



OPEN ACCESS

EDITED BY

Kai-Hua Jia,
Shandong Academy of Agricultural Sciences,
China

REVIEWED BY

Fuguo Cao,
Shenyang Agricultural University, China
Yingzhen Wang,
Lishui Vocational and Technical College, China

*CORRESPONDENCE

Jinyan Yao

✉ jinyan.yao@gxaas.net

Xiuqing Wei

✉ weixiuqing@faas.cn

RECEIVED 11 September 2025

REVISED 12 January 2026

ACCEPTED 12 January 2026

PUBLISHED 03 February 2026

CITATION

Long X, Li L, Fang R, Zhang J, Zhou P, An Z,
Tang W, Yao J and Wei X (2026) Pan-genome
analysis of wax apple (*Syzygium
samarangense*) and its association
with fruit size and cold tolerance.
Front. Plant Sci. 17:1703197.
doi: 10.3389/fpls.2026.1703197

COPYRIGHT

© 2026 Long, Li, Fang, Zhang, Zhou, An, Tang,
Yao and Wei. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Pan-genome analysis of wax apple (*Syzygium samarangense*) and its association with fruit size and cold tolerance

Xing Long¹, Liang Li², Ren Fang¹, Ji Zhang¹, Ping Zhou²,
Zhenyu An¹, Wenzhong Tang¹, Jinyan Yao^{1*} and Xiuqing Wei^{2*}

¹Horticultural Research Institute, GuangXi Academy of Agricultural Sciences, Nanning, China, ²Fruit Research Institute, Fujian Academy of Agricultural Sciences, Fuzhou, Fujian, China

Introduction: Wax apple (*Syzygium samarangense*) is a tropical fruit crop of high economic value, in which fruit size and cold tolerance are key traits affecting cultivation range and market quality. However, the genetic basis underlying these traits remains poorly understood at the genome-wide level.

Methods: We constructed the first wax apple pangenome using whole-genome resequencing data from 27 accessions. Novel non-redundant sequences were identified and annotated, and genes were classified based on presence/absence variation (PAV). Population structure was inferred using both PAV and single-nucleotide polymorphism (SNP) data. Structural variants (SVs) were detected genome-wide, and candidate genes associated with fruit size and cold tolerance were identified by integrating PAV, SNP-based XP-CLR, and SV-based F_{ST} analyses.

Results: The pangenome contained 69 Mb of novel non-redundant sequences and 707 newly predicted genes. PAV analysis classified 35,468 genes as core, 10,789 as dispensable, and 364 as private. Population structure analyses consistently divided the accessions into three subgroups, indicating a multi-lineage domestication history. We identified 44,567 SVs, including 9,999 duplications, 34,593 deletions, and 65 insertions. Integrative selective sweep analyses revealed candidate genes associated with fruit size and cold tolerance. An S-adenosylmethionine-dependent methyltransferase gene was located within a cold-tolerance-related selective sweep region. Genes from the Mol family, as well as genes involved in abscisic acid metabolism and quercetin 3'-O-glucosyltransferase activity, were significantly enriched in large-fruited accessions. Additionally, SVs showed strong genetic differentiation in diterpene phytoalexin biosynthesis genes among cold-tolerant lines.

Discussion: This study provides the first pangenome resource for wax apple and reveals extensive genomic variation associated with key agronomic traits. The identified candidate genes and structural variants offer insights into the genetic mechanisms underlying fruit size and cold tolerance and provide a genomic foundation for trait improvement and molecular breeding of wax apple.

KEYWORDS

cold-tolerance, fruit size, pangenome, population genome, selection analysis, *Syzygium samarangense*

1 Introduction

S. samarangense, commonly known as wax apple, Java apple, or water apple, is a typical tropical evergreen fruit tree belonging to the genus *Syzygium* of the family Myrtaceae. Native to the Andaman and Nicobar Islands and the Malaysian Archipelago (Shü et al., 2011; Khandaker and Boyce, 2016), it is now widely cultivated in regions such as Taiwan (China), Thailand, Indonesia, and Malaysia (Zen-hong Shü et al., 2008). With its vibrant coloration (typically deep red, pink, or milky white), bell- or pear-shaped appearance, and crisp, juicy texture, wax apple is valued for both its ornamental and edible qualities, occupying a significant niche in the high-end fruit markets of Southeast Asia and southern China. In recent years, increasing consumer demand for functional fruits has drawn attention to wax apple due to its rich content of bioactive compounds, including proteins, dietary fiber, sugars, vitamins, flavonoids, and phenolic acids. These constituents offer potential health benefits such as antioxidant, anti-inflammatory, and metabolic regulatory effects (Banadka et al., 2022), further spurring research interest in their nutritional and medicinal applications.

Despite its commercial potential, the industrial development of wax apple still faces several challenges. On one hand, as a typical tropical species, it is highly sensitive to low-temperature stress, and its poor cold tolerance severely limits its introduction and cultivation in subtropical regions. On the other hand, fruit size, a key trait determining market value, is influenced by a complex genetic regulatory network and environmental interactions, resulting in significant variation in single-fruit weight among different cultivars (ranging from 28 to 100 g, with a maximum up to 200 g) (Huang, 2017). Currently, few studies focus on the genetic basis of cold tolerance or fruit size in wax apples. In addition, wax apple exhibits complex genomic characteristics, with existing germplasm resources including both diploid and tetraploid types, which has delayed progress in genetic analysis and molecular breeding compared to other major fruit crops.

Recent advances in high-throughput sequencing have opened new avenues for wax apple genomics. Reference genomes for both diploid (Wei et al., 2023) and tetraploid (Zhang et al., 2024) varieties have been published. However, a single reference genome cannot capture the full genetic diversity of a species. In contrast, the construction of a pan-genome, which integrates population-scale sequence variations such as PAVs and SVs, offers a more comprehensive genomic resource for investigating the genetic basis of important traits (Vernikos et al., 2015). In this study, we constructed the first pan-genome of wax apple using whole-genome resequencing data from diverse cultivars. We systematically identified inter-individual PAVs and SVs, and, by integrating XP-CLR-based selective sweep analyses, we uncovered key genomic regions associated with cold tolerance and fruit size. These findings provide novel insights into the molecular mechanisms underlying environmental adaptation and fruit development in wax apple, and lay a foundation for molecular breeding and trait improvement.

2 Materials and methods

2.1 Data Sources

The diploid reference genome of *S. samarangense* (wax apple) used in this study was obtained from the genome assembly published by Zhang et al (Zhang et al., 2024). Whole-genome resequencing (WGS) data and RNA-Seq data for multiple wax apple accessions (Supplementary Table S1) were retrieved from the National Genomics Data Center (NGDC; <https://ngdc.cncb.ac.cn/>). Specifically, the WGS dataset (accession number: PRJCA011699) (Wei et al., 2023) was used for pan-genome construction, variant calling, and population structure analysis, while the RNA-Seq dataset (accession number: PRJCA020470) was utilized to support novel gene prediction, gene expression analysis, and functional annotation.

2.2 Cold tolerance assessment

Cold tolerance was evaluated based on relative electrolyte conductivity (REC) and malondialdehyde (MDA) content. Plants were subjected to low-temperature treatments in a GXZ-0288 illuminated growth chamber (Ningbo Jiangnan Instrument Factory, Ningbo, China), with temperature fluctuations controlled within $\pm 1^{\circ}\text{C}$. Under dark conditions, plants were exposed to stepwise temperature gradients of 7°C , 4°C , 1°C , and -2°C , with each treatment lasting 12 days. One plant was used per treatment, and each treatment included three biological replicates. Immediately after treatment, fully expanded mature leaves were collected and divided into two portions for REC and MDA measurements.

For REC determination, four mature leaves were randomly collected from different orientations of each plant, washed thoroughly, and rinsed with ultrapure water. After blotting dry, leaf veins were removed and leaf tissues were cut into approximately 0.5 cm^2 fragments and homogenized. A total of 1 g of leaf tissue was transferred into an Erlenmeyer flask containing 20 mL of ultrapure water, followed by vacuum infiltration for 15 min. Samples were incubated at room temperature for 4 h, and the initial electrical conductivity (R1) was measured using a DDS-11A precision digital conductivity meter. The samples were then heated in boiling water for 15 min, cooled to room temperature, and the final conductivity (R2) was measured. REC was calculated as $R1/R2 \times 100\%$.

