More than a spiny morphology: plastome variation in the prickly pear cacti (Opuntieae)

Matias Köhler[1,2,*], Marcelo Reginato[2], Jian-Jun Jin[3] and Lucas C. Majure[4,*]

[1]Departamento de Biologia, Centro de Ciências Humanas e Biológicas, Universidade Federal de São Carlos, Sorocaba, SP, Brazil, [2]Programa de Pós-Graduação em Botânica, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil, [3]Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, NY, USA and [4]University of Florida Herbarium (FLAS), Florida Museum of Natural History, Gainesville, FL, USA
[*]For correspondence. E-mail matias.k@ufrgs.br or lmajure@floridamuseum.ufl.edu

- **Background** Plastid genomes (plastomes) have long been recognized as highly conserved in their overall structure, size, gene arrangement and content among land plants. However, recent studies have shown that some lineages present unusual variations in some of these features. Members of the cactus family are one of these lineages, with distinct plastome structures reported across disparate lineages, including gene losses, inversions, boundary movements or loss of the canonical inverted repeat (IR) region. However, only a small fraction of cactus diversity has been analysed so far.
- **Methods** Here, we investigated plastome features of the tribe Opuntieae, the remarkable prickly pear cacti, which represent one of the most diverse and important lineages of Cactaceae. We assembled *de novo* the plastome of 43 species, representing a comprehensive sampling of the tribe, including all seven genera, and analysed their evolution in a phylogenetic comparative framework. Phylogenomic analyses with different datasets (full plastome sequences and genes only) were performed, followed by congruence analyses to assess signals underlying contentious nodes.
- **Key Results** Plastomes varied considerably in length, from 121 to 162 kbp, with striking differences in the content and size of the IR region (contraction and expansion events), including a lack of the canonical IR in some lineages and the pseudogenization or loss of some genes. Overall, nine different types of plastomes were reported, deviating in the presence of the IR region or the genes contained in the IR. Overall, plastome sequences resolved phylogenetic relationships within major clades of Opuntieae with high bootstrap values but presented some contentious nodes depending on the dataset analysed (e.g. whole plastome vs. genes only). Congruence analyses revealed that most plastidial regions lack phylogenetic resolution, while few markers are supporting the most likely topology. Likewise, alternative topologies are driven by a handful of plastome markers, suggesting recalcitrant nodes in the phylogeny.
- **Conclusions** Our study reveals a dynamic nature of plastome evolution across closely related lineages, shedding light on peculiar features of plastomes. Variation of plastome types across Opuntieae is remarkable in size, structure and content and can be important for the recognition of species in some major clades. Unravelling connections between the causes of plastome variation and the consequences for species biology, physiology, ecology, diversification and adaptation is a promising and ambitious endeavour in cactus research. Although plastome data resolved major phylogenetic relationships, the generation of nuclear genomic data is necessary to confront these hypotheses and assess the recalcitrant nodes further.

Key words: Cactaceae, evolution, *Opuntia*, plastid genome, atypical plastomes, phylogenomics.

INTRODUCTION

Plastids are fundamental components of plants, acting as pluripotent organelles capable of interconversion between different types, such as chromoplasts, amyloplasts and chloroplasts, for distinct functions (e.g. storage, growth, photosynthesis) (Sadali *et al.*, 2019). They represent an ongoing evolutionary trajectory of endosymbiosis from a free-living prokaryote to an organelle of a eukaryotic cell, retaining the bulk of their prokaryotic biochemistry but shrunken by orders of magnitude from the genome size that their ancestors possessed (Timmis *et al.*, 2004). Harbouring one of the three genomes in plants, plastids are by far the most utilized for the investigation of the evolutionary history of plants, in addition to physiological and adaptative features (Daniell *et al.*, 2016; Gitzendanner *et al.* 2018; Ruhlman & Jansen, 2021).

Within land plants, plastid genomes (plastomes) are traditionally reported to be highly conserved in structure, content, arrangement and size (Raubeson and Jansen, 2005; Wicke *et al.*, 2011). Taking angiosperms as a reference, plastomes are represented as circular quadripartite monomers, highly gene dense (~80 protein-coding genes), containing two regions of single copy (SC) distinguished by their length [a large SC (LSC, ~80 kb) and a small SC (SSC, ~20 kb)] separated by a large inverted repeat region (IR, ~25 kb) (Ruhlman and Jansen, 2021).

However, advances facilitating the generation and assembly of high-throughput molecular data of several non-model groups have increasingly reported a salient fraction of variation in plastomes (e.g. Ruhlman and Jansen, 2018; Sinn *et al.*, 2018; Cauz-Santos *et al.*, 2020). In this way, numerous cases of structural rearrangements, including inversions and translocations (e.g. Lin *et al.*, 2015; Li *et al.*, 2016; Rabah *et al.*, 2019; Cauz-Santos *et al.*, 2020; Charboneau *et al.*, 2021), pseudogenization or gene losses (e.g. Kim and Chase, 2017; Xu and Wang, 2021), in addition to the lack of the IR region (e.g. Jin *et al.*, 2020*a*; Lee *et al.*, 2021), have been documented across disparate lineages, suggesting a more dynamic nature of plastome evolution, which had previously been underappreciated.

Members of the cactus family (Cactaceae, ~1800 spp.; Korotkova *et al.*, 2021) are broadly known by their peculiar features, such as succulence, morphological diversity, spines and exuberant flowers, making them one of the most charismatic groups of plants known worldwide (Anderson, 2001). Nonetheless, beyond their morphological, physiological and ecological aspects, molecular components can also reveal intriguing traits of cacti. The study of entire cactus plastomes was initiated recently (Sanderson *et al.*, 2015), and it has since been revealed that plastomes across cacti have undergone significant changes in gene content, order and structure in comparison to canonical angiosperm references (Majure *et al.*, 2019; Solórzano *et al.*, 2019; Köhler *et al.*, 2020; Oulo *et al.*, 2020; Almeida *et al.*, 2021; Amaral *et al.*, 2021; da Silva *et al.*, 2021; Dalla Costa *et al.*, 2022; Qin *et al.*, 2022; Yu *et al.*, 2023). Cactaceae seem to have the smallest plastome for an obligate photosynthetic angiosperm (~104–113 kb; Sanderson *et al.*, 2015; Solórzano *et al.*, 2019; Amaral *et al.*, 2021). Furthermore, independent losses of the inverted repeat and the NADH dehydrogenase-like complex (*ndh*) gene suite have been reported in unrelated lineages [e.g. the saguaro cactus, *Carnegiea gigantea* (Cactoideae: Echinocereeae), the cardo-ananá, *Cereus fernambucensis* (Cactoideae: Cereeae) and the Chacoan-leafy cactus, *Quiabentia verticilata* (Opuntioideae: Cylindropuntieae); Sanderson *et al.*, 2015; Köhler *et al.*, 2020; Amaral *et al.*, 2021]. Additionally, nearly all studied species have shown distinct features involving expansion or contractions of the IR region, rearrangements and gene losses or pseudogenization (Solórzano *et al.*, 2019; Oulo *et al.*, 2020; Almeida *et al.*, 2021; Amaral *et al.*, 2021; Silva *et al.*, 2021; Dalla Costa *et al.*, 2022; Qin *et al.*, 2022; Yu *et al.*, 2023). Nonetheless, only a small fraction of cactus species diversity (<5 %) has had their plastome analysed in a comparative framework.

In this work, we provide a deep analysis of plastome characteristics of the tribe Opuntieae, a lineage with one of the most species-rich genera in Cactaceae (Korotkova *et al.*, 2021), the emblematic prickly pear cacti (*Opuntia* spp.). The group represents a remarkable radiation of cacti broadly distributed across the major arid and semi-arid regions of the Americas (Majure *et al.*, 2012; Majure and Puente, 2014), which still lack phylogenomic information to help elucidate their diversification history. We have assembled the plastome *de novo*, using a comprehensive taxon sampling of major groups of the tribe, and investigated the evolution of their characteristics in a phylogenetic framework. Besides describing major features and discussing the plastome evolution in the tribe, we also investigated conflicting phylogenetic signals along different datasets (full

plastome sequence vs. plastidial genes only) and performed analyses of incongruence to assess the aspects underlying the alternative topologies.

MATERIALS AND METHODS

Taxon sampling, DNA extraction and sequencing

We sampled all seven genera currently recognized and accepted within the tribe Opuntieae (Köhler *et al.*, 2020; Korotkova *et al.*, 2021), covering a comprehensive diversity of each genus (ranging from 15 to 100 % representation), totalling 43 accessions (Supplementary Data Table S1). Samples were from field-collected materials or individuals grown at the Desert Botanical Garden (Phoenix, AZ, USA). DNA was extracted from epidermal tissue dried in silica gel using a modified CTAB method (Doyle and Doyle, 1987) followed by chloroform/isoamyl alcohol precipitation and silica column-based purification steps (for details, see Majure *et al.*, 2019). The quality of DNA was tested using a 1 % agarose gel, and whole genomic DNAs were quantified using the Qubit dsDNA BR Assay Kit and Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). Samples with high molecular weight DNA (>15 kb), showing no degradation, were considered suitable and sent to Rapid Genomics LLC (Gainesville, FL, USA) for library preparation with insert sizes of ~340 bp and sequencing using a genome skimming approach (Straub *et al.*, 2012) on the Illumina HiSeq X platform with 150 bp paired-end reads (for full details, see Supplementary Data Table S1).

De novo assemblies and annotation

Raw reads were quality controlled using BBDuk (Bushnell, 2016), removing low-quality bases ($Q < 20$). Then, quality-controlled paired-end reads were filtered and assembled into complete plastomes using the script 'get_organelle_from_reads. py' from GetOrganelle v.1.7.5 (Jin *et al.*, 2020*b*), which uses Bowtie2 (Langmead and Salzberg, 2012), BLAST (Camacho *et al.*, 2009), and SPAdes 3.1.0 (Bankevich *et al.*, 2012), in addition to Python dependencies, implemented on the HiPerGator SLURM supercomputing cluster housed at the University of Florida (Gainesville, FL, USA). We used default settings, with kmers (-k) set as 21, 45, 65, 85, 105, 115, 127. When necessary, additional parameters were set [e.g. reducing word size (-w), increasing the maximum extension rounds (-R) or providing a close-relative seed database (-s)] to assemble complete graphs. Final assembly graphs were checked in Bandage (Wick *et al.*, 2015) to evaluate visually their overall structures and repeated regions. The boundary junctions between IR and the SC regions, and the putative induced isomers were visually checked in Geneious v.9.0.5 (Biomatters, Auckland, New Zealand) with an *in silico* approach using the library information of paired-end reads with a reference-mapping approach adapted from Jin *et al.* (2020*a*) and Oliver *et al.* (2010). Owing to the relatively short average insert size of our data (Supplementary Data Table S1), the confirmation of isomers in plastomes was limited to samples in which paired reads spanned the entire regions involved (Supplementary Data Fig. S1). Other putative isomers resulting from the flip-flop recombination mediated by the IR

or the short-IRs (sIR; see Martin *et al.*, 2014; Jin *et al.*, 2020c) yielding two different orientations of the SSC were obtained in GetOrganelle, and one of them was selected arbitrarily, because they can coexist in cells and should not impact downstream analysis (see Palmer, 1983; Walker *et al.*, 2015). Annotations were performed with GeSeq (Tillich *et al.*, 2017), using default parameters to predict protein-coding genes by BLAST search, but adding as third-party references NCBI RefSeqs of *Arabidopsis thaliana* (NC_000932), *Spinacia oleracea* (NC_002202), *Solanum lycopersicum* (AC_000188) and *Glycine max* (NC_007942), and tRNAscan-SE v.2.0.7 (Chan *et al.*, 2021) was selected as a third party to annotate tRNA. All sequences were imported into Geneious, and their annotations were curated manually to adjust the boundaries of start and stop codons according to the translated CDS (coding sequence) of *A. thaliana* as a reference for each coding gene, or equivalent open reading frame (ORF) of the sequence. Genes truncated with stop codons within the frame length of the expected coding gene and/or exceedingly divergent protein translations (when compared with *A. thaliana* as reference) were treated as pseudogenes, missing CDS annotation. For *accD*, *ycf1* and *ycf2*, we performed additional BLASTN searches (fragmenting the entire length region in 150–300 bp queries) in the nt database (excluding Cactaceae from the records) to check for significant alignments with the respective CDS feature, assessing putative loss or pseudogenization, as has been found previously in those genes (Sanderson *et al.*, 2015; Köhler *et al.*, 2020; Ruhlman and Jansen, 2021). Additionally, we aligned nucleotide and amino acid translations of these putative pseudogenes with other functional genes of *Eucalyptus globulus* (KC180787; Bayly *et al.*, 2013), *Nicotiana tabacum* (NC_001879; Kunnimalaiyaan and Nielsen, 1997), *Portulaca oleracea* (NC_036236; Liu *et al.*, 2018) and *Spinacia oleracea* (NC_002202; Schmitz-Linneweber *et al.*, 2001) using MAFFT v.7.308 (Katoh and Standley, 2013) to assess their pseudogene status further, checking putative conserved motif regions of *ycf1* (de Vries *et al.*, 2017) and *accD* (Lee *et al.*, 2004). The recognition and annotation of the LSC, SSC, IRs and sIR were performed using Geneious, based on the outputs from GeSeq and graphs analysed in Bandage.

