TWO CHLOROPLAST GENOMES WITH REDUCED INVERTED REPEAT REGIONS IN Mammillaria series Stylothelae (Cactaceae)

Juan Pablo Ortiz-Brunel, Michael R. McKain, Pilar Zamora-Tavares, Eduardo Ruiz-Sánchez, Daniel Sánchez, Isaac Sandoval-Padilla, Bryan N. MacNeill, and Aaron Rodríguez

1 Departamento de Botánica y Zoología, Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Zapopan, Jalisco, México.
2 Doctorado en Ciencias en Biosistemática, Ecología y Manejo de Recursos Naturales y Agrícolas, Universidad de Guadalajara, Zapopan, Jalisco, México.
3 Department of Biological Sciences, The University of Alabama, Tuscaloosa, USA.
4 Laboratorio Nacional de Identificación y Caracterización Vegetal (LaniVeg), Instituto de Botánica, Universidad de Guadalajara, Zapopan, Jalisco, México.
5 CONAHCYT-Universidad de Guadalajara, Departamento de Botánica y Zoología, Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Zapopan, Jalisco, México.

*Author for correspondence: aaron.rodriguez@academicos.udg.mx

Abstract
Background: The chloroplast genomes of Cactaceae exhibit modification in the inverted repeat regions (IRs), gene inversions, and deletions. Among nine Mammillaria species, three distinct chloroplast structures have been identified, although not all of these correspond to the morphology-based classification of the genus.

Question: Is there a distinct chloroplast genome structure in the species of Mammillaria series Stylothelae?

Studied species: Mammillaria bocasana and M. erythrosperma.

Study site and dates: Mexico from 2019 to 2023.

Methods: Chloroplast DNA was sequenced, and chloroplast genomes were de novo assembled using the Fast-Plast program. Complete plastome sequences were annotated and verified. The sequences were aligned in MAUVE program to detect possible structural changes. A maximum likelihood phylogeny was executed to evaluate the relationships of the studied species.

Results: The plastomes ranged from 107,368 bp in Mammillaria bocasana to 108,069 bp in M. erythrosperma. Both presented a quadripartite structure and contained 108 genes. The IRs were ~1,600 bp long and included the genes rpl2, rpl23 (pseudo), and trnL-CAU. MAUVE identified a ~21 kb inversion in the large single copy containing a block of genes related to photosynthesis. The phylogenetic analysis placed both species in a single clade separated from the other species within Mammillaria subg. Mammillaria.

Conclusions: The studied species of Mammillaria series Stylothelae exhibited a different and synapomorphic chloroplast genome structure. Other Mammillaria chloroplast genome structures have evolved independently in different lineages.

Keywords: isomeric plastomes, ndh genes, plastome, structure.
While the chloroplast genome (plastome) remains highly conserved in most angiosperms (Daniell et al. 2016), there are exceptions characterized by significant structural changes. These variations have been reported in some species-rich groups such as Fabaceae, Passifloraceae, and Malpighiales (Cauz-Santos et al. 2020, Jin et al. 2020, Lee et al. 2021), often observed in parasitic plants (Wolf et al. 1992, Braukmann et al. 2013, Frailey et al. 2018, Su et al. 2021) or species adapted to extreme environments (McCoy et al. 2008, Silva et al. 2016, Wei et al. 2021). In this scenario, the Cactaceae family serves as a prominent example within the Caryophyllales order. Plastomes in Caryophyllales members usually range from 151 to 155 kb and exhibit the typical quadripartite structure divided into a large single copy (LSC), a small single copy (SSC), and two inverted repeats (IRa and IRb) regions (Yao et al. 2019). Nevertheless, numerous structural arrangements have been identified within Cactaceae. These arrangements involve plastome reduction derived from expansion/reduction or loss of the inverted repeat regions (IRs), and gene losses (Sanderson et al. 2015, Majure et al. 2019, Köhler et al. 2020, Morais Da Silva et al. 2021, Köhler et al. 2023, Yu et al. 2023).

