Planta (2021) 254:44
https://doi.org/10.1007/s00425-021-03690-5

SHORT COMMUNICATION

The genome of a thorny species: comparative genomic analysis among South and North American Cactaceae

Danilo Trabuco Amaral[1,2] · Juliana Rodrigues Bombonato[1,2] · Sónia Cristina da Silva Andrade[3] · Evandro Marsola Moraes[1] · Fernando Faria Franco[1]

Received: 5 May 2021 / Accepted: 21 July 2021 / Published online: 6 August 2021

Abstract

Main conclusion **The first South American cactus nuclear genome assembly associated with comparative genomic analyses provides insights into nuclear and plastid genomic features, such as size, transposable elements, and metabolic processes related to cactus development.**

Abstract Here, we assembled the partial genome, plastome, and transcriptome of *Cereus fernambucensis* (Cereeae, Cactaceae), a representative species of the South American core Cactoideae. We accessed other genomes and transcriptomes available for cactus species to compare the heterozygosity level, genome size, transposable elements, orthologous genes, and plastome structure. These estimates were obtained from the literature or using the same pipeline adopted for *C. fermabucensis*. In addition to the *C. fernambucensis* plastome, we also performed de novo plastome assembly of *Pachycereus pringlei, Stenocereus thurberi*, and *Pereskia humboldtii* based on the sequences available in public databases. We estimated a genome size of ~ 1.58 Gb for *C. fernambucensis*, the largest genome among the compared species. The genome heterozygosity was 0.88% in *C. fernambucensis* but ranged from 0.36 (*Carnegiea gigantea*) to 17.4% (*Lophocereus schottii*) in the other taxa. The genome lengths of the studied cacti are constituted by a high amount of transposable elements, ranging from ~ 57 to ~ 67%. Putative satellite DNAs are present in all species, excepting *C. gigantea*. The plastome of *C. fernambucensis* was ~ 104 kb, showing events of translocation, inversion, and gene loss. We observed a low number of shared unique orthologs, which may suggest gene duplication events and the simultaneous expression of paralogous genes. We recovered 37 genes that have undergone positive selection along the *Cereus* branch that are associated with different metabolic processes, such as improving photosynthesis during drought stress and nutrient absorption, which may be related to the adaptation to xeric areas of the Neotropics.

Keywords *Cereus* · Nuclear genome · Plastome · Positive selection · Repetitive elements · Transcriptome

Introduction

Whole-genome data are a powerful resource for evolutionary studies in model and non-model plant species, providing information about the processes underlying species diversification, genomic organization, recombination/duplication, etc. (Li and Harkess 2018). Furthermore, comparative analyses benefit from the increasing collections available in public databases, allowing us to explore the genome sequence with unprecedented detail. For instance, new information about nuclear, chloroplastidial, and mitochondrial genome features may provide a structural relationship associated with the complexity of genomic evolution (Szövényi et al. 2021). Thus, a relevant finding for learning about the colonization

Communicated by Dorothea Bartels.

✉ Fernando Faria Franco
franco@ufscar.br

[1] Department of Biology, Center for Human and Biological Sciences, Universidade Federal de São Carlos (UFSCar), Rodovia João Leme dos Santos, Km 110, SP264, Sorocaba 18052-780, Brazil

[2] Graduate Program in Comparative Biology, Faculty of Philosophy, Sciences and Languages of Ribeirão Preto, Universidade de São Paulo (USP), Ribeirão Preto, Brazil

[3] Department of Genetics and Evolutionary Biology, Instituto de Biociências, Universidade de São Paulo (USP), São Paulo, Brazil

and diversification of plant species in harsh environments is the identification of genes or genomic regions involved in adaptations to specific habitat characteristics (Liu et al. 2020; Que et al. 2021).

The family Cactaceae is one of the most conspicuous examples of flowering plant radiation in the Americas. This species-rich group presents remarkable diversity in growth forms and adaptations for survival in stressful xeric conditions (Hunt et al. 2006). The numbers of molecular-based studies are growing up to Cactaceae; six low coverage nuclear genomes were sequenced, as well as several plastomes (most of them from North American taxa), and numerous transcriptomes are available in public databases (Sanderson et al. 2015; Copetti et al. 2017; Wang et al. 2018; Majure et al. 2019; Xu et al. 2019; Köhler et al. 2020). In this regard, the giant saguaro cactus of the Sonoran Desert, *Carnegiea gigantea* (Cactoideae; Pachycereeae), is the most studied, with both a complete plastome (Sanderson et al. 2015) and nuclear genome (Copetti et al. 2017).

