	37	8	2	-	2	3	1	156
C/G	-	-	-	-	-	1	1	1	-	-	-	-	-	-	3
AC/GT	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1
AG/CT	-	13	1	-	-	-	-	-	-	-	-	-	-	-	14
AT/AT	-	15	5	3	1	-	1	-	-	-	-	-	-	-	25
AAG/CTT	-	2	-	-	-	-	-	-	-	-	-	-	-	-	2
AAT/ATT	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1
ACG/CGT	-	-	-	-	-	-	1	-	-	-	-	-	-	-	1
ACT/AGT	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1
AAAG/CTTT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AATC/ATTG	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AATT/AATT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ACCT/AGGT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AGGG/CCCT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AAAGT/ACTTT	2	-	-	-	-	-	-	-	-	-	-	-	-	-	2
AAATT/AATTT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AAAAAT/ATTTTT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AAATTG/AATTTC	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AATACT/AGTATT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1

***Schlumbergera truncata* plastome**

SSR sequence	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Total
A/T	-	-	-	-	-	39	22	33	12	8	4	-	-	1	-	1	-	-	1	1	122
C/G	-	-	-	-	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	2
AC/GT	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AG/CT	-	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	10
AT/AT	-	14	3	5	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	24
AAT/ATT	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AAAG/CTTT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AAAT/ATTT	2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3
AATC/ATTG	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ACCT/AGGT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ACAGAG/CTCTGT	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1

Supplementary Table S7. Distribution of simple sequence repeats (SSR) loci in the *Lepismium cruciforme* plastome

SSR type	SSR	Size	Start	End	Location
mono	(T)11	11	112	122	*trnH-GUG/psbA* (IGS)
mono	(T)14	14	147	160	*trnH-GUG/psbA* (IGS)
mono	(A)9	9	193	201	*trnH-GUG/psbA* (IGS)
mono	(T)8	8	1842	1849	*trnK-UUU/matK* (IGS)
mono	(A)8	8	2426	2433	*matK* (CDS)
mono	(T)9	9	3077	3085	*matK* (CDS)
mono	(A)10	10	3461	3470	*matK/trnK-UUU* (IGS)
mono	(T)9	9	3817	3825	*matK/trnK-UUU* (IGS)
di	(TA)6	12	4270	4281	*trnK-UUU/rps16* (IGS)
mono	(A)9	9	4388	4396	*trnK-UUU/rps16* (IGS)
mono	(A)8	8	4459	4466	*trnK-UUU/rps16* (IGS)
mono	(T)10	10	4490	4499	*trnK-UUU/rps16* (IGS)
mono	(A)8	8	4961	4968	*rps16* (intron)
di	(AT)4	8	5269	5276	*rps16* (intron)
di	(AT)4	8	5754	5761	*rps16/trnQ-UUG* (IGS)
mono	(T)11	11	5776	5786	*rps16/trnQ-UUG* (IGS)
tri	(AAT)4	12	5794	5805	*rps16/trnQ-UUG* (IGS)
mono	(A)10	10	5807	5816	*rps16/trnQ-UUG* (IGS)
mono	(T)9	9	5970	5978	*rps16/trnQ-UUG* (IGS)
mono	(A)9	9	6523	6531	*trnQ-UUG/psbK* (IGS)
mono	(T)8	8	7087	7094	*psbK/psbI* (IGS)
mono	(A)9	9	7116	7124	*psbK/psbI* (IGS)
mono	(T)10	10	7401	7410	*psbK/psbI* (IGS)
mono	(A)8	8	7413	7420	*psbK/psbI* (IGS)
mono	(T)11	11	7596	7606	*psbK/psbI* (IGS)
mono	(A)8	8	7609	7616	*psbK/psbI* (IGS)
mono	(A)9	9	7956	7964	*trnS-GCU/clpP* (IGS)
mono	(A)8	8	8264	8271	*trnS-GCU/clpP* (IGS)
mono	(T)8	8	8719	8726	*clpP* (CDS)
tri	(GAC)9	27	8733	8759	*clpP* (CDS)
mono	(T)9	9	9382	9390	*trnG-UCC* (intron)
mono	(T)10	10	9459	9468	*trnG-UCC* (intron)
mono	(A)10	10	9560	9569	*trnG-UCC* (intron)
mono	(T)8	8	10122	10129	*trnR-UCU/atpA* (IGS)
mono	(A)9	9	11771	11779	*atpA/atpF* (IGS)
di	(CT)5	10	12368	12377	*atpF* (intron)
mono	(A)8	8	12394	12401	*atpF* (intron)
mono	(A)14	14	12413	12426	*atpF* (intron)
mono	(T)11	11	12496	12506	*atpF* (intron)
mono	(A)9	9	12589	12597	*atpF* (intron)

