JARQ 60 (4), 343-356 (2026) https://www.jircas.go.jp

Genomic and Polymorphic Landscape of the Japanese Peach Cultivar Akatsuki

Hiroaki SAKAI[1], Hiroyuki KAWAHIGASHI[2]*, Hiroyuki KANAMORI[3], Utako YAMANOUCHI[3], Hideaki YAEGAKI[2], Hiromi KAJIYA-KANEGAE[3, 4], Ikuko NAKAJIMA[2], Miho TATSUKI[2], Takayoshi YAMANE[2] and Jun-ichi YONEMARU[4]

[1] Research Center for Advanced Analysis, National Agriculture and Food Research Organization (NARO), Tsukuba, Japan
[2] Institute of Fruit Tree and Tea Science, NARO, Tsukuba, Japan
[3] Institute of Crop Science, NARO, Tsukuba, Japan
[4] Research Center for Agricultural Information Technology, NARO, Tsukuba, Japan

Abstract

We present a high-quality genome assembly of the Japanese peach cultivar Akatsuki, a leading commercial and breeding parent. Using PacBio long-read sequencing and Bionano optical mapping, we generated chromosome-scale sequences totaling 241.9 Mbp, anchored to the reference Lovell genome. The Akatsuki genome assembly demonstrated excellent completeness, with a complete BUSCO score of 99.3%, comparable to the high-quality Lovell assembly. Comparative analysis of 38 Japanese cultivars identified genome-wide polymorphisms, which have been incorporated into the TASUKE+ database to facilitate public access and comparative genomic studies. Transcriptome profiling of twelve tissues revealed distinct gene expression patterns associated with fruit development, ripening, and organ identity, providing insights into the molecular basis of tissue-specific functions. Together, these genomic and transcriptomic resources offer a valuable foundation for understanding agronomically important traits in Japanese peaches and will accelerate marker-assisted breeding and cultivar improvement.

Discipline: Horticulture
Additional key words: DNA polymorphisms, gene expression, peach (*Prunus persica*) genome, RNAseq

Introduction

The peach (*Prunus persica*) is a globally important fruit crop known for its excellent flavor and versatile use. It belongs to the Rosaceae family as a diploid species and is self-compatible, which facilitates the development of homozygous lines. Furthermore, its small haploid genome size (< 300 megabases) makes the peach an ideal model for genomic research on rosaceous fruit trees. Significant progress in peach genomics was achieved through the International Peach Genome Initiative, resulting in the release of the Peach v1.0 genome in 2010 (Verde et al. 2013). The subsequent integration of RNA-Seq data improved gene predictions and annotations, leading to the updated Peach v2.1, which includes 26,873 predicted genes (Verde et al. 2017). Whole-genome sequencing has facilitated the identification of genes associated with 22 agronomically important traits, including phenology, fruit quality, allergens, disease resistance, tree architecture, and self-incompatibility (Aranzana et al. 2019). The Genome Database for Rosaceae (GDR, https://www.rosaceae.org/peach/genome) currently provides a comprehensive repository of genomic information for a wide range of species within the Rosaceae family.

Extensive advances in genome research on the peach have recently been achieved in China, the center of origin of the species and home to a vast population of wild relatives and landraces with high genetic diversity. Evolutionary and domestication studies on the peach have provided important insights, with divergence from

*Corresponding author: kawahigashi.hiroyuki565@naro.go.jp
Received 29 August 2025; accepted 15 December 2025; J-STAGE Advanced Epub 30 July 2026.
https://doi.org/10.6090/jarq.25S16

wild species, such as *P. mira*, estimated at approximately 11.5 million years ago, coinciding with the tectonic uplift of the Qinghai-Tibet Plateau (Li & Wang 2020, Li et al. 2021, Cao et al. 2022). The Peach Population Genome Variation database was recently launched, integrating genomic data from 1,378 peach samples, aligned with high-quality reference genomes, such as Lovell, CN14, and Chinese Cling (An et al. 2024). The advances reported by Li et al. (2025) have shed light on the evolutionary origin of peaches and the gene acquisitions that occurred during their domestication and have also enabled the high-resolution identification of genes associated with a wide range of agronomic traits in a single step. The application of genetic diversity studies and sequence-based markers derived from these resources may enhance knowledge-based breeding and other biotechnological approaches for improving *Prunus* crops.

The peach, originally domesticated in China, was introduced to and used in Japan as early as the Yayoi period (approximately 2,000 years ago), as indicated by archaeobotanical findings (Zheng et al. 2014). In the early Meiji era (circa 1872), improved cultivars from China, including the Shanghai (Chinese Cling) variety, were introduced to Japan and became the foundation of modern Japanese peach breeding due to their disease resistance and other desirable traits (Yoshida 1981). The Akatsuki peach is a renowned Japanese cultivar known for its exceptional qualities. It was developed through a meticulous breeding program initiated at the Fruit Tree Experimental Station, Japan (Kanato et al. 1980). The breeding process began in 1952, and Akatsuki was officially recognized and registered as a named cultivar in 1979, cementing its status as a distinguished cultivar within the peach industry. This cultivar resulted from crossing the Hakutou and Hakuhou cultivars, combining desirable traits to produce superior offspring. Akatsuki is now the most widely grown peach in Japan (Yaegaki 2019), with major production areas in Fukushima, Nagano, and Yamanashi Prefectures.

In Japan, peach SSR markers have proven effective in parentage analyses of Japanese peaches, providing a reliable method to trace genetic origins and relationships. A previous study examined the genetic relationships between Hakutou and other peach cultivars utilizing 10 SSR markers. The findings obtained revealed a close relationship between Hakutou and Shanhaishuimitao, suggesting a parent-offspring link (Yamamoto et al. 2003). Additionally, four cultivars presumed to have originated from chance seedlings shared one allele with their putative parent Hakutou at each SSR locus, supporting the hypothesis that they are offspring rather than bud sport mutants (Yamamoto et al. 2003, Ohashi et al. 2012).

Akatsuki is one of the most important Japanese peach cultivars, highly valued for its excellent sweetness, flavor, and firm flesh. It is not only widely cultivated in Japan and exported overseas but is also used as a key parent in modern breeding programs. For example, the cross Akatsuki × Hakuhou and several other breeding lines have inherited high sugar content and firm flesh traits from Akatsuki. Thus, Akatsuki serves as an indispensable genetic resource both as a commercial cultivar and as a breeding parent. However, genomic information on Japanese peach cultivars remains limited, and the genetic basis of Akatsuki has not yet been fully elucidated. The genome sequence of Akatsuki provides a valuable foundation for identifying genes associated with important agronomic traits and developing DNA markers to accelerate peach breeding. Furthermore, understanding the genetic basis of multiple traits in Akatsuki is essential for improving breeding efficiency, adapting to changing cultivation environments, and meeting expanding export demands under global climate change. In addition, previous genome-wide analyses of single-nucleotide polymorphism (SNP) loci have revealed significant changes in the genetic structure of peaches during domestication, as well as differences among domesticated peaches, their wild relatives, and various peach cultivars (Akagi et al. 2016).

