DNA Research, 2025, **32**, dsaf007
https://doi.org/10.1093/dnares/dsaf007
Advance access publication 15 May 2025
Resource Article: Genomes Explored

OXFORD

Resource Article: Genomes Explored

A high-quality genome assembly of *Annona squamosa* (custard apple) provides functional insights into an emerging fruit crop

Manohar S. Bisht, Shruti Mahajan, Abhisek Chakraborty, Vineet K. Sharma*

MetaBioSys Group, Department of Biological Sciences, Indian Institute of Science Education and Research Bhopal, Bhopal – 462066, Madhya Pradesh, India

*Corresponding author. MetaBioSys Group, Department of Biological Sciences, Indian Institute of Science Education and Research Bhopal, Bhopal – 462066, Madhya Pradesh, India –(Email: vineetks@iiserb.ac.in)

Abstract

Annona squamosa, also known as custard apple, is an emerging fruit crop with medicinal significance. We constructed a high-quality genome of *A. squamosa* along with transcriptome data to gain insights into its phylogeny, evolution, and demographic history. The genome has a size of 730.4 Mb with an N50 value of 93.2 Mb assembled into seven pseudochromosomes. The demographic history showed a continuous decline in the effective population size of *A. squamosa*. Phylogenetic analysis revealed that magnoliids were sister to eudicots. Genome syntenic and Ks distribution analyses confirmed the absence of a recent whole-genome duplication event in the *A. squamosa*. Gene families related to photosynthesis, oxidative phosphorylation, and plant thermogenesis were found to be highly expanded in the genome. Comparative analysis with other magnoliids revealed the adaptive evolution in the genes of flavonoid biosynthesis pathway, amino sugar, nucleotide sugar and sucrose metabolism, conferring medicinal value, and enhanced hexose sugar accumulation. In addition, we performed genome-wide identification of *SWEET* genes. Our high-quality genome and evolutionary insights of this emerging fruit crop, thus, serve as a valuable resource for advancing studies in functional genomics, evolutionary biology, and crop improvement.

Keywords: *Annona squamosa*, emerging fruit crop, genome assembly, adaptive evolution, sugar metabolism.

1. Introduction

Fruit crops have been under cultivation for centuries and are an integral part of human life. Many fruiting crops have gained much importance in their breeding and domestication, like Apple (*Malus domesticus*), Mango (*Mangifera indica*), Grape (*Vitis* sp.), and Banana (*Musa* sp.). However, many fruit crops with commercial potential are currently classified as 'emerging' fruit crops.[1,2] One such family of underutilized fruiting crops is the Annonaceae family, the largest family of the order Magnoliales, majorly distributed in tropical regions of the world.[3] Genus *Annona*, *Asimina*, *Rollinia*, and *Uvaria* are the four edible fruit-bearing genera in Annonaceae, among which *Annona* is the most important source of edible fruits, which consist of some important species of this genus (*A. squamosa, A. reticulata, A. muricata, A. cherimola,* and *A. mucosa*), in which, custard apple is the most widely cultivated species.[4]

A. squamosa ($2n = 14$) is a small deciduous tree (about 3 to 6 m in height). The leaves are alternate, ovate, or elliptic-oblong. Flowers are pendulous, fragrant, and yellowish green with three sepals and three petals. The fruit is a syncarp and irregularly heart-shaped[3] (**Fig. 1A–C**). Custard apple is natively from the New World tropics and most widely cultivated *Annona* spp. in the tropical regions of Africa, America, Asia, and the Pacific.[3]

The extracts from various parts of custard apple, such as roots, bark, leaves, and fruit, possess a wide range of ethnomedicinal importance conferring to its anticancer, antioxidant, antiparasitic, antimalarial, antidiabetic, insecticidal, microbicidal antihypertensive, hepatoprotective, and molluscicidal activities.[5] Phytochemical analysis of extracts from seed, leaf, bark, and roots reveals that the plant is filled with secondary metabolites like alkaloids, flavonoids, and tannins.[6] In addition to the immense medicinal properties, the fruit of custard apple is referred to as 'one of the most delicious fruits known to man' because of the high sugar content (~28%) compared to other sweet fruits such as mango (~15%) and banana (~20%).[7,8] Further, among the other *Annona* species, *A. squamosa* has a higher total soluble sugar content.[9] However, the genome of *A. squamosa*, the most widely cultivated species of the genus *Annona*, was not yet available.[10,11] Therefore, in this study, to construct the genome of *A. squamosa*, we used Oxford Nanopore Technology (ONT) and 10× linked reads technologies to sequence the genome and also reported the leaf transcriptome. Further, we examined the demographic history of *A. squamosa*, exploring population dynamics over time, and investigated its phylogenetic placement providing context for its evolutionary relationships with related species. Furthermore, we analysed the evolutionary

Received 15 January 2025; Revised 3 March 2025; Accepted 10 May 2025

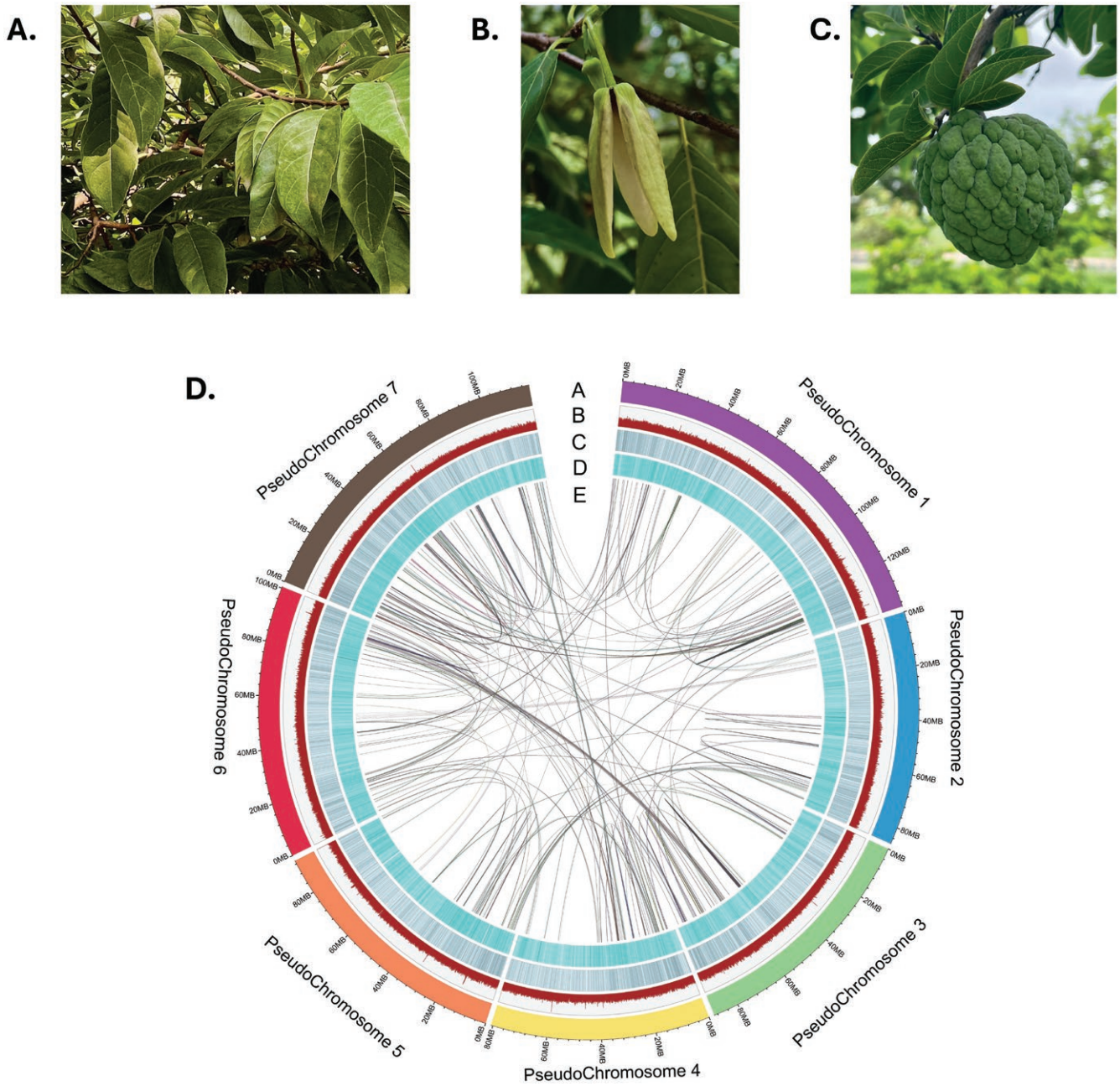

Fig. 1. Morphology and genome assembly characteristics of *A. squamosa*. A. Leaves B. Flower C. Fruit D. Overview of genome assembly a. Assembled pseudochromosomes b. GC density c. Repeat density d. Gene density e. Chromosome synteny.

history of *A. squamosa* with a focus on gene family evolution, whole-genome duplication (WGD) events and adaptive evolution by comparing it with other members of Magnoliales. This genome resource will support future studies on genetic diversity, adaptive traits, and the biosynthesis of bioactive compounds, contributing to a deeper understanding of the medicinal potential and evolutionary biology of custard apple.