mono	(T)10	10	12677	12686	*atpF* (intron)
mono	(T)8	8	15535	15542	*rps2* (CDS)
mono	(T)8	8	16069	16076	*rps2/rpoC2* (IGS)
mono	(A)8	8	16221	16228	*rps2/rpoC2* (IGS)
mono	(T)9	9	16392	16400	*rpoC2* (CDS)
mono	(A)8	8	17452	17459	*rpoC2* (CDS)
mono	(A)9	9	17744	17752	*rpoC2* (CDS)
mono	(T)8	8	18108	18115	*rpoC2* (CDS)
di	(AT)4	8	18242	18249	*rpoC2* (CDS)
mono	(T)11	11	18276	18286	*rpoC2* (CDS)
mono	(A)8	8	18420	18427	*rpoC2* (CDS)
mono	(T)8	8	18747	18754	*rpoC2* (CDS)
di	(AT)5	10	19578	19587	*rpoC2* (CDS)
mono	(A)8	8	22028	22035	*rpoC1* (CDS)
mono	(T)8	8	25054	25061	*rpoB* (CDS)
mono	(T)10	10	25158	25167	*rpoB* (CDS)
mono	(A)9	9	26025	26033	*rpoB/trnC-GCA* (IGS)
mono	(T)10	10	26100	26109	*rpoB/trnC-GCA* (IGS)
mono	(T)10	10	26384	26393	*rpoB/trnC-GCA* (IGS)
di	(AT)4	8	26395	26402	*rpoB/trnC-GCA* (IGS)
di	(TA)5	10	26408	26417	*rpoB/trnC-GCA* (IGS)
mono	(A)9	9	26506	26514	*rpoB/trnC-GCA* (IGS)
mono	(A)8	8	26652	26659	*trnC-GCA/petN* (IGS)
penta	(ATTAA)3	15	27117	27131	*trnC-GCA/petN* (IGS)
mono	(T)9	9	27151	27159	*trnC-GCA/petN* (IGS)
hexa	(TATTAG)3	18	27184	27201	*trnC-GCA/petN* (IGS)
di	(TG)4	8	27868	27875	*petN/psbM* (IGS)
di	(AT)6	12	28438	28449	*psbM/trnD-GUC* (IGS)
mono	(A)8	8	28873	28880	*psbM/trnD-GUC* (IGS)
mono	(T)8	8	29216	29223	*trnD-GUC/ndhJ* (IGS)
mono	(A)8	8	30431	30438	*trnF-GAA/trnL-UAA* (IGS)
di	(TA)5	10	30783	30792	*trnL-UAA* (intron)
mono	(T)8	8	30934	30941	*trnL-UAA* (intron)
mono	(T)10	10	31033	31042	*trnL-UAA* (intron)
mono	(T)10	10	31073	31082	*trnL-UAA* (intron)
mono	(A)10	10	31084	31093	*trnL-UAA* (intron)
di	(AT)4	8	31491	31498	*trnL-UAA/trnT-UGU* (IGS)
mono	(T)9	9	31662	31670	*trnL-UAA/trnT-UGU* (IGS)
mono	(T)9	9	31851	31859	*trnT-UGU/rps4* (IGS)
mono	(T)8	8	31872	31879	*trnT-UGU/rps4* (IGS)
di	(AT)4	8	33002	33009	*rps4/trnS-GGA* (IGS)
tetra	(TGAT)3	12	33339	33350	*trnS-GGA/ycf3* (IGS)
tri	(TAC)4	12	33723	33734	*trnS-GGA/ycf3* (IGS)

