Genome Biology

RESEARCH

Open Access

Chromosome-level genome assembly of the autotetraploid yellow pitaya provides novel insights into evolution of trait patterning in pitaya species with different ploidy

Qamar U. Zaman[1,2,3,11], Ali Raza[4], Liu Hui[1,2,3], Mian Faisal Nazir[5], Vanika Garg[6], Muhammad Ikram[1,2,3], Guoqing Wang[1,2,3], Wei Lv[1,2,3], Darya Khan[1,2,3], Aamir Ali Khokhar[1,2,3], Zhang You[1,2,3], Annapurna Chitikineni[6], Babar Usman[1,2,3], Cui Jianpeng[1,2,3], Xulong Yang[1,2,3], Shiyou Zuo[1,2,3], Peifeng Liu[1,2,3], Sunjeet Kumar[1,2,3], Mengqi Guo[7], Zhi-Xin Zhu[1,2,3], Girish Dwivedi[8,9], Yong-Hua Qin[10], Rajeev K. Varshney[6*] and Hua-Feng Wang[1,2,3*]

*Correspondence: rajeev.varshney@murdoch.edu.au; hfwang@hainanu.edu.cn

[1] School of Breeding and Multiplication (Sanya Institute of Breeding and Multiplication), Hainan Yazhou Bay Seed Laboratory, Hainan University, Sanya 572025, China
[6] Centre for Crop and Food Innovation, WA State Agricultural Biotechnology Centre, Murdoch University, Murdoch, WA 6150, Australia
Full list of author information is available at the end of the article

Abstract

Background: Yellow pitaya (*Selenicereus megalanthus*, 2n = 4x = 44) breeding remains severely hindered due to the lack of a reference genome.

Results: Here, we present a high-quality chromosome-level genome assembly of yellow pitaya using PacBio HiFi sequencing and Hi-C scaffolding technologies. We identify yellow pitaya as an autotetraploid with a genome size of 1.79 Gb, harboring 27,246 high-confidence genes probably from diploid ancestors, red pitaya (*S. undatus*). By comparative analysis of the 3D chromatin architecture, we identify varying number of compartment A/B, topologically associated domains (TADs), and structural variations in diploid (red pitaya) and polyploid (yellow pitaya) species. We find that TAD boundaries are enriched with transcription factor motifs in both species. We find significant alterations in expression of genes in the betalain biosynthesis pathway in both species. We detect differential expression of genes encoding key regulators of pericarp color within the TAD regions of polyploid pitaya and diploid pitaya. We also identify the expression differences in candidate genes that likely influence betacyanin and betaxanthin synthesis in both species.

Conclusions: Our findings suggest that differential 3D genome organization, especially differences in TAD boundaries, may impact gene expression, which may further lead to different trait formation in different pitaya species. This provides theoretical implications for fast-forward breeding.

Keywords: Genome assembly, *S. undatus*, *S. megalanthus*, Genome compartmentalization, TADs, Diploid pitaya, Polyploid pitaya, Structural variations

Background

Dragon fruit or pitaya (*Selenicereus* spp., formerly *Hylocereus*) belongs to the Cactaceae family, an ancient lineage dating back roughly 65 million years [1]. Cacti likely originated during the mid-Tertiary period (~30 million years ago), as supported by chloroplast DNA and internal transcribed sequence (ITS) data [1]. Native to Central America (including southern Mexico, Salvador, Guatemala, and Costa Rica), *Selenicereus* likely emerged during the Early Cretaceous period, aligning with the breakup of the Gondwana continent [2]. The genus includes 28 species classified by morphological traits [3] and is extensively dispersed across America [4]. Among these, *S. megalanthus* (2n = 4x = 44) and *S. undatus* (2n = 22) are the most commercially viable due to their appealing appearance, flavor, and nutritional value, which includes vitamins, antioxidants, minerals [5], phytochemicals (including betacyanins, phenolics, polysaccharides, terpenoids, and alkaloids) [6], dietary fiber, and cancer-preventive properties [7–9]. Fruit morphology, including dimensions, coloration, spine quantity, and areole length in cladodes, is considered as the key criterion for species identification [7].

The Aztecs are believed to have first cultivated pitaya in the thirteenth century [10]. Nowadays, pitaya is widely grown in tropical and subtropical regions, with Vietnam as the leading producer [11], contributing around 50% of the global market (14.11 billion USD https://ap.fftc.org.tw/article/1596), followed by China, Mexico, Colombia, Thailand, Malaysia, India, Indonesia, Australia, Israel, the USA, Pakistan, and many other countries growing on a small scale [4, 7]. Due to its spectacular appearance and antioxidant capacity, its demand and cultivation have surged globally [4, 7]. Pitaya is rich in betalains and tyrosine-originating alkaloid pigments [12], which makes it valuable for the food industry. Its drought tolerance is notable [3], likely stemming from Cactaceae's evolution under low-CO_2 level, arid Gondwana environments [13]. Adaptations like succulent stems, aerial roots, and water-rich fruits support survival under harsh environmental conditions.

Based on their hallmark flavor and morphological and genetic fruit characteristics, the diploid red pitaya (*S. undatus*, red peel and white flesh) and tetraploid yellow pitaya (*S. megalanthus*, yellow peel and white flesh) show unique genetic and morphological features. *S. megalanthus* possesses a yellow peel with large thorns, four or more ribbed cladodes, small fruit scales, small to medium-sized fruits, large seeds, fused parenchyma-chlorenchyma layers, and sweeter flesh (Fig. 1a). In contrast, *S. undatus* has a spineless red peel, three ribs per cladode, smaller seeds, and distinct parenchyma and chlorenchyma layers (Fig. 1b). Yellow pitaya's classification and ploidy have long been debated. First described as *Mediocactus megalanthus* in the 1920s due to its intermediate traits [4, 14], it was later reclassified to *Selenicereus* genus [15], then to *Hylocereus* in 2003 following a proposed allotetraploid origin [16]. In 2013, it was suggested to be an autopolyploid [17]. Ultimately, *Hylocereus* was merged with *Selenicereus* in 2017 [18] (Fig. 1c). Despite distinguished morphological features, the evolutionary divergence between diploid and tetraploid species remains poorly explored and requires more detailed examinations.

The absence of a high-quality reference genome for yellow pitaya limits its fast-forward breeding. Thus, it is crucial to fast-track its breeding by understanding the genetic basis of morphological differences, such as peel color, cladode thickness, and

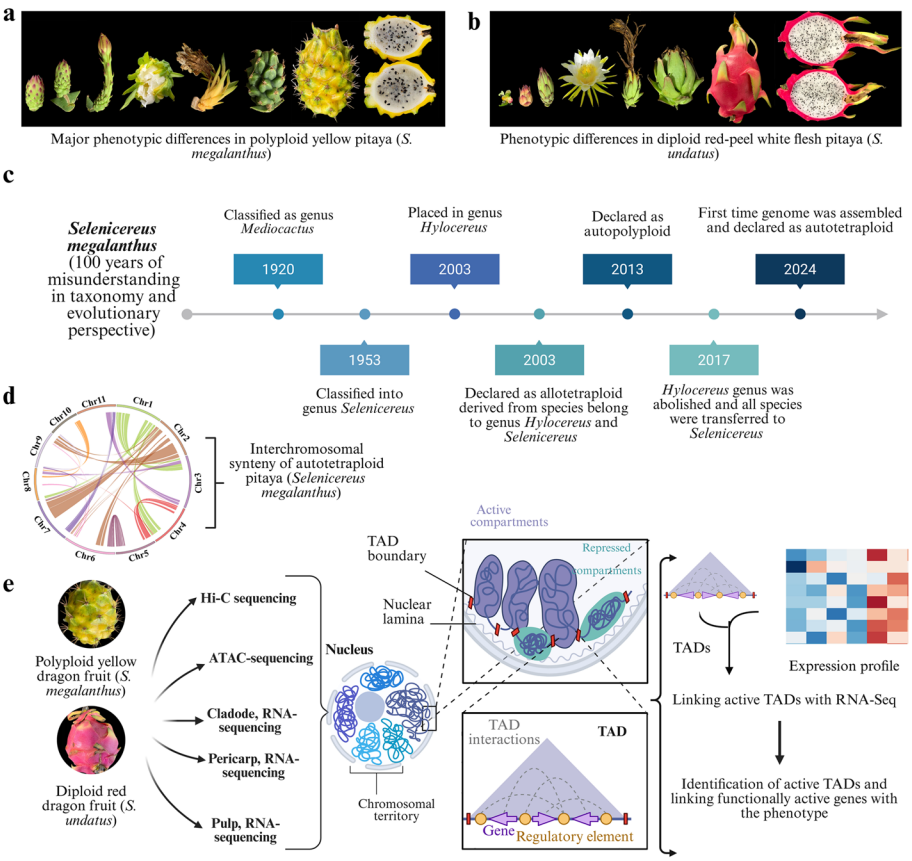

Fig. 1 Misconceptions in pitaya taxonomy and genomic insights into diploid and polyploid species. **a** Phenotypic characteristics of yellow and **b** red pitaya (white flesh). **c** Schematic depiction of 100 years of misunderstood taxonomy of yellow pitaya. Our assembled genomic resource provides insights into the yellow pitaya genome and clarifies the ploidy level achieved in this study. **d** Interchromosomal synteny of a newly assembled genome of yellow pitaya. **e** Comparative genomics of diploid and tetraploid pitaya. Integrative analysis workflow combining Hi-C, ATAC-seq, and RNA-seq data to associate active TADs with gene expression and fruit phenotypic traits

spine presence. The questions remain about how DNA replication, chromatin architecture (topologically associated domain-TAD and compartments A/B), and gene expression differ between diploids and polyploids. In this context, TAD boundaries, enriched in regulatory elements, are central to genome organization and gene regulation [19].

Here, we present the high-quality chromosome-level genome assembly of autotetraploid yellow pitaya (*S. megalanthus*-K. Schum. Ex Vaupel-Ralf Bauer), formerly *Hylocereus megalanthus* (Fig. 1d), developed using PacBio HiFi sequencing. Complementary Hi-C, ATAC-seq, and tissue-specific transcriptome data were also obtained for both yellow and red pitaya (Fig. 1e). Comparative analysis provided insights into chromosomal organization, chromatin structure, TADs, and structural variations (SVs) between both diploid and tetraploid species. Our findings confirm that yellow pitaya is an autotetraploid with high heterozygosity (AAAB). In brief, our work provides novel genomic resources for understanding polyploid genome complexity and evolution and will support future breeding and improvement of these economically important fruit crops.

Results

Chromosome-level genome sequencing, assembly, genome characteristics, and evolution of polyploid pitaya

We assembled a chromosome-level genome of yellow pitaya (*S. megalanthus*) using whole-genome sequencing. For a preliminary overview, *K*-mer analysis (21-mers) estimated the genome size at ~ 3.32 Gb, with 64% AAAB-type, 21% AABB-type, and other minor *K*-mer pair types (Additional file 1: Fig. S1). The estimated results of the 21 *K*-mers with the highest fitting degrees were selected as the final spectrum to predict genome characteristics. A total of ~ 201.97 Gb of sequencing data was generated, which contained 194.17 Gb of high-quality clean data. Smudgeplot analysis confirmed a tetraploid with a dominant "AAAB" pattern (Additional file 1: Fig. S2). The high similarity among the four haplotypes made haploid genome separation difficult, which agrees with a tetraploid karyotype (2n = 4x = 44) [20–22].

We used PacBio HiFi for de novo assembly, Hi-C for scaffolding and TAD identification, ATAC-seq for chromatin accessibility, and RNA-seq for gene prediction. PacBio Revio generated 97.76 Gb circular consensus sequencing (CCS) reads with an average length of 15.95 kb at an N50 of 16 kb (Additional file 2: Table S1), which results in an initial 900 contigs with N50 of 13.18 Mb. Using Purge Haplotigs [23], redundant sequences were removed. A genome-wide landscape of *S. megalanthus* is presented in a circos plot (Fig. 2), showing the chromosomal distribution of GC content, gene density, total repeats, long terminal repeats (LTRs), DNA transposable elements (DNA-TEs), and tandem repeats (TRs). Hi-C scaffolding (477.11 Gb, ~ 100 × coverage) anchored 414 contigs into 11 chromosomes, producing a final genome size of 1.79 Gb with 37.01% GC content (Fig. 2, Table 1) and a 97.73% anchoring rate.

The completeness of assembly was assessed using Compleasm and BUSCO. Genome sequence alignment was carried to the NT database (Additional file 2: Table S2). The initial draft assembly (3.75 Gb) showed 97.65% completeness with 83.4% duplicated BUSCOs (Additional file 2: Table S3). After purging redundant haplotypes and Hi-C scaffolding, the final genome assembly exhibited 92.7% completeness with only 2.7% duplication (Additional file 2: Table S4), suggesting high quality of genome assembly. The distribution of BUSCO copy numbers (Additional file 1: Fig. S3) supports this reduction in duplication across the gene set. Additional details of genome architecture (including read depth, GC content, SNP density, and repeat distribution) are shown in Additional file 1: Fig. S4. Notably, regions with high GC content, such as the distal end of a few chromosomes, showed localized dips in sequencing depth, likely due to assembly challenges in repetitive-rich regions.