In the present study, we conducted a comprehensive genomic analysis of Akatsuki in Japan. This included whole-genome sequencing to reveal the complete genetic makeup, gene expression profiling to understand functional activity, and a polymorphism analysis of 38 cultivars to assess genetic variations within the species. By elucidating the genomic features of Akatsuki, we aim to deepen our understanding of its unique genetic background. This knowledge will provide valuable insights for developing more effective breeding strategies to improve important agronomic traits, such as disease resistance and fruit quality, as well as consumer-preferred fruit qualities.

Materials and methods

1. Materials

All plants used in this study were cultivated at the Institute of Fruit Tree and Tea Science (NIFTS), NARO, in Tsukuba, Japan. The Japanese peach cultivar Akatsuki, a heterozygous hybrid resulting from a cross between Hakutou and Hakuhou, was used. DNA for genome assembly and polymorphism analyses was extracted from young leaves (approximately 6 cm in length) of the Akatsuki Japanese peach tree and 38 additional peach

cultivars, as summarized in Supplementary Table S1, using the cetyl-trimethyl ammonium bromide method (Murray & Thompson 1980).

In the RNA-seq analysis, Akatsuki peach tissues were collected from trees cultivated at NIFTS, NARO in Tsukuba, Japan. The shoot and root tissues of the Akatsuki cultivar were cultured *in vitro* in culture bottles with culture medium for the rapid multiplication of *Prunus* sp. *in vitro* (Quoirin & Lepoiver 1977), supplemented with 5 μM benzyladenine and 1 μM gibberellin 3 for shoot induction, or 2.5 μM indolebutyric acid for root development, following the protocol described by Noguchi (1990).

2. Genome sequence assembly

The genomic DNA of Akatsuki was sequenced using the PacBio Sequel II platform (Pacific Biosciences, Menlo Park, CA). Bionano optical mapping data were obtained through the subcontracted research services provided by AS ONE Corp (Osaka, Japan). DNA was labeled using the DLS DNA Labeling Kit (Bionano Genomics, San Diego, CA), which utilizes the DLE-1 enzyme, and was then processed on the Saphyr Genome Mapping Instrument according to the manufacturer's protocol. HiFi reads were assembled using HiFiasm ver. 0.16.1-r375 and the primary contigs were scaffolded by Bionano Solve ver. 3.6.1 with default settings. Chromosome sequences were constructed by RagTag ver. 2.1.0 with default settings using the Lovell genome sequence as the reference (Verde et al. 2017). Bionano data are also available from the NCBI website with BioProject PRJDB35480.

3. Transcriptome analysis and genome annotation

In the transcriptome analysis, total RNA was extracted from 12 distinct tissue samples of the peach cultivar Akatsuki collected in 2021, 2022, and 2023. To minimize transcriptome variability, three samples of each tissue collected from different trees on the same day were pooled, except for multiple shoots. In 2021, samples included the young fruit mesocarp collected on 19 May (M1), the premature fruit mesocarp on 24 June (M2), and the ripened fruit mesocarp and exocarp on 9 July (M3 and M4). Mature leaves were also collected on 9 July at approximately noon (M5), and dormant winter buds on 1 December (M6). In 2022, flower buds and whole flowers were collected on 28 March (M11 and M12). In 2023, additional samples were collected, including the ripened fruit mesocarp and exocarp on 7 July (M7 and M8), roots cultured under sterile conditions on 2 August (M9), and multiple shoots, including apical stem regions from cultured plants, on the same day (M10). These samples were not biological replicates but represented different tissues and developmental stages. The 2021 (M3 and M4) and 2023 (M7 and M8) ripened fruit samples were collected at comparable maturity stages, as determined by external fruit color and firmness, to represent similar ripening conditions across years. However, these samples were not intended as biological replicates across years, but rather as independent datasets to confirm reproducibility of tissue-specific expression patterns under typical field conditions. The RNA samples were processed in two sequencing batches corresponding to M1-M6 and M7-M12, respectively, using the same analytical platform and standardized procedures to minimize batch effects. Each tissue was immediately frozen in liquid nitrogen. Total RNA was extracted from 12 tissues of Akatsuki using RNeasy (Qiagen, Hilden, Germany) according to the manufacturer's protocol. RNA quality was assessed using the Bioanalyzer 2100 algorithm (Agilent Technologies, Palo Alto, CA, USA). Only high-quality RNA (RNA integrity number > 8) was selected for RNA-seq. cDNA libraries were synthesized from these RNAs, and paired-end 150-bp reads were obtained using a NovaSeq 6000 sequencer (Illumina, San Diego, CA, USA) by Macrogen Japan Corp.

Protein-coding genes were predicted following a previously reported procedure (Takahashi et al. 2019). We used the RNA-Seq data obtained from six samples for *ab initio*- and transcriptome-based gene predictions. Then we combined the data of protein sequences deposited in UniProt (The UniProt Consortium 2021) and those of four Rosaceae species downloaded from the GDR for protein-based gene predictions: *Prunus persica* (v2.0.a1) (Verde et al. 2017), *Fragaria ananassa* (Camarosa Genome v1.0.a2) (Edger et al. 2019), *Malus domestica* (GDDH13 v1.1) (Daccord et al. 2017), and *Pyrus communis* (Bartlett DH Genome v2.0) (Linsmith et al. 2019). These data were used to create a non-redundant set of sequences using MMseqs2 (Steinegger & Soeding 2017) with the "--min-seq-id 0.99 -c 0.99 --cov-mode 0" option. To assess the completeness of gene annotations for *F. ananassa*, *M. domestica*, *P. communis*, and *P. persica* cv. Akatsuki, we performed a Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis (ver. 5.7.1) (Manni et al. 2021) using the embryophyta_odb10 dataset. Orthologous gene clusters among the five genomes were inferred using OrthoFinder ver. 3.1.0 (Emms & Kelly 2019) with default parameters. Repetitive sequences were annotated by Extensive de-novo TE Annotator ver. 2.2.2. Raw sequence reads from the RNA-seq analysis are available at the DNA Data Bank of Japan (DDBJ) under accession numbers DRR700051 to DRR700062.

4. Polymorphisms and lineage analysis

A short-read analysis of 38 peach cultivars was performed using Illumina's next-generation sequencer, Novaseq 6000. Paired-end sequencing was conducted for all samples (150 bp × 2). Library preparation and sequencing were performed by Novogene (Beijing, China) using Illumina's standard protocol. Data analyses were conducted to meet the criterion of Q30 > 85%, and the output data volume for each sample was confirmed to exceed 20 Gb of clean data. Short-read sequences were further trimmed by Trimmomatic ver. 0.39 with the "ILLUMINACLIP: TruSeq3-PE-2.fa:2:30:10 LEADING:20 TRAILING:20 SLIDINGWINDOW:10:20 MINLEN:30" options. DNA variants in each cultivar were detected using the GATK Germline Pipeline implemented in the NVIDIA Clara Parabricks ver. 4.3.1 with the "--dont-use-soft-clipped-bases" option and merged by the GenomicsDBImport and GenotypeGVCFs tools of GATK ver. 4.4.0. The detected variant sites were filtered by the VariantFiltration tool of GATK with the '-filter "QD < 2.0" -filter "QUAL < 30.0" -filter "SOR > 3.0" -filter "FS > 60.0" -filter "MQ < 40.0" -filter "MQRankSum < -12.5" -filter "ReadPosRankSum < -8.0"' options and the '-filter "QD < 2.0" -filter "QUAL < 30.0" -filter "FS > 200.0" -filter "ReadPosRankSum < -20.0"' options for SNPs and short indels, respectively. Each variant site was annotated by SnpEff ver. 4.3t with the custom-made database created from the protein-coding gene structures predicted on the Akatsuki reference genome sequence. A phylogenetic tree of the 38 cultivars was generated by RapidNJ ver. 2.3.3. Genomic polymorphism information for the Akatsuki cultivar was deposited into the TASUKE database (Kumagai et al. 2013, 2019). Short-read next-generation sequencing (NGS) data are available from the DDBJ under accession numbers DRR700012 to DRR700050.

