

A draft genome at chromosome level and metabolomes of leave, root and flowers provide insights into the molecular basis of medicinal ingredients of loquat (*Eriobotrya japonica* (Thunb.) Lindl)

Yunsheng Wang (✉ wys3269@126.com)
Kaili University

Research Article

Keywords: Draft genome, evolution, loquat, medicinal ingredient, Rosaceae

Posted Date: January 19th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-150389/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Eriobotrya japonica, commonly called loquat, is a type of fruit and a famous traditional Chinese medicinal material. Here, a high-quality draft genome of the *E. japonica* cultivar, Big Five-pointed Star, containing 733.32 million bases (Mb) covering approximately 98% of estimated whole genome size (749.25 Mb) is reported. A total of 45,492 protein-coding genes were predicted. Meanwhile, metabolomes of flower, leaf and root of this loquat variety were also determined by UPLC-ESI-MS/M system. 577 metabolites were determined in total, including 98 phenolic acids, 95 flavonoids, 33 alkaloids, 28 terpenoids, one steroids. The accumulation difference of these metabolites among leaf, flower and root of loquat were also clarified. Based on KEGG annotation, genes related to the biosynthesis of medicinal ingredients including some flavonoids and terpenoids were identified. Overall, this study provides global fundamental molecular insights into the medical biology of *Eriobotrya japonica*.

Introduction

E. japonica (Maloideae: Rosaceae), commonly known as loquat, is a kind of evergreen fruit rich in nutrients^[1]. *E. japonica* was first domesticated during the Han dynasty in China 2000 years ago according to documentary records and archaeological relics^[2]. Today, *E. japonica* is planted in more than thirty countries including China, Japan, the United States, France, Italy, and Spain. The annual output exceeds 1.2 million tons worldwide^[3]. *E. japonica* is also an important medicinal plant. Its roots, leaves, and flowers have long been used as traditional Chinese medicine, and have good therapeutic efficacy for inflammation, diabetes, cancer, bacterial infection, aging, pain and allergy. Rich bioactive components, including phenols, flavonoids, terpenoids and polysaccharides, have been found in these tissues^[4-7].

Whole genome sequencing has been performed on important fruit-producing crops and ornamental plants in the Rosaceae family, such as *Malus domestica*^[8], *Prunus persica*^[9], *Pyrus bretschneideri*^[10], *Pyrus betuleafolia*^[11], *Fragaria vesca*^[12], *Prunus mume*^[13], *Prunus avium*^[14], *Prunus yedoensis*^[15], *Rubus occidentalis*^[16], *Fragaria × ananassa*^[17], and *Rosa multiflora*^[18]. The corresponding genome data has provided global genetic information associated with growth, development, ecological, adaptive, and horticulture traits. This has been invaluable for the breeding of new varieties, and to track genetic clues of complex genome evolution hidden in these species. Meanwhile, more and more medicinal plants have appeared on the list of genome sequenced species, which has greatly deepened the understanding of molecular mechanisms and genes related the biosynthesis of medicinal components^[19-24].

A draft genome of a white-fleshed *E. japonica* cultivar named ‘Seventh Star’ has been recently published^[25] however, most *E. japonica* cultivars are yellow/red-fleshed. As a fruit and medicinal plant, studies on the biosynthesis of medicinal ingredients of *E. japonica* at the genome level are still very limited, which

greatly hampers the full understanding and utilization of *E. japonica*. In this study, the whole genome of 'Big Five-pointed Star', a yellow/red-fleshed *E. japonica* cultivar, was sequenced, assembled and analyzed. Meanwhile, metabolomes of flower, leaf and root from loquat tree of this cultivar were also determined and analysis. The major aims were to offer another high-quality reference genome for further research and utilization, and to provide insight into the molecular mechanisms about the biosynthesis of medicinal ingredients in *E. japonica*.

Results

Sequencing and assembly of a high-quality draft genome

An individual of Big Five-pointed Star was selected as sequencing material. Approximately 688.18 million clean short-reads with total 51.54 Gb data was first generated using the Illumina Hiseq 4000 sequencing platform (Table S1). Using these data, a K-mer analysis was performed, and the genome size of Big Five-point Star was evaluated to be 749.25 Mb (Table S2; Fig. 1A), which is almost identical to the result (749 Mb) determined using flow cytometry [26]. The heterozygosity and GC content were evaluated to be 0.31% and 38.58%, respectively (Table S2). Whole genome sequencing was then performed using PacBio long-read sequencing technologies, and more than six million clean subreads were obtained with an average length of 6,121 bp (N50 = 11,469 bp). A total of 36.90 Gb was obtained (Table S4, Fig.S1). With these clean subreads, an initial draft genome composed of 3,677 contigs with 733.32 Mb non-redundant sequences was assembled (Table S4) covering approximately 97.87% of the estimated whole genome size. Three measures were used to evaluate the assembly completeness of the initial draft genome. The first was a screening of 458 core eukaryotic genes and 248 conserved sequence datasets in the CEGMA database [27], which identified 447 (97.06%) and 238 (95.97%) matches, respectively (Table S5). The second was by querying the BUSCO database, which contains 1,440 plant-specific orthologous genes [28]. A total of 1,359 (94.38%) genes, with 921 single and 438 duplicate complete BUSCO genes were identified. The number of missing BUSCO genes was only 66 (4.38%) (Table S5). The third was to map the short-read data onto the draft genome. It was found that 93.81% of the draft genome could be aligned (Table S5). The above results suggested that the initial draft genome had good assembly completeness. Further corrections to the initial assembly and locations of the contigs on the chromosomes were performed using Hi-C technology with approximately 96 Gb of data from about 321.2 million reads generated by the Illumina Hiseq 4000 sequencing platform (Table S6). Among these, approximately 159.7 million read pairs were uniquely mapped on the initial draft genome, and more than 74.6 million read pairs were valid interactions (Table S6). These read pairs were used to scaffold the contigs onto chromosomes, and the number of contigs were finally corrected to 3,938, of which 3,725 with 727.40 Mb covering 99.19% of the draft genome sequences were anchored on chromosomes. The order and direction on the chromosomes of 2,181 contigs, which accounts for 644.88 Mb, could be determined. (Fig.S3; Table S7). These results indicate that the final assembled draft genome has good integrity, and was sufficient as a reference of the whole genome re-sequencing data, and for other molecular markers and genes and for genome utilization.

Genome elements annotation

Approximately 516.11 Mb of repetitive sequences were identified in the *E. japonica* draft genome, accounting for 70.38% of all sequences (Table S8). Of these repetitive sequences, beside 1.22 Mb (0.23%) potential host gene sequence, other are almost transposable element sequences, including 423.60 Mb RNA retrotransposons (Class I) and 113.40 Mb DNA transposons (Class II) (Table S8). Copia and Gypsy types of long terminal repetition are major two RNA retrotransposons, containing 158.90 Mb (30.79%) and 204.74 Mb (39.67%) of total repetitive sequences respectively (Table S8). TIR type accounted for most percentage of DNA transposons repetition sequences, which reached to 88.28 Mb (17.10%) of total repetitive sequences (Table S9).

By integrating the methods of de novo predicting, homologous species predicting and transcriptome predicting, to predict protein-coding genes from non-repetitive regions of the draft genome, a total of 45,492 protein-coding genes with an average length of 3,420 bp and average exon length of 1,532 bp were identified (Table S9-10). Of these, 45,090 (99.12%) could be annotated (Table S11; Data file 1). In addition, 10,426 rRNA genes belonging to 4 different families, 165 microRNA genes belonging to 25 families, 691 tRNA genes belonging to 24 families, 197 snRNA, 1023 snoRNA, and 8,314 pseudogenes were also identified from the final assembly draft genome (Table S12; Data file 2). The distribution pattern of protein-coding genes and RNA genes on the chromosomes was very uneven (fig.2)

Metabolites profiles in flower, leaf and root of loquat

The results of principal component analysis (PCA) and Pearson's correlation analysis based on the content and kind of all metabolites of nine samples showed that samples from the same organ gather more closely and higher correlation (Fig. 3A, 3B). Suggesting that mass spectrometry datum are reliable. Total 577 metabolites including 193 phenols, 33 alkaloids, 28 terpenoids and one steroids, and 322 other kinds metabolites were found in leaf, root and flower of loquat. Relative, more metabolites (573) in flowers were detected, the next (565) were in leafs, and the least (509) were in roots (Fig. 4; Data file 3, 4). Compared the 536 metabolites including 60 organic acids in fruit [29], more metabolites with less organic acids were found in leaf and flower of loquat. Compared by root, the content of 89, 178 and 310 metabolites in flower showed significant up-regulated, down-upregulated and unchanged respectively (Fig. 5A; Data file 5). Corresponding numbers are 76, 94 and 407 when flower compared by leaf (Fig. 5B; Data file 6), and 189, 84 and 304 when leaf compared by root respectively (Fig. 5C; Data file 7). showing that metabolites accumulated in leaf and flower appeared less difference than which in root and flower, and root and leaf. There are 51 metabolites showed a significant difference of content among leaf, flower and root, but 192 metabolites showed no significant change of content among them (Fig. 5D; Data file 8).