Plastome variation and comparative analyses

We investigated the variation of plastomes of our taxon sampling in a phylogenetic framework. Considering that plastomes are represented as circular monomers, we arbitrarily established the 3′-end of the *trnH*^{GUG} gene as the beginning of the monomer as a linear sequence for all plastomes in downstream analyses. Sequences had the second copy of IR or sIR removed and were investigated visually regarding the overall gene order. We also performed analyses with the progressiveMauve algorithm in Mauve v.2.3.1 (Darling *et al.*, 2004; using Geneious plugin) with default settings to double-check our visual inspections. This analysis was performed twice: one analysis using an outgroup as a reference to check rearrangements compared with canonical plastomes (we used *Portulaca oleracea* L., which has a canonical angiosperm plastome and is one of the closest relatives of Cactaceae; see Walker *et al.*, 2018; sequence from GenBank accession KY490694; Liu *et al.*, 2018) and one analysis with Opuntieae samples only. We categorized

the Opuntieae plastomes into different types (arbitrarily named numerically following phylogenetic relationships from the root to tip) based on the pattern of the IR feature compared with our outgroup and among Opuntieae. We looked for tandem repeats in plastomes with Phobos v.3.3.12 (Mayer, 2006), using a perfect search mode and a minimum repeat unit length of 1 bp and a maximum of 1 kb, with default score constraints. We used the phylogenetic tree with more parsimony-informative sites and bootstrap means to map the plastome features of the lineages observed in our study manually (see details below).

Phylogenetic and incongruence analyses

By checking the collinear arrangement of genes within Opuntieae, we performed multiple sequence alignments using MAFFT v.7.308 (Katoh and Standley, 2013), with an automatic search for algorithm selection strategy and the default setting for score matrix and open gap penalty. We built two datasets for independent and comparative analyses: one representing the entire plastome sequences (full ptDNA) and the other with plastidial genes only (ptGenes, including exon, intron and pseudogene sequences extracted from the annotated plastome, with one of the IR/sIR stripped; Supplementary Data Table S2 lists the ptGenes). For each of these datasets, we performed four different alignment strategies for tree inference: (1) raw, as output from MAFFT; and the others trimming the alignment using GBLOCKS v.0.91.1 (Castresana, 2000; Talavera and Castresana, 2007), with default settings (minimum number of sequences for a conserved position, b1 = 50 % + 1; minimum number of sequences for a flank position, b2 = 85 %; maximum number of contiguous non-conserved positions, b3 = 8, minimum length of a block, b4 = 10), but varying in the three options of gap presences: (2) Gblocks, no gaps allowed; (3) Gblocks, half gaps allowed; and (4) with all gaps allowed. We then performed phylogenetic inference using maximum likelihood criteria implemented in RAxML 8.2.4 (Stamatakis, 2014) in the CIPRES Portal (Miller *et al.*, 2010). Given that RAxML is designed mainly to implement generalized time-reversible molecular models (GTR), we used the GTR+G model for the entire sequence, which has been suggested for topological reconstruction skipping model selection (Abadi *et al.*, 2019), and GTR+I+G is not recommended by Stamatakis (see RAxML v.8.2 manual) given the potential interaction between the I and G parameters. Support values were estimated by implementing 1000 bootstrap pseudoreplicates, and the clade names and circumscriptions were derived from previous studies (Majure *et al.*, 2012; Majure and Puente, 2014) and other ongoing projects (M. Köhler, UFSCar, São Paulo, Brazil, & L.C.Majure, FLMNH, Florida, USA, unpublished results).

We assessed putative phylogenetic incongruences among datasets and alignment strategies first visually. By checking incongruences between the relationship of some major clades across our datasets, we performed additional analyses, implementing the framework presented by Smith *et al.* (2015) and Shen *et al.* (2017). We annotated intergenic spacers from the plastomes using a customized R script with functions of the genbankr package (Becker and Lawrence, 2022), extracted 220 genes and intergenic spacer regions, and estimated individual gene trees in RAxML as previously described. We then assessed with PhyParts (Smith *et al.*, 2015) the number

of markers (genes and spacers) supporting each bipartition in our primary topology (full ptDNA, raw, which yielded the highest bootstrap means and lower standard deviation values, see Results; Supplementary Data Table S3), in addition to the number of markers supporting the other main alternative topology, remaining alternative topologies and not supporting any topology (neutral). PhyParts input included the primary topology and gene trees for each marker (based on full ptDNA, raw dataset), under the 'fullconcon' analysis (-a 1). Branches of gene trees with bootstrap values <40 % were collapsed to filter spurious phylogenetic relationships. PhyParts output was summarized with the Python function phypartspiecharts. py (Johnson, 2017) and functions of the R package ape v.5.6.2 (Paradis and Schliep, 2019). Additionally, two major incongruent node relationships involving the positioning of BT (*Brasiliopuntia* + *Tacinga*) and MSA (*Miqueliopuntia* + *Salmonopuntia* + *Airampoa*) clades (delta test 1), and Nopalea and Basilares clades (delta test 2) were investigated (see the Results for details regarding topological differences). For each test, two phylogenies where only the position of these clades varied were selected from our pool of trees. Then, Shimoidara–Hasegawa tests (SH test; Shimodaira and Hasegawa, 1999) were performed on the two alternative topologies using each individual marker alignment (genes and spacers), including raw alignments, and filtered using Gblocks with no gaps allowed. The marker-wise delta log-likelihood, alignment length, number of variant sites and number of parsimony-informative sites were recorded for each comparison, and significance was assessed through 10 000 bootstrap replicates. The SH test was performed with the R package phangorn v.2.10 (Schliep, 2011).

RESULTS

Sequencing and basic assembly results

We sequenced 43 new accessions across Opuntieae, representing all seven genera of the tribe and a comprehensive diversity of each genus (for full details, see Supplementary Data Table S1). Runs on the Illumina HiSeq X resulted in 484 873 592 reads in total, between 5 279 118 (*Tacinga saxatilis*) and 19 016 846 (*Opuntia macrocentra*) per sample, for a mean read number of 11 276 130 sequences per sample. Reads per sample following quality control were between 4 881 904 and 18 531 942, with a mean post-quality control read pool number of 10 989 653 sequences per sample. The GC content following quality control was between 36.6 % (*T. saxatilis*) and 40.2 % (*Opuntia austrina*). For all samples used here, we successfully assembled complete plastomes, with an average base coverage varying between 80.1× (*Brasiliopuntia schulzii*) and 687.1× (*Opuntia cuija*), and the percentage of reads used for plastid assembly was between 1.08 % (*B. schulzii*) and 11.57 % (*Opuntia polyacantha*). The insert size mean of paired reads varied between 175 and 489 bp per sample, with an average of 344 bp (Supplementary Data Table S1).

Plastome features and variation within Opuntieae

Plastid genomes showed striking variation in size, varying in length from ~121 kb (such as in *T. saxatilis* and *O. polyacantha*) to ~160–162 kb (e.g. *Brasiliopuntia* spp. and *Consolea* spp.) (Fig. 1; Table 1). This size variation was remarkably associated with the dynamic movement of expansion or contraction of the gene content in the inverted repeat region across Opuntieae, including the lack of the canonical IR in *Tacinga* spp. and in members of the Basilares, Scheerianeae, Setispina, Macrocentra and Humifusa clades, which have atypical sIR regions (varying from 183 to ~2 kb), and an exceptional case of a short direct repeat (sDR, 793 bp) and a pair of two sIR (657 and 708 bp) in *O. polyacantha* (Figs 1 and 2; Table 1; Supplementary Data Fig. S2).

In total, nine distinct types of plastomes were assembled and annotated based on the pattern of the IR feature, considering their presence or absence, gene content and size (Fig. 1; Table 1; Supplementary Data Fig. S2). Three type of plastomes lack the canonical IR: the type 2, in *Tacinga* spp., presenting only a fragmented remnant $rpl32^{\Psi}$ of ~180 bp; the type 7, in members of Basilares, Scheerianeae, Setispina, Macrocentra and Humifusa clades, which presents short-IRs of ~1–2 kb mostly reduced to $rpl23$ (eventually fragmented), $trnI^{CAU}$, $trnL^{CAA}$ and a pseudogenized $ycf2^{\Psi}$; and the type 8, so far unique in *O. polyacantha*, presenting only an sDR of $rpl23$ and $trnI^{CAU}$, and a pair of sIRs, one of $trnF^{GAA}$ and other of $trnL^{CAA}$ (Table 1). The other six plastome types (types 1, 3–6 and 9) differ in their gene content and size of the IR, varying in the IR size between ~16 kb (Nopalea clade) and ~35 kb (*Consolea* and *Brasiliopuntia* clades), containing most of the typical IR genes, but with notable expansions (encompassing genes usually present in the SSC), or the transfer of typical IR genes to the single copy region.

In general, Opuntieae plastomes contain 110 unique genes, including 76 protein-coding sequences (CDS), 30 transfer RNAs (tRNA) and four ribosomal RNAs (rRNA) (Table 1). Variations are observed when putative pseudogenization or gene losses are presented in some plastomes, such as the $rpl23$ (which is pseudogenized in some lineages with plastomes type 3 and 7: Elatae, Scheerianeae and Macrocentra clades; Supplementary Data Fig. S3D), and some genes of the ndh suite (e.g. $ndhE$ and $ndhF$, which are pseudogenized in *Opuntia basilaris* and *O. polyacantha*; Supplementary Data Fig. S3E; and $ndhG$, $ndhJ$ and $ndhK$, which are lost in *O. polyacantha*). Additionally, all Opuntieae plastomes seem to have undergone pseudogenization of the $ycf1$, $ycf2$ and $accD$; the loss of one intron in $rpl2$, and two introns in $clpP$. The $ycf1$ region varied from ~2 kb (*Opuntia retrorsa* and *Opuntia colubrina*) to ~4 kb (*Consolea* spp.), in all cases containing fragmented ORFs within the length frame of the expected product of the gene (Supplementary Data Fig. S3B), accumulating highly divergent sequences with scattered alignable regions across other angiosperms, and conserving only small fragments with BLAST identity to the conserved $ycf1$ C-terminal motif region (RLEDLACMNRFW) of other angiosperm lineages. Likewise, the $ycf2$ region varied greatly across Opuntieae, from ~328 bp (e.g. *Opuntia rufida*, *Opuntia microdasys* and *Opuntia stenopetala*) to >6 kb (e.g. *O. colubrina*, *Opuntia arechavaletae* and *Brasiliopuntia* spp.), also accumulating highly divergent sequences and fragmented ORFs within the length frame of the expected product of the gene, preserving only small and scattered fragments with BLAST identity to conserved $ycf2$ genes of other angiosperm lineages (Supplementary Data Fig. S3C). The species of the
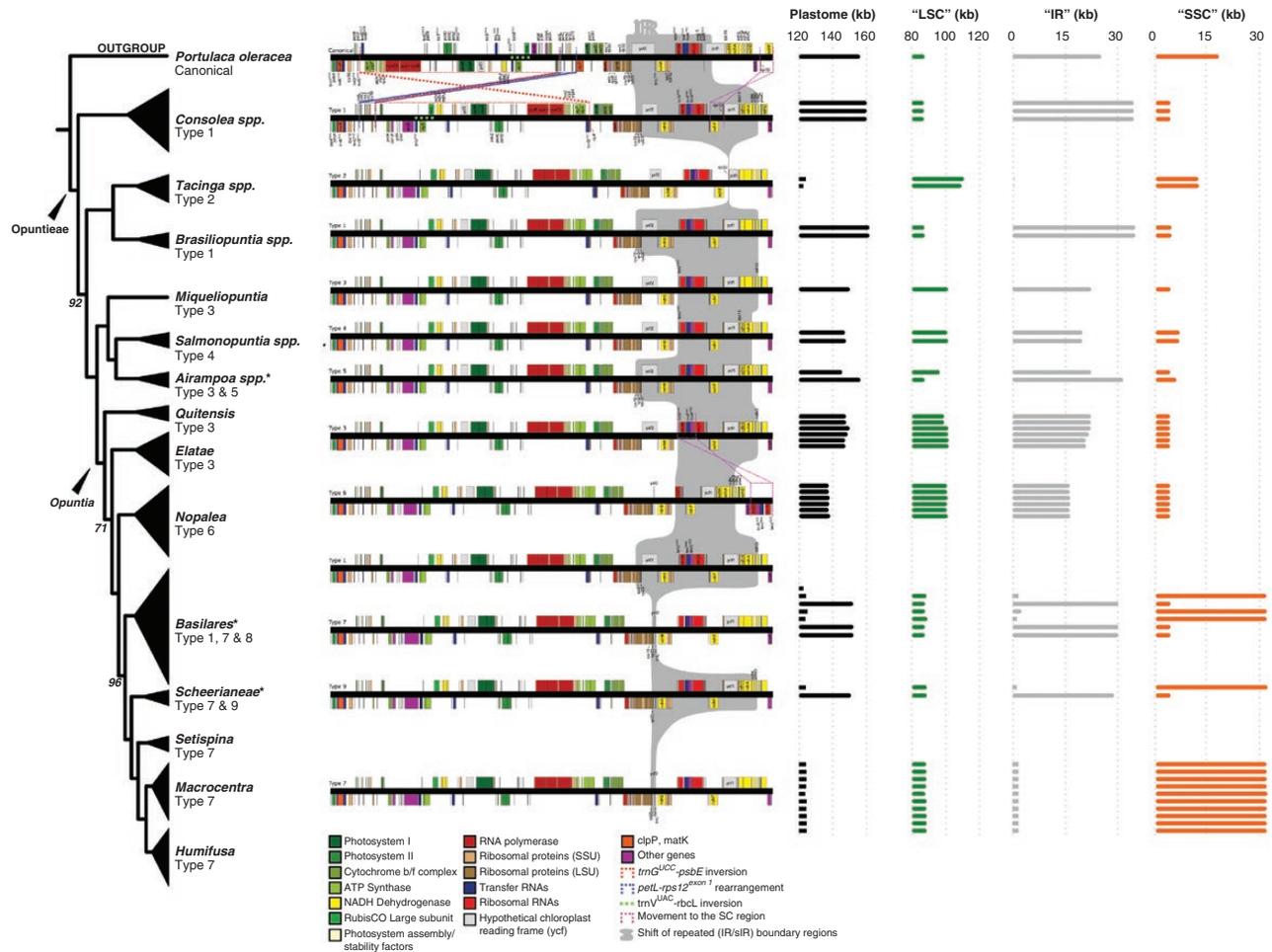
FIG. 1. Plastome variation across Opuntieae lineages. The linear sequences of the plastomes are depicted accompanying the phylogenetic relationships inferred for the tribe (for plastome type 8, see Supplementary Data Fig. S1). On the right, bar plots indicate variation in the size of plastomes and their content. Variations in the IR feature of plastome types are highlighted by the grey box in the linear plastome sequences. On the tree tips, asterisks (*) indicate that more than one plastome type is recovered in the clade. On the bar plot titles, quotation marks indicate that the term is not fully coherent with some plastomes considering that they lack an IR, hence the use of LSC or SSC is for convenience across the dataset. Phylogenetic nodes have total bootstrap support values, except when noted.