Unanchored sequences may represent genomic segments that were acquired through lateral gene transfer (LGT) during domestication. We predicted 27,246 high-confidence genes using a combination of homology-based and de novo methods, including reference genomes (*S. undatus* and *Carnegiea gigantea*) and the BGI HiFAP pipeline (Additional file 2: Table S3). Among 36,277 transcripts, 91.48% were functionally annotated (Additional file 2: Tables S5–S7). The average gene length was 6255.31 bp (mean length of transcripts-7247.29 bp), the average CDS length was 1187.89 bp, and the mean exons per gene was 5.42 (Additional file 1: Fig. S5). Comparative analysis with 10 species identified 146,064 gene families, 6751 shared

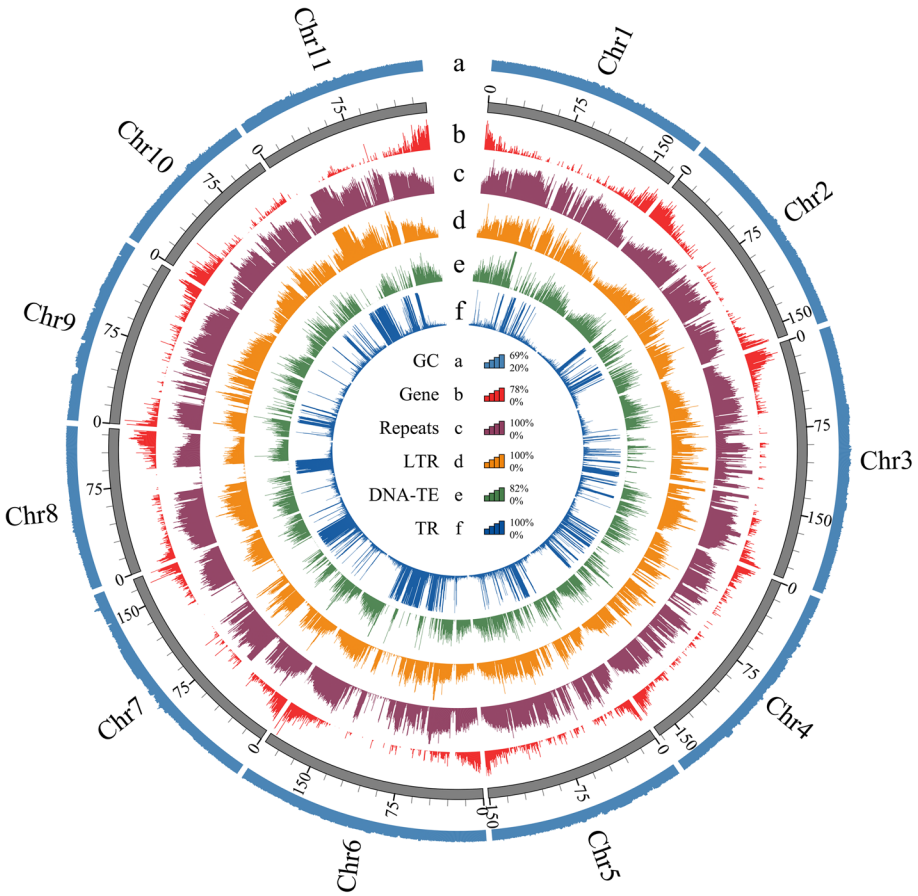

Fig. 2 The genomic characteristics of the autotetraploid yellow pitaya genome (*S. megalanthus*). The outermost track shows the 11 assembled chromosomes. The six inner concentric tracks represent the genomic distribution of (**a**) GC content, (**b**) gene density, (**c**) total repeat content, (**d**) long terminal repeats (LTRs), (**e**) DNA transposable elements (DNA-TEs), and (**f**) tandem repeats (TRs). Each track is scaled individually, with corresponding color-coded gradients shown in the center

Table 1 Statistics of genome assembly and annotation of yellow pitaya (*Selenicereus megalanthus*)

Features	*Selenicereus megalanthus*
Total assembly size (Gb)	1.7916
Number of contigs by HiFi	900
Number of contigs by HiFi + Hi-C + Purge	425
Assembly anchored in chromosomes (%)	97.73
GC content (%)	37.01
Genes with defined alleles	27,246 (100%)
Functionally annotated genes	24,924 (91.48%)

families, and 1149 single-copy orthologs that were used to construct the phylogenetic tree (Additional file 2: Table S8, Additional file 1: Fig. S3b). The contig distribution across chromosomes (Fig. 3a) highlights the assembly and coverage across the 11 chromosomes and unanchored regions. *S. megalanthus* contained 17,885 gene families and 25,268 genes, including 81 unique families. Divergence between diploid and

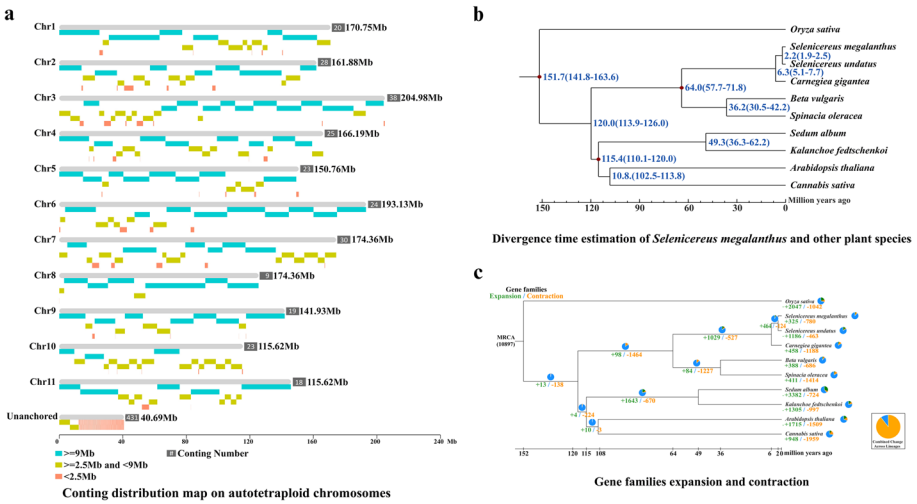

Fig. 3 Contig distribution map on autotetraploid chromosomes, divergence time of *S. megalanthus*, and the estimation of gene family expansion and contraction. **a** The gray color indicates the length of each chromosome, whereas other colors represent contigs of varying length. **b** Divergence time of *Selenicereus megalanthus* and other plant species. The number at the node position represents the divergence time of plant species from their predecessors in million years ago (MYA). The divergence time supported by the 95% highest posterior density (HPD) is also indicated in brackets. The branch length of a phylogenetic tree represents only the number of base substitutions or genetic distance per point. **c** The expansion and contraction of gene families in plant species. The green color indicates the gene families incorporated into the genome over the species evolutionary history, and the red color represents the loss of gene families from the plant species. The blue color in the pie charts represents the proportion of gene families (copy number) that remained unchanged during the evolutionary lineages

polyploid occurred ~ 1.9-2.2 million years ago (MYA), both sharing a WGD event, while divergence from *C. gigantea* occurred ~ 5.1-7.7 MYA (Fig. 3b).

We identified 464 expended and 124 contracted gene families between diploid and polyploid pitayas (Fig. 3c). Importantly, expanded genes were significantly enriched in "protein dimerization activity," "terpene synthase activity," "diterpenoid biosynthetic pathway," and "glycosyltransferase activity" in *S. megalanthus*, suggesting their role in yellow pigmentation via betaxanthin biosynthesis (Additional file 1: Fig. S6). Using WGDI, synonymous substitution rate (*Ks*) distribution discovered a single peak, suggesting a WGD ~ 1.9–2.2 MYA, followed by autopolyploidization of *S. megalanthus*. Collinearity analysis confirmed conserved genomic blocks within and across *S. megalanthus* and *S. undatus* (Additional file 1: Fig. S7, Additional file 1: Fig. S8), which resulted in 777 syntenic blocks (average of 37 genes/block) (Supplementary Table 9). We removed synteny blocks associated with ancient WGD signals to clarify the evolutionary relationship further. The filtered collinearity analysis discovered a clean 1:1 chromosomal correspondence between *S. undatus* and *S. megalanthus* (Additional file 1: Fig. S9), supporting the hypothesis that the diploid species or a closely related ancestor contributed to the autotetraploid origin of *S. megalanthus*. These results suggest their shared evolutionary history and genomic rearrangements (Additional file 2: Table S9). The nucleotide substitution rate (r) was calculated as 2.29×10^{-9}/site/year using the *Ks* and divergence time. LTR insertion time was computed using $T = K/2r$ (or $r = K/2T$), linking genome evolution to autopolyploidization and mutation accumulation.

Genome-wide comparative analysis of diploid and polyploid pitaya

We conducted a comparative genomic analysis of diploid and polyploid pitaya plants to investigate differences in 3D genome architecture, including chromatin compartments and TADs that influence gene expression. Interaction matrices were converted into Z-score matrices, and their differences were computed to construct 200-kb resolution heatmaps for diploid (Fig. 4a) and polyploid (Fig. 4b) pitaya, which highlights

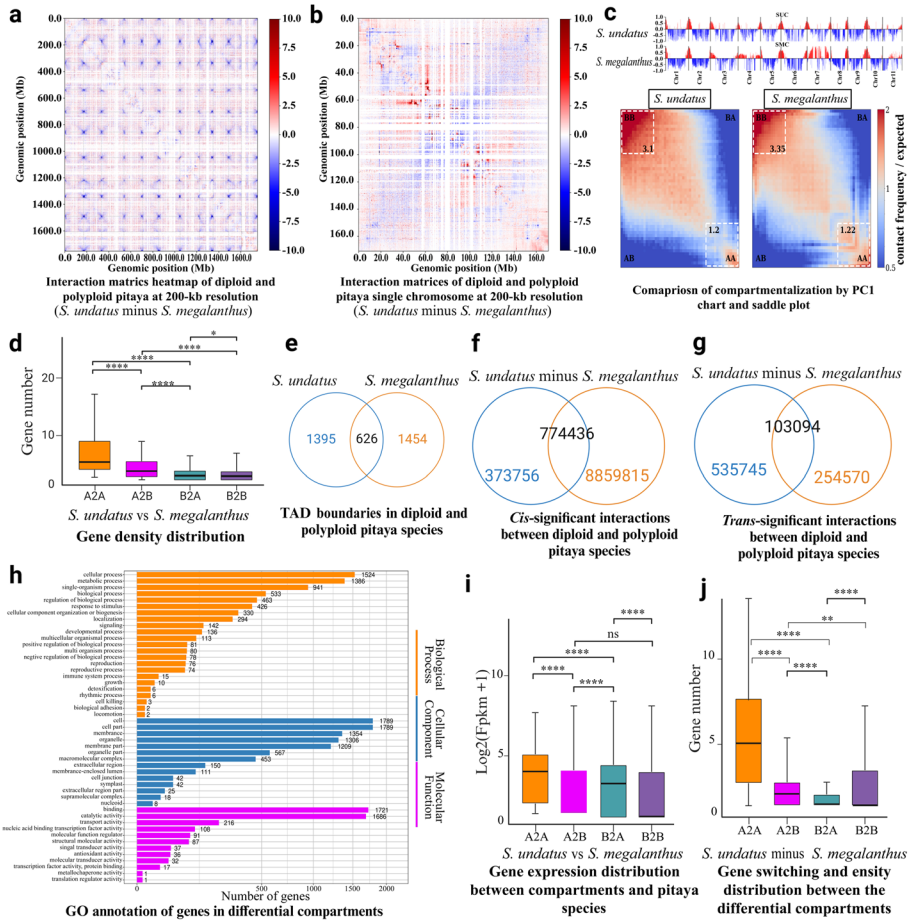

Fig. 4 Genome-wide comparative analysis of diploid and polyploid pitaya species. **a** Interaction heatmap at 200-kb resolution showing the differences between the genomes of polyploid and diploid pitaya (*S. megalanthus* subtracted from *S. undatus*). Red indicates greater interaction intensity in *S. megalanthus*, and blue indicates stronger interactions in *S. undatus*. Both axes represent genomic positions. **b** Reduced heatmap of the interaction of single chromosomes at 200-kb resolution for both pitaya species. The red color bar indicates that the interaction intensities of *S. megalanthus* were greater than those of *S. undatus*. A higher value in the blue bar indicated higher differences in interaction intensity between the two species. **c** PC1 comparison chart and compartment saddle plot. Red indicates compartment A, and blue represents compartment B. In the saddle plot, color intensity reflects interaction signals between AA, BB, and AB compartments. **d** Gene density distribution across four compartment types. **e** Venn diagram of common and unique TAD boundaries in diploid (*S. undatus*) and polyploid (*S. megalanthus*) pitaya species. **f** Venn diagram of *cis*-significant interactions between two species, including common and unique interactions. **g** Venn diagram of *trans*-significant interactions, including common and unique interactions between two species. **h** GO annotation of genes in differential compartments of diploid and polyploid species. The horizontal axis represents the number of genes annotated to the GO terms. The vertical axis denotes GO terms in three categories: biological processes, cellular components, and molecular functions. **i** Distribution of gene numbers across conserved (A2A and B2B) and switching (A2B and B2A) compartments. **j** Distribution of gene density across distinct compartments and gene switching between the compartments of diploid and polyploid pitaya. Statistical significance: ****$p < 0.0001$, ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, and ns: not significant

chromosome-level interactions. At this resolution, both genomes showed increasing genomic distance with decreasing interaction frequency, suggesting structural complexity (Additional file 1: Fig. S9, Additional file 1: Fig. S10). Moreover, saddle plots discovered stronger A-A compartment interactions in polyploid pitaya (1.2-1.22) and slightly enhanced B-B interactions (3.35 vs. 3.1 in diploid) (Fig. 4c) which suggests ploidy-dependent chromatin compartmentalization. These compartments were dynamic, with A/B state switching associated with large-scale transcriptional shifts [24]. Further analysis at 100-kb resolution showed dynamic A/B switching and TAD reorganization, reflecting evolutionary trajectory (Fig. 4d and Additional file 2: Table S10). We identified 626 shared, 1395 diploid-specific, and 1454 polyploid-specific TAD boundaries (Fig. 4e and Additional file 2: Table S11). Using the Fit-Hi-C technique [25], we detected 362,158 significant *cis*- (Fig. 4f) and 144,1230 *trans*-interactions (Fig. 4g) in both pitaya species, highlighting differences in long-range interactions. Genes within the TAD boundaries were enriched in cellular process, binding functions, and signal transduction pathways (Fig. 4h; Additional file 2: Table S12, Additional file 1: Fig. S11). Gene expression analysis across conserved (A2A and B2B) and switching compartments (A2B and B2A) displayed that most differentially expressed genes were located in A2A regions (Additional file 2: Table S13 and Fig. 4i), with A2A correlating more with gene density compared to B2B. In contrast, B2A regions had greater gene density than A2B (Fig. 4j).