Phenolic acids in leaf, root and flower of loquat

Phenolic compounds are aromatic secondary metabolites in plants, which can be clustered into different families such as phenolic acids, flavonoids, lignans, coumarins and tannins. The phenolic compounds in plant have been proved to be great potential for human health including delaying aging, and reducing oxidative stress risk and inflammation which related with chronic diseases^[30, 31]. Previous studies showed that phenolics including phenolic acids and flavonoids are abundant in tissues of loquat tree^[32]. Here, 98 in total and 95, 96, 85 phenolic acids separately were detected in flower, leaf and root, including methyl chlorogenic acid, chlorogenic acid and caffeic acid (Fig. 4). Of 98 Phenolic acids, nine including methyl caffeate and menzoylferuloyltartaric acid were found to be difference accumulation among flower, leaf and root of loquat, and 46 including p-Coumaric acid, vanillin, 6-O-Caffeoylarbutin, caffeic acid, dibutyl phthalate, isosalicylic acid O-glycoside showing no significant difference accumulation, and the next four show high content in them (Data file 8). Chlorogenic acid, caffeic acid and their analogues and derivatives are main compound of phenolic acids in plant, and play a significant role in human health. Previous research suggested the 3- and 5-caffeoylquinic acids were two main bioactive ingredients in loquat fruit^[33]. Here, 6-O-Caffeoylarbutin, caffeic acid, chlorogenic acid, caffeic aldehyde, 3,4-Dicaffeoylquinic acid, O-Caffeoyl maltotriose, dicaffeoylquinic acid-glucoside, syringoylcaffeoylquinic acid O-glucose, 4-O-(6'-O-Glucosylcaffeoylglucosylferuloyl)-4-hydroxybenzyl alcohol, neochlorogenic acid (5-O-Caffeoylquinic acid), chlorogenic acid methyl ester were found in leaf, flower and root of loquat, except 4-O-(6'-O-Glucosylcaffeoylglucosylferuloyl)-4-hydroxybenzyl alcohol is lack in roots. 6-O-Caffeoylarbutin, caffeic acid, 5-O-Caffeoylshikimic acid, chlorogenic acid and neochlorogenic acid(5-O-Caffeoylquinic acid) appeared high content in leaf, flower and root of loquat (Data file 3, 8). These results partly provided reasonable explanation why flower and root of loquat, as well as leaf, were used as medicinal materials in the folk of China, and all emerged a good curative effect.

Flavonoids are an important kind of phenolics. Quercetin-type flavonols (primarily as quercetin glycosides), the most abundant of the flavonoid molecules distributed in plants, have a wide range of biological actions for human health^[34]. Up to now, at least sixteen flavonoids including quercetin, isoquercitrin, rutin, hyperoside, quercitrin, have been found and isolated in loquat^[35]. In this study, total 81 flavonoids including 2 anthocyanins, 1 biflavones, 2 chalcones, 5 dihydroflavone, 3 dihydroflavonol, 10 flavanols, 15 flavonoid, 3 flavonoid carbonoside, 38 flavonols and 2 isoflavones were detected (Fig. 4; Data file 8). Most of these flavonoids are new reports. Among them, quercetin-3-O-(2-O-rhamnosyl)-galactoside, quercetin-3-O-neohesperidoside, quercetin-3-O-xylosyl(1→2)-galactoside, quercetin-3-O-glucoside(Isoquercitrin) were accumulated with significant difference content among leaf, flower and root of loquat, and dihydroquercetin(Taxifolin), quercetin-3-O-(6"-trans-p-Coumaroyl)-glucoside, quercetin-5-O-glucuronide, quercetin-3-O-(2"-acetyl)-glucuronide acculmated in leaf, flower and root with no significant different (Data file 8).

Terpenoids in leaf, root and flower of loquat

Beside phenolic compounds, terpenoids are also a main class of active ingredients in loquat leaf. Those, especial ursolic acid, endow loquat with strong anti-inflammatory activity and antitussive effect [36, 37]. At least fourteen triterpene acids have been isolated from the loquat leaf and showed marked anti-inflammatory effects [38, 39]. Among them, ursolic acid and oleanolic acid have been well studied and were proved to be strong bioactives for great potential healthful benefits [40, 41]. Here, 28 terpenoids (2 sesquiterpenoids, 25 triterpene and one triterpene saponin) were detected in leaf, flower and root of loquat, including oleanonic acid, ursolic acid and their four derivatives (2-Hydroxyoleanolic acid, 2,3-Dihydroxy 5(6),12(13)diene ursolic acid, 27,28-Dicarboxyl ursolic acid, Ursolic acid-OCH₃) (Fig. 4; Data file 3,4). δ -Amyrenone, Ursolic acid-OCH₃, Medicagenic acid 3-O-GlcA-28-O-Rha(1,2)-Ara showed a significant accumulation differences, and betulinic acid, ursonic acid, betulonic acid, oleanonic acid, maslinic acid, 24,30-Dihydroxy-12(13)-enolupinol, 2-Hydroxyoleanolic acid showed no significant accumulation differences in leaf, flower and roots. Some metabolites such as ligupleurol geniposide, β -Amyrenone were firstly reported in loquat. ursonic acid, pomolic acid, asiatic acid and caffeoyl hawthorn acid showed relatively high-content in loquat (Fig. 4; Data file 8). The result offered another evidence that leaf, flower and root of loquat were all used as medicinal materials in the folk of China, and all emerged a good curative effect.

Alkaloids in leaf, root and flower of loquat

Alkaloids are nitrogen-containing natural bioactive compounds in plant, which also possessed a wide range of biological and pharmacological benefits (Rasouli et al., 2020) [42]. There were still no reports about alkaloids in loquat. In this article, total 33 alkaloids (19 alkaloids, 7 phenolamine, 6 plumerane and 1 pyrrole alkaloids), 33 in flower, 30 in leaf, 28 in root of loquat respectively were found (Fig. 4; Data file 3,4). Among them, the bioactive compounds such as betaine, stachydrine, N,N'-Bis(Sinapoyl)Spermidine, N',N'',N'''-p-Coumaroyl-cinnamoyl-caffeoyl, spermidine and choline have been proved to be good medicinal effect for human [43-45]. Stachydrine and 3-((2-Aminoethoxy)(hydroxy)phosphoryl)oxy-2-hydroxypropyl palmitate showed a significant accumulation differences, and spermine, choline, cocamidopropyl betaine, nicotine-hex, tryptamine, indole-3-carboxaldehyde, indole-3-carboxylic acid showed no significant accumulation differences in leaf, flower and root of loquat. choline, betaine, n-Benzylmethylene isomethylamine, 3-((2-Aminoethoxy)(hydroxy)phosphoryl)oxy-2-hydroxypropyl palmitate, spermine and stachydrine showed a relatively high content in all three organs of loquat (Data file 8).

Other metabolites in leaf, flower and root of loquat

Except the phenolic compounds, terpenoid, alkaloids and so on secondary metabolites, polysaccharide also plays an important role in medicinal application of loquat [46]. Study showed that the contents of

polysaccharides in different part of tree were significantly different among difference loquat cultivars [47]. In this study, 16 monosaccharides (D-(-)-Threose, D-(-)-Arabinose, D-Arabitol, D-Glucose, sedoheptulose, N-Acetyl-D-glucosamine, 5-O-Feruloyl-L-arabinose, melibiose, D-(+)-Sucrose, D-(+)-Trehalose, isomaltulose, turanose, solatriose, raffinose, D-(+)-Melezitose, Panose) and two glycoside (galactinol, D(+)-Melezitose O-rhamnoside) were found in all three organs of loquat (Data file 3,4). All of them were accumulated in all three organs of loquat with significant difference. 5-O-Feruloyl-L-arabinose, Solatriose, Isomaltulose, D(+)-Melezitose O-rhamnoside, D-(+)-Melezitose, D-(+)-Sucrose, N-Acetyl-D-glucosamine were steadily accumulated in all leaf, flower and root of loquat with no significant difference content. Moreover, D-Glucose, D-(+)-Trehalose, Galactinol*, Isomaltulose, and Turanose accumulated in all three organs of loquat with higher content (Data file 3,4).