North American clade presented a second degradation in the *ycf2*, reducing the region to ~400 bp, and the apparent loss in *O. basilaris* and *O. polyacantha*. The *accD* region is presented with a long ORF of ~3.5 kb, accumulating long, scattered and divergent fragments of alignable sequences (Supplementary Data Fig. S3A) and conserving the five C-terminal motifs with identity to canonicals *accD* of other angiosperm lineages.

All Opuntieae plastomes assembled here shared rearrangements involving two blocks, an apparent translocation of the *petL–rps12*exon 1 (~4 kb; Fig. 1, blue dotted line) and an inversion involving the *trnG*UCC*–psbE* region (~58 kb; Fig. 1, red dotted line), compared with canonical angiosperm plastomes, as represented by *Portulaca oleracea* (Liu *et al.*, 2018). Considering the adjacency of the two rearranged blocks, the rearrangement might also be caused by two inversion events, an inversion of the *trnG*UCC*–rps12*exon 1 region followed by an inversion of the *rps12*exon 1*– petL* region. Within the trnG^UCC*–petL* block, a second inversion occurred involving the *trnV*UAC*–rbcL* inversion (~5 kb; Fig. 1, green dotted line), which placed the *accD* region adjacent to *trnV*UAC, and not to *rbcL*, as it is in canonical angiosperm plastomes.

Opuntieae plastomes harbour a significant number of tandem repeats [simple sequence repeats (ptSSRs)], which are relatively conserved in number among lineages, varying from 816 (*Opuntia guatemalensis*) to 857 (*Salmonopuntia schickendantzii*), totalling from ~7.4 to 8.4 % of the plastome content (*O. colubrina* and *Tacinga palmadora*, respectively) (Table 2; for full data, see Supplementary data Table S4). Most of the SSRs are mononucleotide repeats derived from A/T, varying in length from 7 to 29 bp, but a remarkable presence of more complex repeats (penta, hexa and more than seven nucleotides) is noteworthy in number and length.

Phylogenetic relationships and plastome evolution

Plastid genome sequences were shown to be informative to infer phylogenetic relationships within Opuntieae lineages, supporting 11 major clades, based on our sampling, with high support values (Fig. 2, highlighted and annotated clades). The full ptDNA dataset resulted in a raw alignment of 137 064 bp, with 5135 distinct patterns, 3769 parsimony-informative sites (PIS)

TABLE 1. *Plastome features of Opuntieae members. Asterisks (*) indicate that the gene is in the boundary of IR and SC, being fragmented in one of the IR/SC junctions. Psi (Ψ) indicates pseudogene; NA, not applicable.*

Clade/species	Plastome length (bp)	IR feature	IR/sIR/sDR gene content	IR/sIR length (bp)	LSC (bp)	SSC (bp)	Number of unique genes (CDS, tRNA, rRNA)	Pseudogenization (Ψ) or loss (X)
Consolea								
C. macracantha	160 649	Type 1: expanded	ndhA, ndhB, ndhG, ndhF, ndhH, ndhI, rpl2, rpl23, rpl32, rps12, rps15, rps19*, rps7, rrn16, rrn23, rrn4.5, rrn5, trnA^{UGC}, trnI^{CAU}, trnI^{GAU}, trnL^{CAA}, trnN^{GUU}, trnR^{ACG}, trnV^{GAC}, ycf1^{Ψ}, ycf2^{Ψ}	34 824	86 905	4096	76 CDS, 30 tRNA, 4 rRNA	accD^{Ψ}, ycf1^{Ψ}, ycf2^{Ψ}
C. moniliformis	160 839	expanded		34 823	87 097	4096		
C. rubescens	160 830			34 823	87 084	4100		accD^{Ψ}, ycf1^{Ψ}, ycf2^{Ψ}
Tacinga								
T. palmadora	123 381	Type 2: IR-lacking	rpl32*	183	110 837	12 178	76 CDS, 30 tRNA, 4 rRNA	
T. saxatilis	121 937			182	109 273	12 300		
Brasiliopuntia								
B. brasiliensis	162 410	Type 1: expanded	ndhA, ndhB, ndhF, ndhG*, ndhH, ndhI, rpl2, rpl23, rpl32, rps12, rps15, rps19, rps7, rrn16, rrn23, rrn4.5, rrn5, trnA^{UGC}, trnI^{CAU}, trnI^{GAU}, trnL^{CAA}, trnN^{GUU}, trnR^{ACG}, trnV^{GAC}, ycf1^{Ψ}, ycf2^{Ψ}	35 320	87 379	4391	76 CDS, 30 tRNA, 4 rRNA	accD^{Ψ}, ycf1^{Ψ}, ycf2^{Ψ}
B. schulzii	162 564			35 320	87 531	4393		
Miqueliopuntia								
M. miquelii	150 503	Type 3: modified (gene content)	ndhA, ndhF, ndhG, ndhH, ndhI, rpl32, rps15, rrn16, rrn23, rrn4.5, rrn5, trnA^{UGC}, trnI^{GAU}, trnN^{GUU}, trnR^{ACG}, trnV^{GAC}, ycf1^{Ψ}	22 590	101 231	4092	76 CDS, 30 tRNA, 4 rRNA	accD^{Ψ}, ycf1^{Ψ}, ycf2^{Ψ}
Salmonopuntia								
S. salmiana	147 474	Type 4: modified (gene content)	ndhA*, ndhF, ndhH, rpl32, rps15, rrn16, rrn23, rrn4.5, rrn5, trnA^{UGC}, trnI^{GAU}, trnN^{GUU}, trnR^{ACG}, trnV^{GAC}, ycf1^{Ψ}	19 919	100 911	6725	76 CDS, 30 tRNA, 4 rRNA	accD^{Ψ}, ycf1^{Ψ}, ycf2^{Ψ}
S. schickendantii	148 035			20 027	101 361	6620		
Airampoa								
A. erectoclada	145 686	Type 3: modified (gene content)	ndhA, ndhF, ndhG, ndhH, ndhI, rpl32, rps15, rrn16, rrn23, rrn4.5, rrn5, trnA^{UGC}, trnI^{GAU}, trnN^{GUU}, trnR^{ACG}, trnV^{GAC}, ycf1^{Ψ}	22 571	96 447	4097	76 CDS, 30 tRNA, 4 rRNA	accD^{Ψ}, ycf1^{Ψ}, ycf2^{Ψ}
A. soehrensii	156 992	Type 5: modified (gene content)	ndhA*, ndhB, ndhF, ndhH, ndhI, rpl2, rpl23, rpl32, rps12, rps15, rps19*, rps7, rrn16, rrn23, rrn4.5, rrn5, trnA^{UGC}, trnI^{CAU}, trnL^{CAA}, trnN^{GUU}, trnR^{ACG}, trnV^{GAC}, ycf1^{Ψ}, ycf2^{Ψ}	31 821	87 505	5845		
Opuntia								
Quitensis								
O. macbridei	148 031	Type 3: modified (gene content)	ndhA, ndhF, ndhG, ndhH, ndhI, rpl32, rps15, rrn16, rrn23, rrn4.5, rrn5, trnA^{UGC}, trnI^{GAU}, trnN^{GUU}, trnR^{ACG}, trnV^{GAC}, ycf1^{Ψ}	22 476	98 975	4104	76 CDS, 30 tRNA, 4 rRNA	accD^{Ψ}, ycf1^{Ψ}, ycf2^{Ψ}
O. quitensis	148 114			22 476	99 057	4105		
Elatae								
O. arechavaletae	149 275	Type 3: modified (gene content)	ndhA, ndhF, ndhG, ndhH, ndhI, rpl32, rps15, rrn16, rrn23, rrn4.5, rrn5, trnA^{UGC}, trnI^{GAU}, trnN^{GUU}, trnR^{ACG}, trnV^{GAC}, ycf1^{Ψ}	21 916	101 335	4108	76 CDS, 30 tRNA, 4 rRNA	accD^{Ψ}, ycf1^{Ψ}, ycf2^{Ψ}
O. colubrina	147 853			21 051	101 636	4115	76 CDS, 30 tRNA, 4 rRNA	accD^{Ψ}, ycf1^{Ψ}, ycf2^{Ψ}, rpl23^{Ψ}
O. quimilo	150 374			22 392	101 475	4115	75 CDS, 30 tRNA, 4 rRNA	
O. retrorsa	147 640			20 921	101 684	4114		