Chloroplast genomes within Cactaceae exhibit a wide range in length, spanning from 107 to 162 kb, and their IRs vary from around 350 bp to 37 kb (Solórzano et al. 2019, Köhler et al. 2023, Yu et al. 2023). It is worth noting that Carnegiea gigantea (Engelm.) Britton & Rose (the saguaro) and Lophocereus schoenii (Engelm.) Britton & Rose, lack the IRs (Sanderson et al. 2015, Solórzano et al. 2019), as well as other Opuntioideae such as Quiabentia verticalata (Vaupel) Borg (Köhler et al. 2020). Within the genus Mammillaria Haw., nine chloroplast genomes have been sequenced with lengths ranging from 107 to 116 kb and three distinct structural variations (Solórzano et al. 2019, Hinojosa-Alvarez et al. 2020, Yu et al. 2023).

In Mammillaria, the structure of the chloroplast genome displays variation. Solórzano et al. (2019) assembled chloroplast genomes for seven species representing three out of the eight subgenera proposed by Hunt et al. (2006): Krainzia, Mammillaria, and Phellosperma. Interestingly, species from different subgenera share the same chloroplast genome structure (Solórzano et al. 2019). Consequently, the different chloroplast genome structures seem not to be related to the morphology-based classification. Given the variation of the plastome structure in Mammillaria, other unknown structures may exist in the remaining subgenera or in other groups within the subgenus Mammillaria.

Plastome sequences and information are lacking for the Mammillaria series Stylothelae, which is part of M. subg. Mammillaria and encompasses 19 taxa (Hunt 2016, González-Zamora et al. 2023, Ortiz-Brunel et al. 2023). This series can be distinguished by the presence of axillary bristles, slightly embedded flowers and fruits, generally uncinated central spines, and the production of Luethy’s alkaloid. The combination of these characters differentiates Stylothelae from the other series (Fitz-Maurice & Fitz-Maurice 2006, Hunt et al. 2006). Additionally, all species of M. series Stylothelae share the absence of the rpl16 intron, which is proposed as a synapomorphy of the group (Butterworth & Wallace 2004, Butterworth et al. 2007). This could be an indicator of chloroplast structure variation. Based on this evidence, we anticipate that species of M. subg. Mammillaria series Stylothelae may exhibit a different chloroplast genome structure.

Analyses of chloroplast genome structures and phylogenetic hypothesis based on whole chloroplast sequences have demonstrated their utility in elucidating phylogenetic relationships within Mammillaria and its related genera (Solórzano et al. 2019, Breslin et al. 2021, Chincoya et al. 2023). Our objective was to sequence and assemble the plastome of M. bocasana Poseg. and M. erythrospersperma Boed., which belong to M. subg. Mammillaria series Stylothelae (Hunt et al. 2006). Additionally, we also compared their structures with those previously described for the genus to gain insights into a broad overview of plastome evolution within the genus.

**Material and methods**

**Sampling.** Two living individuals of *Mammillaria bocasana* (J. P. Ortiz-Brunel 922, IBUG) and two of *M. erythrospersperma* (J. P. Ortiz-Brunel 410, IBUG) (herbarium acronym according to Thiers 2023) were collected nearby their type locality (**Figure 1**). The morphology of the specimens was compared with the morphological descriptions to corroborate their identity (Bravo-Hollis & Sánchez-Mejorada 1991, Reppenhagen 1991, Fitz-Maurice & Fitz-Maurice 2006). We preserved the plants in a greenhouse at the University of Guadalajara until tissue collection.
Figure 1. Studied species in their habitat. A) *Mammillaria bocasana*, B) *M. erythrosperma*.

**DNA extraction and sequencing.** Using freshly collected tissue from the tubercles of only one individual per species, chloroplasts were isolated and cpDNA extracted as described in Shi *et al.* (2012), with slight modifications. The cpDNA quantity and quality were evaluated with a Qubit 3.0 and NanoDrop 2000 (Thermo Fisher Scientific, Waltham, Massachusetts), respectively. We prepared the libraries using the Ion Plus Fragment Library Kit (Thermo Fisher Scientific) and selected DNA fragments of approximately 250 bp using an E-Gel Sizeselect Agarose Gel (Thermo Fisher Scientific). Sequencing of single-end reads of 250 bp was performed on a Personal Genome Machine (Thermo Fisher Scientific) in the Laboratorio Nacional de Identificación y Caracterización Vegetal (LaniVeg) at the University of Guadalajara.