Identification of genes responsible for metabolism of medicinal ingredients using KEEG annotation

Previous studies suggested that the main components that endow *E. japonica* with medicinal value are flavonoids, terpenoids and polysaccharides [32-34]. Among 577 metabolites identified, 271 could be annotated by KEGG database (Data file 4, 9), but only 110 including 15 phenolic acids, six flavonoids and 12 alkaloids could be assigned into specific pathway such as 'Metabolic pathways' (ko01100), 'Biosynthesis of secondary metabolites' (ko01110), 'Phenylpropanoid biosynthesis' (ko00940), 'Flavonoid biosynthesis' (ko00941), 'Stilbenoid, diarylheptanoid and gingerol biosynthesis' (ko00945), 'Caffeine metabolism', 'Isoflavonoid biosynthesis' (ko00950), etc (Data file 10). According to the KEGG annotation, 71 genes responsible for the biosynthesis of flavonoids (Fig S2; Data file 10), and 92, 32, 56, and 37 genes involved into the biosynthesis pathways of terpenoid backbones, monoterpenoids, diterpenoids, sesquiterpenoid-triterpenoids were identified respectively (Fig. S3-6; Data file 10). Quercetin is an important flavonoid which has been shown to modify eicosanoid biosynthesis (anti-prostanoid and anti-inflammatory responses) and other therapeutic functions [48]. Which are abundant and almost exist with the form of quercetin glycosides in loquat as mentioned above. Genes encoding key enzymes in the pathway of biosynthesis of quercetin, including three genes (EVM0007289.1; EVM0018354.1; EVM0040197.1) coding flavonoid 3'-monooxygenase-like [1.14.13.21] enzymes were identified in loquat genome (Data file 10). Chlorogenic acid is one of phenolic acids rich in leaf, flower and root of loquat, which play key role in Phenylpropanoid biosynthesis (ko00940), Flavonoid biosynthesis (ko00941), Stilbenoid, diarylheptanoid and gingerol biosynthesis (ko00945). Shikimate O-hydroxycinnamoyltransferase [EC:2.3.1.133] is a key enzyme for biosynthesis of chlorogenic acid, And 15 putative genes (EVM0001669.1, EVM0005803.1, EVM0006222.1, EVM0006273.1, EVM0013101.1, EVM0014121.1, EVM0016248.1, EVM0019449.1, EVM0022718.1, EVM0025581.1, EVM0026375.1, EVM0026790.1, EVM0042804.1, EVM0044744.1 and EVM0044979.1) encoded this enzyme were detected in loquat (Data file 10). Uridine diphosphate (UDP) glycosyltransferases transfer glycosyl residues from activated nucleotide sugars to acceptor molecules (aglycones), and play a key role in regulating the solubility of the acceptors within cells and throughout the organism [49]. In this study, several UDP-glucuronosyl coding genes involved in the biological metabolism pathways of

polysaccharides, including five genes (EVM0001836.3; EVM0024320.2; EVM0026146.1; EVM0036105.1; EVM0038009.1) encoding UDP-N-acetylglucosamine transferase subunit ALG14-like enzymes [EC:2.4.1.141] and two genes (EVM0001695.1; EVM0001956.1) encoding UDP-N-acetylglucosaminephosphotransferase-like enzymes [EC:2.7.8.15] were found (Data file 10). These results offer a possible molecular clue to explain why water soluble polysaccharides in *E. japonica* are high and could play an important role in its healthful effects.

Discussion

E. japonica blooms in late autumn/winter, and ripens in spring/early summer. At this time, most fresh fruits are not on the market in the Northern Hemisphere, so the fresh fruit of *japonica* plays an important role in filling the gap of the fruit basket, greatly adding to its commodity value. In recent years, *E. japonica* has increasingly becoming an important fruit worldwide. However, due to the lack of information on its genome, studies on the genetics and molecular biology of *E. japonica* are limited. This has greatly impeded understanding of growth, development, and varieties breeding. Recently, a draft genome of *E. japonica* cultivar Seventh Star has been published [25]. Here, a new high-quality draft genome of the *E. japonica* cultivar, Big Five-pointed Star, is described. Big Five-pointed Star is a cultivar with yellow/red fresh and has the largest planting area in China [50]. In contrast, Seventh Star is a mutant with white flesh and was bred recently. Results of the assembly and annotation of these two draft genomes were not exactly the same. For example, the estimated genome size of Seventh Star is ~710.83 Mb, while Big Five-pointed Star is ~749.25 Mb. The number of predicted coding genes in Seventh Star is 45,743, while the number in Big Five-pointed Star is 45,450. This new *E. japonica* draft genome from Big Five-pointed Star will provide a more solid foundation and more choices about the reference genome for further *E. japonica* studies in molecular biology, genetics, and breeding.

More than fourth metabolites including ursolic acid, ursolic acid methyl ester, acetyl ursolic acid, and oleanolic acid, chlorogenic acid, neochlorogenic acid, caffeic acid with potential medicinal value have been detected in different organs of loquat by classical instrumental analyses [51, 52]. However, Traditional phytochemistry methods are very time-consuming, and labor intensive and low-throughput. So far, there is still a limited of global determination about the metabolites including phenols, flavonoids, terpenes and so on secondary metabolites which have potential health effects of loquat tissues or organs. And this is not conducive to the further research and utilization of medicinal value of loquat. In recent years, traditional Chinese medicine (TCM) has been greatly accelerated by the analytical technologies and methodologies of genomics, proteomics and metabonomics [53, 54]. Among these Omics, metabolomics is a very valuable for the analysis of components including various metabolites of TCM [55]. Here, the metabolomes of leaf, flower and root of loquat were determined using a widely targeted metabolomic analysis based on the based on the liquid chromatography and series mass spectrometry (LC-MS/MS) which firstly offered by Chen et al [56], and many new metabolites including those belong to phenols,

terpenoids and alkaloids were detected in loquat, and this would greatly deepen the knowledge of the biochemical substances in loquat that produce medicinal efficacy. Meanwhile, lots of genes encoding the enzymes those related to the biosynthesis of metabolites including phenols, terpenoids and alkaloids, and this laid a good foundation for further study on molecular pharmacology of loquat.

Conclusion

In this study, a high-quality genome of loquat with yellow-fleshed were assembled, and three high-throughput metabolomes of leaf, flower and root of loquat were determined. 45490 putative protein-encoding genes and 577 metabolites were detected. Among them, some metabolite belong to phenols, terpenoids and alkaloids and other kind with potential healthful value and genes related to the biosynthese of these metabolites were identified and analyzed. Overall, this study provides global fundamental molecular insights into the nutrient and medical biology of *Eriobotrya japonica*.

Material And Methods

Brief introduction of sequencing objective: loquat cultivar, Big Five-pointed Star

Big Five-pointed Star, also called Dawuxing, shows excellent characteristics, including high average single fruit weight, edible rate, and soluble solids content. The peel and flesh are easily stripped, and has a sweet taste with juicy, soft, and delicate meat. Big Five-pointed Star has become a popular cultivar with the most rapid development and widest planting area in China ^[50].

DNA extraction

Total DNA were extracted from young leaves of Big Five-pointed Star using a modified protocol based on the CTAB method ^[57]. RNAase was then used to remove RNA contamination from the total DNA in a 37 °C water bathfor 1 h. The quality of and content of extracted total genomic DNA was checked using agarose gel electrophoresis and spectrophotometry (Nanodrop WND-1000, Nano-Drop Technologies Inc, Delaware, USA).

Illumina short-read library construction, sequencing and raw data statistics

The Illumina paired-end reads library (350 bp) was constructed strictly according to the guideline of sequencer manufacturer using the following steps: Qualified total DNA was fragmented into small segment; segments approximately 350 bp in length were selected on a 3% agarose gel for further analysis; end repair and A-tailing was performed, and Illumina compatible adaptors were added to the selected DNA fragments; PCR-amplification was executed using Illumina adapter-specific primers, and the paired-end sequencing library was then finished. The quality and quantity of the sequencing library was

performed to sequence by using 150 base-length read v3 chemistry in paired-end flow cell on the Illumina HiSeq 2000 (Illumina, San Diego, CA, USA).

Survey of Big Five-pointed Star based on short-reads data by K-mer analysis

A K-mer analysis was performed using the KAT program ^[58] to make an initial estimate based on C-value, heterozygosity, and repetitive rate of the genome of Big Five-pointed Star according to the following formula: genome size = (total nucleotides number)/(average sequencing depth) = (total number of K-mer)/(average K-mer depth). The K value used the maximum number of odd numbers that met the following formula: $4^K/\text{genome} > 200$.

Pacbio long-read library construction, sequencing, and raw data statistics

The long-read sequencing library (20 kb) was constructed strictly according to the guidelines of the Pacbio sequencer (Menlo Park, CA, USA)) using the following steps. The genomic-DNA of Big Five-pointed Star was fragmented by g-TUBE; Performed a damage repairing for fragmented DNA; Executed end-repairing for the broken DNA; Ligated the broken DNA with the dumbbell-shaped adaptor; Digested the DNA segment by using exonuclease; Selected the target segment by using BluePippin. Finally, the sequencing library was successfully constructed.

Assembly and integrity evaluation of draft genome

To assembly the loquat draft genome using the long-read sequencing data, subreads with low-quality (<Q20) and short length (<500) were filtered, and the remaining subreads were corrected by Canu software ^[59]. The corrected data were then assembled into a draft genome sequence by WTDBG ^[60], Falcon ^[61], and Canu software, respectively. Then, the three assembled results were optimized using the Quickmerge ideology ^[62]. Finally, the assembly was further improved by correcting errors joining the short-read data using Pilon software ^[63].