mono	(A)10	10	33841	33850	*trnS-GGA/ycf3* (IGS)
tri	(AGA)4	12	33852	33863	*trnS-GGA/ycf3* (IGS)
mono	(T)11	11	34145	34155	*ycf3* (intron)
mono	(A)8	8	34525	34532	*ycf3* (intron)
mono	(A)10	10	35184	35193	*ycf3* (intron)
mono	(A)8	8	35326	35333	*ycf3* (intron)
mono	(T)10	10	36206	36215	*ycf3/psaA* (IGS)
di	(AT)4	8	36382	36389	*ycf3/psaA* (IGS)
coumpand	(A)8(GA)4	16	36564	36579	*ycf3/psaA* (IGS)
di	(CT)4	8	43110	43117	*trnS-UGA* (CDS)
mono	(A)10	10	43218	43227	*trnS-UGA/psbC* (IGS)
mono	(A)10	10	43259	43268	*trnS-UGA/psbC* (IGS)
mono	(A)10	10	46113	46122	*psbD/trnT-GGU* (IGS)
di	(AT)9	18	46746	46763	*trnT-GGU/trnE-UUC* (IGS)
mono	(A)10	10	46954	46963	*trnT-GGU/trnE-UUC* (IGS)
mono	(A)8	8	47452	47459	*trnY-GUA/ndhK* (IGS)
mono	(T)8	8	47514	47521	*trnY-GUA/ndhK* (IGS)
mono	(A)11	11	47565	47575	*trnY-GUA/ndhK* (IGS)
mono	(T)8	8	47713	47720	*ndhK* (CDS)
mono	(T)15	15	48283	48297	*ndhK/ndhC* (IGS)
di	(AT)4	8	48899	48906	*ndhC/rbcL* (IGS)
di	(GA)4	8	49575	49582	*rbcL* (CDS)
di	(TA)5	10	50656	50665	*rbcL/atpB* (IGS)
di	(TA)6	12	50667	50678	*rbcL/atpB* (IGS)
di	(AT)7	14	51111	51124	*rbcL/atpB* (IGS)
mono	(A)10	10	51441	51450	*atpB* (CDS)
di	(TA)4	8	53444	53451	*atpE/trnM-CAU* (IGS)
mono	(A)9	9	53771	53779	*trnM-CAU/accD* (IGS)
mono	(A)8	8	54371	54378	*trnM-CAU/accD* (IGS)
mono	(T)10	10	54472	54481	*trnM-CAU/accD* (IGS)
mono	(T)9	9	54503	54511	*trnM-CAU/accD* (IGS)
mono	(T)9	9	55230	55238	*accD* (CDS)
mono	(A)10	10	55367	55376	*accD* (CDS)
mono	(A)15	15	55935	55949	*accD* (CDS)
mono	(A)10	10	55951	55960	*accD* (CDS)
mono	(A)8	8	56343	56350	*accD* (CDS)
mono	(T)9	9	57716	57724	*accD/psaI* (IGS)
di	(TA)4	8	57948	57955	*accD/psaI* (IGS)
mono	(A)8	8	58031	58038	*accD/psaI* (IGS)
mono	(G)10	10	58342	58351	*psaI/ycf4* (IGS)
mono	(C)9	9	58663	58671	*ycf4* (CDS)
mono	(T)10	10	58678	58687	*ycf4* (CDS)
mono	(A)9	9	59167	59175	*ycf4/cemA* (IGS)