2. Materials and methods

2.1 Sample collection, species identification, and DNA extraction

The leaves from *A. squamosa* were collected from IISER Bhopal campus, Bhopal, India (23.2599°N, 77.4126°E). The DNA extraction was performed using DNeasy Plant mini kit (Qiagen). The extracted DNA was used for amplifying *matK* DNA marker region for species identification (Fig. S1). The amplicons were sequenced on Sanger sequencer, followed by aligning with NCBI nucleotide (nt) database using blastn. The species was identified as *A. squamosa* with 99.88% identity of *matK*. After species identification, high molecular weight DNA was extracted from the leaves sample. So, the extraction buffer previously used for DNA extraction from *A. muricata* was modified for *A. squamosa* DNA extraction.[12] The extraction buffer [3% CTAB (cetyl trimethyl ammonium bromide), 1M Tris-HCl, 0.7M NaCl, 0.05M EDTA (ethylenediamine tetra acetic acid), 3% PVP-40, 1% β-mercaptoethanol (freshly added)] was pre-heated for 30 min at 65 °C. The leaves used

were already frozen for a few days to reduce the amount of photosynthesis products. The leaves were homogenized in liquid nitrogen using a pre-cooled autoclaved mortar pestle. The powdered leaves were added to 1 mL of pre-heated extraction buffer and mixed by inversion. Proteinase K (25 µL) and RNase A (2 µL) were added for protein and RNA hydrolysis, respectively. The tubes were mixed by inversion and incubated for 12 h at 65 °C. Upon lysis, the DNA was purified twice using chloroform: iso-amyl alcohol (24:1) in equal volume to remove the organic contaminants. The aqueous phase was taken and 0.7× ice-cold isopropanol was added to precipitate the DNA. An overnight incubation at –20 °C was done to facilitate DNA precipitation. After this incubation, the DNA was pelleted down and followed by washing with 70% ethanol. The washed DNA was dissolved in the G2 buffer of Blood and cell culture DNA mini kit (Qiagen), which was facilitated by incubation for 30 min at 50 °C. The G2 buffer with dissolved DNA was passed through equilibrated Genomic tip 20 and allowed to pass under gravitation pull. The Genomic tip 20 column was three times washed with 1 mL of QC buffer. The DNA was eluted in QF buffer (1 mL). For DNA precipitation, 0.7× isopropanol was added and kept at –20 °C for overnight. After incubation, the DNA was pelleted down, washed thrice with 70% ethanol, and air-dried to remove remnant ethanol. Finally, the pellet was suspended in 50 µL of nuclease-free water. The DNA sample was quantified on Qubit 2.0 fluorometer using a qubit ds DNA broad-range assay kit (Invitrogen, United States). The DNA samples were purified using 0.45× magnetic beads (Beckman Coulter, USA) for nanopore sequencing.

2.2 Genome sequencing

To prepare a linked read library on a Chromium platform, the extracted genomic DNA was processed with the help of Chromium Genome Library Kit and Gel Bead Kit v2 (10× Genomics, USA) and the resultant linked read library was sequenced using the NovaSeq 6000 (Illumina, Inc., USA) with 150 bp paired-end sequencing protocol. The library preparation for nanopore sequencing was performed using ONT library preparation kits (SQK-LSK109 and SQK-LSK110). The libraries were sequenced on an in-house MinION Mk1C sequencer using FLO-MIN106 flowcells (**Table S1**).

2.3 RNA extraction and sequencing

The extraction buffer used for RNA extraction from *A. squamosa* leaves was the same as that used for *A. muricata* fruits.[13] The protocol starts with the leaf homogenate, which was transferred into the RNA extraction buffer. The lysis was performed for 15 min at 60 °C. After incubation, the samples were cooled down on an ice bath for 10 min. 150 µL of 5M Potassium acetate was added to cooled samples, mixed carefully, and centrifuged at 9,500 ×g for 20 min at room temperature. The supernatant was taken and mixed with 0.5 volume of 95% ethanol. The mixture was passed through the column of RNeasy mini kit (Qiagen). Further process was followed as per the manufacturer's protocol. Qubit 2.0 fluorometer was used to quantify the RNA using RNA HS assay kit (Invitrogen, USA). The extracted RNA was used for library preparation using TruSeq stranded total RNA library preparation kit along with Ribo-Zero Plant workflow. The library quality assessment was performed on Agilent TapeStation 4150 with a High Sensitivity D1000 ScreenTape. The library was loaded on NovaSeq 6000 (Illumina, Inc., USA) and sequenced for generating paired-end reads of 150 bp (**Table S1**).

2.4 Genome size and ploidy estimation

The genome size of *A. squamosa* was computationally estimated by preprocessed 10× linked reads using Jellyfish v2.3.0[14] that created 21-mer count histograms, which were later employed in GenomeScope 2.0[15] for genome size and heterozygosity estimation. Further, the ploidy estimation was performed using Smudgeplot v0.2.2.[15]

2.5 Genome assembly and assessment

Oxford Nanopore sequencing raw reads were base-called using Guppy v3.2.1 (Oxford Nanopore Technologies), and adapters were removed using Porechop v0.2.4 (Oxford Nanopore Technologies). *De novo* assembly was performed with different genome assemblers using default parameters: Wtdbg2,[16] Canu,[17] and Flye v2.9.1.[18] Based on Quast v5.0.2[19] statistics (N50, L50, assembled genome length and number of contigs), the Flye genome assembly was selected and polished using Pilon v1.23[20] in three iterations. The scaffolding of contig assembly was performed by ARCS v1.2.2[21] and LINKS v2.0.0[22] using barcode-filtered 10× linked reads, and adapter-free Nanopore reads, respectively. AGOUTI v0.3.3[23] was used to further scaffold the obtained assembly using trimmed and filtered paired-end transcriptome data produced from Trimmomatic v0.39[24] (assembly v01).

10× linked reads were assembled using Supernova v2.1.1[25] with the '-maxreads' option set to 681 million paired-end reads and other default parameters. The obtained assembly was corrected using barcode-processed 10× reads by Tigmint v 1.2.6.[26] Further, the assembly was scaffolded using ARCS v1.2.2[21] and LINKS v2.0.0[22] with the input of Longranger basic linked reads and adapter-free Nanopore reads, respectively. Further, scaffolding was done by AGOUTI v0.3.3[23] using trimmed paired-end transcriptome reads (assembly v02).

Assembly v01 and assembly v02 were merged using Quickmerge v0.3[27] to produce a hybrid assembly. Gap-closing was done by LR_Gapcloser[28] in five iterations and TGS_Gapcloser v1.2.1[29] using Nanopore reads. In this gap-closed assembly, final polishing was done using Pilon v1.23.[20] The assembly was further length-based filtered with ≥ 5kbp cut-off (assembly v03). Further, to orient and align the scaffolds into the pseudochromosomes, chromosome-level assembly of *A. cherimola* was used using Chromosemble in Satsuma v2[30,31] with default parameters. The obtained pseudochromosome-level assembly (assembly v04) was then gap-filled using TGS_Gapcloser v1.2.1[29] and polished with Pilon v1.23[20] (**Table S2**).

To assess the quality of assembly quality, Quast v5.2.2[19] and BUSCO v5.4.3[32] with the embryohyta_odb10 dataset were used. To further assess the contiguity of the assembly, barcode-filtered 10× linked reads, nanopore raw reads, and quality filtered transcriptomic reads were mapped to the final draft assembly using MiniMap2 v2.17,[33] BWA-MEM v0.7.17,[34] and HISAT2 v2.2.1,[35] respectively, and SAMtools v1.13[36] 'flagstat' utility was used to calculate mapping percentage.

2.6 Chloroplast assembly and annotation of *A. squamosa*

The chloroplast genome was assembled by GetOrganlle v1.7.7.0,[37] using filtered 10× reads and the resulting circular assembly was annotated by GeSeq[38] and visualized using the OGDRAW v1.3.1[39] program of CHLOROBOX.

2.7 Demographic history

To gain insights into the demographic history of *A. squamosa*, Pair-wise Sequentially Markovian Coalescent (PSMC)[40] was used. The filtered 10x short reads were mapped to the genome using BWA-MEM.[34] SAMtools v1.13[36] 'mpileup' and BCFtools v1.9[41] were used to form the consensus diploid sequence. Sites with quality scores < 20 were filtered from consensus sequences. PSMC analysis was performed using the parameters 'N30 -t5 -r5 -p4+25*2 + 4+6' with 100 bootstrap values. A generation time of 15 years and a per-generation mutation rate of 7×10^{-9} were used to infer the demographic history.[42,43]

2.8 Gene set construction and functional annotation

The pseudochromosome-level genome assembly *A. squamosa* was used to create the *de novo* repeat library using RepeatModeler v2.0.3.[44] Further, these repeats were clustered to remove the redundant sequences using CD-HIT-EST v4.8.1,[45] and the remaining unknown TEs were classified using TEclass2[46] software with a probability threshold of > 0.65 and resulting library was used to soft mask the repeats using RepeatMasker v4.1.2 (http://www.repeatmasker.org).