Results and discussion

1. Sequencing and assembly

The *de novo* assembly of 1,866,942 HiFi reads of Akatsuki produced 583 contigs, with a total length of 257.13 Mbp and an N50 length of 12.10 Mbp. Of the 583 contigs, 36 were assembled into 14 scaffolds using Bionano optical mapping data, spanning 240.87 Mbp. By aligning the scaffolds and remaining contigs to the Lovell reference genome sequence, 13 scaffolds and 37 contigs spanning 241.90 Mbp were anchored to the eight chromosomes. In addition, the plastid genome sequence was reconstructed from the three contigs by mapping the unanchored contigs to the publicly available chloroplast genome sequence. The mitochondrial genome was not fully connected and remained fragmented, and no master circle has been obtained (scaffold001, contig130, contig312). The contigs that were not anchored to the chromosomes tended to show a significantly higher proportion of repetitive sequences than the eight chromosomes ($p < 0.01$). The fraction of repetitive elements in the Akatsuki reference genome sequence was 42.54% (109.31 Mbp), which was higher than that in the reference genome of another peach cultivar, Lovell (36.10%, 81.10 Mbp). The most abundant class of repetitive elements was the Gypsy-type long terminal repeat (LTR)-retrotransposon (8.23%), followed by the Copia-type LTR-retrotransposon (4.31%).

A diagonal line extending from the top left to the bottom right was observed in the dot plot analysis comparing Akatsuki (vertical axis) and Lovell (horizontal axis). This indicates that synteny is largely conserved in the regions of contigs and scaffolds anchored to chromosomes (Fig. 1). However, notable structural differences were observed in the centromere region of chromosome 3, where several inversions were identified. These structural variations in this specific region suggest potential divergence between the two cultivars, which may reflect distinct evolutionary paths or functional adaptations in this chromosomal region. However, the central region of chromosomes is recognized to generally have a low gene density, abundant repetitive sequences, and suppressed recombination. In the peach, quantitative trait loci (QTLs) and genes associated with traits such as semi-dwarf growth, weeping habit, skin color, self-incompatibility, and male sterility (Aranzana et al. 2019, Tan et al. 2021) have been identified, but they are located on chromosomal arms rather than in the central region.

We predicted 25,681 protein-encoding genes and 31,523 transcripts within the Akatsuki reference genome sequence by combining *ab initio*–, transcriptome–, and protein–based gene prediction approaches (Supplementary Table S2-S4). The results of the BUSCO analysis of various plant genomes, including Akatsuki, Peach Lovell, *P. communis* (BartlettDHv2.0), *M. domestica* (GDDH13_1-1), and *F. ananassa* (v1.0.a2), are shown in Table 1. The Akatsuki and Lovell genomes had similar numbers of loci (25,681 vs. 26,873, respectively), with Lovell showing a slightly higher count. Akatsuki had 31,523 predicted isoforms, compared with 47,089 in Lovell. The genome assembly of Akatsuki demonstrated high completeness, with a complete BUSCO score of 99.3%, comparable to that of the reference Lovell genome (Table 1).

A total of 27,371 orthogroups were identified among the five Rosaceae genomes using OrthoFinder. Among

them, 14,522 orthogroups were shared across all species, representing the conserved core gene set (Fig. 2). Additionally, a number of species-specific orthogroups were detected in each species, suggesting lineage-specific gene gain or retention events. Interestingly, 1,372 orthogroups were shared exclusively between two peach cultivars, but were absent from the other three species. A Gene Ontology (GO) enrichment analysis was conducted on 1,770 Akatsuki genes uniquely shared with Lovell. Several GO terms were significantly overrepresented (Benjamini-Hochberg adjusted $q < 0.01$), including biological processes such as proteolysis (GO:0006508), cell surface receptor signaling pathway (GO:0007166), and response to auxin (GO:0009733), as well as molecular functions including cysteine-type peptidase activity (GO:0008234) and ADP binding (GO:0043531). These enriched terms suggest roles in hormonal responses and protein turnover that may be associated with peach-specific physiological traits, such as fruit softening, sugar accumulation, and hormonal regulation during ripening. In addition, no significantly enriched GO categories were detected for the Akatsuki-specific orthologs. However, the genes uniquely present in Akatsuki are listed in Supplemental Table S5.

2. Polymorphisms of peach cultivars

Short-read sequencing was performed on 38 cultivars, including Japanese, Chinese, and Western cultivars, as well as rootstock and ornamental types. Through this sequencing, we aimed to identify genetic polymorphisms across the peach genome. This analysis enabled us to assess the genetic diversity of each cultivar and map the distribution of polymorphisms among cultivars, providing foundational information for evaluating peach genetic resources and understanding traits relevant to cultivation. An analysis of the WGS data from 38 cultivars identified 4,758,036 biallelic polymorphic sites. On average, each cultivar exhibited 212,602 homozygous SNPs and 330,931 heterozygous SNPs. After filtering the polymorphic sites, heterozygosity and the inbreeding coefficient (F value) were calculated. Heterozygosity ranged between 0.07 and

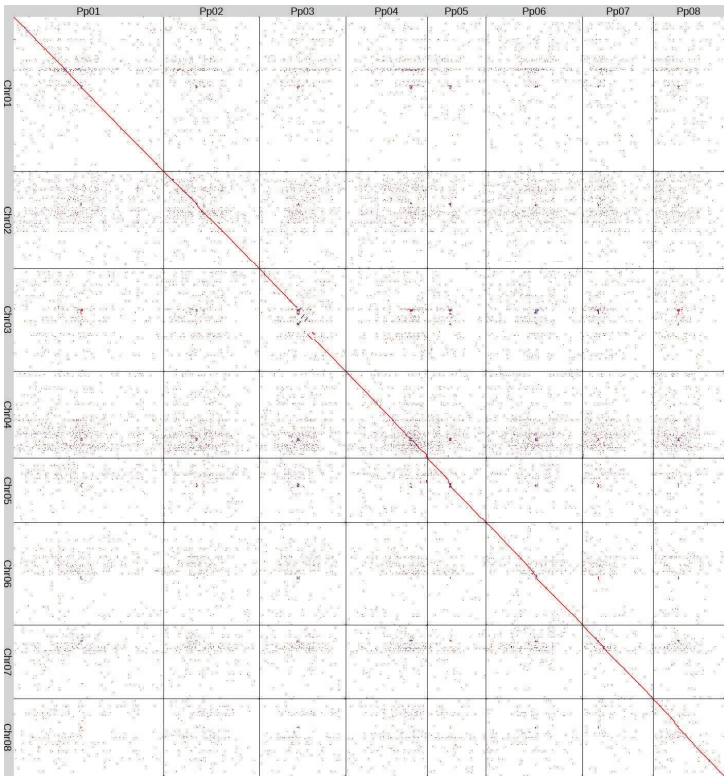


Fig. 1. Dot plot analysis of genomes of Akatsuki (vertical axis) and Lovell (horizontal axis)
Each dot represents a homologous genomic region shared between the two cultivars. The diagonal line from the top left to the bottom right indicates a high degree of conserved synteny between the Akatsuki and Lovell genomes. The centromeric region of chromosome 3 appears to exhibit structural differences.

