Full-length transcriptome sequencing of the identified fruit shape-related genes of the Olecranon honey peach

Jianliang Liu
Zhongkai University of Agriculture and Engineering
Yao Bao
Zhongkai University of Agriculture and Engineering
Qin Wang
Zhongkai University of Agriculture and Engineering
Huifan Liu (lm_zkng@163.com)
Zhongkai University of Agriculture and Engineering https://orcid.org/0000-0002-0306-4523

Research article

Keywords: Olecranon honey peach, Oxford Nanopore Technologies, gene, fruit shape

Posted Date: April 1st, 2020

DOI: https://doi.org/10.21203/rs.3.rs-17108/v2

Abstract

Background: The Olecranon honey peach shaped like an eagle's beak. In this study, full-length transcriptome sequencing of the Olecranon honey peach and its bud mutant variant was performed by Oxford Nanopore Technologies to comparatively analyze the differentially expressed genes and the transcriptome to identify important fruit shape-related genes.

Results: Full-length transcriptome sequencing was performed to analyze the two peaches. A consensus isoform was obtained and compared to the reference genome for an elimination of redundancy analysis. As a result, a final set of 58 596 transcript sequences was obtained. A total of 21 745 simple sequence repeats were obtained, and 18 322 alternative splicing (AS) events were identified. The comparative analysis of the non-redundant transcript revealed 2530 new gene loci and 37 364 novel transcripts. A total of 457 genes were differentially expressed in the two groups, including 169 up-regulated genes and 288 down-regulated genes. A total of 1519 transcripts were differentially expressed, of which 552 were up-regulated and 997 were down-regulated.

Conclusions: In the case of the plant hormone signal transduction pathway identified by the KEGG annotation, our analyses revealed that differential expression of genes 9229 , 26004 , 22504 , 2822 , 2826 , 2824 , ONT.1953 , ONT.1950 , and ONT.1953 was related to the shape of the peach and may regulate the production of large fruit, via endogenous hormones, secondary metabolites, and signal transduction. This study provides useful information on the shape-related genes and transcripts in the Olecranon honey peach.

Background

The peach is a genetic and genomic reference species of the genus *Prunus* because of its small genome and relatively short juvenile period [1]. The cultivated peach is a diploid (2n=2x=16) with a relatively small genome (265 Mb) (*www.rosaceae.org*). The Olecranon honey peach was named after its unique surface features, as its tip resembles the beak of an eagle. It is considered as one of the best peach varieties and is very popular in China because it is sweet and juicy, and has a crisp texture [2]. The shape of the peach has been evaluated in a comprehensive genome-wide association study using 129 different types of peaches. The results suggest that the *PpCAD1* gene (*ppa003772m*, designated as *PpCAD1*) plays an important role in the peach fruit shape [3]. Breeding has predominantly focused on fruit size for both domestication and improvement of peach farming [4]. Oxford Nanopore Technologies (ONT) sequencing is a next-generation technology for nanopore sequencing [5]. DNA/RNA strands bind to the biological membrane-embedded nanopore protein under the control of motor proteins, and then the strands unwind. Based on the voltage difference across the biological membrane, the DNA/RNA strand passes through the nanopore protein at a specific rate. Because of differences in the chemical properties of different bases on the DNA/RNA chain, when a single base or DNA molecule passes through the nanopore channel, changes occur in the electrical signals [6]. By detecting and responding to these signals, the corresponding base can be deciphered, enabling real-time sequencing [7]. Nanopore-based interpretation

uses the current signal, rendering the decision process complex. The nanopore interprets bases using the sophisticated "Recurrent Neural Network" algorithm according to the magnitude of the current [8].

Bud mutation is a reportedly important tool to establish new cultivars, particularly for vegetative propagation of fruits, and has been recognized as a critical resource in functional genomics studies of model plant species [9, 10]. In Lianping, in northern Guangdong Province, China, bud mutation was found on a tree of the Olecranon honey peach. Lianping is a city with a subtropical monsoon climate, located at the longitude 114°14'–114°56' and the latitude 24°06'–24°36'. The annual average air temperature is 18.5–20.7 ℃, the precipitation is 1700 mm–1800 mm, and the sunshine duration is 1668 h. It is well-known that high protein levels or enzyme activities are not necessarily determined by a high gene transcript level [11]. In the present study, we sequenced the transcriptome of the peach by third-generation sequencing using the Oxford Nanopore Technologies (ONT) method, which is based on nanopore sequencing and single-molecule real-time (SMRT) sequencing [5]. We not only compared the transcriptome of the Olecranon honey peach with its bud variant but also predicted genes related to fruit type and shape.

Results

Comparison of two peach fruit quality indices

In this study, a marked difference in appearance between the Olecranon honey peach and its bud variant was observed (Fig. 1A and 1B). The tip of the Olecranon honey peach is shaped like an eagle's beak, and is large and warped. The two types of peaches also differ significantly in some quality indicators, such as weight, longitudinal diameter, transverse diameter, hardness, juice rate, color and luster, nutrients, and mineral elements (Table 1). The quality indicators, such as cluster analysis and NMDS analysis, of the two types of peaches also differ (Fig. 5A and 5B). Notably, the weight of the Olecranon honey peach was 205.50 ± 0.98 g, whereas that of the budding variant was 116.26 ± 1.23 g. The fruit shape index was calculated as the ratio of the longitudinal diameter to the transverse diameter. Their fruit shape indices did not differ significantly at 1.02 and 1.06, respectively. This implies that the shape of both the Olecranon honey peach and the bud mutant variant is almost circular. However, there were large differences in the longitudinal and transverse diameters of these peaches.

Transcriptomic analysis from ONT sequencing

Young fruits were sequenced because fruit shape, textural properties, and cutting content depend on events that occur early during fruit development [12]. To compare the characteristics based on the fruit type, gene expression dynamics and full-length transcripts of the young fruit of the two types of peaches were determined by ONT, and total clean data in the fastq format were further filtered to eliminate short reads and low-quality reads. After quality filtering, the BeadNum ranged from 5 687 141 to 7 035 534 with N50 ranging from 1263 to 1423; the mean q-score was Q9 in each library (Table 2). Through cDNA sequencing, reads can be identified as full-length sequences when primers have recognized both ends of the reads. The numbers of full-length reads in the six libraries were 5 195 328, 481 897, 4 795 588, 5 346

222, 4 312 262, and 4 847 747, respectively. The full-length percentages (FL%) were found to be 77.57%, 77.81%, 77.20%, 77.86%, 77.68%, and 76.97%, respectively (Table 3). The consensus isoform was obtained by polishing the full-length reads. Fusion transcript prediction was performed on obtained consistent transcripts, and the fusion transcripts from 7 to 12 were obtained for each sample. (S-Table 1).

Using the minimap2 software (Li 2018), all consensus isoforms were aligned to the reference genome for elimination of redundancy analysis, after obtaining 58 596 consensus isoforms. Transcripts above 500 base pairs were screened from the non-redundant transcripts, and MISA software was used for SSR analysis. A total of 21 745 SSRs were obtained, including 6 372 mononucleotide, 10 271 dinucleotide, 4 718 trinucleotide, 235 tetranucleotide, 48 pentanucleotide, and 101 hexanucleotide repeats. Density distribution of the different SSR types was statistically analyzed (Fig. 1D); the perfect dinucleotide (p2) was found to have the highest density, followed by the perfect mononucleotide (p1) and the perfect trinucleotide (p3). The perfect pentanucleotide (p5) and the hybrid SSR (p6) were found to have the lowest density. The non-redundant transcripts of all the samples were compared with the known annotations of the reference genome, revealing 2530 new gene loci and 37 364 new transcripts. Sequence analysis of the newly discovered transcripts revealed 24 205 complete open reading frame (ORF) sequences, 2603 transcription factors, and 1005 lncRNAs. By comparison with eight protein databases, 34 665 new transcripts were functionally annotated. In total, 8502 transcripts were annotated in the Cluster of Orthologous Groups of proteins (COG) database; 20 467 were annotated in the Gene Ontology Consortium (GO) database; 15 074 were annotated in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database; 19 471 were annotated in the euKaryotic Ortholog Groups (KOG) database; 14 296 were annotated in protein family (Pfam); 20 851 were annotated in Swiss-Prot; 31 324 were annotated in eggNOG; and 34 595 were annotated in NR. Ultimately, 34 665 transcripts were annotated in the eight databases (Table 4).

Alternative splicing

Alternative splicing (AS), the process of removing introns from pre-mRNA and rearranging exons to produce several types of mature transcripts, occurs before protein synthesis [13]. In the present study, 18,322 AS events were identified in the transcripts by AStalavista [14]. The main types of AS included exon skipping (ES), alternative 3' splice site (A3SS), mutually exclusive exon (MEE), alternative 5' splice site (A5SS), and intron retention (IR) (Fig. 1E). Among the five main types of alternatively spliced transcripts, IR was found to be predominant, accounting for 36.18% of the AS transcripts, followed by A3SS (32.58%), A5SS (15.97%), ES (14.16%), and MEE (1.11%) (S-Table 2). This finding differs from those of previous studies in various plants, such as tea, maize, cotton, rice, moso bamboo, *Arabidopsis*, and *Populus* [15-18], in which IR was predominant, followed by ES, A3SS, A5SS, and MEE. AS events were significantly more frequent in the Olecranon honey peach (from 3581 to 3860, an average of 3731) than those in the bud peach (from 2310 to 2446, an average of 2376) (S-Table 2). In the two groups of samples, the percentages of A5SS and ES in the Olecranon honey peach were significantly lower than those in the bud peach. However, IR showed the opposite results (Fig. 1C).