MDA content was quantified using a commercial assay kit (Suzhou Michy Biomedical Technology Co., Ltd., Suzhou, China). Briefly, 0.3 g of fresh leaf tissue was homogenized in 1 mL of extraction buffer at room temperature and centrifuged at 10,000 rpm for 10 min. An aliquot of 100 μL supernatant was mixed with 300 μL reaction reagent and incubated in a 92°C water bath for 30 min. After cooling, the mixture was centrifuged at 12,000 rpm for 10 min at 25°C . Absorbance of the supernatant was measured at 532 nm and 600 nm using a spectrophotometer, and ΔA was calculated as $A_{532} - A_{600}$. MDA content (nmol/g fresh weight)

was calculated as $25.58 \times (\Delta A + 0.0076)/W$, where W represents the fresh weight of the sample.

2.3 Pangenome construction

To construct the *S. samarangense* pangenome, we utilized WGS data from 27 wax apple accessions. Raw sequencing reads were first subjected to quality control using fastp (v0.23.4) (Chen et al., 2018) with default parameters. Cleaned reads were then aligned to the diploid reference genome using BWA (Li and Durbin, 2009), and the resulting alignments were sorted using the sort function in SAMtools (v1.13) (Danecek et al., 2021). Unmapped reads were extracted using the fastq module of SAMtools for subsequent *de novo* assembly. The unmapped reads from each accession were independently assembled using MaSuRCA (v3.2.1) (Zimin et al., 2013) to generate contigs. All contigs from the 27 samples were merged, and redundant sequences were removed using CD-HIT-EST (v4.8.1) (Fu et al., 2012) with default settings. To further eliminate redundancy, an all-vs-all sequence comparison was performed using BLASTN (E-value $\leq 1e-5$) and nucmer (Kurtz et al., 2004); sequences with over 90% identity and 90% coverage to the reference genome were removed. The resulting non-redundant novel sequences were then filtered against the NCBI NT database using BLASTN, and sequences that came from non-Viridiplantae organisms (e.g., archaea, viruses, bacteria, fungi, and animals) were discarded. To exclude organellar sequences, we downloaded the mitochondrial and chloroplast genome sequences of *S. samarangense* from NCBI (Supplementary Table 2) and used BLASTN to remove any contigs of mitochondrial or chloroplast origin. Finally, the filtered novel sequences were once again aligned to the reference genome using BLASTN, and any remaining sequences with high similarity to the reference were excluded. The final set of unique novel sequences was then merged with the reference genome to form the *S. samarangense* pangenome.

2.4 Gene structure annotation

To annotate the novel sequences of the *S. samarangense* pangenome, transcriptome data from 27 wax apple samples were first downloaded from the NCBI database. Transcriptome assembly for each sample was performed using SOAPdenovo-Trans (v1.0.5) (Xie et al., 2014) and Trinity (v2.15.1) (Grabherr et al., 2011). The assembled transcriptomes from all samples were then merged and redundant sequences were removed using CD-HIT-EST, producing a non-redundant transcript set to serve as transcriptomic evidence for gene annotation. To train a homologous protein model for gene prediction, we used AUGUSTUS (v3.5.0) (Stanke et al., 2006), which was trained on protein sequences from the reference genome. To minimize the interference of repetitive elements during gene structure prediction, we identified repetitive sequences in the novel genomic regions using three complementary approaches. First, RepeatMasker (v4.1.7) (Chen, 2004) was applied with the RepBase transposable element library

(v17.01, <http://www.girinst.org/repbase>) to detect known transposable elements. Second, a *de novo* repeat library was constructed using RepeatModeler (v2.0.5) (Flynn et al., 2020), and repetitive elements were annotated using RepeatMasker. Third, tandem repeats were identified using Tandem Repeats Finder (TRF) (v4.09) (Benson, 1999). The outputs from all three methods were merged, and the identified repeats were masked in the novel sequences. Gene structure prediction was then performed on the repeat-masked novel sequences using MAKER2 (v2.31.10) (Cantalapiedra et al., 2021). Protein-coding genes with amino acid sequences longer than 50 residues and annotation edit distance (AED) ≤ 0.5 were retained as high-confidence gene models.

For functional annotation, eggNOG-mapper (v2.1.2) (Cantalapiedra et al., 2021) was used to assign Gene Ontology (GO) terms to each gene. KOBAS (v2.0) (Xie et al., 2011) was employed to annotate KEGG Orthology (KO) terms and assign genes to KEGG pathways. Additionally, HMMER (hmmsearch, v3.4) (Eddy, 2023) was used to annotate Pfam domains in the protein-coding genes. Gene enrichment analyses were conducted using the clusterProfiler R package (Wu et al., 2021).

2.5 Population analysis based on PAV

To investigate population structure based on gene PAV, WGS reads were first remapped to the constructed wax apple pangenome using BWA. Gene-level PAVs were identified using SGSGeneLoss (v0.1) (Golicz et al., 2015), with parameters set to minCov = 2 and lostCutoff = 0.2, meaning a gene was considered present if at least two reads covered more than 20% of its exon regions. A binary PAV matrix was then generated and used for multidimensional analyses of population genetic structure. Principal component analysis (PCA) was conducted using the vegan R package (Oksanen et al., 2015). A maximum likelihood phylogenetic tree was constructed using IQ-TREE (v2.3.3) (Minh et al., 2020) with the options -st BIN -alrt 1000, which are optimized for binary input data. Additionally, population structure was inferred using the binary matrix in STRUCTURE (v2.3.4) (Hubisz et al., 2009).

Based on the PAV matrix, genes in the pangenome were categorized as core genes (present in all 27 accessions), dispensable genes (present in 1–26 accessions), and private genes (present in only one accession). Gene frequencies were calculated to assess potential selection pressures acting on PAVs during population divergence. Differences in gene frequencies between groups were evaluated using Fisher's exact test, with p-values adjusted by the Benjamini-Hochberg (BH) method. Statistical significance was defined as a false discovery rate (FDR) < 0.001 and $|\log_2 \text{fold change}| > 1$.

2.6 Population genetic analysis based on SNPs

SNP calling was performed using GATK (v4.6.1) (McKenna et al., 2010) with the following pipeline: First, potential PCR

duplicates were identified and removed using the “MarkDuplicates” tool. Variant calling was then conducted using HaploTypeCaller to generate GVCF files. The raw SNPs were filtered using VariantFiltration with parameters “FS>30.0 || QD<2.0” to remove low-quality variants. We subsequently applied SelectVariants with parameters “-exclude-filtered true -restrict-alleles-to BIALLELIC” to retain only high-quality biallelic SNPs. Further filtering was performed using VCFtools (v0.1.16) (Danecek et al., 2011) with parameters “-max-missing 0.7 -maf 0.5” to remove SNPs with >30% missing data across populations and minor allele frequency (MAF) <5%.

Population structure was analyzed using ADMIXTURE (v1.3.0) (Alexander et al., 2009) with default parameters. For phylogenetic analysis, SNP sequences were extracted and used to construct a maximum-likelihood tree with CASTER-site (v1.19.1.4) (Zhang et al., 2023) under default parameters, with subsequent visualization performed using iTOL (<https://itol.embl.de/upload.cgi>).

2.7 Detection of selective signatures

We employed the XP-CLR (Python version, <https://github.com/hardingnj/xpclr>) (Chen et al., 2010) to identify selective sweeps between different wax apple populations. The genome was divided into non-overlapping 10 kb windows, and the average XP-CLR likelihood score was calculated for each window. Windows with scores in the top 5% genome-wide were considered strong selective signals. Adjacent windows or those separated by a single window (within the top 10% XP-CLR scores) were merged, with the maximum average score assigned to the new merged region. For candidate genes and their ±2 kb flanking regions, we evaluated population differentiation using F_{ST} values computed by VCFtools.

2.8 SV calling and filtering

We performed SV detection using Delly (v1.2.6) (Rausch et al., 2012) with the “call” module for individual samples, followed by merging results using Delly’s “merge” tool. The merged SVs were converted to VCF format using bcftools (v1.18) (Danecek et al., 2021). Subsequent filtering was conducted with VCFtools using parameters “-max-missing 0.5 -maf 0.01” to retain variants with ≤50% missing data across samples and minor allele

frequency (MAF) ≥1%. Only variants exceeding 50 bp in length were classified as SVs for downstream analyses.

2.9 SV hotspot identification

SV hotspots were identified following the method described by Qin et al (Qin et al., 2021). Briefly, we calculated SV breakpoint distributions across the genome using 200 kb sliding windows with 100 kb steps along each chromosome. Windows were ranked in descending order based on SV counts, with the top 5% of windows containing the highest SV breakpoint frequencies designated as hotspots. Adjacent hotspot windows were subsequently merged into contiguous “hotspot regions”.

2.10 Genome-wide genetic differentiation analysis

Genome-wide population differentiation was assessed by calculating fixation index (F_{ST}) and nucleotide diversity (π) values for both SVs and SNPs using VCFtools. The genome was partitioned into non-overlapping 100 kb windows, with average F_{ST} and π values computed for each window. Regions with $F_{ST} > 0.15$ were considered to exhibit significant genetic differentiation between wax apple populations.