Three methods were used to evaluate completeness of draft genome. The first was by blasting the assembled draft genome with a standard of more than 70% identity against the Core Eukaryotic Genes Mapping Approach database (CEGMA) ^[27], and included 458 core eukaryotic genes (CEGs) and 248 highly conserved core eukaryotic genes. The second method was by blasting the assembled draft genome with at least 70% identity against the embryophyta_odb9 dataset in the BUSCO v2.0 database

[28], which includes 1,440 conserved core genes of plant. Finally, the short-reads sequencing data by Illumina technology were mapped to the assembly draft genome using BWA software [64].

Hi-C sequencing library construction

A Hi-C sequencing library was constructed according to standard protocols described by Servant et al (2015) [65]. Briefly, the cells of young leaves were fixed with formaldehyde. The Dissociated the fixed cell and cut the cross-linking product with restricted endonuclease enzyme to produce viscous ends. The biotin marker was introduced into the viscous ends and repaired to produce blunt ends; Ligated the blunt end; Unlashed the cross-linking to separate DNA from protein, and protein from protein; Extracted the DNA; Used the Covaris interruptor to interrupt DNA to the right size then repaired the end; Purified and recycled the interrupt DNA segment by gel electrophoresis; Removed those DNA segment without biotin marker; Added A to the remaining DNA segment including biotin marker; Added PCR adaptor; PCR; Purified and recycled PCR produce by gel electrophoresis. the Hi-C sequencing library was finally finished.

Hi-C sequencing, data summary and estimation

The Illumina HiSeq 4000 (Illumina) sequencing platform was used for paired sequencing by synthesis. The paired reads were mapped on the assembled loquat genome draft by using the BWA program (version: 0.7.10-r789; aln model; other parameters were set to default) [64].

Hi-C assembly and quality estimation

LACHESIS software [65] was used to scaffold contigs onto the chromosomes with the following parameters: CLUSTER_MIN_RE_SITES = 53; CLUSTER_MAX_LINK_DENSITY = 2; CLUSTER_NONINFORMATIVE_RATIO = 2; ORDER_MIN_N_RES_IN_TRUN = 21; ORDER_MIN_N_RES_IN_SHREDS = 22.

Repetitive sequence predicting and annotation

To annotate the genomic sequence, a special database for identifying repetitive sequences in the loquat genome was constructed with the help of LTR FINDER v1.05 [66], MITE-Hunter [67], RepeatScout v1.0.5 [68] and PILER-DF v2.4 software [69] based on the principles of structure prediction and de novo prediction. This special database was then merged with the REPBASE database [70] as the final repetitive sequence database. The PASTEClassifier software [71] was used to classify the database. Finally, the Repeatmasker

V4.0.6 software ^[72] was used to predict repetitive sequences based on a well-constructed repeating sequence database.

Protein-coding gene prediction and functional annotation

Protein-coding genes based on non-repetitive sequences of the draft genome were predicted using three methods: De novo predicting (Ab initio) by Genscan software ^[73], Augustus v2.4 ^[674], GlimmerHMM v3.0.4 ^[75], GeneID v1.4 ^[76], SNAP ^[77], homologous species predicting by software GeMoMa v1.3.1 ^[78] (based on the gene sequences of *A. thaliana*, *Oryza sativa Japonica*, *Malus domestica*, *Pyrus bretschneideri*, and *Fragaria vesca* from gene and expression databases in NCBI (<https://www.ncbi.nlm.nih.gov/guide/genes-expression/>)). The method of prediction used TransDecoder v2.0.1 software (<http://transdecoder.github.io/> [last accessed May 8, 2019]) and GeneMarkS-T v5.1 ^[79] based on loquat unigenes assembled using Hisat v2.0.4 ^[80] and StringTie v1.2.3 ^[81] software from the transcriptome data with a reference transcript based on loquat expressed sequence tags collected from the NCBI dbEST database (<http://www.ncbi.nlm.nih.gov/dbEST/>), and used PASA v2.4.1 software (<https://github.com/PASAPipeline/PASAPipeline/releases/tag/pasa-v2.4.1>) based on these loquat unigenes from the transcriptome data with non-referenced transcripts.

Results from above three methods were integrated using EVM v1.1.1 software (<https://github.com/EvidenceModeler/EvidenceModeler/releases/tag/v1.1.1>). The gene sequence in gff3 format has been deposited into the China National Center for Bioinformation database (<https://bigd.big.ac.cn/gwh>: Submission ID: WGS010381).

Five gene databases including Nr (<https://www.ncbi.nlm.nih.gov/refseq/>), TrEMBL (<https://www.uniprot.org/statistics/TrEMBL>), GO (<http://geneontology.org/>), KOG (<http://www.ncbi.nlm.nih.gov/COG/>) and KEGG (<https://www.genome.jp/kegg/>) were used to annotate the molecular function of predicted protein-coding gene using BLAST v2.10.0 software (<https://www.ncbi.nlm.nih.gov/books/NBK131777/>.) with a threshold e-value of (1e-5).

RNA gene prediction and annotation

Infenal 1.1 software ^[82] was used to predict rRNA genes based on Rfam (<https://rfam.xfam.org/>), and MicroRNA based on the miRBase database (<http://www.mirbase.org/>). The program tRNAscan-SE v1.3.1 ^[83] was used to identify tRNA loci. The GenBlastA v1.0.4 software ^[84] was used to search homologous gene sequences by blasting the remaining draft genome regions after shielding the prediction gene

sequences. Pseudogenes were identified by searching immature termination codons and shift code mutations using GeneWise v2.4.1 software^[85].

Sampling, sample preparation and extraction for metabolites

Nine samples including nine samples including three from leaf, three from flower and three from root of Big Five-pointed Star at different developmental stages were collected for metabolome analysis. The freeze-dried sample were crushed using a mixer mill (MM 400, Retsch) with a zirconia bead for 1.5 min at 30 Hz. 100mg powder was weighted and extracted overnight at 4 with 0.6 ml 70% aqueous methanol. Following centrifugation at 10,000g for 10 min, the extracts were absorbed (CNWBOND Carbon-GCB SPE Cartridge, 250mg, 3ml; ANPEL, Shanghai, China, www.anpel.com.cn/cnw) and filtrated (SCAA-104, 0.22µm pore size; ANPEL, Shanghai, China, <http://www.anpel.com.cn/>) before UPLC-MS/MS analysis.

UPLC Conditions

The sample extracts were analyzed using an UPLC-ESI-MS/MS system (UPLC, Shim-pack UFLC SHIMADZU CBM30A system, www.shimadzu.com.cn/; MS, Applied Biosystems 4500 Q TRAP, www.appliedbiosystems.com.cn/). The analytical conditions were as follows, UPLC: column, Agilent SB-C18 (1.8 µm, 2.1 mm*100 mm); The mobile phase was consisted of solvent A, pure water with 0.1% formic acid, and solvent B, acetonitrile. Sample measurements were performed with a gradient program that employed the starting conditions of 95% A, 5 % B. Within 9 min, a linear gradient to 5% A, 95% B was programmed, and a composition of 5% A, 95% B was kept for 1min. Subsequently, a composition of 95% A, 5.0 % B was adjusted within 1.10 min and kept for 2.9 min. The column oven was set 1 to 40°C; The injection volume was 4µl. The effluent was alternatively connected to an ESI-triple quadrupolelinear ion trap (QTRAP)-MS.

ESI-Q TRAP-MS/MS

LIT and triple quadrupole (QQQ) scans were acquired on a triple quadrupole-linear ion trap mass spectrometer (Q TRAP), API 4500 Q TRAP UPLC/MS/MS System, equipped with an ESI Turbo Ion-Spray interface, operating in positive and negative ion mode and controlled by Analyst 1.6.3 software (<https://sciex.com/products/software/analyst-software>; AB Sciex). The ESI source operation parameters were as follows: ion source, turbo spray; source temperature 550; ion spray voltage (IS) 5500 V (positive ion mode)/-4500 V (negative ion mode); ion source gas I (GSI), gas II(GSII), curtain gas (CUR) were set at 50, 60, and 30.0 psi, respectively; the collision gas(CAD) was high. Instrument tuning and mass calibration were performed with 10 and 100 µmol/L polypropylene glycol solutions in QQQ and LIT

modes, respectively. QQQ scans were acquired as MRM experiments with collision gas (nitrogen) set to 5 psi. DP and CE for individual MRM transitions were done with further DP and CE optimization. A specific set of MRM transitions were monitored for each period according to the metabolites eluted within this period [56].

Qualitative and quantitative analysis of metabolites

Based on mwdb (metaware database) built by Maiwei company, the qualitative analysis is carried out according to the secondary spectrum information, and the repeated signals including K^+ ion, Na^+ ion and NH_4^+ ion, as well as the repeated signal of fragment ion which itself is larger molecular weight substance are removed. Metabolite quantification was performed by multiple reactions monitoring mode analysis using triple quadrupole mass spectrometry [86].

Data analysis

The mass spectrum data were processed by Software Analyst 1.6.3. Samples from same organ were regarded as repetition. Principal component analysis (PCA), Pearson's correlation analysis, differential expression analysis, and drawing of heat map were performed using statistical module of R software package (<https://www.r-project.org/>). The Kyoto Encyclopedia of genes and genomes database [87] was used to annotate and display the biosynthetic pathways of different metabolites.