Table 1. *Continued*

Clade/species	Plastome length (bp)	IR feature	IR/sIR/sDR gene content	IR/sIR length (bp)	LSC (bp)	SSC (bp)	Number of unique genes (CDS, tRNA, rRNA)	Pseudogenization (Ψ) or loss (X)
Nopalea								
O. auberi	138 379	Type 6: modified (gene content)	*ndhA, ndhF, ndhG, ndhH, ndhI, rpl32, rps15, rrn23*, rrn4.5, rrn5, trnN^{GUU}, trnR^{ACG}, ycf1$^{\Psi}$*	16 604	101 064	4107	76 CDS, 30 tRNA, 4 rRNA	*accD$^{\Psi}$, ycf1$^{\Psi}$, ycf2$^{\Psi}$*
O. caracassana	138 141			16 503	101 028	4107		
O. dejecta	137 452			16 368	100 609	4107		
O. gaumeri	137 818			16 500	100 711	4107		
O. guatemalensis	137 494			16 377	100 636	4104		
O. jamaicensis	137 596			16 376	100 735	4109		
Basilares								
O. basilaris	123 500	Type 7: short-IR	*ndhB*, rpl23*, trnI^{CAU}, trnL^{CAA}*	1730	88 469	31 571	74 CDS, 30 tRNA, 4 rRNA	*accD$^{\Psi}$, ycf1$^{\Psi}$, ycf2^{X}, ndhE$^{\Psi}$, ndhF$^{\Psi}$*
O. microdasys	124 337		*rpl2, rpl23, rps19*, trnI^{CAU}, trnL^{CAA}*	2500	87 528	31 809	76 CDS, 30 tRNA, 4 rRNA	*accD$^{\Psi}$, ycf1$^{\Psi}$, ycf2$^{\Psi}$*
O. rufida	123 145		*trnI^{CAU}, trnL^{CAA}*	1150	89 040	31 805		
O. pachyrrhiza	152 690	Type 1: expanded	*ndhA, ndhB, ndhF, ndhG, ndhH, ndhI, rpl2, rpl23, rpl32, rps12, rps15, rps19*, rps7, rrn16, rrn23, rrn4.5, rrn5, trnA^{UGC}, trnI^{CAU}, trnL^{CAU}, trnL^{CAA}, trnN^{GUU}, trnR^{ACG}, trnV^{GAC}, ycf1$^{\Psi}$*	30 469	87 641	4111		
O. pycnantha	152 441			30 382	87 562	4115		
O. stenopetala	152 717			30 530	87 545	4112		
O. polyacantha	121 985	Type 8: one short-DR, two short-IR	sDR: *rpl23*, trnI^{CAU}*. sIR$_1$: *trnF^{GAA}*, sIR$_2$: *trnL^{CAA}*	NA	NA	NA	71 CDS, 30 tRNA, 4 rRNA	*accD$^{\Psi}$, ycf1$^{\Psi}$, ndhE$^{\Psi}$, ndhF$^{\Psi}$, ycf2^{X}, ndhG^{X}, ndhI^{X}, ndhK^{X}*
Scheerianeae								
O. cuija	123 275	Type 7: short-IR	*rpl23*, trnI^{CAU}, trnL^{CAA}*	1203	88 784	32 085	75 CDS, 30 tRNA, 4 rRNA	*accD$^{\Psi}$, ycf1$^{\Psi}$, ycf2$^{\Psi}$, rpl23$^{\Psi}$*
O. scheeri	151 001	Type 9: expanded	*ndhA, ndhB, ndhF, ndhG, ndhH, ndhI, rpl23*, rpl32, rps12, rps15, rps7, rrn16, rrn23, rrn4.5, rrn5, trnA^{UGC}, trnI^{CAU}, trnL^{CAA}, trnN^{GUU}, trnR^{ACG}, trnV^{GAC}, ycf1$^{\Psi}$**	29 089	88 718	4105	76 CDS, 30 tRNA, 4 rRNA	*accD$^{\Psi}$, ycf1$^{\Psi}$, ycf2$^{\Psi}$*
Setispina								
O. setispina	123 480	Type 7: short-IR	*rpl23, trnI^{CAU}, trnL^{CAA}*	1752	88 297	31 679	76 CDS, 3$^{\Psi}$, 30 tRNA, 4 rRNA	*accD$^{\Psi}$, ycf1$^{\Psi}$, ycf2$^{\Psi}$*
Macrocentra								
O. aureispina	124 016	Type 7: short-IR	*rpl23$^{\Psi}$*, trnI^{CAU}, trnL^{CAA}*	1763	88 759	31 731	75 CDS, 30 tRNA, 4 rRNA	*accD$^{\Psi}$, ycf1$^{\Psi}$, ycf2$^{\Psi}$, rpl23$^{\Psi}$*
O. chisosensis	123 987			1763	88 729	31 732		
O. chlorotica	123 748			1767	88 511	31 703		
O. macrocentra	123 988			1763	88 730	31 732		
O. strigil	123 994			1783	88 723	31 705		

TABLE 1. *Continued*

Clade/species	Plastome length (bp)	IR feature	IR/sIR/sDR gene content	IR/sIR length (bp)	LSC (bp)	SSC (bp)	Number of unique genes (CDS, tRNA, rRNA)	Pseudogenization (Ψ) or loss (X)
Humifusa								
O. austrina	123 091	Type 7: short-IR	rpl23, trnI^CAU, trnL^CAA	1767	87 823	31 734	76 CDS, 30 tRNA, 4 rRNA	accD^Ψ, ycf1^Ψ, ycf2^Ψ
O. drummondii	123 611			1767	88 346	31 731		
O. macrorhiza	123 881			1765	88 566	31 785		
O. mesacantha	123 784			1766	88 535	31 717		

TABLE 2. *Summary information of number, variation in length (in base pairs) and type of tandem repeats [simple sequence repeats (SSRs)] within Opuntieae plastomes. Numbers with dashes (–) indicate unit variations through the extreme of the recorded values.*

Clade	Total number of repeats	Total length of repeats (bp)	Plastome repeat content (%)	Type of repeat Mononucleotide Total	Length	Dinucleotide Total	Length	Trinucleotide Total	Length	Tetranucleotide Total	Length	Pentanucleotide Total	Length	Hexanucleotide Total	Length	>7 Total	Length
Consolea	838–846	10 005–10 233	7.9–8.1	336–338	7–17	41	8–16	59	9–13	86–87	10–14	99–100	11–19	131–132	12–19	85–89	14–348
Brasiliopuntia	843–852	9826–9982	7.7–7.8	332–334	7–18	42	8–26	57	9–15	89–92	10–14	94–96	11–17	133–134	12–17	96–97	14–183
Tacinga	832–835	10 111–10 361	8.3–8.4	332–334	7–19	41–42	8–20	53	9–15	88–93	10–14	95–97	11–19	125–126	12–80	91–97	14–287
MST	829–857	9599–9949	7.5–7.8	336–348	7–25	40–41	8–25	57–62	9–13	87–91	10–14	93–101	11–133	125–130	12–126	85–90	14–220
Quitensis	850–852	9820–9923	7.8	340	7–17	44	8–28	57–58	9–13	91–92	10–15	101	11–19	126	12–42	91	14–168
Elatae	838–857	9434–9711	7.4–7.6	335–343	7–19	42–44	8–20	58–62	9–26	88–95	10–14	95–102	11–20	125–127	12–17	84–88	14–254
Nopaleae	816–824	9359–9969	7.7–8.1	327–332	7–26	40–43	8–22	56	9–13	88–91	10–14	102–103	11–25	117–119	12–33	81–86	14–360
Basilares	826–845	99 487–10 104	7.7–8.3	328–336	7–24	39–43	8–24	57–63	9–20	87–91	10–16	92–104	11–70	120–125	12–63	87–95	14–372
Scheerianeae	831–837	9462–9937	7.7–8.1	339–340	7–17	43	8–20	54–55	9–13	87–89	10–14	101–102	11–25	123–125	12–17	81–86	14–357
Setispina	829	9316	7.6	336	7–29	44	8–26	55	9–13	86	10–14	99	11–19	128	12–17	81	14–198
Humifusa	823–834	9251–9812	7.6–8	332–334	7–16	41–42	8–20	55–57	9–13	86	10–14	97–99	11–19	126–129	12–19	83–89	14–191
Macrocentra	823–828	9468–9796	7.7–8	331–333	7–21	42	8–24	55–56	9–15	85–87	10–16	98–99	11–20	126–128	12–17	83–86	14–215
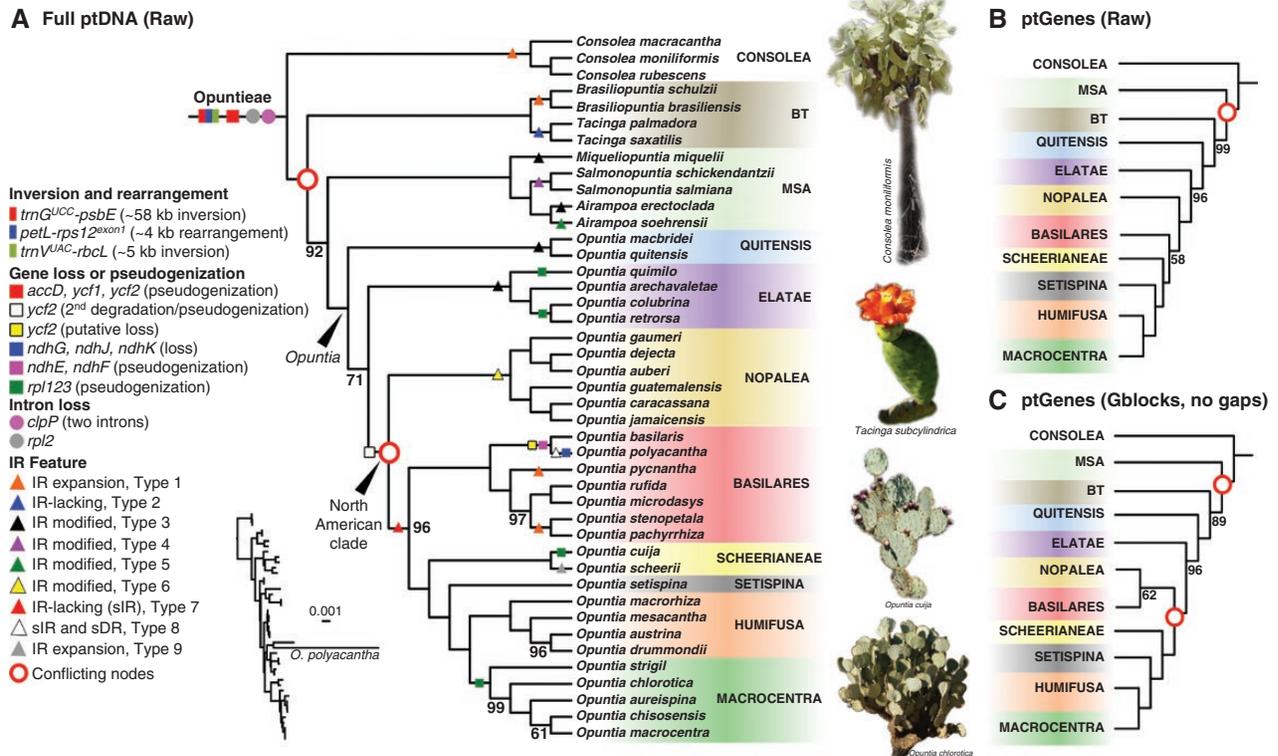
FIG. 2. Phylogenetic inference of tribe Opuntieae based on plastome sequences with the maximum likelihood (ML) criterion, with the major plastome evolution features mapped on the tree. (A) Inference based on the full plastome sequences (genes and intergenic regions, raw alignment). (B) Inference based on chloroplast genes (raw alignment). (C) Inference based on plastome genes trimmed with Gblocks (no gaps allowed). Phylogenetic nodes have full bootstrap support (100), except when depicted; incongruent nodes across different datasets are highlighted by red circles.

and 129 283 constant sites, whereas the ptGenes resulted in a raw alignment of 85 574 bp, 2677 distinct patterns, 2155 PIS and 82 324 constant sites (results for all alignment scenarios are given in Supplementary Data Table S3). All full plastome alignment scenarios (raw and three Gblocks settings) recovered the same backbone topology (Supplementary Data Fig. S4), with the raw alignment yielding higher bootstrap means and lower standard variations (Supplementary Data Table S3). However, contrasting topologies were recovered by comparing the full plastome dataset (full ptDNA) with the genes-only (ptGenes) datasets involving two nodes (Fig. 2A–C; Supplementary Data Fig. S4). *Consolea* was sister to the rest of Opuntieae, while the relationship of the BT or MSA clades as sister to the *Opuntia* clade was conflicting when comparing the full ptDNA dataset with the ptGenes dataset, despite both scenarios recovering high bootstrap support values. A similar conflicting topology was recovered within *Opuntia* lineages, with the full ptDNA and the ptGenes (raw alignment) supporting the Nopalea clade as sister to the rest of the North American (NA) *Opuntia* clades (Basilares + Scheerianeae + Setispina + Humifusa + Macrocentra), while the ptGenes (Gblocks no gaps, half gaps and all gaps allowed) recovered a third topology, suggesting a sister relationship between the Nopalea and Basilares clades, but with low bootstrap support (Fig. 2C).

Concordance analyses revealed that most markers of the full ptDNA (raw) lack phylogenetic power (Fig. 3A, grey pie chart), while some nodes of the phylogeny are supported by a few regions of the plastome sequence (Fig. 3A, blue pie

chart), even when they have maximum bootstrap support (bs) value (bs = 100; e.g. the North American clade of *Opuntia* has total bootstrap value and is supported by 14 plastome regions). Likewise, the alternative topologies regarding the two contentious relationships involving (1) the MSA or BT clade as sister to *Opuntia* (Fig. 3B), or (2) the Nopalea clade as sister to Basilares or sister to the rest of the NA *Opuntia* clades (Fig. 3C) are mainly driven by a handful of markers (Supplementary Data Table S5). Although the MSA/BT recalcitrant relationship seems to be a hard incongruence, with around a dozen markers disputing the alternative topologies, the Nopalea/Basilares contentious relationship seems to be more by a lack of resolution across the analysed regions than a dispute for alternative topologies (Supplementary Data Table S5).