This study characterized the partial nuclear genome, plastome, and transcriptome of the South American cactus *Cereus fernambucensis* Lem. (Cereeae), a wide-ranging species found in xeric vegetation patches along the Brazilian Atlantic forest (Franco et al. 2017). These are the first nuclear genome and transcriptome assemblies from a South American cactus, which represents an important source of information for evolutionary, structural, and bioprospection studies. We compare the heterozygosity, genome size, repetitive DNA content, and plastome structure in our *C. fernambucensis* to those available in public databases. Finally, we investigated the presence of positive selection in the *Cereus* branch by the inclusion of orthologs available for North American cactus species.

Materials and methods

Sampling and sequencing

An individual of *C. fernambucensis* from Porto de Galinhas, Brazil (− 8.42 S, − 34.98 W, voucher SORO2672), was used for the whole-genome and transcriptome sequencing. DNA was extracted from the roots using a DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) and sequenced on an Illumina HiSeq2500. RNA extraction and cDNA library construction from the cladodes epidermis were performed at the BGI Americas facilities (San Jose, CA, USA) using the DNBSeq protocol. In both cases, a short-read (150 bp) paired-end library construction was used. We generated ~ 220 million and ~ 95 million reads for the genomic and transcriptomic datasets, respectively. Access numbers of genome/transcriptome species data are shown in Table S1.

Nuclear genome size estimation, de novo assembly, gene prediction, and annotation

The sequenced paired-end genomic reads were filtered with SeqyClean v.1.10.09 (Zhbannikov et al. 2017). We estimated the genome size with Jellyfish2 v.2.2.3 (Marcaiz and Kingsford 2011), heterozygosity with GenomeScope2 (Vurture et al. 2017), and ploidy with Smudgeplots v.0.1.3. (Ranallo-Benavidez et al. 2020). The *de novo* assembly was conducted in Velvet v.1.1 (Zerbino and Birney 2008) with the default settings (for Velvet assembly details see Table S2). Gene prediction was performed using Maker v.3.01.03 (Cantarel et al. 2008) in two rounds of annotation. In the first one, we conducted an ab initio gene prediction, using the AUGUSTUS v.2.5.5 (Stanke and Waack 2003), followed by a filtering step to recover the coding sequences using TransDecoder v.5.5.0 (Grabherr et al. 2011). In the second round, we included predicted proteins from the ab initio prediction and the *C. fernambucensis* transcript (generated in this study) for the AUGUSTUS. The annotation was conducted using Blastx (Altschul et al. 1997) against the Viridiplantae database (retrieved on 04/04/2019), recovering the five best hits for each gene. We estimated the genome completeness of all genome (and transcriptome) assemblies using BUSCO v.4.0.1 (Simão et al. 2015).

Plastome assembly and annotation

The plastome assembly was carried out using the GetOrganelles software (Jin et al. 2020) with the default settings. We also assembled the plastome of *Pachycereus pringlei*, *Stenocereus thurberi*, and *Pereskia humboldtii* using sequences from Copetti et al. (2017; Table 1) and using their scaffolds as reference genomes to improve *Cereus* assembly. The annotations were performed in GeSeq (Tillich et al. 2017) and CPGAVAS2 (Shi et al. 2019).

Repetitive DNA analysis

Repetitive DNA elements were identified/annotated using RepeatModeler v.1.0.8, RepeatMasker v.4.0.9 (http://www.repeatmasker.org), and RepeatExplorer (Novák et al. 2013) software in default settings. RepeatModeler and RepeatMasker were also used to identify single-sequence repetitions, and RepeatExplorer was used to detect putative satellite DNA among Cactaceae.