mono	(A)8	8	59297	59304	*ycf4/cemA* (IGS)
di	(AT)4	8	60287	60294	*petA* (CDS)
mono	(A)8	8	60647	60654	*petA* (CDS)
mono	(T)10	10	62951	62960	*petL/petG* (IGS)
mono	(A)9	9	63405	63413	*trnW-CCA/trnP-UGG* (IGS)
di	(AT)4	8	63709	63716	*trnP-UGG/psaJ* (IGS)
di	(AT)5	10	64429	64438	*psaJ/rps18* (IGS)
mono	(T)8	8	64489	64496	*psaJ/rps18* (IGS)
mono	(T)8	8	65207	65214	*rpl20* (CDS)
mono	(A)8	8	65336	65343	*rpl20* (CDS)
mono	(T)8	8	65603	65610	*rpl20* (CDS)
mono	(A)9	9	65881	65889	*rpl20/rps12* (IGS)
mono	(A)8	8	67176	67183	*clpP/psbB* (IGS)
mono	(T)8	8	68896	68903	*psbB/psbT* (IGS)
mono	(T)8	8	68911	68918	*psbB/psbT* (IGS)
mono	(T)10	10	68961	68970	*psbB/psbT* (IGS)
mono	(A)9	9	70016	70024	*petB* (intron)
mono	(T)10	10	70121	70130	*petB* (intron)
mono	(A)10	10	70306	70315	*petB* (intron)
mono	(A)11	11	71418	71428	*petB/petD* (IGS)
di	(AG)4	8	71686	71693	*petD* (intron)
mono	(T)8	8	71695	71702	*petD* (intron)
mono	(T)15	15	71707	71721	*petD* (intron)
di	(AT)4	8	71863	71870	*petD* (intron)
di	(TA)4	8	71872	71879	*petD* (intron)
di	(TA)4	8	72337	72344	*petD* (CDS)
di	(GA)4	8	73759	73766	*rpoA* (CDS)
mono	(T)12	12	74423	74434	*rps11/rpl36* (IGS)
penta	(TTACT)3	15	74609	74623	*rpl36/infA* (IGS)
mono	(T)9	9	75525	75533	*rps8/rpl14* (IGS)
mono	(A)9	9	75548	75556	*rps8/rpl14* (IGS)
mono	(T)8	8	75664	75671	*rpl14* (CDS)
tetra	(TTTC)3	12	77355	77366	*rpl16* (intron)
hexa	(TATTTT)3	18	77400	77417	*rpl16* (intron)
mono	(A)8	8	77448	77455	*rpl16* (intron)
mono	(T)8	8	77587	77594	*rpl16* (intron)
mono	(T)9	9	78308	78316	*rps3* (CDS)
mono	(T)9	9	78572	78580	*rpl22* (CDS)
mono	(T)8	8	79174	79181	*rps19* (CDS)
mono	(T)8	8	79195	79202	*rps19* (CDS)
mono	(T)8	8	79216	79223	*rps19* (CDS)
mono	(T)12	12	79489	79500	*rps19/rpl2* (IGS)
di	(GA)4	8	81177	81184	*ycf2* (CDS)

di	(GA)4	8	81189	81196	*ycf2* (CDS)
di	(GA)4	8	82122	82129	*ycf2* (CDS)
hexa	(TTTCAA)3	18	83326	83343	*ycf2* (CDS)
mono	(A)9	9	84085	84093	*ycf2* (CDS)
di	(GA)4	8	84106	84113	*ycf2* (CDS)
tri	(TCT)4	12	84759	84770	*ycf2* (CDS)
mono	(T)8	8	88862	88869	*ycf1* (CDS)
mono	(T)9	9	90384	90392	*ycf1* (CDS)
mono	(A)8	8	91548	91555	*ycf1* (CDS)
mono	(A)8	8	91986	91993	*ycf1* (CDS)
mono	(A)10	10	92037	92046	*ycf1* (CDS)
mono	(A)8	8	92060	92067	*ycf1* (CDS)
tetra	(AATT)3	12	92069	92080	*ycf1* (CDS)
mono	(A)9	9	92924	92932	*ycf1* (CDS)
mono	(A)8	8	93082	93089	*ycf1* (CDS)
mono	(A)8	8	93280	93287	*ycf1* (CDS)
mono	(A)8	8	93956	93963	*ycf1/rps15* (IGS)
mono	(A)9	9	94488	94496	*rps15/psaC* (IGS)
mono	(T)9	9	94799	94807	*psaC/ndhD* (IGS)
mono	(A)10	10	96148	96157	*ndhD* (CDS)
mono	(T)9	9	96404	96412	*ndhD/ccsA* (IGS)
mono	(A)10	10	96427	96436	*ndhD/ccsA* (IGS)
mono	(A)10	10	97607	97616	*ccsA/trnL-UAG* (IGS)
mono	(A)8	8	98347	98354	*trnL-UAG/rpl32* (IGS)
mono	(A)8	8	98390	98397	*trnL-UAG/rpl32* (IGS)
mono	(T)10	10	98597	98606	*trnL-UAG/rpl32* (IGS)
mono	(A)10	10	98680	98689	*trnL-UAG/rpl32* (IGS)
mono	(T)10	10	98711	98720	*trnL-UAG/rpl32* (IGS)
mono	(T)8	8	99073	99080	*rpl32* (CDS)
mono	(T)8	8	99333	99340	*rpl32/ndhF* (IGS)
mono	(A)10	10	99420	99429	*rpl32/ndhF* (IGS)
mono	(T)8	8	99566	99573	*rpl32/ndhF* (IGS)
mono	(G)8	8	101059	101066	*trnN-GUU/trnR-ACG* (IGS)
mono	(A)8	8	101134	101141	*trnN-GUU/trnR-ACG* (IGS)
di	(AG)4	8	103267	103274	*rrn23* (CDS)
tetra	(CTAC)3	12	103299	103310	*rrn23* (CDS)
mono	(T)8	8	104890	104897	*rrn23* (CDS)
penta	(AAAGT)3	15	107159	107173	*trnI-GAU/rrn16* (IGS)
di	(AG)4	8	107207	107214	*trnI-GAU/rrn16* (IGS)
di	(TC)4	8	107223	107230	*trnI-GAU/rrn16* (IGS)
tetra	(AGGG)3	12	107256	107267	*trnI-GAU/rrn16* (IGS)
di	(AG)4	8	107729	107736	*trnI-GAU/rrn16* (IGS)
mono	(A)9	9	110107	110115	*trnV-GAC/rps12* (IGS)

