

Study of the dynamic changes in the chemical constituents of Soapberry (*Sapindus mukorossi* Gaertn.) pericarp during fruit development by a non-targeted metabolomics approach

Yuanyuan Xu

Beijing Forestry University

Yuan Gao

Beijing Forestry University

Zhong Chen

Beijing Forestry University

Guochun Zhao

Beijing Forestry University

Jiming Liu

Beijing Forestry University

Xin Wang

Beijing Forestry University

Shilun Gao

Beijing Forestry University

Duanguang zhang

Yuanhua Forestry Biotechnology Co. , Ltd

Liming Jia (✉ jlm@bjfu.edu.cn)

Beijing Forestry University

Research Article

Keywords: Sapindus mukorossi, pericarp, non-targeted metabolomics, UHPLC-QE-HF-MS, HILIC

Posted Date: January 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-148385/v1>

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

[Read Full License](#)

Abstract

Soapberry (*Sapindus mukorossi* Gaertn.) is a multi-functional tree, which is widely used in daily chemicals, biomedicine, biomass energy and landscaping. The pericarp of soapberry can be used as medicine or detergent. However, there is no systematic study on chemical constituents of soapberry pericarp in fruit development, and the dynamic changes of these constituents are far from clear. In this study, we applied a non-targeted metabolomics approach using an ultra-high performance liquid chromatography-Q Exactive HF hybrid quadrupole-Orbitrap mass spectrometer (UHPLC-QE-HF-MS) to comprehensively profile the variations of metabolites in soapberry pericarp at eight fruit development stages. The metabolome coverage of UHPLC-QE-HF-MS on a HILIC column was higher than that of a C18 column. A total of 111 metabolites were putatively identified, and these metabolites showed three accumulation patterns (pre-accumulation, mid-accumulation and post-accumulation) with fruit development. Twenty-five of these 111 metabolites (including amino acids and their derivatives, flavonoids, organic acids, fatty acids, nucleotides and their derivatives, alkaloids, carbohydrates, terpenoids, vitamins, phosphorylated intermediates) were present at significantly different levels between the two adjacent stages, which were involved in 13 KEGG pathways, among them 5 pathways (flavonoid biosynthesis; histidine metabolism; aminoacyl-tRNA biosynthesis; flavone and flavonol biosynthesis; and phenylalanine, tyrosine and tryptophan biosynthesis) were most relevant. S8 stage (fruit ripening stage) is the most suitable stage for fruit harvesting to utilize the pericarp, during which the accumulation of many bioactive and valuable metabolites (e.g., furamizole, alpha-tocopherol quinone, sucrose) in the pericarp was highest. To the best of our knowledge, this was the first time that the metabolomics in soapberry pericarp during the whole fruit development was profiled. This study will be beneficial to guide the harvesting, processing and application, and pave the way for further studies on the biosynthesis mechanism of the main metabolites of the soapberry pericarp.

Background

Metabolomics/metabonomics aims to study the components and dynamic changes of all small molecule metabolites ($MW < 1000$) in an organism or tissue, even in a single cell¹. Among them, the purpose of non-targeted metabolomics is to profile all the compounds that can be detected. The differential metabolites are screened through multivariate statistical analysis, and the pathway analysis is carried out to reveal the physiological mechanism². Liquid chromatography-mass spectrometry (LC-MS) is the most common analytical platform in metabolomics, and suitable for the analysis of compounds with poor stability and low volatility in complex samples because of its low requirements for sample pretreatment and no need for chemical derivatization. Moreover, analysis of different polar metabolites can be realized by using different chromatographic columns³.