Differential expression analysis of gene and transcripts

Gene and transcript expression are time and location-specific. Under different conditions, the expression levels of genes and transcripts significantly differ. There were 457 differential expression genes (DEGs) in the two groups (Olecranon honey peach and the bud peach), including 169 up-regulated genes and 288 down-regulated genes (Fig. 2A). Up-regulated genes showed higher expression levels in the bud peach than in the Olecranon honey peach, whereas the down-regulated genes showed the opposite results. Hierarchical clustering analysis was performed for the screened DEGs, and genes with the same or similar expression patterns were clustered. The genes were divided into two subclasses (Fig. 2B). The 457 DEGs were annotated using the best BLAST hit from eight protein databases. A total of 390 genes were annotated in the eight databases, and 226 genes were annotated using the GO database. These 226 genes were annotated as DEGs and divided into three main GO functional categories (biological process, cellular component, and molecular function) and 36 subcategories (Fig. 2C). In the cellular component (CC) classification, the major subcategories were “membrane,” “membrane part,” “cell part,” and “cell.” For the molecular function (MF) classification, “catalytic activity” and “binding” were the dominant subgroups. “Metabolic process,” “cellular process,” and “single-organism process” were the major subcategories in the biological process (BP) classification. The GO annotation classifications of all genes were divided into 50 subcategories. The major subcategories of all genes were consistent with those of the DEGs. Ninety-nine DEGs were annotated in the COG database; “Carbohydrate transport and metabolism,” “transcription,” and “secondary metabolites biosynthesis, transport and catabolism” were the three significant subcategories in the COG functional classification analysis of the DEGs (Fig. 2D). To better explain the function of these DEGs, the DEGs were searched in the KEGG pathway database. Among the genes, 118 DEGs were annotated in the KEGG database, and 55 KEGG pathways were identified. Most were metabolism-related pathways, such as brassinosteroid biosynthesis (ko00905), starch and sucrose metabolism (ko00500), cysteine and methionine metabolism (ko00270), and other secondary metabolic pathways of peach plants. Additionally, there were 10 DEGs involved in genetic information processing, and 2 DEGs involved in cellular processes, environmental information processing, and organismal systems (Fig. 2E). Moreover, 164 DEGs were annotated in the KOG database; 385 DEGs were annotated in NR; 242 DEGs were annotated in Pfam; 256 DEGs were annotated in Swiss-Prot; and 325 DEGs were annotated in eggNOG.

Among the two groups (Olecranon honey peach and the bud mutant variant), 1519 transcripts were differentially expressed, of which 552 were up-regulated and 997 were down-regulated. These were divided into four subclasses (Fig. 3A). When the FDR was found to be more than 2, the transcripts were considered to be different. The expression of some transcripts in the two groups showed a fold-change greater than 10 (Fig. 3B). A total of 1332 differentially expressed transcripts (DETs) were annotated, with 796 annotated by GO. These 796 DETs were assigned into three main GO functional categories (BP, CC, and MF) and 41 subcategories (Fig. 3C). The major subcategories were “membrane,” “cell part,” and “cell” in the CC classification. As observed for the DEGs, for MF classification, “catalytic activity” and “binding” were the dominant major subgroups. “Metabolic process,” “cellular process,” and “single-organism process” were the major subcategories in the BP classification. GO annotation classifications of all

transcripts were divided into 52 subcategories. A total of 470 of the 1332 DETs were annotated by KEGG and matched 91 pathways (S-Table 3); most annotated pathways in the transcriptome were related to carbon metabolism and amino acid biosynthesis in a total of 17 DETs. Additionally, 349 DETs were annotated in the COG database; 611 DETs were annotated in the KOG database; 1320 DETs were annotated in NR; 721 DETs were annotated in Pfam; and 805 DETs were annotated in Swiss-Prot. Further, 1125 DETs were annotated in eggNOG, and most DETS were related to the post-translational modification, protein turnover, chaperone classes, and unknown functions (Fig. 3D). On the basis of these results, the number of annotated DEGs was found to be significantly lower than the number of DETs.

Differential expression of the genes for the fruit shape in two types of peaches

Fruit shape is an important trait affecting the appearance of peaches [3], including the Olecranon honey peach. In the bud sports, we found that both the peach diameter and the part resembling an eagle's beak became smaller. Studies have shown that plant hormones affect plant growth and are closely related to fruit development, including the shape of the fruit. For example, auxin and cytokinin are important plant hormones that regulate fruit development [19]. The concentration of auxin in the center of the fruit is higher than that of the peel, and gibberellin is related to cell division [20] and is responsible for swelling in the tomato fruit development [21]. We identified 18 DEGs related to plant hormones (S-Table 4), which were divided into up-regulated genes (*ONT.7681*, gene *5670*, gene *25091*, *ONT.9923*) and down-regulated genes (*ONT.16582*, gene *2826*, *ONT.13765*, gene *22504*, *ONT.16845*, gene *6609*, gene *2822*, gene *2460*, *ONT.1953*, gene *2824*, *ONT.1950*, gene *9229*, gene *2966*, gene *26004*); these genes were divided into three subgroups (S-Table 5). According to the GO annotation, the genes were related to biological regulation, reproductive processes, development processes, and growth (S-Table 5). The SAUR family protein (SAUR) is related to gene *9229* regulation in the signal transduction of plant hormones (ko04075) according to the KEGG annotation. It may play some role in the auxin signal transduction pathway involving calcium and calmodulin [22]. Gene *26004* regulates indole-3-pyruvate monooxygenase (YUCCA) in the tryptophan metabolism pathway (ko00380). This gene can promote the growth factors because it plays a role as a key auxin biosynthesis enzyme [23]. Gene *22504* codes for 3-epi-6-deoxocathasterone 23-monooxygenase (CYP90C1/ROT3) in the brassinosteroid biosynthesis pathway (ko00905). CYP90C1/ROT3 indicates the redundant brassinosteroid C-23 hydroxylase [24]. *ONT.1950*, *ONT.1953*, gene *2822*, gene *2824*, and gene *2826* regulate the PHYB activation-tagged suppressor named CYP73A1/BAS1 in ko00905, based on the KEGG annotation. *ONT.1950*, *ONT.1953*, gene *2822*, gene *2824*, and gene *2826* were shown to be involved in the brassinosteroid biosynthesis pathway with gene *22504*. CYP90C1 and CYP73A1 play a key role in brassinolide biosynthesis.

Differential expression of the transcripts for the fruit shape in two types of peaches

In the analysis of the signal transduction pathways for plant hormones, we identified seven DETs by the KEGG annotation: *ONT.5631.5*, *ONT.5631.9*, *rna13157*, *ONT.942.6*, *ONT.4473.2*, *rna4079*, and *rna4080*. Both *ONT.5631.5* and *ONT.5631.9* were annotated as AUX1/LAX as well as auxin influx carriers (AUX1/LAX family) in the ko04075 database. *Rna13157* was annotated as a SAUR family protein that

has been known to promote cell enlargement and plant growth (Fig. 4). *ONT.942.6* was annotated as AHP, which is a histidine-containing phosphotransfer protein (Fig. 4). This protein is important in promoting cell division. *ONT.4473.2* was annotated as EIN3 (Fig. 4) and was defined as an ethylene-insensitive protein 3. This protein facilitates fruit ripening and senescence. *Rna4079* and *rna4080* were annotated as TGA and defined as the transcription factor TGA by the KO annotation. TGA affects the resistance of plants to diseases [25]. In the analysis of the tryptophan metabolism pathway by the KEGG annotation, we identified three DETs: *rna36192*, *ONT.10241.19*, and *ONT.10241.4*. *Rna36192* is annotated as YUCCA, an indole-3- pyruvate monooxygenase (Fig. 4), and *ONT.10241.19* and *ONT.10241.4* were annotated as katE/CAT, belonging to the catalase family. In the brassinosteroid biosynthesis pathway (ko00905), we identified 11 DETs, which were divided into two classes: CYP734A1 or BAS1, including *ONT.1922.5*, *ONT.1923.2*, *ONT.1923.3*, *ONT.1949.1*, *ONT.1950.1*, *ONT.1953.1*, *rna3897*, *rna3899*, and *rna3900* (Fig. 4) and CYP90C1/ROT3, including *ONT.15392.3* and *rna31428*. The former is defined as PHYB activation-tagged suppressor 1, whereas the latter is a 3-epi-6-deoxocathasterone 23-monooxygenase (Fig. 4). Using DataViz visual data analysis software to further analyze the relationship between differential transcriptome and fruit shape, we discovered that the fruit shape index was related to *ONT.4473.2*, *ONT.5631.5*, *rna13157*, *rna4080*, *ONT.10241.19*, *ONT.10241.4*, *rna36192*, *ONT.15392.3*, *ONT.1923.2*, *ONT.1923.3*, *ONT.1949.1*, *ONT.1950.1*, *ONT.1953.1*, *rna3897*, *rna3899*, and *rna3900*, but not to *ONT.5631.9*, *ONT.942.6*, *rna4079*, or *ONT.1922.5*. Moreover, the transverse diameter was related to all transcriptomes, expect for *ONT.5631.9*. In contrast, the longitudinal diameter did not correlate with any feature of the transcriptome (fig. 5C).

