BMC Genomics

RESEARCH

Open Access

Comparative analysis of chloroplast genomes and phylogenetic relationships of different pitaya cultivars

Enting Zheng[1,2†], Gulbar Yisilam[1,2,3†], Chuanning Li[1,2], Fangfang Jiao[1,2,3], Yulan Ling[1,2], Shuhua Lu[4], Qiuyan Wang[1,2*] and Xinmin Tian[1,2*]

Abstract

Background Pitaya is an important tropical fruit highly favoured by consumers owing to its good and juicy characteristics. It contains a large amount of betacyanin, which is a natural food-colouring agent, in the peel and pulp. However, few studies have focused on the pitaya chloroplast (cp) genomes.

Results To explore the genetic differences and phylogenetic relationships among the cp genomes of the six pitaya cultivars, we assembled, annotated, and performed a comparative genomic analysis. The cp genomes of the six cultivars exhibited a typical circular structure, ranging in length from 133,146 to 133,617 bp, with a GC content of 36.4%. All individual cp genomes were annotated with 123 genes, including 80 protein-coding genes, 38 tRNA genes, four rRNA genes, and one pseudogene (*ycf*68). Six mutated hotspot regions (*trn*F-GAA-*rbc*L, *trn*M-CAU-*acc*D, *rpl*20-*psb*B, *acc*D, *rpl*22, *ycf*1) were detected, which could be considered potential molecular markers for population genetics and molecular phylogeny studies. Phylogenetic analysis showed that pitaya cultivars clustered into a single branch in the phylogenetic tree of the Cactaceae family. Furthermore, the observed phylogenetic patterns suggest a complex genetic basis for colour variation among pitaya cultivars.

Conclusions The study findings expand our understanding of the cp genome of pitaya and the phylogenetic relationships among different cultivars. The genomic data obtained provide important information for the breeding and genetic improvement of pitaya.

Keywords Pitaya, Chloroplast genome, Structural characteristics, Phylogenetic analysis

†Enting Zheng and Gulbar Yisilam contributed equally to this study.

*Correspondence:
Qiuyan Wang
qiuyanw2015@163.com
Xinmin Tian
tianxm333333@foxmail.com
[1] Key Laboratory of Ecology of Rare and Endangered Species and Environmental Protection (Ministry of Education) & Guangxi Key Laboratory of Landscape Resources Conservation and Sustainable Utilization in Lijiang River Basin, Guangxi Normal University, Guilin 541006, China
[2] Guangxi University Engineering Research Center of Bioinformation and Genetic Improvement of Specialty Crops, Guangxi Normal University, Guilin 541006, China
[3] Xinjiang Key Laboratory of Biological Resources and Genetic Engineering, College of Life Science and Technology, Xinjiang University, Urumqi 830046, China
[4] Guangxi Institute of Botany, Guangxi Zhuang Autonomous Region and Chinese Academy of Sciences, Guilin 541006, China

Introduction

Pitaya is the fruit of a class of climbing plants belonging to the genera *Selenicereus* and *Hylocereus* in the family Cactaceae. These plants originated from tropical and subtropical Central America [1]. Recently, owing to its rich nutritional value and pharmacological effects, the worldwide cultivation of pitaya has gradually spread from the Americas to tropical and subtropical countries such as Vietnam and China [2–4]. Based on the colour of the fruit peel and pulp, pitaya can be classified into the following three main types: *Selenicereus monacanthus* with red peel and pulp, *Selenicereus undatus* with red peel and white pulp, and *Selenicereus megalanthus* with yellow peel and white pulp [5]. Pitaya is rich in a variety of nutrients such as betaine, polyphenols, flavonoids, and anthocyanins [6, 7], which are important for the treatment of a variety of diseases such as diabetes, cardiovascular disease, and cancer [8–10]. Furthermore, pitaya peel has great potential in the food industry, such as food packaging and coatings [11]. Current research on pitaya focuses on cultivation techniques [12, 13], nutrient composition [6, 14], pests and diseases [15, 16], and medical value [9, 10]. Despite the potential economic value and health benefits of pitaya, chloroplast (cp) genomics research on pitaya still lags behind that of some traditional fruits. Currently, plant breeding is gradually moving toward the 4.0 era, and the addition of big data and artificial intelligence will provide more efficient and accurate methods for plant breeding [17, 18]. Adequate genomic data is the important foundation for realising "smart breeding", so we still need to obtain more abundant genomics data including cp genomes to form a more systematic database [19].

The species classification of the genera *Hylocereus* and *Selenicereus* is controversial. Britton and Rose separated these two genera based on morphological differences [20]. However, because most fruits of plants in both genera are edible, natural hybridisation between the genera occurs. Some hybridised individuals possess characteristics of both the genera *Hylocereus* and *Selenicereus*, presenting significant taxonomic difficulties; therefore, researchers have suggested merging these two genera [21]. Korotkova et al. [22] conducted a molecular phylogenetic with four plastid region segments and supports the taxonomic treatment of transferring all species of *Hylocereus* to *Selenicereus*. The chloroplast is important organelle for photosynthesis in plants, and comparative analysis of plant cp genomes is important for the study of evolutionary relationships between relatives and species identification [23, 24]. In recent years, only two articles have been published on the complete cp genome of pitaya, revealing the cp genome sequences of five *Selenicereus* species and determining the taxonomic positions of these plants in Cactaceae [25, 26]. Adding the complete cp genomic information of different pitaya species can help expand the database of pitaya genomes and better explore the genome evolution of individuals within *Selenicereus* and their phylogenetic relationships.

In this study, we analysed the sequence structure of six pitaya cultivars cp genomes and performed comparative genomic and phylogenetic studies to investigate the genetic differences and relationships among cp genomes across various cultivars. This study's findings contribute to our understanding of chloroplast genomes and evolution within the genus *Selenicereus* and related species in Cactaceae, while providing a valuable genomic resources that could contribute to future pitaya breeding programs. The identified highly variable sites and SSR sites identified through screening could serve as molecular markers for molecular-assisted selection in cultivar development. Meanwhile, comparative genetic analysis of cultivars may enable targeted screening of superior germplasm and optimize hybridization strategies for trait improvement.

Materials and methods

Plant materials

The six different pitaya cultivars samples for the study were provided by the Guangxi Institute of Botany, namely *Selenicereus megalanthus* 'Yanwoguo' and *Selenicereus megalanthus* 'Wucihuanglong' with yellow peel white pulp, *Selenicereus monacanthus* 'Sijihong' and *Selenicereus monacanthus* 'Jingduyihao' with red peel red pulp, *Selenicereus undatus* 'Putongbairou' and *Selenicereus undatus* 'Baishuijing' with red peel white pulp. Voucher specimens (voucher numbers: GZ202302401 – GZ202302406) were identified by Shuhua Lu and deposited at the herbarium of the Guangxi Institute of Botany, Guilin, China.

DNA extraction and sequencing

Total genomic DNA was extracted from the stems of fresh plants. After DNA extraction, samples from the six pitaya cultivars were sent to the Anhui Double Helix Gene Technology Company (Anhui, China) for genomic library construction and Illumina sequencing. Illumina paired-end libraries (150 bp read length) were generated in a single lane on an Illumina HiSeq2500 (2500 Illumina Way, San Diego, USA), and the paired-end raw reads were processed using Trimmomatic v0.39 [27] to remove adapters and low-quality reads to produce high-quality clean data.

Chloroplast assembly and annotation

The genome assembly and annotation processes were conducted using well-established bioinformatics tools, ensuring reliable and high-quality results for all six pitaya

cultivars. Red peel white pulp pitaya *Selenicereus undatus* (Haw.) D. R. Hunt (GB: NC.053698) was downloaded from the National Center for Biotechnology Information (NCBI: https://www.ncbi.nlm.nih.gov/) as a reference sequence, and high-quality resequencing data obtained by screening were assembled using GetOrganelle v1.7.5 [28] with the parameters set to -R 15 -k 21, 45, 65, 85, 105, 127 -F embplant_pt. The completed assembled sequences were preliminarily annotated using Geseq [29] and CPGAVAS2 [30]. The annotation results were manually corrected using Geneious Prime v2024.0.5 [31] to obtain the complete cp genome. Finally, the cp genomes of these six pitaya cultivars were circularly mapped using OGDRAW v1.3.1 (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) [32]. All fully annotated cp genome sequences were uploaded to the NCBI GenBank database under the accession numbers PQ824054 – PQ824059.

Structural characterization of the chloroplast genome

The total length and guanine-cytosine (GC) content of the cp genome, large single-copy region (LSC), inverted repeat regions (IRs) and small single-copy region (SSC), and gene composition were analysed using Geneious Prime v2024.0.5.

Repeat sequence analysis

Online software MISA-web (https://webblastipk-gatersleben.de/misa/.) was used to identify simple sequence repeats (SSRs) in the target sequences [33], and the thresholds for mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeat sequences were set to 10, 5, 4, 3, 3, and 3, respectively.

Dispersed repetitive sequences of the cp genome, including forward, reverse, complementary, and palindromic repeats, were identified using REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer/) [34]. The parameters were set as follows: minimum sequence length was 30 bp; Hamming distance was 3, maximum number of computed repeats was 5000 bp. Finally, the number of dispersed repeat sequence types and the distribution of cp genomes were obtained.

Analysis of IRs contraction and expansion

The boundaries of the LSC, IRb, SSC, and IRa regions among six pitaya cultivars cp genomes were visualised using the CPJSdraw tool [35] to analyse the contraction and expansion of four regions in different pitaya cultivars as well as the correlation between the boundary regions and the genes.

Genomic variation analysis

To explore the genomic sequence variation in the cp genome of pitaya, the mVISTA program (http://genome.lbl.gov/vista/index.shtml) was used to analyse the genomic variation of different pitaya cultivars using *Selenicereus megalanthus* (GB: NC.087625.1) as a reference in the Shuffle-LAGAN mode for sequence similarity comparison of cp genomes of different pitaya cultivars [36].