**Chloroplast genome assembly.** Raw reads quality was assessed in the FastQC v. 0.11.7 program (Andrews 2010). Then, we used the Trimmomatic tool (Bolger *et al.* 2014) to discard reads with low quality (PHRED quality score < 15) using the leading, trailing and avgcall tools. The resulting reads were *de novo* assembled into contigs following the Fast-Plast pipeline (McKain 2017) with slight modifications. Reads were mapped against the genome of *Mammillaria pectinifera* F.A.C. Weber as a reference using Bowtie2 v. 2.5.1 (Langmead & Salzberg 2012) under the very-sensitive-local parameter set to filter for chloroplast-like sequences. These filtered reads were *de novo* assembled using SPAdes 3.15.0 (Bankevich *et al.* 2012) using k-mer sizes of 21, 35, 57, and 89 with the “only-assembler” option. The assembled contigs were merged and extended using the “afin” script available from the Fast-Plast program using default parameters with 50 extension loops and the chloroplast-mapped reads. Once a single contig was reached, the sequence_based_ir.pl script packaged with Fast-Plast was used to find putative IR regions. We used Sequencher v. 4.1.4 (Gene Codes) to verify the IRs. A final coverage analysis to verify the accuracy of our assemblies was conducted using scripts from Fast-Plast and supported by Jellyfish 2 (Marçais & Kingsford 2011). Gene annotation was performed in the GeSeq platform (Tillich *et al.* 2017) and every annotation was manually confirmed. The chloroplast genome circular representation was produced with OGDraw v. 1.3.1 (Greiner *et al.* 2019).

**Chloroplast genome structure comparison.** We selected one sequence for each known chloroplast genome structure in *Mammillaria* (Solórzano *et al.* 2019) for comparison to our new assemblies. *Mammillaria pectinifera* represented the structure 1 (S1), *M. crucigera* Mart. characterized the structure 2 (S2), and *M. zephyranthoides* Scheidw. corresponded to the structure 3 (S3) (Solórzano *et al.* 2019). These representative chloroplast genomes were converted into linear genomes and then aligned to our *M. bocasana* and *M. erythrosperma* plastomes. All genomes were aligned using the plastome of *M. bocasana* as the reference in MAUVE v. 2.4.0 (Darling *et al.* 2004) with the Progressive Mauve Tool using the default parameters. GenBank accession numbers are indicated in Table 1.
Plastomes of Mammillaria series Stylothelae

Table 1. List of the species used in the analyses.

<table>
<thead>
<tr>
<th>Species</th>
<th>GenBank accession number</th>
<th>Infrageneric classification sensu Hunt et al. (2006)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carnegiea gigantea</td>
<td>NC_027618.1</td>
<td></td>
</tr>
<tr>
<td>Mammillaria albiflora</td>
<td>MN517610.1</td>
<td>Subgenus Krainzia, series Herrerae-Pectinifera</td>
</tr>
<tr>
<td>M. bocasana</td>
<td>OR863748</td>
<td>Subgenus Mammillaria, series Stylothelae</td>
</tr>
<tr>
<td>M. crucigera</td>
<td>MN517613.1</td>
<td>Subgenus Mammillaria, series Supertextae</td>
</tr>
<tr>
<td>M. erythrosperma</td>
<td>OR863749</td>
<td>Subgenus Mammillaria, series Stylothelae</td>
</tr>
<tr>
<td>M. huitzilopochtli</td>
<td>MN517612.1</td>
<td>Subgenus Mammillaria, series Supertextae</td>
</tr>
<tr>
<td>M. pectinifera</td>
<td>MN519716.1</td>
<td>Subgenus Krainzia, series Herrerae-Pectinifera</td>
</tr>
<tr>
<td>M. solisiodex</td>
<td>MN518341.1</td>
<td>Subgenus Krainzia, series Herrerae-Pectinifera</td>
</tr>
<tr>
<td>M. supertexta</td>
<td>MN508963.1</td>
<td>Subgenus Mammillaria, series Supertextae</td>
</tr>
<tr>
<td>M. zephyranthoides</td>
<td>MN517611.1</td>
<td>Subgenus Phellosperma</td>
</tr>
</tbody>
</table>