Table 1. Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis of various fruit plant genomes

	Akatsuki	*Prunus persica* (Lovell v2.0.a1)	*Pyrus communis* (BartlettDHv2.0)	*Malus domestica* (GDDH13_1-1)	*Fragaria ananassa* (Camarosa v1.0.a2)
Number of loci	25,681	26,873	37,445	45,116	108,447
(Isoform numbers)	31,523	47,089	37,445	45,116	127,746
Complete	99.30%	99.30%	81.20%	96.60%	97.80%
Complete and single-copy	80.30%	55.70%	64.50%	64.90%	7.60%
Complete and duplicated	19.00%	43.60%	16.70%	31.70%	90.20%
Fragmented	0.20%	0.20%	9.30%	1.50%	1.20%
Missing	0.50%	0.50%	9.50%	1.90%	1.00%

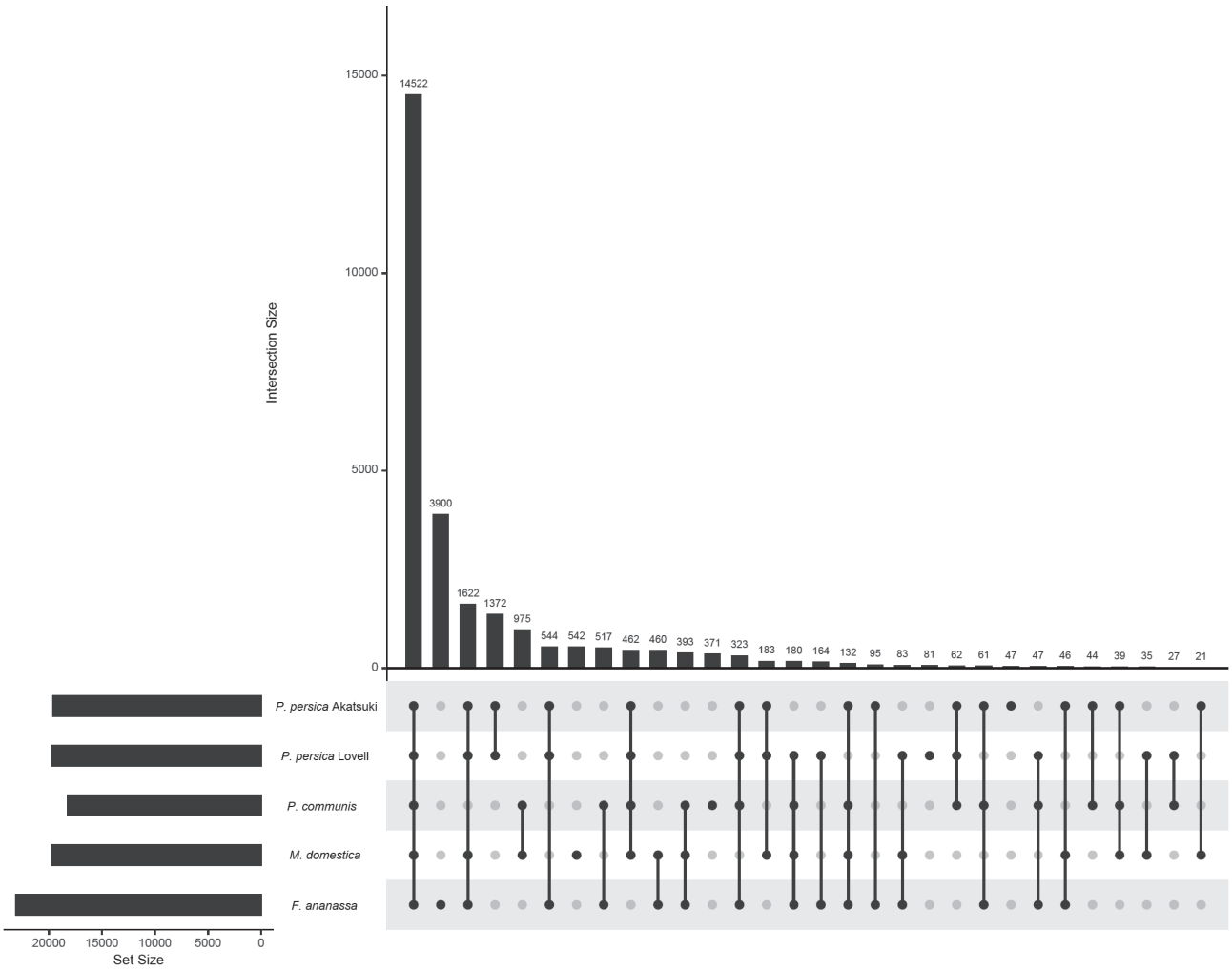


Fig. 2. UpSet plot showing the distribution of orthogroups among five Rosaceae genomes

The plot summarizes the presence and absence patterns of orthologous gene groups identified by OrthoFinder across five Rosaceae genomes: *F. ananassa*, *M. domestica*, *P. communis*, *P. persica* Lovell, and *P. persica* Akatsuki. Each bar on the upper histogram indicates the number of orthogroups shared by the corresponding combination of species (highlighted by filled dots below). The matrix below the bars displays which species are present (filled circles) or absent (blank) in each intersection.

0.36, and the F value between −0.14 and 0.76. The cultivar with the lowest heterozygosity was Houkimomo, while Coral had the highest.

A phylogenetic tree was constructed based on polymorphisms detected in the 38 cultivars (Fig. 3). Japanese cultivars, including Akatsuki, formed a distinct clade that also contained Hakutou and Hakuhou, two foundational progenitors from which many other cultivars have descended. This clustering pattern supports the validity of the polymorphism analysis using short-read sequencing, as it accurately reflects known genetic relationships among these cultivars. In contrast, Western cultivars, such as Lovell, formed a separate clade, reinforcing the tree's phylogenetic accuracy. Additionally, rootstock and ornamental cultivars clustered into distinct clades that exhibited greater genetic diversity and were phylogenetically distant from the cultivated fruit cultivars. This differentiation underscores the genetic uniqueness of rootstock and ornamental groups relative to the primary cultivated cultivars.

Cultivars originating from China exhibited a greater genetic diversity and were broadly distributed across the phylogenetic tree, indicating a wide genetic base. Notably, Shanghaishuimitao, a cultivar instrumental in the modern breeding of Japanese peaches, was positioned in a clade closely related to Japanese cultivars, reflecting its genetic contribution to these cultivars. In contrast, Akame, the origin of which remains unclear, clustered within the clade of Western-derived cultivars, suggesting its introduction from Western germplasm. This pattern shows both the genetic diversity present in Chinese peach cultivars and historical gene flow among regions, which has contributed to the development of modern cultivars.