Discussion

Bud mutations (bud sports), a consequence of somatic genetic variation, can lead to the occurrence of phenotypic alteration in plants [26]. Based on quality indicators, the bud variant of the peach differs remarkably from the Olecranon honey peach, particularly in fruit type. These differences indicate that although buds and the Olecranon honey peaches are grown on similar trees, they do not belong to the same variety. In this study, we evaluated the fruit type-related indicators, including weight, horizontal and vertical diameters, fruit type index, and the part of the fruit that resembles the shape of an eagle's beak. The bud mutant variant of the peach was typically smaller than the Olecranon honey peach based on these indicators. The Olecranon honey peach is a new variety of peach with high consumer value in China [2]. However, not all buds have a regulatory effect on the fruit type, such as the Beni Shogun apple mutation. The Beni Shogun mutation reportedly affect traits like skin coloration, fruit softening, and starch hydrolysis, whereas the loss of acidity, sugar accumulation, and weight are seemingly unaffected by the mutation [27].

To confirm that these two groups of peaches do not belong to the same variety, they were sequenced using the ONT SMRT long read isoform platform. This method overcomes the technical hurdles inherent in accurate prediction of the full-length splicing subtypes in second-generation sequencing short-read data [28]. The ONT platform is a new approach commonly used to evaluate plants and animals, such as the polar bear [12], *Hiptage benghalensis* [29], zebrafish [30], and *Sus scrofa* [31].

Three biological replicates for each cultivar were made into six cDNA libraries. For each replicate, approximately 6.10 GB of clean bases were generated using the BIOMARKER platform. More than 99.7% of the clean reads were successfully mapped to the peach reference genome (minimap 2.0) in each sample. A total of 2603 transcription factors were predicted to be new transcripts, belonging to 19 different families of transcription factors, including auxin-related genes expressing AUX/IAA family members.

Fruit ripening is a developmental process, and the fruit shape depends on events occurring early in fruit development [12, 32, 33]. The first stage post-anthesis, that is, the development process of the whole fruit development after fertilization, exhibits a rapid increase in cell division and cell proliferation [34]. This stage comprises differing cell division rates and durations in fruit tissues, and significantly impacts the final fruit shape [12]. Fruit shape formation initiates a complex responsive network, ranging from a transcriptional control to a post-translational control. In the signal transduction pathway for plant hormones, the down-regulated gene *9229* and the pair transcriptome *rna13157* were shown to affect auxin processing, promoting cell enlargement and plant growth with *ONT.5631.5* and *ONT.5631.9*. The down-regulated *ONT.942.6* was shown to act on the processing of cytokinin to promote cell division and shoot initiation. In a study of apple fruit type, fruit size was related to the ability in cellular expansion [35] and division [36]. In the tryptophan metabolism pathway, the gene *26004* showed a higher expression in the Olecranon honey peaches than that in the bud mutant variant. Notably, the gene *26004* codes for flavin monooxygenase, which was first identified as a key auxin biosynthesis enzyme [23]. Overexpression of YUC in *Arabidopsis* leads to auxin overproduction [37]. Brassinolide is an active brassinosteroid that promotes plant growth. In the brassinosteroid biosynthesis pathway, some DEGs (including gene *22504*, *ONT.1950*, *ONT.1953*, gene *2822*, gene *2824*, gene *2826*) and DETs (including *ONT.15392.3*, *rna31428*, *ONT.1922.5*, *ONT.1923.2*, *ONT.1923.3*, *ONT.1949.1*, *ONT.1950.1*, *ONT.1953.1*, *rna3897*, *rna3899*, *rna3900*) were shown to be important in the synthesis of brassinolides (Fig. 4 and Fig. 5C), which promote cell elongation and cell division. Reportedly, the *Arabidopsis* proteins CYP90C1/ROT3 and CYP734A1/BAS1 have important functions in leaf elongation [24, 38]. As described above, the primary gene families (Fig. 4) identified in this study coded for AUX1, SAUR, AHP, YUCCA, CYP90C1/ROT3, and CYP734A1/BAS1. These proteins regulate the production of large fruits by regulating endogenous hormones, secondary metabolites, and signal transduction. Gene *22504*, *ONT.1950*, *ONT.1953*, gene *2822*, gene *2824*, and gene *2826* are the primary factors affecting the shape of the Olecranon honey peach.

Conclusion

In this study, full-length transcriptome sequencing of the Olecranon honey peach and the bud mutant variant of the peach was performed by SMRT sequencing, to compare the DEGs and the DETs. Among them, down-regulation of *gene 9229, gene 26004,* gene *22504*, gene *2822*, gene *2826*, gene *2824*, *ONT.1953*, *ONT.1950*, and *ONT.1953* were related to the size of the Olecranon honey peach. Compared to an ideally round peach, the Olecranon honey peach is malformed, and its tip that resembles the shape of an eagle's beak determines fruit quality to some extent. Some studies have shown that an uneven

endogenous hormone content in fruits can lead to fruit asymmetry, such as in the Chinese gooseberry [39] and the flat peach. Further studies are required to determine whether down-regulation of certain DEGs (*gene 9229, gene 26004, gene 22504, gene 2822, gene 2826, gene 2824, ONT.1953, ONT.1950, and ONT.1953)*, as well as specific plant hormones are related to the beak structure of the Olecranon honey peach. Collectively, our findings may be useful for farm production of peaches. This study has potential for application in boosting the production of peaches for economic benefits.

Methods

Plant materials

The cultivars used in this study were Olecranon honey peach and bud variant chick peach, which were provided by the Lianping Shangping Chick Peach Base in Guangdong. Young fruits were immediately frozen in liquid nitrogen and stored at -80 ℃ for full-length transcriptome sequencing. Ripe fruit were selected by experienced fruit growers at maturity (fruit hanging time of approximately 75 days) and quickly transported to a well-ventilated, sterile laboratory for determination of fruit quality index.

Determination of the fruit quality index

Three free ripe fruits were randomly selected from each variety to determine the fruit quality index.

Weight and index of fruit shape

Peaches were weighed using an electronic balance (Tianli Instrument Co. Ltd., Changshu, China). The maximum longitudinal diameter (L) and the maximum transverse diameter (D) of the fruit were measured with a Vernier caliper. The fruit shape index was calculated according to the ratio of longitudinal diameter to transverse diameter of the fruit.

Determination of soluble solids

The soluble solid content of fruits was estimated by refractometry (LYT-390; Linyu Trading Co., Ltd., Shanghai, China). The fruits were cut into small pieces and put into a juicer (SJ21-150; Supor Co. Ltd., Zhengjiang, China) to obtain a homogenous juice. Two to three drops of the juice were placed on the surface of the prism, the sweeper was positioned to align with the light source, and the achromatic knob was adjusted to divide the field of vision into two.

Hardness assay

Fruit hardness was estimated using a durometer (GY-4; Zhejiang Tuopu Instrument Company, Zhejiang, China) according to previously published methods [40]. Briefly, the durometer was preheated for 10 min and set to zero, and hardness was subsequently measured after recording the peak value. Three points were chosen and placed parallel to each other on the measuring table, after which the handle was

pressed at a constant speed to read the peak value. The resulting readings were recorded as the hardness values in kg/cm^2.

Color variation

Three fruits from each group were selected prior to treatment to monitor color variation throughout the study according to the methods of Poel et al. [41] with minor modifications. Briefly, fruit skin color was evaluated at consistent positions (selected from the mouth, left of the concave line of the peach body, and right of the concave line of the peach body) using a colorimeter (CS-260; CHNSpec Technology Co., Ltd., Hangzhou, China).

Titratable acid

Following the method of Sandie et al. [42], titratable acidity was measured by blending 50 g of diced flesh with 50 ml of deionized water for 2 min. The puree was titrated with 0.1 M NaOH to pH 8.1 using an autotitrator. Results were expressed as mg/g of malic acid equivalent (MAE). The measurements for each treatment were repeated three times and the results were averaged.

Full-length transcriptome sequencing

The experimental procedure was used the standard protocol from ONT, including sample quality testing, library construction, library quality testing, and library sequencing. The construction of the ONT library uses the method for the connection sequencing kit (SQK-LSK109), which is directly constructed by end repair and adding sequencing adapters, using the PromethION sequencer. The database-building process is as follows: First, RNA was extracted for detection of purity, concentration, and sample integrity by advanced molecular biology equipment to ensure the use of qualified samples for transcriptome sequencing. Thereafter, libraries were constructed, by primer annealing, reversing transcription into cDNA and addition of switching oligos, then synthesizing complementary strands, DNA damage repair, end repair, and magnetic bead purification. Finally, the sequencing linker was added before machine sequencing.