Phylogenetic analysis. We downloaded seven chloroplast genome sequences of Mammillaria from GenBank published by Solórzano et al. (2019). These sequences represented the three different chloroplast genome structures within the genus. The plastome of Carnegiea gigantea (Sanderson et al. 2015) was included as an outgroup. Detailed GenBank accession numbers for all sequences used in the analysis are listed in Table 1. We aligned the ten complete plastome sequences using MAFFT v. 7.52 (Katoh et al. 2019) with default parameters. A maximum likelihood (ML) search was executed in MEGA v. 11 (Tamura et al. 2021), employing the GTR + G + I model. Supporting branch values were obtained through 1,000 bootstrap replications. To discard a possible influence of the chloroplast structures in the results, we performed two additional phylogenetic analyses with the same parameters but different datasets. The first analysis used a matrix in which the 21 kb inversion block was reverted for M. bocasana and M. erythrosperma. In the second one, we filtered for 54 CDS regions shared among all taxa (Table S1). All phylogenetic analyses were performed with the same parameters.

Results

Chloroplast genome structure comparison. The chloroplast genomes of Mammillaria bocasana and M. erythrosperma exhibited a quadripartite structure, including an LSC, an SSC, and two small IRs. The chloroplast genome of M. bocasana was 107,368 bp long with an LSC of 75,290 bp, an SSC of 28,896 bp, and two IRs of 1,591 bp. Mammillaria erythrosperma chloroplast genome was 108,069 bp long, from which 76,393 bp conformed the LSC, the SSC was of 28,402 bp, and two IRs of 1,637 bp (Figure 2). The Guanine-Cytosine (GC) content was 37 % in M. bocasana and 36.6 % in M. erythrosperma.

Both plastomes shared identical gene content and order with 108 protein coding genes, tRNAs, and rRNAs (Figure 2, Table 2). Ten of those genes were pseudogenized in Mammillaria bocasana and 11 in M. erythrosperma. The latter exhibited pseudogenization of the rps16 gene, resulting from a partial loss of the first exon. The IRs of both species contained rpl2, pseudogene rpl23, and trnL-CAU with IR lengths of 1,591 bp in M. bocasana and 1,637 in M. erythrosperma. In both plastomes, the IRs was delimited by the rps19 and the ycf2 genes and the IRb by the ndhB (pseudogene) and trnH-GUG genes. Both plastomes lacked functional NADH dehydrogenase-like (NDH) complex (ndh) genes, though we detected only pseudogenes of ndhB, ndhD, and ndhF.

When compared to other Mammillaria plastome structures, the progressive MAUVE alignment identified a new chloroplast genome structure (S4) characterized by a unique inversion within the LSC region in both species of Mammillaria series Stylothelae (Figure 3). This inversion was 21,248 bp long in M. bocasana and 21,469 bp in M. erythrosperma and encompassed 21 coding regions, extending from perG to atpE genes (Figure 2). The majority of these inverted genes are associated with photosynthetic functions (Table 2).
the S1, S2, and S4, had all the other gene blocks in the same order and arrangement. However, the alignment showed remarkable reductions and inversions in the S3, compared with all the other structures.

Phylogenetic analysis. The chloroplast genome recovered robust phylogenetic relationships, with bootstrap support values (BS) exceeding 95% for all tree branches (Figure 4). The analyses with the complete plastomes, the complete plastomes with the 21 kb reverted block, and the shared CDS, recovered mostly the same topology. These differed in the placement of *Mammillaria albiflora* and *M. zephyranthoides*, and the relationships within *M. series Supertextae* (Figures 4, S1, S2). Here, we used the phylogeny inferred with the complete plastomes to describe and discuss our results (Figure 4). The resulting phylogeny placed our two plastome sequences in a monophyletic group within all other *Mammillaria* species (BS = 100%). Interestingly, species with the same chloroplast structure were dispersed across different clades (Figure 4), except for *Mammillaria bocasana* and *M. erythrosperma*, which had a sister spe-
Plastomes of *Mammillaria* series *Stylothelae*

**Table 2.** Gene content of the chloroplast genome of *M. bocasana* and *M. erythrosperma*. Ψ indicates a pseudogene in both species; *Complete gene in *M. bocasana*, pseudogene in *M. erythrosperma*; **One copy functional and the other a pseudogene in both species.