DNA polymorphism data were visualized using the multi-genome browser TASUKE+ (Fig. 4). TASUKE+ is a web-based genome browser designed to visualize large-scale resequencing data generated by NGS technologies. It enables the simultaneous display of variations, read depths, and annotations across multiple genomes at various scales. Variant frequencies and the depth of coverage are depicted using gradations of color within blocks that range in size from 1 bp to 100 kb. Up to 200 blocks can be displayed in a single window, allowing visualization of up to 20 Mb of genomic data.

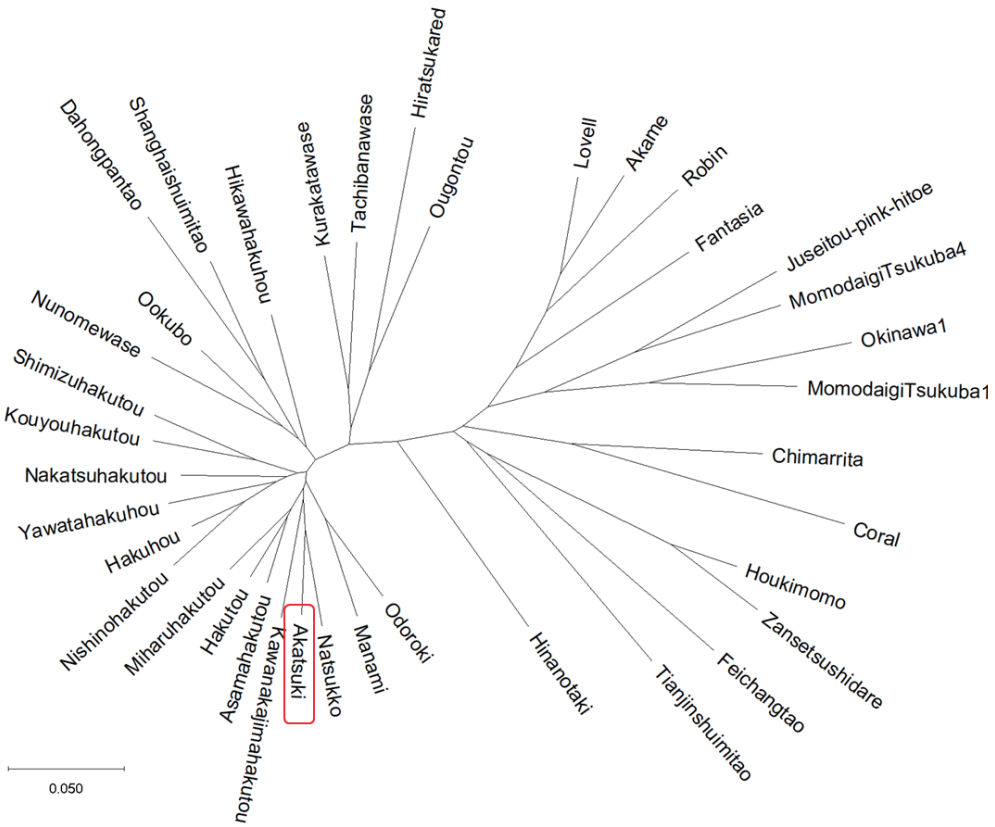


Fig. 3. Phylogenetic tree constructed based on polymorphisms detected in 38 peach cultivars
Akatsuki is highlighted in red. Branch lengths represent genetic distances. The clustering pattern highlights genetic relatedness among cultivars, providing insights into the breeding history and lineage of modern peach cultivars.

Variant data from 38 peach cultivars, including Akatsuki, are displayed with reference to the Akatsuki genome sequence (https://tasuke02.dna.naro.go.jp/tasuke/naropeach/).

Using TASUKE+, we evaluated polymorphisms in the peach tonoplast sugar transporter 1 (PpTST1) gene (Prupe.aktk.05G005900.01 / Prupe_5G006300), which has been reported to play a role in sugar accumulation in peach fruit (Fig. 4). PpTST1 is located within a QTL region associated with sucrose content and explains approximately 29% of the phenotypic variation in this trait (Dirlewanger et al. 1999). This gene also overlaps with the D locus, a major QTL associated with fruit acidity (Boudehri et al. 2009). Previous studies showed that cultivars carrying the G/T or T/T genotypes in the third exon exhibited significantly higher sucrose and total sugar contents than those with the G/G genotype (Peng et al. 2020). Although sugar content data were not available for the cultivars analyzed in this study, the observed distribution of PpTST1 genotypes was generally consistent with previously reported trends. Cultivars with the G/G genotype included older cultivars, such as Shanghaishuimitao, Tianjinshuimitao, and Tachibanawase, as well as rootstock, including Momodaigi Tsukuba 1 and 4 and ornamental types, such as Zansetsushidare, which are generally characterized by lower sweetness or higher acidity. In contrast, many modern table cultivars, including Akatsuki, carried the G/T or T/T genotypes, which have been associated with higher sugar accumulation in previous reports. Therefore, while the present analysis supports a qualitative association between PpTST1 polymorphisms and fruit sweetness, further validation using quantitative sugar content data will be required to confirm this association.

3. Gene expression in various tissues

A gene expression analysis was performed using RNA extracted from 12 representative tissues of the Akatsuki peach cultivar. The sampled tissues included young fruits, pre-mature fruits, ripened fruits, ripened fruit skin, leaves, winter buds, ripened fruits (collected in 2021 and 2023), ripened fruit skin (collected in 2021 and 2023), roots, multiple shoots, flower buds, and flowers (Supplementary Table S7). Each sample represented a pool of tissues collected from three independent trees on the same day, except for multiple shoots. Because these samples were designed to provide a broad overview of tissue-specific gene expression rather than to support biological replication, the analysis focused on descriptive and exploratory comparisons rather than on differential

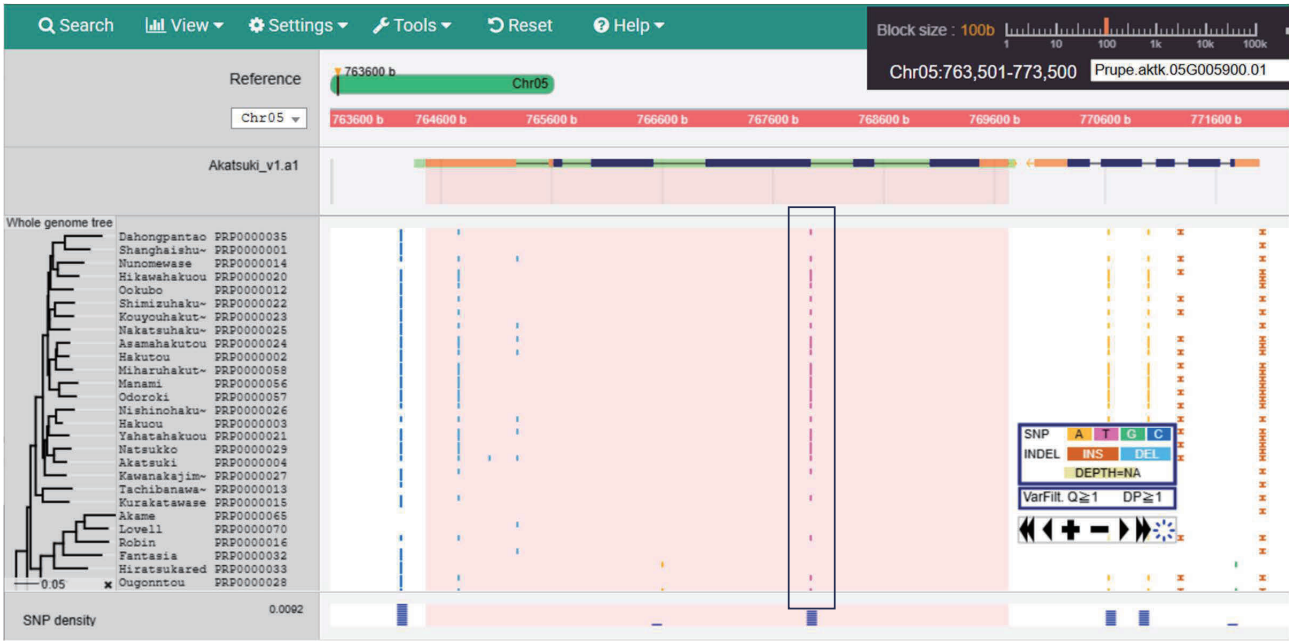


Fig. 4. Screenshot of TASUKE+, a web-based genome browser designed for visualizing large-scale resequencing data generated by next-generation sequencing technologies