Declarations

Data availability:

Draft genome sequence with fasta format and protein-coding genes sequence with gff format were deposited in China National Center for Bioinformation <https://bigd.big.ac.cn/gsub/>: accessible ID: GWHAOTB00000000. The reviewer can log in with username: wys3269@126.com code: wys123456 to check it.

Acknowledgments

This work was funded by National Natural Science Foundation of China (31560091) and Science and technology plan project of Guizhou Province [2019]4318. I sincerely thank the experimental personnel and bioinformatics analysts in Biomarker company (www.biomarker.com.cn) who participated in this project.

Author Contributions

Y.S.W designed the experiments, collected samples, analyzed data and wrote the paper

Conflict of Interest

The author declare no conflict of interest.

References

- [1] Li X, Xu C, Chen K. 2016. Chapter 16 - Nutritional and Composition of Fruit Cultivars: Loquat (*Eriobotrya japonica* Lindl.). Nutritional Composition of Fruit Cultivars. 2016: 371–394.
- [2] Lin S, Sharpe RH, Janick J. 1999. Loquat: Botany and horticulture. Hort. Rev. 23: 233–276.
- [3] Badenes ML, Blasco M, Naval MM. 2015. Loquat: Progress and Expectations. Proc. IV International Symposium on Loquat. Acta Hort. 1092, ISHS, 19–24.
- [4] Tommasi ND, Shimone FD, Pizza C. 1992. Constituents of *Eriobotrya japonica* : a study of their antiviral properties. J. Nat. Prod. 55: 1067–1073.
- [5] Lu SM, Wang WH, Zheng MY. 2013. Research progress on the bioactive components and pharmacological effects of *Eriobotrya japonica*. Curr. Biotechnol. 3: 453–458.
- [6] Ramírez-Gómez Xóchitl Sofía, Monroy-Torres Rebeca and Linares-Segovia Benigno. 2014. Anti-inflammatory and antitumor properties of *Eriobotrya japonica* Lindl: Mini-review. Immun. Endoc. & Metab. Agents in Med. Chem. 14: 15–20
- [7] Liu Y, Zhang W, Xu C, Li X. 2016. Biological activities of extracts from Loquat (*Eriobotrya japonica* Lindl.): A review. Int. J. Mol. Sci. 17: 1983.
- [8] Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, et al. 2010. The genome of the domesticated apple (*Malus × domestica* Borkh.). Nat. Genet. 42: 833–839.
- [9] Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, Zhebentyayeva T, Dettori MT, Grimwood J, Cattonaro F, et al. 2013. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat. Genet. 45: 487–494.
- [10] Wu J, Wang Z, Shi Z, Zhang S, Ming R, Zhu S, Khan MA, Tao S, Korban SS, Wang H, et al. 2013. The genome of the pear (*Pyrus bretschneideri* Rehd.). Genome Res. 23: 396–408.

- [11] Dong X, Wang Z, Tian L, Zhang Y, Qi D, Huo H, Xu J, Li Z, Liao R, Shi M, et al. 2020. De novo assembly of a wild pear (*Pyrus betuleafolia*) genome. *Plant Biotechnol. J.* 18: 581–595.
- [12] Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP et al. 2011. The genome of woodland strawberry (*Fragaria vesca*). *Nat. Genet.* 43: 109–116.
- [13] Zhang Q, Chen W, Sun L, Sun L, Zhao F, Huang B, Yang W, Tao Y, Wang J, Yuan Z, Fan G, et al. 2012. The genome of *Prunus mume*. *Nat. Commun.* 3: 1318
- [14] Shirasawa K, Isuzugawa K, Ikenaga M. 2017. The genome sequence of sweet cherry (*Prunus avium*) for use in genomics-assisted breeding. *DNA Res.* 24: 499–508.
- [15] Baek S, Choi K, Kim GB, Yu HJ, Cho A, Jang H, Hyuk-Jin C, Sun K, Joo-Hwan C, Munet JH. 2018. Draft genome sequence of wild *Prunus yedoensis* reveals massive inter-specific hybridization between sympatric flowering cherries. *Genome boil.* 19: 127.
- [16] VanBuren R, Bryant D, Bushakra JM, Vining KJ, Edger PP, Rowley ER, Priest HD, Michael TP, Lyons E, Filichkin SA, Dossett M, et al. 2016. The genome of black raspberry (*Rubus occidentalis*). *Plant J.* 87: 535–547.
- [17] Edger PP, Poorten TJ, VanBuren R, Hardigan MA, Colle M, McKain MR, Smith RD, Teresi SJ, Nelson ADL, Wai CM, et al. 2019. Origin and evolution of the octoploid strawberry genome. *Nat. Genet.* 51: 541–547.
- [18] Raymond O, Gouzy J, Just J, Verdenaud M, Lemainque A, Vergne P, Moja S, Choisne N, Pont C, et al. 2018. The Rosa genome provides new insights into the domestication of modern roses. *Nat. Genet.* 50: 772–777.
- [19] Yan L, Wang X, Liu H, Tian Y, Lian J, Yang R, Hao S, Wang X, Yang S, Li Q, et al. 2015. The Genome of *Dendrobium officinale* Illuminates the Biology of the Important Traditional Chinese Orchid Herb. *Mol. Plants* 8: 922–934.
- [20] Xu H, Song J, Luo H, Zhang Y, Li Q, Zhu Y, Xu J, Li Y, Song C, Wang B, et al. 2016. Analysis of the genome sequence of the medicinal plant *Salvia miltiorrhiza*. *Mol. Plant* 9: 949–952.
- [21] Urasaki N, Takagi H, Natsume S. 2017. Draft genome sequence of bitter melon (*Momordica charantia*), a vegetable and medicinal plant in tropical and subtropical regions. *DNA Res.* 24: 8.
- [22] Shen Q, Zhang L, Liao Z, Wang S, Yan T, Shi P, Liu M, Fu X, Pan Q, Wang Y, et al. 2018. The Genome of *Artemisia annua* provides insight into the evolution of Asteraceae family and artemisinin biosynthesis. *Mol. Plants.* 11: 776–788.

- [23] Hoopes GM, Hamilton JP, Kim J, Zhao D, Wiegert-Rininger K. 2018. Genome assembly and annotation of the medicinal plant *Calotropis gigantea*, a producer of anticancer and antimalarial cardenolides. *G3: GENES GENOM. GENET.* 8: 385–391.
- [24] Sun W, Leng L, Yin Q, Xu M, Huang M. 2019. The genome of the medicinal plant *Andrographis paniculata* provides insight into the biosynthesis of the bioactive diterpenoid neoandrographolide. *Plant J.* 97: 841–857.
- [25] Jiang S, An H, Xu F, Zhang X. 2020. Chromosome-level genome assembly and annotation of the loquat (*Eriobotrya japonica*) genome. *Gigascience.* 9(3): g1aa015.
- [26] Zhang Z, Wang Y, Lin S, Du K. 2012. Ploidy identification of loquats for genome sequencing project by flowcytometry. *J. Fruit Sci.* 29: 498–504.
- [27] Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics,* 23: 1061–1067.
- [28] Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 31: 3210.
- [29] Zou S, Wu J, Shahida MQ, He Y, Lin S, Liu Z, Yang X. 2020. Identification of key taste components in loquat using widely targeted metabolomics. *Food Chemistry* 323: 126822.
- [30] Sarawong C, Schoenlechner R, Sekiguchi K, Berghofer E, Ng PKW. 2014. Effect of extrusion cooking on the physicochemical properties, resistant starch, phenolic content and antioxidant capacities of green banana flour. *Food Chem.* 143:33–39.
- [31] Ambriz-Pérez DL, Leyva-López N, Gutierrez-Grijalva EP, Heredia JB. 2016. Phenolic compounds: Natural alternative in inflammation treatment. A Review. *Cogent Food And Agriculture.* 2(1): 1131412,
- [32] Zhou C, Sun C, Chen K, Li X. 2011. Flavonoids, phenolics, and antioxidant capacity in the flower of *Eriobotrya japonica* Lindl. *Int. J. Mol. Sci.* 12: 2935–2945.
- [33] Pande G, Akoh CC. 2010. Organic acids, antioxidant capacity, phenolic content and lipid characterisation of Georgia-grown underutilized fruit crops. *Food Chem.* 120: 1067–1075.
- [34] Li Y, Yao J, Han C, Yang J, Chaudhry MT, Wang S, Liu H, Yin Y. 2016. Quercetin, inflammation and immunity. *Nutrients* 8(3): 167.
- [35] Sagar NA, Pareek S, Bhardwaj R, and Vyas N. 2019. Bioactive compounds of loquat (*Eriobotrya japonica* (Thunb.) L. Reference Series in Phytochemistry, https://doi.org/10.1007/978-3-030-06120-3_10-1.