The soft-masked genome was used to construct the gene set using MAKER v3.01.04[47] in three rounds. In the first round, the evidence-based gene was predicted, and a *de novo* transcriptome assembly was constructed using Trinity[48] with transcriptome data from our study and previous study[49] (**Table S1**). Further, protein sequences from other magnoliids species *(A. muricata*, *A. cherimola*, *L. chinense*, and *M. grandiflora*) were used. In the second and third rounds of MAKER *ab initio* gene prediction was performed using AUGUSTUS v3.3.3[50] and SNAP v1.0.[51] The obtained gene models were then filtered for AED value < 0.5 and transcript length ≥ 150 bp.

Additionally, Barrnap v0.9 (https://github.com/tseemann/barrnap), miRBase database[52] (sequence identity 80%, e-value 10–3) and tRNAscan v2.0.9,[53] for rRNA, miRNA, and tRNA predictions, respectively.

The protein coding gene set was functionally annotated by performing BLASTP against NCBI-nr, Pfam-A,[54]and Swiss-Prot databases[55] with an e-value of 10^{-5}. Further, *A. squamosa* gene set, including the genes in highly expanded gene families and evolutionary signatures, were annotated using eggNOG-mapper v2.1.12[56] and KAAS v2.1[57]

2.9 Phylogeny construction and analysis of evolution in gene families

To resolve the phylogenetic position of *A. squamosa*, protein sequences of 23 plant species along with *A. squamosa* proteins were selected, which include eight Magnoliids (*Magnolia grandiflora, Liriodendron chinense*, *Annona muricata*, *Annona montana*, *Annona cherimola*, *Laurelia sempervirens*, *Aristolochia fimbriata*, *Piper nigrum*), six Eudicots (*Cynara cardunculus*, *Nicotiana attenuata*, *Glycine max*, *Manihot esculenta, Arabidopsis thaliana*, *Vitis vinifera*), 6 Monocots (*Dioscorea rotundata*, *Brachypodium distachyon, Musa acuminata*, *Oryza sativa*, *Sorghum bicolour*, *Zea mays*), 2 ANA-grade angiosperms (*Amborella trichopoda, Nymphaea colorata*), and 1 Gymnosperm as outgroup (*Ginkgo biloba*) (**Table S3**). For each protein, gene family clustering was performed on its longest isoforms using OrthoFinder v2.5.4.[58] The obtained orthogroups were further filtered to extract the fuzzy one-to-one orthogroups using KinFin v1.1.[59] The resulting orthogroups were aligned using MAFFT v7.310.[60] BeforePhylo v0.9.0 (https://github.com/qiyunzhu/BeforePhylo) was used to filter and concatenate the alignments. A maximum-likelihood tree was constructed using RAxML v8.2.12[61] with 500 bootstrap values and 'PROTGAMMAAUTO' substitution model. The divergence time was estimated using the MCMCtree implemented in PAML v4.10.6[62] with the following calibration points: *A. trichopoda* and *N. colorata* (179.9 to 205.0) million years ago (mya) and *Z. mays* and *M. acuminata* (103.2 to 117.1) mya.

To examine the evolution in gene families, proteome files containing the longest protein isoforms were analysed using CAFÉ v5.[63] As suggested for CAFÉ analysis, all vs all BLASTP was performed to cluster and filtering of gene families. Using the 163 million-year calibration period between *A. squamosa* and *N. attenuata*, the phylogenetic tree was transformed into an ultrametric tree (https://timetree.org/). The CAFÉ v5 with two-lambda (λ) model was performed using the ultrametric tree and filtered gene, in which species from Annonaceae family were assigned separated λ-value compared to other species. In addition, the protein sequences of five *Annona* species were used to find core gene family clusters and species-specific gene clusters using OrthoVenn3.[64]

2.10 Genome synteny and whole-genome duplication analysis

Synteny analysis was performed for inter-species (*A. squamosa—A. cherimola*) and intra-species (*A. squamosa-A. squamosa*) using MCScanX.[65] Further, to infer WGD history in *A. squamosa* genome, whole-genome Ks (substitutions per synonymous site) distribution was performed for paralogs and orthologs including *A. cherimola, A. montana,* and *L. chinense* using wgd2.[66] Further, the rough dating of the genome duplication events was done using the Ks values. The peaks of Ks profiles were determined and translated into divergence time (*T*) in millions of years using the following formula: $T = Ks/2r$, where *T* is the time of divergence and r is the synonymous-site mutation rate (7×10^{-9} sites/year).[42,43]

2.11 Adaptive evolutionary signatures in *A. squamosa*

For comprehensive insights into the adaptive evolutionary traits in *A. squamosa* genome, comparative evolutionary analysis was performed with 10 other magnoliids species, in which 6 species were from Magnoliales order (*A. montana*, *A. muricata*, *A. cherimola, M. biondii, M. grandiflora, L. chinense*), 2 from Piperales (*P. nigrum*, *A. fimbriata*) 2 from Laurales (*L. sempervirens*, *C. kanehirae*).

2.11.1 *A. squamosa* genes with unique substitution with functional impact

The orthogroups were constructed from the protein sequence of these 11 species using OrthoFinder v2.5.4.[58] Each orthogroup was filtered to keep the longest sequence per species, which was later aligned using MAFFT v7.310.[60] The genes of *A. squamosa* that had different amino acid in position as compared to other species were called genes with unique amino acid substitution. Gaps and 10 positions around the gap were excluded from the analysis. Further, to screen the functional impact of these uniquely

Table 1. Genome assembly and annotation statistics of *Annona squamosa*.

Feature	*A. squamosa*
Final pseudochromosome assembly	
Total length (bp)	730,367,479
Number of scaffolds	58
Longest scaffold (Mb)	137.68
Scaffold N50 (Mb)	93.21
Scaffold L50	4
Scaffold GC content	34.32%
BUSCO (complete)	94.1%
Gene models	
Number of coding genes obtained from MAKER	48,633
Number of coding genes with AED value < 0.5 length-based filtering (high confidence gene set)	40,435
Repeat content of the genome (%)	64.82
Non-protein-coding genes	
Number of rRNAs	491
Number of tRNAs (decoding standard amino acids)	1,076
Number of miRNAs (homology-based)	95

substituted genes, SIFT was used with UniProt as a reference database.[67]

2.11.2 *A. squamosa* genes with positive selection

The orthogroups nucleotide sequences of all 11 species were aligned using MAFFT v7.310.[60] The resulting PHYLIP format alignments with the phylogenetic tree of 11 species were used for the positive selection analysis by implementing the branch-site model in the 'codeml' program of PAML v4.10.6.[62] For the statistical significance, a likelihood-ratio test and chi-square analysis were performed, and genes that qualified the threshold against the null model with FDR-corrected ($P < 0.05$) were named positively selected genes. These positively selected genes were then further identified for the positively selected codon sites (> 95% probability) for the foreground lineage using Bayes empirical Bayes analysis.[68,69]

2.12 Genome-wide Identification of *SWEET* genes and phylogenetic analysis

Protein sequences for *SWEET* (sugars will eventually be exported transporter) gene were identified in the proteome of *A. squamosa* and other two *Annona* species (*A. montana* and *A. cherimola*), using the hidden Markov model profiles of the MtN3_slv domain for the *SWEET* gene family (PF03083) which were downloaded from the Pfam database (http://pfam.xfam.org/) and used as query in HMMER software with e-value 10^{-5}. All the resulting putative genes were then screened for the presence of the MtN3_slv domain using SMART (http://smart.embl-heidelberg.de/), and sequences with at least one MtN3_slv domain were retained.

To investigate and categorize these SWEET genes, the TAIR database (https://www.arabidopsis.org/) was used to retrieve the *Arabidopsis thaliana* AtSWEET proteins. MAFFT v7.310[60] was used to align the genes, and BeforePhylo v0.9.0 (https://github.com/qiyunzhu/BeforePhylo) was used to filter and concatenate the alignments. RAxML v8.2.12 with 'PROTGAMMAAUTO' substitution model and 500 bootstrap values was used to create a maximum-likelihood tree.[61]

3. Results

3.1 Sequencing summary, genome size, ploidy, and heterozygosity estimation

DNA sequencing generated 111.8 Gb (150× coverage) using 10× linked reads sequenced on Illumina platform and 16.5 Gb (22× coverage) data from Nanopore sequencing. Transcriptome sequencing generated 15.8 Gb of raw short-reads data from the leaf tissues (**Table S1**). The estimated genome size was 737.7, with diploid ploidy and 0.18% heterozygosity (**Fig. S2**).

3.2 Genome assembly construction and quality assessment

The final pseudochromosome-level genome assembly had a genome size of 730.4 Mbp with N50 of 93.24 Mbp and GC of 34.32% (**Fig. 1D**, **Table 1**). The assembly showed 94.1% complete BUSCO (**Table S4**), indicating its completeness. Further, 91.05% of the barcode-filtered linked reads, 91.33% of nanopore reads, and 95.57% of transcriptome data mapped to the final genome assembly.