Regardless of incongruent topologies, *Opuntia* is well resolved as monophyletic, including eight major clades (Fig. 2). The South American (SA) Quitensis clade (including *Opuntia macbridei* and *O. quitensis*) is sister to the rest of the genus, while the other SA clade, Elatae (including *O. retrorsa*, *O. colubrina*, *O. arechavaletae* and *O. quimilo*), is resolved as sister to the NA clade. The NA clade encompasses the Nopalea clade (including *Opuntia dejecta*, *O. auberi*, *O. gaumeri*, *O. guatemalensis*, *O. caracassana* and *O. jamaicensis*) as sister to the rest of the NA clades (in the full ptDNA and ptGenes raw datasets) or as sister to the Basilares clade (including *Opuntia pachyrrhiza*, *O. stenopetala*, *O. microdasys*, *O. rufida*, *O. pycnantha*, *O. polyacantha*, and *O. basilaris*) in the ptGenes (Gblocks scenarios). The
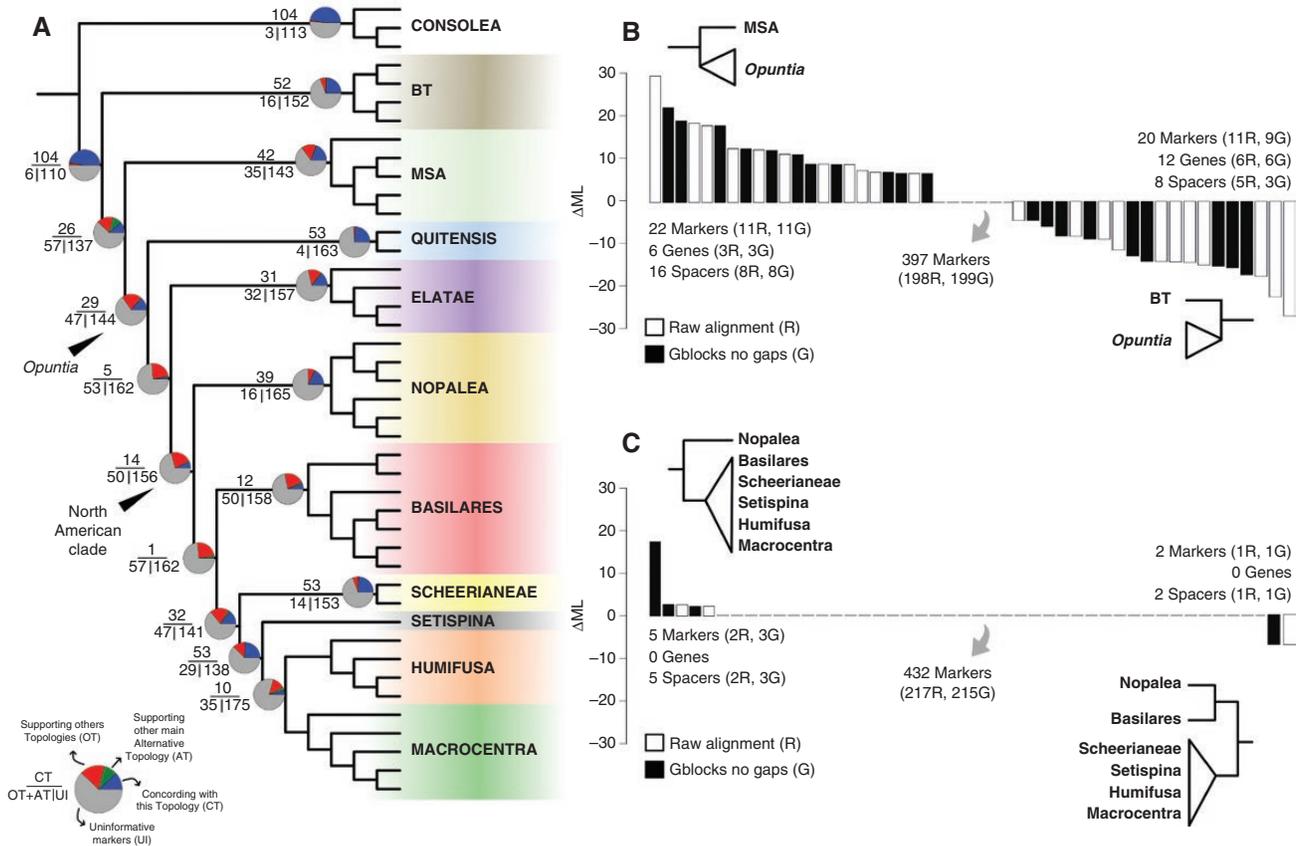
FIG. 3. Incongruence analyses. (A) Pie charts depicting the number of plastome markers (gene or spacer regions) supporting the topology shown (blue pie chart, numbers above), supporting other topologies (green and red pie charts, number below, before pipe, |), or uninformative markers (grey pie chart, number below after pipe). (B, C) Marker-wise delta log-likelihood (ΔML) for the alternative topologies of raw and Gblock alignments; bar plots shown with positive values are markers significantly supporting the schematic topology on the left, while negative values are markers supporting the right one. Grey bar plots are neutral markers, with no significative support for any of the compared topologies.

Scheerianeae clade (*Opuntia scheerii* and *O. cuija*) and the Setispina (*Opuntia setispina*) clades are subsequent sisters to a clade with the Humifusa (*Opuntia drummondii*, *O. austrina*, *O. mesacantha* and *O. macrorhiza*) + Macrocentra clades (*Opuntia macrocentra*, *O. chisosensis*, *O. aureispina*, *O. chlorotica* and *O. strigil*).

The evolution of plastome types across Opuntieae appears to be both homoplasious (presenting recurrent features in several different clades) and clade or species specific. The largest plastomes of Opuntieae (~160–162 kb, plastome type 1), which have incorporated some genes that are typically in the SSC into the IR region, are present in the *Consolea* and *Brasiliopuntia* clades but are also found in distantly related species of the Basilares clade (*Opuntia pycnantha*, *O. stenopetala* and *O. pachyrrhiza*). Likewise, plastome type 3, which is ubiquitous in the Elatae clade, is also observed in the Quitensis clade and members of the MST clade. In contrasts, plastome type 6, which has transferred some rRNA genes that are typically in the IR region to the SC region, is exclusive of the Nopalea clade, while the IR-lacking plastome type 2 is exclusive of the *Tacinga* clade. Likewise, the IR-lacking plastome type 7 is the most common in the NA *Opuntia* clade (pervasive in the Macrocentra, Humifusa and Setispina clades), although not exclusive to all lineages.

A similar pattern of independent occurrence is observed with gene loss and some pseudogenizations. Despite the major inversions (*trnG^{UCC}–psbE* and *trnV^{UAC}–rbcL*), rearrangement (*petL–rps12^{exon1}*), pseudogenization (*accD*, *ycf1* and *ycf2*) and intron loss (*clpP* and *rpl2*), which were observed in all Opuntieae plastomes, other events seem to have occurred independently. The *rpl23* pseudogenization is characteristic of the Macrocentra clade but is also present in independent lineages of the Scheerianeae and Elatae clades. In contrast, the second degradation/pseudogenization of *ycf2*, reducing the large region of ~3–6 kb to only ~400 bp, has occurred once for the NA clade. Likewise, the putative loss of *ycf2* and the *ndhE–ndhF* pseudogenization is observed only in the sister taxa, *O. basilaris* + *O. polyacantha*, while the *ndhG*, *ndhJ* and *ndhK* loss is unique in *O. polyacantha*.

DISCUSSION

The structure, content and arrangement of plastomes have been widely studied in the last decades, increasing especially with the advent of the high-throughput sequencing era, which has revolutionized innumerable aspects of evolutionary biology. Despite numerous advances supporting the overall conserved structure

of plastomes in land plants, there are also accumulating reports of variation in the theme across disparate lineages, which have challenged the idea of an overall conserved structure among plant lineages (Mower and Vickrey, 2018; Ruhlman and Jansen, 2018, 2021). Furthermore, other misconceptions can obfuscate our understanding of the evolution of this organelle (Gonçalves *et al.*, 2020). For example, the very idea of the plastome as a circular molecule (which seemed intuitive as bacterial descendants through the endosymbiotic origin of the organelle) has been rethought with the understanding of predominantly linear and branched molecules covalently linked through repeating units that undergo recombination both within and between units and molecules (Bendich and Smith, 1990; Lilly *et al.*, 2001; Bendich, 2004; Oldenburg and Bendich, 2004, 2015, 2016; Lee *et al.*, 2021). In this scenario, the mechanisms underlying such diversity of variation in content and structure of plastomes, as reported here, can be more complex to discern, but at the same time are more likely to occur. However, the incorporation of these emerging ideas in the ongoing flux of data with robust bioinformatics tools is still a bottleneck to providing a deeper understanding of plastome evolution.

The mechanisms by which structural variations in plastomes arise and are maintained have been discussed (reviewed by Wicke *et al.*, 2011; Jansen and Ruhlman, 2012; Ruhlman and Jansen, 2021). Inversions and deletions can occur owing to intra- or intermolecular homologous recombination when repeats are palindromic (Ogihara *et al.*, 1988; Maliga, 1993), and it has been suggested to be one of the generating causes of atypical plastomes (Guisinger *et al.*, 2008, 2011; Haberle *et al.*, 2008; Cauz-Santos *et al.*, 2020). Plastomes seem to replicate by a few concomitantly or sequentially different mechanisms (e.g. displacement-loop, rolling circle and recombination-dependent replication; Oldenburg and Bendich, 2015; Ruhlman and Jansen, 2021), which can also drive a certain amount of variability (Maréchal and Brisson, 2010), especially in the case of recombination-dependent replication (Ruhlman *et al.*, 2017; Choi *et al.*, 2019; Charbonneau *et al.*, 2021). The integrity of ptDNA is thought to decline during plant development, owing to the degradation of molecules that are damaged but not repaired (Oldenburg and Bendich, 2015), making mechanisms of plastid DNA recombination, replication and repair crucial steps for its maintenance. Cacti are exposed to many factors that can lead to DNA damage, owing to the harsh environments they usually occupy, which could be linked at some point with the noticeable variation in cactus plastomes that has been reported (Sanderson *et al.*, 2015; Majure *et al.*, 2019; Solórzano *et al.*, 2019; Köhler *et al.*, 2020; Oulo *et al.*, 2020; Almeida *et al.*, 2021; Amaral *et al.*, 2021; Silva *et al.*, 2021; Dalla Costa *et al.*, 2022; Qin *et al.*, 2022; Yu *et al.*, 2023). Additionally, ecophysiological constraints and the impacts of specialized photosynthesis systems (i.e. crassulacean acid metabolism) in plastome evolution are still scarcely explored. To propose further hypotheses of the putative mechanisms underlying plastome variations in Opuntieae, we encourage the generation of more data, including a comprehensive sampling of other cactus lineages (including emblematic forms of cacti, such as *Pereskia* spp., *Leuenbergeria* spp., *Blossfeldia* and *Maihuenia* spp.), which could represent additional plastome types and intermediate variations of that analysed here.

The transfer of genes between SC regions and the IR (shifts in the IR boundaries), and vice versa, is a frequent mechanism involved in plastome evolution, and it seems to have a major impact on plastome variation across disparate and closely related species (Zhu *et al.*, 2016; Ruhlman and Jansen, 2018; Choi *et al.*, 2020). In Opuntieae, this seems to be the main mechanism associated with the variation of plastome types observed here, accompanied by the events of loss of the IR region. Interestingly, plastomes within Opuntieae are well conserved regarding gene arrangement, conserving a collinear syntenic block of genes, which stresses a standing idea that rearrangements are more frequent when the large inverted repeat is lost (Palmer and Thompson, 1982). Although the presence of repetitive sequences in plastomes has been reported as an important mechanism in plastome size evolution and genomic rearrangements (Jo *et al.*, 2011; Dugas *et al.*, 2015; Wu *et al.*, 2021), our data do not support these hypotheses fully, considering that Opuntieae plastomes do not present significant rearrangements. Furthermore, despite Opuntieae presenting a striking variation in plastome size across major clades, the number of tandem repeats is similar among the clades. However, our data might help in further assessments of the impact of tandem repeats in pseudogenization and shifts in the inverted repeat region (Sinn *et al.*, 2018).

In contrast, in comparison to canonical angiosperms, as represented by the outgroup *Portulaca oleracea*, Opuntieae plastomes have notable rearrangements involving two blocks of sequences with putative inversions and translocations. All Opuntieae samples presented the *trnV*^{UAC}–*rbcL* inversion (~5 kb), which has long been proposed as a synapomorphy of Cactaceae (Downie and Palmer, 1994; Wallace, 1995). Some of these inversions might well have influenced a few features observed in some of the genes adjacent to the regions involved in the inversions, such as *clpP* (which has lost two introns and is reduced to a small, conserved domain, ~150 bp) and *accD* (which is presented as a long ORF of ~3.5 kb, accumulating a long and divergent fragment of sequence and only a small conserved domain, putatively a pseudogene). However, these genes are also widely reported as pseudogenes, lost or under positive selection in several disparate lineages not necessarily related to inversions (e.g. Jansen *et al.*, 2008; Wicke *et al.*, 2011; Harris *et al.*, 2013; Dugas *et al.*, 2015; Wang *et al.*, 2016; Ruhlman and Jansen, 2018), suggesting that other mechanisms might be involved in their modifications. Also, we have reported pseudogenization of *ycf1* and *ycf2*, which are not directly involved in rearrangement breakpoints but are in shift movements of SC/IR regions, especially *ycf2*. Although both genes appear to be essential for plastid function in most plants (Drescher *et al.*, 2000; de Vries *et al.*, 2015; Yang *et al.*, 2016; Kikuchi *et al.*, 2018), their loss or pseudogenization is relatively frequent (see Graham *et al.*, 2017; Ruhlman and Jansen, 2021). For example, all Poales have lost *ycf1* and *ycf2* in a progressive degradation of the gene sequence (Guisinger *et al.*, 2010; de Vries *et al.*, 2015), which could also have occurred in Opuntieae, because the loss of *ycf2* and small degraded fragments (~300 bp) is a shared feature of the most derived clades of *Opuntia* (the NA *Opuntia* clade: Nopalea + Basilares + Scheerianeae + Setispina + Humifusa + Macrocentra). However, the mechanisms responsible for such aspects remain elusive.