mono	(T)16	16	112734	112749	*ndhB* (CDS)
mono	(T)9	9	112766	112774	*ndhB* (CDS)

CDS, coding sequences; IGS, intergenic spacers

Supplementary Table S8. Distribution of simple sequence repeats (SSR) loci in the *Schlumbergera truncata* plastome

SSR type	SSR	Size	Start	End	Location
tetra	(AAAT)3	12	345	356	*trnH-GUG/psbA* (IGS)
mono	(A)11	11	386	396	*trnH-GUG/psbA* (IGS)
mono	(A)8	8	2622	2629	*matK* (CDS)
mono	(T)9	9	3279	3287	*matK* (CDS)
mono	(A)10	10	3663	3672	*matK/trnK-UUU* (IGS)
mono	(T)9	9	4019	4027	matK/trnK-UUU (IGS)
di	(TA)5	10	4472	4481	*trnK-UUU/rps16* (IGS)
mono	(A)9	9	4602	4610	*trnK-UUU/rps16* (IGS)
mono	(A)8	8	4673	4680	*trnK-UUU/rps16* (IGS)
mono	(T)13	13	4710	4722	*trnK-UUU/rps16* (IGS)
mono	(A)18	18	5184	5201	*rps16* (intron)
di	(AT)4	8	5502	5509	*rps16* (intron)
tri	(ATT)4	12	6008	6019	*rps16/trnQ-UUG* (IGS)
mono	(A)22	22	6027	6048	*rps16/trnQ-UUG* (IGS)
mono	(T)9	9	6191	6199	*rps16/trnQ-UUG* (IGS)
mono	(A)11	11	6308	6318	*rps16/trnQ-UUG* (IGS)
mono	(A)10	10	6749	6758	*trnQ-UUG/psbK* (IGS)
mono	(T)8	8	7313	7320	*psbK/psbI* (IGS)
mono	(A)9	9	7342	7350	*psbK/psbI* (IGS)
mono	(A)12	12	7527	7538	*psbK/psbI* (IGS)
compound	(TA)3(T<A>)(A)7	15	7993	8007	*trnS-GCU/clpP* (IGS)
hexa	(AGAGAC)4	24	8626	8649	*clpP* (CDS)
mono	(T)10	10	9213	9222	*trnG-UCC* (intron)
mono	(T)8	8	9292	9299	*trnG-UCC* (intron)
mono	(A)8	8	11607	11614	*atpA/atpF* (IGS)
mono	(A)8	8	12225	12232	*atpF* (intron)
mono	(A)10	10	12247	12256	*atpF* (intron)
mono	(T)10	10	12326	12335	*atpF* (intron)
mono	(A)10	10	12422	12431	*atpF* (intron)
mono	(T)10	10	12511	12520	*atpF* (intron)
mono	(A)8	8	12743	12750	*atpF* (intron)
mono	(T)10	10	15631	15640	*rps2/rpoC2* (IGS)
mono	(A)8	8	16037	16044	*rps2/rpoC2* (IGS)
mono	(T)9	9	16203	16211	*rpoC2* (CDS)
mono	(A)8	8	17266	17273	*rpoC2* (CDS)
di	(AT)4	8	18074	18081	*rpoC2* (CDS)
mono	(T)11	11	18108	18118	*rpoC2* (CDS)
mono	(T)10	10	18302	18311	*rpoC2* (CDS)
mono	(T)12	12	18317	18328	*rpoC2* (CDS)
mono	(T)8	8	18618	18625	*rpoC2* (CDS)