Nucleotide diversity analysis

To assess the nucleotide diversity (Pi) among the cp genomes of the six pitaya cultivars, the sequences of the six cp genomes were aligned using the MAFFT function in Geneious Prime v2024.0.5. The genome-wide nucleotide diversity was subsequently computed using a sliding window of DnaSP v6.12.03, with a window length of 600 bp and a step size of 200 bp [37].

Analysis of the codon usage bias

The coding sequence (CDS) genes in the cp genomes of six pitaya cultivars were extracted using PhyloSuite v1.2.3 software [38]. According to a previous research method [39], one of the genes with duplicates or gene < 300 bp in length were eliminated, and genes with ATG as the start codon and TAA\TAG\TGA as the stop codon were selected. Finally, 45 gene sequences were used for subsequent codon usage bias analysis, and these 45 CDSs were integrated into one sequence for relative synonymous codon usage (RSCU) analysis. The codon usage frequency and RSCU values of different pitaya cultivars were calculated using CodonW v1.4.2 [40]. Finally, the results were visualised using TBtools v2.154 [41].

Phylogenetic analysis

To construct a phylogenetic tree of Cactaceae, we downloaded 40 cp genomes of Cactaceae plants from the NCBI, together with six fully assembled pitaya from this study and one outgroup of Portulacaceae (*Portulaca oleracea*, GB: NC.036236). The 45 shared CDSs from these 47 species were selected to construct the maximum likelihood (ML) and Bayesian inference (BI) trees. Phylogenetic analysis of the genus *Selenicereus* was based on the six pitaya cultivars used in the current study and the four complete cp genomes of pitaya available in the NCBI database. Two data matrices (complete cp genomes and 63 shared CDSs) were selected for ML and BI analyses, with *Carnegiea gigantea* (GB: NC.027618) as the outgroup. The GenBank numbers and classifications of the cp genome sequences downloaded from the NCBI database are shown in Table S1.

Shared single-copy CDSs were identified and extracted using PhyloSuite v1.2.3. Before constructing the phylogenetic tree, both the extracted shared CDSs and the complete cp genome sequences were aligned using MAFFT v7.505 [42] in PhyloSuite v1.2.3. These

aligned nucleotide sequences were rejoined, and unaligned sequences were clipped using Gblocks [43], and phylogenetic trees were created based on the optimised data using the ML method of IQ-TREE v2.2.2.6 [44], with the bootstrap (BS) parameter set to 1000, and Modelfinder v2.2.0 [45] to determine the best alternative model.

BI analysis was performed using MrBayes v3.2.7a Markov chain Monte Carlo method row [46], and the best alternative model was determined using Modelfinde software and run for 200,000 generations, sampling the tree every 1000 generations. The first 20% of the trees were discarded as aged, and the remaining trees were used to generate consensus trees. Finally, the phylogenetic tree was visualised using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree).

Results

Characteristics of the chloroplast genome

In this study, the complete cp genomes of six pitaya cultivars with typical tetrameric structures were assembled, including an LSC region, two IR regions (IRa and IRb), and an SSC region. Based on consistent gene content, order, and orientation, we use one cp gene map to represent all six pitaya cultivars cp genomes map (Fig. 1). The complete cp genome of pitaya ranged in size from 133,146 to 133,617 bp. The total GC content of the cp genome was 36.4% in both cases. The size of the LSC region ranged from 68,076 to 68,528 bp, with a GC content ranging from 36.2% to 36.3%. The size of the SSC region ranged from 21,716 to 21,808 bp, with a GC content ranging from 39.6% to 39.7%. The size of the IR region ranged from 21,677 to 21,806 bp, with a GC content from 34.9% to 35.0%. Among the samples examined, *S. megalanthus* 'Wucihuanglong' exhibited the largest cp genome, while *S. monacanthus* 'Jingduyihao' had the smallest (Table S2).

In each of the six pitaya cultivar genomes, we identified 123 genes, including 80 protein-coding genes (PCGs), 38 tRNA genes, four rRNA genes, and one pseudogene (*ycf*68) (Table S2). The gene structures and contents of the six pitaya cp genomes are highly conserved. The number of genes in the six pitaya cultivars was counted, and 19 duplicated genes were identified in the IR regions, including nine tRNA genes and 10 PCGs (*rps*16, *atp*A, *atp*F, *psb*A, *psb*I, *psb*K, *clp*P, *mat*K, *ycf*1, and *ycf*2). The identified genes could be categorized into four groups according to their functions: the first group was photosynthesis-related genes totalling 35 types; the second group was self-replication-related genes totalling 57 types; the third group was other genes totalling 6 types; the fourth group was unknown genes totalling 6 types (Table 1).

Analyses of simple sequence repeats and dispersed repeats

Overall, 66–69 SSRs were identified in the cp genomes of the six pitaya cultivars. *S. megalanthus* 'Yanwoguo' had the highest number of repeat sequences (69), followed by *S. undatus* 'Putongbairou' with 67. The remaining cultivars contained 66 SSRs. All six pitaya cultivars contained mono-, di-, tri-, tetra-, penta-nucleotide repeat sequences. With the exception of two red peel white pulp cultivars (*S. undatus* 'Putongbairou' and *S. undatus* 'Baishuijing') which did not have hexanucleotide repeats, hexanucleotide repeats were identified in all the remaining individuals. The identified SSR exhibited the highest number of mononucleotide repeat sequences, followed by di-, tetra-, tri-, penta-, and hexa-nucleotide repeat sequences (Fig. 2A). Mononucleotide repeat sequences consisting of A/T motifs were the most prevalent, accounting for 73.75%. This was followed by dinucleotide repeat sequences based on AT/AT motifs, which accounted for 13.5% (Fig. 2B). Further statistics revealed that, among these six pitaya plants, most of SSRs were distributed in the LSC region, followed by the IR region, and were least distributed in the SSC region (Fig. 2C).

In this study, we analysed repetitive sequences of more than 30 bp in all samples, and these repetitions appeared in a dispersed form in the genome. We found that the presence of 1,097 dispersed repeat sequences in the cp genomes of the six pitaya cultivars. The cultivar *S. undatus* 'Putongbairou' exhibited the highest number of dispersed repeat sequences, with a total of 214. In contrast, *S. monacanthus* 'Jingduyihao' had the lowest number, with 161 dispersed repeat sequences. The analysis identified four distinct categories of dispersed repeat sequences in all six pitaya cultivars: forward, reverse, palindromic, and complementary. The six cultivars under consideration had the highest number of forward repeats (96–152), and the lowest number of complementary repeats, with only one identified in each cultivar (Fig. 2D). Further statistical analysis revealed that most of the identified dispersed repeat sequences were less than 50 bp in length (98–151), followed by a range of 50–99 bp (30–43) (Fig. 2E).

IRs contraction and expansion

In this study, we compared the contraction and expansion of the IR/SC boundaries in the cp genomes of six pitaya cultivars. The analysis revealed a high degree of similarity between the six pitaya cultivars. Both the SSC and IR regions of *S. megalanthus* 'Yanwoguo' showed expansion

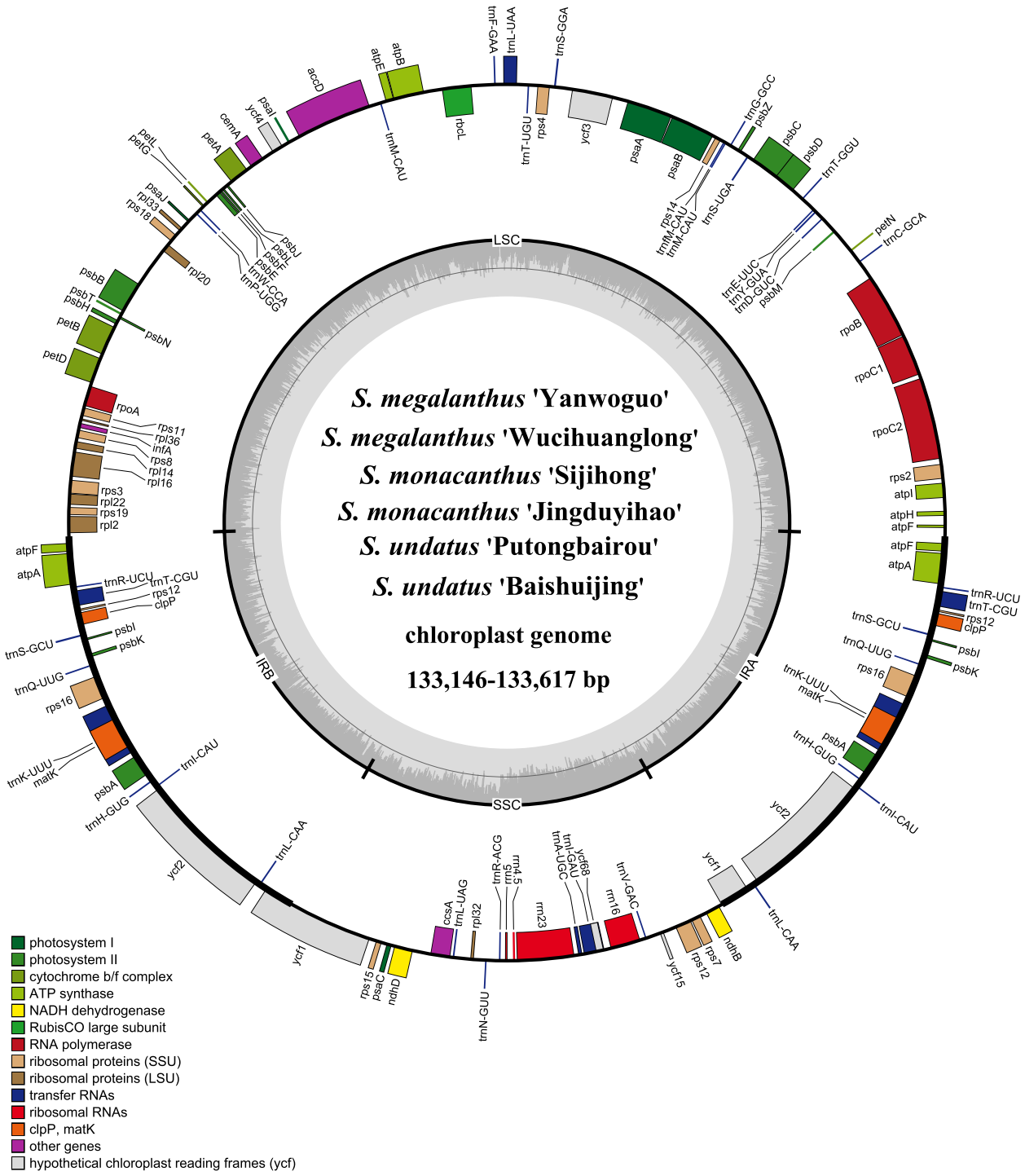

Fig. 1 Gene map of the chloroplast genome among six pitaya cultivars. The direction of gene transcription is counterclockwise on the outside of the circle and clockwise on the inside of the circle. Genes of different functional types are indicated by different colours. The dark gray colour in the inner circle corresponds to GC content