3.3 Gene set construction and functional annotation

A total of 64.82% of repetitive regions were identified in the *A. squamosa* genome, with 52.13% as interspersed repeats. Retroelements consisted of 29.96% LTR (long terminal repeat) elements (12.98% Ty1/Copia and 19.00% Gypsy/DIRS1 elements) (**Table S5**). MAKER predicted 48,633 protein-coding genes. After AED value-based (< 0.5) and length-base (≥ 150 bp) filtering, 40,435 high-confidence genes were retained, and overall, 88% of these genes were functionally annotated against the publicly available databases (**Table S6**). Further, these genes showed 88.3% of complete and fragmented BUSCO (**Table S4**). The distribution of coding genes in KEGG pathways and KOG categories is listed in **Tables S7 and S8**, respectively. Moreover, non-coding RNAs were also identified in *A. squamosa* genome assembly (**Table 1**).

3.4 Chloroplast assembly and annotation

The resulting chloroplast genome of *A. squamosa* is a typical quadripartite circular structure of length 202,915 bp consisting of a large single copy of 69,733 bp, a small single copy (SSC) of 3,028 bp, and two inverted repeats (IRA and IRB) of length 65,077 bp each and coding for 136 protein-coding genes and 37 tRNA genes (**Fig. S3**). The highly expanded IRA and IRB regions make the chloroplasts bigger in size compared to other reported plastomes from the magnoliids clade.[70]

3.5 Demographic history of *A. squamosa*

PSMC results indicated a continuous decline in the effective population size of *A. squamosa*, correlating with its low heterozygosity of ~0.18% (**Fig. 2**, **Fig. S2A**). This might link to the Quaternary period's contraction of tropical regions, suggesting that *A. squamosa* faced significant challenges from climate changes, like many other tropical species.[43,71] In addition, *A. squamosa* might have experienced pronounced genetic bottlenecks during its domestication, further contributing to its reduced genetic diversity.[43,72,73]

3.6 Phylogeny construction and gene family evolution

The maximum likelihood (ML) tree based on 433 one-to-one fuzzy orthogroups with 500 bootstrap replicates across 23 angiosperms and one gymnosperm (outgroup) revealed that *A. squamosa* was most closely related to *A. cherimola*, which was sister to the subclade formed by *A. montana* (Annonaceae). Additionally, magnoliids were positioned as sisters to eudicots (**Fig. 3A**).

Gene family evolution analysis identified a total of 9,443 filtered gene families, among which 1,204 gene families were expanded, and 1,506 gene families were contracted in *A. squamosa*. Further, among these expanded gene families in *A. squamosa*, 17 were highly expanded (> 10 expanded genes). These highly expanded gene families were involved in photosynthesis, oxidative phosphorylation, biosynthesis of secondary metabolites, and thermogenesis (**Table S9**). In *A. squamosa,* expansion of ATPase subunit genes (*ATP1FA*, *ATP1FB*, *ATP1E*, *ATPeF1A*, and *ATPeF1B*), NAD(P)H-quinone oxidoreductase subunits (*ndhc*, *ndhj*, and *ndhk*) and *NDUFS7* was observed. These genes are actively involved in oxidative phosphorylation and thermogenesis, supporting the floral thermogenesis present in *A. squamosa*, a character linked with beetle-pollinated flowers.[74–76]

Gene clustering among *A. squamosa* and four other *Annona* species showed 9,806 core gene family clusters and 1,035 species-specific gene clusters in *A. squamosa* (**Fig. 3B**, **Table S10**). Gene ontology (GO) of a species-specific gene cluster of *A. squamosa* was enriched in biological processes like DNA recombination, transcription regulation, photosynthesis, and response to heat and salt stresses (**Table S11**). These specific genes associated with molecular mechanisms and both biotic and abiotic stress tolerance may enhance *A. squamosa* adaptability and resilience to environmental challenges.

3.7 Whole-genome duplication and synteny

The Ks distribution of orthologs revealed *A. squamosa* to be diverged from *L. chinense* at around Ks ~ 0.75 (53.4 mya), whereas *A. squamosa* diverged from *A. montana* at Ks ~ 0.14 (10 mya) and from *A. cherimola* around Ks ~ 0.06 (4.2 mya) (**Fig. 4A**), which is congruent with species phylogeny (**Fig. 3A**). Further, the paranome Ks distribution revealed no signs of independent WGD in *Annona* species after diverging. However, *L. chinense* experienced an independent WGD event at Ks ~0.69 after diverging from *Annona* (**Fig. 4B**).[77,78]

The intra-genomic synteny of *A. squamosa* showed 9.30% of collinearity, while the inter-genomic synteny between *A. squamosa* and *A. cherimola* showed 35.21% of collinearity. This high level of interspecies synteny indicates a significant similarity between the genomes of the two species, reflecting their relatively recent divergence. In general, the majority of the chromosomes of *A. squamosa* were aligned with corresponding regions of the chromosomes of *A. cherimola* in a one-to-one relationship (**Fig. 4C**).

3.8 Genes with adaptive evolution signatures in *A. squamosa*

In total, 1,781 orthogroups were found across 11 selected species of magnoliid clade to screen the genes with signatures

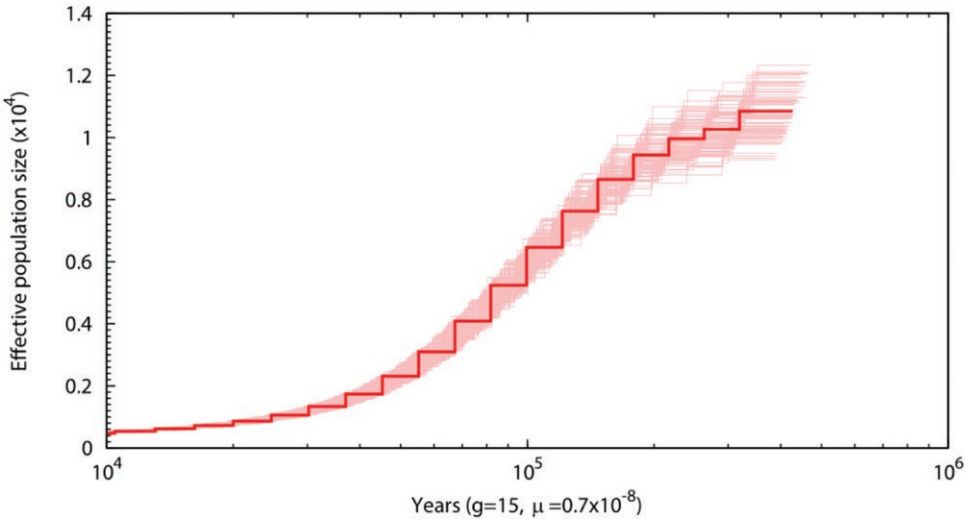

Fig. 2. Demographic history of *A. squamosa*. Light red lines represent the bootstrap values used in PSMC analysis.

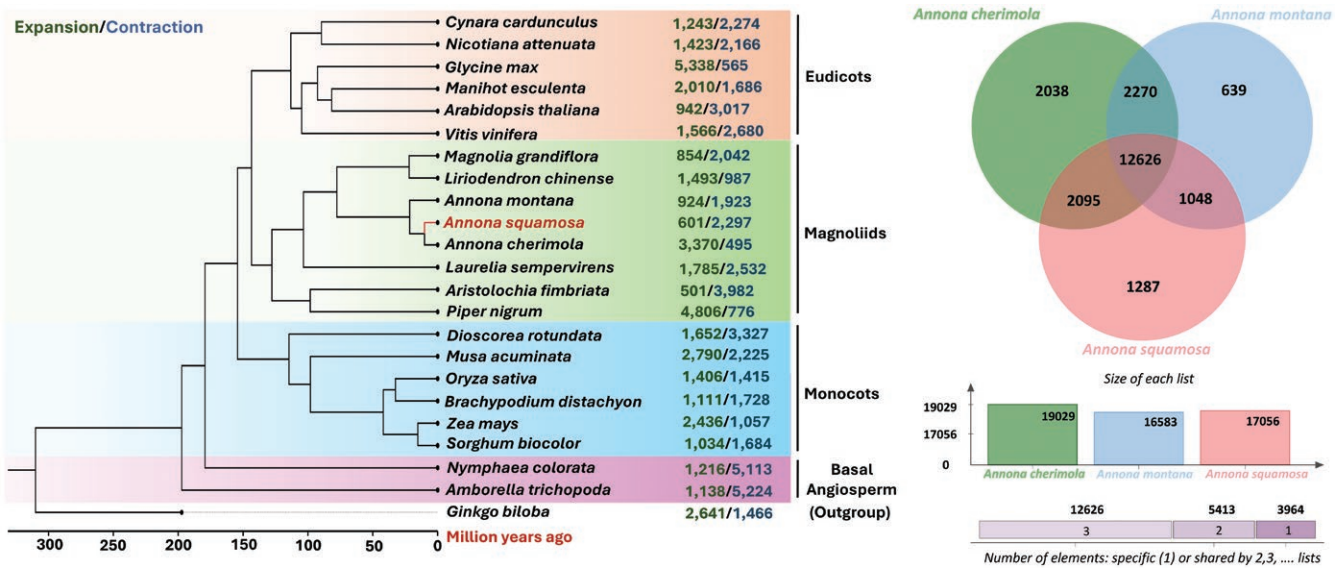

Fig. 3. A. Genome-wide species phylogenetic tree of *A. squamosa* with respect to eudicots and monocots. Numbers in green and blue represent the number of expanded and contracted gene families in each species, respectively. **B. Orthologous gene family cluster analysis of among *Annona* species.**