De novo transcriptome assembly and functional annotation

The RNA-Seq reads were filtered also using SeqyClean v.1.10.09 and de novo assembled using Trinity v.2.11.0 (Grabherr et al. 2011). The transcripts were converted to the coding sequence and translated to the amino acid by TransDecoder. Annotation was performed using Blastp against the SWISS-PROT database, recovering the five best hits for each gene. Gene ontology (GO) categories were assigned in Blast2GO v.5.2.5 (Conesa et al. 2005) and plotted using WEGO2.0 (Ye et al. 2018).

Orthologous protein clustering and tests of genes under positive selection

Raw genomic and transcriptomic amino acid sequences (Table S1) were clustered in orthogroups with OrthoFinder v.2.0.9 (Emms and Kelly 2019). The 74 orthologous transcripts shared across Cereeae + *Eryosices* (outgroup) were concatenated (supergene) and used for phylogenetic reconstruction in IQ-TREE v2.0.3, using ultrafast bootstrap (Nguyen et al. 2015), which inferred the substitution model to the supergene (LG + G4 + I + F; Fig. S1). We used the concatenated gene approach to produce an ultra-metric tree, which was used as the input tree in the test of positive selection. We investigated signatures of positive selection with the PAMLv4.8 program (Yang 2007) using the branch model (M2) and the site models M7 and M8 to calculate ω within *C. fernambucensis*, and tested whether positive selection affected specific residues in each gene. We also evaluated the null model (M0) for the entire tree using the likelihood ratio test (LRT) and the Bayes empirical Bayes (BEB) methods. False positives were checked using LRTs and multiple tests (FDR of 5%) based on the script proposed by Lee et al. (2017).

Results and discussion

Sequencing and genome size estimation

Our estimates agree with a diploid genome for *C. fernambucensis* (Fig. S2), with a size of ~ 1.58 Gb, ~ 38.0% GC, and 0.88% heterozygosity (Table 1). The calculated genome size fits within the range estimated for other *Cereus* species using flow cytometry (range from 1.87 Gb in *C. hexagonus* to 2.01 Gb in *C. jamacaru*; Silva 2015) and in silico estimation of other cacti (range from 0.98 Gb in *Pereskia humboldtii* to 1.47 Gb in *Lophocereus schottii*; Copetti et al. 2017). The genomes studied range in heterozygosity from ~ 0.3% (*C. gigantea*) to ~ 17% (*L. schottii*) and display a high amount of transposable elements (> 50%; Table 1).

Genome and transcriptome assembly and annotation

A total of 13,545 genes were predicted in the *C. fernambucensis* genome. Genome completeness as measured by BUSCO was only ~ 56%, which was similar to other public cactus genomes, with the exception of *C. gigantea* and *P. humboldtii* (Table 1). RNA-Seq analysis yielded 186,099 putative transcript products after assembly, showing the completeness of ~ 98%, 92,661 (~ 50%) of which were annotated (Fig. S3).

The *C. fernambucensis* plastome showed gene loss, translocation, and inversion (Fig. 1a), as well as the smallest plastome (104,273 bp) among the Cactoideae members, analyzed here. We identified in the *C. fernambucensis* plastome the presence of the *ndhJ*, *ndhD*, *ndhH*, and *ndhB* genes (the last one in a possible pseudogenization process in *C. gigantea*, see Sanderson et al. 2015), and two regions displayed inversion followed by translocation (between *matk/psbM* and *rpoC1/petA*) when compared to the *Opuntia quimilo* plastomes. We were not able to annotate the two inverted repeat (IR) regions, indicating possible problems in plastome assembly and/or increased plastome complexity in *Cereus*, as suggested for other cacti (Köhler et al. 2020).

Table 1 Comparison of genome features among Cactaceae species. TE, transposable element

	C. fernambucensis	*C. gigantea*	*L. schottii*	*P. pringlei*	*S. thurberi*	*P. humboldti*
In silico diploid genome size estimated (Gb)	1.58	1.3*	1.47*	1.41*	1.42*	0.98*
Genome completeness as measured in BUSCO[#] (%)	~ 56	~ 75	~ 58	~ 43	~ 53	~ 29
Heterozygosity level (%)	0.88	0.34–0.36	0.16–17.4	3.25–12.4	3.1–12.1	3.19–3.63
TE (%)	58.43	57.6*	64.2	58.51	66.57	58.3
Genomic coverage	22.43	> 97*	24.2*	20.4*	24.09*	17.2*
Putative Satellite (high Confident number)	Yes (2)	No	Yes (2)	Yes (2)	Yes (3)	Yes (3)