compared to the other five cultivars, with the IR region being 21,806 bp in length and SSC region being 21,808 bp in length, respectively (Fig. 3). In six cultivars, both copies of the *ycf*1 gene span the IR/SSC border regions. The *ycf*1 gene in the IRb/SSC border region extended from the IRb region to the SSC region, with extensions

Table 1 Gene composition of chloroplast genome of six pitaya cultivars

Gene function	Group of gene	Gene names	Amount
rRNA	Ribosomal RNAs	*rrn*16, *rrn*23, *rrn*4.5, *rrn*5	4
tRNA	Transfer RNAs	*trn*A-UGC, *trn*C-GCA, *trn*D-GUC, *trn*E-UUC, *trn*F-GAA, *trnf*M-CAU, *trn*G-GCC , *trn*I-GAU, *trn*L-UAA, *trn*L-UAG, *trn*N-GUU, *trn*P-UGG, *trn*R-ACG, *trn*S-GGA, *trn*S-UGA, *trn*T-GGU, *trn*T-UGU, *trn*V-GAC, *trnW-CCA*, *trn*Y-GUA, *trn*I-CAU(× 2), *trn*K-UUU(× 2), *trn*L-CAA(× 2), *trn*H-GUG(× 2), *trn*M-CAU(× 2), *trn*Q-UUG(× 2), *trn*R-UCU(× 2), *trn*S-GCU(× 2), *trn*T-CGU(× 2)	38
Self replication	DNA dependent RNA polymerase	*rpo*A, *rpo*B, *rpo*C1, *rpo*C2	4
	Small subunit of ribosome	*rps*11, *rps*12, *rps*14, *rps*15, *rps*16(× 2), *rps*18, *rps*19, *rps*2, *rps*3, *rps*4, *rps*7, *rps*8	13
	Large subunit of ribosome	*rpl*14, *rpl*16, *rpl*2, *rpl*20, *rpl*22, *rpl*32, *rpl*33, *rpl*36	8
Gene for photosynthesis	Subunits of photosystem I	*psa*A, *psa*B, *psa*C, *psa*I, *psa*J	5
	Subunits of photosystem II	*psb*A(X2), *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*H, *psb*I(X2), *psb*J, *psb*K(X2), *psb*L, *psb*M, *psb*N, *psb*T, *psb*Z	18
	Subunits of cytochrome b/f complex	*pet*A, *pet*B, *pet*D, *pet*G, *pet*L, *pet*N	6
	Subunits of ATP synthase	*atp*A(X2), *atp*B, *atp*E, *atp*F(X2), *atp*H, *atp*I	8
	Subunit of rubisco	*rbc*L	1
	Subunits of NADH-dehydrogenase	*ndh*B, *ndh*D	2
Other gene	Maturase	*mat*K(X2)	2
	Envelop membrane protein	*cem*A	1
	c-type cytochrom synthesis gene	*ccs*A	1
	Subunit of Acetyl-CoA-carboxylase	*acc*D	1
	Protease	*clp*P(X2)	2
	Translational initiation factor	*inf*A	1
Unknown	Conserved open reading frames	*ycf*1(X2), *ycf*15, *ycf*2(X2), *ycf*3, *ycf*4	7
	pseudo	*ycf*68	1
Total			123

(X2) indicates that the gene located in the IRs and thus had two complete copies

ranging from 4,076 to 4,127 bp. Simultaneously, the *ycf*1 genes in the SSC/IRa border region both have 45 bp extensions into the SSC region.

Comparative analysis of chloroplast genomes

In this study, published *Selenicereus megalanthus* (GB: NC.087625) was used as a reference, and the mVISTA online tool was used to conduct a genome-wide comparative analysis of the cp genomes of the six pitaya cultivars. These six pitaya cultivars were similar to the reference sequence in terms of gene structure and alignment order, and the variant sites were mainly found in the LSC region, followed by the SSC region, with no obvious variation in the IR regions. The genes with more significant variations in the protein-coding region were *acc*D, *rps*18, *rpl*22, *rps*19, and *ycf*1, with the highest degree of sequence variability found in the *acc*D gene. More sequence variation were found in the non-coding regions than in the protein-coding regions, such as *atp*H-*atp*I, *trn*F(GAA)-*rbc*L, *trn*M(CAU)-*acc*D, *trn*L(CAA)-*ycf*1, *ndh*D-*ccs*A, *ycf*1-*trn*L(CAA), and *trn*L(CAA)-*ycf*2 intergenic regions all showed variability (Fig. 4A).

To elucidate the level of sequence variation, Pi analysis of the six pitaya cultivars was performed in this study using DnaSP software, and the Pi values among the sequences were calculated. The results showed that the Pi values ranged from 0.00000 to 0.08511 with an average value of 0.00363, and the maximum peak appeared in the *acc*D gene (Fig. 4B). The LSC region had the highest average nucleotide diversity (Pi = 0.00481), followed by the SSC (Pi = 0.00405) and IR regions (Pi = 0.00159). The highly variable sites were mostly distributed in the LSC and SSC regions, whereas the IR region was more conserved and had a lower mutation rate, which was consistent with the results of genome-wide comparative analysis. In this study, six different highly variable sites (Pi ≥ 0.015) were screened out, namely *trn*F-GAA-*rbc*L, *trn*M-CAU-*acc*D, *acc*D, *rpl*20-*psb*B, *rpl*22, *ycf*1.

Codon usage bias analysis

In this study, six pitaya cp genomes were analysed for codon usage, frequency, and preference. A total of PCGs > 300 bp in length were encoded by 19,189 (*S. monacanthus* 'Jingduyihao') to 19,350 (*S. megalanthus* 'Wucihuanglong') codons, including stop codons. The total number

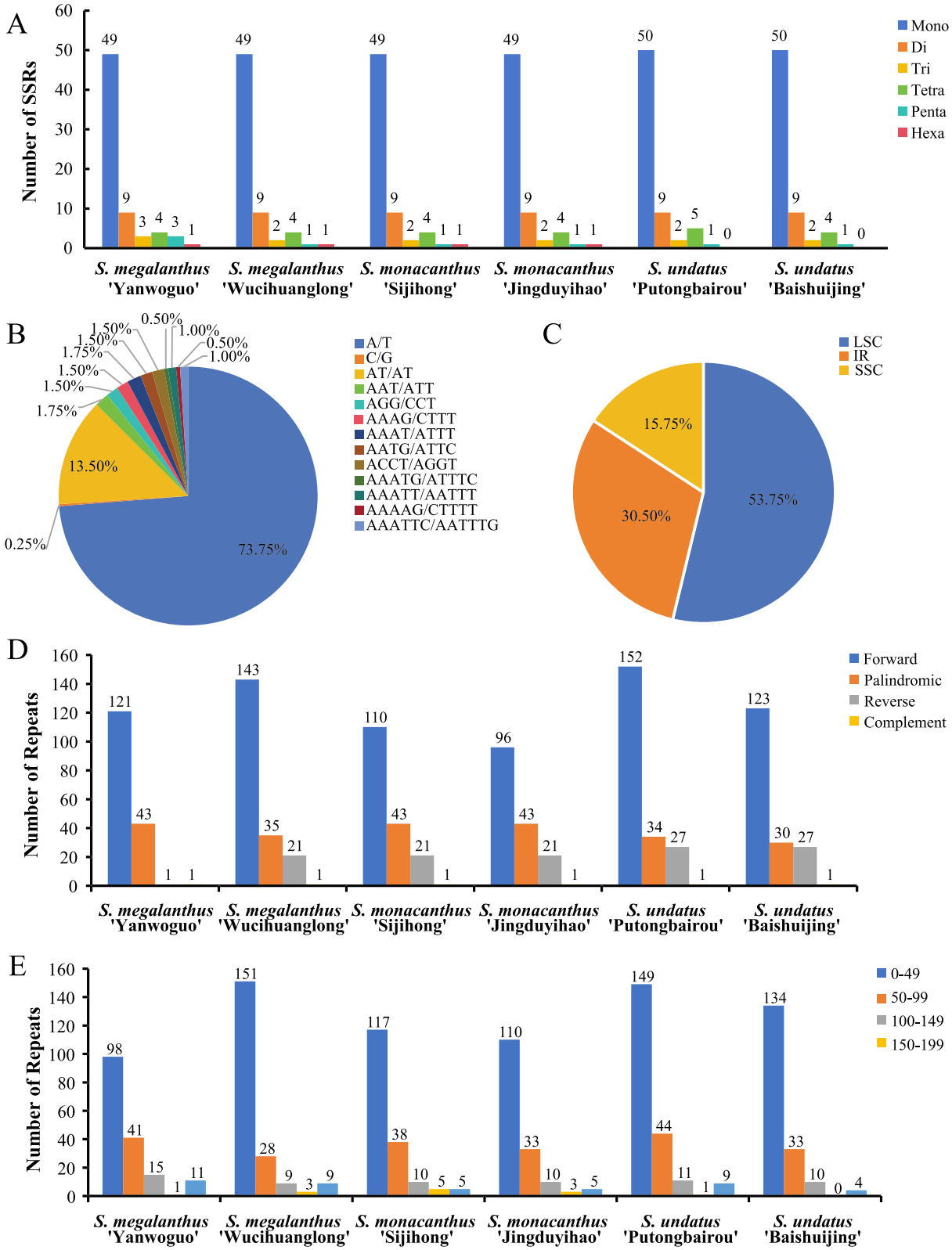

Fig. 2 Analysis of SSRs and Dispersed repeats in the chloroplast genomes of six pitaya cultivars. **A** Number of different types of SSRs; **B** Percentage of different SSRs motifs; **C** Percentage of SSRs in LSC, SSC, and IR regions; **D** Number of different types of Dispersed repeats; **E** Number of Dispersed repeat types by length. Mono = mononucleotide repeat, Di = dinucleotide repeat, Tri = trinucleotide repeat, Tetra = tetranucleotide repeat, Penta = pentanucleotide repeat, Hexa = hexanucleotide repeat