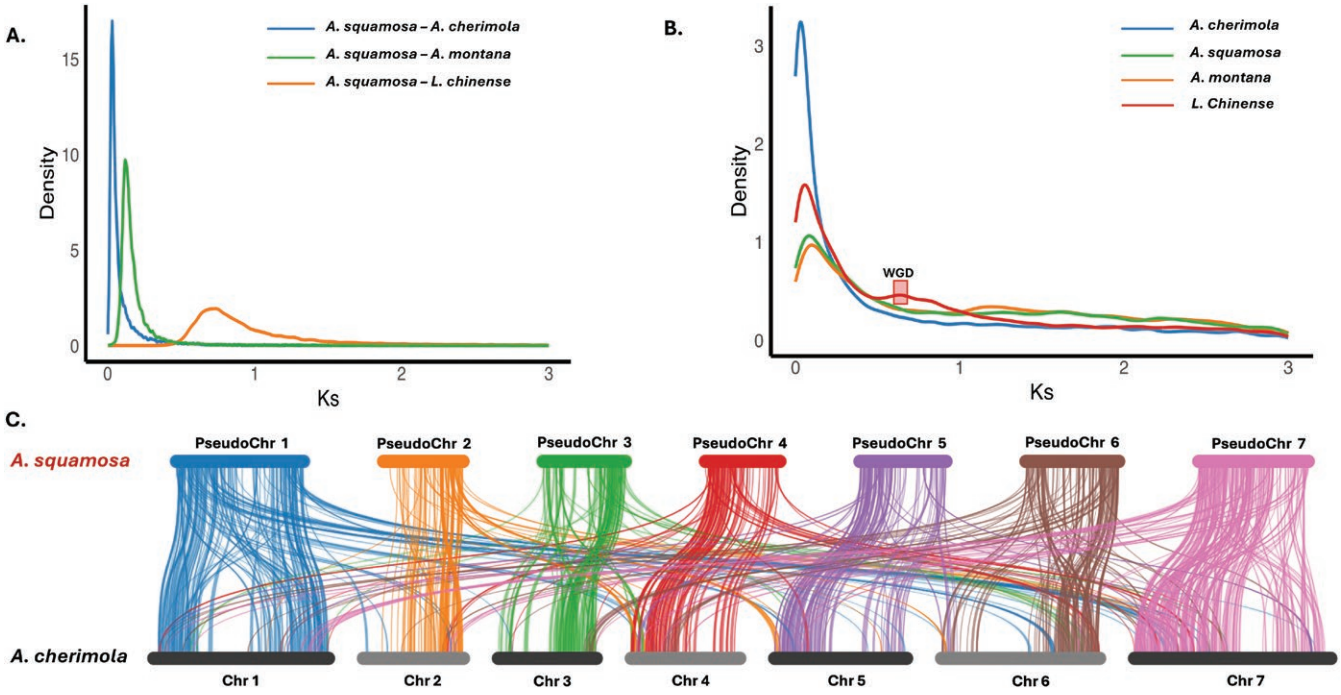

Fig. 4. Whole genome duplication history of *A. squamosa*. A. Ks distribution of paralogs B. Ks distribution of orthologs C. whole-genome synteny between *A. squamosa* and *A. cherimola*.

of adaptive evolution in *A. squamosa*. Comparative analysis revealed 260 genes with unique amino acid substitution with functional impact and 75 positively selected genes (P-value < 0.05) in *A. squamosa*. KEGG pathway analysis of the genes with signatures of adaptive evolution revealed that genes with unique amino acid substitution with functional impact were involved in starch and sucrose metabolism, glyoxylate and dicarboxylate metabolism, flavonoid biosynthesis, plant hormone signal transduction, and plant–pathogen interaction. In addition, positively selected genes were involved in starch and sucrose metabolism, purine metabolism, plant hormone signal transduction, and amino sugar and nucleotide sugar metabolism (**Tables S12 and S13**).

3.8.1 Flavonoid biosynthesis pathway

Flavonoids are structurally diverse secondary metabolites in plants that contribute significantly to plant growth and development.[79,80] Further, various phytochemical studies on *A. squamosa* revealed that various flavonoids, including rutin, kaempferol, quercetin, isorhamnetin, and farmarixetin, were rich in the leaves of plants in *Annona*, governing the medicinal importance.[6,81] In *A. squamosa,* genes involved in the

biosynthesis of various flavonoids, including anthocyanins, showed evolutionary signatures. The *F3'H*, *DFR*, and *ANS* genes showed unique amino acid substitution with functional impact, while the *F3'5'H* gene was positively selected (**Fig. 5**, **Table S14**).

3.8.2 Amino sugar, nucleotide sugar, and sucrose metabolism

A. squamosa fruit is one of the sweetest fruits, containing up to 28% of total sugars.[7,11] Comparative evolutionary analysis revealed the adaptive evolution in the genes involved in amino sugar, nucleotide sugar, and sucrose metabolism pathways. The *HK, UGP2*, and *TPP* had unique amino acid substitutions with functional impact, *GAUT* and *INV* were positively selected, and TPS showed both positive selection and unique amino acid substitution (**Fig. 6**, **Table S14**). These evolutionary modifications in the genome of *A. squamosa* likely contribute to the metabolic efficiency in facilitating high sugar content and enhancing the plant's adaptive responses to environmental pressures.

3.9 Genes encoding *SWEET* transporters

The SWEET family is a novel class of sugar transporters that enable the diffusion of sugars along a concentration gradient across cell membranes by acting as bidirectional uniporters.[82,83] A total of 13, 20, and 14 *SWEET* genes were Identified in the *A. squamosa, A. cherimola*, and *A. montana* genomes, respectively (**Table S15**). Phylogeny grouped these genes into four clades (**Fig. 7**). Among 13 genes of *A. squamosa*, 8 genes were categorized into clades I and II, primarily transporting glucose. Three genes were in clade III, which mainly transports sucrose, and 2 genes were in clade IV, which transports fructose.[84] Of these, clade I and IV genes were highly expressed in the leaf tissue compared to clade II and III (**Table S16**), suggesting a key role of fructose and glucose accumulation over sucrose in the sink tissues.

4. Discussion

In this study, we performed whole-genome sequencing of *A. squamosa*, which is an emerging fruit crop and provides the first high-quality genome of this species. The genome was diploid with an assembly size of 730 Mbp (**Table 1** and **Fig. S2**). The heterozygosity of *A. squamosa* (0.18%) was low compared to 1.05% of *A. cherimola*, but higher than *A. muricata* (0.06%). This reduced heterozygosity of *A. squamosa* may result from extensive domestication and hybridization.[72,73] Further, the low heterozygosity could also be a possible reason for the reduced effective population size (**Fig. 2**), particularly in comparison to the higher effective population size observed in the more heterozygous *A. cherimola* (**Fig. S4**). This decline in population size may raise challenges in the future in terms of the breeding of the plant.[2,72,85]

The pseudochromosome assembly of the *A. squamosa* genome provides a deeper understanding of the timing of the WGD event of the Annonaceae family. Our genome syntenic and Ks distribution analyses confirmed the absence of a recent WGD event in the *A. squamosa*, also observed in other *Annona* genomic studies[77,86] (**Fig. 4A** and **4B**). This absence of WGD may suggest that the genus retained a more ancestral genome structure (**Fig. 4C**).

Gene family evolution analysis reflects the evolutionary changes with respect to the ancestral species that are derived from various adaptive forces. These forces result in the expansion and contraction of gene families, resulting in lineage-specific adaptations that confer their distinct ecological and functional requirements.[87,88] The observed expansion of gene families involved in oxidative phosphorylation and thermogenesis in *A. squamosa* provides valuable insights into its floral biology. Thermogenesis facilitates the release of floral scents, serving as signals to attract insects by indicating the availability of food resources.[74,75]

Custard apple has been used in traditional medicines, governed by extracts obtained from various sections of the custard apple that contain phenol-based compounds, mainly alkaloids or flavonoids.[6] The presence of evolutionary signatures in the enzymes involved in flavonoid biosynthesis suggested the adaptive evolution of these pathways in *A. squamosa* (**Fig. 5**, **Table S14**).