*Estimated in Copetti et al. (2017); [#]these values comprise partial sequence and duplicated genes

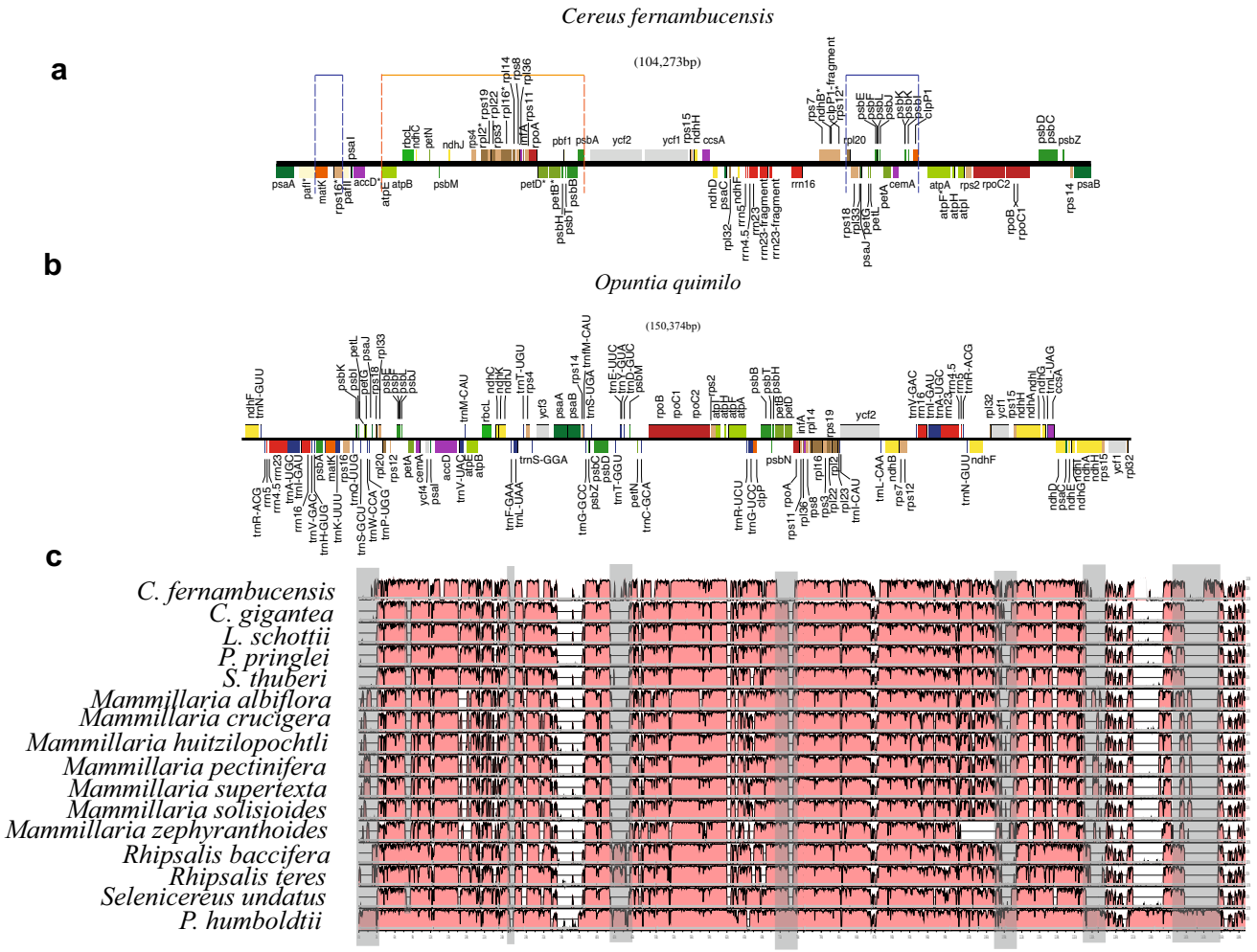

Fig. 1 Plastid genome structure (linear) of *Cereus fernambucensis* (**a**) and the *Opuntia quimilo* (**b**), which was used as the reference plastome. The orange and blue lines represent structural rearrangements (inversion and translocation regions, respectively). **c** Alignment and coverage mapping among Cactaceae plastome regions. The gray boxes are the regions with putative losses and/or pseudogenization relative to *O. quimilo* plastome. The inverted regions (IRA and IRB) in the *C. fernambucensis* plastome were not annotated and are not described in the figure