2.11 PCR and quantitative real-time PCR analysis

Total RNA was extracted using the UPure Tissue RNA Kit (Biokeystone, China), and RNA quality and concentration were assessed by spectrophotometry and agarose gel electrophoresis. First-strand cDNA was synthesized from 1 µg of total RNA using a reverse transcription kit according to the manufacturer’s instructions.

Conventional PCR was performed to validate the presence and absence of the pangenome novel genes. Primers were designed using Primer3 software (Table 1). PCR reactions were carried out in a total volume of 20 µL containing cDNA template, gene-specific primers, and PCR master mix. The amplification program consisted of an initial denaturation at 95°C for 3 min, followed by 35 cycles of denaturation at 95°C for 30 s, annealing at 60°C for 30 s, and

TABLE 1 Primers used in this study.

Gene	Forward (5'-3')	Reverse (5'-3')
evm.TU.group11.2151	TGTGGACTGAGCATCTTCACT	TCAAGGCCTCCGAAGACTTCT
novel_gene00049	GTCAGCATGTATGGGCTAACTC	ATGCTTCACTCTACTCCATATT
evm.TU.group6.4267	CAGAACCAGGGGCTCGAAAAAG	ATGGAGGCCTTCTCTGTCTCTC
evm.TU.group5.517	TGGCTACAACTTTAGAGTCTT	TCAGATCTTCTCCATCTTCAG
evm.TU.group6.4268	TGAAGATGGAATCGCTCTCGT	CTAGGGTCTCAGGAAGGCCAATC
evm.TU.group7.4063	TACAGGAGAAGAACTAAGACCA	ATGACCGAACACATCAATGCAG

extension at 72°C for 30 s, with a final extension at 72°C for 5 min. PCR products were analyzed by 1.5% agarose gel electrophoresis.

Quantitative real-time PCR (qRT-PCR) was conducted using a SYBR Green-based detection system on a real-time PCR instrument. Total RNA was reverse-transcribed into cDNA with TUREscript 1st Stand cDNA SYNTHESIS Kit (Aidlab, China). Gene-specific primers were designed using Primer3 software (F: TGTGACCGACTCTTGTTTC, R: CAGCAGCATCAACTCTTC). Each reaction was performed in a 20 μ L volume containing diluted cDNA, gene-specific primers, and SYBR Green PCR master mix. Melting curve analysis was performed to confirm amplification specificity. Relative gene expression levels were calculated using the $2^{-\Delta\Delta C_t}$ method, with an internal reference gene (SsSKIP16, F: GGAACCTCCACTCTGTTCCA, R: AGTCGTAGGGCATTCCATTG) used for normalization.

3 Results

3.1 Construction of the wax apple pangenome

In this study, we constructed the *S. samarangense* pangenome using WGS data from 27 cultivars. In addition to the reference genome, 69 Mb of novel sequences were assembled, comprising

41,045 contigs. Gene structure annotation identified 707 novel protein-coding genes within these sequences. Based on the PAV identification of all genes in the pangenome, a total of 46,621 genes were categorized into 35,468 core genes, 10,789 dispensable genes, and 364 private genes (Figures 1A, B). We examined the PAV profiles across all samples (Figure 1C), finding that the number of genes present in each accession ranged from 43,817 to 44,906. Core genes accounted for approximately 80% of the genes in each individual, whereas private genes represented only a very small proportion, suggesting their highly restricted distribution and potential cultivar specificity. As the number of samples increased, the total number of pangenome genes gradually rose and approached saturation, indicating that the constructed pangenome captures the genetic diversity of *S. samarangense*. In contrast, the number of core genes steadily declined with increasing sample size, implying differences in the core gene repertoire across populations (Figure 1D). Furthermore, PCR amplification of randomly selected sequences confirmed the reliability of the predicted PAVs in this study (Supplementary Figure S1).

Given the widespread PAV of dispensable genes, which may underlie phenotypic differences among populations, we performed functional enrichment analysis on these genes. GO enrichment revealed significant associations with terms related to asymmetric cell division (e.g., zygote asymmetric cell division, asymmetric cell division), nuclear pore complex components (e.g., nuclear pore

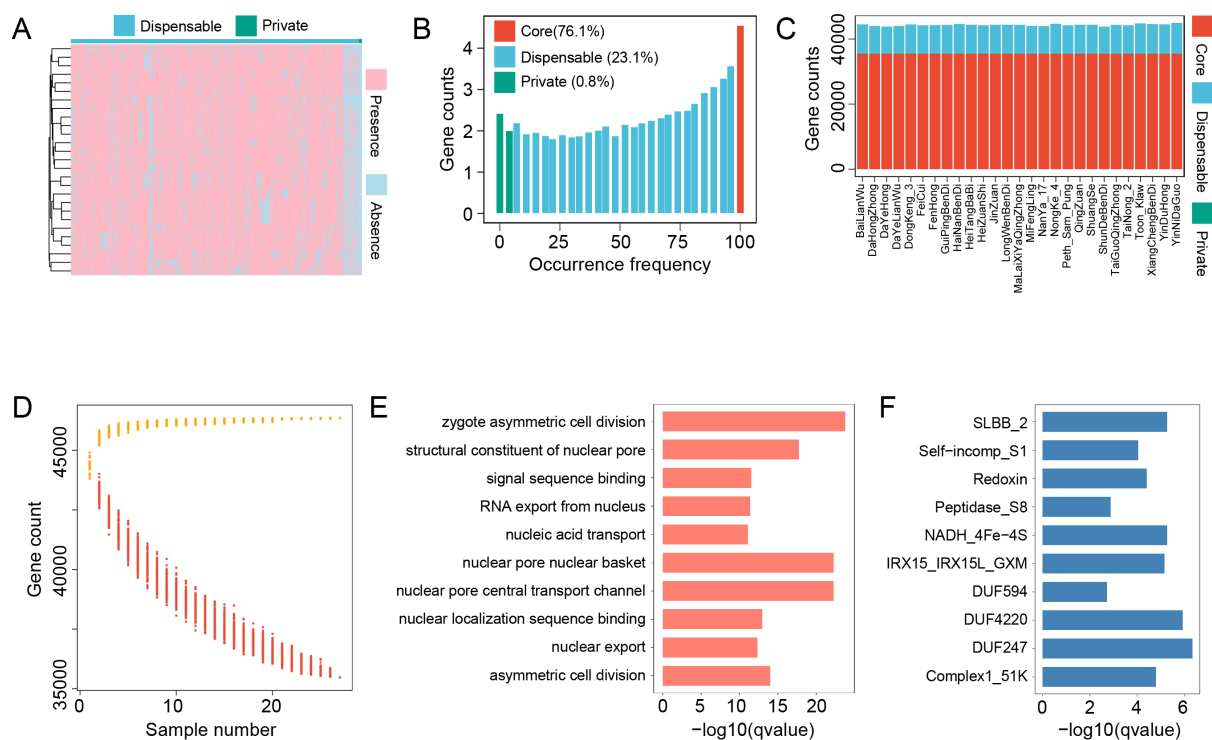


FIGURE 1

Construction of the wax apple pangenome. (A) PAV heatmap of dispensable and private genes. Rows represent 27 samples along with a hierarchical clustering tree, and columns represent genes. (B) Frequency distribution of core, dispensable, and private genes. The x-axis indicates gene occurrence frequency, while the y-axis shows the number of genes at each frequency. Different colors represent different PAV types. (C) Gene counts of different PAV categories across all samples. (D) Curve fitting of pangenome gene number and core gene number in the pangenome as the sample size increases. (E) GO enrichment analysis of dispensable genes. (F) PFAM domain enrichment analysis of dispensable genes.

central transport channel, nuclear pore nuclear basket), and nucleic acid transport (e.g., RNA export from nucleus, nucleic acid transport) (Figure 1E, Supplementary Table 3). Pfam domain enrichment showed overrepresentation of domains such as Complex1_51K, Peptidase_S8, IRX15_IRX15L_GXM, DUF247, and Self-incomp_S1 (Figure 1F, Supplementary Table 4).