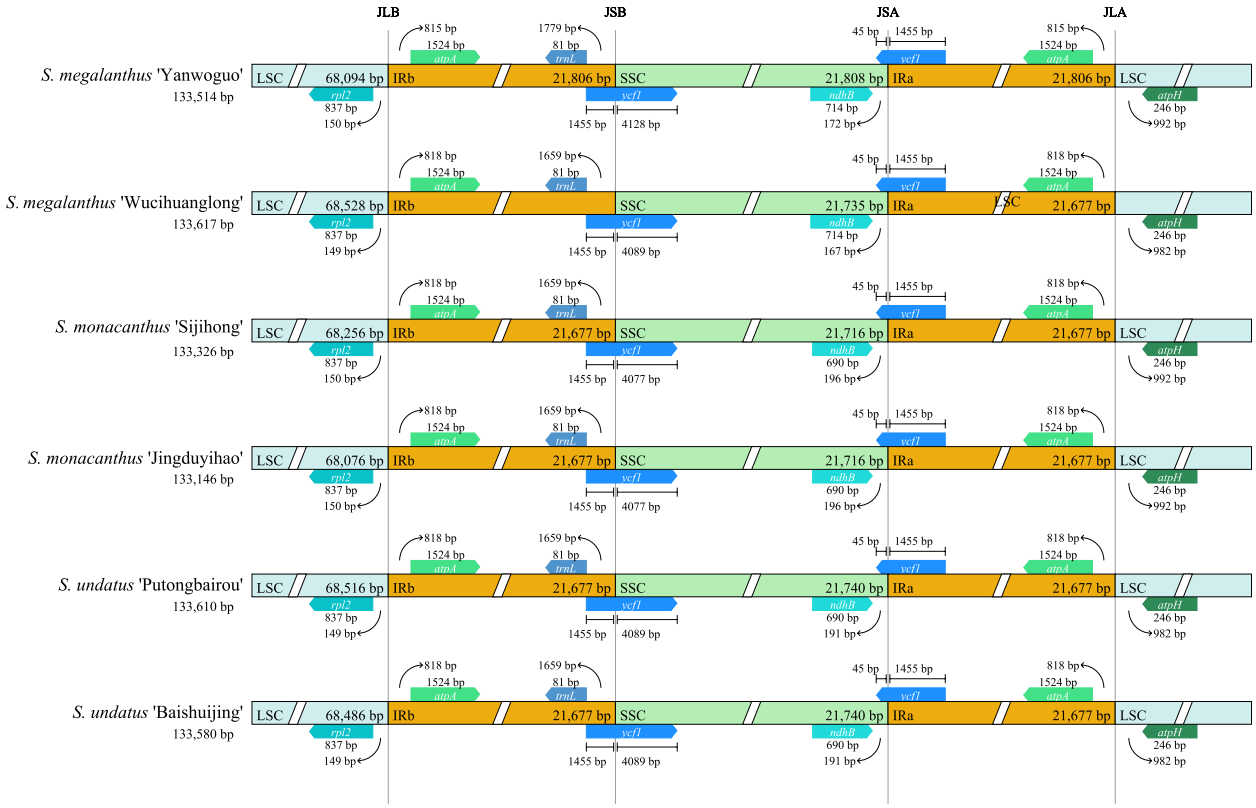

Fig. 3 Comparison of the LSC, IR, SSC junction positions among six pitaya cultivars cp genomes. JLB: junction of LSC and IRb; JSB: junction of SSC and IRb; JSA: junction of SSC and IRa; JLA: junction of LSC and IRa. Boxes above and below the mainline indicate the adjacent border genes. The gaps between the genes and the boundaries are indicated by the base lengths (bp)

of codons did not change significantly and the types of codons were consistent with the types of amino acids. Leucine (Leu: 1898–1932 codons) was the most abundant amino acid, whereas cysteine (Cys: 230–243 codons) was the least abundant (Fig. 5A and Table S3). Based on the results of the RSCU analysis, it was shown that among the 64 codons identified, RSCU values ranged from 0.33 to 1.88 (Fig. 5B). Among the six analysed sequences, UUA and CUC encoding leucine showed the largest and smallest RSCU values, respectively. The RSCU value of 1.00 for Met and Trp indicated no bias in methionine and tryptophan codons. The number of high-frequency codons (RSCU >1) was 31, with 29 ending in A or U bases. Furthermore, three stop codons, UAA, UAG, and UGA, were present. The RSCU value of UAA was >1, suggesting that the stop codons preferred UAA in the analysed sequences.

Phylogenetic analysis

To clarify the phylogenetic relationships of pitaya in the Cactaceae, we performed a phylogenetic analysis of 46 species (Fig. 6A and Fig. 6B). Phylogenetic analysis of ML and BI yielded the same topology, with the results presented in the form of a single tree. As shown in the tree topologies, Cactaceae was divided into three major clades representing the subfamilies Cactoideae, Opuntioideae, and Pereskioideae. The nine pitaya cultivars formed a monophyletic clade within the Hylocereeae tribe. They are sister groups of the tribe Echinocereaea, and all belong to the largest subfamily, Cactoideae. To show the phylogenetic relationships of the different pitaya cultivars with different peel and flesh colours more clearly, we performed phylogenetic analyses of the 10 pitaya cultivars based on their complete cp genomes (Fig. 6D) and CDSs (Fig. 6E). The ML and BI trees constructed based on these two datasets exhibited completely consistent topologies. In the phylogenetic tree, the red peel red pulp of *S. monacanthus* 'Sijihong', *S. monacanthus* 'Jingduyihao' and *S.* sp. fenhonglong were in the same branch. Furthermore, the published *Selenicereus undatus* with red peel white pulp did not form a separate branch with *S. undatus* 'Baishuijing' and *S. undatus* 'Putongbairou', which are the red-peel white pulp cultivar, but rather served as the sister of

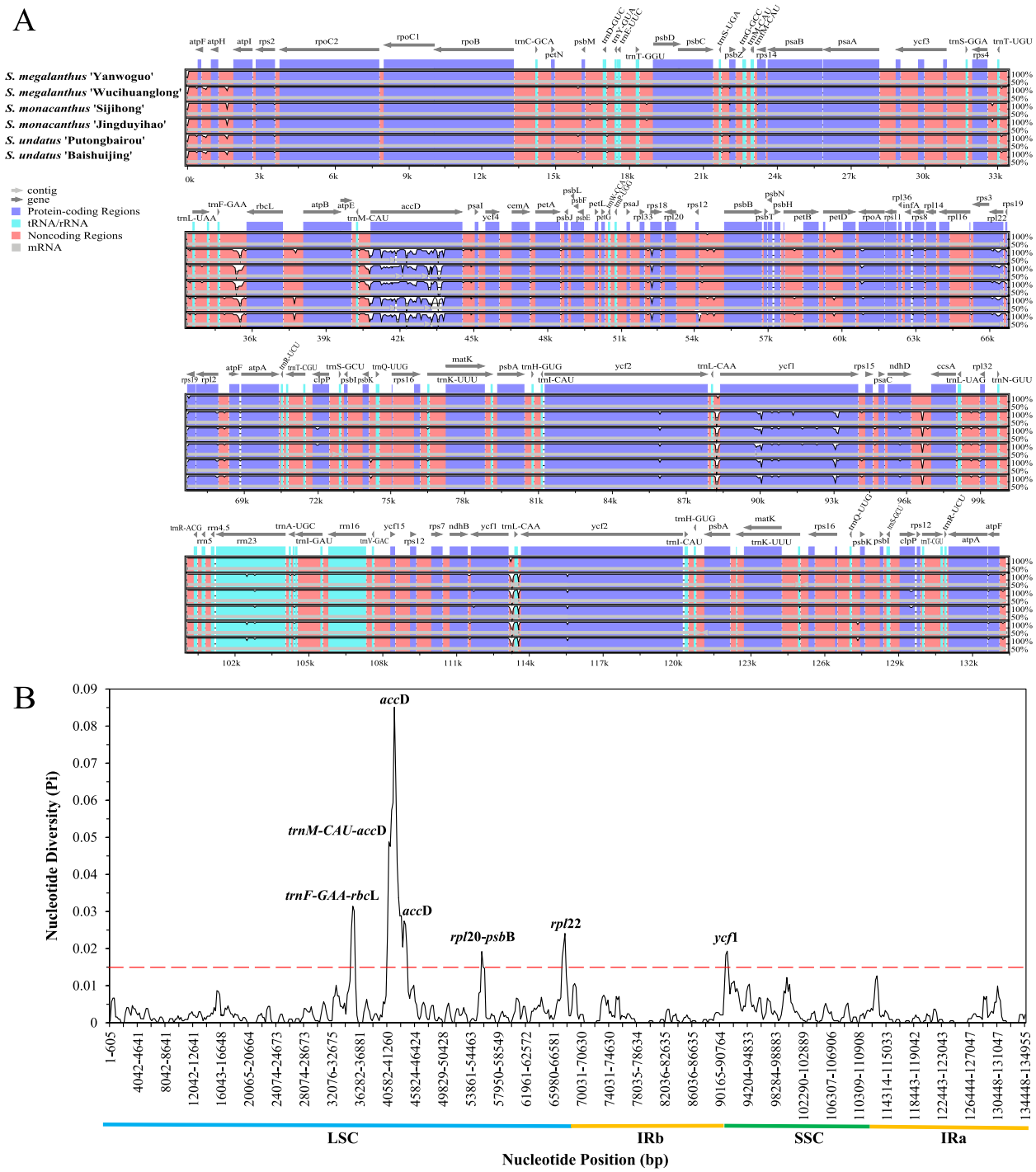

Fig. 4 Comparative analysis of the chloroplast genomes of six pitaya cultivars. **A** Comparison of the chloroplast whole genomes of seven pitaya cultivars using *Selenicereus megalanthus* annotation as a reference. Vertical scale depicts sequence similarity in aligned regions with percentage identity ranging from 50%– 100%; **B** Sliding window analysis of the chloroplast genomes of six pitaya cultivars. X-axis: position of the midpoint of the window; Y-axis: each window's nucleotide diversity. Highly variable sites (Pi ≥ 0.015) are labelled above the corresponding positions