Oxford Nanopore Technologies long read processing

Raw reads were first filtered with a minimum average read quality score of 7 and a minimum read length of 500 bp. Ribosomal RNA were discarded after mapping to the rRNA database. Next, full-length, non-chimeric (FLNC) transcripts were determined by searching for primers at both ends of reads. Clusters of FLNC transcripts were obtained after mapping to the reference genome with mimimap2, and consensus isoforms were obtained after polishing within each cluster by pinfish.

Removal of redundancies

Consensus sequences were mapped to the reference genome using minimap2 [43]. Mapped reads were further collapsed by the cDNA Cupcake package with min-coverage=85% and min-identity=90%. The 5

difference was not considered when collapsing redundant transcripts.

Search for fusion transcripts

The criteria for fusion candidates are that a single transcript must have: mapping to 2 or more loci; minimum coverage for each locus of 5% and minimum coverage in bp of ≧1 bp; total coverage of ≧95%; and distance between the loci of at least 10 kb.

Structure analysis

Transcripts were validated against known reference transcript annotations with gffcompare. AS events, including IR, ES, AD, AA, and MEE, were identified by the AStalavista tool. SSRs of the transcriptome were identified using MISA. APA analysis was conducted with TAPIS [28].

Transcription factor prediction

Plant transcription factors were identified with iTAK [44].

lncRNA analysis

Four computational approaches, including CPC/CNCI/CPAT/Pfam, were combined to sort non-protein-coding RNA candidates from putative protein-coding RNAs in the transcripts. Putative protein-coding RNAs were filtered out using a minimum length and exon number threshold. Transcripts having lengths more than 200 nt and more than two exons were selected as lncRNA candidates and further screened using CPC/CNCI/CPAT/Pfam, which can distinguish the protein-coding genes from the non-coding genes.

Gene functional annotation

Gene function was annotated based on the following databases: NR (NCBI non-redundant protein sequences); Pfam; KOG/COG/eggNOG (Clusters of Orthologous Groups of proteins); Swiss-Prot (a manually annotated and reviewed protein sequence database); KEGG; and GO.

Quantification of gene/transcript expression levels and differential expression analysis

Full-length reads were mapped to the reference transcriptome sequence. Reads with a match quality above 5 were further used for quantification. Expression levels were estimated by reads per gene/transcript per 10 000 reads mapped.

Samples with biological replicates

A differential expression analysis of two conditions/groups was performed using the DESeq R package (1.10.1). DESeq provides statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. The resulting P values were adjusted using Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with an FDR <0.01 and fold change ≥2 found by DESeq were assigned as differentially expressed.

Samples without biological replicates

Prior to the differential gene expression analysis, for each sequenced library, the read counts were adjusted by edgeR program package through one scaling normalized factor. Differential expression analysis of two samples was performed using the EBSeq R package. The resulting FDR (false discovery rate) was adjusted using the PPDE (posterior probability of being DE). FDR <0.01 and fold change ≥2 were set as the thresholds for significantly differential expression.

Functional enrichment analysis

GO enrichment analysis of the DEGs was implemented by the GO seq R packages based on Wallenius' non-central hyper-geometric distribution [45], which can adjust for gene length bias in DEGs. KEGG [46] is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism, and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies (http://www.genome.jp/kegg/). We used KOBAS [47] software to test the statistical enrichment of differential expression genes in KEGG pathways.

Protein-protein interaction

The sequences of the DEGs were BLAST searched (BLASTX) against the genome of related species (the protein-protein interaction of which exists in the STRING database: http://string-db.org/) to get the predicted PPI of these DEGs. Then the PPI of these DEGs was visualized in Cytoscape [48].

Statistical analyses

Data are expressed as the mean ± standard deviation (SD) of three replicates. Significant differences between the means of different parameters were calculated using Duncan's multiple-range test using SPSS 17.0 software (SPSS Inc., Chicago, IL). The threshold of $p < 0.05$ was considered to indicate statistical significance.

Preprocessed GeoChip data were further analyzed with different statistical methods: (i) hierarchical clustering for peach-population phenotype and transcriptome; (ii) detrended correspondence analysis combined with analysis of similarities, nonparametric multivariate analysis of variance (Adonis) and multi-response permutation procedure for determining the overall functional changes in the peaches; (iii) significant Pearson's linear correlation (r) analysis, as well as analysis of variance; (iv) use of DataViz visual data analysis software to further analyze the relationship between differential transcriptome and fruit shape.

Abbreviations

ONT: Oxford Nanopore Technologies; SMRT: single-molecule real-time; AS: alternative splicing; ORF: open reading frame; COG: Cluster of Orthologous Groups of proteins; GO: Gene Ontology Consortium; KEGG:

Kyoto Encyclopedia of Genes and Genomes; KOG: euKaryotic Ortholog Groups; Pfam: protein family; AS: alternative splicing; ES: exon skipping; A3SS: alternative 3' splice site; MEE: mutually exclusive exon; A5SS: alternative 5' splice site; IR: intron retention; DEGs: differential expression genes; MF: molecular function; CC: cellular component; BP: biological process; DETs: differentially expressed transcripts; SAUR: SAUR family protein; YUCCA: indole-3-pyruvate monooxygenase; CYP90C1/ROT3: 3-epi-6-deoxocathasterone 23-monooxygenase; L: maximum longitudinal diameter; D: maximum transverse diameter; MAE: malic acid equivalent; FLNC: full-length, non-chimeric; NR: NCBI non-redundant protein sequences; FDR: false discovery rate; PPDE: posterior probability of being DE; PPI: protein-protein interaction; SD: standard deviation.

Declarations

Acknowledgments

We thank Biomarker Technologies (Beijing) for helping with transcriptome sequencing.

Authors' contributions

JLL, YB, QW, and HFL designed the study. JLL, YB, and HFL performed the experiments and analyzed the results. YB and HFL wrote the manuscript. JLL, YB, QW, and HFL participated in the result analyses and interpretation of data. HFL revised the manuscript critically. All authors have read and approved the final manuscript.

Funding

This work was funded by the Key-Area Research and Development Program of Guangdong Province (2019B020238003) awarded to JLL and the Guangdong Provincial Agricultural Science and Technology Special Commissioner Precision Poverty Alleviation Rural Rejuvenation Project (KA1901003) awarded to HFL. The funder had no role in the design of the study; in the collection, analysis, and interpretation of data; or in writing the manuscript.

Availability of data and materials

The sequencing data are available in the NCBI genome database under the accession number GCF000346465.2. https://www.ncbi.nlm.nih.gov/genome/?term=prunus+persica. The datasets supporting the results of this article are included within the article and the additional files. These samples were authenticated by experts from the Department of Modern Agriculture Research Center, Zhongkai University of Agriculture and Engineering, China.

Ethics approval and consent to participate

Not applicable.

Consent to publish

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Authors' information

[1] College of Light Industry and Food, Zhongkai University of Agriculture and Engineering, Guangzhou, Guangdong, 510225, China. [2] Modern Agriculture Research Center，Zhongkai University of Agriculture and Engineering, Guangzhou, Guangdong, 510225, China.

References

1. Shulaev V, Korban SS, Sosinski B, Abbott AG, Aldwinckle HS, Folta KM, Iezzoni A, Main D, Arus P, Dandekar AM *et al*: **Multiple models for Rosaceae genomics**. *Plant Physiol* 2008, **147**(3):985-1003.https://doi.org/10.1104/pp.107.115618.
2. Wang Q, Zhang H, Liu H, Liu J: **Effect of different preservation treatments on olecranon honey peach**. *Journal of Food Processing and Preservation* 2019, **43**(7).https://doi.org/10.1111/jfpp.13960.
3. Cao K, Zhou Z, Wang Q, Guo J, Zhao P, Zhu G, Fang W, Chen C, Wang X, Wang X *et al*: **Genome-wide association study of 12 agronomic traits in peach**. *Nat Commun* 2016, **7**:13246.https://doi.org/10.1038/ncomms13246.
4. Li Y, Cao K, Zhu G, Fang W, Chen C, Wang X, Zhao P, Guo J, Ding T, Guan L *et al*: **Genomic analyses of an extensive collection of wild and cultivated accessions provide new insights into peach breeding history**. *Genome Biol* 2019, **20**(1):36.https://doi.org/10.1186/s13059-019-1648-9.
5. Deamer D, Akeson M, Branton D: **Three decades of nanopore sequencing**. *Nat Biotechnol* 2016, **34**(5):518-524.https://doi.org/10.1038/nbt.3423.
6. Magi A, Semeraro R, Mingrino A, Giusti B, D'Aurizio R: **Nanopore sequencing data analysis: state of the art, applications and challenges**. *Brief Bioinform* 2018, **19**(6):1256-

1272.https://doi.org/10.1093/bib/bbx062.