<table>
<thead>
<tr>
<th>Genes group</th>
<th>Name</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rubisco Large Subunit</td>
<td><em>rbc</em> L</td>
<td>1</td>
</tr>
<tr>
<td>Acetyl-CoA carboxylase beta subunit</td>
<td><em>acc</em> DΨ</td>
<td>1</td>
</tr>
<tr>
<td>Other</td>
<td><em>ycf</em> 1, <em>ycf</em> 2, <em>ycf</em> 3, <em>ycf</em> 4Ψ, <em>ycf</em> 68Ψ</td>
<td>5</td>
</tr>
<tr>
<td>Cytochrome C synthesis</td>
<td><em>ccs</em> A</td>
<td>1</td>
</tr>
<tr>
<td>Chloroplast envelope membrane protein</td>
<td><em>cem</em> A</td>
<td>1</td>
</tr>
<tr>
<td>Maturase</td>
<td><em>clp</em> P (2**), <em>mat</em> K</td>
<td>3</td>
</tr>
<tr>
<td>Translation initiation factor</td>
<td><em>inf</em> A</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>108</td>
<td></td>
</tr>
</tbody>
</table>

Species relationship with strong support (BS = 100 %). Further, this clade was sister to a group that included species from *M. subg. Mammillaria* series *Supertextae* (*M. supertexta*, *M. crucigera* and *M. huitzilopochtli*) and *M. subg. Krainzia* (*M. pectinifera* and *M. solisioides*). The phylogeny did not support the monophyly of *Mammillaria* subg. *Mammillaria* nor *M. subg. *Krainzia* (Figure 4, Table 1). While *M. subg. Mammillaria* appeared to be paraphyletic, *M. series Supertextae* (BS = 100 %) and *M. series Stylothelae* (BS = 100 %) formed two monophyletic groups.

**Discussion**

Chloroplast genome structure comparison. The chloroplast genome sequences of *Mammillaria bocasana* and *M. erythrosperma* (*Mammillara series Stylothelae*) exhibit an undescribed plastome structure (S4) (Figure 2). This new
structure shares similar gene content and arrangement with the S1, represented by *M. pectinifera* (Solórzano et al. 2019). It differs, however, by a ~ 21 kb inversion within the LSC that had not been reported in other *Mammillaria* species (Figure 2). The inversion contains mainly protein coding genes related to photosynthesis. On the other hand, the IRs of *M. bocasana* and *M. erythrosperma* (S4) were similar in length and gene content to those of *M. albiflora* and *M. pectinifera* (S1) (Solórzano et al. 2019). The distinctive features of this new structure are the inversion within the LSC and the presence of the *rpl2* gene in the IRs instead of a partial sequence of the *ycf2* gene (Table 3). Our results support the idea that *rnl*-CAU is involved in reconfiguring of IRs in *Mammillaria* (Solórzano et al. 2019). Boundary shifts seem to result from gene rearrangements, but further research is needed.

![Figure 3. MAUVE alignment. A) Mammillaria bocasana, B) M. erythrosperma, C) M. pectinifera, D) M. crucigera, E) M. zephyranthoides. The codes S1, S2, and S3 represent the chloroplast genome structures found by Solórzano et al. (2019) and S4 indicates the new structure found in this work.](image-url)

Compared with other Cactaceae, the *Mammillaria* plastomes display high variation associated with gene translocations and inversions (Solórzano et al. 2019, Yu et al. 2023). Plastome variation is common in Cactaceae and typically involves changes within the IRs. This variation is given by gene rearrangements, expansion and contraction of coding and non-coding regions, and changes in the boundaries of the IRs (Yu et al. 2023). Palmer (1986) and Walker et al. (2015) suggested that the IR could induce isomers, particularly within the SSC. This phenomenon is common in Cupressaceae (Guo et al. 2014, Qu et al. 2017) and was recently discovered within the LSC in Cactaceae (Yu et al. 2023). In cacti, the frequency of isomers is estimated around 1% (Yu et al. 2023). Another source of plastome variation is the high presence of short repeat sequences that promote chloroplast structure differentiation mediated by intramolecular recombination (Ruhlman et al. 2017, Qu et al. 2017). When comparing the plastomes in the MAUVE alignment, we found short sequence repeats near the boundaries of the 21 kb inverted block of genes within the LSC of *M. bocasana* and *M. erythrosperma*. High short repetitive regions are common within the LSC of *Mammillaria* plastomes (Chincoya et al. 2020). It is possible that repetitive sequences in *Mammillaria* may serve as recombination points and cause the rearrangements. As stated and demonstrated by Yu et al. (2023) more studies are needed to confirm this or to discard the presence of plastome isomers. Given this, it is possible that some *Mammillaria* plastome structures are merely isomers, but further evaluation is needed.