Genomic structural differences in diploid and polyploid pitaya species

Polyploids arise from the genome duplication (autopolyploid) or hybridization of divergent genomes (allopolyploid) [26] and are often accompanied by rapid SVs, which are vital for evolutionary adaptation. We compared autotetraploid yellow pitaya to diploid pitaya and identified 41,409 and 82,394 presence-absence variations (PAVs), 1245 inversions, 15,725 translocations, 23,325,494 SNPs, 2,294,810 InDels, and 19,928 copy number variations (CNVs). Most SVs, including PAVs (31,506–66,158), inversions (710), translocations (13,533), and CNVs (14,890), were found in intergenic regions of yellow pitaya (Additional file 2: Table S14), contributing to phenotypic differences (Fig. 1A). Despite extensive synteny between the genomes, frameshift insertions and SVs in UTRs, exons, introns, and intergenic regions further expanded the polyploid genome.

Notably, large inversions were noticed on chromosomes 1-5, 7-9, and extensive translocations were found across nearly all chromosomes, except Chr 10, which had the fewest duplications 10 (Fig. 5). High SNP densities on Chr 1 (up to 48 Mb) and low densities at chromosomal ends suggest evolutionary hotspots and conserved regions, respectively (Additional file 1: Fig. S12). SNP counts ranged from 25,752,287 (Chr 1) to 15,658,201 (Chr 11), reflecting heterozygosity in yellow pitaya.

The diploid genome is ~ 1.41 Gb [27], while the yellow pitaya genome expanded to 1.79 Gb, indicating extensive structural reorganization. Clonal propagation [28] or somatic mutations [29] may introduce further variation. In this context, evidence from other species suggests DNA loss or retention varies, e.g., *Phlox drummondii* shows rapid loss post-autopolyploidization [30], while *Arabidopsis* retains genome size initially [31].

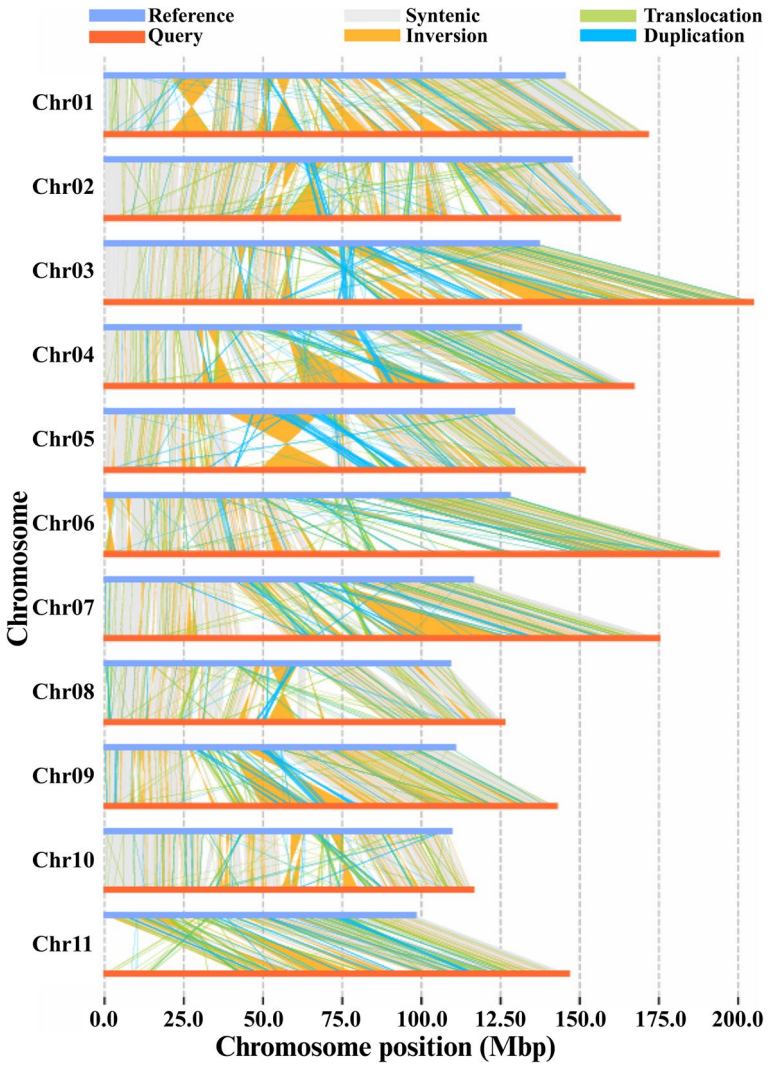

Fig. 5 Synteny analysis of *S. undatus* and *S. megalanthus*. The horizontal blue and orange bars represent the genome of *S. undatus* (reference sequence) and *S. megalanthus* (query sequence). The gray, yellow, green, and light blue bars denote the structural variations between the two genomes, including syntenic regions, inversions, translocations, and duplications, respectively

Three-dimensional genomic architecture of diploid and polyploid pitaya species

The eukaryotic chromatin genome is organized into hierarchical structures, including chromatin loops, A/B compartments, and TADs. To explore the 3D genome organization in both species, we generated and analyzed Hi-C datasets aligned to the respective reference genomes. Following adapter trimming and quality assessment, including base content, quality, and GC distribution (Additional file 1: Figs. S13–S16), we obtained 958–997 million clean reads for diploid and 1582–1916 million for polyploid pitaya (Additional file 2: Table S15). Data reproducibility was confirmed by high inter-sample correlations (0.992 for diploid and 0.989 for polyploid).

Hi-C interaction matrices were constructed by binning the genome (100-kb resolution for genome-wide and 40 kb for chromosome-level resolution), filtering valid

pairs (Q30) (Additional file 2: Table S16), and applying normalization via Juicer to correct for GC bias and restriction site distribution. This yielded high-quality interaction matrices for diploid and polyploid pitayas (Additional file 1: Figs. S17–S19). Resolution metrics showed ≥ 80% of bins reaching 1000 contacts (Additional file 1: Fig. S20), and interaction decay curves confirmed contact decay with genomic distance (Additional file 1: Fig. S21).

Using PCA and C-score analysis, genomic regions were classified into active (A) and inactive (B) compartments, with A regions associated with higher gene density, open chromatin, and active histone marks, and B regions showing the opposite (Fig. 6a and b). Compartment statistics discovered 648 A and 728 B compartments in diploid, and 519 A and 1064 B compartments in polyploid genomes (Table 2). Moreover, gene distribution within compartments showed more genes in A compartment (e.g., 19,811 in diploid vs. 7696 in B) (Additional file 2: Table S17). Boxplots highlighted significant size differences between A and B compartments ($p < 0.0001$; Additional file 1: Fig. S22), although GC content differences were non-significant (Additional file 1: Fig. S23).

TAD structures were further mapped at 40-kb resolution using insulation scores [32], which identify TAD boundaries based on local interaction minima (Fig. 6c and f; Additional file 1: Fig. S24, Additional file 1: Fig. S25). In total, 3376 TADs were identified in the diploid, and 2031 in the polyploid pitaya (Additional file 2: Table S18), with mean lengths of 401,056 bp and 861,897 bp, respectively.

To ensure a comprehensive comparison, TAD analysis was independently performed using both the diploid and tetraploid genomes as references. To characterize TAD-associated gene distributions, genes were classified as within TADs (inter) or at TAD boundaries (border). Diploid genomes had 24,384 inter and 4045 border genes, while polyploid genomes had 20,453 and 2264 genes, respectively (Fig. 6d and e; Additional file 2: Table S12, Additional file 1: Fig. S26). Notably, no significant GC content differences were observed between TAD interiors and boundaries (Additional file 2: Table S19, Additional file 1: Fig. S27). The ability of TAD boundaries to functionally restrict the genome applied to both active and inactive genome compartments (A and B). While mammalian TADs involve structural proteins (CTCF) [33], such insulators are lacking in plants [19, 34]. However, plant transcription factors like *TEOSINTE BRANCHED 1*, *CYCLOIDEA*, and *PCF1* (*TCP1*) may play a similar role [35]. In plants (including maize, rice, sorghum, foxtail millet, and tomato), TADs mediate enhancer-promoter loops [36, 37] and are enriched for active chromatin features [38]. Using Fit-Hi-C, we identified statistically significant *cis*- and *trans*-interactions (10-kb resolution, $p \leq 0.01$). *S. undatus* exhibited 952,744 *cis*- and 4,913,539 *trans*-interactions, while *S. megalanthus* showed 9,634,251 *cis*- and 357,664 *trans*-interactions (Fig. 6g and h).

Finally, TAD boundaries were screened for enriched transcription factor motifs using JASPAR and MEME Suite. In diploid pitaya, significant enrichment was noticed for motifs such as *PHYPADRAFT_64121* (90.55%), *RAV1* (90.28%), *AT3G24120* (90.01%), *WRKY60* (89.72%), and *WRKY18* (89.6%). However, in polyploid pitaya, higher enrichment was observed as *PHYPADRAFT_64121* (91.21%), *RAV1* (90.41%), *WRKY75* (89.67%), *WRKY60* (89.61%), and *MYB3* (89.53%) motifs (Additional file 2: Table S20 and Fig. 6j), indicating their possible functional significance in both species.

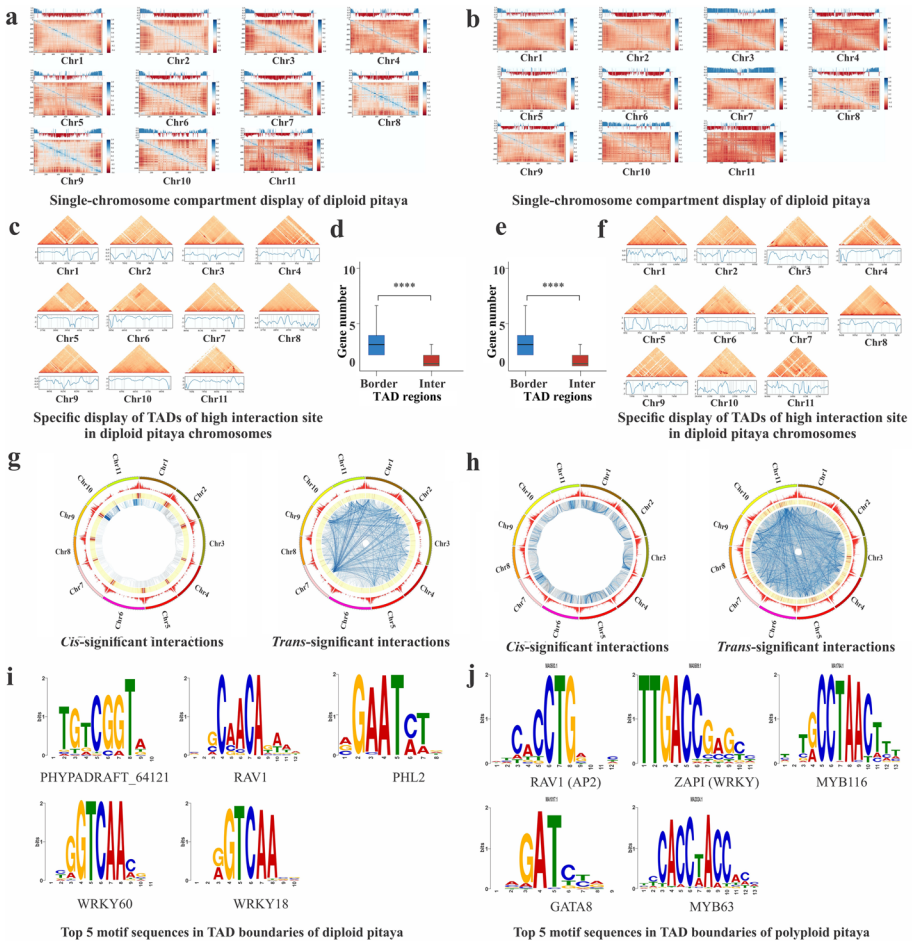

Fig. 6 3D genome structural analysis of diploid and polyploid pitaya. **a** Single chromosome compartments of *S. undatus*. **b**Single chromosome compartments of *S. megalanthus* (top panels: distribution of A/B compartment values, with blue indicating active (A) and red indicating inactive (B) regions; bottom panels: correlation matrices derived from chromosome interaction matrices). **c** TAD structures across single chromosomes in *S. undatus* at high interaction sites (top panels: interaction signals; bottom panels: insulation scores and TAD borders). Gene number distribution at TAD boundaries (border) and within TADs (inter) in diploid (**d**) and polyploid (**e**) pitaya (****$p < 0.0001$). **f**TAD structures across single chromosomes in *S. megalanthus*, similar to **c**. **g** Circos plots showing genome-wide *cis*- and *trans*-significant interaction sites in *S. undatus*. **h** Circos plots showing genome-wide *cis*- and *trans*-significant interaction sites in *S. megalanthus* (chromosome names and numbers are arranged clockwise; red bands indicate gene density; blue lines represent interaction strength with darker shades for smaller *p* values). **i** Top 5 enriched motif sequences at TAD boundaries in diploid pitaya (*S. undatus*). **j** Top 5 enriched motif sequences at TAD boundaries in polyploid pitaya (*S. megalanthus*). In **i** and **j**, the horizontal axis indicates motif length (bp); the vertical axis indicates nucleotide frequency