S. sp. fenhonglong. The yellow peel white pulp cultivar *S. megalanthus* 'Yanwoguo' was clustered with the published *Selenicereus megalanthus*, but the *S. megalanthus* 'Wucihuanglong' showed similarities to the red peel white pulp of *S. undatus* 'Baishuijing' and *S. undatus* 'Putongbairou'.

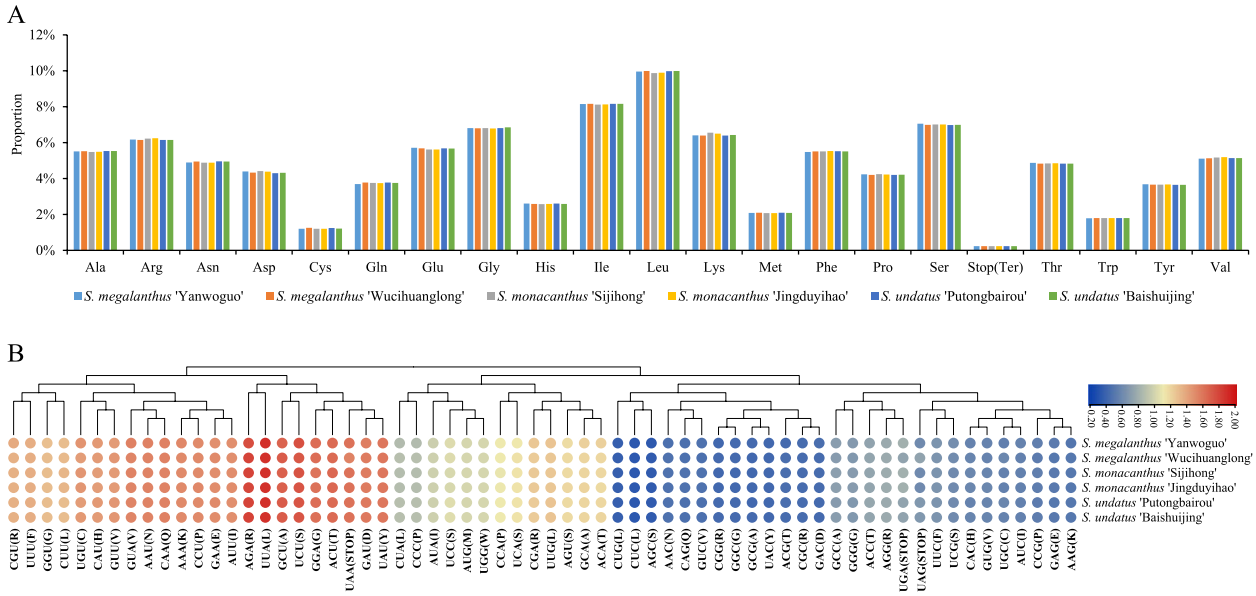

Fig. 5 Codon usage of six pitaya chloroplast genome. **A** Proportion of amino acids and stop codons in the protein coding genes. **B** Codon preference heat map of six pitaya chloroplast genome. The RSCU values of codons were used as the basis for tree clustering. As the red colour deepens, the RSCU value increases. As the blue colour deepens, the RSCU value decreases

Discussion

Characteristics of the chloroplast genome

Understanding plant cp genomes provides an important basis for species identification, evolutionary relationships, genetic engineering and other research [47, 48]. We analysed the cp genomes of six pitaya cultivars and the results indicated a 471 bp discrepancy in the sequences under consideration. No substantial rearrangement was observed between pitaya, the difference in sequence length was minimal, and all sequences exhibited a characteristic tetrameric structure. This finding indicates that the structure of six pitaya cp genomes is similar to most angiosperm cp genomes and that the genomes are largely conserved [2, 49]. Notably, we observed that specific IR boundary expansions in the cp genome of *S. megalanthus* 'Yanwoguo' may indicate potential adaptive evolution in this cultivar. These differences in length may be related to the contraction and expansion of the IR region or variability in the non-coding region [50, 51], warrant further validation through functional studies.

We identified a total of 123 genes in the sequence, of which four single-copy rRNAs are one of the characteristics shared by Cactaceae [52]. Plastid *ndh* genes and peptides encoded by nuclear genes form a higher plant NADH dehydrogenase complex, which provides energy for the recycling pathway in photosynthesis [53]. However, only two *ndh* gene family genes (*ndh*B and *ndh*D) were found in the pitaya PCGs, which is consistent with previous studies [26] and confirms the existence of a large number of loss events in the *ndh* gene family in the pitaya genome. One hypothesis posits that the absence of plastid *ndh* genes in certain plants may be attributed to their translocation from plastids to nuclear genes, a phenomenon that occurs in response to adverse environmental conditions, or to ensure the continuity of photosynthesis, a process vital for plant growth [54, 55]. Alternatively, complex alternative pathways within the nucleus may facilitate the loss of plastid *ndh* genes [56]. Additionally, extant angiosperms devoid of the *ndh* gene may represent an evolutionary endpoint on the phylogenetic tree, with plants lacking the *ndh* gene being progressively eliminated during evolution [57]. Further research is necessary to elucidate the molecular regulatory mechanisms that ensure normal growth despite the substantial loss of the *ndh* gene family in pitaya plastids.

Repeat sequences

SSRs, also known as microsatellite sequences, comprise 1–6 bp nucleotides that appear multiple times in succession, thus forming short tandem repeat sequences. This type of sequence offers certain advantages including marker abundance, high repetitiveness, and co-occurring inheritance. Therefore, it is often used as a molecular marker, and is of great significance in the studies of species identification, genetic map construction, and population polymorphism [58–60].

A total of 400 SSRs were identified in this study, and the SSRs detected in the cp genomes of pitaya of different