7. Jain M, Olsen HE, Paten B, Akeson M: **The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community**. *Genome Biol* 2016, **17**(1):239.https://doi.org/10.1186/s13059-016-1103-0.
8. Krishnakumar R, Sinha A, Bird SW, Jayamohan H, Edwards HS, Schoeniger JS, Patel KD, Branda SS, Bartsch MS: **Systematic and stochastic influences on the performance of the MinION nanopore sequencer across a range of nucleotide bias**. *Sci Rep* 2018, **8**(1):3159.https://doi.org/10.1038/s41598-018-21484-w.
9. Liu Q, Zhu A, Chai L, Zhou W, Yu K, Ding J, Xu J, Deng X: **Transcriptome analysis of a spontaneous mutant in sweet orange [Citrus sinensis (L.) Osbeck] during fruit development**. *J Exp Bot* 2009, **60**(3):801-813.https://doi.org/10.1093/jxb/ern329.
10. Chatelet P, Laucou V, Fernandez L, Sreekantan L, Lacombe T, Martinez-Zapater JM, Thomas MR, Torregrosa L: **Characterization of Vitis vinifera L. somatic variants exhibiting abnormal flower development patterns**. *J Exp Bot* 2007, **58**(15-16):4107-4118.https://doi.org/10.1093/jxb/erm269.
11. Lu P, Vogel C, Wang R, Yao X, Marcotte EM: **Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation**. *Nat Biotechnol* 2007, **25**(1):117-124.https://doi.org/10.1038/nbt1270.
12. Gallusci P, Hodgman C, Teyssier E, Seymour GB: **DNA Methylation and Chromatin Regulation during Fleshy Fruit Development and Ripening**. *Front Plant Sci* 2016, **7**:807.https://doi.org/10.3389/fpls.2016.00807.
13. Chen MX, Wijethunge B, Zhou SM, Yang JF, Dai L, Wang SS, Chen C, Fu LJ, Zhang J, Hao GF *et al*: **Chemical Modulation of Alternative Splicing for Molecular-Target Identification by Potential Genetic Control in Agrochemical Research**. *J Agric Food Chem* 2019, **67**(18):5072-5084.https://doi.org/10.1021/acs.jafc.9b02086.
14. Foissac S, Sammeth M: **ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets**. *Nucleic Acids Res* 2007, **35**(Web Server issue):W297-299.https://doi.org/10.1093/nar/gkm311.
15. Qiao D, Yang C, Chen J, Guo Y, Li Y, Niu S, Cao K, Chen Z: **Comprehensive identification of the full-length transcripts and alternative splicing related to the secondary metabolism pathways in the tea plant (Camellia sinensis)**. *Sci Rep* 2019, **9**(1):2709.https://doi.org/10.1038/s41598-019-39286-z.
16. Thatcher SR, Zhou W, Leonard A, Wang BB, Beatty M, Zastrow-Hayes G, Zhao X, Baumgarten A, Li B: **Genome-wide analysis of alternative splicing in Zea mays: landscape and genetic regulation**. *Plant Cell* 2014, **26**(9):3472-3487.https://doi.org/10.1105/tpc.114.130773.
17. Dong C, He F, Berkowitz O, Liu J, Cao P, Tang M, Shi H, Wang W, Li Q, Shen Z *et al*: **Alternative Splicing Plays a Critical Role in Maintaining Mineral Nutrient Homeostasis in Rice (Oryza sativa)**. *Plant Cell* 2018, **30**(10):2267-2285.https://doi.org/10.1105/tpc.18.00051.
18. Marquez Y, Brown JWS, Simpson C, Barta A, Kalyna M: **Transcriptome survey reveals increased complexity of the alternative splicing landscape inArabidopsis**. *Genome Research* 2012, **22**(6):1184-

1195.https://doi.org/10.1101/gr.134106.111.

19. Seymour GB, Ostergaard L, Chapman NH, Knapp S, Martin C: **Fruit development and ripening**. *Annu Rev Plant Biol* 2013, **64**:219-241.https://doi.org/10.1146/annurev-arplant-050312-120057.
20. Pattison RJ, Catala C: **Evaluating auxin distribution in tomato (Solanum lycopersicum) through an analysis of the PIN and AUX/LAX gene families**. *Plant J* 2012, **70**(4):585-598.https://doi.org/10.1111/j.1365-313X.2011.04895.x.
21. Srivastava A, Handa AK: **Hormonal Regulation of Tomato Fruit Development: A Molecular Perspective**. *Journal of Plant Growth Regulation* 2005, **24**(2):67-82.https://doi.org/10.1007/s00344-005-0015-0.
22. Guilfoyle GHT: **Auxin-responsive gene expression: genes, promoters and regulatory factors**. *Plant Mol Biol* 2002(49:):373-385
23. Zhao Y: **Auxin biosynthesis: a simple two-step pathway converts tryptophan to indole-3-acetic acid in plants**. *Mol Plant* 2012, **5**(2):334-338.https://doi.org/10.1093/mp/ssr104.
24. Ohnishi T, Szatmari AM, Watanabe B, Fujita S, Bancos S, Koncz C, Lafos M, Shibata K, Yokota T, Sakata K *et al*: **C-23 hydroxylation by Arabidopsis CYP90C1 and CYP90D1 reveals a novel shortcut in brassinosteroid biosynthesis**. *Plant Cell* 2006, **18**(11):3275-3288.https://doi.org/10.1105/tpc.106.045443.
25. Johnson C, Boden E, Arias J: **Salicylic Acid and NPR1 Induce the Recruitment of trans-Activating TGA Factors to a Defense Gene Promoter in Arabidopsis**. *The Plant Cell* 2003, **15**(8):1846-1858.https://doi.org/10.1105/tpc.012211.
26. NAITHANI SP, RAGHUVANSHI SS: **Cytogenetical Studies in the Genus Citrus**. *NATURE* 1968, **181**(4620):1406-1407.https://doi.org/10.1038/1811406b0.
27. Dong QL, Yan ZY, Liu Z, Yao YX: **Early ripening events caused by bud mutation in Beni Shogun apple**. *Russian Journal of Plant Physiology* 2011, **58**(3):439-447.https://doi.org/10.1134/s1021443711030034.
28. Abdel-Ghany SE, Hamilton M, Jacobi JL, Ngam P, Devitt N, Schilkey F, Ben-Hur A, Reddy AS: **A survey of the sorghum transcriptome using single-molecule long reads**. *Nat Commun* 2016, **7**:11706.https://doi.org/10.1038/ncomms11706.
29. Tian B, Lu T, Xu Y, Wang R, Chen G: **Identification of genes associated with ricinoleic acid accumulation in Hiptage benghalensis via transcriptome analysis**. *Biotechnol Biofuels* 2019, **12**:16.https://doi.org/10.1186/s13068-019-1358-2.
30. Nudelman G, Frasca A, Kent B, Sadler KC, Sealfon SC, Walsh MJ, Zaslavsky E: **High resolution annotation of zebrafish transcriptome using long-read sequencing**. *Genome Res* 2018, **28**(9):1415-1425.https://doi.org/10.1101/gr.223586.117.
31. Li Y, Fang C, Fu Y, Hu A, Li C, Zou C, Li X, Zhao S, Zhang C, Li C: **A survey of transcriptome complexity in Sus scrofa using single-molecule long-read sequencing**. *DNA Res* 2018, **25**(4):421-437.https://doi.org/10.1093/dnares/dsy014.

32. Chaib J, Devaux MF, Grotte MG, Robini K, Causse M, Lahaye M, Marty I: **Physiological relationships among physical, sensory, and morphological attributes of texture in tomato fruits**. *J Exp Bot* 2007, **58**(8):1915-1925.https://doi.org/10.1093/jxb/erm046.

33. van der Knaap E, Chakrabarti M, Chu YH, Clevenger JP, Illa-Berenguer E, Huang Z, Keyhaninejad N, Mu Q, Sun L, Wang Y *et al*: **What lies beyond the eye: the molecular mechanisms regulating tomato fruit weight and shape**. *Front Plant Sci* 2014, **5**:227.https://doi.org/10.3389/fpls.2014.00227.

34. Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E: **A Retrotransposon-Mediated Gene Duplication Underlies Morphological Variation of Tomato Fruit**. *Science* 2008, **319**(5869):1527-1530.https://doi.org/10.1126/science.1153040.

35. Malladi A, Hirst PM: **Increase in fruit size of a spontaneous mutant of 'Gala' apple (Malus x domestica Borkh.) is facilitated by altered cell production and enhanced cell size**. *J Exp Bot* 2010, **61**(11):3003-3013.https://doi.org/10.1093/jxb/erq134.

36. Harada T, Kurahashi W, Yanai M, Wakasa Y, Satoh T: *Scientia Horticulturae* 2005, **105**(4):447-456.https://doi.org/10.1016/j.scienta.2005.02.006.

37. Zhao. Y, K. S, R. J, D J: **A Role for Flavin Monooxygenase-Like Enzymes in Auxin Biosynthesis**. *Science* 2001, **291**:306-309.https://doi.org/10.1126/science.291.5502.306.

38. Turk EM, Fujioka S, Seto H, Shimada Y, Takatsuto S, Yoshida S, Wang H, Torres QI, Ward JM, Murthy G *et al*: **BAS1 and SOB7 act redundantly to modulate Arabidopsis photomorphogenesis via unique brassinosteroid inactivation mechanisms**. *Plant J* 2005, **42**(1):23-34.https://doi.org/10.1111/j.1365-313X.2005.02358.x.