Genome-level chromatin accessibility of diploid and polyploid pitaya plants

Assays for transposase-accessible chromatin using sequencing (ATAC-seq) enable high-resolution mapping of open chromatin regions and *cis*-regulatory elements, thus shedding light on transcriptional regulation mechanisms [39]. In this study, we applied ATAC-seq via the sucrose precipitation method [40] and generated 0.375 billion clean reads from diploid and polyploid pitaya (Additional file 2: Table S21). Quality control included base content, base mass, GC content, and insert length distributions (Additional file 1: Figs. S28–S30, Additional file 2: Table S22). Chromatin accessibility was

Table 2 Summary of the number and length distribution of compartments

Compartments	Compartment No	Bin No	Minimum (bp)	Median (bp)	Mean (bp)	Max (bp)	Wilcoxon test
Compartment A (*S. undatus*)	648	4697	4.5692e + 04	2.0e + 05	7.24703e + 05	1.59e + 07	3.233–16
Compartment B (*S. undatus*)	728	8576	5.5212e + 04	4.0e + 05	1.177879e + 06	2.02e + 07	
Compartment A (*S. megalanthus*)	519	6114	1.0e + 05	2.0e + 05	1.177373e + 06	2.26e + 07	1.28e − 01
Compartment B (*S. megalanthus*)	1064	8834	5.5408e + 04	2.0e + 05	8.3012e + 05	2.53e + 07	

enriched around transcriptional start sites (TSSs) in both genomes (Fig. 7a). Peak calling with MACS3 software identified 43,186 peaks in diploid and 30,276 in polyploid pitaya (Additional file 2: Table S23). Replicates were assessed for overlap and peaks were visualized within ± 3 kb of TSSs (Fig. 7b and Additional file 1: Fig. S31). We also plotted signal distribution across gene bodies and TSS-proximal regions (Additional file 1: Fig. S32). After merging replicate peak sets per species, we obtained 125,321 total peaks in *S. undatus* and 30,276 in *S. megalanthus*. Among these, 13,136 peaks were shared between both species (Additional file 1: Fig. S33). The increase in total peaks in *S. undatus* reflects combined replicates and higher overall chromatin openness in this species. Differential chromatin accessibility between species was examined using DiffBind, which resulted in 13,977 differentially accessible regions (DARs), including 3355 gain DARs and 10,622 loss DARs. Clustering of DAR signal values highlighted pronounced differences between species (Fig. 7b) and annotation guided that > 5% of gain DARs and > 24% of loss DARs were located in promoter regions (Fig. 7c). Further analysis of promoter-linked DARs showed association with 159 genes for gain DARs and 2362 genes for loss DARs (*S. undatus* vs. *S. megalanthus*). GO and KEGG analysis (Additional file 1: Figs. S34–S37) indicated that gain DAR genes were predominantly enriched in "cation transport," "ATP metabolism," and "carbohydrate biosynthesis," "response to a toxic substance," "ATP synthesis," "triphosphate biosynthesis," "herbicide response," and "energy." In contrast, loss DARs were enriched in glucose pathways, circadian, stimulus–response, and photosynthesis (Fig. 7d).

Motif enrichment analysis identified 109 motifs associated with gain DARs and 425 with loss DARs. Gain DARs were enriched in *WRKY75/40*, *UIF1*, A*T5G02460*, *GATA10*, and *OBP3/4*, whereas loss DARs displayed enrichment in *ABI5*, *EmBP-1*, *TCP8*, *bZIP910*, and *AP1* (Additional file 1: Fig. S38, Additional file 1: Fig. S39). We further compared DAR-associated genes with differentially expressed genes (DEGs). Overall, 15 genes overlapped between gain DAR and upregulated DEGs (Fig. 7e), while 6 overlapped with downregulated DEGs. In contrast, 146 and 296 genes overlapped between loss DARs and up/downregulated DEGs, respectively (Additional file 1: Fig. S40). Genes associated with gain DARs showed increased expression in *S. megalanthus*, which agrees with chromatin accessibility dynamics (Fig. 7f). Furthermore, the integration of ATAC-seq and RNA-seq data enabled genome-wide visualization that confirmed the significant chromatin accessibility differences between diploid and polyploid pitaya species (Fig. 7g).

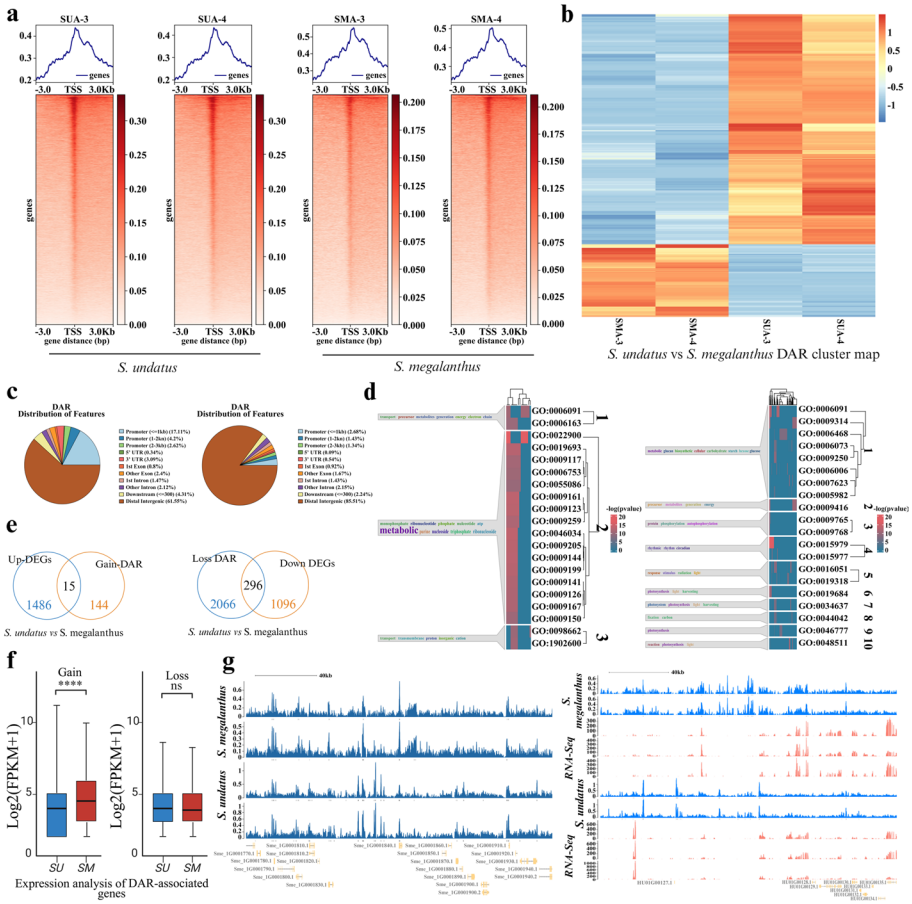

Fig. 7 ATAC-seq analysis of the whole genomes of diploid and polyploid pitaya plants. **a** ATAC-seq signal enrichment at TSS regions in S. undatus and S. megalanthus (heatmaps: 3 kb upstream/downstream of TSS on the x-axis, genes on the y-axis. Line plots: signal enrichment relative to TSS position). **b** DAR (differentially accessible regions) cluster map (X-axis: species; Y-axis: DARs; color scale represents signal intensity). **c** Distribution of the gain and loss of the DARs. The map illustrates the location of DARs on genomes and the proportion of DARs at each site. **d** GO terms in gain DAR and loss DAR-associated genes located on the promoter region. Genes are clustered according to their term, and their function is displayed on the left side. **e** Comparison of DAR-associated genes and differentially expressed genes to identify the overlapping genes in the promoter region. **f** Box plot of DAR-related gene expression. The left box plot shows significantly higher expression in S. megalanthus, indicating that genes associated with DAR gain exhibit greater transcriptional activity in S. megalanthus than in S. undatus. The right box plot shows a comparison that does not exhibit significant differences. **g** The multi-omics data displayed significant differences in chromatin accessibility between diploid and polyploid pitaya species

Comprehensive analysis of chromatin architecture in diploid and polyploid pitaya species to develop unique phenotypes

To identify tissue-specific DEGs and open chromatin regions in pitaya, we integrated the Hi-C and pericarp RNA-seq data from both pitaya species by following Liu and colleagues [41]. We categorized all 40-kb genomic bins into three groups based on sub-compartment switching and expression changes. All 11 chromosomes were classified into compartment A (open chromatin and high expression), compartment B (closed chromatin and low expression), TAD, and TAD boundaries. In yellow pitaya, 16,685 genes were found in compartment A and 5279 genes in compartment B (Additional file 2: Table S24). In contrast, in red-peel pitaya, 17,759 and 5978 genes were found in A

and B compartments, respectively (Additional file 2: Table S25). TADs and their boundaries were enriched in insulator-binding proteins. We identified 20,453 and 2264 genes in TAD and TAD boundaries in yellow pitaya (Additional file 2: Table S26), and 2438 and 4045 genes in red-peel pitaya, respectively (Additional file 2: Table S27). This methodology assisted the discovery of betalain pathway genes associated with red and yellow pigmentation. Initially, we assessed whether the "down" and "up" bins overlapped with more DEGs and examined consistency across replicates only shared chromatic features. By classifying genes based on A/B compartments, TADs, and boundaries, we linked expression profiles and genome annotations to genes that are involved in the betalain pathway.

The gene associated with betalain biosynthesis in diploid pitaya closely matched those in yellow pitaya at corresponding chromosomes, though some showed variation in copy number. Most genes were differentially expressed between diploid and polyploid species, correlating with morphological differences. Particularly, genes involved in the shikimate pathway (e.g., *HU10G00264*, *HU09G01263*, *HU07G00551*, and *HU08G02124* in red peel and *Sme_10G0002470*, *Sme_9G0013340*, *Sme_7G0004800*, and *Sme_8G0020650* in yellow peel) were identified based on sequence homology and functional annotations. These genes likely have distinct, non-redundant roles in color regulation (Fig. 8). Betalains biosynthesis involves tyrosine conversion via two enzymatic steps. Predicted genes responsible for this in diploid pitaya include *HU02G00163*, *HU10G00810*, *HU10G00173*, *HU07G00908*, *HU02G00811*, *HU05G00517*, and *HU11G01112*, and in polyploid pitaya include *Sme_2G0001630*, *Sme_10G0007740*, *Sme_10G0001580*, *Sme_7G0008480*, *Sme_2G0008230*, *Sme_5G0004900*, and *Sme_11G0010540*. Tyrosine itself is synthesized from prephenate or arogenate via prephenate dehydrogenase and arogenate dehydrogenase (ADH) genes such as *HU01G02069* and *HU07G00319* (diploid) and *Sme_1G0021320* and *Sme_7G0002630* (polyploid) [42]. Subsequently, cytochrome P450 enzyme hydroxylate tyrosine to *L-DOPA* in diploid (*HU10G01783*, *HU08G01572*, and *HU08G01571*) and polyploid (*Sme_10G0019300*, *Sme_8G0014610*, and *Sme_8G0014620*) (Fig. 8a). *L-DOPA* is oxidized by DODA to form *4,5-DOPA* [43], mediated by *HU07G00239* and *HU07G00240* (diploid) and *Sme_7G0001870* and *Sme_7G0001880* (polyploid) genes. For the synthesis of red violet betacyanin, we predicted genes. In red-peel pitaya, genes like *HU06G02516*, *HU04G00198*, *HU07G01678*, and *HU07G00475* regulate *cyclo-DOPA* conversion into red–purple betacyanin [43]. In contrast, yellow pitaya contains genes (e.g., *Sme_6G0025350*, *Sme_4G0001820*, *Sme_7G0017600*, and *Sme_7G0004100*) from the *SmeABA2*, *SmeFG2/3*, and *SmeHD-ZIP* families that condense betalamic acid with amino groups to produce yellow betaxanthins [44]. Comparative expression analysis highlighted significant differences: *shikimate* pathway genes (*Sme_10G0002470* and *Sme_7G0004800*) in yellow pitaya had threefold higher expression compared to their red-peel counterparts (*HU10G00264* and *HU07G00551*). Prephenate-related genes showed no significant expression differences between the species (Fig. 8b). Expression divergence in the betalain pathway suggests their distinct phenotypes. In the betalain pathway, we identified genes with varying degrees of expression in both pitaya species, leading to two contrasting phenotypes. We found that the gene encoding *CYP76AD1*, which regulates the enzyme oxidase that oxidizes tyrosine into L-DOPA and produces betacyanins in diploid pitaya,