In Cactaceae, the length of chloroplast genomes exhibits high variability due to gene losses, gene duplications, pseudogenization, and expansions/contractions of the IRs (Solórzano et al. 2019, Köhler et al. 2023, Yu et al. 2023). *Mammillaria* is known for having some of the shortest plastomes within the family, and our findings were consistent with this pattern. In the plastome of *M. bocasana* and *M. erythrosperma*, we observed the pseudogenization of the following genes: *accD*, *ndhB*, *ndhD*, *ndhF*, *rpl23* (both copies), *rpl33*, *rps16*, *ycf4*, *ycf68*, and *clpP* (one copy) (Table 2). All these pseudogenes have been previously found in other species of the genus (Solórzano et al. 2019) and are common within Cactaceae (Yu et al. 2023). The plastomes of *M. bocasana* (S4) and *M. zephyranthoides* (S3) are
Plastomes of *Mammillaria* series *Stylothelae*

Table 3. Length and gene content of the inverted repeat regions of *Mammillaria*. The symbol Ψ indicates a pseudogene. The S1 to S3 structures are those reported by Solórzano et al. (2019). S4 denotes the structure found in this work.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Species</th>
<th>IR length (bp)</th>
<th>IR genes content</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td><em>Mammillaria albiflora, M. pectinifera</em></td>
<td>1,348; 1,544</td>
<td>rpl23, trnl-CAU, ycf2</td>
</tr>
<tr>
<td>S3</td>
<td><em>Mammillaria zephyranthoides</em></td>
<td>28,252</td>
<td>psbA, trnH-GUG, trnl-CAU, ycf2, ndhB, rps7, rps12, trnV-GAC, rrn16, trnI-GAU, rrn23, rrn4.5, rrn5, trnR-ACG, trnN-GUU, ndhF, rpl32</td>
</tr>
<tr>
<td>S4</td>
<td><em>Mammillaria bocasana, M. erythrosperma</em></td>
<td>1,591; 1,637</td>
<td>rpl2, rpl23Ψ, trnl-CAU</td>
</tr>
</tbody>
</table>

the shortest among *Mammillaria* (~107 kb), but they differ from each other in their gene content (108 and 130, respectively) and arrangement, suggesting a different evolutionary pathway. According to Chincoya et al. (2023), the divergence of most *Mammillaria* clades occurred ~4.5 Ma, which implies a recent diversification of chloroplast structures within the genus. It is necessary to assemble more *Mammillaria* plastomes to trace an accurate evolutionary history.

The evolutionary implications of losing *ndh* genes are not fully understood. These genes play a role in the cyclic electron flow of ATP production (Martín et al. 2009, Strand et al. 2019). The partial or complete loss of the *ndh* genes suite is common in gymnosperms (Braukmann et al. 2009, Martín & Sabater 2010) and frequently occurs in angiosperms (Blazier et al. 2011, Sun et al. 2017, Sun et al. 2018, Könyves et al. 2021, Mower et al. 2021, Cao et al. 2022). In Cactaceae, the complete loss or pseudogenization of multiple *ndh* genes is common (Sanderson et al. 2015, Solórzano et al. 2019, Morais da Silva et al. 2021, Köhler et al. 2023, Yu et al. 2023). In general, it is not yet clear whether these genes have been transferred to the mitochondrial or nuclear genomes or if there are alternative metabolic pathways that compensate for their absence (Lin et al. 2015, Ruhlman et al. 2015, Sanderson et al. 2015, Ranade et al. 2016, Strand et al. 2019). These genes appear to be dispensable under favorable conditions but become crucial when plants are exposed to abiotic stress conditions (Ruhlman et al. 2015, Lin et al. 2017, Sabater 2021). All *Mammillaria* species lacking the *ndh* genes inhabit arid or semiarid regions and it remains unknown how they grow in harsh environments and compensate for the absence of some or all of these genes.