di	(AT)5	10	19449	19458	*rpoC2* (CDS)
mono	(T)10	10	25041	25050	*rpoB* (CDS)
mono	(A)9	9	25981	25989	*rpoB/trnC-GCA* (IGS)
mono	(A)10	10	26126	26135	*trnC-GCA/petN* (IGS)
mono	(T)8	8	26618	26625	*trnC-GCA/petN* (IGS)
mono	(T)8	8	27450	27457	*petN/psbM* (IGS)
di	(TG)4	8	27683	27690	*petN/psbM* (IGS)
di	(AT)7	14	28283	28296	*psbM/trnD-GUC* (IGS)
mono	(A)10	10	28718	28727	*psbM/trnD-GUC* (IGS)
mono	(T)8	8	29059	29066	*trnD-GUC/ndhJ* (IGS)
mono	(A)8	8	29790	29797	*trnF-GAA/trnL-UAA* (IGS)
mono	(T)8	8	30197	30204	*trnL-UAA* (intron)
mono	(T)9	9	30235	30243	*trnL-UAA* (intron)
mono	(A)10	10	30245	30254	*trnL-UAA* (intron)
di	(AT)4	8	30643	30650	*trnL-UAA/trnT-UGU* (IGS)
di	(AT)4	8	30652	30659	*trnL-UAA/trnT-UGU* (IGS)
mono	(T)9	9	30822	30830	*trnL-UAA/trnT-UGU* (IGS)
mono	(T)12	12	31011	31022	*trnT-UGU/rps4* (IGS)
mono	(T)9	9	31035	31043	*trnT-UGU/rps4* (IGS)
mono	(A)9	9	31132	31140	*trnT-UGU/rps4* (IGS)
tetra	(TGAT)3	12	32343	32354	*trnS-GGA/ycf3* (IGS)
mono	(A)10	10	32698	32707	*trnS-GGA/ycf3* (IGS)
mono	(T)11	11	32998	33008	*ycf3* (intron)
mono	(A)8	8	33386	33393	*ycf3* (intron)
mono	(A)10	10	34045	34054	*ycf3* (intron)
mono	(T)12	12	35089	35100	*ycf3/psaA* (IGS)
di	(AT)4	8	35241	35248	*ycf3/psaA* (IGS)
di	(AT)5	10	35272	35281	*ycf3/psaA* (IGS)
coumpond	(A)20(<A>G)(AG)3	28	35456	35483	*ycf3/psaA* (IGS)
di	(CT)4	8	41913	41920	*trnS-UGA* (CDS)
mono	(A)9	9	42021	42029	*trnS-UGA/psbC* (IGS)
mono	(A)10	10	42061	42070	*trnS-UGA/psbC* (IGS)
mono	(A)10	10	44915	44924	*psbD/trnT-GGU* (IGS)
di	(AT)7	14	45543	45556	*trnT-GGU/trnE-UUC* (IGS)
mono	(A)8	8	46533	46540	*trnY-GUA/ndhK* (IGS)
mono	(A)13	13	46645	46657	*trnY-GUA/ndhK* (IGS)
mono	(T)8	8	46795	46802	*ndhK* (CDS)
mono	(T)10	10	47358	47367	*ndhK/ndhC* (IGS)
di	(AT)4	8	48169	48176	*ndhC/rbcL* (IGS)
di	(GA)4	8	48845	48852	*rbcL* (CDS)
mono	(A)10	10	49815	49824	*rbcL/atpB* (IGS)
di	(TA)6	12	49926	49937	*rbcL/atpB* (IGS)
mono	(A)10	10	50224	50233	*rbcL/atpB* (IGS)