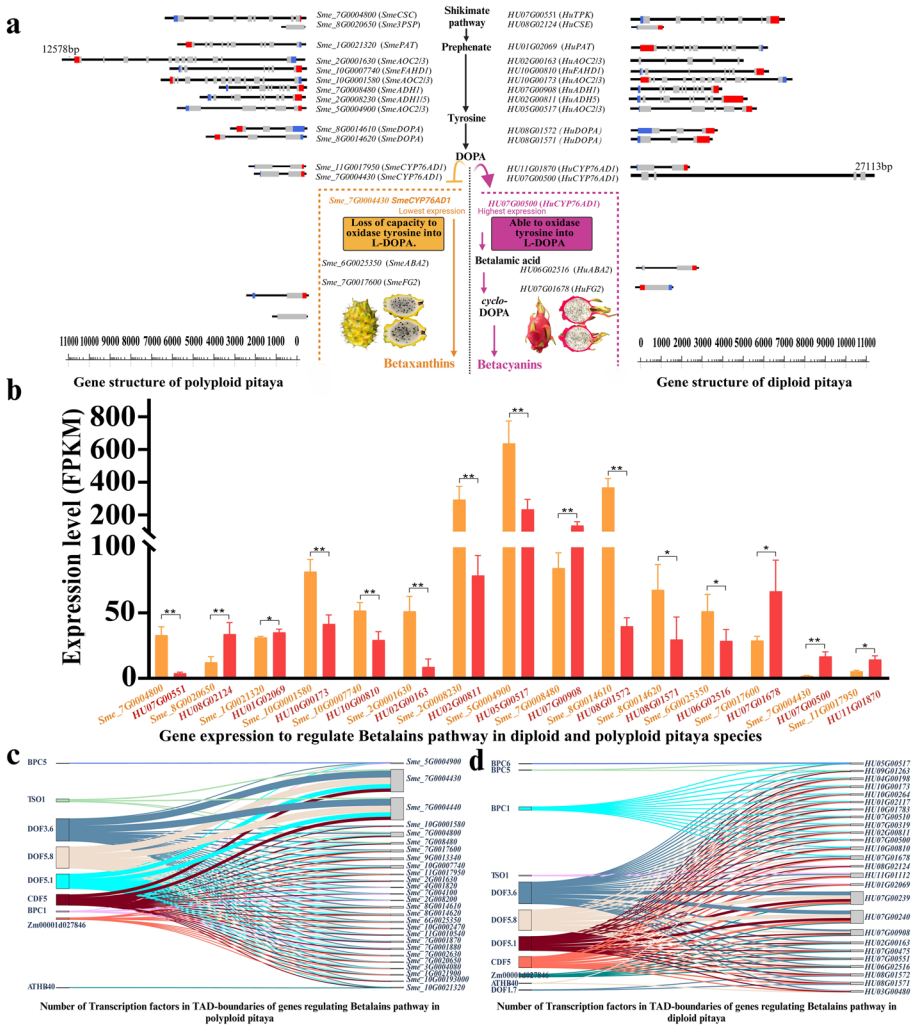

Fig. 8 Linking tissue-specific gene expression to open chromatin of diploid and polyploid pitaya. **a** Genes involved in betalain biosynthesis show high sequence similarity between yellow (*S. megalanthus*) and red-peel (*S. undatus*) pitaya genomes but differ in copy number. In *S. undatus*, genes regulate cyclo-DOPA conjugation to produce red–purple betacyanins, while in *S. megalanthus*, genes from *SmeABA2*, *SmeFG2/3*, and *SmeHD-ZIP* families promote betaxanthin production. **b** Expression levels of shikimate pathway genes (*Sme_10G0002470*, *Sme_7G0004800*) are higher in yellow pitaya compared to homologs in red pitaya. Duplication and compartmentalization of *CYP736A12* genes contributed to reduced oxidation of tyrosine into *L-DOPA* and favored betaxanthin biosynthesis in polyploid pitaya. **c** TFs identified from the 40-kb region of TAD boundaries and the top 5 TFs are presented in polyploid pitaya. **d** Top TFs identified within TAD boundaries associated with betalain biosynthesis in diploid pitaya

exhibited two fold greater expression. The gene (*HU07G00500-CYP76AD1*) encoding an oxidase-producing L-DOPA and betacyanin showed twofold higher expression than *Sme_7G0004430-CYP76AD1* in yellow pitaya (Fig. 8b). Similarly, *HU01G02117-CYP736A12* had 1.2-fold higher expression than its homolog *Sme_1G0021900-CYP736A12*. Duplication of *CYP736A12* in both compartments A and B of yellow pitaya genome suggests evolutionary divergence that may reduce red pigment synthesis.

The above observations lead to key questions: how are chromatin folding, compartmentalization, and TAD borders associated with phenotype differences? Does motif count at TAD boundaries influence gene expression? We observed major differences in

TADs and boundaries between diploid and polyploid pitaya, potentially driving transcriptional changes. We examined the gene positions within TADs and mapped the TAD boundaries on diploid and polyploid pitaya species genomes (Additional file 2: Tables S24 and S25). We mapped gene positions within TADs and identified 40-kb regions of interest in both species (Additional file 2: Tables S28–S30), highlighting regions involved in the betalain pathway. TFs from these TAD boundaries were cataloged, and motif counts were computed to identify the top 5 TFs per species (Additional file 2: Table S31). We observed that variations in motif types and counts at TAD boundaries were associated with differences in expression and pigmentation traits. We statistically analyzed gene expression within the betalain pathway using *t*-tests ($p < 0.05$) and removed the non-significant genes from further consideration (Fig. 8a). Figure 8b illustrates genes with significant expression variation that drive pigmentation differences between the two pitaya species. The *Sme_8G0014610* gene showed reduced gene expression within the TAD and a greater number of TFs at the TAD boundaries in yellow pitaya compared to red-peel pitaya (Fig. 8c). Furthermore, the *CYP76AD1* genes (*Sme_7G0004430* and *Sme_7G0004440*) displayed the highest number (same number of TFs) of TFs [*DOF3.6* (11,890), *DOF5.8* (11,650), *DOF5.1* (7780), *CDF5* (5605), and *TSO1* (2220)]. Still, they showed the lowest gene expression in yellow pitaya compared to red-peel pitaya (Fig. 8d). This difference may explain why these genes have lost the ability to oxidase tyrosine and produce betaxanthins. These findings indicate that TAD boundaries act as regulators and repressors simultaneously and mechanistically control genes within TADs. Furthermore, we found that the *Hu01G02117* gene of diploid pitaya is in compartment A of the genome. Conversely, the counterpart gene (*Sme_1G0021900*) is located in compartment A and compartment B of polyploid pitaya. This might be a reason for the low expression of *CYP736A12* genes in polyploid pitaya, which led to the loss of the capacity to oxidize tyrosine and ultimately produce yellow fruit.

Discussion

Autopolyploidy and WGDs in yellow pitaya may explain phenotypic and metabolic variations, such as changes in fruit color, cladode thickness, fruit size, and seed size. Following polyploidization, chromosome number, genome size, gene count, and expression evolve, accelerating plant diversification and novel trait development. Gene redundancy often originates from unreduced gametes in diploids, which are frequently induced by temperature fluctuation in plants [45–47], fish, and amphibians [48]. These gametes are regulated by heritable genes whose expression increases under environmental stresses. Grafting has also played a role in polyploid evolution, particularly in the Cactaceae family, where nuclear genomes can move from rootstock to scion, potentially producing polyploid cells and fertile shoots [49]. Yellow pitaya may represent an "interracial autopolyploid" formed through crossing genetically divergent yet structurally similar populations. These genomic changes enhance morphology, physiology, and ecology, which guide speciation and environmental adaptations. Multiple maternal lineages further increase genetic diversity [50]. Polyploids benefit from multiple alleles per locus, thereby enhancing stress tolerance and persistence across diverse environments [50]. Moreover, gene dosage effects from duplicated genomes support robust regulatory networks [51]. Domestication may drive gene gain or loss, e.g., *GA20-OX* loss-of-function

mutations trigger dwarf wheat and the Green Revolution. In synthetic autopolyploid of *Phlox drummondii*, 17–25% DNA was lost by the third generation, improving the seed set (30–66%) and stabilizing tetraploids [30]. However, further studies are required to characterize genome loss/gain patterns during pitaya domestication and identify functional vs. cryptic genes. Vegetative reproduction, perennial growth, and outcrossing mating systems may also contribute to autopolyploid formation [52, 53].

Betalains in diploid and polyploid pitaya have accelerated research into pigment evolution and biosynthesis. Plant pigmentation, particularly in *Caryophyllales*, uniquely produces either betalains or anthocyanins but never both. To understand pigment differences between both pitaya species, we analyzed 3D chromatin structures (loops, TADs, A/B compartments, and chromosome territories) using insulation scores. We identified 3376 TADs in diploid and 2031 in polyploid genomes, with mean lengths of 401,056 bp and 861,897 bp, respectively. TADs organize genomic functions and regulate gene expression patterns [35]. Intra-TAD interactions are more frequent than inter-TAD, and boundaries prevent the spread of activity [19]. Here, we observed A/B compartment switching and TAD reorganization between diploid and polyploid species. Repeatability analysis documented 626 shared TAD boundaries, with 1395 and 1454 unique to diploid and polyploid genomes, respectively. We identified 24,384 and 24,775 genes inside TADs, and 4045 and 2162 at TAD boundaries in diploid and polyploid genomes, respectively. TADs in plants lack *CTCF* but may be regulated by TCP TFs (e.g., *TEOSINTE BRANCHED 1*, *CYCLOIDEA*, and *PCF1*) [19, 34, 35]. Furthermore, TAD boundary motif enrichment in diploid included *PHYPADRAFT_64121* (90.55%), *RAV1* (90.28%), *AT3G24120* (90.01%), *WRKY60* (89.72%), and *WRKY18* (89.6%), while polyploid boundaries featured *PHYPADRAFT_64121* (91.21%), *RAV1* (90.41%), *WRKY75* (89.67%), *WRKY60* (89.61%), and *MYB3* (89.53%) motifs. Higher-order chromatin loops connect enhancers and promoters, as seen in several plant species (e.g., maize, rice, sorghum, foxtail millet, and tomato), often characterized by active genes and open chromatin [36–38]. Plant TAD boundaries are enriched in *TCP* and *bZIP* TF binding sites, indicating their key role in TAD formation [54]. In *Arabidopsis*, loop disruption affects *FLC* gene expression [55], and in *Marchantia polymorpha*, TCP1 loss alters expression in *TCP1-rich* TADs without changing chromatin patterns, indicating *TCP1* may not be essential for TAD formation [56]. In rice and pepper, TCP-enriched GC-rich motifs and gene-to-gene loops at TADs regulate gene expression and spatial clustering [57, 58].

Differential gene expression between diploid and polyploid pitaya correlates with fruit morphology differences. We identified shikimate pathway genes for betalain biosynthesis in red peel (*HU10G00264*, *HU09G01263*, *HU07G00551*, and *HU08G02124*) and yellow peel (*Sme_10G0002470*, *Sme_9G0013340*, *Sme_7G0004800*, and *Sme_8G0020650*) pitaya. We also noticed that motif variations at TAD boundaries appear to influence TF binding, gene expression, and, ultimately, peel coloration. TFs identified from boundaries support this regulatory role. However, an important question remains: do these genes function individually, or do they act in coordination as regulatory modules? To address and understand this gap, future studies should focus on functionally validating the genes and associated motifs and TFs to examine their precise role in peel color and other trait improvements, such as stress tolerance [59] by genome engineering [60], preferably by mobile CRISPR [61]. Given the challenges in engineering the pitaya genome,

such integrated approaches may deliver new insights into regulatory mechanisms in pitaya or other model plants. Therefore, this study reports the first chromosome-level reference genome for yellow pitaya, assembled using PacBio HiFi-Seq, with additional Hi-C, ATAC-seq, and RNA-seq data for both (diploid and polyploid) species.

Conclusions

Our study presents the first chromosome-level genome assembly of autotetraploid yellow pitaya that provides a foundational reference and insights into how polyploidization and 3D chromatin reorganization (e.g., TAD boundary shifts and A/B compartment switching) may drive differential gene expression, particularly in the betalain biosynthesis pathway, leading to phenotypic divergence between diploid and polyploid pitaya species. The enrichment of transcription factor motifs at TAD boundaries highlights their potential role in regulating trait-specific gene networks. However, the diploid pitaya genome has fast-tracked trait improvement efforts [3, 27]. The lack of a polyploid reference genome has hindered research on *S. megalanthus*, and this high-quality assembly now addresses this gap. Therefore, we suggest that future work should functionally validate candidate genes and associated regulatory modules to discover their collective impact on peel color of pitaya species and stress adaptation. This resource provides a foundation for exploiting genomic architecture in pitaya breeding that offers insights into polyploid crop improvement and sustainable horticultural innovation.

Methods

Plant material

Plant material of yellow pitaya [*Selenicereus megalanthus*, formerly *Hylocereus megalanthus* (K. Schum. Ex Vaupel) Ralf Bauer] and red peel-white pulp pitaya [*Selenicereus undatus*-Britton & Rose, previously known as *Hylocereus undatus*] were collected from Hainan-Shengda Modern Agriculture Development Company (latitude/longitude 19°N/110°E), Qionghai, Hainan, China.