3.2 Population structure and phylogenetic analysis of the wax apple

To investigate the genetic structure of *S. samarangense* cultivars, we first performed PCA based on the binary PAV matrix. The results revealed a clear tri-group differentiation: Subgroup I comprised FenHong, HeiZuanShi, and TaiNong_2; Subgroup II included HeiTangBaBi, QingZuan, and ShuangSe; while the remaining 21 cultivars clustered into Subgroup III (Figure 2A). A phylogenetic tree constructed from the binary PAV matrix further supported this population structure: members within Subgroups I and II formed well-supported monophyletic clades with short branch lengths, indicating close genetic relationships within each subgroup (Figure 2B). Population structure analysis using STRUCTURE also supported the three-group division, with the optimal K value (K = 3) corresponding to the clustering observed in PCA and phylogenetic analysis. Notably, the two smaller subgroups

(each with three cultivars) were clearly differentiated from the main cultivar group (21 accessions), suggesting that these accessions may have undergone distinct domestication trajectories or breeding selection (Figure 2C).

However, population structure analysis based on SNPs revealed a partially different pattern (Figure 2D). While HeiTangBaBi, QingZuan, and ShuangSe remained a separate cluster, the subgroup containing FenHong, HeiZuanShi, and TaiNong_2 did not form an independent group and instead clustered with the main cultivars. Interestingly, the SNP-based phylogenetic tree still maintained a three-clade structure. This discrepancy may be attributed to the different temporal sensitivities of molecular markers: PAVs may retain deeper historical signals, such as ancient lineage divergence or early selection events, whereas SNPs may better capture recent gene flow and convergence driven by cultivation practices.

3.3 Population selection analysis in wax apple

As a key tropical fruit crop, *S. samarangense* exhibits substantial phenotypic diversity across cultivars, particularly in fruit size and environmental adaptability. For example, cultivars such as HeiTangBaBi and YinNiDaGuo produce notably larger fruits,

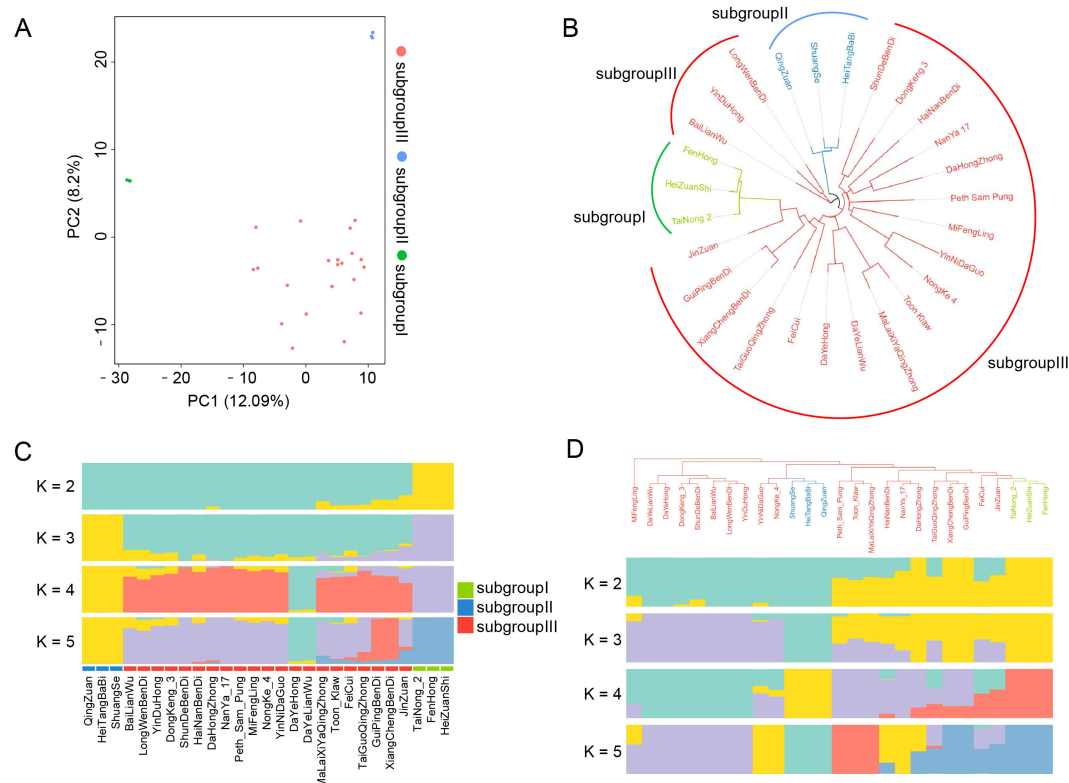


FIGURE 2

Phylogenetic and population structure analysis of wax apple based on PAV and SNP. (A) PCA based on PAVs. Each dot represents a sample, and colors indicate groupings derived from PCA. (B) Phylogenetic tree constructed using the binary PAV matrix. (C) Population structure analysis based on PAVs, showing inferred clusters under K = 2, 3, 4, and 5. (D) Phylogenetic tree and population structure analysis based on SNPs.

while DaYeHong and DongKeng 3 demonstrate enhanced cold tolerance under low-temperature conditions (Supplementary Table 5, Supplementary Figures S2, S3). To investigate the genetic basis of these phenotypic differences, we first analyzed gene frequency based on the PAV across distinct population groups. A total of 212 genes showed significantly higher presence frequencies in cold-tolerant varieties, while 311 genes were showed significantly higher presence frequencies in large-fruited cultivars (Supplementary Tables 6, 7). Functional annotations revealed that some of these genes, which are under PAV selection, are associated with phenotype-relevant functions. For example, *evm.TU.group2.3467* exhibits a significantly higher frequency in cold-tolerant cultivars and is annotated with a PPR domain in the Pfam database, which is associated with cold resistance. Similarly, *evm.TU.group5.517* shows a significantly higher frequency in large-fruited cultivars and contains a P450 domain, which is linked to fruit size. Importantly, several key genes were located on novel sequences uncovered by pangenome assembly. For instance, *novel_gene00318*, annotated with a PFAM domain LRR_1, is potentially involved in disease resistance. These findings underscore the power of the pangenome in capturing gene variants missed by the reference genome.

In addition, we performed XP-CLR analysis to identify genomic regions under selection between contrasting groups for cold tolerance and fruit size. We identified ~8.7 Mb and ~9.2 Mb of regions with significant selection signals between cold-tolerant vs. cold-sensitive (Figure 3A), and large-fruited vs. small-fruited populations (Figure 3B), respectively. In the cold tolerance comparison, 21 genes had selection signals (Supplementary Table 8), several of which are functionally relevant to cold response. For example, *evm.TU.group2.1803* was found in a peak region with an XP-CLR likelihood score of 37.6. PFAM annotation identified it as a SAM-dependent methyltransferase involved in quercetin metabolism. Methyltransferases can influence plant development and stress response through epigenetic modifications. Moreover, this gene exhibited a high F_{ST} (>0.6) between cold-tolerant and sensitive populations, indicating strong genetic differentiation and suggesting its potential role in cold adaptation. Similarly, *evm.TU.group2.3887*, located in a region with a score of 13.6, belongs to the LRR-RLK gene family, known to play key roles in plant development, hormone signaling, abiotic stress response, and pathogen defense (Shiu and Bleeker, 2001; Xiang et al., 2006; de Lorenzo et al., 2009; Li et al., 2018). For fruit size, 16 genes were located within selected regions (Supplementary Table 9). Notably, we identified a homologous gene