Custard apple fruits are largely rich in reducing sugars (glucose and fructose) compared to non-reducing sugar (sucrose),[11,89] contributing to their sweetness and shorter shelf life, as sucrose generally enhances fruit shelf life.[90,91] Our genomic analysis revealed an adaptive evolution in genes associated with amino sugar, nucleotide sugar, and sucrose metabolism, suggesting a selective shift towards this high hexose

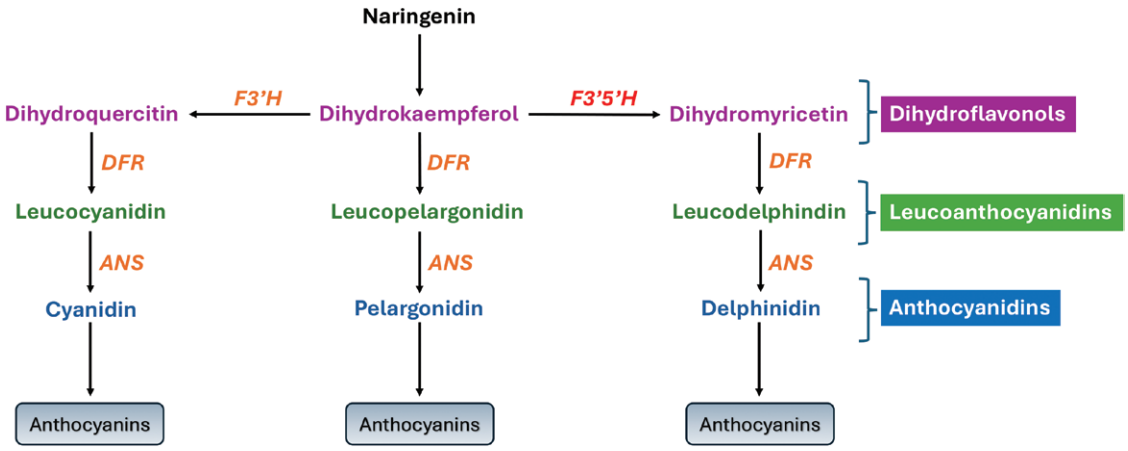

Fig. 5. Evolutionary signatures in flavonoid biosynthesis pathways. Red represents genes with positive selection, orange represents genes with amino acid substitution with functional impact. *F3'H*; flavanone 3-hydroxylase, *F3'5'H*; Flavonoid 3'5' hydroxylase, *DFR*; dihydroflavonol reductase, *ANS*; Anthocyanidin Synthase.

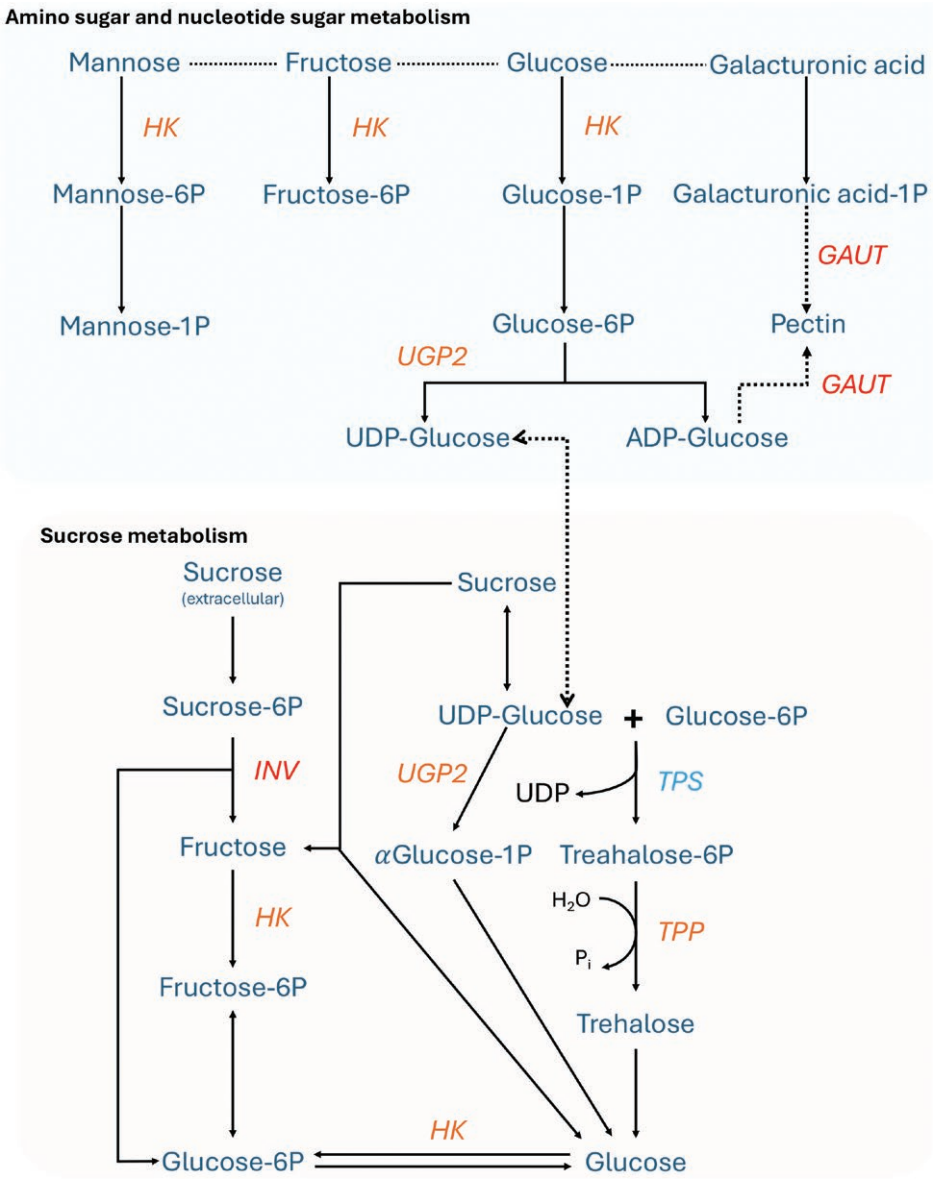

Fig. 6. Evolutionary signatures in Amino sugar, nucleotide sugar and sucrose metabolism pathway. Red represents genes with positive selection, orange represents genes with amino acid substitution with functional impact, and blue represents genes with both positive selection and unique amino acid substitution. *HK*; Hexokinase, *GAUT*; Galacturonosyltransferase, *INV*; Inevrtase, *UGP2*; UTP-glucose-1-phosphate uridylyltransferase 2, *TPS*; Trehalose-6-phosphate synthase, *TPP*; Trehalose-6-phosphate phosphatases.

sugar accumulation (Fig. 6, Table S14). Specifically, the *INV* gene, which encodes a key enzyme that hydrolyses sucrose into glucose, is positively selected.[92,93] Further, the *TPS* and *TPP* genes, which regulate sucrose levels by serving as signalling and homeostatic regulators, also exhibit adaptive changes.[94] In addition to that, the high expression of *SWEET* transporters genes related to clade I and IV over clade III suggest a preponderant role of fructose and glucose transport over sucrose transport in the sink tissues (Table S16). This, high hexose-to-sucrose ratio in early fruit development has been reported to increase mitotic activity through hexokinase (*HK*) signalling that increases cell number. Additionally, hexoses have higher osmotic potential than sucrose, causing them to retain more water, which increases fruit size and cell volume—a property that affects shelf life, nutritional qualities, and consumer acceptance.[91]

5. Conclusion

In summary, the high-quality genome of this emerging sweet fruit crop, along with evolutionary insights and pattern of gene family expansion, with phylogenetic relation, provides a deeper understanding of the adaptations that shape the evolution of *A. squamosa* within the Annonaceae family. This study, thus, serves as a valuable resource that will assist in the field of functional genomics, evolutionary biology, and crop improvement in the Annonaceae family.

Acknowledgements

MSB thanks the Ministry of Education, Govt. of India, for the Prime Minister Research Fellowship (PMRF). S.M. and A.C. thank the Council of Scientific and Industrial Research