In recent studies, metabolomics has been used to analyze the metabolic characteristics of different growth or processing periods. For example, Toffali et al⁴ generated a detailed picture of the changing metabolic profiles during late berry development in the important Italian grapevine cultivar Corvina based

on liquid chromatography-electrospray ionization-mass spectrometry (LC-ESI-MS) analysis on an analytical Alltima HP C18 column. The changes of constituents of black tea⁵ and oolong tea^{6,7} during their manufacturing process were also studied by using LC-MS on a C18 column or T3 column. Additionally, non-targeted LC-MS is widely used in biomarker discovery and disease diagnosis^{8,9}. Moreover, non-targeted metabolomics analysis based on LC-MS combined with gas chromatography-time-of-flight mass spectrometry (GC-MS), has been also used to study the characterization of biochemical variation within species or the relationship between metabolites and phenotypes or genotypes¹⁰. As we know, most of the chromatographic columns used in the non-targeted LC-MS study are C18 or T3 columns, and the use of Hydrophilic interaction liquid chromatography (HILIC) columns is still rare. However, the application of HILIC in bioanalysis has increased gradually, because of its advantages for separating polar compounds¹¹.

Soapberry (*Sapindus mukorossi* Gaertn., Sapindaceae, *Sapindus*) is an economic tree with multifunctional comprehensive utilization value which is widely used in toiletries, biomedicine, biomass energy and landscaping. It is mainly distributed in the south of the Yangtze River in China, Indochina Peninsula, India and Japan¹². In addition, soapberry is a traditional medicinal plant in China. According to the Compendium of Materia Medica, an ancient Chinese pharmaceutical book, its pericarp can be used to wash hair and face to cure dandruff and freckles¹³. Modern pharmacological studies have also shown that the pericarp of soapberry has anti-inflammatory, anti-tumor, anti-bacterial, anti-viral, hepatoprotective, insecticidal and the other bioactivity¹⁴⁻¹⁸. Besides, soapberry pericarp is rich in saponins, with good non-ionic surface activity, high foaming property and strong decontamination ability, which can replace raw materials for petrochemical products to produce detergent¹⁹. Previous studies reported that the main chemical constituents of the soapberry pericarp are terpenoids (especially triterpenoid saponins and sesquiterpenoid glycosides), phenylpropanoids, steroids and saccharides^{15,20-23}. At present, the researches on soapberry pericarp mainly focus on the saponin components and their bioactivity. The study of the non-targeted metabolomics of the pericarp of soapberry at different fruit development stages has not yet been reported.

The objectives of this study were: 1) to comprehensively describe the dynamic changes of the soapberry pericarp metabolites at eight different fruit development stages via non-targeted metabolomics using an ultra-high performance liquid chromatography-Q Exactive HF hybrid quadrupole-Orbitrap mass spectrometer (UHPLC-QE-HF-MS) on an Accucore Vanquish C18 column and a SeQuant ZIC pHILIC column; 2) to identify the differential metabolites in the two adjacent periods; 3) to analyze the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway of these differential metabolites. To the best of our knowledge, this was the first time that the metabolomics in soapberry pericarp during the whole fruit development was profiled. The results could offer valuable information for harvesting, processing and application of soapberry pericarp, and the biosynthesis of the main metabolites.

Materials And Methods

Plant materials. Three ten-year-old soapberry select tree (average tree height 6.5 m, average DBH 13.5cm, and average annual output 20 kg) were used, which cultivated in an orchard located in Jianning County, Fujian Province, China (latitude 26°49' N and longitude 116°52' E, altitude 300 m above sea level). The soapberry fruits were sampled during the period from June to November 2018 at eight time points, corresponding to initial fruit stage (S1, 15 days after flowering), cotyledons growth stage (S2, 45 days after flowering), fruit expanding stage (S3 and S4, 75 and 90 days after flowering, respectively), seed hard shell stage (S5, 105 days after flowering), fruit turning color stage (S6, 120 days after flowering), fruit near ripening stage (S7, 135 days after flowering) and fruit ripening stage (S8, 150 days after flowering) (Fig. 1). Three biological replicates were taken at each time point resulting in a total of 24 samples. Fruit was randomly picked from the East, South, West and North directions of the middle and upper part of the crown of the trees, and the number of fruits picked was determined by their size. Each biological replicate comprised fruit randomly extracted from a pool of soapberry fruit collected from each tree. The pericarp and the seed were separated, and the pericarps were immediately frozen in liquid nitrogen and stored at -80 °C. The pericarps were dried using a vacuum freeze dryer (LGJ-10, Beijing songyuan huaxing biotechnology co., Lto., Beijing, China). The dried pericarps were then ground into uniform powder by a ball mill (MM400, Retsch, Germany).