The *de novo* assembled plastomes of *Mammillaria bocasana* and *M. erythrosperma* lacked the rpl16 intron. However, the gene seems to be completely functional because only the main intron is excised and a complete gene remains (Butterworth et al. 2007). The same case has been rarely documented in Amaryllidaceae, Geraniaceae, Goodeniaceae, Papaveraceae, and Plumbaginaceae, but it is infrequent even within them (Campagna & Downie 1998, Zhang et al. 2020, Kim et al. 2023). Consequently, this feature is considered a robust signal of common ancestry (Campagna & Downie 1998). Up to this point, all the other chloroplast genomes known for *Mammillaria* have the rpl16 intron. Therefore, it is highly plausible that the absence of the rpl16 intron could be a synapomorphy for *M.* series *Stylothelae*, as suggested by Butterworth et al. (2007). To evaluate this, additional taxonomic sampling is necessary, including recently described species within the series (González-Zamora et al. 2022, 2023, Ortiz-Brunel et al. 2023).

**Phylogenetic analysis.** The ML tree confirmed the inclusion of our newly sequenced chloroplast genomes within *Mammillaria* (Figure 4). *Mammillaria* subg. *Mammillaria* is paraphyletic, partly due to the inclusion of *M. pectinifera* and *M. solisoides*, which belong to *M.* subg. *Krainzia*. Similar results have been reported in recent, more comprehensive studies, indicating the need for further research (Chincoya et al. 2023). Our results agreed with the monophyly of series *Supertextae*, which was also identified by Cervantes et al. (2021). In all three phylogenetic analyses performed with different datasets, *Mammillaria bocasana* and *M. erythrosperma* grouped as sister to the
Ortiz-Brunel et al. / Botanical Sciences

clade containing the subgenera *Mammillaria* and *Krainzia*. It is possible that *M*. series *Stylothelae* is monophyletic, as well as other series within *M*. subg. *Mammillaria*. Taxon sampling was limited in the present and previous works, and thereby only limited conclusions can be drawn until denser sampling can be done.

Our phylogenetic analyses support that using complete chloroplast genome sequences or only the shared CDS can produce well-supported hypotheses. An independent study based solely on chloroplast protein-coding genes yielded a similar topology (Solórzano et al. 2019). In this study, we performed a phylogenetic analysis with the original full plastomes aligned, other with the ~ 21 kb inversion of *M. bocasana* and *M. erythrosperma* reverted, and another only with 54 shared CDS for all taxa. The full original plastomes dataset produced a better resolved phylogeny, but the hypothesis generated with the inverted block reverted retrieved the same topology and support (Figures 4, S1). However, using the CDS dataset, the placement of *M. albiflora* and *M. zephyranthoides* was different (Figures 4, S2). Any of the three approaches is useful in establishing insights into the evolutionary history of the *Mammillaria* chloroplast genomes. However, it is necessary to increase the taxon sampling to test if some structures have a unique origin (Figure 4). Different chloroplast genome structures have been identified in some species-rich groups with significant morphological variation (Cauz-Santos et al. 2020, Köhler et al. 2020, Lee et al. 2021). With the discovery of a fourth structure characterizing until now the *Mammillaria* series *Stylothelae*, it becomes evident that more extensive taxon sampling across *Mammillaria* is required. Other chloroplast structures might exist within the genus.

**Figure 4.** Phylogenetic analysis based on Maximum Likelihood. The numbers above the branches represent the Bootstrap support values. Codes for the structures (S) are the same as in Figure 3.

**Supplementary material**

Supplemental data for this article can be accessed here: https://doi.org/10.17129/botsci.3446

**Acknowledgements**

The authors thank Jessica Pérez-Alquicira for her help and feedback. Also, the authors thank two anonymous reviewers and the associated editor for improving the quality of this manuscript.
Literature cited


Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S,  
new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational  
Biology 19: 455-477. DOI: https://doi.org/10.1089/cmb.2012.0021

Blazier JC, Guisinger MM, Jansen RK. 2011. Recent loss of plastid-encoded ndh genes within Erodium  
(Geraniaceae). Plant Molecular Biology 76: 263-272. DOI: https://doi.org/10.1007/s11103-011-9753-5

Blazier JC, Guisinger MM, Jansen RK. 2011. Recent loss of plastid-encoded ndh genes within  
Erodium (Geraniaceae). Plant Molecular Biology 76: 263-272. DOI: https://doi.org/10.1007/s11103-011-9753-5


Breslin PB, Wojciechowski MF, Majure LC. 2021. Molecular phylogeny of the Mammillloid clade (Cactaceae)  
resolves the monophyly of Mammillaria. TAXON 70: 308-323. DOI: https://doi.org/10.1002/tax.12451

rpl16 intron in Mammillaria series Stylothelae (Cactaceae) delineates members of the M. crinita group.  