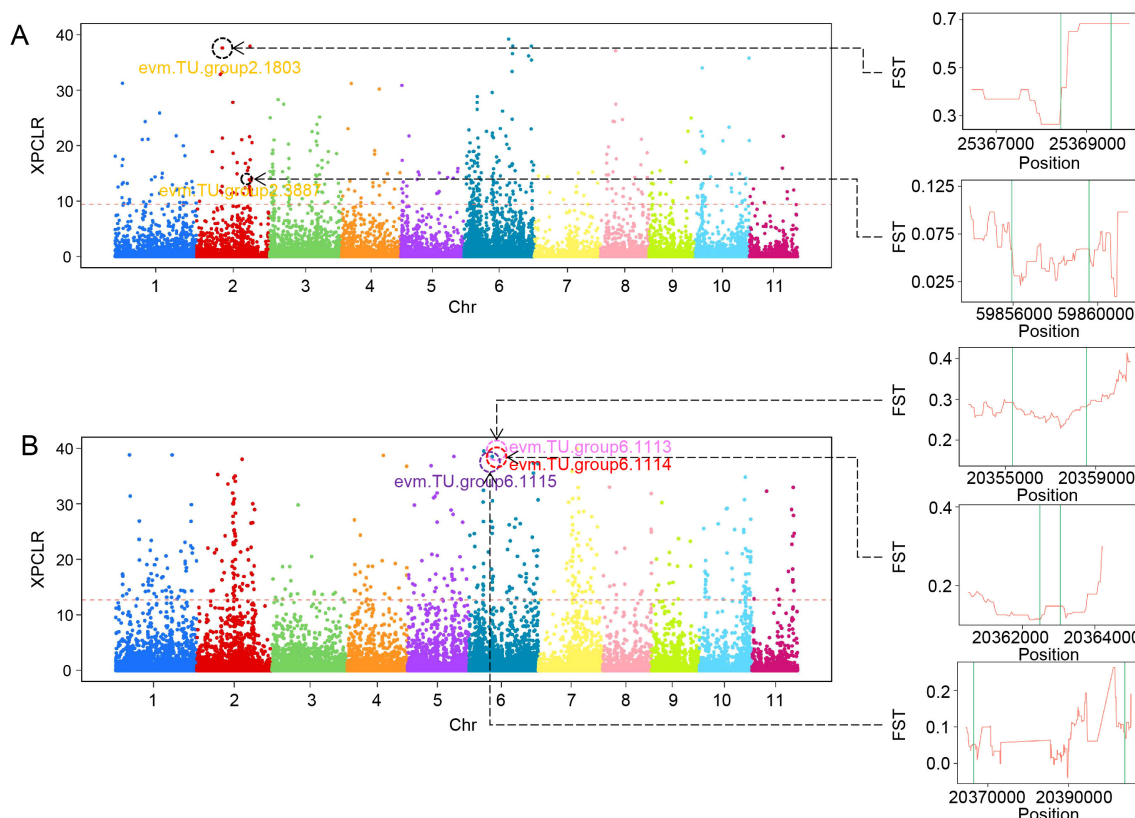


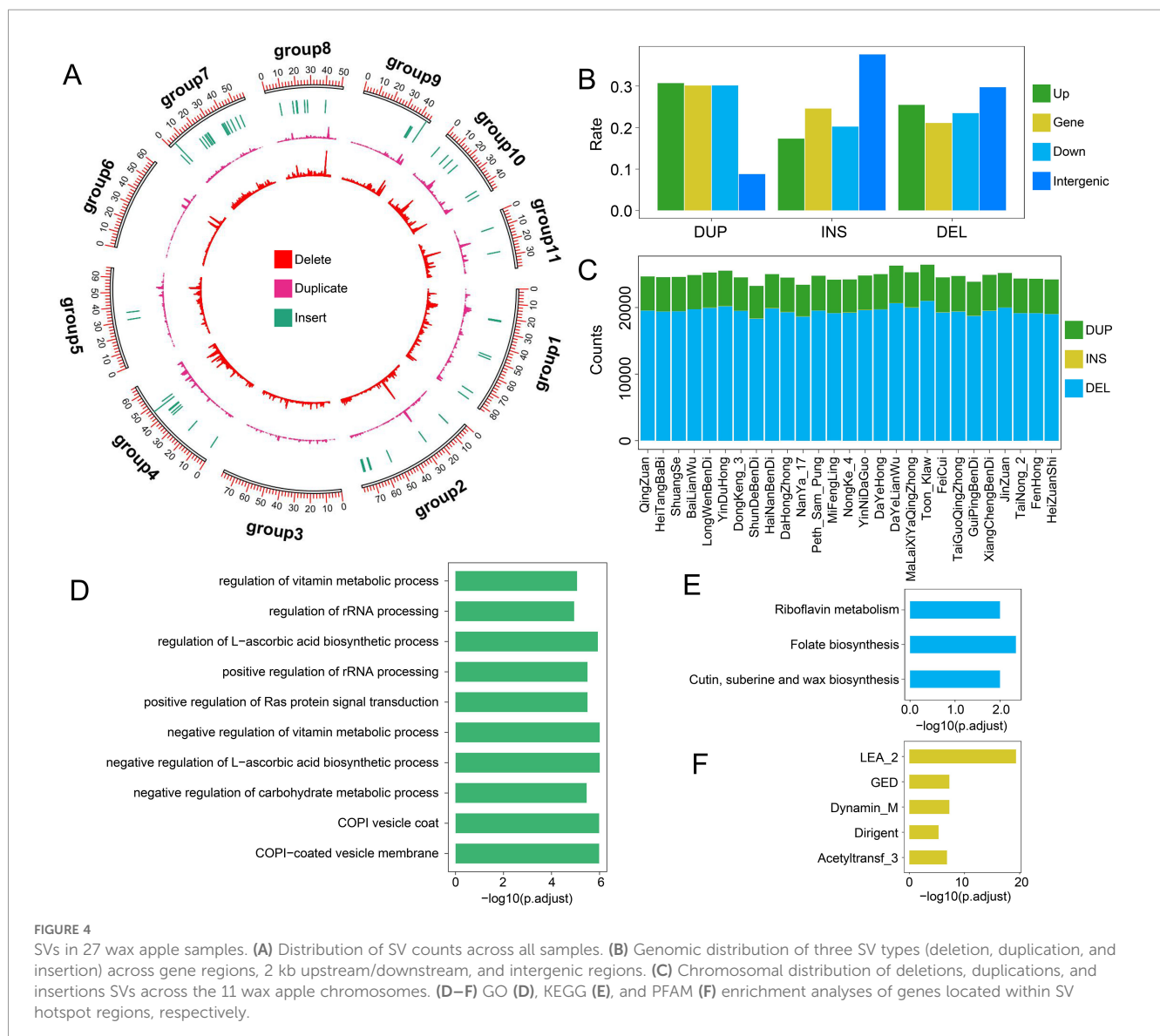
FIGURE 3

Selective sweeps in wax apple. (A) Genome-wide distribution of XP-CLR likelihood scores between cold-tolerant (most cold-resistant and cold-resistant) and non-cold-tolerant populations. The line chart on the right represents F_{ST} values across the gene regions and 2 kb upstream and downstream of *evm.TU.group2.1803* and *evm.TU.group2.3887*. (B) Genome-wide distribution of XP-CLR likelihood scores between large-fruited (single fruit weight ≥ 100 g) and small-fruited (single fruit weight ≤ 70 g) populations. The line chart on the right represents F_{ST} values across the gene region and 2 kb upstream and downstream of *evm.TU.group6.1113*, *evm.TU.group6.1114*, and *evm.TU.group6.1115*.

cluster of Mol-like proteins on chromosome 6, including *evm.TU.group6.1113*, *evm.TU.group6.1114*, and *evm.TU.group6.1115* (Figure 3B). This region had an XP-CLR score of 38.5, especially the gene region and upstream and downstream of *evm.TU.group6.1113* showed F_{ST} values above 0.15, suggesting that this gene cluster may contribute to the regulation of fruit size. *evm.TU.group6.1115* exhibited a higher number of Mol domains across the three genes in the cluster (Supplementary Figure S4), and its expression level was significantly elevated in large-fruited cultivars compared with small-fruited cultivars (Supplementary Figure S5), suggesting that this gene may be associated with fruit size. Moreover, SVs were detected within the genomic regions of several genes located in these selection signal intervals. For example, both an SV deletion and an SV duplication were identified within the genomic region of *evm.TU.group2.1803*, while two SV deletions were detected within the regions of *evm.TU.group6.1113* and *evm.TU.group6.1115*, indicating that SVs may be associated with adaptive evolution of genes under selection.

3.4 SV landscape in the wax apple population

SVs play crucial roles in plant genome evolution, agronomic trait regulation, and environmental adaptation. In this study, we comprehensively identified SVs across the wax apple population. A total of 44,657 SVs were detected, including 9,999 duplications, 34,593 deletions, and 65 insertions. The number of SVs per sample ranged from 23,252 to 26,411, with deletions accounting for approximately 78.35%–79.6% of all SVs. Insertions were rare, with fewer than 50 per sample (Figure 4A). Regarding their genomic locations, most duplication SVs were found within gene regions or in the gene upstream and downstream regions of genes. Deletion SVs were distributed more evenly among gene regions, upstream/downstream regions, and intergenic regions. Notably, 31.1% of insertion SVs were located in intergenic regions (Figure 4B). As SVs were not uniformly distributed across the genome (Figure 4C),



we performed a hotspot analysis to identify regions with high SV density. A total of 67 SV hotspots were identified. Although these hotspots accounted for only 6% of the genome (~41 Mb), they harbored 87.7% of all SVs (39,063), suggesting their potential importance in genome structure and function.

To investigate the functional implications of these SV-enriched regions, we performed enrichment analysis of genes located within the hotspots. The results showed that these genes were significantly enriched in pathways related to secondary metabolism, particularly those involving vitamin biosynthesis and metabolism. GO enrichment analysis highlighted terms such as COPI-coated vesicle membrane, COPI vesicle coat, negative regulation of carbohydrate metabolic process, negative regulation of L-ascorbic acid biosynthetic process, positive regulation of Ras protein signal transduction, and regulation of rRNA processing (Figure 4D). KEGG analysis revealed enrichment in folate biosynthesis, riboflavin metabolism, and cutin, suberine and wax biosynthesis pathways (Figure 4E). PFAM domain enrichment indicated a prevalence of domains such as LEA_2, GED, Dynamin_M, Dirigent, and Acetyltransf_3 (Figure 4F). These findings provide

new insights into the potential genetic basis underlying fruit quality and stress tolerance in wax apple.