39. Hopping ME: **Effect of exogenous auxins, gibberellins, and cytokinins on fruit development in Chinese gooseberry (Actinidia chinensisPlanch.)**. *New Zealand Journal of Botany* 1976, **14**(1):69-75.https://doi.org/10.1080/0028825x.1976.10428652.

40. Scheidt TB, Silva FVM: **High pressure processing and storage of blueberries: effect on fruit hardness**. *High Pressure Research* 2017, **38**(1):80-89.https://doi.org/10.1080/08957959.2017.1402895.

41. Van de Poel B, Bulens I, Hertog MLATM, Van Gastel L, De Proft MP, Nicolai BM, Geeraerd AH: **Model-based classification of tomato fruit development and ripening related to physiological maturity**. *Postharvest Biology and Technology* 2012, **67**:59-67.https://doi.org/10.1016/j.postharvbio.2011.12.005.

42. Møller SM, Travers S, Bertram HC, Bertelsen MG: **Prediction of postharvest dry matter, soluble solids content, firmness and acidity in apples (cv. Elshof) using NMR and NIR spectroscopy: a comparative study**. *European Food Research and Technology* 2013, **237**(6):1021-1024.https://doi.org/10.1007/s00217-013-2087-6.

43. Li H: **Minimap2: pairwise alignment for nucleotide sequences**. *Bioinformatics* 2018, **34**(18):3094-3100.https://doi.org/10.1093/bioinformatics/bty191.

44. Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, Zhang P, Banf M, Dai X, Martin GB, Giovannoni JJ *et al*: **iTAK: A Program for Genome-wide Prediction and Classification of Plant Transcription Factors,**

Transcriptional Regulators, and Protein Kinases. *Mol Plant* 2016, **9**(12):1667-1670.https://doi.org/10.1016/j.molp.2016.09.014.

45. Tatusov RL, Galperin MY, Natale DA: **The COG database: a tool for genome-scale analysis of protein functions and evolution**. *Nucleic Acids Res* 2000, **28**(1):33-36
46. Kanehisa M: **The KEGG resource for deciphering the genome**. *Nucleic Acids Research* 2004, **32**(90001):277D-280.https://doi.org/10.1093/nar/gkh063.
47. Wu J, Shen J, Mao X, Liu K, Wei L, Liu P, Yang G: **Isolation and analysis of differentially expressed genes in dominant genic male sterility (DGMS) Brassica napus L. using subtractive PCR and cDNA microarray**. *Plant Science* 2007, **172**(2):204-211.https://doi.org/10.1016/j.plantsci.2006.08.010.
48. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks**. *Genome Res* 2003, **13**(11):2498-2504.https://doi.org/10.1101/gr.1239303.

Tables

Table 1 Comparison of the quality indices of the two peach fruits.

Project		Bud mutant variant of the peach	Olecranon honey peach
Weight (g)		116.26 ± 1.23	205.50 ± 0.98
Longitudinal diameter (cm)		62.92 ± 1.89	73.50 ± 1.67
Transverse diameter (cm)		59.26 ± 1.09	72.09 ± 1.34
Fruit shape index		1.06 ± 1.56	1.02 ±1.08
Hardness		1.634 ± 0.40	1.925 ± 0.51
Juice rate (%)		52.29 ± 0.12	42.88 ± 0.05
Soluble solids (%)		10.55 ± 0.00	10.33 ± 0.01
Color and luster	△a	-2.71 ± 1.91	-12.52 ± 0.19
	△b	25.82 ± 1.25	25.25 ± 1.38
	△L	38.95 ± 1.39	36.00 ± 3.58
	△E	53.35 ± 2.38	47.28 ± 3.72

Table 2 Clean data statistical table. After the original fastq data were further filtered for short reads and low-quality reads, the total clean data were obtained. A-1; A-2; A-3: Olecranon honey peach; B-1; B-2; B-3: Bud mutant variant of the peach.

Sample Name	Read Num	Base Num	N50	Mean Length	Max Length	Mean Q-score
A-1	6,855,766	8,140,421,280	1,347	1,187	10,170	Q9
A-2	6,344,028	7,573,280,925	1,339	1,193	11,367	Q9
A-3	6,366,073	8,018,115,098	1,448	1,259	10,014	Q9
B-1	7,035,534	8,059,133,822	1,263	1,145	21,540	Q9
B-2	5,687,141	7,102,833,899	1,423	1,248	13,080	Q9
B-3	6,458,809	8,229,558,235	1,410	1,274	21,190	Q9

Table 3 Statistics of the full-length sequence data. In cDNA sequencing, the primer identified at both ends of the reads was considered as a full-length sequence.

Sample ID	Number of clean reads (except rRNA)	Number of full-length reads	Full-length percentage (FL%)
A-1	6,698,079	5,195,382	77.57%
A-2	6,193,058	4,818,907	77.81%
A-3	6,211,916	4,795,588	77.20%
B-1	6,866,714	5,346,222	77.86%
B-2	5,551,215	4,312,262	77.68%
B-3	6,298,495	4,847,747	76.97%

Table 4 Statistics on the number of new transcripts annotated. Functional annotation was performed on new transcripts obtained from an alternative splicing analysis

Annotated database	New isoforms
COG	8502
GO	20,467
KEGG	15,074
KOG	19,471
Pfam	14,296
Swiss-Prot	20,851
eggNOG	31,324
nr	34,595
All	34,665

Supplementary Information

Additional files 1: Table S1. Fusion transcript of each sample in the GFF file table.

Additional files 2: Table S2. Alternative splicing statistics.

Additional files 3: Table S3. KEGG path annotation.

Additional files 4: Table S4. Gene enrichment in the selected samples.

Additional files 5: Table S5. GO annotation classification of the selected genes.

Figures

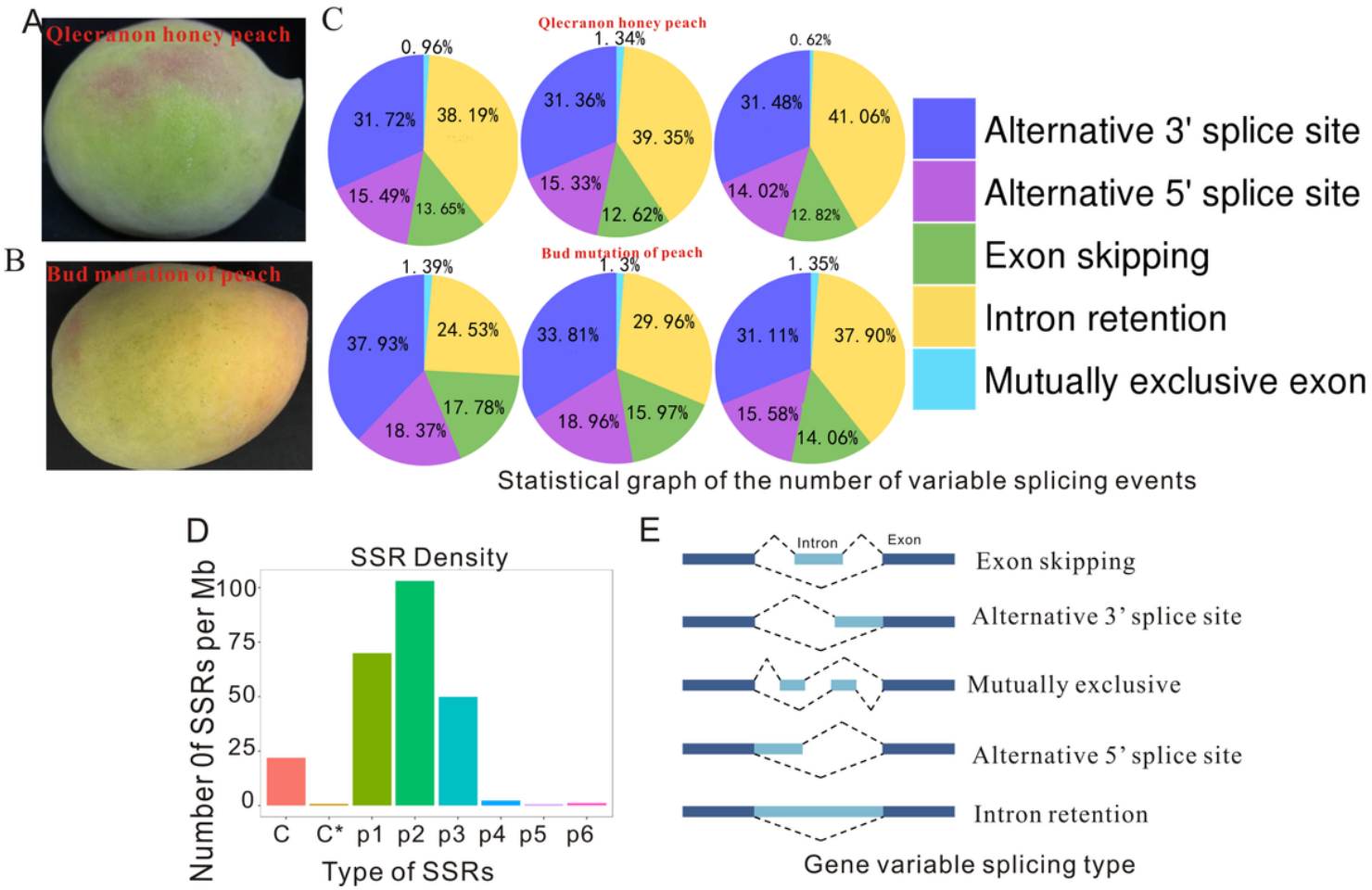

Figure 1

Appearance of the peach and its buds, and transcriptomic analysis from ONT sequencing. (A) Olecranon honey peach. (B) Bud mutant variant of the peach. (C) Number of variable splicing events. The two samples were evaluated 3 times in parallel. (D) Density distribution of different SSR types. (E) Gene variable splicing type.