mono	(T)11	11	50376	50386	*rbcL/atpB* (IGS)
mono	(A)10	10	50698	50707	*atpB* (CDS)
di	(TA)4	8	52690	52697	*atpE/trnM-CAU* (IGS)
mono	(A)10	10	53750	53759	*accD* (CDS)
mono	(A)9	9	55415	55423	*accD* (CDS)
mono	(T)9	9	57626	57634	*accD/psaI* (IGS)
di	(TA)4	8	57858	57865	*accD/psaI* (IGS)
mono	(C)10	10	58436	58445	*ycf4* (CDS)
mono	(T)12	12	58451	58462	*ycf4* (CDS)
mono	(A)11	11	58872	58882	*ycf4/cemA* (IGS)
mono	(A)12	12	58905	58916	*ycf4/cemA* (IGS)
di	(AT)4	8	59938	59945	*petA* (CDS)
mono	(A)8	8	60298	60305	*petA* (CDS)
mono	(A)10	10	61544	61553	*psbF* (CDS)
mono	(T)10	10	62579	62588	*petL/petG* (IGS)
mono	(A)10	10	63033	63042	*trnW-CCA/trnP-UGG* (IGS)
di	(TA)6	12	63339	63350	*trnP-UGG/psaJ* (IGS)
mono	(A)10	10	63426	63435	*trnP-UGG/psaJ* (IGS)
di	(AT)6	12	64545	64556	*rpl33/rps18* (IGS)
mono	(T)8	8	64607	64614	*rpl33/rps18* (IGS)
mono	(A)8	8	65503	65510	*rpl20* (CDS)
mono	(T)8	8	65770	65777	*rpl20* (CDS)
mono	(A)9	9	66009	66017	*rpl20/rps12* (IGS)
mono	(A)8	8	66915	66922	*rps12/psbB* (IGS)
mono	(T)11	11	68635	68645	*psbB/psbT* (IGS)
mono	(A)11	11	69980	69990	*petB* (intron)
mono	(A)11	11	71104	71114	*petB/petD* (IGS)
di	(AG)4	8	71245	71252	*petD* (intron)
mono	(T)8	8	71389	71396	*petD* (intron)
di	(AT)4	8	71541	71548	*petD* (intron)
di	(TA)6	12	71550	71561	*petD* (intron)
di	(TA)4	8	72021	72028	*petD* (CDS)
di	(GA)4	8	73450	73457	*rpoA* (CDS)
mono	(T)13	13	74114	74126	*rps11/rpl36* (IGS)
mono	(T)8	8	75235	75242	*rps8/rpl14* (IGS)
mono	(A)16	16	75253	75268	*rps8/rpl14* (IGS)
mono	(T)8	8	75376	75383	*rpl14* (CDS)
mono	(A)11	11	75777	75787	*rpl14/rpl16* (IGS)
mono	(T)10	10	76969	76978	*rpl16* (intron)
tetra	(TTTC)3	12	77097	77108	*rpl16* (intron)
mono	(T)11	11	77160	77170	*rpl16* (intron)
mono	(A)10	10	77198	77207	*rpl16* (intron)
mono	(T)8	8	77339	77346	*rpl16* (intron)

mono	(T)9	9	78321	78329	*rpl22* (CDS)
tetra	(ATAA)3	12	78465	78476	*rpl22* (CDS)
tetra	(ATAA)4	16	80683	80698	*trnI-CAU/ycf2* (IGS)
di	(GA)4	8	80789	80796	*ycf2* (CDS)
di	(GA)4	8	80801	80808	*ycf2* (CDS)
di	(GA)4	8	81734	81741	*ycf2* (CDS)
mono	(A)9	9	83694	83702	*ycf2* (CDS)
di	(GA)4	8	83715	83722	*ycf2* (CDS)
di	(TA)4	8	87640	87647	*ycf2* (CDS)
mono	(T)8	8	88470	88477	*ycf1* (CDS)
mono	(A)9	9	89434	89442	*ycf1* (CDS)
mono	(T)8	8	89969	89976	*ycf1* (CDS)
mono	(A)10	10	90625	90634	*ycf1* (CDS)
mono	(A)10	10	91186	91195	*ycf1* (CDS)
mono	(A)12	12	91665	91676	*ycf1* (CDS)
di	(AT)4	8	92531	92538	*ycf1* (CDS)
mono	(A)9	9	92555	92563	*ycf1* (CDS)
mono	(A)9	9	93596	93604	*ycf1/rps15* (IGS)
mono	(T)11	11	93991	94001	*rps15/psaC* (IGS)
mono	(A)8	8	94132	94139	*psaC* (CDS)
mono	(T)8	8	95391	95398	*ndhD/ccsA* (IGS)
mono	(A)13	13	95425	95437	*ndhD/ccsA* (IGS)
mono	(A)12	12	96596	96607	*ccsA/trnL-UAA* (IGS)
mono	(A)8	8	97333	97340	*trnL-UAA/rpl32* (IGS)
mono	(A)10	10	97661	97670	*trnL-UAA/rpl32* (IGS)
mono	(A)8	8	97717	97724	*trnL-UAA/rpl32* (IGS)
mono	(T)8	8	97891	97898	*rpl32* (CDS)
di	(AT)6	12	98122	98133	*rpl32/ndhF* (IGS)
mono	(T)8	8	98296	98303	*rpl32/ndhF* (IGS)
mono	(A)8	8	98383	98390	*rpl32/ndhF* (IGS)
mono	(T)8	8	98759	98766	*ndhF* (CDS)
mono	(A)8	8	99688	99695	*trnN-GUU/trnR-ACG* (IGS)
di	(AG)4	8	102214	102221	*rrn23* (CDS)
tetra	(CTAC)3	12	102246	102257	*rrn23* (CDS)
mono	(A)9	9	105576	105584	*trnI-GAU* (intron)
mono	(C)8	8	106550	106557	*rrn16* (CDS)
mono	(A)10	10	108647	108656	*trnV-GAC/rps12* (IGS)
mono	(T)10	10	111139	111148	*ndhB/ycf1* (IGS)
mono	(T)9	9	111173	111181	*ndhB/ycf1* (IGS)