TASUKE+ facilitates comparative genomic analyses by integrating large datasets and providing an interactive platform for examining sequence diversity and genomic features. The gene structure and polymorphism of the tonoplast sugar transporter 1 (PpTST1) gene (Prupe.aktk.05G005900.01) are shown. A G/T polymorphism is located at the end of the third exon. The reference allele is G, and the pink line within the black box indicates a substitution to T.

expression testing.

A principal component analysis (PCA) of TPM-normalized data revealed clear clustering patterns corresponding to the tissue type (Fig. 5A). The first and second principal components accounted for 25.4 and 18.2% of the total variance, respectively. Fruit-related tissues (M2, M3, M4, M7, and M8), representing various developmental stages and sampling years, were closely clustered, indicating similar transcriptional profiles across fruit tissues. This result was further supported by hierarchical clustering shown in the heatmap (Fig. 5B). In contrast, M1, corresponding to immature green fruit, was plotted slightly apart from the cluster of more developed fruit tissues, suggesting transcriptional differences at earlier stages of fruit development. Floral organs, including flower buds (M11) and flowers (M12), also formed a separate cluster, indicating closely related gene expression patterns within reproductive tissues. Vegetative tissues, such as leaves (M5) and shoots (M10), clustered together, reflecting shared transcriptional features associated with photosynthesis and vegetative growth.

The sequencing was performed in two separate batches (M1-M6 and M7-M12), corresponding to different sampling years and sequencing runs. However, PCA indicated that tissue identity, rather than batch, was the dominant factor explaining variation in gene expression. This suggests that biological differences among tissues were much greater than technical or seasonal variation, supporting the comparability of these datasets.

Gene Ontology (GO) enrichment analysis further characterized tissue-specific transcriptional trends (Supplementary Fig. S1). In young fruit (SupFig. S1A), genes related to cell wall organization, including cellulose biosynthesis and degradation, were highly expressed within the Biological Process category, suggesting active cell wall remodeling during early fruit growth. In premature fruit (SupFig. S1B), a number of auxin-responsive genes were detected, indicating that auxin signaling was active at this stage during fruit expansion and maturation. In ripened fruit (SupFig. S1C), genes associated with proteolysis showed elevated expression levels. This implies that during the ripening and softening process, proteins responsible for maintaining the cell wall and cellular structure are degraded, which may be related to changes in fruit texture, such as increased softness.

These findings collectively demonstrate that the RNA-seq data successfully captured biologically meaningful tissue-specific expression patterns in peach. Although biological replicates were not included, the clear clustering of related tissues and consistent GO enrichment results suggest that the dataset is robust for

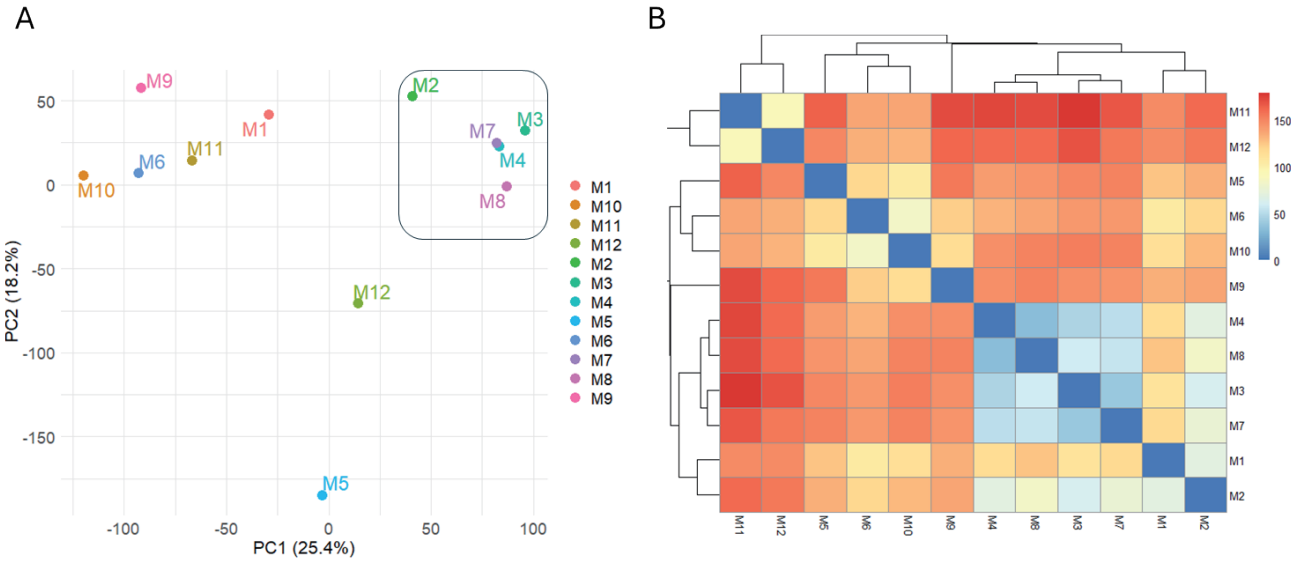


Fig. 5. Clustering of gene expression profiles across 12 tissues in the Akatsuki peach

(A) A principal component analysis (PCA) showing sample distribution along the first two principal components (25.4 and 18.2% variance, respectively). Black circles indicate fruit and fruit skin samples. Samples cluster according to the tissue type, reflecting distinct expression patterns.

(B) Heatmap of pairwise sample correlations based on gene expression. Colors indicate correlation strengths (blue: high, red: low). Hierarchical clustering reveals transcriptional similarity among samples.

descriptive and exploratory analyses. This comprehensive expression atlas provides a useful resource for identifying tissue-preferential genes and for exploring molecular mechanisms underlying organ development, ripening, and physiological differentiation in *Prunus persica*.

Additionally, expression data for specific genes, including those related to fruit quality and allergenicity, were obtained in this dataset. Although these data were not analyzed in detail in the present study, they can serve as a valuable resource for future research focusing on gene-level functional analyses in peaches.