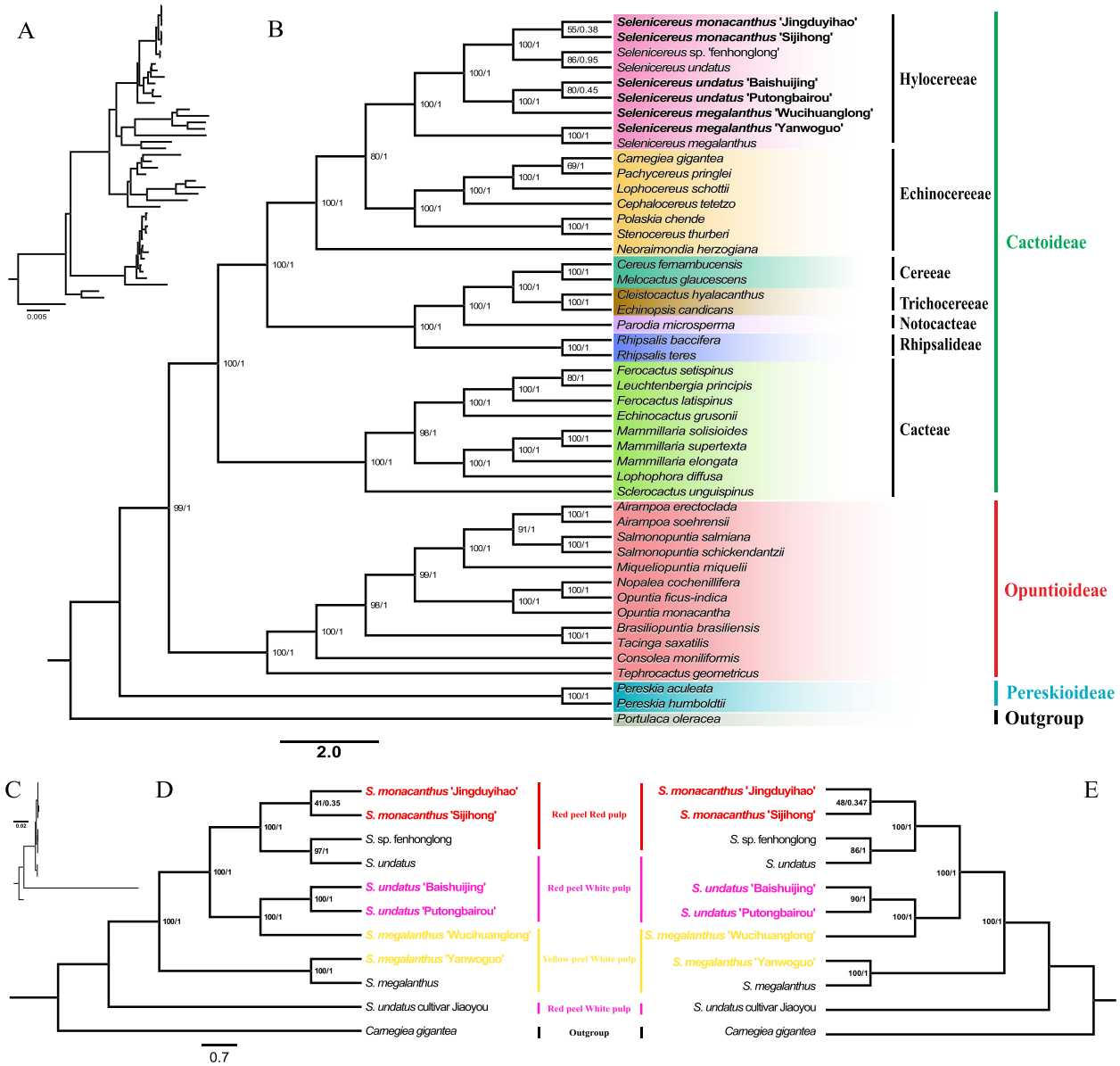

Fig. 6 Phylogenetic tree of plant chloroplast genomes. Phylogram **A** and Cladogram **B** of 46 Cactaceae species inferred from 45 shared coding sequences (CDSs); **C** Phylogram of ten pitaya cultivars inferred from whole chloroplast sequences; Cladogram constructed based on the whole chloroplast sequences **D** and 63 shared CDSs **E** of the ten pitaya cultivars, with *Camegiea gigantea* as the outgroup. Branch lengths of phylogram represent evolutionary divergence, with longer branches indicating greater genetic change and shorter branches indicating higher conservation. The newly sequenced pitaya cultivars in this study are indicated by different colours of bold font. Red: red peel red pulp pitaya; Pink: red peel white pulp pitaya; Yellow: yellow peel white pulp pitaya. The branch nodes numbers represent the Bootstrap (BS) and Posterior Probability (PP) values

colours exhibited minimal variation in number. Single-nucleotide SSRs were the most prevalent repeats in all samples, and the A/T motif was the most common. This is one of the reasons for the enrichment of AT in cp genome sequences, which is in agreement with the results of studies on plant cultivars [61, 62]. Localisation of SSRs revealed that they were predominantly distributed in the LSC region, followed by the IR and SSC regions. This distribution is inconsistent with that of the SSRs in LSC > SSC > IR found in most plants. The differences in the number and distribution of SSRs might be related to genetic variants or the contraction and expansion of the reverse repeat region [52]. This was confirmed in a study of *S. megalanthus* 'Yanwoguo', which had a significantly expanded IR region and the highest number of SSRs. Furthermore, The unique repetitive elements in *S.*

megalanthus 'Yanwoguo' consist of C/G base pairs and AATGA/ATTTC base pairs. These SSR sites can be used to identify *S. megalanthus* 'Yanwoguo' in the future [63].

The distribution and variation of dispersed repeat sequences in the genome can reveal genetic differences between cultivars and provide important information for genetic diversity studies [64]. Dispersed repeat sequence analysis of the six pitaya cp genomes revealed that forward and palindromic repeats were the most abundant, a phenomenon that is common in most plants [26, 62]. All four repeat types were present in the six pitaya cp genomes, but there were some differences in their numbers. One of the most interesting things to study is the fact that the reverse repeat sequences of the *S. megalanthus* 'Yanwoguo' showed a significant reduction, with only one identified, which may be related to the occurrence of sequence rearrangements [65, 66].

IRs contraction and expansion

Contraction and expansion of the IR regions is one of the reasons for the variation in the length of the cp genome, which is commonly observed in the subfamily Cactoideae [52]. Analysis of the IR boundaries revealed that the IR/SC boundaries were highly conserved in the cp genomes of the six pitaya cultivars. The IR-SSC boundary positions were spanned by the *ycf*1 gene. Furthermore, *rpl*2 and *atp*F were found in the neighbouring positions of the boundary of the LSC and IRb (JLB) and the boundary of the LSC and IRa (JLA), respectively. Only these boundaries and the positions of neighbouring genes away from the boundaries differed slightly. This result is consistent with the cp genomes of four *Selenicereus* spp. studied by Qin et al. [26]. The analysis of the IR region's length revealed that, the IR regions in the five cultivars exhibited high conservation, except for the expansion observed in the IR region of *S. megalanthus* 'Yanwoguo'. This suggests that *S. megalanthus* 'Yanwoguo' may have undergone a distinct evolutionary trajectory, resulting in the expansion of the IR regions.

Comparative genome analysis

A full sequence comparison of the six pitaya cp genomes revealed that these sequences were highly conserved. Most of the sites with variability were distributed in non-coding regions, with the coding regions showing more conservation, as demonstrated previously for several species [67]. *acc*D, *ycf*1, and *rpl*22 are highly variable sites that have been widely used in species identification and phylogenetic studies of various angiosperms [68–70]. A substantial body of research has identified the prevalence of repeat sequences upstream of the *acc*D gene in cactus. These repeat sequences have been shown to influence the rearrangement of plastid genes, resulting in high variability [52]. In our study, we also found high variability in *acc*D, which may have facilitated the rearrangement of the cp genome in pitaya, and thus had some effect on fruit pericarp flesh colour. Organelle genome mutation is a complex process that is influenced by multiple factors, which may also be interconnected and influenced by each other [71]. Therefore, further experimental studies are still needed to explore the mutation mechanism and evolutionary significance of organelle DNA in pitaya in the future.

To further analyse these highly variable sites, we calculated the Pi values of the six pitaya cp genomes. The results also verified a previous statement that more highly variable sites were found in the LSC and SSC regions of the pitaya than in the IR region. Among the six highly variable sites screened, *trn*M-CAU-*acc*D, *rpl*20-*psb*B, and *rpl*22 have not been reported in the genus *Selenicereus*. These have been identified as highly variable sites suitable for developing specific DNA barcodes for the identification of *Selenicereus* spp. [62, 72]. To further verify the value of these sites for *Selenicereus* species identification, it will be necessary to collect additional genetic resources and conduct molecular experiments. Furthermore, the highly variable sites identified may reflect adaptive evolution driven by environmental pressures. For instance, the polymorphism in *acc*D, a gene critical for fatty acid synthesis, could enhance chloroplast membrane stability under drought conditions by regulating lipid composition[73]. Similarly, the variability in *ycf*1, a multifunctional gene associated with stress resistance, might contribute to oxidative stress tolerance [74]. These adaptive mutations could underlie the ecological success of pitaya in tropical and subtropical regions. Future studies combining selection pressure analysis and functional assays are needed to validate the adaptive significance of these genomic regions.

Codon usage bias analysis

Codons are critical in linking genetic material, amino acids, and proteins in an organism [75–77]. Gene mutations, translation efficiency, gene expression levels, and natural selection pressures have all been suggested to contribute to codon usage bias [78, 79]. We found that Leu was the most abundant amino acid of the six pitaya cp genomes, followed by Ile and then Ser. Calculation of RSCU values in the cp genome yielded UUA-Leu as the most commonly used codons, followed by AGA-Arg. The six pitaya cp genome codons showed a strong preference for A/U endings, which also resulted in overall higher AT content in the cp genomes, a phenomenon that is also prevalent in other angiosperms, showing a broadly similar trend in the evolution of plant cp genomes [51, 61, 80]. In this study, highly similar codon usage patterns

were observed in the cp genomes of six pitaya varieties, which may be indicative of their close evolutionary relationships. Studies have shown that codon preference is closely related to tRNA abundance [81]. Codon with high preference promotes high abundance of the corresponding tRNA, reducing the risk of translation errors, and fast ribosomal translational movement thereby increasing the efficiency of protein synthesis. This optimisation mechanism essentially reflects the dynamic equilibrium strategy developed by organisms during long-term evolution [82]. Thus, Codon usage bias in the pitaya cp genome suggests potential influences on gene expression and may reflect evolutionary pressures, but further studies are needed to confirm these associations.

Phylogenetic analysis

Owing to the simple structure and matrilineal inheritance of the cp genome, it is often used to decipher the phylogenetic relationships of species [83]. Past research has shown that the use of complete cp genome data shows higher support and better reflects the phylogenetic relationships of some closely related species than the construction of phylogenetic trees using data from a single or a few cp genes [51, 70, 84]. Phylogenetic analysis revealed that the 46 individuals could be classified into four clades, including the outgroup *Portulaca oleracea* from the Portulacaceae family. Furthermore, we explored the positional relationships of pitaya cultivars with different peel and pulp colours in the genus *Selenicereus*. Phylogenetic analysis provided valuable insights into the evolutionary relationships among pitaya cultivars, although some branches require further resolution due to paraphyletic patterns and low support values. The paraphyletic relationship of yellow peel white pulp and red peel white pulp pitaya suggests that the *S. megalanthus* 'Wucihuanglong' may be hybridized between the *S. megalanthus* 'Yanwoguo' (yellow peel white pulp pitaya) and the red peel white pulp pitaya, giving it both a yellow peel and a larger fruit size. We suspect that the paraphyletic relationship in the pitaya cp genome may cause by the common interspecific hybridisation events within *Selenicereus* species. Frequent gene flow allows genetic introgression in the progeny, resulting in these cultivars not being ideally segregated based on their phenotypic characteristics [85, 86]. Moreover, due to the limited material in this study and the low support for some branches, we hope that more complete cp genomic data of pitaya will help us better understand the genus *Selenicereus* and suggest that the evolutionary relationships of pitaya populations can be further investigated in the future by combining nuclear genomic and biogeographic studies.