We report the complete loss of the *ndh* gene family in the *P. pringlei* and *S. thurberi* genomes, as in *C. gigantea* (Sanderson et al. 2015), while in *P. humboldtii,* all copies of *ndh* were observed. The putative circular plastome structures of these recent assemblies are available in Fig. S4.

The plastome mapping comparison among the investigated cactus species showed regions of putative losses and pseudogenization, which are proposed by Sanderson et al. (2015), Majure et al. (2019), and Köhler et al. (2020) (Fig. 1b, c), such as *accD*, *ycf1*, and *ycf2*. Majure et al. (2019) observed the loss of *ndhJ* and *rpl33* and the absence of IR regions in *C. bigelovii*; Sanderson et al. (2015) identified the absence of IR regions in *C. gigantea*, while Köhler et al. (2020) described the loss of the *ndh* gene suite and events of translocation. Altogether, these results indicate that the plastome may have experienced very active structural evolution in Cactaceae, involving changes in genome size and synteny and gene loss events in *C. fernambucensis* when compared to the *O. quimilo* plastomes.

Orthogroups were obtained for the genomic and transcriptomic data (Fig. 2). The number of orthogroups containing shared genes among the genome dataset (11,527) was higher than that observed among the transcriptome dataset (3053). The difference was more accentuated for the orthogroups that contained a single copy shared among them (2859 in the genome and 7 in the transcriptome), which may be explained by the unbalanced amount of transcriptomic and genomic data available for the comparative analyses and/or by the multiple transcripts and isoforms generated for each gene, even single-copy gene that is quite common