3.5 SV-driven population differentiation analyses

SV as a major source of genomic variation may play a crucial role in adaptive evolution and phenotypic diversity. However, the contribution of SVs to population differentiation and key traits such as cold tolerance and fruit size in wax apple remains unclear. In this study, we calculated SV-based F_{ST} values between populations with different levels of cold tolerance and fruit size, identifying 216 and 244 significantly differentiated regions, respectively (Figure 5A). These regions contained 1,514 and 1,516 genes, respectively. Compared with SNP-based F_{ST} analysis under the same parameters, SV-based F_{ST} identified a larger number of significantly differentiated genomic regions (Figure 5B). Moreover, the number of genes associated with SVs exceeded those associated with SNPs. This phenomenon is likely because

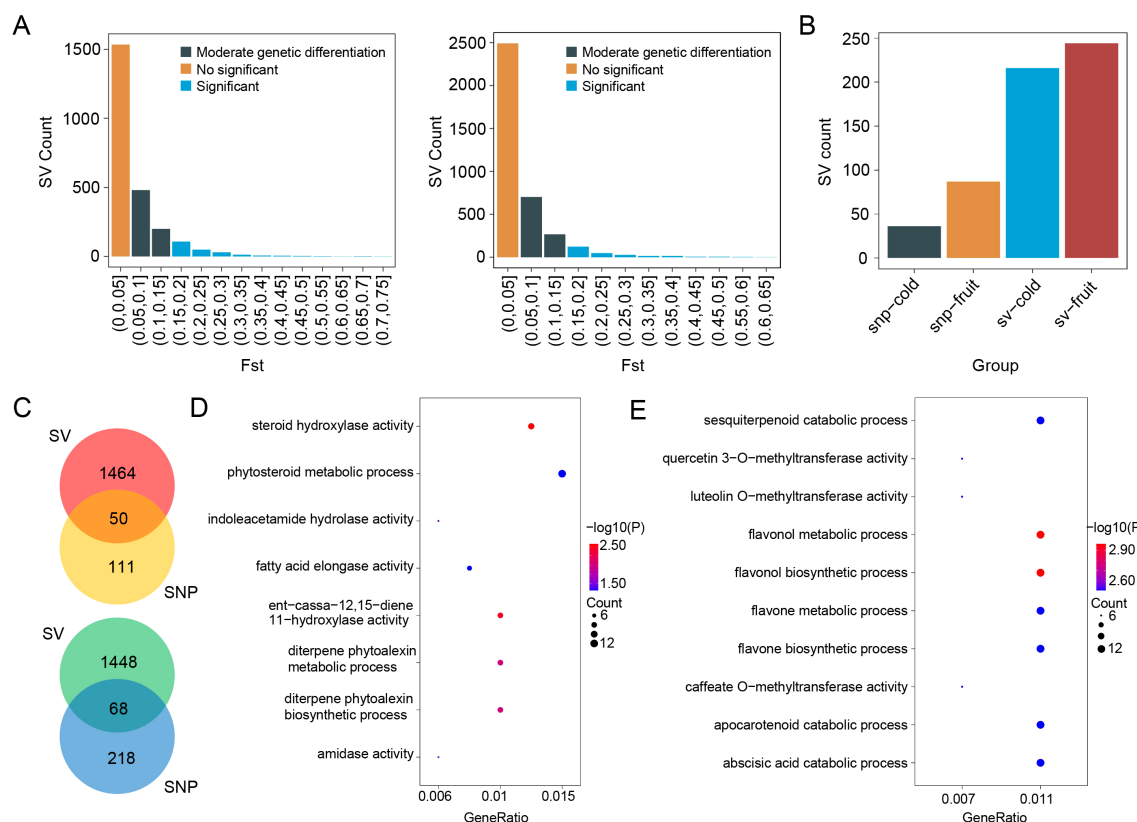


FIGURE 5

Population differentiation based on SVs. (A) Distribution of F_{ST} values between groups with differing cold tolerance and fruit size. When $F_{ST} < 0.05$, genetic differentiation between groups is minimal and considered non-significant. F_{ST} values between 0.05 and 0.15 indicate moderate genetic differentiation, while $F_{ST} \geq 0.15$ suggests significant genetic differentiation. (B) Comparison of the number of significantly differentiated genomic windows (window size = 100,000 bp) identified using SNPs and SVs. (C) Comparison of the number of genes located within significantly differentiated regions identified using SNPs and SVs. (D) GO enrichment analysis of genes located in significantly differentiated regions between groups with differing cold tolerance. (E) GO enrichment analysis of genes located in significantly differentiated regions between groups with differing fruit size.

SVs span longer sequences than SNPs, which involve only a single base change. The overlap between SV-associated and SNP-associated genes was minimal (Figure 5C), suggesting that these two types of variants may influence trait differentiation through distinct genetic mechanisms. To further explore the functions of genes located in SV-driven differentiated regions, we performed gene enrichment analysis. In the cold-tolerant and cold-sensitive populations, genes within significantly differentiated regions were enriched in GO terms related to plant growth, development, and environmental adaptation, including amidase activity, diterpene phytoalexin biosynthetic process, diterpene phytoalexin metabolic process, ent-cassa-12,15-diene 11-hydroxylase activity, fatty acid elongase activity, indoleacetamide hydrolase activity, phytosteroid metabolic process, and steroid hydroxylase activity (Figure 5D).

4 Discussion

Compared to a single reference genome, a pangenome provides a more comprehensive representation of the genetic diversity within a species. In recent years, significant progress has been made in pangenome research across various plant species. For example, the tomato pan-genome (1,179 Mb) revealed 351 Mb of novel sequences and 4,873 new genes (Gao et al., 2019); the sunflower pangenome identified 17,061 new genes (Hübner et al., 2019); and analysis of 3,010 accessions of Asian cultivated rice uncovered 268 Mb of non-redundant novel sequences containing 12,465 new genes (Wang et al., 2018). In this study, we assembled a pangenome based on 27 diploid *S. samarangense* accessions and assembled 69 Mb of novel sequences and 707 new genes. Compared to other plant species, the scale of novel sequences and gene numbers in the wax apple is relatively limited, which may be attributed to its low genetic diversity. The natural distribution of wax apple is geographically restricted to tropical regions such as Malaysia, Indonesia, the Philippines, and parts of southern China, including Guangdong and Taiwan. This narrow distribution likely constrains the accumulation of genetic variation. From a functional perspective, previous studies have shown that variable genes in pangenomes (i.e., flexible or dispensable genes) are often closely associated with environmental adaptation. For instance, in tomato, variable genes are significantly enriched in defense response pathways (Gao et al., 2019), while in Asian rice, dispensable genes are primarily involved in immune and defense regulation (Wang et al., 2018). In contrast, the variable genes identified in wax apple showed no significant enrichment in pathways related to environmental adaptation or defense, which may reflect its limited distribution and relatively weak adaptability. Interestingly, several unique functional domains were enriched in the Pfam analysis of wax apple genes, including DUF4220, DUF594, Self-incomp_S1, and DUF247. Among them, DUF4220 and DUF594 often occur in tandem and are plant-specific. Although the functions of these domains remain largely unknown, studies suggest they may be involved in sugar metabolism regulation. For example, in rice, the *OsSAC1* encodes a protein containing both DUF4220 and DUF594 domains, and its mutation leads to sugar accumulation in leaves, implying a potential

role for these domains in sugar metabolism and distribution (Zhu et al., 2018). Additionally, the Self-incomp_S1 and DUF247 domains are closely linked to plant self-incompatibility (SI). In grasses, DUF247 is widely present in self-incompatible species but often lost or mutated in self-compatible ones (Foote et al., 1994). The absence of Self-incomp_S1 and DUF247 domains in the wax apple genome may be associated with the evolutionary development of its self-compatibility (Manzanares et al., 2016; Herridge et al., 2022). These findings offer insights into the potential roles of dispensable genes in wax apple and provide a foundation for further functional investigations.