Butterworth CA, Wallace RS. 2004. Phylogenetic studies of Mammillaria (Cactaceae): insights from chloro-  
plast sequence variation and hypothesis testing using the parametric Bootstrap. American Journal of Botany 91: 1086-1098. DOI: https://doi.org/10.3732/ajb.91.7.1086

Braukmann TWA, Kuzmina M, Stefanović S. 2009. Loss of all plastid ndh genes in Gnetales and conifers: ex-  
tent and evolutionary significance for the seed plant phylogeny. Current Genetics 55: 323-337. DOI: https://  
doi.org/10.1007/s00294-009-0249-7

Braukmann T, Kuzmina M, Stefanović S. 2013. Plastid genome evolution across the genus Cuscuta (Convolvulac-  
caeae): two clades within subgenus Grammica exhibit extensive gene loss. Journal of Experimental Botany 64: 977-989. DOI: https://doi.org/10.1038/jxb.ers391

Campagna ML, Downie SR. 1998. The intron in chloroplast gene rpl16 is missing from the flowering plant families  

Cao DL, Zhang XJ, Qu XJ, Fan SJ. 2022. Plastid phylogenomics sheds light on divergence time and ecological  
doi.org/10.3389/fpls.2022.1046253

ia series Supertextae (Cactaceae). PhytoKeys 177: 25-42. DOI: https://doi.org/10.3897/phytoscientia.177.62915

Chincoya DA, Arias S, Vaca-Paniagua F, Dávila P, Solórzano S. 2023. Phylogenomics and biogeography of the  
DOI: https://doi.org/10.3390/biology12040512

Chincoya DA, Sanchez-Flores A, Estrada K, Díaz-Velásquez CE, González-Rodríguez A, Vaca-Paniagua F,  
Dávila P, Arias S, Solórzano S. 2020. Identification of high molecular variation loci in complete chloro-  
plast genomes of Mammillaria (Cactaceae, Caryophyllales). Genes 11: 830. DOI: https://doi.org/10.3390/  
genese11070830


González-Zamora P, Aquino D, Rodríguez A, Sánchez D. 2023. *Mammillaria monochrysacantha* (Cactaceae), a new endemic species from Guanajuato, Mexico. *Phytotaxa* 618: 243-253. DOI: [https://doi.org/10.11646/phytotaxa.618.3.2](https://doi.org/10.11646/phytotaxa.618.3.2)


Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357-359. DOI: [https://doi.org/10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923)


Palmer JD. 1986. Chloroplast DNA exists in two orientations. *Nature* 301: 92-93. DOI: [https://doi.org/10.1038/301092a0](https://doi.org/10.1038/301092a0)


Ruhlman TA, Chang WJ, Chen JJW, Huang Y-T, Chan M-T, Zhang J, Liao DC, Blazier JC, Jin X, Shih MC,


Strand DD, D’Andrea L, Bock R. 2019. The plastid NAD(P)H dehydrogenase-like complex: structure, function and evolutionary dynamics. *Biochemical Journal* 476: 2743-2756. DOI: [https://doi.org/10.1042/BCJ20190365](https://doi.org/10.1042/BCJ20190365)


Plastomes of *Mammillaria* series *Stylothelae*

Wolfe KH, Morden CW, Palmer JD. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proceedings of the National Academy of Sciences* **89**: 10648-10652. DOI: [https://doi.org/10.1073/pnas.89.22.10648](https://doi.org/10.1073/pnas.89.22.10648)


**Associate editor:** Monserrat Vazquez Sánchez

**Author contributions:** JPOB, PZT, ERS, DSC, and AR designed the study. JPOB, PZT and ISP isolated and sequenced DNA. JPOB, MRM, BNM and ISP assembled and annotated plastomes. JPOB analyzed data and prepared figures and tables. JPOB and AR wrote the first draft manuscript. All authors reviewed the final version of the manuscript.

**Supporting agencies:** Consejo Nacional de Humanidades Ciencias y Tecnologías (CONAHCyT), doctoral scholarship 812934 provided to Juan Pablo Ortiz Brunel.

**Conflict of interest:** The authors declare that there is no conflict of interest, financial or personal, in the information, presentation of data, and results of this article.