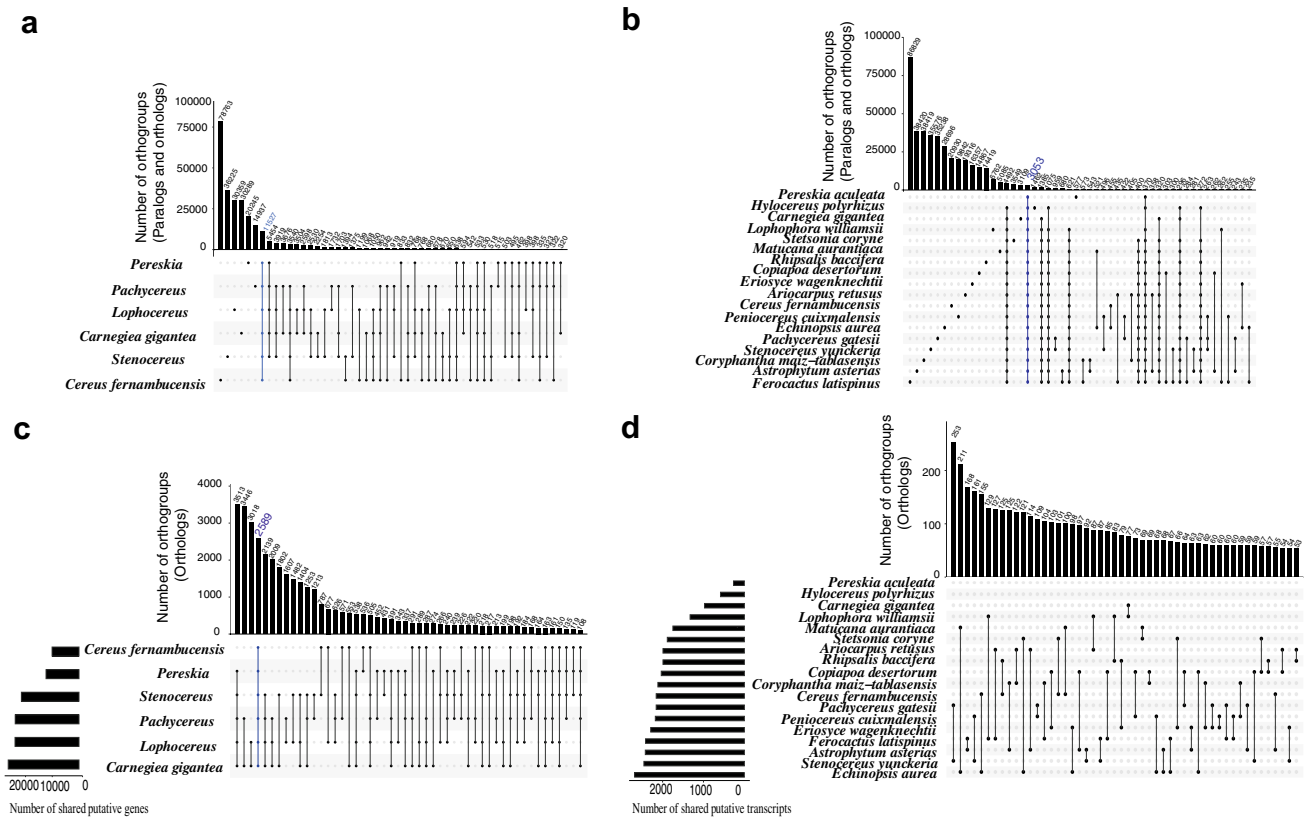

Fig. 2 Plot graph shared and unique orthogroups across genome and transcriptome data of Cactaceae species using the *UpsetR* package in R (Conway et al. 2017). The vertical bars represent the numbers of orthogroups, while the horizontal bars represent the numbers of orthogroups per species. The dot or dots linked under the histograms represent the intersections among species. Plots **a** and **c** display the number of orthogroups that contain multiple copies per species, in the genome and transcriptome datasets, respectively. Plots **b** and **d** show the number of orthogroups that contain a single copy per species, in the genome and in transcriptome datasets, respectively. Shared orthogroups across all species are highlighted in blue. Due to the enormous possibilities of combinations, we restricted the plots to the first 50 intersections

among transcriptome assembly. However, given that eukaryotic genomes are highly dynamic, including events of the gene or even whole-genome duplication (Peer et al. 2009), we cannot reject that the main difference in the number of single-copy orthogroups could be associated with gene duplication events at the genome level or with the expression of paralogous genes and isoforms at the transcriptome level.

Repetitive sequence annotation

Repetitive sequence and transposable element (TE) evaluations in Cactaceae remain underexplored. For *C. fernambucensis*, we obtained a total of 1,790,192 elements of repetitive DNA, representing ~ 55% of the total assembly. The numbers of TEs in the other studied cacti were similarly high, ranging from ~ 55 to ~ 70% (Table 1). The majority of these sequences were classified as interspersed repeats of Class I (retrotransposons; Table S3). The percentages among the types of elements were similar among the studied species, with the exception of *C. gigantea*, which displayed a low number of unclassified elements; this result was probably due to the completeness of the genome assembly. Despite a large number of TEs in Cactaceae genomes (more than 55%), we did not observe a direct association between the number of repetitive elements and genome size in the studied species (Table S3).

RepeatExplorer analyses recovered two putative tandemly arranged repetitive elements with a high probability of being satellite DNA families in *C. fernambucensis*, *L. schottii*, and *P. pringlei* and three of these elements in *S. thurberi* and *P. humboldti*. We did not recover any high confidence satellite DNA (only low confidence; data not shown) for *C. gigantea*. The complete details of the putative satDNAs, including the consensus length, are available in Table S4. To the best of

our knowledge, this is the first report of satellite DNA in Cactaceae.

Positive selection of single-copy genes

We found evidence that 37 orthologs were under positive selection along the *Cereus* branch (Table S5), of which 12 were annotated and associated with possible physiological responses to drought stress in plants. Among them, we identified *Light-harvesting chlorophyll a/b-binding* (*Lhc*) and *Maintenance of PSII under high light 1* (*MPH1*), which are genes associated with increased tolerance to drought stress (Semedo et al. 2020; Zhao et al. 2020). LHC is regulated by phosphorylation-driven state transitions, increasing the responsiveness of stomata and thereby reducing evapotranspiration (Zhao et al. 2020). MPH1 acts as a photo-inhibitor under high radiation, preventing photo-oxidative damage (Semedo et al. 2020). Both enzymes present critical functions during photosynthetic activity and homeostasis.