DNA extraction, library preparation, and quality control

Young cladode from yellow pitaya were collected for the DNA nano Ball (DNB) sequence survey. Total genomic DNA was isolated from fresh cladodes using the standard cetyltrimethylammonium bromide (CTAB) method and processed on the DNBSEQ™ platform (BGI, zero-PCR protocol). DNA was randomly sheared to ~ 350 bp fragments using a Covaris ultrasonic system. These fragments were end-repaired, A-tailed, ligated to adaptors, and circularized by rolling circle amplification (RCA) to generate DNBs. Sequencing was performed on the BGI DNBSEQ™ platform. The library was sequenced, and the raw data were obtained in FASTQ format. Using fastp-v0.20.04 software, clean reads were generated after removing the adapter, and low-quality reads using Trimmomatic software (v 0.38) [62]. We used multiple *K*-mers for the analysis. The *K*-mer frequency was rapidly assessed by Jellyfish-(v2.2.6) software (http://www.cbcb.umd.edu/software/jellyfish) [63] and genome size, heterozygosity, repeat sequences, and sequence depth were evaluated using GenomeScope2 software (https://github.com/tbenavi1/genomescope2.0) [64].

PacBio HiFi sequencing

High-molecular-weight genomic DNA from *S. megalanthus* cladodes was fragmented using a Megaruptor to an average size of 15–20 kb. SMRTbell libraries were prepared by ligating adapters and removing linear/damaged DNA, followed by size selection using BluePippin. The libraries were sequenced on the PacBio Sequel II platform in CCS mode.

Chromosome conformation capture (Hi-C) library construction and sequencing

Hi-C libraries were constructed for both diploid (*S. undatus*) and polyploid (*S. megalanthus*) pitayas to explore chromatic interactions [65, 66]. Young cladodes were fixed with 1% formaldehyde to cross-link DNA following the protocols of Wuhan-Fraser Gene Information Co., Ltd. Following cell lysis, the cross-linked chromatin adjacent sites were digested using the specific restriction endonuclease enzyme DpnII. The sticky ends of nucleic acids generated by DpnII were covalently labeled with biotin-dCTP, producing blunt-ended repaired DNA strands. DNA was physically fragmented, and biotin-dCTP-labeled fragments were purified using streptavidin beads. T4 DNA polymerase removed the terminal alkaloid tag from the unlinked fragments. The resulting libraries were quality-checked using Qubit 2.0 (determine the library concentration) and an Agilent 2100 (detect the integrity of the DNA fragments and insert size) and sequenced on the Illumina HiSeq (BGI PE150) platform. Hi-C reads were filtered using Trim Galore (v0.6.7) (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and mapped using Burrows–Wheeler alignment algorithm (BWA) [67], and yellow pitaya Hi-C reads were aligned with the newly assembled reference genome of polyploid yellow pitaya (*S. megalanthus*). HiC-Pro (v2.8.0) was used to identify uniquely aligned paired-end reads [68]. Reads from *S. undatus* were aligned to its published reference genome [27, 69], while *S. megalanthus* reads were aligned to the newly assembled genome.

ATAC sequencing

An assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) was carried out for both species of pitaya, diploid (*S. undatus*) and polyploid (*S. megalanthus*), to analyze the chromatin regions accessible to the Tn5 transposase [70]. The sucrose precipitation method was used for the ATAC-seq experiment [40]. Young cladode tissues from diploid and polyploid pitaya species were prepared for cell suspension to isolate the nucleus. Tn5 transposase exclusively cleaves open chromatin and simultaneously adds sequencing adapters to both ends of the cleaved DNA fragments. The digested DNA fragments were retrieved and amplified by PCR, and the library size was selected by a gel-free double-sided size-selection protocol using Agencourt AMPure XP beads (Cat. 63,881) at 0.5 × and 1.2 × concentration. Libraries were sequenced using next-generation high-throughput sequencing (Illumina PE150). The raw data were filtered using Trimmomatic [62]. Clean reads of diploid pitaya were aligned with the reference genome (http://www.pitayagenomic.com/download.php) [27, 69], and clean reads of polyploid pitaya were aligned with the reference genome (newly assembled genome of *Selenicereus megalanthus*) using Bowtie 2 (v 2.2.6). For differential accessibility analysis, DiffBind with DESeq2 was used to compare chromatin accessibility between the two species. Replicate consistency was ensured by assessing overlap

between biological replicates prior to merging, and merged peak sets were used for species-level DAR detection. Significant DARs were defined using log2FoldChange > 0.58 and FDR < 0.05. All processing steps followed the pipeline described by Zhang et al. [71].

Transcriptome sequencing

Young-developing cladodes (30-day-old stems), fruit pericarp (when the fruit turns its color), and pulp (before maturity) were sampled from both pitaya species (*S. undatus* and *S. megalanthus*) for transcriptome sequencing. Total RNA was extracted from the stem, pericarp, and fruit pulp samples. The concentration and purity of the RNA were detected with a Nanodrop 2000, and the RNA integrity number (RIN) was determined with an Agilent 2100 spectrophotometer. A NanoDrop spectrophotometer was used to determine the purity (OD260/280), concentration, and normal nucleic acid absorption peak. A total of 2 μg of RNA was added at a concentration of ≥ 300 ng/μL, with the OD260/280 ranging from 1.8 to 2.2. Raw reads were generated using the BGI Illumina platform.

Quality assessment, reference genome assembly, and chromosome construction

The genome size of *Selenicereus megalanthus* was estimated to be approximately 3.32 Gb using a 21-mer frequency distribution curve based on 194.17 Gb of clean BGI short reads. GenomeScope2 and Smudgeplot analysis confirmed its autotetraploid nature, with AAAA-type *k*-mer pairs accounting for 97% (Additional file 1: Figs. S1 and S2). High-fidelity (HiFi) long-read data were assembled using Hifiasm (v0.19.6), a haplotype-resolved assembler optimized for HiFi reads (https://github.com/chhylp123/hifiasm). To reduce redundancy caused by heterozygous contigs, Purge Haplotigs (v1.0.4) was used to identify and remove redundant sequences based on read depth, distribution, and sequence similarity [23] that yielded a high-quality haploid assembly of 1.79 Gb with a contig N50 of 13.21 Mb (Additional file 2: Table S32).

Short reads were aligned to the genome using BWA (v0.7.17), and long reads with Minimap2 (v2.24) [67, 72]. Mapping results were processed with SAMtools [73] to generate sorted and indexed BAM files, and Picard Toolkit [2019, Broad Institute, GitHub Repository (https://broadinstitute.github.io/picard/)] was used to mark duplicates. SNPs and InDels were identified using the GATK Toolkit (v4.4.0.0) with recalibration and filtering to reduce false positives. Genome assembly quality was evaluated based on contiguity, completeness, and correctness (3-Cs). Contiguity was assessed by contig N50; completeness was measured with BUSCO (v0.2.4) [74], yielding scores above 95%. Hi-C data were used to scaffold the contigs onto chromosomes. The final assembly achieved a genome mapping rate of 96.96% and a coverage rate of 99.81%. Genome-wide features, including read depth, SNP density, and repeat distribution were visualized using Circos (v0.69–6) (Additional file 1: Fig. S4).

S. megalanthus gene annotation

Repeat sequences were predicted using TRF, PILER (http://www.drive5.com/piler) [75], LTR Finder [76], and RepeatModeler (https://www.repeatmasker.org/), in combination with Repbase. Gene prediction was performed using a combination of de novo,

homology-based, and transcriptome-supported methods. De novo predictions were conducted with GlimmerHMM [77].

Homologous gene models were derived from related species, including *S. undatus* (http://www.pitayagenomic.com/download.php) [27], *Carnegiea gigantea* (GCA_029747015.1_UA_SGP5p_2) [78], *Arabidopsis thaliana* (GCA_000001735.1_TAIR10) [79], *Oryza sativa* (GCF_001433935.1_IRGSP-1.0) [80], *Solanum lycopersicum* (GCF_000188115.5_SL3.1), and *Solanum tuberosum* (GCF_000226075.1_SolTub_3.0) [81]. Transcriptome data from stem, pericarp, and pulp tissues were aligned with HISAT2 [82] and assembled with StringTie [83]. PacBio full-length transcripts were obtained using ISOseq3, and coding regions were identified using TransDecoder (https://github.com/TransDecoder/TransDecoder).

All gene evidence was integrated using MAKER2 (v2.31.11) [84], and further refined with AUGUSTUS (v3.4.0) [85] and GeneMark-EP + (v4.6.3) (http://topaz.gatech.edu/GeneMark/license_download.cgi) [84, 86]. Final gene models were filtered to remove transposon-related domains and redundant sequences. Functional annotation was performed by comparing predicted proteins to databases including SWISS-PROT [87], InterProScan [88], Gene Ontology [89], KEGG [90], and KOBAS-I [91], using a cutoff value of $1e^{-5}$ to assign gene functions, pathways, and structural domains (Additional file 2: Table S33).

Hi-C heatmaps of red (diploid) and yellow (polyploid) pitaya

Clean data from polyploid yellow pitaya were aligned to the newly constructed reference genome, and diploid white-fleshed pitaya Hi-C data were aligned to the public reference genome (http://www.pitayagenomic.com/download.php, accessed December 13, 2023) using Juicer software [92]. Messy reads, including single-ended, low-quality, and unaligned reads, were filtered out. For Hi-C analysis, valid interaction pairs were retained, while self-circles, dangling ends, and PCR duplicates were discarded. Library quality was assessed based on the proportion of valid pairs in non-PCR replicates. After filtering, valid pairs (Q30) were used for interaction matrix construction and downstream analysis (Additional file 2: Table S34). The 3D genome structure was reconstructed by aligning paired-end reads across genomic loci. Hi-C data correlation between replicates was analyzed using GenomeDISCO [93]. The reference genome was binned, and filtered reads were assigned to bins to construct interaction matrices. To correct biases due to GC content, enzyme sites, and fragment sizes, matrices were normalized with Juicer [92]. Genome-wide interaction matrices at 100-kb resolution and single-chromosome matrices at 40-kb resolution were generated for diploid and polyploid species. Differences in genome interactions were analyzed by converting matrices into Z-score matrices [32], subtracting diploid and polyploid Z-scores, and displaying the results as heatmaps at 200-kb resolution. Interaction frequency and distance were also compared between genomes [94, 95].

Compartment identification in both pitaya species

Genome compartments were investigated to identify potential links within a single chromosome after constructing an interaction map [96]. PC1 values at 100-kb resolution were obtained via the bin C-score method [97], and all bins in the reference genomes of

polyploid yellow and diploid red pitaya were divided into active (compartment A) and inactive (compartment B) regions. Positive and negative bins were designated as compartments A and B to compute their distribution across chromosomes. Consecutive bins of the same compartment were merged to calculate the number and length distributions of compartments genome-wide. The distributions were visualized using boxplots, and the significance of length differences was assessed with Wilcoxon's rank-sum test [98]. The number of genes per bin in compartments A and B was counted, and distributions were displayed using box plots. The significance of gene number differences between compartments was also tested. To explore the association between GC content and compartment structure [99, 100], GC content per bin was measured, visualized via boxplots, and tested for significance by the Wilcoxon test [98].

TAD structural analysis, classification of genes, and motif analysis in red and yellow pitaya species

Topologically associating domain (TAD) boundaries were calculated using the insulation score method with cworld-dekker [32], reflecting interactions between the two sides of each bin. Interactions were recognized as TAD boundaries. TADs were identified with "cooltools" at a 40-kb resolution, aligning boundary definitions with the bin size. Motif enrichment analysis was performed with "FIMO" by scanning 40-kb TAD boundary bins for motifs. To normalize, a uniform 40-kb bin size was used across the genome, mitigating biases due to varying boundary lengths. All genome bins were classified as either TAD boundaries (border) or within-TAD regions (inter). Gene numbers in each bin type were counted, and distributions were visualized using boxplots. The significance of gene number differences was assessed by Wilcoxon's test [98]. The number of TAD boundaries containing motifs was quantified, and the top five motifs were reported.

Gene function annotation in differential compartment regions

Gene functional annotations were performed by aligning the assembled genome against the Nr (non-redundant) protein database using DIAMOND [101]. DIAMOND internally applies statistical measures, including the *e* value. The expected value of the hit quantifies the number of alignments of comparable or higher quality that we expect to find searching this query against a database of random sequences, the same size as the actual target database. By default, DIAMOND reports all alignments with *e* value < 0.001. The gene sets within the compartment region was functionally annotated, and the 20 most significantly enriched genes were identified and displayed on GO and KEGG scatter plots.

Analysis of significant interaction sites in diploid and polyploid pitaya species

At 10-kb resolution, Fit-Hi-C software (v1.0.1) (L resolution) was employed to examine the significant interactions between the two loci within the genome of the pitaya species [25]. The significant cis and trans interaction sites were identified based on the number of reads supporting the interaction with a p value ≤ 0.01 and q value ≤ 0.01. The corresponding cumulative probability and interacting loci were sorted. A p value and q value < 0.01, along with a read count > 2, were considered to indicate significant interactions [39].