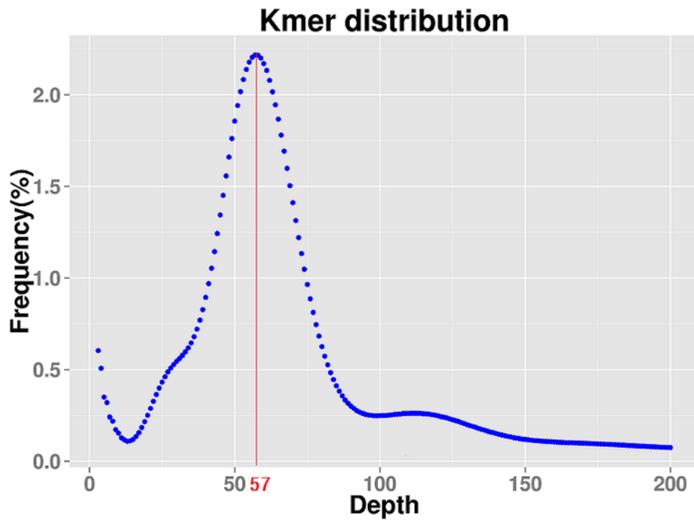
- [36] Cha DS, Eun JS, Hoon J. 2011. Anti-inflammatory and antinociceptive properties of the leaves of *Eriobotrya japonica* J. Ethnopharmacol. 134: 305–312.
- [37] Huang Y, Li J, Cao Q, Yu SC, Lv WX, Jin Y, Zhang L, Zou YH, Ge JF. 2006. Anti-oxidative effect of triterpene acids of *Eriobotrya japonica* (Thunb.) Lindl. leaf in chronic bronchitis rats. Life Sci. 78: 2749–2757.
- [38] Banno N, Akihisa T, Tokuda H, Yasukawa K, Taguchi Y, Akazawa H, Ukiya M, Kimura Y, Suzuki T, Nishino H. 2005. Anti-inflammatory and antitumor-promoting effects of the triterpene acids from the leaves of *Eriobotrya japonica*. Biol. Pharm. Bull. 28: 1995–1999.
- [39] Yang Y, Huang Y, Huang C, Lv XW, Liu L, Wang YY, Li J. 2012. Antifibrosis effects of triterpene acids of *Eriobotrya japonica* (Thunb.) Lindl. leaf in a rat model of bleomycin-induced pulmonary fibrosis. J. Pharm. Pharmacol. 64: 1751–1760.
- [40] Zhou C, Chen K, Sun C, Chen Q, Zhang W, Li X. 2007. Determination of oleanolic acid, ursolic acid and amygdalin in the flower of *Eriobotrya japonica* Lindl. by HPLC. Biomed Chromatogr. 21(7): 755–761.
- [41] Tan H, Zhao C, Zhu Q, Katakura Y, Tanaka H, Ohnuki K, Shimizu K. 2019. Ursolic Acid Isolated from the Leaves of Loquat (*Eriobotrya japonica*) Inhibited Osteoclast Differentiation through Targeting Exportin 5. J. Agric. Food Chem. 67(12): 3333–3340.
- [42] Rasouli H, Yarani R, Pociot F, Popović-Djordjević J. 2020. Anti-diabetic potential of plant alkaloids: Revisiting current findings and future perspectives. Pharmacological Research 155: 104723.
- [43] Alcalde E, Dinarès I, Pérez-García L. 1999. The betaine pool: molecular guests in medicinal chemistry and molecular hosts in supramolecular chemistry, Il Farmaco, 54(5): 297–308.
- [44] Yin J, Zhang ZW, Yu WJ, Liao JY, Luo XG, Shen YJ. 2010. Stachydrine, a major constituent of the Chinese herb *Leonurus heterophyllus* sweet, ameliorates human umbilical vein endothelial cells injury induced by anoxia-reoxygenation. Am J Chin Med. 38(1):157–171.
- [45] Corbin KD, Zeisel SH. 2012. Choline metabolism provides novel insights into nonalcoholic fatty liver disease and its progression, Curr. Opin. Gastroenterol. 28: 159.
- [46] Fu Y, Yuan Q, Lin S, Liu W, Du G, Zhao L, Zhang Q, Lin DR, Liu Y T, Qin W, Li D Q, Wu D T. 2019. Physicochemical characteristics and biological activities of polysaccharides from the leaves of different loquat (*Eriobotrya japonica*) cultivars. 135: 274–281.
- [47] Jiang F, Zheng SQ, Gao HY, Huang AP, Chen XP. 2007. Polysaccharides and flavones in loquat cultivars. Acta Hortic. 750: 321–324.
- [48] Formica JV, Regelson W. 1995. Review of the biology of quercetin and related bioflavonoids. Food Chem. Toxicol. 33: 1061–1080.

- [49] Ross J, Li Y, Lim E, Bowles DJ. 2001. Higher plant glycosyltransferases. *Genome Biol.* 2: 1–6.
- [50] Zhang ZZ, Fu TT, Li YL, Peng CH, Qin W. 2011. Effects of ozone treatment on the preservation of loquat cultivar big five-pointed star during storage. *Food Sci.* 32: 282–285.
- [51] Liu Y, Zhang W, Xu C, Li X. 2016. Biological activities of extracts from loquat (*Eriobotrya japonica* Lindl.): A review. *Int. J. Mol. Sci.* 17: 1983.
- [52] Li J, Ding X, Jian T, Lü H, Zhao L, Li J, Liu Y, Ren B, Chen J. 2020. Four sesquiterpene glycosides from loquat (*Eriobotrya japonica*) leaf ameliorates palmitic acid-induced insulin resistance and lipid accumulation in HepG2 Cells via AMPK signaling pathway. *Peer J.* 8: e10413.
- [53] Sun H, Zhang A, Wang X. 2012. Potential role of metabolomic approaches for Chinese medicine syndromes and herbal medicine. *Phytother Res.* 26(10): 1466-1471.
- [54] Guo R, Luo X, Liu J, Liu L, Wang X, Lu H. 2020. Omics strategies decipher therapeutic discoveries of traditional Chinese medicine against different diseases at multiple layers molecular-level, *Pharmacol. Res.* 52: 104627.
- [55] Allwood JW, Goodacre R. 2010. An Introduction to Liquid Chromatography–Mass Spectrometry Instrumentation Applied in Plant Metabolomic Analyses. *Phytochem. Anal.* 21: 33–47.
- [56] Chen W, Gong L, Guo Z, Wang W, Zhang H, Liu X, Yu S, Xiong L, Luo; J. 2013. A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: Application in the study of rice metabolomics. *Mol. Plants* 6: 1769–1780.
- [57] Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytl. Bull.* 19: 11–15.
- [58] Mapleson D, Accinelli GG, Kettleborough G, Wright J, Clavijo BJ. 2017. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies, *Bioinformatics* 33: 574–576.
- [59] Koren, S. [Walenz BP](#), [Berlin K](#), [Miller JR](#), [Bergman NH](#), [Phillippy AM](#). 2017. Canu: scalable and accurate long-read assembly via adaptive K-mer weighting and repeat separation. *Genome Res.* 27: 722–736.
- [60] Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. *Nat. Meth.* 17: 1–4.
- [61] Chin CS, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion, GT. et al. 2016. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Meth.* 13: 1050–1054.
- [62] Chakraborty M, Baldwinbrown JG, Long AD, et al. 2016. Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. *Nucleic. Acids Res.* 44: e147–e147.

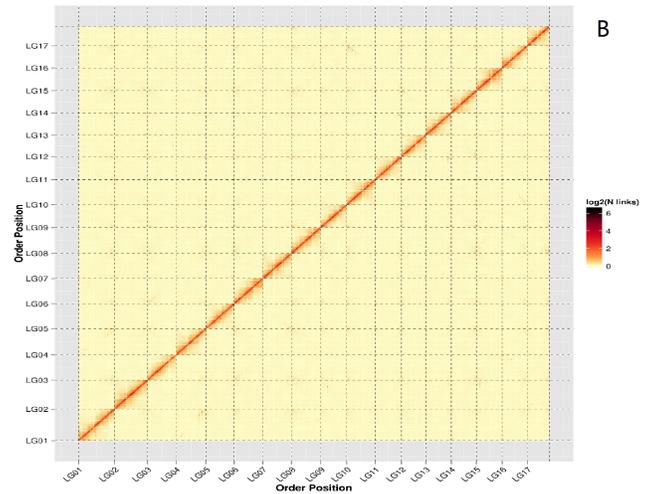
- [63] Walker BJ, Abeel T, Shea T, Priest M, Amr A, et al. 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *Plos One* 9: e112963.
- [64] Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.
- [65] Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Jay S. 2013. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* 31: 1119–1125.
- [66] Xu Z, Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic. Acids Res.* 35: W265–W268.
- [67] Han Y, Wessler SR. 2010. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic. acids res.* 38: e199
- [68] Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes. *Bioinformatics* 21: i351–i358.
- [69] Edgar RC, Myers EW. 2005. PILER: identification and classification of genomic repeats. *Bioinformatics* 21: i152–i158.
- [70] Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* 110: 462–467.
- [71] Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O. 2007. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8: 973–982.
- [72] Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* 4.10. 11–14.10. 14
- [73] Burge C, Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268: 78–94.
- [74] Stanke M, Morgenstern B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 33: W465–W467.
- [75] Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* 20: 2878–2879.
- [76] Blanco E, Parra G, Guigó R. 2007. Using geneid to identify genes. *Curr. Protoc. bioinformatics* 4.3.1–4.3.28
- [77] Kor I. 2004. Gene finding in novel genomes. *BMC bioinformatics* 5: 59.