Other events of pseudogenization, gene and intron loss are remarkable within Opuntieae, representing unique or shared occurrences. The *rpl2* intron loss is ubiquitous in Opuntieae and has been reported as synapomorphic within Centrospermae lineages (Caryophyllales; Palmer *et al.*, 1988; Yao *et al.*, 2019). Likewise, the events of loss or pseudogenization of *ndh* suite genes have been reported in some cactus lineages (Sanderson *et al.*, 2015; Solórzano *et al.*, 2019; Köhler *et al.*, 2020; Amaral *et al.* 2021; Silva *et al.*, 2021; Dalla Costa *et al.*, 2022; Qin *et al.*, 2022), because this is especially associated with hemi- or holoparasitism, carnivory, xerophytes and submersed plants (Braukmann *et al.*, 2009; Wicke *et al.*, 2011; Peredo *et al.*, 2013; Silva *et al.*, 2016). Previous authors have suggested that retention of the *ndh* complex is associated with the transition of plants to stressful environments and that *ndh* loss would be associated with decreased environmental stressors but of limited biological significance in contemporary plants (Ruhlman *et al.*, 2015; Lin *et al.*, 2017). The association between *ndh* loss and the presence of crassulacean acid metabolism photosynthesis has also been speculated (Strand *et al.*, 2019; Köhler *et al.*, 2020) but is still elusive. In contrast, a dynamic transfer of segments of the plastid genome to the nuclear or mitochondrial genome, and vice versa, has been reported increasingly (Stegemann *et al.* 2003; Cui *et al.*, 2021; Hertle *et al.*, 2021), which could suggest the transition of these genes to other genomes. Sanderson *et al.* (2015) found many non-plastid copies of plastid *ndh* genes in the nuclear genome of the saguaro cactus, but none had intact reading frames. Based on the presence of other nuclear genes, they conclude that the existence of an alternative pathway, made redundant with the function of the plastid *ndh*, might have facilitated the loss of the plastid *ndh* gene suite in photoautotrophs, such as the saguaro, and putatively, other cacti.

The knowledge about plastome evolution across Opuntieae can help to elucidate the remarkable radiation of cactus diversity. *Opuntia* is the most species-rich lineage of the tribe, and likewise, across all of Cactaceae (Korotkova *et al.*, 2021). Despite this high diversity, which is considered to be young in origin (Arakaki *et al.*, 2011), plastome sequences are shown to be extremely informative in recovering the phylogenetic relationships of its major clades. Additionally, the identification of different plastome features according to the presence or absence of the IR, content and variation within the group promotes new insights into the diversification of the lineages and the putative drivers underlying such aspects, besides providing helpful signatures for barcoding and taxonomic assessment. For example, Chen *et al.* (2022) recently assembled a prickly pear plastome assigned to the *Opuntia sulphurea* taxon, which is a southern South American species. We analysed their assembly in our dataset and detected it as a problematic identification based on plastome features. The plastome from the study by Chen *et al.* (2022) is an IR-lacking sample (122 kb), similar to the IR-lacking type 7 of our dataset (or the type 8, present only in *O. polyacantha*), typical of the North American *Opuntia* species that are not part of the Nopalea clade, whereas *O. sulphurea* is part of the Elatae clade, with which it shares their type 3 plastome features (Köhler M., unpublished data). We confirmed this by checking phylogenetic analyses of the GenBank accession MW927506 in our dataset using maximum likelihood approaches with a reduced matrix using the

chloroplast markers listed by Köhler *et al.* (2020) as phylogenetically informative within Opuntioideae (Supplementary Data Fig. S5), and we suggest that the sample used by Chen *et al.* (2022) might represent *O. polyacantha*, which slightly resembles *O. sulphurea*, leading to erroneous identification.

Our phylogenetic analyses provided new and robust relationships within Opuntieae lineages with a comprehensive sampling when compared with previous studies (Griffith and Porter, 2009; Majure *et al.*, 2012; Majure and Puente, 2014; Köhler *et al.*, 2020, 2021). However, in our study, we revealed that different datasets (whole plastome sequences vs. genes only) and alignment strategies (raw or trimmed with different allowances of gaps in Gblocks) have yielded distinct topologies involving mainly two nodes. Phylogenomic analyses are sensitive to systematic or alignment errors and trimming biases (Philippe *et al.*, 2011, 2017; Zhong *et al.*, 2011; Walker *et al.*, 2019; Portik and Wiens, 2021), a pattern that was corroborated here. Furthermore, we demonstrated that a few plastid markers can support a major supported topology (high bootstrap values), in addition to underlying contentious relationships. Some of these recalcitrant nodes can be the result of hard incongruences, with a set of different markers supporting alternative topologies, whereas others can result from the lack of resolution across molecular sequences, usually leading to weak bootstrap support according to our analyses. Parins-Fukuchi *et al.* (2021) have demonstrated that phylogenomic conflicts coincide with rapid morphological innovations. The two contentious nodes recovered in our analyses might be more representative of this phenomenon. The MSA clade represents a peculiar group within Opuntieae regarding morphology (with putative plesiomorphic characters, such as the terete stems in *Miqueliopuntia* and *Salmonopuntia*a) and geography (being endemic to southern South American arid regions of Atacama and inter-Andean valleys), which represents contrasting transitions to the other mostly ubiquitous characteristics of flattened stems and broader distribution of members of the BT and *Opuntia* clades. Likewise, part of the Nopalea clade is characterized by members with unique floral features within Opuntieae showing hummingbird pollination syndrome, with short and erected tepals forming a tube with exserted stamens and styles. However, we are still lacking a nuclear phylogeny of the Opuntieae lineages to assess a more robust phylogenetic hypothesis, which will aid in understanding the evolutionary radiation of the group.

The issues involving topological incongruence among chloroplast markers have been explored recently (Gonçalves *et al.*, 2019, 2020; Walker *et al.*, 2019; Xiao *et al.*, 2020; Zhang *et al.*, 2020*a*, *b*), including within the cactus subfamily Opuntioideae (Köhler *et al.*, 2020), and topological incongruence has been suggested to result from systematic errors, to which phylogenomic analyses are sensitive, or a product of biological events (e.g. heteroplasmy and horizontal gene transfer) that require further investigation. In recent years, some studies have explored using the multispecies coalescent approach for plastome markers stemming from the evidence of heteroplasmic recombination; however, most of the results tend to be confounding (Thode *et al.*, 2021), while there are substantial arguments to continue treating plastomes as a single estimate of the underlying species phylogeny (Doyle, 2022). However, we submit that quantifying and filtering the phylogenetic signal in

plastome data is an important step for evaluating topological concordance, especially when performing downstream comparative analyses based on these data.

Conclusions

Cacti are one of the most charismatic groups of plants, but with remarkable difficulties for taxonomic and systematic understanding because of several challenges, such as hybridization, polyploidy, rapid diversification, the evolution of homoplasious characters and the intimidating prickly morphology, which leads to a general lack of primary data, because few scientists feel comfortable putting their hands on them. In this study, we reinforce that more than a prickly morphology, cacti harbour intriguing molecular aspects related to their plastome variation, putatively linked with their diversification history. Opuntieae, one impressive lineage of cacti, present a striking variation in plastome size related to contractions, expansions and the loss of the IR region, in addition to several cases of pseudogenization or gene loss. Despite some contentious signals across markers, analyses of incongruence can leverage plastome sequences to provide a robust framework to gain a deeper understanding of the evolutionary radiation of this group. Further studies, putting our dataset into a broader scale across the cacti tree of life or even exploring more closely related lineages, are necessary. Besides, innovative analyses should be carried out to address how ecological drivers, physiological constraints and morphological traits of cacti might be related to the variation in plastomes that has been reported across the family.

SUPPLEMENTARY DATA

Supplementary data are available at *Annals of Botany* online and consist of the following.

Table S1: taxa sampled in this study, voucher information, raw read and post-quality control numbers, %GC content, mean insert size and library content resulting from the chloroplast genome *de novo* assembly. **Table S2:** plastidial genes extracted from the plastomes and aligned for the ptGenes dataset. **Table S3:** statistics of alignments of each dataset. **Table S4:** total number and variation of tandem repeats identified in Opuntieae platomes. **Table S5:** complete delta gene-wise log-likelihood scores of topology tests, including associated data of alignment length and the number of variant and parsimony-informative sites. **Figure S1:** partial read mapping results confirming the existence of two candidate isomers. **Figure S2:** types of plastomes assembled in Opuntieae samples. **Figure S3:** status of the putative pseudogenes reported in Opuntieae based on a representative sampling of our data. **Figure S4:** maximum likelihood inference of all datasets and alignment strategies analysed. **Figure S5:** phylogenetic placement of the GenBank accession MW927506 (from Chen *et al.*, 2022) within the Basilares clade.

AUTHOR CONTRIBUTIONS

M.K.: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data Curation, Writing—original draft, Writing—review & editing, Visualization, Funding acquisition. M.R.: Methodology, Software, Validation, Formal analysis, Investigation, Writing—Review & Editing. J.-J.J.: Software, Methodology, Validation, Writing—review & editing. L.C.M.: Conceptualization, Methodology, Validation, Investigation, Resources, Writing—review & editing, Funding acquisition. All authors read and approved the final manuscript.

DATA AVAILABILITY

The dataset generated and analysed in this study can be found in the GenBank submission OQ613412–OQ613411. Supplemental information cited is included in this published article. The codes, alignments and information associated with the plastome assemblies are available on GitHub (https://github.com/tunasdelsur/OpuntieaePlastomes).

LITERATURE CITED

Abadi S, **Azouri D**, **Pupko T**, **Mayrose I. 2019**. Model selection may not be a mandatory step for phylogeny reconstruction. *Nature Communications* **10**: 934. doi:10.1038/s41467-019-08822-w.

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. *Brazilian Journal of Botany* **44**: 877–888. doi:10.1007/s40415-021-00772-2.

Amaral DT, **Bombonato JR**, **da Silva Andrade SC**, **Moraes EM**, **Franco FF. 2021**. The genome of a thorny species: comparative genomic analysis among South and North American Cactaceae. *Planta* **254**: 44. doi:10.1007/s00425-021-03690-5.

Anderson EF. 2001. *The cactus family*. Portland, Oregon: Timber Press.

Arakaki M, **Christin P-A**, **Nyffeler R**, *et al*. **2011**. Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 8379–8384. doi:10.1073/pnas.1100628108.

Bankevich A, **Nurk S**, **Antipov D**, *et al*. **2012**. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**: 455–477. doi:10.1089/cmb.2012.0021.

Bayly MJ, **Rigault P**, **Spokevicius A**, *et al*. **2013**. Chloroplast genome analysis of Australian eucalypts – *Eucalyptus*, *Corymbia*, *Angophora*, *Allosyncarpia* and *Stockwellia* (Myrtaceae). *Molecular Phylogenetics and Evolution* **69**: 704–716. doi:10.1016/j.ympev.2013.07.006.

Becker G, **Lawrence M. 2022**. *genbankr: parsing GenBank files into semantically useful objects. R package version 1.27.0*. https://bioconductor.org/packages/devel/bioc/html/genbankr.html.

Bendich AJ. 2004. Circular chloroplast chromosomes: the grand illusion. *The Plant Cell* **16**: 1661–1666. doi:10.1105/tpc.160771.

Bendich AJ, **Smith SB. 1990**. Moving pictures and pulsed-field gel electrophoresis show linear DNA molecules from chloroplasts and mitochondria. *Current Genetics* **17**: 421–425. doi:10.1007/bf00334522.

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. *Current Genetics* **55**: 323–337. doi:10.1007/s00294-009-0249-7.

Camacho C, **Coulouris G**, **Avagyan V**, *et al*. **2009**. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421. doi:10.1186/1471-2105-10-421.

Bushnell B. 2016. *BBMap short read aligner, and other bioinformatic tools*. https://sourceforge.net/projects/bbmap/ (4 June 2021, date accessed).

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17**: 540–552. doi:10.1093/oxfordjournals.molbev.a026334.

Cauz-Santos LA, **da Costa ZP**, **Callot C**, *et al*. **2020**. A repertory of rearrangements and the loss of an inverted repeat region in *Passiflora* chloroplast genomes. *Genome Biology and Evolution* **12**: 1841–1857. doi:10.1093/gbe/evaa155.

Chan PP, **Lin BY**, **Mak AJ**, **Lowe TM. 2021**. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Research* **49**: 9077–9096. doi:10.1093/nar/gkab688.

Charboneau JLM, **Cronn RC**, **Liston A**, **Wojciechowski MF**, **Sanderson MJ. 2021**. Plastome structural evolution and homoplastic inversions in neo-astragalus (Fabaceae). *Genome Biology and Evolution* **13**: evab215. doi:10.1093/gbe/evab215.