Conclusion

In this study, we assembled and annotated the cp genomes of six pitaya cultivars with three different colours types of peel and pule. We analysed the basic features of the six cp genomes, including their basic structure, gene composition, dispersed repeat sequences, and SSR sites. The results showed that the cp genomes of six pitaya cultivars were largely conserved in their overall structure and gene content, although variability was observed in several non-coding regions and hotspot genes such as *acc*D and *ycf*1. A significant reduction in reverse repeat sequences and expansion of the IR regions were observed in *S. megalanthus* 'Yanwoguo'. mVISTA analysis revealed more variable sites in the intergenic regions than in the coding regions. Calculating the nucleotide diversity values, we screened six different highly variable sites (Pi ≥ 0.015) and found three previously unreported highly variable sites (*trn*M-CAU-*acc*D, *rpl*20-*psb*B, *rpl*22) that may serve as potential molecular markers for species identification and phylogenetic studies, pending further validation. The observed phylogenetic patterns suggest a complex genetic basis for colour variation among pitaya cultivars, potentially influenced by hybridization and gene flow. This study provides valuable information for further understanding of the phylogenetic relationships and cp genome variation in the genus *Selenicereus*.

Abbreviations

cp	Chloroplast
LSC	Large single-copy region
SSC	Small single-copy region
IRs	Inverted repeat regions
NCBI	National Center for Biotechnology Information
SSRs	Simple sequence repeats
RSCU	Relative synonymous codon usage
Pi	Nucleotide diversity
ML	Maximum Likelihood
BS	Bootstrap value
BI	Bayesian inference
PP	Posterior probability
GC	Guanine-cytosine

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11581-2.

Supplementary Material 1. Table S1. The genebank number and classifieds of the chloroplast genome sequence downloaded from NCBI. Table S2. Summary of Chloroplast genome features of the six Selenicereus species. Table S3. Codon usage number and relative synonymous codon usage (RSCU) values of protein-coding genes of the six pitaya chloroplast genome.

Acknowledgements

We thank editors and anonymous reviewers for their valuable comments and suggestions on the manuscript.

Authors' contributions
E.T.Z. generated and analyzed the data, wrote the original draft and revised it. G.Y. analyzed the data and help to revise the manuscript. F.F.J. and C.N.L. revise the manuscript. Y.L.L. organized the data. S.H.L. collected plant materials. Q.Y.W. and X.M.T. planned and directed the study and revised the manuscript. All authors contributed to the experiments and approved the final draft of the manuscript.

Funding
This work was supported by the National Natural Science Foundation of China (32360058); The Central Government Guides Local Science and Technology Development Projects, China (2023ZYZX1224); Basic research fund of Guangxi Institute of Botany (Guizhi Ye 23007); The Key Research and Development Program of Guangxi Province of China (2024AB33097; 2024AB05024).

Data availability
Sequence data that support the findings of this study have been deposited in the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/) with the accession numbers: PQ824054–PQ824059.

Declarations

Ethics approval and consent to participate
The materials involved in the article does not an endangered or protected species; therefore, permission is not required to collect this species.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 16 January 2025 Accepted: 8 April 2025
Published online: 09 May 2025

References

1. Nunes EN, De Sousa ASB, De Lucena CM, Silva S de M, De Lucena RFP, Alves CAB, Alves RE. Pitaia (Hylocereus sp.): Uma revisão para o Brasil. Gaia Scientia. 2014;8(1):90–98.
2. Ibrahim SRM, Mohamed GA, Khedr AIM, Zayed MF, El-Kholy AA-ES. Genus *Hylocereus*: Beneficial phytochemicals, nutritional importance, and biological relevance—A review. J Food Biochem. 2018;42:e12491.
3. Luu H, Le T-L, Huynh N, Quintela-Alonso P. Dragon fruit: A review of health benefits and nutrients and its sustainable development under climate changes in Vietnam. Czech J Food Sci. 2021;39(2):71–94.
4. Chen J, Ran W, Zhao Y, Zhao Z, Song Y. Effects of fertilization on soil ecological stoichiometry and fruit quality in Karst pitaya orchard. Sci Rep. 2024;14(1):18307.
5. Chen J, Sabir I, Qin Y. From challenges to opportunities: Unveiling the secrets of pitaya through omics studies. Sci Hortic. 2023;321: 112357.
6. Martinez RM, Melo CPB, Pinto IC, Mendes-Pierotti S, Vignoli JA, Verri WA, Casagrande R. Betalains: A Narrative Review on Pharmacological Mechanisms Supporting the Nutraceutical Potential Towards Health Benefits. Foods. 2024;13(23):3909.
7. Kim H, Choi H, Moon JY, Kim YS, Mosaddik A, Cho SK. Comparative antioxidant and antiproliferative activities of red and white pitayas and their correlation with flavonoid and polyphenol content. J Food Sci. 2011;76(1):C38–45.
8. Nishikito DF, Borges ACA, Laurindo LF, Otoboni AMMB, Direito R, Goulart RA, Nicolau CCT, Fiorini AMR, Sinatora RV, Barbalho SM. Anti-Inflammatory, antioxidant, and other health effects of dragon fruit and dotential delivery systems for its bioactive compounds. Pharmaceutics. 2023;15(1):159.
9. El-Nashar HAS, Al-Azzawi MA, Al-Kazzaz HH, Alghanimi YK, Kocaebli SM, Alhmammi M, Asad A, Salam T, El-Shazly M, Ali MAM. HPLC-ESI/MS-MS metabolic profiling of white pitaya fruit and cytotoxic potential against cervical cancer: Comparative studies, synergistic effects, and molecular mechanistic approaches. J Pharm Biomed Anal. 2024;244: 116121.
10. Flores-Verastegui MIM, Coe S, Tammam J, Almahjoubi H, Bridle R, Bi S, Thondre PS. Effects of Frozen Red Dragon Fruit Consumption on Metabolic Markers in Healthy Subjects and Individuals at Risk of Type 2 Diabetes. Nutrients. 2025;17(3):441.
11. Jiang H, Zhang W, Li X, Shu C, Jiang W, Cao J. Nutrition, phytochemical profile, bioactivities and applications in food industry of pitaya (*Hylocereus* spp.) peels: A comprehensive review. Trends Food Sci Tech. 2021;116:199–217.
12. Shah K, Zhu XY, Zhang TT, Chen JY, Chen JX, Qin YH. Gibberellin-3 induced dormancy and suppression of flower bud formation in pitaya (Hylocereus polyrhizus). BMC Plant Biol. 2025;25(1):47.
13. Moreira RA, Rodrigues MA, Souza RC, Silva ADD, Silva FOR, Lima CG, Pio LAS, Pasqual M. Natural and artificial pollination of white-fleshed pitaya. Acad Bras Cienc. 2022;94(suppl 3): e20211200.
14. Shah K, Chen J, Chen J, Qin Y. Pitaya nutrition, biology, and biotechnology: A review. Int J Mol Sci. 2023;24(18):13986.
15. Luo RZ, Zhang R, Chen JX, Peng SJ, Sabir IA, Li ZQ, Wu LF, Hu GB, Shah K, Qin YH. Transcription factors HmeWRKY33 and HmeWRKY51 regulate the susceptibility of pitaya to canker disease. Plant Dis. 2025.
16. Zhang SW, Liu Y, Liu J, Li EC, Xu BL. Characterization and Pathogenicity of *Colletotrichum truncatum* Causing *Hylocereus undatus* Anthracnose through the Changes of Cell Wall-Degrading Enzymes and Components in Fruits. J Fungi (Basel). 2024;10(9):652.
17. Wallace JG, Rodgers-Melnick E, Buckler ES. On the Road to Breeding 4.0: Unraveling the Good, the Bad, and the Boring of Crop Quantitative Genomics. Annu Rev Genet. 2018;52:421–444.
18. Ansari R, Manna A, Hazra S, Bose S, Chatterjee A, Sen P. Breeding 4.0 vis-à-vis application of artificial intelligence (AI) in crop improvement: an overview. NZ J Crop Hortic Sci. 2024;1–43.
19. Jiang S, Cheng Q, Yan J, Fu R, Wang XF. Genome optimization for improvement of maize breeding. Theor Appl Genet. 2020;133:1491–502.
20. Britton NL, Rose JN. The Cactaceae: Descriptions and Illustrations of Plants of the Cactus family. 3rd ed. Washington: Carnegie Institution; 1922.
21. Gómez-Hinostrosa C, Hernández HM, Terrazas T, Correa-Cano ME. A new subspecies of *Hylocereus undatus* (Cactaceae) from southeastern México. Haseltonia. 2009;11:11–7.
22. Korotkova N, Borsch T, Arias S. A phylogenetic framework for the Hylocereeae (Cactaceae) and implications for the circumscription of the genera. Phytotaxa. 2017;327:1–46.
23. Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee SB, Peery R, McNeal JR, Kuehl JV, Boore JL. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc Natl Acad Sci USA. 2007;104(49):19369–74.
24. Gitzendanner MA, Soltis PS, Wong GK, Ruhfel BR, Soltis DE. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. Am J Bot. 2018;05(3):291–301.
25. Liu J, Liu ZY, Zheng C, Niu YF. Complete chloroplast genome sequence and phylogenetic analysis of dragon fruit (*Selenicereus undatus* (Haw.) D.R. Hunt). Mitochondrial DNA Part B. 2021;6(3):1154–1156.
26. Qin Q, Li J, Zeng S, Xu Y, Han F, Yu J. The complete plastomes of red fleshed pitaya (*Selenicereus monacanthus*) and three related *Selenicereus* species: Insights into gene losses, inverted repeat expansions and phylogenomic implications. Physiol Mol Biol Plants. 2022;28:123–37.
27. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):114–2120.
28. Jin JJ, Yu WB, Yang JB, Song Y, De Pamphilis CW, Yi TS, Li DZ. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21:241.
29. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq: Versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45:W6–11.
30. Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C. CPGAVAS2: An integrated plastome sequence annotator and analyzer. Nucleic Acids Res. 2019;47(W1):W65–73.
31. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. Geneious Basic: An integrated and extendable desktop software