- [78] Keilwagen J, Wenk M, Erickson JL, Schattat MH, Jan G, Frank H. 2016. Using intron position conservation for homology-based gene prediction. *Nucleic acids res.* 44: e89–e89.
- [79] Tang S, Lomsadze A, Borodovsky M. 2015. Identification of protein coding regions in RNA transcripts. *Nucleic. Acids Res.* 43: e78.
- [80] Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast splicedaligner with low memory requirements. *Nat. Methods* 12, 357–360.
- [81] Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seqreads. *Nat. Biotechnol* 33: 290–295.
- [82] Nawrocki EP, Eddy SR. 2013. [Infernal 1.1: 100-fold faster RNA homology searches](#), *Bioinformatics* 29: 2933–2935.
- [83] Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic acids res.* 25: 0955–0964.
- [84] She R, Chu J K, Pei J, et al. 2009. GenBlastA: enabling BLAST to identify homologous Gene sequences. *Genome Res.* 19: 143.
- [85] Birney E, Clamp M, Durbin R. 2004. GeneWise and genomewise. *Genome res.* 14: 988–995.
- [86] Fraga C.G., Clowers B.H., Moore R.J., et al. 2010. Signature-discovery approach for sample matching of a nerve-agent precursor using liquid chromatography-mass spectrometry, XCMS, and chemometrics. *Anal. Chem.* 82(10): 4165–4173.
- [87] Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic. Acids Res.* 28(1): 27–30.

Figures



A



B

Figure 1

Genome survey based on frequency distribution of k-mers ($k = 21$) of *Eriobotrya japonica* genome based short-reads (A) and Hi-C assembled based on contigs from long-reads (B) Note: A total of 44,651,396,663 K-mers were identified and the peak of K-mer depth observed was 57.1. Genome size can be estimated as (total K-mer number)/(the volume peak). Therefore, the genome size of loquat was evaluated as 749.25 Mb. Hi-C map showing genome-wide all-by-all interactions of the *E. japonica* genome and a high resolution of each of the 17 chromosomes that are scaffolded and assembled independently.

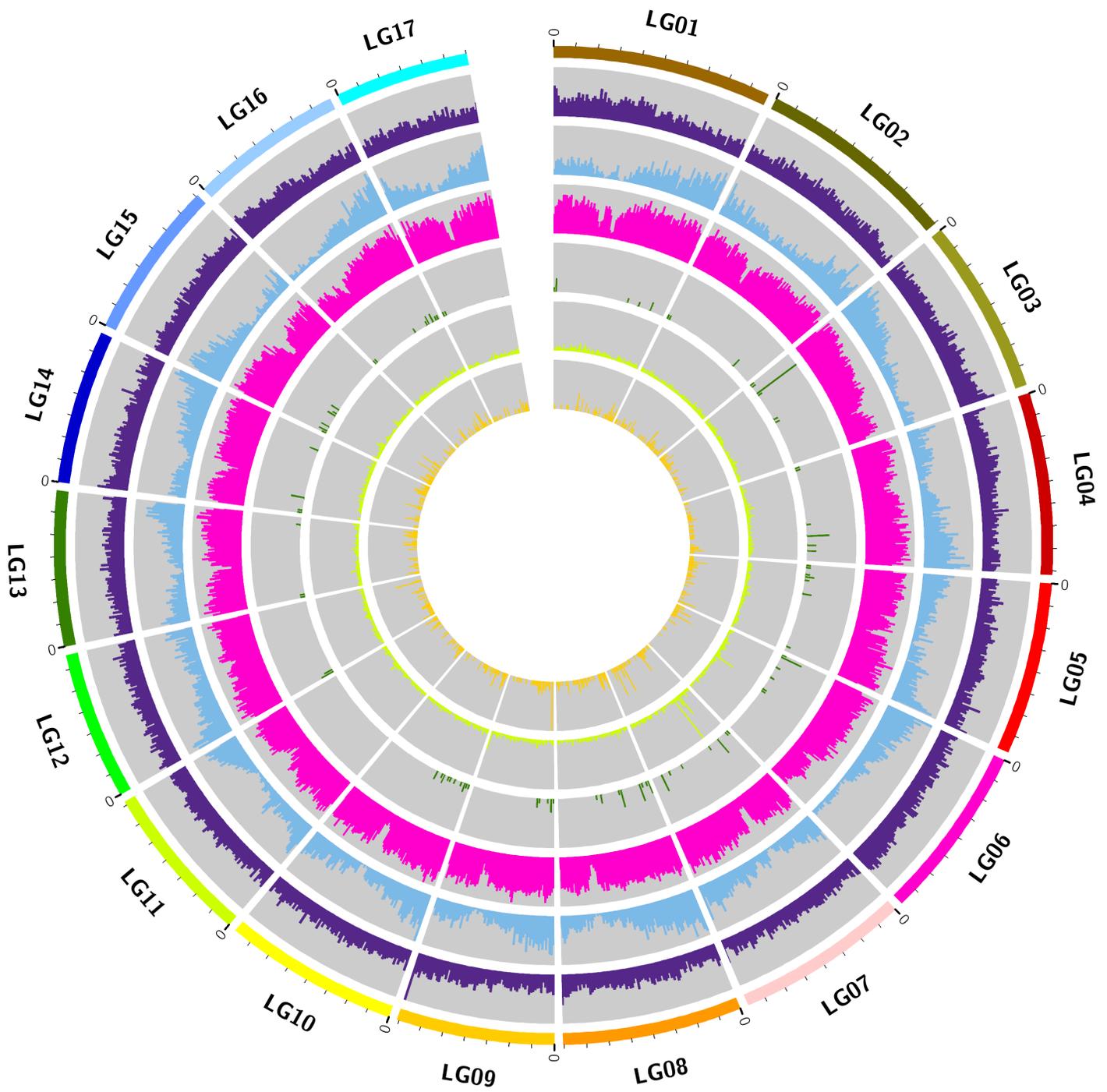


Figure 2

Chromosome-wide distribution of tRNA, rRNA, miRNA, repeats, genes, and GC from inside out

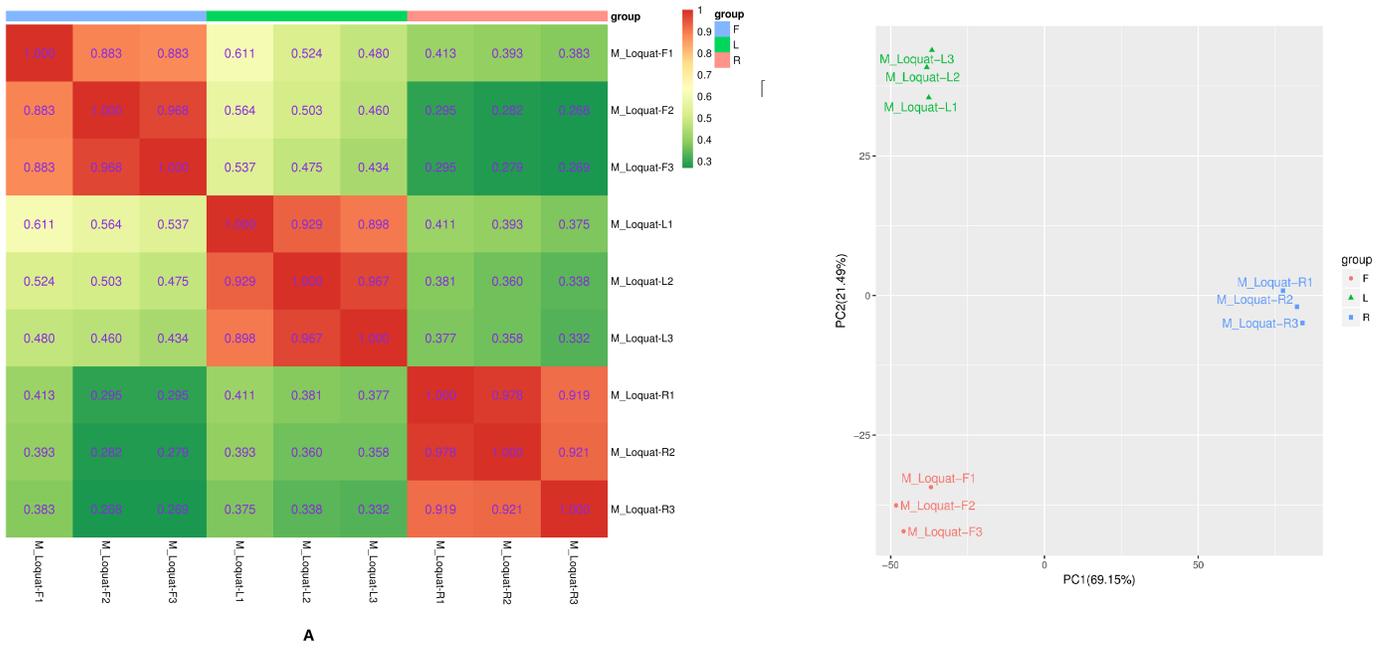


Figure 3

Principal component (A) and Pearson`s correlation (B) analysis based on Metabolites of all samples