Chen J, **Zhang S**, **Tang W**, **Du X**, **Yuan Y**, **Wu S. 2022**. The complete chloroplast genome sequence of *Opuntia sulphurea* (Cactaceae). *Mitochondrial DNA Part B* **7**: 361–362. doi:10.1080/23802359.2022.2035837.

Choi I-S, **Jansen R**, **Ruhlman T. 2019**. Lost and found: return of the inverted repeat in the legume clade defined by its absence. *Genome Biology and Evolution* **11**: 1321–1333. doi:10.1093/gbe/evz076.

Choi I-S, **Jansen R**, **Ruhlman T. 2020**. Caught in the act: variation in plastid genome inverted repeat expansion within and between populations of *Medicago minima*. *Ecology and Evolution* **10**: 12129–12137. doi:10.1002/ece3.6839.

Cui H, **Ding Z**, **Zhu Q**, **Wu Y**, **Qiu B**, **Gao P. 2021**. Comparative analysis of nuclear, chloroplast, and mitochondrial genomes of watermelon and melon provides evidence of gene transfer. *Scientific Reports* **11**: 1595. doi:10.1038/s41598-020-80149-9.

Dalla Costa TP, **Silva MC**, **de Santana Lopes A**, *et al*. **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.

Daniell H, **Lin C-S**, **Yu M**, **Chang W-J. 2016**. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biology* **17**: 134. doi: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 Research* **14**: 1394–1403. doi:10.1101/gr.2289704.

Downie SR, **Palmer JD. 1994**. A chloroplast DNA phylogeny of the caryophyllales based on structural and inverted repeat restriction site variation. *Systematic Botany* **19**: 236–252. doi:10.2307/2419599.

Doyle JJ. 2022. Defining coalescent genes: theory meets practice in organelle phylogenomics. *Systematic Biology* **71**: 476–489. doi:10.1093/sysbio/syab053.

Doyle JJ, **Doyle JL. 1987**. A rapid DNA isolation procedure from small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**: 11–15.

Drescher A, **Ruf S**, **Calsa T**, **Carrer H**, **Bock R. 2000**. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *The Plant Journal: for Cell and Molecular Biology* **22**: 97–104. doi:10.1046/j.1365-313x.2000.00722.x.

Dugas DV, **Hernandez D**, **Koenen EJM**, *et al*. **2015**. Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions and accelerated rate of evolution in *clpP*. *Scientific Reports* **5**: 16958. doi:10.1038/srep16958.

Gitzendanner MA, **Soltis PS**, **Wong GK-S**, **Ruhfel BR**, **Soltis DE. 2018**. Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. *American Journal of Botany* **105**: 291–301. doi:10.1002/ajb2.1048.

Gonçalves DJP, **Simpson BB**, **Ortiz EM**, **Shimizu GH**, **Jansen RK. 2019**. Incongruence between gene trees and species trees and phylogenetic signal variation in plastid genes. *Molecular Phylogenetics and Evolution* **138**: 219–232. doi:10.1016/j.ympev.2019.05.022.

Gonçalves DJP, **Jansen RK**, **Ruhlman TA**, **Mandel JR. 2020**. Under the rug: abandoning persistent misconceptions that obfuscate organelle evolution. *Molecular Phylogenetics and Evolution* **151**: 106903. doi:10.1016/j.ympev.2020.106903.

Graham SW, **Lam VKY**, **Merckx VSFT. 2017**. Plastomes on the edge: the evolutionary breakdown of mycoheterotroph plastid genomes. *New Phytologist* **214**: 48–55. doi:10.1111/nph.14398.

Griffith MP, **Porter JM. 2009**. Phylogeny of Opuntioideae (Cactaceae). *International Journal of Plant Sciences* **170**: 107–116. doi:10.1086/593048.

Guisinger MM, **Kuehl JV**, **Boore JL**, **Jansen RK. 2008**. Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 18424–18429. doi:10.1073/pnas.0806759105.

Guisinger MM, **Chumley TW**, **Kuehl JV**, **Boore JL**, **Jansen RK. 2010**. Implications of the plastid genome sequence of *Typha* (Typhaceae, Poales) for understanding genome evolution in Poaceae. *Journal of Molecular Evolution* **70**: 149–166. doi:10.1007/s00239-009-9317-3.

Guisinger MM, **Kuehl JV**, **Boore JL**, **Jansen RK. 2011**. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. *Molecular Biology and Evolution* **28**: 583–600. doi:10.1093/molbev/msq229.

Haberle RC, **Fourcade HM**, **Boore JL**, **Jansen RK. 2008**. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *Journal of Molecular Evolution* **66**: 350–361. doi:10.1007/s00239-008-9086-4.

Harris ME, **Meyer G**, **Vandergon T**, **Vandergon VO. 2013**. Loss of the acetyl-CoA carboxylase (*accD*) gene in *Poales*. *Plant Molecular Biology Reporter* **31**: 21–31. doi:10.1007/s11105-012-0461-3.

Jansen RK, **Ruhlman TA. 2012**. Plastid genomes of seed plants In: **Bock R**, **Knoop V**, eds. *Advances in photosynthesis and respiration. Genomics of chloroplasts and mitochondria*. Dordrecht: Springer Netherlands, 103–126.

Hertle AP, **Haberl B**, **Bock R. 2021**. Horizontal genome transfer by cell-to-cell travel of whole organelles. *Science Advances* **7**: eabd8215. doi:10.1126/sciadv.abd8215.

Jansen RK, **Wojciechowski MF**, **Sanniyasi E**, **Lee S-B**, **Daniell H. 2008**. Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps12* and *clpP* intron losses among legumes (Leguminosae). *Molecular Phylogenetics and Evolution* **48**: 1204–1217. doi:10.1016/j.ympev.2008.06.013.

Jin D-M, **Jin J-J**, **Yi T-S. 2020***a*. Plastome structural conservation and evolution in the clusioid clade of Malpighiales. *Scientific Reports* **10**: 9091. doi:10.1038/s41598-020-66024-7.

Jin D-M, **Wicke S**, **Gan L**, **Yang J-B**, **Jin J-J**, **Yi T-S. 2020***b*. The loss of the inverted repeat in the putranjivoid clade of Malpighiales. *Frontiers in Plant Science* **11**: 942. doi:10.3389/fpls.2020.00942.

Jin J-J, **Yu W-B**, **Yang J-B**, *et al*. **2020***c*. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* **21**: 241. doi:10.1186/s13059-020-02154-5.

Jo YD, **Park J**, **Kim J**, *et al*. **2011**. Complete sequencing and comparative analyses of the pepper (*Capsicum annuum* L.) plastome revealed high frequency of tandem repeats and large insertion/deletions on pepper plastome. *Plant Cell Reports* **30**: 217–229. doi:10.1007/s00299-010-0929-2.

Johnson M. **2017**. *PhypartsPieCharts*. https://github.com/mossmatters/phyloscripts/tree/master/phypartspiecharts (8 March 2023, date last accessed).

Katoh K, **Standley DM. 2013**. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780. doi:10.1093/molbev/mst010.

Kikuchi S, **Asakura Y**, **Imai M**, *et al*. **2018**. A Ycf2-FtsHi heteromeric AAA-ATPase complex is required for chloroplast protein import. *The Plant Cell* **30**: 2677–2703. doi: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**: e0187318. doi:10.1371/journal.pone.0187318.

Köhler M, **Reginato M**, **Souza-Chies TT**, **Majure LC. 2020**. Insights into chloroplast genome evolution across Opuntioideae (Cactaceae) reveals robust yet sometimes conflicting phylogenetic topologies. *Frontiers in Plant Science* **11**: 729. doi:10.3389/fpls.2020.00729.

Köhler M, **Font F**, **Puente-Martinez R**, **Majure LC. 2021**. "That's *Opuntia*, that was!", again: a new combination for an old and enigmatic *Opuntia* s.l. (Cactaceae). *Phytotaxa* **505**: 262–274. doi:10.11646/phytotaxa.505.3.2.

Korotkova N, **Aquino D**, **Arias S**, *et al*. **2021**. Cactaceae at Caryophyllales.org – a dynamic online species-level taxonomic backbone for the family. *Willdenowia* **51**: 251–270. doi:10.3372/wi.51.51208.

Kunnimalaiyaan M, **Nielsen BL. 1997**. Fine mapping of replication origins (*ori*A and *ori*B) in *Nicotiana tabacum* chloroplast DNA. *Nucleic Acids Research* **25**: 3681–3686. doi:10.1093/nar/25.18.3681.

Lee C, **Choi I-S**, **Cardoso D**, *et al*. **2021**. The chicken or the egg? Plastome evolution and an independent loss of the inverted repeat in papilionoid legumes. *The Plant Journal* **107**: 861–875. doi:10.1111/tpj.15351.

Langmead B, **Salzberg S. 2012**. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357–359. doi:10.1038/nmeth.1923.

Lee SS, **Jeong WJ**, **Bae JM**, **Bang JW**, **Liu JR**, **Harn CH. 2004**. Characterization of the plastid-encoded carboxyltransferase subunit (accD) gene of potato. *Molecules and Cells* **17**: 422–429.

Li F-W, **Kuo L-Y**, **Pryer KM**, **Rothfels CJ. 2016**. Genes translocated into the plastid inverted repeat show decelerated substitution rates and elevated GC content. *Genome Biology and Evolution* **8**: 2452–2458. doi:10.1093/gbe/evw167.

Lilly JW, **Havey MJ**, **Jackson SA**, **Jiang J. 2001**. Cytogenomic analyses reveal the structural plasticity of the chloroplast genome in higher plants. *The Plant Cell* **13**: 245–254. doi:10.1105/tpc.13.2.245.

Lin C-S, **Chen JJW**, **Huang Y-T**, *et al*. **2015**. The location and translocation of *ndh* genes of chloroplast origin in the Orchidaceae family. *Scientific Reports* **5**: 9040. doi:10.1038/srep09040.

Lin C-S, **Chen JJW**, **Chiu C-C**, *et al*. **2017**. Concomitant loss of NDH complex-related genes within chloroplast and nuclear genomes in some orchids. *The Plant Journal* **90**: 994–1006. doi:10.1111/tpj.13525.

Liu X, **Yang H**, **Zhao J**, **Zhou B**, **Li T**, **Xiang B. 2018**. The complete chloroplast genome sequence of the folk medicinal and vegetable plant purslane (*Portulaca oleracea* L.). *The Journal of Horticultural Science and Biotechnology* **93**: 356–365. doi:10.1080/14620316.2017.1389308.

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

Majure LC, **Puente R**, **Griffith MP**, **Judd WS**, **Soltis PS**, **Soltis DE. 2012**. Phylogeny of *Opuntia* s.s. (Cactaceae): clade delineation, geographic origins, and reticulate evolution. *American Journal of Botany* **99**: 847–864. doi:10.3732/ajb.1100375.

Majure LC, **Baker MA**, **Cloud-Hughes M**, **Salywon A**, **Neubig KM. 2019**. Phylogenomics in Cactaceae: a case study using the chollas sensu lato (Cylindropuntieae, Opuntioideae) reveals a common pattern out of the Chihuahuan and Sonoran deserts. *American Journal of Botany* **106**: 1327–1345. doi:10.1002/ajb2.1364.

Maliga P. 1993. Towards plastid transformation in flowering plants. *Trends in Biotechnology* **11**: 101–107. doi:10.1016/0167-7799(93)90059-i.

Maréchal A, **Brisson N. 2010**. Recombination and the maintenance of plant organelle genome stability. *New Phytologist* **186**: 299–317. doi:10.1111/j.1469-8137.2010.03195.x.

Martin GE, **Rousseau-Gueutin M**, **Cordonnier S**, *et al*. **2014**. The first complete chloroplast genome of the Genistoid legume *Lupinus luteus*: evidence for a novel major lineage-specific rearrangement and new insights regarding plastome evolution in the legume family. *Annals of Botany* **113**: 1197–1210. doi:10.1093/aob/mcu050.

Mayer C. 2006. Phobos 3.3.11 (https://www.ruhr-uni-bochum.de/ecoevo/cm/cm_phobos.htm).

Miller MA, **Pfeiffer W**, **Schwartz T. 2010**. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *2010 Gateway Computing Environments Workshop (GCE)*: 1–8.

Mower JP, **Vickrey TL. 2018**. Structural diversity among plastid genomes of land plants. In: **Chaw S-M**, **Jansen RK**, eds. *Plastid genome evolution. Advances in botanical research*. Cambridge, MA: Academic Press, 263–292.

Ogihara Y, **Terachi T**, **Sasakuma T. 1988**. Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. *Proceedings of the National Academy of Sciences of the United States of America* **85**: 8573–8577. doi:10.1073/pnas.85.22.8573.

Oldenburg DJ, **Bendich AJ. 2004**. Most chloroplast DNA of maize seedlings in linear molecules with defined ends and branched forms. *Journal of Molecular Biology* **335**: 953–970. doi:10.1016/j.jmb.2003.11.020.