The other annotated genes under selection are implicated in stress response, hormone regulation, and development: *pectinesterase inhibitors*, *dirigent protein*, *calcium-dependent kinase*, and *receptor kinase protein (ZAR1)*. In particular, *pectinesterase* and *ZAR1* are influenced by physiological stress and play important roles in cell wall development during fruit ripening, pollen tube growth, and root development (Wormit and Usadel 2018; Atif et al. 2019). One group of genes (*Ring-H2 finger*, *E3 ubiquitin-protein ligase*, and *phospholipase D zeta 1*) is associated with phosphorus supply mechanisms in phosphate-starved soils (Pan et al. 2019). The first two genes contribute to the regulation of protein ubiquitination, playing a central role in the regulation of the phosphate starvation response and phosphate acquisition; *phospholipase D zeta* is a regulated root hair morphogenesis protein that contributes to the acquisition of inorganic phosphorus within the galacto-lipid pathway and is also involved in root elongation in phosphate starvation periods (Wang et al. 2020). It is possible that the adaptability to xeric South American biomes may have been facilitated by the ability of this cactus lineage to increase the assimilation of phosphorus and other metabolites from the soil.

Conclusion

We performed comparative analyses, evaluating the genomic features and shared orthologs among Cactaceae. We report the sequencing of the *C. fernambucensis* partial nuclear genome, plastome, and transcriptome, the first for a South American columnar cactus clade. We provide insights into the repetitive elements, gene duplication, and transcripts within Cactaceae; gene loss and translocation in the Cactaceae plastome; and putative genes under selection in *Cereus*. The results presented here and the availability of a new cactus genome, plastome, and transcriptome provide resources for future studies of Neotropical cacti and related lineages, which is an important clade for comparative genomics, demographic, and evolutionary studies.

Author contribution statement FFF and DTA conceived the idea; DTA led both the analyses and manuscript writing; JRB, DTA, and SCSA performed data collection and comparative genomic analyses; EMM helped with the initial conceptions of this study and writing. FFF and EMM provided funding resources. All authors contributed to the intellectual development of the paper, made multiple revisions, and approved the final draft.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00425-021-03690-5.

Acknowledgements We thank the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; 2014/25227-0 and 2018/03428-5), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001 (fellowship to D.T.A. and J.R.B.) for financial support. Additional funds were provided to EMM by the National Council for Scientific and Technological Development (CNPq; 303940/2019-0). We thank Luiz Henrique M. Fonseca (USP) for help in preliminary analyses; Gustavo C. S. Kuhn for help in the interpretation of the RepeatExplorer results; and Deren Eaton (Columbia University) and Luciano Digiampietri (USP) for server resources.

Data availability The datasets generated during and/or analyzed during the current study are available in the NCBI SRA database under project number PRJNA587492.

Declarations

Conflict of interest The authors declare no competing interest.

References

Altschul SF, Madden TL, Schäffer AA et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Atif RM, Shahid L, Waqas M et al (2019) Insights on calcium-dependent protein kinases (CPKs) signaling for abiotic stress tolerance in plants. Int J Mol Sci 20:5298

Cantarel BL, Korf I, Robb SM et al (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res 18:188–196

Conesa A, Gotz S, García-Gómez JM et al (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676

Conway JR, Lex A, Gehlenborg N (2017) UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics 33:2938–2940