Currently, population genetics and genomic studies of wax apple remain in the early exploratory stages. Wax apple exhibits both diploid and tetraploid forms, with a diverse and complex range of chromosome numbers ($2n = 33, 42, 44, 66$, and 88), suggesting that the species may have undergone polyploidization and/or hybridization events. Although high-quality genomes of diploid and tetraploid wax apple were published in 2023 (Wei et al., 2023) and 2024 (Zhang et al., 2024), respectively, their population structure, origin, and evolutionary history remain largely unresolved. A previous study classified local landraces and cultivated varieties into two major groups (Wei et al., 2023), but only reported population structure results for $K = 2$. Increasing the K value may reveal finer substructures, indicating that the genetic diversity of wax apple may be more complex than previously recognized. In this study, PAV-based population analysis of 27 local landraces revealed two distinct subgroups. Notably, earlier SNP-based PCA also divided the local landraces into three groups (Wei et al., 2023), consistent with our PAV-based PCA findings. Importantly, both SNP- and PAV-based PCA and phylogenetic analyses support the existence of three major groups: (1) FenHong, HeiZuanShi, and TaiNong_2; (2) HeiTangBaBi, QingZuan, and ShuangSey; and (3) other varieties. These results suggest that current wax apple populations may have originated from three distinct ancestral lineages, pointing to multiple independent domestication or dispersal events. However, research on the origin and evolution of the wax apple is still limited. This hypothesis requires further support from expanded genomic data, population genetics analyses, and paleobotanical evidence. Future studies combining whole-genome resequencing, demographic modeling, and comparative genomics will be critical to unraveling the domestication history and adaptive evolutionary mechanisms of wax apple.

Fruit size and cold tolerance are two key agronomic traits that directly impact the yield and cultivation potential of *S. samarangense*. In this study, we integrated three complementary approaches, including PAV selection, SNP-based XP-CLR selection sweep detection, and SV-based F_{ST} population differentiation, to identify candidate genetic loci potentially associated with these traits. PAV selection analysis identified *evm.TU.group5.517*, which exhibits significant frequency differences between different fruit size populations and encodes a protein containing a P450 domain. Previous studies have demonstrated that RNAi-mediated suppression of *P450* alters fruit size in cherry, suggesting a potential conserved role of this gene family in fruit development (Qi et al., 2017). Similarly, *evm.TU.group2.3467*, which shows significant frequency differences between cold-tolerant and cold-sensitive groups, harbors a PPR domain. PPR proteins have been implicated in cold

tolerance in rice by regulating mitochondrial superoxide levels (Zu et al., 2023), indicating their likely involvement in cold adaptation in wax apple. XP-CLR analysis further supported the involvement of these genes in adaptive differentiation. For instance, in cold-tolerant populations, the selective region contained *evm.TU.group2.1803*, which is annotated as an S-adenosylmethionine-dependent methyltransferase. A gene with the same domain called *WPEAMT* has been shown to contribute to cold adaptation in wheat (Charron et al., 2002), reinforcing its relevance in low-temperature stress responses. Additionally, the XP-CLR region associated with fruit size contained a *Mol* gene cluster, members of which have been previously linked to fruit development (Feechan et al., 2008; Chen et al., 2014). SV-based F_{ST} analysis revealed genomic regions with significant genetic differentiation between groups with contrasting cold tolerance or fruit size. In cold-tolerant vs. cold-sensitive groups, genes within highly differentiated regions were enriched for terms related to diterpene phytoalexin biosynthesis and metabolism. Diterpene phytoalexins are known to act as crucial defense-related secondary metabolites, playing vital roles in plant responses to both biotic and abiotic stresses (Sashankar and Chakraborty, 2023). In contrast, genes showing differentiation between fruit size groups were enriched for terms related to fruit development and hormone signaling. For example, abscisic acid (ABA) is known to regulate ethylene biosynthesis and signaling during fruit ripening, and its crosstalk with other hormones is a major determinant of fruit size and maturation, although the underlying regulatory mechanisms remain largely unclear (Gupta et al., 2022). Moreover, quercetin, a flavonol belonging to the polyphenol family, has been implicated in influencing fruit peel color and ripening in apple, alongside anthocyanin glycosides (Reay and Lancaster, 2001), highlighting its potential role in wax apple fruit quality traits. Collectively, these findings provide valuable genetic insights into the molecular mechanisms underlying fruit development and cold tolerance in wax apple. The candidate genes and pathways identified here offer promising targets for functional validation and molecular breeding. However, the roles of these candidate genes are currently supported primarily by genomic analysis, and their precise biological functions require further experimental verification through approaches such as gene expression perturbation, transgenic analysis, or genome editing. Future studies integrating functional genomics and multi-omics approaches will be instrumental in constructing regulatory networks of key agronomic traits and accelerating varietal improvement in this species.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

XL: Conceptualization, Methodology, Software, Validation, Writing – original draft, Funding acquisition. LL: Software,

Writing – original draft. RF: Investigation, Software, Writing – original draft. JZ: Data curation, Resources, Writing – original draft. PZ: Investigation, Resources, Writing – original draft. ZA: Data curation, Software, Writing – original draft. WT: Resources, Software, Writing – original draft. JY: Conceptualization, Software, Visualization, Writing – original draft, Writing – review & editing. XW: Project administration, Resources, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was funded by Guangxi Key R&D Program Project (No. Guike AB18221035); Special Project on Basic Scientific Research of Guangxi Academy of Agricultural Sciences (No. Gui Agricultural Science 2021YT049); the Fundamental Scientific Research at Nonprofit Research Institutions in Fujian province (2021R10280010).

Acknowledgments

We would like to acknowledge Chengdu Genepre Co., Ltd. (www.genepre.com) for providing assistance in data analysis.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2026.1703197/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Gel electrophoresis validation of pangenome genes.

SUPPLEMENTARY FIGURE 2

Fruit morphology of wax apple.

SUPPLEMENTARY FIGURE 3

Cold tolerance assessment of wax apple. **(A)** Phenotypic responses of wax apple leaves under cold stress treatment. **(B)** MDA content in wax apple leaves under cold stress. **(C)** Relative electrolyte conductivity of wax apple leaves under cold stress.

SUPPLEMENTARY FIGURE 4

Gene structure of fruit size and cold tolerance candidates.

SUPPLEMENTARY FIGURE 5

The expression level of *evm.TU.group6.1115* in the species with different fruit size. ** represents that there is a significant change in the two samples.

SUPPLEMENTARY TABLE 1

Information on wax apple accessions and sequencing data.

SUPPLEMENTARY TABLE 2

Reference organellar genomes of *S. samarangense*.

SUPPLEMENTARY TABLE 3

GO enrichment analysis of dispensable genes.

SUPPLEMENTARY TABLE 4

Pfam domain enrichment analysis of dispensable genes.

SUPPLEMENTARY TABLE 5

Cultivar names and their key phenotypic traits.

SUPPLEMENTARY TABLE 6

Genes associated with cold tolerance.

SUPPLEMENTARY TABLE 7

Genes associated with fruit size.

SUPPLEMENTARY TABLE 8

XP-CLR candidates for cold tolerance.

SUPPLEMENTARY TABLE 9

XP-CLR candidates for fruit size.