Conclusion

We conducted comprehensive genome, transcriptome, and polymorphism analyses of Akatsuki, a representative Japanese peach cultivar. The reference genome sequence of Akatsuki serves as a highly valuable foundation for advancing genome-assisted breeding in Japan. This sequence provides high-resolution genetic information on key agronomic traits, including fruit quality, yield, and regional adaptability, and is expected to significantly enhance breeding efficiency. As modern agriculture faces increasing demands to address climate change and diverse cultivation environments, the Akatsuki reference genome serves as a critical resource for guiding scientifically informed cultivar improvement strategies.

Data Availability

Genome sequence and annotation data have been deposited in the DDBJ under the BioProject accession number PRJDB35480, and transcriptome data are available under the accession number PRJDB35498. Gene Annotation and Correspondence Table

Gene annotation of the Akatsuki genome was conducted, and the results obtained, including InterPro domain assignments and GO terms, are summarized in Supplementary Table S2. In addition, gene correspondence between Akatsuki and Lovell is presented in Supplementary Tables S3 and S4.

Acknowledgements

This research was partially supported by the Advanced Analysis Center Research Supporting Program of the National Agriculture and Food Research Organization (NARO). The authors acknowledge the support for NGS analyses from the Genome Breeding Support Office of the Institute of Crop Science, NARO. We would like to express our gratitude to the staff in the Operations Department and the part-time employees for their work in the field. Part of this research was conducted using the NARO AI Research Supercomputer “Shiho.”

References

Akagi, T. et al. (2016) Genome-wide view of genetic diversity reveals paths of selection and cultivar differentiation in peach domestication. *DNA Res.*, **23**, 271-282.

An, Y. et al. (2024) PPGV: a comprehensive database of peach population genome variation. *BMC Plant Biol.*, **24**, 701.

Aranzana, M. J. et al. (2019) Prunus genetics and applications after de novo genome sequencing: achievements and prospects. *Hort. Res.*, **6**, 58.

Boudehri, K. et al. (2009) Phenotypic and fine genetic characterization of the D locus controlling fruit acidity in peach. *BMC Plant Biol*., **19,** 59.

Cao, K. et al. (2022) Chromosome-level genome assemblies of four wild peach species provide insights into genome evolution and genetic basis of stress resistance. *BMC Biol.*, **20**, 139.

Daccord, N. et al. (2017) High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nature Genet.*, **49**, 1099-1106.

Dirlewanger, E. et al. (1999) Mapping QTLs controlling fruit quality in peach (*Prunus persica* (L.) Batsch) *Theor. Appl. Genet*., **98**, 18-31.

Edger, P. P. et al. (2019) Origin and evolution of the octoploid strawberry genome. *Nature Genet*., **51**, 541-547.

Emms, D. M. & Kelly, S. (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.*, **20**, 238.

Kanato, K. et al. (1980) New peach cultivar Akatsuki. *Kaju shikenjyou houkoku* (*Bul. Fruit Tree Res. Stn. Ser. A*), **7**, 1-6 [In Japanese with English summary].

Kumagai, M. et al. (2013) TASUKE: a web-based visualization program for large-scale resequencing data. *Bioinfo.*, **29**, 1806-1808.

Kumagai, M. et al. (2019) TASUKE+: a web-based platform for exploring GWAS results and large-scale resequencing data. *DNA Res.*, **26**, 445-452.

Li, Y. & Wang, L. (2020) Genetic resources, breeding programs in China, and gene mining of peach: A review. *Hort. Plant J.*, **6**, 205-215.

Li, Y. et al. (2021) Genomic analyses provide insights into peach local adaptation and responses to climate change. *Genome Res.*, **31**, 592-606.

Li, Y. et al. (2025) Panvariome and pangenome of 1,020 global peach accessions shed light on evolution patterns, hidden natural variations, and efficient gene discovery. *Mol. Plant*, **18**, 995-1013.

Linsmith, G. et al. (2019) Pseudo-chromosome–length genome assembly of a double haploid “Bartlett” pear (*Pyrus communis* L.). *GigaSci*., **8**, giz138.

Manni, M. et al. (2021) BUSCO Update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.*, **38**, 4647-4654.

Murray, M. G. & Thompson, W. F. (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.*, **8**, 4321-4326.

Noguchi, K. (1990) Kajunaeseisan to baiotekunoroji (*Fruit Tree Prod. Biotech.*). In I. Kozaki & Y. Noma (eds.) pp. 59-78 [In Japanese].

Ohashi, Y. et al. (2012) Discrimination of peach varieties using SSR markers. *Fukushimaken nougyou sougousenta kenkyu houkoku* (*Fukushima Pref. Agric. Res. Cent. Res. Rep.*), **4**, 29-38 [In Japanese].

Peng, Q. et al. (2020) Functional analysis reveals the regulatory role of *PpTST1* encoding tonoplast sugar transporter in sugar accumulation of peach fruit. *Int. J. Mol. Sci.*, **21**, 1112.

Quoirin, M. & Lepoivre, P. (1977) Improved media for in vitro culture of *Prunus* sp. *Acta Hort.*, **78**, 437-442.

Steinegger, M. & Söding, J. (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotech.*, **35**, 1026-1028.

Takahashi, Y. et al. (2019) Domesticating *Vigna stipulacea*: A potential legume crop with broad resistance to biotic stresses. *Front. Plant Sci.*, **10**, 1607.

Tan, Q. et al. (2021) Chromosome-level genome assemblies of five *Prunus* species and genome-wide association studies for key agronomic traits in peach. *Hort. Res.*, **8**, 213.

Thermo Fisher Scientific (2021) *f95 Peach*, Allergen Encyclopedia. https://www.thermofisher.com/phadia/wo/en/resources/allergen-encyclopedia/f95.html. Accessed on 4 February 2025.

Verde, I. et al. (2013) The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication, and genome evolution. *Nature Genet.*, **45**, 487-494.

Verde, I. et al. (2017) The Peach v2.0 release: high-resolution linkage mapping and deep resequencing improve chromosome-scale assembly and contiguity. *BMC Genomics*, **18**, 225.