Copetti D, Búrquez A, Bustamante E et al (2017) Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. Proc Natl Acad Sci USA 114:12003–12008
Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20:1–14
Franco FF, Silva GR, Moraes EM (2017) Plio-Pleistocene diversification of *Cereus* (Cactaceae, Cereeae) and closely allied genera. J Linn Soc Bot 183:199–210
Grabherr MG, Haas BJ, Yassour M et al (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat Biotechnol 29:644–652
Hunt D, Taylor N, Charles G (2006) The new cactus lexicon, atlas & text. DH Books, Milborne Port
Jin JJ, Yu WB, Yang JB, Song Y, Depamphilis CW, Yi TS, Li DZ (2020) GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol 21:1–31
Köhler M, Reginato M, Souza-Chies TT, Majure LC (2020) Insights into chloroplast genome evolution across Opuntioideae (Cactaceae) reveals robust yet sometimes conflicting phylogenetic topologies. Front Plant Sci 11:729
Lee R, Wiel L, van Dam TJP, Huynen MA (2017) Genome-scale detection of positive selection in nine primates predicts human-virus evolutionary conflicts. Nucleic Acids Res 45:10634–10638
Li F, Harkess A (2018) A guide to sequence your favorite plant genomes. Appl Plant Sci 6:e1030
Liu Z, Zhang L, Yan Z et al (2020) Genomic mechanisms of physiological and morphological adaptations of limestone langurs to karst habitats. Mol Biol Evol 37(4):952–968
Majure LC, Baker MA, Cloud-Hughes M, Salywon A, Nuebug 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. Am J Bot 106:1–19
Marcaiz G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27:764–770
Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274
Novák P, Neumann P, Pech J, Steinhaisl J, Macas J (2013) RepeatExplorer: a galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics 29:792–793
Pan W, Wu Y, Xie Q (2019) Regulation of ubiquitination is central to the phosphate starvation response. Trends Plant Sci 24:755–769
Peer YV, Maere S, Meyer A (2009) The evolutionary significance of ancient genome duplications. Nat Rev Genet 10:725–732
Que T, Wang H, Yang W et al (2021) The reference genome and transcriptome of the limestone langur, *Trachypithecus leucocephalus*, reveal expansion of genes related to alkali tolerance. BMC Biol 19(1):1–15
Ranallo-Benavidez TR, Jaron KS, Schatz MC (2020) GenomeScope 2.0 and Smudgeplots: reference-free profiling of polyploid genomes. Nat Commun 11:1432
Sanderson MJ, Copetti D, Búquez A et al (2015) Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): loss of the ndh gene suite and inverted repeat. Am J Bot 102:1115–1127
Semedo JN, Rodrigues AP, Lidon FC et al (2020) Intrinsic non-stomatal resilience to drought of the photosynthetic apparatus in *Coffea* spp. is strengthened by elevated air [CO2]. Tree Physiol 41(5):708–727
Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C (2019) CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res 47:W65–W73
Silva RGN (2015) Genome size of invasive and non-invasive succulent species. PhD Thesis, University of Coimbra, Portugal
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212
Stanke M, Waack S (2003) Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19(suppl_2):ii215–ii225
Szövényi P, Gunadi A, Li FW (2021) Charting the genomic landscape of seed-free plants. Nat Plants 7:554–565
Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S (2017) GeSeq—versatile and accurate annotation of organelle genomes. Nucleic Acids Res 45:W6–W11
Vurture G, Sedlazeck FJ, Nattestad M et al (2017) GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33:2202–2204
Wang N, Yang Y, Moore MJ et al (2018) Evolution of Portulacineae marked by gene tree conflict and gene family expansion associated with adaptation to harsh environments. Mol Biol Evol 36:112–126
Wang X, Bi S, Wang L et al (2020) GLABRA2 regulates actin bundling protein VILLIN1 in root hair growth in response to osmotic stress. Plant Physiol 184:176–193
Wormit A, Usadel B (2018) The multifaceted role of pectin methylesterase inhibitors (PMEIs). Int J Mol Sci 19:2878
Xu M, Liu CL, Luo J et al (2019) Transcriptomic de novo analysis of pitaya (*Hylocereus polyrhizus*) canker disease caused by *Neoscytalidium dimidiatum*. BMC Genomics 20:10
Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24(8):1586–1591
Ye J, Zhang Y, Cui H et al (2018) WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. Nucleic Acids Res 46:W71–W75
Zerbino DR, Birney E (2008) Velvet: algorithm for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829
Zhao S, Gao H, Luo J et al (2020) Genome-wide analysis of the light-harvesting chlorophyll *a/b*-binding gene family in apple (*Malus domestica*) and functional characterization of *MdLhcb4. 3*, which confers tolerance to drought and osmotic stress. Plant Physiol Biochem 154:517–529
Zhbannikov I, Hunter S, Foster J, Settles M (2017) SeqyClean: A pipeline for high throughput sequence data preprocessing. In: Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics. Boston, MA, 20–24 Aug. 2017. Association for Computer Machinery, New York, pp 407–416