Comparative genome structural variation and genome evolution analysis

Comparative genome alignment was performed using Nucmer (MUMmer4) with "-c 40 -l 90" parameters [102]. Alignments were filtered using delta-filter "−1," and SNPs/InDels (1–50 bp) were identified using Show-SNPs (-ClrTH). Structural variants, including inversions, translocations, CNVs, and PAVs, were detected by SyRI (v1.6.3) [103] and annotated with ANNOVAR [104]. For yellow pitaya genome evolutionary analysis, orthologous gene families of *S. megalanthus*, a closely related cactus species [*S. undatus* (http://www.pitayagenomic.com/download.php), *Carnegiea gigantea* (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_029747015.1/, GCA_029747015.1_UA_SGP5p_2)] and plant species, including *Beta vulgaris* (https://phytozome-next.jgi.doe.gov/info/Bvulgaris_EL10_1_0, Bvulgaris_548_EL10_1.0), *Spinacia oleracea* (https://phytozome-next.jgi.doe.gov/info/Soleracea_Spov3, Soleracea_575_Spov3), *Sedum album* (https://genomevolution.org/CoGe/SearchResults.pl?s=Sedum%20album&p=genome, V3), *Arabidopsis thaliana* (https://phytozome-next.jgi.doe.gov/info/Athaliana_TAIR10, Athaliana_167_TAIR10), *Kalanchoe fedtschenkoi* (https://phytozome-next.jgi.doe.gov/info/Kfedtschenkoi_v1_1, Kfedtschenkoi_382_v1.1), *Oryza sativa* (https://phytozome-next.jgi.doe.gov/info/Osativa_v7_0, Osativa_323_v7.0), and *Cannabis sativa* (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_029168945.1/, GCF_029168945.1_ASM2916894v1), were obtained from OrthoMCL (V1.4, http://orthomcl.org/orthomcl/). A super alignment matrix was generated by aligning the single-copy families with MUSCLE (v.3.8.31) and following the methodologies reported in the literature [105–107].

Synonymous substitution rates (*Ks*) for paralogs in *S. megalanthus* and orthologs between *S. undatus* and *S. megalanthus* were estimated with MCScanX software (http://chibba.pgml.uga.edu/mcscan2/). Selection pressure (*Ka*/*Ks*) was assessed, and WGDI software was used to compute and filter *Ks* values for detecting whole-genome duplication (WGD) events. Single-copy families were also aligned using MUSCLE (v3.8.31, http://www.drive5.com/muscle/) for nucleic acid difference analysis, and the single-base difference rate was calculated with vcf2dis as 0.010014.

By distinguishing the synonymous substitution (*Ks*) and nonsynonymous substitution (*Ka*) rates, we assessed the selection pressure at the protein level, typically equal to 1 ($Ka/Ks = 1$). We used WGDI software to compute the *Ks* values, filtered the data based on collinear blocks (to reduce the error caused by tandem repeats), and analyzed the distribution of the *Ks* frequency to determine the number of peaks for observing the WGD events.

We selected single-copy gene families for alignment using MUSCLE (v3.8.31, http://www.drive5.com/muscle/). To identify single-base differences, we used vcf2dis software for each gene family, and the nucleic acid difference rate was calculated to be 0.010014.

Linking Hi-C data with RNA-seq data

To identify whether sub-compartment switching correlates with changes in expression levels, we classified all 40-kb genomic bins into three groups based on changes in the sub-compartment of pitaya pericarp tissues. To link the Hi-C data with the RNA-seq data, we followed previous studies [41, 58] and approaches to explore the reorganization

of chromatin domains (TADs) and loops are correlated with changes in expression levels. We correlated this classification of TAD features and loops with changes in gene expression profiles, including DEGs and fold changes. We conducted a *t*-test to compare the two groups based on p value < 0.05. If any gene group between yellow and red pitaya showed a non-significant difference, we removed that one and contained only those that showed a significant difference at a significance level < 0.05.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-025-03695-3.

Additional file 1: Figures S1–S40.

Additional file 2: Tables S1–S34.

Acknowledgements
We thank Hainan-Shengda Modern Agriculture Development Company, Hainan, China, for providing plant materials. We also thank two anonymous reviewers and our handling editors for providing critical and helpful feedback, which helps us improve our manuscript. Thanks to BGI Tech Solutions Co., Ltd. and Wuhan Frasergen Bioinformatics Co., Ltd. for sequencing and providing technical support. Biorender was used to compile panels and improve the figures for the readers.

Peer review information
Eduard Akhunov and Wenjing She were the primary editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Authors' contributions
Q.U.Z., M.F.N., H.-F.W., and R.K.V. designed the research. Q.U.Z., V.G., M.F.N. M.I., D.K., L.H., A.C., A.A.K., S.K., M. G., G. D., Y-H. Q. and B.U. analyzed the data. W. L., AR, C.J., G.W., X.Y., S.Z., Z.Y., P.L., and Z.X.Z. generated the plant materials. H.-F.W., R.K.V. supervised the whole-genome sequencing work. Q.U.Z. wrote the article with input from the other authors. All authors reviewed the manuscript.

Authors' X and Wechat handles
Qamar U. Zaman: @qamaruzaman1 (X), qamaruzamanch (Wechat); Ali Raza: Ale_Raza6 (X), Ali_Raza6 (Wechat); Rajeev K. Varshney: @rajvarshney (X), varshney_raj (Wechat); Hua-Feng Wang: Sargentodoxa (Wechat).

Funding
This study was funded by the Hainan Province Science and Technology Special Fund (ZDYF2022XDNY190), the Project of Sanya Yazhou Bay Science and Technology City (Grant Number: SCKJ-JYRC-2022–83), the Innovational Fund for Scientific and Technological Personnel of Hainan Province, the Collaborative Innovation Center of Nanfan and High-Efficiency Tropical Agriculture, Hainan University Funding (XTCX2022NYB09), and the Hainan Provincial Natural Science Foundation of China (421RC486). This study was also supported by the Food Futures Institute of Murdoch University to RKV.

Data availability
All PacBio, ATAC, and RNA-Seq raw sequence data have been deposited in the National Center for Biotechnology Information (NCBI) (SRA database under BioProject accession PRJNA1117350) [108] and Hi-C data are available at the China National Center for Bioinformation-CNCB (Under GSA database Bioproject ID PRJCA026647) [109]. Reference genome assembly and annotation files of Selenicereus megalanthus-Sme (Sme.fasta.gz, Sme gene annotation.xls, Sme annotation_gene_transcript.xls, Sme.gene.cds, Sme.gene.gff and Sme.gene.pep) are available at https://doi.org/10.6084/m9.figshare.25943530 [110]. No custom scripts and software were used besides those mentioned in the Methods section.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
Mengqi Guo is affiliated with Wuhan Frasergen Bioinformatics Co., Ltd. The author declares that this affiliation did not influence the study design, data collection, analysis, decision to publish, or preparation of the manuscript. All other authors declare no competing interests.

Author details
[1]School of Breeding and Multiplication (Sanya Institute of Breeding and Multiplication), Hainan Yazhou Bay Seed Laboratory, Hainan University, Sanya 572025, China. [2]Collaborative Innovation Center of Nanfan and High-Efficiency Tropical Agriculture, School of Tropical Crops, Hainan University, Haikou 570228, China. [3]College of Tropical Agriculture

and Forestry, Hainan University, Haikou 570228, China. [4]Guangdong Key Laboratory of Plant Epigenetics, College of Life Sciences and Oceanography, Shenzhen University, Shenzhen 518060, China. [5]Jiangxi Provincial Key Laboratory of Ex-Situ Plant Conservation and Utilization, Lushan Botanical Garden, Chinese Academy of Sciences, Jiujiang, Jiangxi, China. [6]Centre for Crop and Food Innovation, WA State Agricultural Biotechnology Centre, Murdoch University, Murdoch, WA 6150, Australia. [7]Wuhan Frasergen Bioinformatics Co., Ltd, , Wuhan, Hubei, China. [8]Harry Perkins Institute of Medical Research, The University of Western Australia, Murdoch, WA 6150, Australia. [9]Department of Cardiology, Fiona Stanley Hospital, Murdoch, WA 6150, Australia. [10]Key Laboratory of Biology and Genetic Improvement of Horticultural Crops (South China), College of Horticulture, South China Agricultural University, Guangzhou, Guangdong 510642, China. [11]Present Address: Yazhouwan National Laboratory, 572025 Sanya, China.

Received: 27 June 2024 Accepted: 11 July 2025
Published online: 06 August 2025

References

1. Hershkovitz MA, Zimmer EA. On the evolutionary origins of the cacti. Taxon. 1997;46:217–32.
2. Khan D, Harris A, Zaman QU, Wang H-X, Wen J, Landis JB, Wang H-F. The evolutionary history and distribution of cactus germplasm resources, as well as potential domestication under a changing climate. J Syst Evol. 2024;62(5):858–75.
3. Zheng J, Meinhardt LW, Goenaga R, Zhang D, Yin Y. The chromosome-level genome of dragon fruit reveals whole-genome duplication and chromosomal co-localization of betacyanin biosynthetic genes. Hortic Res. 2021;8:63.
4. Tel-Zur N. Breeding an underutilized fruit crop: a long-term program for Hylocereus. Hortic Res. 2022;9:uhac078.
5. Tenore GC, Novellino E, Basile A. Nutraceutical potential and antioxidant benefits of red pitaya (*Hylocereus polyrhizus*) extracts. J Funct Foods. 2012;4:129–36.
6. 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 Technol. 2021;116:199–217.
7. Abirami K, Swain S, Baskaran V, Venkatesan K, Sakthivel K, Bommayasamy N. Distinguishing three dragon fruit (*Hylocereus* spp.) species grown in Andaman and Nicobar Islands of India using morphological, biochemical and molecular traits. Sci Rep. 2021;11:2894.
8. Wu L-c, Hsu H-W, Chen Y-C, Chiu C-C, Lin Y-I, Ho J-aA. Antioxidant and antiproliferative activities of red pitaya. Food Chem. 2006;95:319–27.
9. Paśko P, Galanty A, Zagrodzki P, Luksirikul P, Barasch D, Nemirovski A, Gorinstein S. Dragon fruits as a reservoir of natural polyphenolics with chemopreventive properties. Molecules. 2021;26:2158.
10. Watanabe K. Bringing stability to world food supplies with Japanese technology and Mexican genetic resources. In Crop production under stressful conditions: application of cutting-edge science and technology in developing countries. Kokubun M, Asanuma S, editors. Singapore: Singapore: Springer; 2018;177–193
11. Morillo-Coronado AC, Manjarres-Hernández EH, Saenz-Quintero ÓJ, Morillo-Coronado Y. Morphoagronomic evaluation of yellow pitahaya (*Selenicereus megalanthus* Haw) in Miraflores, Colombia. 2022;12:1582.
12. Gandía-Herrero F, García-Carmona FJ. The dawn of betalains. New Phytol. 2020;227:664–6.
13. Arakaki M, Christin P-A, Nyffeler R, Lendel A, Eggli U, Ogburn RM, Spriggs E, Moore MJ, Edwards EJ. Contemporaneous and recent radiations of the world's major succulent plant lineages. Proceed Natl Acad Sci USA. 2011;108:8379–84.
14. Britton NL, Rose JN. The Cactaceae: descriptions and illustrations of plants of the cactus family. Courier Corporation; 1963.
15. Moran RJGH. *Selenicereus megalanthus* (Schumann) Moran. 1953;8:325.
16. Bauer R. A synopsis of the tribe Hylocereeae F. Buxb. Hunt; 2003.
17. Plume O, Straub SC, Tel-Zur N, Cisneros A, Schneider B, Doyle JJ. Testing a hypothesis of intergeneric allopolyploidy in vine cacti (Cactaceae: Hylocereeae). Syst Bot. 2013;38:737–51.
18. 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.
19. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012;485:376–80.
20. Tel-Zur N, Abbo S, Bar-Zvi D, Mizrahi Y. Genetic relationships among *Hylocereus* and *Selenicereus* vine cacti (Cactaceae): evidence from hybridization and cytological studies. Ann Bot. 2004;94:527–34.
21. Tel-Zur N, Dudai M, Raveh E, Mizrahi Y. In situ induction of chromosome doubling in vine cacti (Cactaceae). Sci Hortic. 2011;129:570–6.
22. Lichtenzveig J, Abbo S, Nerd A, Tel-Zur N, Mizrahi Y. Cytology and mating systems in the climbing cacti *Hylocereus* and *Selenicereus*. Am J Bot. 2000;87:1058–65.
23. Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics. 2018;19:460.
24. Wang M, Wang P, Lin M, Ye Z, Li G, Tu L, Shen C, Li J, Yang Q, Zhang X. Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. Nat Plants. 2018;4:90–7.
25. Ay F, Bailey TL, Noble WS. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. Genome Res. 2014;24:999–1011.
26. Schiessl S-V, Katche E, Ihien E, Chawla HS, Mason AS. The role of genomic structural variation in the genetic improvement of polyploid crops. Crop J. 2019;7:127–40.