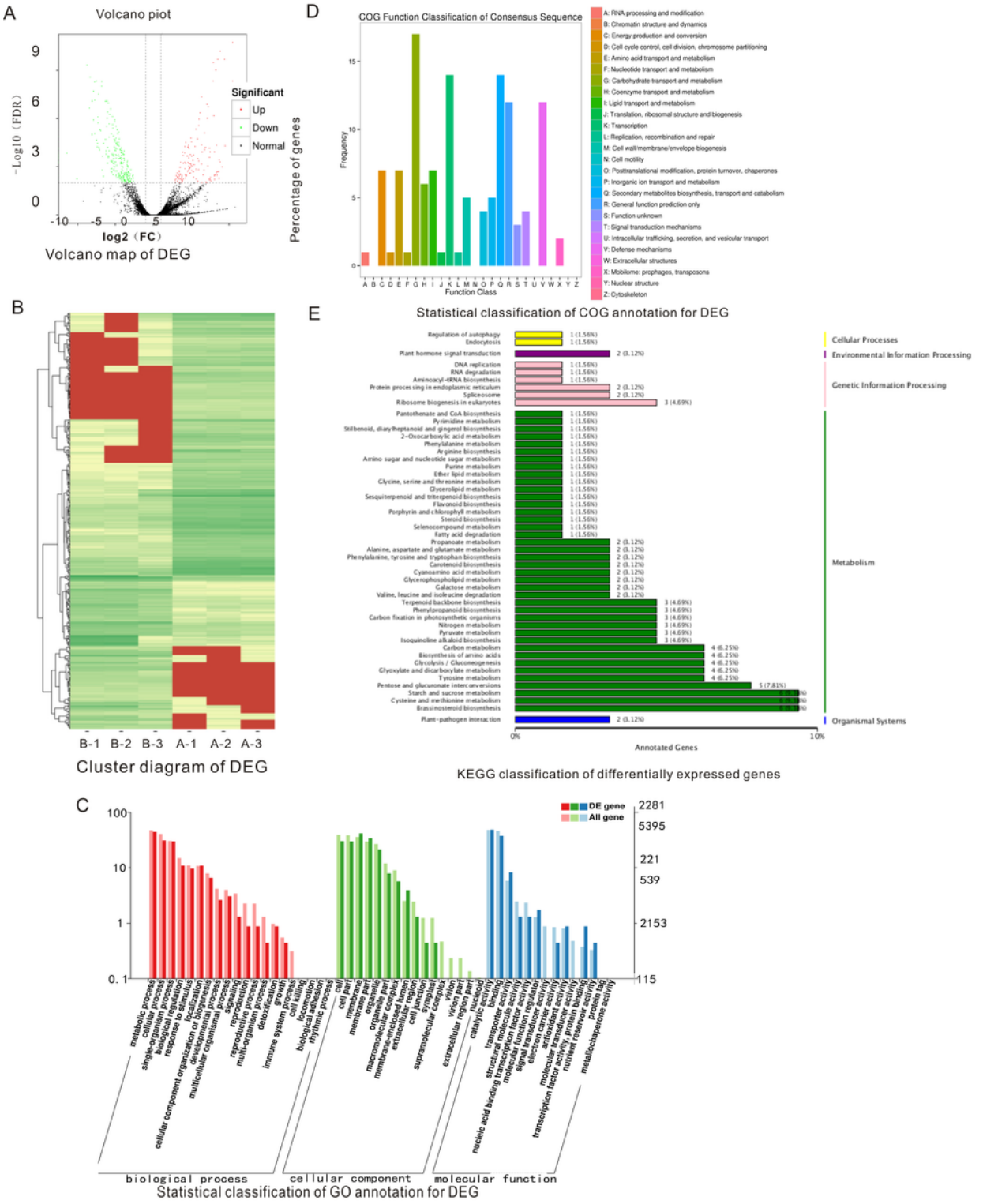

Figure 2

Analysis of DEGs in the two peach species. (A) Volcano map of the DEGs. Each point in the differential expression volcano chart represents a gene, and the abscissa represents the logarithm of the multiple of the difference in expression of a certain gene in the two samples; the ordinate represents the negative logarithm of statistical significance of the change in gene expression. A larger absolute value of the abscissa indicates a greater fold-difference in expression between the two samples; a larger ordinate

value indicates more significant differential expression and greater reliability of the DEGs obtained by screening. The green dots in the figure represent the down-regulated DEGs, red dots represent the up-regulated DEGs, and black dots represent the non-DEGs. (B) Cluster diagram of the DEGs. The abscissa represents the sample name and clustering result of the sample, and the ordinate represents the differential gene and clustering result of the gene. Different columns in the figure represent different samples, and different rows represent different genes. The color represents the expression level log2 (CPM + 1e-6) of the gene in the sample. (C) Statistical classification of the GO annotation for the DEG. The abscissa represents the GO classification, left of the ordinate represents the percentage of the number of genes, and right represents the number of genes. (D) Statistical classification of the COG annotation for DEG. The abscissa represents the content of each classification of COG, and the ordinate represents the number of genes. (E) The KEGG classification of DEGs. The ordinate represents the name of the KEGG metabolic pathway, and the abscissa represents the number of genes annotated to the pathway and their proportion to the total number of genes annotated.

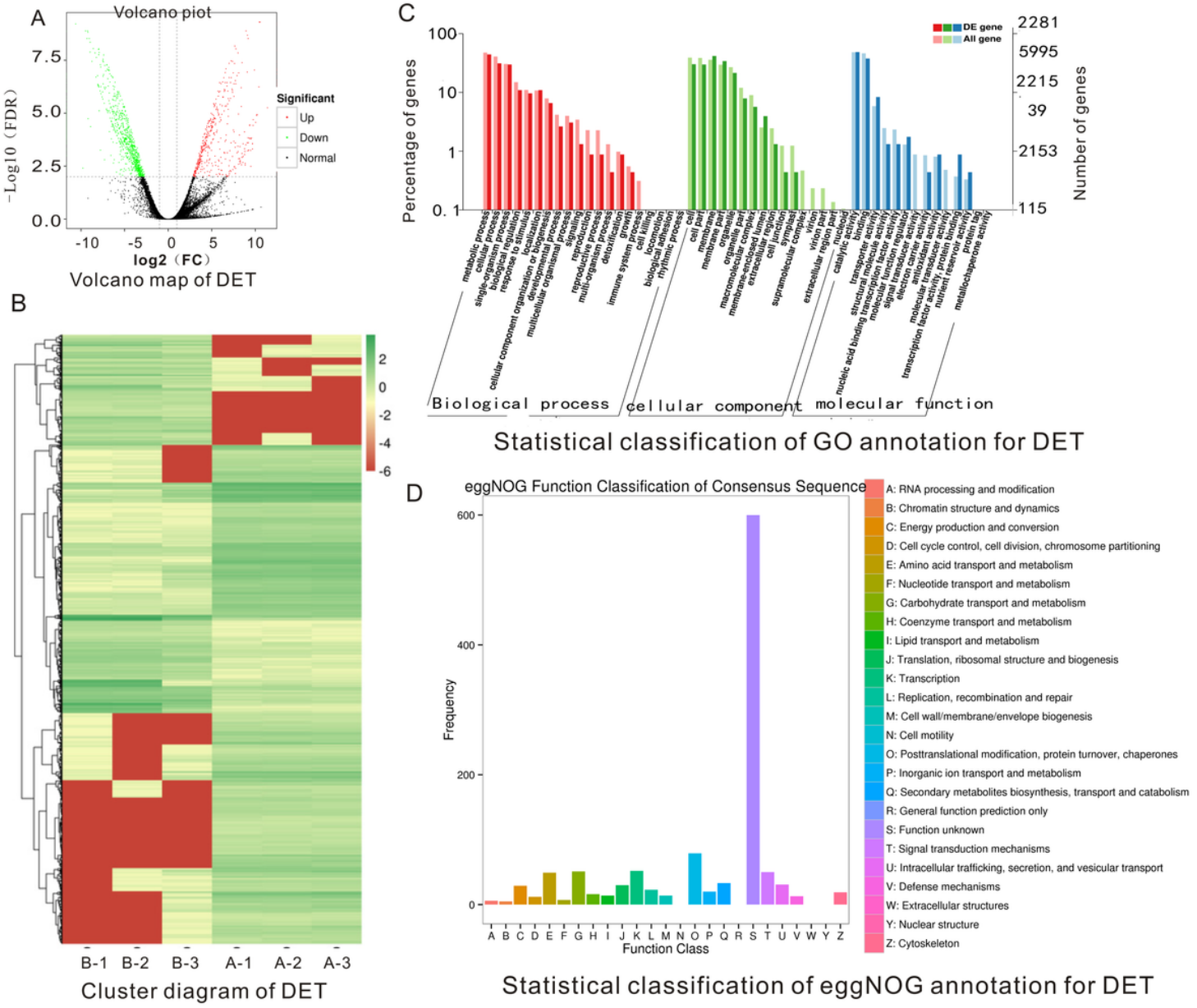

Figure 3

Analysis of the differentially expressed transcriptome in the two peach species. (A) Volcano map of DET. (B) Cluster diagram of DET. (C) Statistical classification of the GO annotation for DET. (D) Statistical classification of the eggNOG annotation for DET. The abscissa represents the contents of the eggNOG classification, and the ordinate represents the number of genes.