CDS, coding sequences; IGS, intergenic spacers

4. GENERAL CONCLUSIONS

Here, the complete plastomes of *Melocactus glaucescens*, *Lepismium cruciforme*, and *Schlumbergera truncata*, three species of the subfamily Cactoideae, were reported. The detailed analysis of these genomes revealed a series of unique features that bring new insights into the atypical evolution of the plastomes of the family Cactaceae, most likely imposed by the environment for the adaptation and diversification of these species to extreme environmental conditions. Structurally, the plastome of *M. glaucescens* presented unique rearrangements, culminating in the alteration of the gene order of the LSC region and reconfiguration of the IR region, whose expansion resulted in the unusual duplication of 22 genes, such as *accD* and *rbcL*, generally located in the LSC. On the other hand, the analysis of the plastomes of *L. cruciforme* and *S. truncata* revealed a conserved structure among the tribe Rhipsalideae, with tribe-specific rearrangements and some variations in the size and content of the IR region.

Extensive gene loss and pseudogenization occurred in the plastomes analyzed here. Shared losses among the three species include the genes of the *ndh* complex, *rpl23*, and *trnV-UAC*. Additionally, *S. truncata* lost the gene functionality of *ycf4* (pseudogene). In turn, *L. cruciforme* lost the *rpl33* and *trnT-GGU* genes. The *trnT-GGU* is a pseudogene in *L. cruciforme* plastome due to a nucleotide insertion and several mutation events in the sequence. These mutations also occurred in three other species of the tribe Rhipsalideae, indicating a possible process of degeneration of the *trnT-GGU* gene in the tribe and the action of the superwobbling mechanism (trnT-UGU) in *L. cruciforme* plastome to supply the loss of gene functionality. *M. glaucescens* lost two other genes, *trnA-UGC* and *trnV-GAC*. The *trnA-UGC* and *trnV-UAC* gene losses did not change the codon usage frequency of protein-coding genes and, unlike the *trnT-GGU* gene, they cannot be compensated by superwobbling, indicating the importation of tRNAs from the cytosol into the plastid in Cactoideae.

Regarding molecular evolution, the plastid genes of the family Cactaceae showed high gene divergence, with the *accD*, *clpP*, and *ycf1* genes being the most divergent. Furthermore, almost half of the plastid genes had sites under putative positive selection, most of them involved in plastid gene expression. Another characteristic of the family is the polymorphism of RNA editing sites since several gains and losses were identified in the plastomes, revealing the rapid evolution of this molecular mechanism in Cactaceae. Although phylogenetic analyzes based on concatenated genes resulted in well-supported trees capable of resolving close relationships, a curious relationship was found in the tribe Rhipsalideae. The genus *Rhipsalis*

commonly described as monophyletic did not cluster but formed two distinct clades with *L. cruciforme* and *S. truncata*. Nine highly polymorphic regions were identified among the plastomes of the tribe Rhipsalideae, which may be useful to improve the phylogenetic resolutions of the tribe, in addition to the sequencing and availability of new plastomes. Finally, hundreds of plastid molecular markers were mapped here, wich will be useful both for population genetics studies and to assist in the access and conservation of natural resources, mainly for endangered species such as *M. glaucescens*.