platform for the organization and analysis of sequence data. Bioinformatics. 2012;28:647–1649.
32. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47: W59–W64.
33. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: A web server for microsatellite prediction. Bioinformatics. 2017;33:2583–5.
34. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: The manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001;29(22):4633–42.
35. Li HE, Guo QQ, Xu L, Gao HD, Liu L, Zhou XY. CPJSdraw: analysis and visualization of junction sites of chloroplast genomes. PeerJ. 2023;11: e15326.
36. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: Computational tools for comparative genomics. Nucleic Acids Res. 2004;32(Web Server issue):W273–W279.
37. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 2017;34:3299–302.
38. Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 2020;20(1):348–55.
39. Wang J, Wang T, Wang L, Zhang J, Zeng Y. Assembling and analysis of the whole chloroplast genome sequence of *Elaeagnus angustifolia* and its codon usage bias. Acta Botan Boreali-Occiden Sin. 2019;39(09):1559–72.
40. Sharp PM, Li WH. The codon adaptation index—a measure of directional synonymous codon usage bias and its potential applications. Nucleic Acids Res. 1987;15(3):1281–95.
41. Chen C, Wu Y, Li J, Wang X, Zeng Z, Xu J, Liu Y, Feng J, Chen H, He Y, Xia R. TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. Mol Plant. 2023;16:1733–42.
42. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
43. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 2007;56:564–77.
44. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74.
45. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–9.
46. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61(3):539–542.
47. Li DM, Zhao CY, Liu XF. Complete chloroplast genome sequences of *Kaempferia galanga* and *Kaempferia elegans*: Molecular structures and comparative analysis. Molecules. 2019;24(3):474.
48. Lv SY, Ye XY, Li ZH, Ma PF, Li DZ. Testing complete plastomes and nuclear ribosomal DNA sequences for species identification in a taxonomically difficult bamboo genus *Fargesia*. Plant Divers. 2023;45(2):147–55.
49. Zhang L, Yi C, Xia X, Jiang Z, Du LH, Yang SX, Yang X. *Solanum aculeatissimum* and *Solanum torvum* chloroplast genome sequences: a comparative analysis with other Solanum chloroplast genomes. BMC Genomics. 2024;25:412.
50. Li X, Ding Z, Miao H, Bao J, Tian X. Complete chloroplast genome studies of different apple varieties indicated the origin of modern cultivated apples from *Malus sieversii* and *Malus sylvestris*. PeerJ. 2022;10: e13107.
51. Li Z, Duan B, Zhou Z, Fang H, Yang M, Xia C, Zhou Y, Wang J. Comparative analysis of medicinal plants *Scutellaria baicalensis* and common adulterants based on chloroplast genome sequencing. BMC Genomics. 2024;25:39.
52. Yu J, Li J, Zuo Y, Qin Q, Zeng S, Rennenberg H, Deng H. Plastome variations reveal the distinct evolutionary scenarios of plastomes in the subfamily Cereoideae (Cactaceae). BMC Plant Biol. 2023;23(1):132.
53. Kharabian-Masouleh A, Furtado A, Alsubaie B, Al-Dossary O, Wu A, Al-Mssalem I, Henry R. Loss of plastid *ndh* genes in an autotrophic desert plant. Comput struct biotechnol j. 2023;21:5016–27.
54. Martín M, Sabater B. Plastid *ndh* genes in plant evolution. Plant Physiology and Biochem. 2010;48(8):636–45.
55. Ranade SS, García-Gil MR, Rosselló JA. Non-functional plastid *ndh* gene fragments are present in the nuclear genome of Norway spruce (*Picea abies* L. Karsch): Insights from in silico analysis of nuclear and organellar genomes. Mol Genet Genomics. 2016;291(2):935–941.
56. Sanderson MJ, Copetti D, Búrquez A, Bustamante E, Charboneau JL, Eguiarte LE, Kumar S, Lee HO, Lee J, McMahon M, Steele K, Wing R, Yang TJ, Zwickl D, Wojciechowski MF. Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): Loss of the *ndh* gene suite and inverted repeat. Am J Bot. 2015;102(7):1115–27.
57. Sabater B. On the edge of dispensability, the chloroplast *ndh* genes. Int J Mol Sci. 2021;22(22):12505.
58. Deguilloux MF, Pemonge MH, Petit RJ. Use of chloroplast microsatellites to differentiate oak populations. Ann For Sci. 2004;61(8):825–30.
59. Provan J, Powell W, Hollingsworth PM. Chloroplast microsatellites: New tools for studies in plant ecology and evolution. Trends Ecol Evol. 2001;16(3):142–7.
60. Zhang J, Yang J, Lv Y, Zhang X, Xia C, Zhao H, Wen C. Genetic diversity analysis and variety identification using SSR and SNP markers in melon. BMC Plant Biol. 2023;23(1):39.
61. Oulo MA, Yang JX, Dong X, Wanga VO, Mkala EM, Munyao JN, Onjolo VO, Rono PC, Hu GW, Wang QF. 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. 2020;9(8):979.
62. Alawfi MS, Alzahrani DA, Albokhari EJ. Complete plastome genomes of three medicinal Heliotropiaceae species: Comparative analyses and phylogenetic relationships. BMC Plant Biol. 2024;24(1):654.
63. Hu J, Yao J, Lu J, Liu W, Zhao Z, Li Y, Jiang L, Zha L. The complete chloroplast genome sequences of nine melon varieties (*Cucumis melo* L.): Insights into comparative analysis and phylogenetic relationships. Front Genet. 2024;15:1417266.
64. Turdi R, Mu L, Tian X. Characteristics of the chloroplast genome of *Isopyrum anemonoides*. Chin J Biotech. 2022;38(8):2999–3013.
65. Kawata M, Harada T, Shimamoto Y, Oono K, Takaiwa F. Short inverted repeats function as hotspots of intermolecular recombination giving rise to oligomers of deleted plastid DNAs (ptDNAs). Curr Genet. 1997;31(2):179–84.
66. Guisinger MM, Kuehl JV, Boore JL, Jansen RK. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: Rearrangements, repeats, and codon usage. Mol Biol Evol. 2011;28(1):583–600.
67. Song X, Ting S, Luo WJ, Ni XP, Iqbal S, Ni ZJ, Huang X, Yao D, Shen ZJ, Gao ZH. Comparative analysis of the complete chloroplast genome among *Prunus mume*, *P. armeniaca*, and *P. salicina*. Hortic Res. 2019;6:89.
68. Vinitha MR, Kumar US, Aishwarya K, Sabu M, Thomas G. Prospects for discriminating Zingiberaceae species in India using DNA barcodes. J Integr Plant Biol. 2014;56(8):760–73.
69. Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, Cheng T, Guo J, Zhou S. *ycf*1, the most promising plastid DNA barcode of land plants. Sci Rep. 2015;5:8348.
70. Song BN, Liu CK, Zhao AQ, Tian RM, Xie DF, Xiao YL, Chen H, Zhou SD, He XJ. Phylogeny and diversification of the genus *Sanicula* L. (Apiaceae): Novel insights from plastid phylogenomic analyses. BMC Plant Biol. 2024;24(1):70.
71. Wang J, Zou Y, Mower JP, Reeve W, Wu Z. Rethinking the mutation hypotheses of plant organellar DNA. Genomics Commun. 2024;1: e003.
72. Newmaster SG, Fazekas AJ, Steeves RA, Janovec J. Testing candidate plant barcode regions in the Myristicaceae. Mol Ecol Resour. 2008;8(3):480–90.
73. Huang C, Liu D, Li ZA, Molloy DP, Luo ZF, Su Y, Li HO, Liu Q, Wang RZ, Xiao LT. The PPR protein RARE1-mediated editing of chloroplast accD transcripts is required for fatty acid biosynthesis and heat tolerance in Arabidopsis. Plant commun. 2023;4(1): 100461.
74. Khandelwal NK, Millan CR, Zangari SI, Avila S, Williams D, Thaker TM, Tomasiak TM. The structural basis for regulation of the glutathione transporter Ycf1 by regulatory domain phosphorylation. Nat Commun. 2022;13:1278.
75. Fu H, Liang Y, Zhong X, Pan Z, Huang L, Zhang H, Xu Y, Zhou W, Liu Z. Codon optimization with deep learning to enhance protein expression. Sci Rep. 2020;10:17617.
76. Outeiral C, Deane CM. Codon language embeddings provide strong signals for use in protein engineering. Nat Mach Intell. 2024;6:170–9.

77. Diez M, Medina-Muñoz SG, Castellano LA, Da Silva PG, Wu Q, Bazzini AA. iCodon customizes gene expression based on the codon composition. Sci Rep. 2022;12:12126.
78. Plotkin J, Kudla G. Synonymous but not the same: The causes and consequences of codon bias. Nat Rev Genet. 2011;12:32–42.
79. Chen SL, Lee W, Hottes AK, Shapiro L, McAdams HH. Codon usage between genomes is constrained by genome-wide mutational processes. Proc Natl Acad Sci USA. 2004;101(10):3480–5.
80. Lin XM, Huang JQ. Codon Preference Analysis of the Chloroplast and Nuclear Genomes in Cactaceae. Mol Plant Breed. 2024;22(22):7400–12.
81. Ren GP, Dong YY, Dang YK. Codon codes: Codon usage bias influences many levels of gene expression. Sci Sin Vitae. 2019;49(7):839–47.
82. Parvathy ST, Udayasuriyan V, Bhadana V. Codon usage bias. Mol Biol Rep. 2022;49:539–65.
83. Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17:134.
84. Chu ZZ, Yisilam G, Qu ZZ, Tian XM. Comparative analyses on the chloroplast genome of three sympatric *Atraphaxis* species. Chin Bull Bot. 2023;58(3):417–32.
85. Cisneros A, Tel-Zur N. Genomic analysis in three *Hylocereus* species and their progeny: Evidence for introgressive hybridization and gene flow. Euphytica. 2013;194:109–24.
86. Guo M, Pang X, Xu Y, Jiang W, Liao B, Yu J, Xu J, Song J, Chen S. Plastid genome data provide new insights into the phylogeny and evolution of the genus *Epimedium*. J Adv Res. 2021;36:175–85.

Publisher's Note