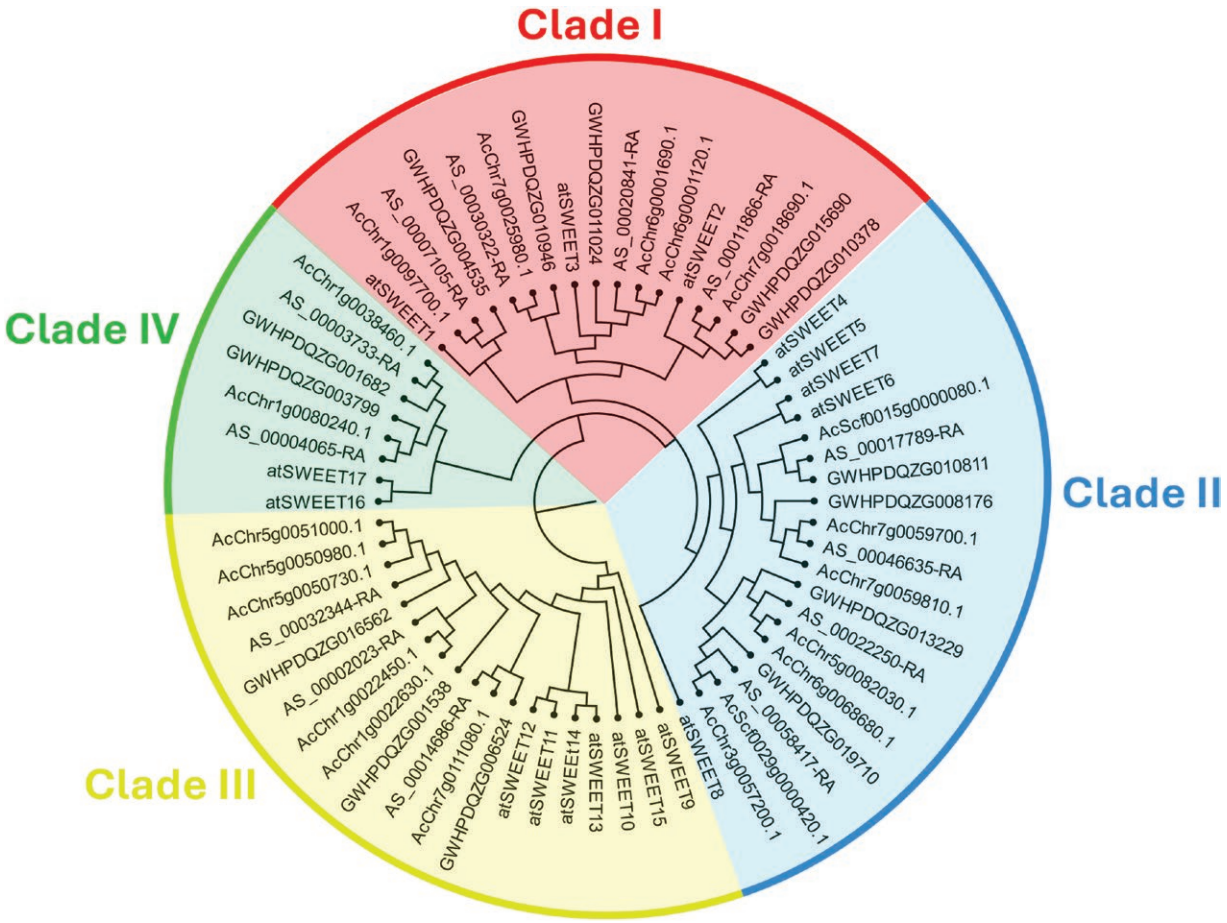

Fig. 7. Phylogenetic relationship analysis of 64 SWEET proteins from *Annona squamosa*, *Annona cherimola*, *Annona montana*, and *Arabidopsis thaliana*.

(CSIR) for the fellowship. The authors also thank the Sanger sequencing at IISER Bhopal and the intramural research funds provided by IISER Bhopal.

Author contributions

V.K.S. conceived and coordinated the project. S.M. performed DNA-RNA extraction, prepared the samples for sequencing, performed Nanopore sequencing, and species identification. M.S.B. and V.K.S. designed the computational framework of the study. M.S.B. performed all the computational analyses presented in the study with inputs from A.C. M.S.B. constructed all the figures. M.S.B. and V.K.S. interpreted the results. M.S.B., A.C., S.M., and V.K.S. wrote the manuscript. All the authors have read and approved the final version of the manuscript.

Funding

This project is funded by the intramural research funds provided by IISER Bhopal.

Data availability

Data associated with this study has been deposited in the NCBI SRA database under the BioProject accession – PRJNA1184485 and BioSample accession – SAMN44671926.

Conflict of interest

The authors declare no competing interests.

Supplementary material

Supplementary data are available at *DNARES* online.

References

1. Hummer KE, et al. Emerging fruit crops. *Fruit Breeding*. 2012:**8**:97–147.
2. Wang R, et al. Genomic insights into domestication and genetic improvement of fruit crops. *Plant Physiol*. 2023:**192**:2604–2627. https://doi.org/10.1093/plphys/kiad273
3. Datiles MJ, Acevedo-Rodríguez P. Annona squamosa (sugar apple). *CABI Compendium*. 2022.
4. Padmanabhan P, Paliyath G. Annonaceous fruits. *Encyclopedia of Food and Health*. 2016:169–173. https://doi.org/10.1016/B978-0-12-384947-2.00031-3
5. Ma C, Chen Y, Chen J, et al. A review on *Annona squamosa* L.: phytochemicals and biological activities. *The American journal of Chinese Medicine* 2017:**45**:933–964. https://doi.org/10.1142/S0192415X17500501
6. Kumar M, et al. Custard apple (*Annona squamosa* L.) leaves: nutritional composition, phytochemical profile, and health-promoting biological activities. *Biomolecules*. 2021:**11**:614. https://doi.org/10.3390/biom11050614
7. Dar MS, et al. Nutrient and flavor content of mango (*Mangifera indica* L.) cultivars: an appurtenance to the list of staple foods. *Nutritional Composition of Fruit Cultivars*. 2016:445–467.

8. Forster M, Rodríguez Rodríguez E, Darias Martín J, and Romero CD. Distribution of nutrients in edible banana pulp. *Technology and Biotechnology*. 2003:**41**:167–171.
9. Anuragi H, et al. Molecular diversity of Annona species and proximate fruit composition of selected genotypes. *3 Biotech*. 2016:**6**:1–10.
10. Gupta Y, et al. De novo assembly and characterization of transcriptomes of early-stage fruit from two genotypes of *Annona squamosa* L. with contrast in seed number. *BMC Genomics*. 2015:**16**:1–14.
11. Fang R, et al. Characterization of full-length transcriptome and mechanisms of sugar accumulation in *Annona squamosa* fruit. *Biocell*. 2020:**44**:737–750. https://doi.org/10.32604/biocell.2020.012933
12. Lira-Ortiz R, et al. Comparison of three genomic DNA extraction methods from soursop leaves (*Annona muricata* L.). *Mex J Biotechnol*. 2020:**5**:106–119.
13. Montenegro Brasil I, et al. Isolation of total RNA from ripe and unripe soursop (Annona muricata L.) fruit. *Afr J Plant Sci*. 2008:**2**:94–098.
14. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 2011:**27**:764–770. https://doi.org/10.1093/bioinformatics/btr011
15. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020:**11**:1–10.
16. Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. *Nat Methods*. 2019:**17**:155–158.
17. Koren S, et al. Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. *Genome Res*. 2017:**27**:722–736. https://doi.org/10.1101/gr.215087.116
18. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 2019:**37**:540–546.
19. Gurevich A, Saveliev V, Vyahhi N, Tesler G. Genome analysis QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013:**29**:1072–1075.
20. Walker BJ, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 2014:**9**:e112963. https://doi.org/10.1371/journal.pone.0112963
21. Yeo S, et al. ARCS: scaffolding genome drafts with linked reads. *Bioinformatics*. 2018:**34**:725–731. https://doi.org/10.1093/bioinformatics/btx675
22. Warren RL, et al. LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. *GigaScience*. 2015:**4**:1–11.
23. Zhang SV, Zhuo L, Hahn MW. AGOUTI: improving genome assembly and annotation using transcriptome data. *GigaScience*. 2016:**5**:1–12.
24. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014:**30**:2114–2120. https://doi.org/10.1093/bioinformatics/btu170
25. Weisenfeld NI, et al. Direct determination of diploid genome sequences. *Genome Res*. 2017:**27**:757–767. https://doi.org/10.1101/gr.214874.116
26. Jackman SD, et al. Tigmint: Correcting assembly errors using linked reads from large molecules. *BMC Bioinf*. 2018:**19**:1–10.
27. Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ. Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Res*. 2016:**44**:e147–e147. https://doi.org/10.1093/nar/gkw654
28. Xu G-C, et al. LR Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. *GigaScience*. 2018:**8**:1–14.
29. Xu M, et al. TGS-GapCloser: a fast and accurate gap closer for large genomes with low coverage of error-prone long reads. *BMC Genomics*. 2018:**9**:1–11.
30. Grabherr MG, et al. Genome-wide synteny through highly sensitive sequence alignment: satsuma. *Bioinformatics*. 2010:**26**:1145–1151. https://doi.org/10.1093/bioinformatics/btq102
31. Chakraborty A, et al. Genome sequencing and de novo and reference-based genome assemblies of Bos indicus breeds. *Genes Genomics*. 2023:**45**:1399–1408. https://doi.org/10.1007/s13258-023-01401-w
32. Manni M, et al. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021:**38**:4647–4654. https://doi.org/10.1093/molbev/msab199
33. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018:**34**:3094–3100. https://doi.org/10.1093/bioinformatics/bty191
34. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009:**25**:1754–1760. https://doi.org/10.1093/bioinformatics/btp324
35. Kim D, et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019:**37**:907–915.
36. Li H, et al.; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009:**25**:2078–2079. https://doi.org/10.1093/bioinformatics/btp352
37. Jin JJ, Yu WB, Yang JB, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol*. 2020:**21**:1–31.
38. Tillich M, et al. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 2017:**45**:W6–W11. https://doi.org/10.1093/nar/gkx391
39. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res*. 2019:**47**:W59–W64. https://doi.org/10.1093/nar/gkz238
40. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature*. 2011:**475**:493–496.
41. Li H, Barrett J. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011:**27**:2987–2993.
42. Collevatti RG, et al. Contrasting spatial genetic structure in *Annona crassiflora* populations from fragmented and pristine savannas. *Plant Sys Evol*. 2014:**300**:1719–1727. https://doi.org/10.1007/s00606-014-0997-9
43. Strijk JS, et al. Chromosome-level reference genome of the soursop (*Annona muricata*): a new resource for Magnoliid research and tropical pomology. *Mol Ecol Resour*. 2021:**21**:1608–1619. https://doi.org/10.1111/1755-0998.13353
44. Flynn JM, et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA*. 2020:**117**:9451–9457. https://doi.org/10.1073/pnas.1921046117
45. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006:**22**:1658–1659. https://doi.org/10.1093/bioinformatics/btl158
46. Bickmann, L., Rodriguez, M., Jiang, X., and Makalowski, W. 2023, TEclass2: Classification of transposable elements using Transformers. *bioRxiv*, 2023.10.13.562246.
47. Cantarel BL, et al. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res*. 2008:**18**:188–196.
48. Haas BJ, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protocols*. 2013:**8**:1494–1512. https://doi.org/10.1038/nprot.2013.084
49. Liu K, et al. Transcriptome analysis and identification of genes associated with floral transition and flower development in sugar apple (*Annona squamosa* L.). *Front Plant Sci*. 2016:**7**:217236.
50. Stanke M, et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res*. 2006:**34**:W435–W439. https://doi.org/10.1093/nar/gkl200
51. Korf I. Gene finding in novel genomes. *BMC Bioinf*. 2004:**5**:1–9.

52. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res*. 2019:**47**:D155–D162. https://doi.org/10.1093/nar/gky1141
53. Chan PP, Lowe TM. tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods Mol Biol*. 2019:**1962**:1–14. https://doi.org/10.1007/978-1-4939-9173-0_1
54. Finn RD, et al. Pfam: the protein families database. *Nucleic Acids Res*. 2014:**42**:D222–D230. https://doi.org/10.1093/nar/gkt1223
55. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res*. 2000:**28**:45–48. https://doi.org/10.1093/nar/28.1.45
56. Cantalapiedra CP, et al. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol*. 2021:**38**:5825–5829.
57. Moriya Y, et al. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res*. 2007:**35**:W182–W185. https://doi.org/10.1093/nar/gkm321
58. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019:**20**:1–14.
59. Laetsch DR, Blaxter ML. KinFin: Software for taxon-aware analysis of clustered protein sequences. *G3 (Bethesda, Md.)*. 2017:**7**:3349–3357. https://doi.org/10.1534/g3.117.300233
60. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013:**30**:772–780. https://doi.org/10.1093/molbev/mst010
61. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014:**30**:1312–1313. https://doi.org/10.1093/bioinformatics/btu033
62. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007:**24**:1586–1591. https://doi.org/10.1093/molbev/msm088
63. Mendes FK, Vanderpool D, Fulton B, Hahn MW. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics*. 2021:**36**:5516–5518. https://doi.org/10.1093/bioinformatics/btaa1022
64. Sun J, et al. OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Res*. 2023:**51**:W397–W403. https://doi.org/10.1093/nar/gkad313
65. Wang Y, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012:**40**:e49–e49. https://doi.org/10.1093/nar/gkr1293
66. Chen H, Zwaenepoel A, Van De Peer Y. wgd v2: a suite of tools to uncover and date ancient polyploidy and whole-genome duplication. *Bioinformatics*. 2024:**40**:1–11.
67. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003:**31**:3812–3814. https://doi.org/10.1093/nar/gkg509
68. Mahajan S, Bisht MS, Chakraborty A, Sharma VK. Genome of Phyllanthus emblica: the medicinal plant Amla with super antioxidant properties. *Front Plant Sci*. 2023:**14**:1210078. https://doi.org/10.3389/fpls.2023.1210078
69. Bisht MS, Singh M, Chakraborty A, Sharma VK. Genome of the most noxious weed water hyacinth (*Eichhornia crassipes*) provides insights into plant invasiveness and its translational potential. *iScience*. 2024:**27**:110698. https://doi.org/10.1016/j.isci.2024.110698
70. Ping J, Hao J, Wang T, Su Y. Comparative analysis of plastid genomes reveals rearrangements, repetitive sequence features, and phylogeny in the Annonaceae. *Front Plant Sci*. 2024:**15**:1351388. https://doi.org/10.3389/fpls.2024.1351388
71. Barlow J, et al. , The future of hyperdiverse tropical ecosystems. *Nature*. 2018:**559**:517–526. https://doi.org/10.1038/s41586-018-0301-1
72. Doebley JF, Gaut BS, Smith BD. The molecular genetics of crop domestication. *Cell*. 2006:**127**:1309–1321.
73. Zhu Q, et al. Multilocus analysis of nucleotide variation of oryza sativa and its wild relatives: severe bottleneck during domestication of rice. *Mol Biol Evol*. 2007:**24**:875–888. https://doi.org/10.1093/molbev/msm005
74. Wang R, et al. Thermogenesis, flowering and the association with variation in floral odour attractants in magnolia sprengeri (Magnoliaceae). *PLoS One*. 2014:**9**:e99356. https://doi.org/10.1371/journal.pone.0099356
75. Kishore K, et al. Pollination biology of *Annona squamosa* L. (Annonaceae): evidence for pollination syndrome. *Sci Horticult*. 2012:**144**:212–217. https://doi.org/10.1016/j.scienta.2012.07.004
76. Gottsberger G. Pollination and evolution in neotropical Annonaceae. *Plant Species Biol*. 1999:**14**:143–152. https://doi.org/10.1046/j.1442-1984.1999.00018.x
77. Tang G, et al. The *Annona montana* genome reveals the development and flavor formation in mountain soursop fruit. *Ornam Plant Res*. 2023:**3**:0–0. https://doi.org/10.48130/opr-2023-0014
78. Chen J, et al. Liriodendron genome sheds light on angiosperm phylogeny and species–pair differentiation. *Nat Plants*. 2018:**5**:18–25. https://doi.org/10.1038/s41477-018-0323-6
79. Mahajan S, et al. Genome sequencing and functional analysis of a multipurpose medicinal herb *Tinospora cordifolia* (Giloy). *Sci Rep*. 2024:**14**:1–17.
80. Chakraborty A, Mahajan S, Bisht MS, Sharma VK. Genome sequencing of *Syzygium cumini* (jamun) reveals adaptive evolution in secondary metabolism pathways associated with its medicinal properties. *Front Plant Sci*. 2023:**14**:1260414. https://doi.org/10.3389/fpls.2023.1260414
81. Gupta RK, et al. In vivo evaluation of anti-oxidant and anti-lipidimic potential of Annona squamosa aqueous extract in Type 2 diabetic models. *J Ethnopharmacol*. 2008:**118**:21–25. https://doi.org/10.1016/j.jep.2008.03.008
82. Ji J, et al. Plant SWEET family of sugar transporters: structure, evolution and biological functions. *Biomolecules*. 2022:**12**:205. https://doi.org/10.3390/biom12020205
83. Breia R, et al. Plant SWEETs: from sugar transport to plant–pathogen interaction and more unexpected physiological roles. *Plant Physiol*. 2021:**186**:836–852. https://doi.org/10.1093/plphys/kiab127
84. Chen LQ, et al. Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. *Science (New York, N.Y.)*. 2012:**335**:207–211. https://doi.org/10.1126/science.1213351
85. Zamir D. Improving plant breeding with exotic genetic libraries. *Nat Rev Genet*. 2001:**2**:983–989. https://doi.org/10.1038/35103589
86. Talavera A, et al. Genomics in neglected and underutilized fruit crops: a chromosome-scale genome sequence of cherimoya (*Annona cherimola*). *Plants People Planet*. 2023:**5**:408–423.
87. Fang Y, et al. Plant protein-coding gene families: their origin and evolution. *Front Plant Sci*. 2022:**13**:995746. https://doi.org/10.3389/fpls.2022.995746
88. Luna SK, Chain FJJ. Lineage-specific genes and family expansions in dictyostelid genomes display expression bias and evolutionary diversification during development. *Genes (Basel)*. 2021:**12**:1628. https://doi.org/10.3390/genes12101628
89. Brandão APME, Santos DYAC. Nutritional value of the pulp of different sugar apple cultivars (*Annona squamosa* L.). *Nutritional Composition of Fruit Cultivars*. 2016, 195–214.
90. Cardoso CP, et al. Modification of sugar profile and ripening in atemoya (Annona × atemoya Mabb.) fruits through copper hydroxide application. *Plants (Basel, Switzerland)*. 2023:**12**:768. https://doi.org/10.3390/plants12040768
91. Yu J, et al. Starch and sugars as determinants of postharvest shelf life and quality: some new and surprising roles. *Curr Opin Biotechnol*. 2022:**78**:102844. https://doi.org/10.1016/j.copbio.2022.102844
92. Stein O, Granot D. An overview of sucrose synthases in plants. *Front Plant Sci*. 2019:**10**:435701.
93. Ruan YL, et al. Sugar input, metabolism, and signaling mediated by invertase: roles in development, yield potential, and response to drought and heat. *Mol Plant*. 2010:**3**:942–955. https://doi.org/10.1093/mp/ssq044
94. Figueroa CM, Lunn JE. A tale of two sugars: trehalose 6-phosphate and sucrose. *Plant Physiol*. 2016:**172**:7–27. https://doi.org/10.1104/pp.16.00417