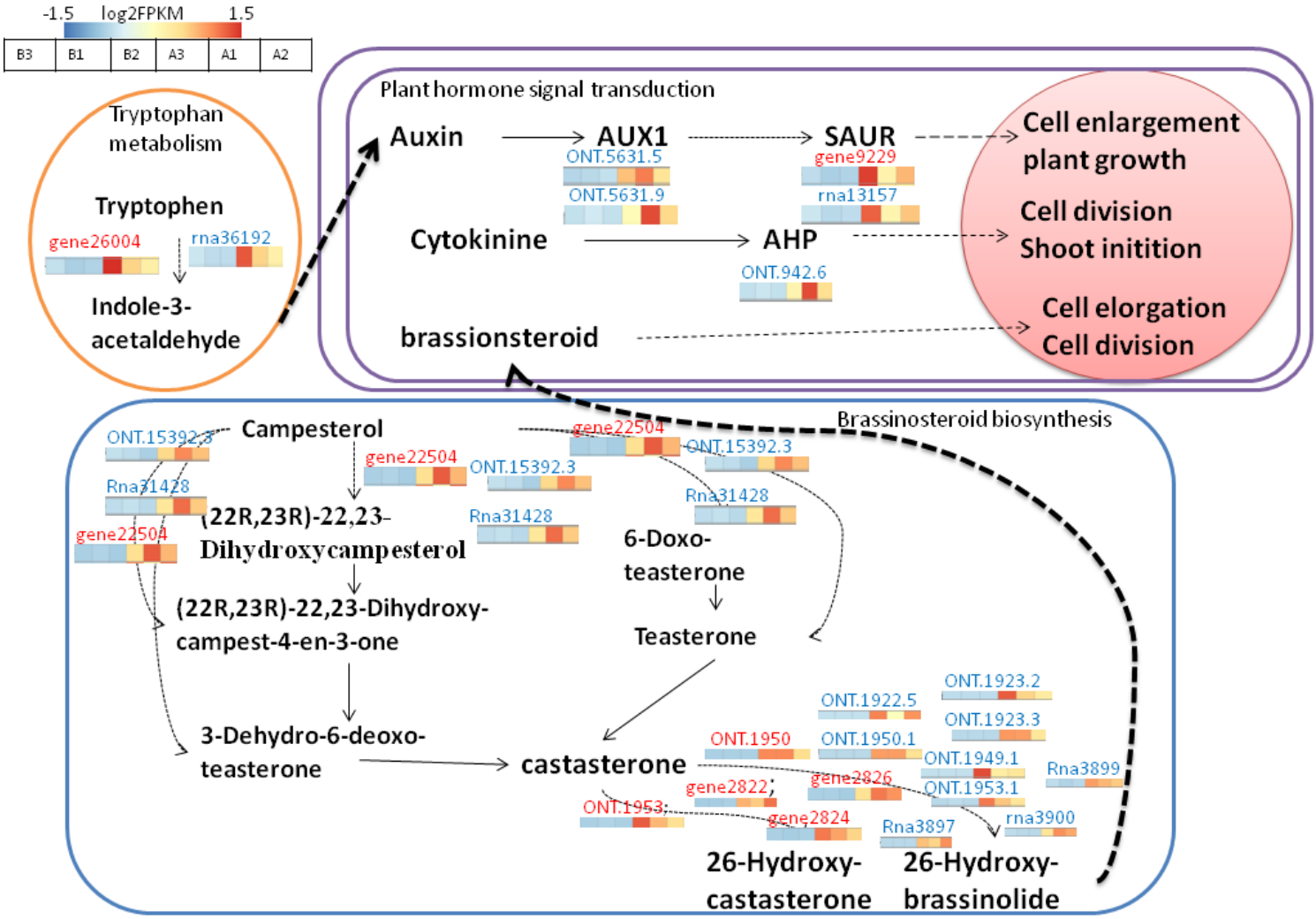

Figure 4

Proposed gene networks involved in the fruit shape in the Olecranon honey peach. The expression levels (represented by the Log2FPKM) of possible candidates have been highlighted in color scales (blue to red) in the two peach species. (A) represents the Olecranon honey peach; (B) represents the bud mutant variant of the peach; and 1, 2, and 3 represent three studies in parallel. In the networks, the letters are in red for genes and in blue for transcriptomes.

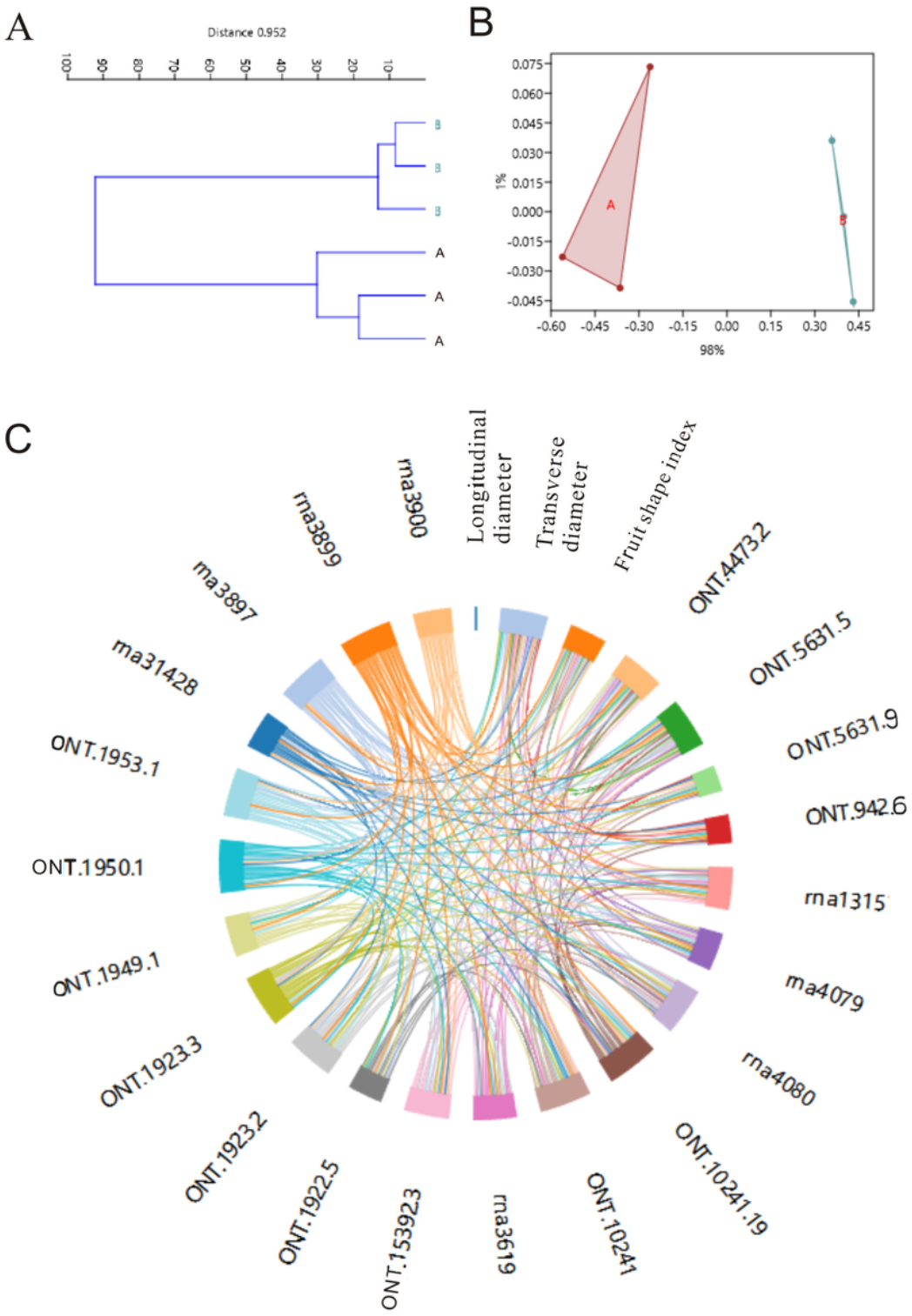

Figure 5

Preprocessed GeoChip data were further analyzed. (A) Two groups of the peach quality index tree classification diagram. (B) NMDS analysis of the peach quality index. (C) Correlation analysis between different transcriptomes and fruit shape.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementarydocument.docx