References

- Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. Available online at: <http://genome.cshlp.org/content/19/9/1655.abstract> (Accessed July 31, 2009).
- Banadka, A., Wudali, N. S., Al-Khayri, J. M., and Nagella, P. (2022). The role of *Syzygium samarangense* in nutrition and economy: An overview. *South Afr. J. Bot.* 145, 481–492. doi: 10.1016/j.sajb.2022.03.014
- Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. doi: 10.1093/nar/27.2.573
- Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* 38, 5825–5829. doi: 10.1093/molbev/msab293
- Charron, J.-B. F., Breton, G., Danyluk, J., Muzac, I., Ibrahim, R. K., and Sarhan, F. (2002). Molecular and biochemical characterization of a cold-regulated phosphoethanolamineN-methyltransferase from wheat. *Plant Physiol.* 129, 363–373. doi: 10.1104/pp.001776
- Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. *Genome Res.* 20, 393–402. doi: 10.1101/gr.100545.109
- Chen, N. (2004). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinforma.* 5, 4–10. doi: 10.1002/0471250953.bi0410s05
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560
- Chen, Y., Wang, Y., and Zhang, H. (2014). Genome-wide analysis of the mildew resistance locus o ('MLO') gene family in tomato ('*Solanum lycopersicum*' L.). *Plant Omics* 7, 87–93. doi: 10.3316/informit.319919300137901
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *Gigascience* 10, giab008. doi: 10.1093/gigascience/giab008
- de Lorenzo, L., Merchan, F., Laporte, P., Thompson, R., Clarke, J., Sousa, C., et al. (2009). A novel plant leucine-rich repeat receptor kinase regulates the response of medicago truncatula roots to salt stress. *Plant Cell* 21, 668–680. doi: 10.1105/tpc.108.059576
- Eddy, S. R. (2023). HMMER: biosequence analysis using profile hidden Markov models. Available online at: <http://Hmmer.Org/> (Accessed August 2023).
- Feechan, A., Jermakow, A. M., Torregrosa, L., Panstruga, R., and Dry, I. B. (2008). Identification of grapevine MLO gene candidates involved in susceptibility to powdery mildew. *Funct. Plant Biol.* 35, 1255–1266. doi: 10.1071/FP08173
- Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., et al. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U. S. A.* 117, 9451–9457. doi: 10.1073/pnas.1921046117
- Foot, H. C., Ride, J. P., Franklin-Tong, V. E., Walker, E. A., Lawrence, M. J., and Franklin, F. C. (1994). Cloning and expression of a distinctive class of self-incompatibility (S) gene from *Papaver rhoeas* L. *Proc. Natl. Acad. Sci.* 91, 2265–2269. doi: 10.1073/pnas.91.6.2265
- Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565
- Gao, L., Gonda, I., Sun, H., Ma, Q., Bao, K., Tieman, D. M., et al. (2019). The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.* 51, 1044–1051. doi: 10.1038/s41588-019-0410-2
- Golicz, A. A., Martinez, P. A., Zander, M., Patel, D. A., Van De Wouw, A. P., Visendi, P., et al. (2015). Gene loss in the fungal canola pathogen *Leptosphaeria maculans*. *Funct. Integr. Genomics* 15, 189–196. doi: 10.1007/s10142-014-0412-1
- Grabherr, M., Haas, B., Yassour, M., Levin, J. Z., Thompson, D. A., and Amit, I. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883
- Gupta, K., Wani, S. H., Razzaq, A., Skalicky, M., Samantara, K., Gupta, S., et al. (2022). Abscissic acid: role in fruit development and ripening. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.817500
- Herridge, R., McCourt, T., Jacobs, J. M. E., Mace, P., Brownfield, L., and Macknight, R. (2022). Identification of the genes at S and Z reveals the molecular basis and evolution of grass self-incompatibility. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1011299
- Huang, C. C. (2017). "Innovation of wax apple industry in Taiwan," in *Acta Horticulturae* (International Society for Horticultural Science (ISHS, Leuven, Belgium), 1–6. doi: 10.17660/ActaHortic.2017.1166.1
- Hubisz, M. J., Falush, D., Stephens, M., and Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.* 9, 1322–1332. doi: 10.1111/j.1755-0998.2009.02591.x
- Hübner, S., Bercovich, N., Todesco, M., Mandel, J. R., Odenheimer, J., Ziegler, E., et al. (2019). Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. *Nat. Plants* 5, 54–62. doi: 10.1038/s41477-018-0329-0
- Khandaker, M. M., and Boyce, A. N. (2016). Growth, distribution and physiochemical properties of wax apple (*Syzygium samarangense*): A Review. *Aust. J. Crop Sci.* 10, 1640–1648. doi: 10.21475/ajcs.2016.10.12.PNE306
- Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5, R12. doi: 10.1186/gb-2004-5-2-r12

- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, X., Ahmad, S., Guo, C., Yu, J., Cao, S., Gao, X., et al. (2018). Identification and characterization of LRR-RLK family genes in potato reveal their involvement in peptide signaling of cell fate decisions and biotic/abiotic stress responses. *Cells* 7, 120. doi: 10.3390/cells7090120
- Manzanares, C., Barth, S., Thorogood, D., Byrne, S. L., Yates, S., Czaban, A., et al. (2016). A gene encoding a DUF247 domain protein cosegregates with the S self-incompatibility locus in perennial ryegrass. *Mol. Biol. Evol.* 33, 870–884. doi: 10.1093/molbev/msv335
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015
- Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P., O'Hara, B., et al. (2015). Package “vegan” - community ecology package. *R News* 8, 48–50. Available online at: <https://github.com/vegandevs/vegan>.
- Qi, X., Liu, C., Song, L., Li, Y., and Li, M. (2017). PaCYP78A9, a cytochrome P450, regulates fruit size in sweet cherry (*Prunus avium* L.). *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.02076
- Qin, P., Lu, H., Du, H., Wang, H., Chen, W., Chen, Z., et al. (2021). Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell* 184, 3542–3558.e16. doi: 10.1016/j.cell.2021.04.046
- Rausch, T., Zichner, T., Schlattl, A., Stütz, A. M., Benes, V., and Korbel, J. O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28, i333–i339. doi: 10.1093/bioinformatics/bts378
- Reay, P. F., and Lancaster, J. E. (2001). Accumulation of anthocyanins and quercetin glycosides in 'Gala' and 'Royal Gala' apple fruit skin with UV-B-Visible irradiation: modifying effects of fruit maturity, fruit side, and temperature. *Sci. Hortic. (Amsterdam)* 90, 57–68. doi: 10.1016/S0304-4238(00)00247-8
- Sashankar, P., and Chakraborty, B. (2023). Phytoalexins from Poaceae : their types, biosynthesis, role in biotic and abiotic stress adaptation and application in human health. *J. Mycopathol. Res.* 61, 287–300. doi: 10.57023/JMycR.61.3.2023.287
- Shiu, S. H., and Bleecker, A. B. (2001). Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. *Proc. Natl. Acad. Sci. U. S. A.* 98, 10763–10768. doi: 10.1073/pnas.181141598
- Shü, Z.-h., Meon, Z., Tirtawinata, R., and Thanarut, C. (2008). “wax apple production in selected tropical asian countries,” in *Acta Horticulturae* (International Society for Horticultural Science (ISHS, Leuven, Belgium), 161–164. doi: 10.17660/ActaHortic.2008.773.22
- Shü, Z.-H., Shiesh, C.-C., and Lin, H.-L. (2011). “23 - Wax apple (*Syzygium samarangense* (Blume) Merr. and L.M. Perry) and related species,” in *Woodhead Publishing Series in Food Science, Technology and Nutrition*. Ed. S. F. Yahia (Cambridge, UK: Woodhead Publishing), 458–475e. E. M. B. T.-P. BT. @ of T. doi: 10.1533/9780857092618.458
- Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* 34, W435–W439. doi: 10.1093/nar/gkl200
- Vernikos, G., Medini, D., Riley, D. R., and Tettelin, H. (2015). Ten years of pan-genome analyses. *Curr. Opin. Microbiol.* 23, 148–154. doi: 10.1016/j.mib.2014.11.016
- Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., et al. (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557, 43–49. doi: 10.1038/s41586-018-0063-9
- Wei, X., Chen, M., Zhang, X., Wang, Y., Li, L., Xu, L., et al. (2023). The haplotype-resolved autotetraploid genome assembly provides insights into the genomic evolution and fruit divergence in wax apple (*Syzygium samarangense* (Blume) Merr. and Perry). *Hortic. Res.* 10, uhad214. doi: 10.1093/hr/uhad214
- Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innov* 2, 10041. doi: 10.1016/j.xinn.2021.100141
- Xiang, Y., Cao, Y., Xu, C., Li, X., and Wang, S. (2006). Xa3, conferring resistance for rice bacterial blight and encoding a receptor kinase-like protein, is the same as Xa26. *Theor. Appl. Genet.* 113, 1347–1355. doi: 10.1007/s00122-006-0388-x
- Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., et al. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 39, W316–W322. doi: 10.1093/nar/gkr483
- Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., et al. (2014). SOAPdenovo-Trans: *de novo* transcriptome assembly with short RNA-Seq reads. *Bioinformatics* 30, 1660–1666. doi: 10.1093/bioinformatics/btu077
- Zhang, J., Li, Z., Liang, Y., He, S., Guo, G., Yang, S., et al. (2024). Diploid wax apple (*Syzygium samarangense*) genome identified NAC genes regulating fruit development. *Hortic. Res.* 11, uhae025. doi: 10.1093/hr/uhae025
- Zhang, C., Nielsen, R., and Mirarab, S. (2023). CASTER: Direct species tree inference from whole-genome alignments. *bioRxiv* 2023, 10.04.560884. doi: 10.1101/2023.04.560884
- Zhu, X., Shen, W., Huang, J., Zhang, T., Zhang, X., Cui, Y., et al. (2018). Mutation of the osSAC1 gene, which encodes an endoplasmic reticulum protein with an unknown function, causes sugar accumulation in rice leaves. *Plant Cell Physiol.* 59, 487–499. doi: 10.1093/pcp/pcx203
- Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., and Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics* 29, 2669–2677. doi: 10.1093/bioinformatics/btt476
- Zu, X., Luo, L., Wang, Z., Gong, J., Yang, C., Wang, Y., et al. (2023). A mitochondrial pentatricopeptide repeat protein enhances cold tolerance by modulating mitochondrial superoxide in rice. *Nat. Commun.* 14, 6789. doi: 10.1038/s41467-023-42269-4