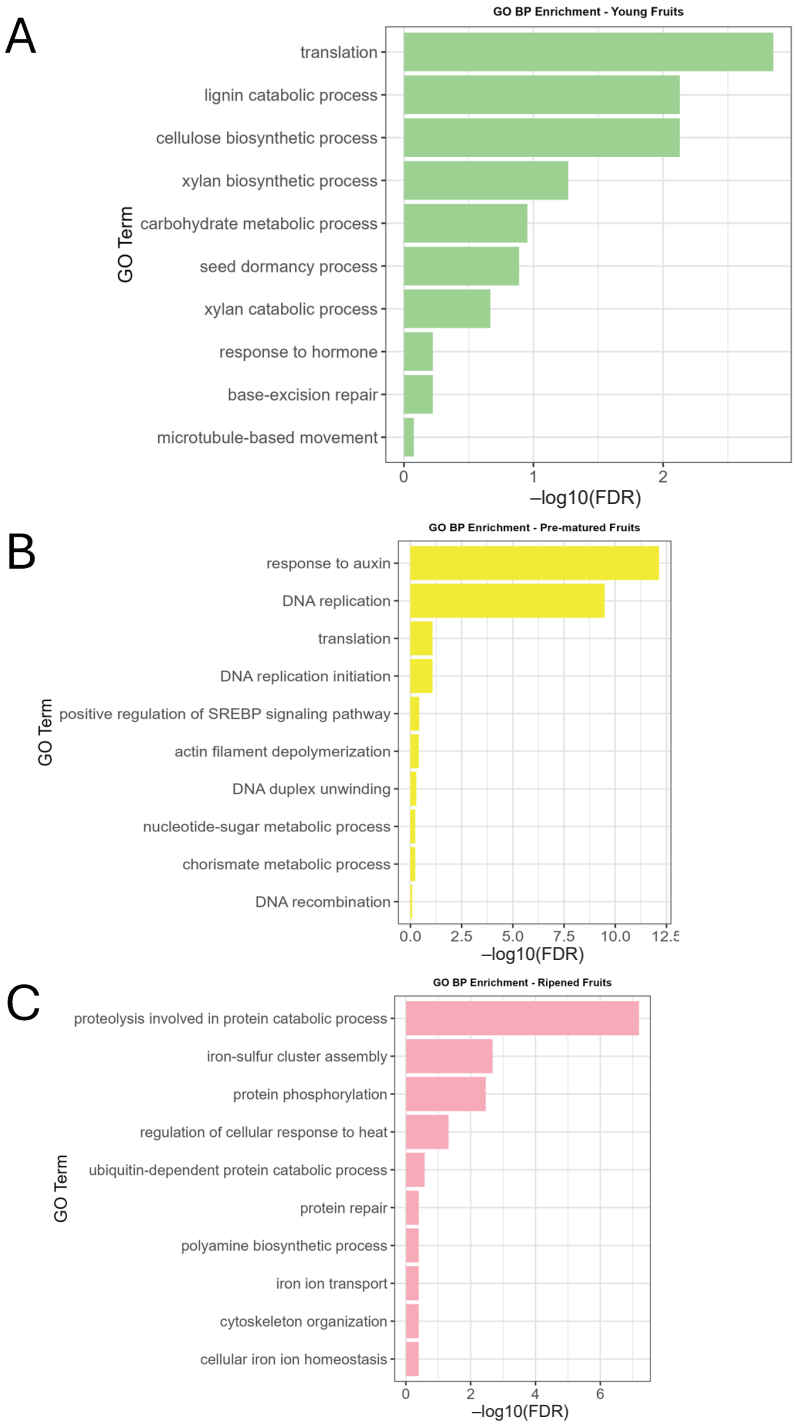
Yamamoto, T. et al. (2003) Parentage analysis in Japanese peaches using SSR markers. *Breed. Sci.*, **53**, 35-40.

Yamamoto, T. et al. (2003) Shanhai Suimitsuto, one of the origins of Japanese peach cultivars. *J. Japan. Soc. Hort. Sci.*, **72**, 116-121.

Yaegaki, H. (2019) Trends in peach breed development at the National Agriculture and Food Research Organization (NARO). *Chuo kajitsu kyokai nyusu reta* (*Cent. Fruit Assoc. Newsletter*), **52**, 1-3 [In Japanese].

Yoshida, M. (1981) Recent trend of peach breeding in Japan. *Jap. Agric. Res. Q.*, **15**, 106-109.

Zheng, Y. et al. (2014) Archaeological evidence for peach (*Prunus persica*) cultivation and domestication in China. *PLOS ONE*, **9**: e106595.



Supplementary Fig. S1. Top 10 significantly enriched Gene Ontology (GO) biological process (BP) terms for each peach fruit tissue analyzed

(A) Young fruits, (B) premature fruits, (C) ripened fruits. In each tissue, transcripts per million (TPM) expression values were transformed into Z-scores across all genes, and the top 1,000 genes with the highest Z-scores were selected for a GO enrichment analysis. Bar plots show the negative log-transformed FDR values ($-\log_{10}$(FDR)) for the most enriched BP terms. Bar colors correspond to individual tissues. The GO enrichment analysis was performed using R.

Supplementary Table S1. List of Peach Varieties Used for NGS

Variety name	NGS_ID	PF_ID	Usage* and Origin	JP number in NARO Genbank
Shanghaishuimitao	NGSp021	PRP0000001	China	112630
Hakutou	NGSp022	PRP0000002	Japan	112534
Hakuhou	NGSp023	PRP0000003	Japan	112532
Akatsuki	NGSp024	PRP0000004	Japan	112519
Tianjinshuimitao	NGSp030	PRP0000010	China	112634
Feichangtao	NGSp069	PRP0000011	China	112629
Ookubo	NGSp031	PRP0000012	Japan	112567
Tachibanawase	NGSp032	PRP0000013	Japan	112578
Nunomewase	NGSp033	PRP0000014	Japan	112559
Kurakatawase	NGSp070	PRP0000015	Japan	112550
Robin	NGSp034	PRP0000016	U.S.A.	112670
Hikawahakuhou	NGSp038	PRP0000020	Japan	225999
Yawatahakuhou	NGSp039	PRP0000021	Japan	225998
Shimizuhakutou	NGSp040	PRP0000022	Japan	112574
Kouyouhakutou	NGSp041	PRP0000023	Japan	112548
Asamahakutou	NGSp042	PRP0000024	Japan	226001
Nakatsuhakutou	NGSp043	PRP0000025	Japan	112557
Nishinohakutou	NGSp071	PRP0000026	Japan	152009
Kawanakajimahakutou	NGSp044	PRP0000027	Japan	-
Ougontou	NGSp045	PRP0000028	Japan	-
Natsukko	NGSp046	PRP0000029	Japan	230006
Fantasia	NGSp072	PRP0000032	U.S.A.	112867
Hiratsukared	NGSp073	PRP0000033	Japan	112836
Dahongpantao	NGSp074	PRP0000035	China	112695
Coral	NGSp055	PRP0000045	Brasil	118372
Chimarrita	NGSp056	PRP0000046	Brasil	236168
Manami	NGSp066	PRP0000056	Japan	236144
Odoroki	NGSp067	PRP0000057	Japan	226000
Miharuhakutou	NGSp068	PRP0000058	Japan	236140
Hinanotaki	NGSp079	PRP0000059	Ornamental, Japan	239299
Zansetsushidare	NGSp080	PRP0000060	Ornamental, Japan	112830
Houkimomo	NGSp081	PRP0000061	Ornamental, Japan	112816
Momodaigi Tsukuba1	NGSp082	PRP0000062	Rootstock, Japan	112794
Momodaigi Tsukuba4	NGSp083	PRP0000063	Rootstock, Japan	112797
Okinawa1	NGSp084	PRP0000064	Landrace, Japan	112785
Akame	NGSp085	PRP0000065	Rootstock, Unknown	112806
Juseitou-pink-hitoe	NGSp086	PRP0000066	Ornamental, Japan	112817
Lovell	NGSp087	PRP0000070	U.S.A.	112655

*Unless otherwise specified, cultivars are classified as table peach varieties.

The following tables are large and provided as large files. Therefore, only their titles are shown. Please view these files on the website.

Supplementary Table S2. Gene annotations for the Akatsuki genome, including InterPro domains and Gene Ontology terms

Supplementary Table S3. Gene Correspondence from Akatsuki to Lovell

Supplementary Table S4. Gene Correspondence from Lovell to Akatsuki

Supplementary Table S5. List of genes uniquely present in the Japanese peach cultivar Akatsuki

Supplementary Table S6. Heterozygosity and inbreeding coefficient (F value) of the 38 Japanese peach varieties analyzed in this study

Supplementary Table S7. Gene expression of 12 samples in the Akatsuki peach