27. Chen J-y, Xie F-f, Cui Y-z, Chen C-b, Lu W-j, Hu X-d, Hua Q-z, Zhao J, Wu Z-j, Gao D, et al. A chromosome-scale genome sequence of pitaya (*Hylocereus undatus*) provides novel insights into the genome evolution and regulation of betalain biosynthesis. Hortic Res. 2021;8:164.
28. Van Drunen WE, Husband BC. Evolutionary associations between polyploidy, clonal reproduction, and perenniality in the angiosperms. New Phytol. 2019;224:1266–77.
29. Otto SP. The evolutionary consequences of polyploidy. Cell. 2007;131:452–62.
30. Raina SN, Parida A, Koul KK, Salimath SS, Bisht MS, Raja V, Khoshoo TN. Associated chromosomal DNA changes in polyploids. Genome. 1994;37:560–4.
31. Ozkan H, Tuna M, Galbraith D. No DNA loss in autotetraploids of *Arabidopsis thaliana*. Plant Breed. 2006;125:288–91.
32. Crane E, Bian Q, McCord RP, Lajoie BR, Wheeler BS, Ralston EJ, Uzawa S, Dekker J, Meyer BJ. Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature. 2015;523:240–4.
33. Alipour E, Marko JF. Self-organization of domain structures by DNA-loop-extruding enzymes. Nucleic Acids Res. 2012;40:11202–12.
34. Szabo Q, Bantignies F, Cavalli G. Principles of genome folding into topologically associating domains. Sci Adv. 2019;5:eaaw1668.
35. Tourdot E, Grob S. Three-dimensional chromatin architecture in plants-general features and novelties. Eur J Cell Biol. 2023;102: 151344.
36. Doğan ES, Liu C. Three-dimensional chromatin packing and positioning of plant genomes. Nat Plants. 2018;4:521–9.
37. Li E, Liu H, Huang L, Zhang X, Dong X, Song W, Zhao H, Lai J. Long-range interactions between proximal and distal regulatory regions in maize. Nat Commun. 2019;10:2633.
38. Dong P, Tu X, Chu PY, Lü P, Zhu N, Grierson D, Du B, Li P, Zhong S. 3D chromatin architecture of large plant genomes determined by local a/b compartments. Mol Plant. 2017;10:1497–509.
39. Zhang L, Liu L, Li H, He J, Chao H, Yan S, Yin Y, Zhao W, Li M. 3D genome structural variations play important roles in regulating seed oil content of *Brassica napus*. Plant Commun. 2024;5:100666.
40. Tannenbaum M, Sarusi-Portuguez A, Krispil R, Schwartz M, Loza O, Benichou JIC, Mosquna A, Hakim O. Regulatory chromatin landscape in *Arabidopsis thaliana* roots uncovered by coupling INTACT and ATAC-seq. Plant Methods. 2018;14:113.
41. Liu Z, Chen Y, Xia Q, Liu M, Xu H, Chi Y, Deng Y, Xing D. Linking genome structures to functions by simultaneous single-cell Hi-C and RNA-seq. Science. 2023;380:1070–6.
42. Lopez-Nieves S, Yang Y, Timoneda A, Wang M, Feng T, Smith SA, Brockington SF, Maeda HA. Relaxation of tyrosine pathway regulation underlies the evolution of betalain pigmentation in Caryophyllales. New Phytol. 2018;217:896–908.
43. Timoneda A, Feng T, Sheehan H, Walker-Hale N, Pucker B, Lopez-Nieves S, Guo R, Brockington S. The evolution of betalain biosynthesis in Caryophyllales. New Phytol. 2019;224:71–85.
44. Schliemann W, Kobayashi N, Strack D. The decisive step in betaxanthin biosynthesis is a spontaneous reaction. Plant Physiol. 1999;119:1217–32.
45. De Storme N, Copenhaver GP, Geelen D. Production of diploid male gametes in Arabidopsis by cold-induced destabilization of postmeiotic radial microtubule arrays. Plant Physiol. 2012;160:1808–26.
46. Mason AS, Nelson MN, Yan G, Cowling WA. Production of viable male unreduced gametes in Brassica interspecific hybrids is genotype specific and stimulated by cold temperatures. BMC Plant Biol. 2011;11:103.
47. Pécrix Y, Rallo G, Folzer H, Cigna M, Gudin S, Le Bris M. Polyploidization mechanisms: temperature environment can induce diploid gamete formation in Rosa sp. J Exp Bot. 2011;62:3587–97.
48. Mable BK, Alexandrou MA, Taylor MI. Genome duplication in amphibians and fish: an extended synthesis. J Zool. 2011;284:151–82.
49. Gaut BS, Miller AJ, Seymour DK. Living with two genomes: grafting and its implications for plant genome-to-genome interactions, phenotypic variation, and evolution. Ann Rev Genet. 2019;53:195–215.
50. Parisod C, Holderegger R, Brochmann C. Evolutionary consequences of autopolyploidy. New Phytol. 2010;186:5–17.
51. Dwivedi SL, Heslop-Harrison P, Spillane C, McKeown PC, Edwards D, Goldman I, Ortiz R. Evolutionary dynamics and adaptive benefits of deleterious mutations in crop gene pools. Trends Plant Sci. 2023;28:685–97.
52. Stebbins GL. Variation and evolution in plants. Columbia University Press; 1950.
53. Soltis PS, Soltis DE. The role of genetic and genomic attributes in the success of polyploids. Proc Natl Acad Sci USA. 2000;97:7051–7.
54. Ouyang W, Xiong D, Li G, Li X. Unraveling the 3D genome architecture in plants: present and future. Mol Plant. 2020;13:1676–93.
55. Crevillén P, Sonmez C, Wu Z, Dean C. A gene loop containing the floral repressor FLC is disrupted in the early phase of vernalization. Embo J. 2013;32:140–8.
56. Karaaslan ES, Wang N, Faiß N, Liang Y, Montgomery SA, Laubinger S, Berendzen KW, Berger F, Breuninger H, Liu C. Marchantia TCP transcription factor activity correlates with three-dimensional chromatin structure. Nat Plants. 2020;6:1250–61.
57. Liu C, Cheng Y-J, Wang J-W, Weigel D. Prominent topologically associated domains differentiate global chromatin packing in rice from Arabidopsis. Nat Plants. 2017;3:742–8.
58. Liao Y, Wang J, Zhu Z, Liu Y, Chen J, Zhou Y, Liu F, Lei J, Gaut BS, Cao B, et al. The 3D architecture of the pepper genome and its relationship to function and evolution. Nat Commun. 2022;13:3479.
59. Raza A, Li Y, Prakash CS, Hu Z. Panomics to manage combined abiotic stresses in plants. 2025. Trends Plant Sci. https://doi.org/10.1016/j.tplants.2025.03.00.
60. Zaman QU, Raza A, Lozano-Juste J, Chao L, Jones MGK, et al. Engineering plants using diverse CRISPR-associated proteins and deregulation of genome-edited crops. Trends Biotechnol. 2024;42:560–74.

61. Zaman QU, Raza A, Gill RA, Hussain MA, et al. New possibilities for trait improvement via mobile CRISPR-RNA. Trends Biotechnol. 2023;41:1335–8.
62. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
63. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27:764–70.
64. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 2020;11:1432.
65. Liu C. In situ Hi-C library preparation for plants to study their three-dimensional chromatin interactions on a genome-wide scale. Methods Mol Biol. 2017;1629:155–66.
66. Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure JJ. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat Biotechnol. 2013;31:1119–25.
67. Li H, Durbin RJ. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
68. Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, Heard E, Dekker J, Barillot E. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. 2015;16:1–11.
69. Chen C, Li F, Xie F, Chen J, Hua Q, Chen J, Wu Z, Zhang Z, Zhang R, Zhao J, et al. Pitaya Genome and Multiomics Database (PGMD): a comprehensive and integrative resource of *Selenicereus undatus*. Genes (Basel). 2022;13:745.
70. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013;10:1213–8.
71. Zhang X, Pandey MK, Wang J, Zhao K, Ma X, Li Z, Zhao K, Gong F, Guo B, Varshney RK, Yin D. Chromatin spatial organization of wild type and mutant peanuts reveals high-resolution genomic architecture and interaction alterations. Genome Biol. 2021;22:315.
72. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–100.
73. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.
74. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2.
75. Edgar RC, Myers EW. PILER: identification and classification of genomic repeats. Bioinformatics. 2005;21:152–8.
76. Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 2007;35:W265-268.
77. Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics. 2004;20:2878–9.
78. Copetti D, Búrquez A, Shimizu KK, Wing RA, Sanderson MJ, Wojciechowski MF. An improved genome assembly of the saguaro cactus (*Carnegiea gigantea* (Engelm.) Britton Rose). bioRxiv (preprint) 2023. https://doi.org/10.1101/2023.04.11.536419.
79. Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, et al. The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Res. 2008;36:D1009-1014.
80. Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S, et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. Rice. 2013;6:4.
81. Xu X, Pan S, Cheng S, Zhang B, Mu D, Ni P, Zhang G, Yang S, Li R, Wang J, et al. Genome sequence and analysis of the tuber crop potato. Nature. 2011;475:189–95.
82. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37:907–15.
83. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33:290–5.
84. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 2011;12:491.
85. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. 2008;24:637–44.
86. Brůna T, Lomsadze A, Borodovsky M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genom Bioinform. 2020;2:lqaa026.
87. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. 2000;28:45–8.
88. Zdobnov EM, Apweiler R. InterProScan–an integration platform for the signature-recognition methods in InterPro. Bioinformatics. 2001;17:847–8.
89. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene Ontology: tool for the unification of biology. Nat Genetics. 2000;25:25–9.
90. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000;28:27–30.
91. Bu D, Luo H, Huo P, Wang Z, Zhang S, He Z, Wu Y, Zhao L, Liu J, Guo J, et al. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. Nucleic Acids Res. 2021;49:W317-w325.
92. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 2016;3:95–8.
93. Ursu O, Boley N, Taranova M, Wang YXR, Yardimci GG, Stafford Noble W, Kundaje A. GenomeDISCO: a concordance score for chromosome conformation capture experiments using random walks on contact map graphs. Bioinformatics. 2018;34:2701–7.

94. Du Z, Zheng H, Huang B, Ma R, Wu J, Zhang X, He J, Xiang Y, Wang Q, Li Y, et al. Allelic reprogramming of 3D chromatin architecture during early mammalian development. Nature. 2017;547:232–5.
95. Ke Y, Xu Y, Chen X, Feng S, Liu Z, Sun Y, Yao X, Li F, Zhu W, Gao L, et al. 3D chromatin structures of mature gametes and structural reprogramming during mammalian embryogenesis. Cell. 2017;170:367-381.e320.
96. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326:289–93.
97. Zheng X, Zheng Y. CscoreTool: fast Hi-C compartment analysis at high resolution. Bioinformatics. 2018;34:1568–70.
98. Haynes W. Wilcoxon rank sum test. In: Dubitzky W, Wolkenhauer O, Cho K-H, Yokota H, editors. Encyclopedia of systems biology. New York: Springer, New York; 2013. p. 2354–5.
99. Lajoie BR, Dekker J, Kaplan N. The Hitchhiker's guide to Hi-C analysis: practical guidelines. Methods. 2015;72:65–75.
100. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014;159:1665–80.
101. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12:59–60.
102. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: a fast and versatile genome alignment system. PLOS Comput Biol. 2018;14: e1005944.
103. Goel M, Sun H, Jiao W-B, Schneeberger K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol. 2019;20:277.
104. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38: e164.
105. Liao N, Hu Z, Miao J, Hu X, Lyu X, Fang H, Zhou Y-M, Mahmoud A, Deng G, Meng Y-Q, et al. Chromosome-level genome assembly of bunching onion illuminates genome evolution and flavor formation in Allium crops. Nat Commun. 2022;13:6690.
106. Hu G, Feng J, Xiang X, Wang J, Salojärvi J, Liu C, Wu Z, Zhang J, Liang X, Jiang Z, et al. Two divergent haplotypes from a highly heterozygous lychee genome suggest independent domestication events for early and late-maturing cultivars. Nat Genet. 2022;54:73–83.
107. Marchant DB, Chen G, Cai S, Chen F, Schafran P, Jenkins J, Shu S, Plott C, Webber J, Lovell JT, et al. Dynamic genome evolution in a model fern. Nat Plants. 2022;8:1038–51.
108. Zaman QU, Raza A, Hui L, Nazir MF, Garg V, Ikram M, Wang G, Lv W, Khan D, Khokhar AA, You Z, Chitikineni A, Usman B, Jianpeng C, Yang X, Zuo S, Liu P, Kumar S, Guo M, Zhu ZX, Dwivedi G, Qin YH, Varshney RK, Wang HF. Chromosome-level genome assembly of the autotetraploid yellow pitaya provides novel insights into evolution of trait patterning in pitaya species with different ploidy. 2025. https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA1117350.
109. Zaman QU, Raza A, Hui L, Nazir MF, Garg V, Ikram M, Wang G, Lv W, Khan D, Khokhar AA, You Z, Chitikineni A, Usman B, Jianpeng C, Yang X, Zuo S, Liu P, Kumar S, Guo M, Zhu ZX, Dwivedi G, Qin YH, Varshney RK, Wang HF. Chromosome-level genome assembly of the autotetraploid yellow pitaya provides novel insights into evolution of trait patterning in pitaya species with different ploidy. 2025. https://ngdc.cncb.ac.cn/gsa/browse/CRA016827.
110. Zaman QU, Raza A, Hui L, Nazir MF, Garg V, Ikram M, Wang G, Lv W, Khan D, Khokhar AA, You Z, Chitikineni A, Usman B, Jianpeng C, Yang X, Zuo S, Liu P, Kumar S, Guo M, Zhu ZX, Dwivedi G, Qin YH, Varshney RK, Wang HF. Chromosome-level genome assembly of the autotetraploid yellow pitaya provides novel insights into the evolution of trait patterning in pitaya species with different ploidy. 2025. DOI: https://doi.org/10.6084/m9.figshare.25943530. [https://figshare.com/articles/dataset/_b_Chromosome-level_genome_assembly_of_the_autotetraploid_yellow_pitaya_provides_novel_insights_into_evolution_of_trait_patterning_in_pitaya_species_with_different_ploidy_b_/25943530].

Publisher's Note