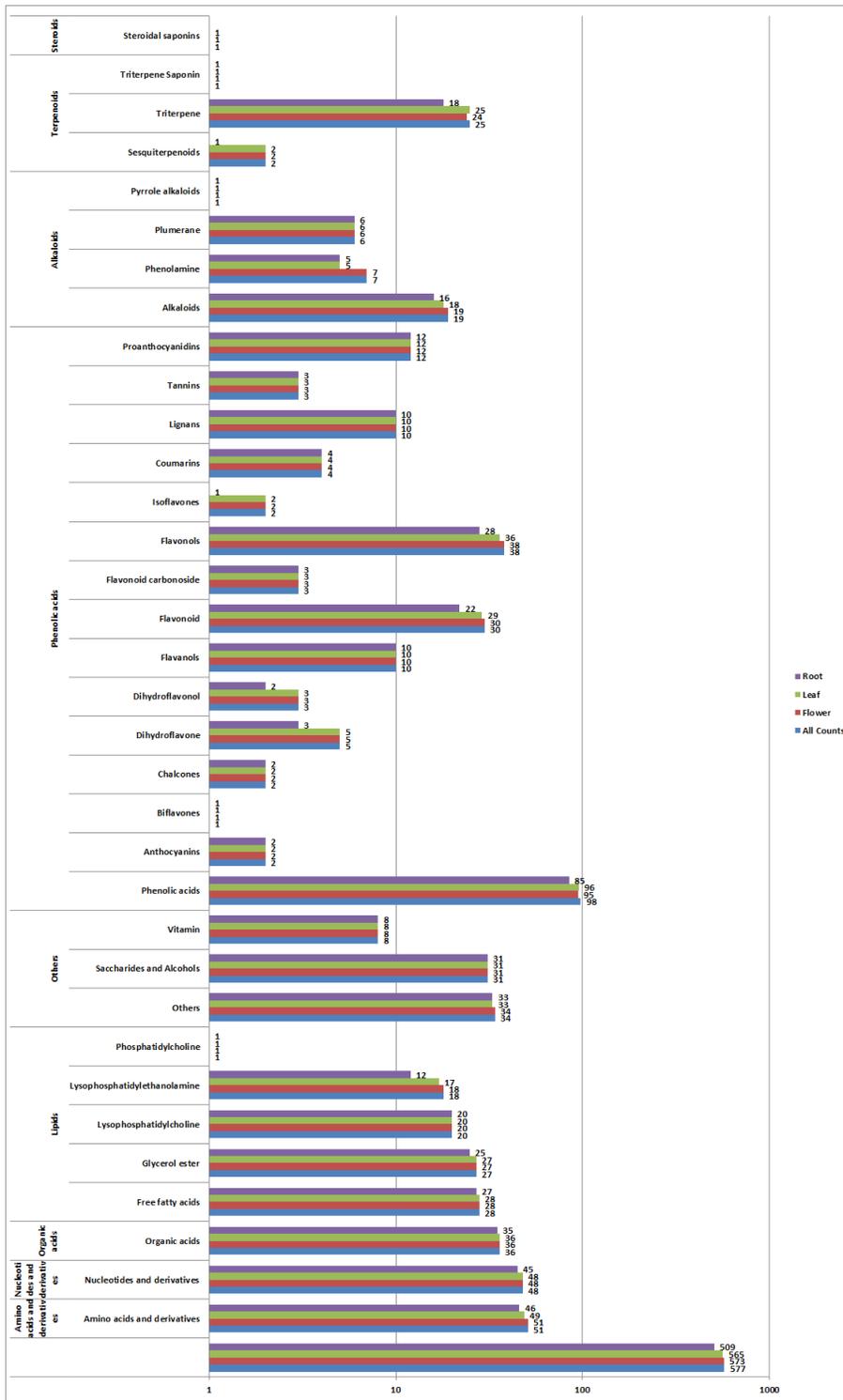


Figure 4

The category (indicating by Y-axis) and quantity (indicating by X-axis) of Metabolites in leaf, flower and root of loquat

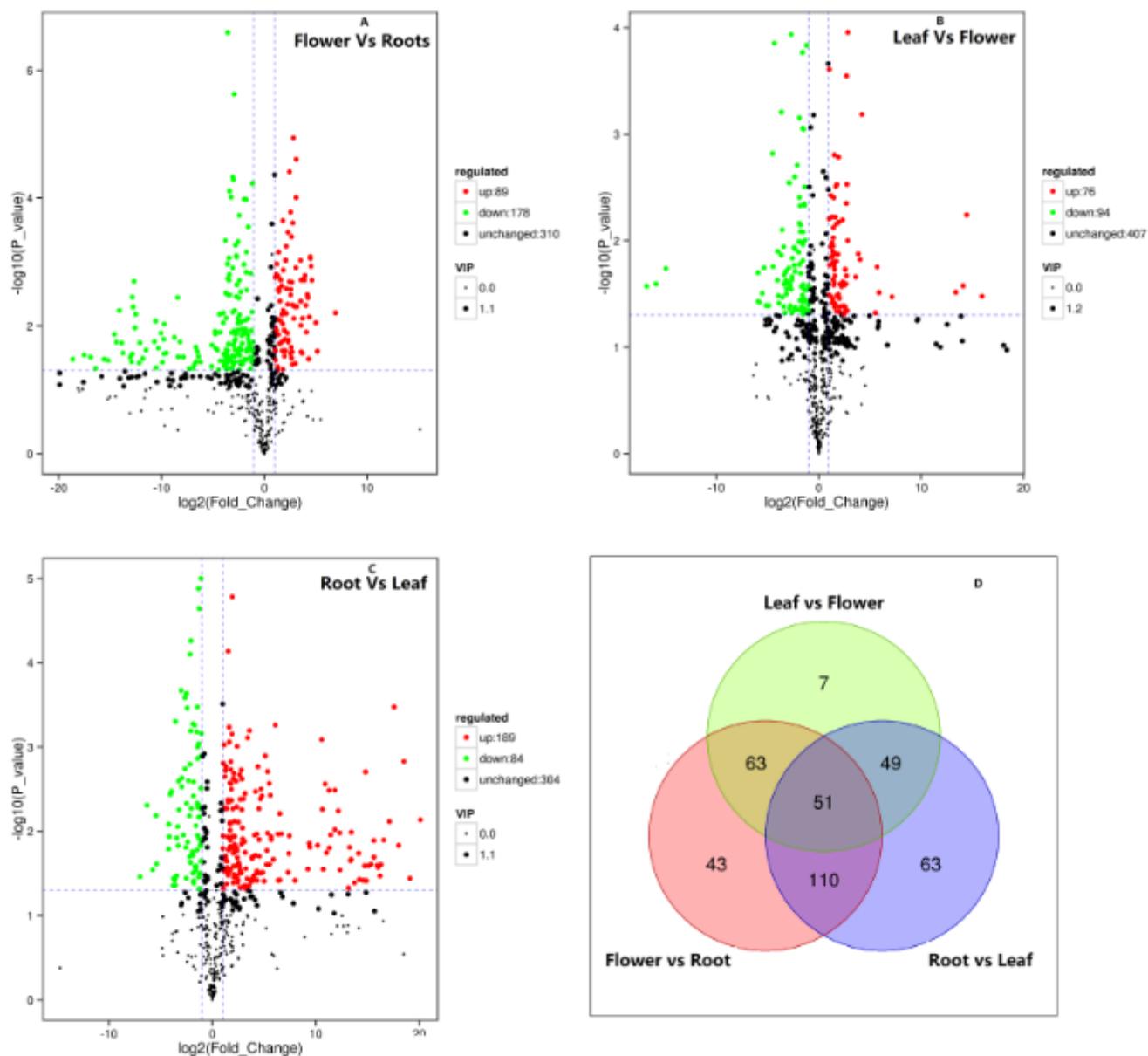


Figure 5

Hotmapping of metabolites with significant difference between flower and root (A), between root and leaf (B), and between leaf and flower (C) and venn statistics of three pairs (D)

Supplementary Files

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

- [8SupplimentaryTableSandfigureS.pdf](#)
- [Datafile.zip](#)
- [FigureS6Genesrelatedwithsesquiterpenoidandtriterpenoidbiosynthesis.tif](#)
- [FigureS5Genesrelatedwithditerpenoidbiosynthesis.tif](#)

- [FigureS4Genesrelatedwithmonoterpenoidbiosynthesis.tif](#)
- [FigureS3Genesinvolvedintothe pathway of terpenoid backbone biosynthesis.tif](#)
- [FigureS2Genesinvolvedintothe pathway of flavonoid biosynthesis.tif](#)
- [FigureS1LengthdistributionpatternofsubreadproducedbyPacbiotechnology.tif](#)