Oldenburg DJ, **Bendich AJ. 2015**. DNA maintenance in plastids and mitochondria of plants. *Frontiers in Plant Science* **6**: 883. doi:10.3389/fpls.2015.00883.

Oldenburg DJ, **Bendich AJ. 2016**. The linear plastid chromosomes of maize: terminal sequences, structures, and implications for DNA replication. *Current Genetics* **62**: 431–442. doi:10.1007/s00294-015-0548-0.

Oliver MJ, **Murdock AG**, **Mishler BD**, *et al*. **2010**. Chloroplast genome sequence of the moss *Tortula ruralis*: gene content, polymorphism, and structural arrangement relative to other green plant chloroplast genomes. *BMC Genomics* **11**: 143. doi:10.1186/1471-2164-11-143.

Oulo MA, **Yang J-X**, **Dong X**, *et al*. **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**: 979. doi:10.3390/plants9080979.

Palmer JD. 1983. Chloroplast DNA exists in two orientations. *Nature* **301**: 92–93. doi:10.1038/301092a0.

Palmer JD, **Thompson WF. 1982**. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell* **29**: 537–550. doi:10.1016/0092-8674(82)90170-2.

Palmer JD, **Jansen RK**, **Michaels HJ**, **Chase MW**, **Manhart JR. 1988**. Chloroplast DNA variation and plant phylogeny. *Annals of the Missouri Botanical Garden* **75**: 1180–1206. doi:10.2307/2399279.

Paradis E, **Schliep K. 2019**. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**: 526–528. doi:10.1093/bioinformatics/bty633.

Parins-Fukuchi C, **Stull GW**, **Smith SA. 2021**. Phylogenomic conflict coincides with rapid morphological innovation. *Proceedings of the National Academy of Sciences of the United States of America* **118**: e2023058118. doi:10.1073/pnas.2023058118.

Peredo EL, **King UM**, **Les DH. 2013**. The plastid genome of *Najas flexilis*: adaptation to submersed environments is accompanied by the complete loss of the NDH complex in an aquatic angiosperm. *PLoS One* **8**: e68591. doi:10.1371/journal.pone.0068591.

Philippe H, **Brinkmann H**, **Lavrov DV**, *et al*. **2011**. Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biology* **9**: e1000602. doi:10.1371/journal.pbio.1000602.

Philippe H, **Vienne DM de**, **Ranwez V**, **Roure B**, **Baurain D**, **Delsuc F. 2017**. Pitfalls in supermatrix phylogenomics. *European Journal of Taxonomy* **283**: 1–25. doi:10.5852/ejt.2017.283.

Portik DM, **Wiens JJ. 2021**. Do alignment and trimming methods matter for phylogenomic (UCE) analyses? *Systematic Biology* **70**: 440–462. doi:10.1093/sysbio/syaa064.

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. *Physiology and Molecular Biology of Plants: An International Journal of Functional Plant Biology* **28**: 123–137. doi:10.1007/s12298-021-01121-z.

Rabah SO, **Shrestha B**, **Hajrah NH**, *et al*. **2019**. *Passiflora* plastome sequencing reveals widespread genomic rearrangements. *Journal of Systematics and Evolution* **57**: 1–14. doi:10.1111/jse.12425.

Raubeson LA, **Jansen RK. 2005**. Chloroplast genomes of plants. In: **Henry RJ**, ed. *Plant diversity and evolution: genotypic and phenotypic variation in higher plants*. Wallingford: CABI, 45–68.

Ruhlman TA, **Jansen RK. 2018**. Aberration or analogy? The atypical plastomes of Geraniaceae. In: **Chaw S-M**, **Jansen RK**, eds. *Plastid genome evolution. Advances in Botanical Research*. Cambridge, MA: Academic Press, 223–262.

Ruhlman TA, **Jansen RK. 2021**. Plastid genomes of flowering plants: essential principles. In: **Maliga P**, ed. *Methods in molecular biology. Chloroplast biotechnology: methods and protocols*. New York, NY: Springer US, 3–47.

Ruhlman TA, **Chang W-J**, **Chen JJ**, *et al*. **2015**. NDH expression marks major transitions in plant evolution and reveals coordinate intracellular gene loss. *BMC Plant Biology* **15**: 100. doi:10.1186/s12870-015-0484-7.

Ruhlman TA, **Zhang J**, **Blazier JC**, **Sabir JSM**, **Jansen RK. 2017**. Recombination-dependent replication and gene conversion homogenize repeat sequences and diversify plastid genome structure. *American Journal of Botany* **104**: 559–572. doi:10.3732/ajb.1600453.

Sadali NM, **Sowden RG**, **Ling Q**, **Jarvis RP. 2019**. Differentiation of chromoplasts and other plastids in plants. *Plant Cell Reports* **38**: 803–818. doi:10.1007/s00299-019-02420-2.

Sanderson MJ, **Copetti D**, **Búrquez A**, *et al*. **2015**. Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): loss of the *ndh* gene suite and inverted repeat. *American Journal of Botany* **102**: 1115–1127. doi:10.3732/ajb.1500184.

Schliep KP. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**: 592–593. doi:10.1093/bioinformatics/btq706.

Schmitz-Linneweber C, **Maier RM**, **Alcaraz J-P**, **Cottet A**, **Herrmann RG**, **Mache R. 2001**. The plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization. *Plant Molecular Biology* **45**: 307–315. doi:10.1023/a:1006478403810.

Shen X-X, **Hittinger CT**, **Rokas A. 2017**. Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nature Ecology & Evolution* **1**: 0126. doi:10.1038/s41559-017-0126.

Shimodaira H, **Hasegawa M. 1999**. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* **16**: 1114–1116. doi:10.1093/oxfordjournals.molbev.a026201.

Silva SR, **Diaz YCA**, **Penha HA**, *et al*. **2016**. The chloroplast genome of *Utricularia reniformis* sheds light on the evolution of the *ndh* gene complex of terrestrial carnivorous plants from the Lentibulariaceae family. *PLoS One* **11**: e0165176. doi:10.1371/journal.pone.0165176.

da Silva GM, **de Santana Lopes A**, **Gomes Pacheco T**, *et al*. **2021**. Genetic and evolutionary analyses of plastomes of the subfamily Cactoideae (Cactaceae) indicate relaxed protein biosynthesis and tRNA import from cytosol. *Brazilian Journal of Botany* **44**: 97–116. doi:10.1007/s40415-020-00689-2.

Sinn BT, **Sedmak DD**, **Kelly LM**, **Freudenstein JV. 2018**. Total duplication of the small single copy region in the angiosperm plastome: rearrangement and inverted repeat instability in *Asarum*. *American Journal of Botany* **105**: 71–84. doi:10.1002/ajb2.1001.

Smith SA, **Moore MJ**, **Brown JW**, **Yang Y. 2015**. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evolutionary Biology* **15**: 150. doi:10.1186/s12862-015-0423-0.

Solórzano S, **Chincoya DA**, **Sanchez-Flores A**, *et al*. **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.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313. doi:10.1093/bioinformatics/btu033.

Stegemann S, **Hartmann S**, **Ruf S**, **Bock R. 2003**. High-frequency gene transfer from the chloroplast genome to the nucleus. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 8828–8833. doi:10.1073/pnas.1430924100.

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:10.1042/BCJ20190365.

Straub SCK, **Parks M**, **Weitemier K**, **Fishbein M**, **Cronn RC**, **Liston A. 2012**. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *American Journal of Botany* **99**: 349–364. doi:10.3732/ajb.1100335.

Talavera G, **Castresana J. 2007**. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* **56**: 564–577. doi:10.1080/10635150701472164.

Thode VA, **Oliveira CT**, **Loeuille B**, **Siniscalchi CM**, **Pirani JR. 2021**. Comparative analyses of *Mikania* (Asteraceae: Eupatorieae) plastomes and impact of data partitioning and inference methods on phylogenetic relationships. *Scientific Reports* **11**: 13267. doi:10.1038/s41598-021-92727-6.

Tillich M, **Lehwark P**, **Pellizzer T**, *et al*. **2017**. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* **45**: W6–W11. doi:10.1093/nar/gkx391.

Timmis JN, **Ayliffe MA**, **Huang CY**, **Martin W. 2004**. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics* **5**: 123–135. doi:10.1038/nrg1271.

de Vries J, **Archibald JM**, **Gould SB. 2017**. The carboxy terminus of YCF1 contains a motif conserved throughout >500 Myr of streptophyte evolution. *Genome Biology and Evolution* **9**: 473–479. doi:10.1093/gbe/evx013.

de Vries J, **Sousa FL**, **Bölter B**, **Soll J**, **Gould SB. 2015**. YCF1: a green TIC? *The Plant Cell* **27**: 1827–1833. doi:10.1105/tpc.114.135541.

Walker JF, **Jansen RK**, **Zanis MJ**, **Emery NC. 2015**. Sources of inversion variation in the small single copy (SSC) region of chloroplast genomes. *American Journal of Botany* **102**: 1751–1752. doi:10.3732/ajb.1500299.

Walker JF, **Walker-Hale N**, **Vargas OM**, **Larson DA**, **Stull GW. 2019**. Characterizing gene tree conflict in plastome-inferred phylogenies. *PeerJ* **7**: e7747. doi:10.7717/peerj.7747.

Walker JF, **Yang Y**, **Feng T**, *et al*. **2018**. From cacti to carnivores: improved phylotranscriptomic sampling and hierarchical homology inference provide further insight into the evolution of Caryophyllales. *American Journal of Botany* **105**: 446–462. doi:10.1002/ajb2.1069.

Wallace RS. 1995. Molecular systematic study of the Cactaceae: using chloroplast DNA variation to elucidate cactus phylogeny. *Bradleya* **13**: 1–12. doi:10.25223/brad.n13.1995.a1.

Wang W-C, **Chen S-Y**, **Zhang X-Z. 2016**. Chloroplast genome evolution in Actinidiaceae: *clpP* loss, heterogenous divergence and phylogenomic practice. *PLoS One* **11**: e0162324. doi:10.1371/journal.pone.0162324.

Wick RR, **Schultz MB**, **Zobel J**, **Holt KE. 2015**. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31**: 3350–3352. doi:10.1093/bioinformatics/btv383.

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 Molecular Biology* **76**: 273–297. doi:10.1007/s11103-011-9762-4.

Wu S, **Chen J**, **Li Y**, *et al*. **2021**. Extensive genomic rearrangements mediated by repetitive sequences in plastomes of *Medicago* and its relatives. *BMC Plant Biology* **21**: 421. doi:10.1186/s12870-021-03202-3.

Xiao T-W, **Xu Y**, **Jin L**, **Liu T-J**, **Yan H-F**, **Ge X-J. 2020**. Conflicting phylogenetic signals in plastomes of the tribe Laureae (Lauraceae). *PeerJ* **8**: e10155. doi:10.7717/peerj.10155.

Xu X, **Wang D. 2021**. Comparative chloroplast genomics of *Corydalis* species (Papaveraceae): evolutionary perspectives on their unusual large scale rearrangements. *Frontiers in Plant Science* **11**: 600354. doi:10.3389/fpls.2020.600354.

Yang X-F, **Wang Y-T**, **Chen S-T**, **Li J-K**, **Shen H-T**, **Guo F-Q. 2016**. PBR1 selectively controls biogenesis of photosynthetic complexes by modulating translation of the large chloroplast gene *Ycf1* in *Arabidopsis*. *Cell Discovery* **2**: 16003. doi:10.1038/celldisc.2016.3.

Yao G, **Jin J-J**, **Li H-T**, *et al*. **2019**. Plastid phylogenomic insights into the evolution of Caryophyllales. *Molecular Phylogenetics and Evolution* **134**: 74–86. doi:10.1016/j.ympev.2018.12.023.

Yu J, **Li J**, **Zuo Y**, *et al*. **2023**. Plastome variations reveal the distinct evolutionary scenarios of plastomes in the subfamily Cereoideae (Cactaceae). *BMC Plant Biology* **23**: 132. doi:10.1186/s12870-023-04148-4.

Zhang X, **Sun Y**, **Landis JB**, *et al*. **2020**a. Plastome phylogenomic study of Gentianeae (Gentianaceae): widespread gene tree discordance and its association with evolutionary rate heterogeneity of plastid genes. *BMC Plant Biology* **20**: 340. doi:10.1186/s12870-020-02518-w.

Zhang R, **Wang Y-H**, **Jin J-J**, *et al*. **2020**b. Exploration of plastid phylogenomic conflict yields new insights into the deep relationships of Leguminosae. *Systematic Biology* **69**: 613–622. doi:10.1093/sysbio/syaa013.

Zhong B, **Deusch O**, **Goremykin VV**, *et al*. **2011**. Systematic error in seed plant phylogenomics. *Genome Biology and Evolution* **3**: 1340–1348. doi:10.1093/gbe/evr105.

Zhu A, **Guo W**, **Gupta S**, **Fan W**, **Mower JP. 2016**. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytologist* **209**: 1747–1756. doi:10.1111/nph.13743.