Metabolite profiling

(1) Sample preparation. The sample (30 mg) was extracted with 0.27 mL methanol/H₂O (75:25, v/v) by ultrasonic treatment for 30 min. The solution was centrifuged at 14,000 g for 10 min at 4 °C and the supernatant was evaporated to dryness under vacuum. The samples were reconstituted in 0.15 mL of either acetonitrile (ACN)/H₂O/formic acid (FA) (50:50:0.1, v/v/v) or ACN/water (60:40, v/v) containing 1.5 ppm of Sulfadimethoxine, L-Valine-1-¹³C and Pyruvic-1-¹³C acid as internal standards for subsequent liquid chromatography-mass spectrometry (LC-MS) analysis on reverse phase and HILIC separation respectively.

(2) LC-MS analysis on C18 column. Chromatographic separation was performed on a Vanquish UHPLC system with a Accucore Vanquish C18+ column (100×2.1 mm i.d., 1.5 μm) coupled to a Q Exactive™ Hybrid Quadrupole-Orbitrap High Resolution Mass Spectrometer (Thermo Fisher Scientific, San Jose, CA, USA). The mobile phase consisted of (A) water/0.1% FA and (B) ACN/0.1%FA. The gradient was as follows: (0% B), 2.0-4.0 min (0-15% B), 4.0-14.0 min (15-32% B), 14.0-19.0 min (32-50% B), 19.0-19.1 min (50-100% B), 19.1-21.0 min (100% B), 21.0-21.1 min (100-0% B), 21.1-25.0 min (0% B), then initial conditions were maintained for 5 min to equilibrate the column. The flow rate was 320 μL/min and the injection volumes were set to 2 μL. To avoid possible bias, the sequence of injections was randomized with a QC run after injection of 8 samples used for normalization in quantitation data analysis. The conditions for mass spectrometry have been previously described²⁴. In brief, nitrogen as sheath, auxiliary, and sweep gas was set at 50, 8, and 1 U, respectively. Data were acquired under resolving power of 120,000 (at *m/z* 200); automatic gain control target, 3×10⁶ ions; maximum injection time, 100 ms; scan range, 70-1050 *m/z*, spray voltage, 3.50 kV; and capillary temperature, 275°C. ESI-/+ data-dependent MS/MS spectra were generated for the QC samples and used for identification purposes, MS/MS data

were acquired with a full scan followed by top 15 MS/MS scans with resolving power of 15,000 (at m/z 200); automatic gain control target, 1×10^5 ions; maximum injection time, 50 ms; isolation window, 0.4 m/z ; and NCE 20, 30, 40. The acquired raw files were processed using Compound Discoverer 2.1. An untargeted metabolomics workflow with putative identification through in-house mass list, ChemSpider and mzCloud databases were used for processing the raw data and for compound annotation. The software parameters for alignment were 5 ppm mass tolerance for the adaptive curve model and 0.5 min maximum shift for alignment. The software parameters for detecting unknown compounds were 5 ppm mass tolerance for detection, 30% intensity tolerance, 3 for the signal to noise threshold, and 2×10^6 as minimum peak intensity.

(3) LC-MS analysis on HILIC column. Chromatographic separation was performed on a Vanquish UHPLC system with a SeQuant ZIC pHILIC column (150×2.1 mm i.d., 5 μm) coupled to a QE-HF Mass Spectrometer (Thermo Fisher Scientific, San Jose, CA, USA)²⁴. The mobile phase consisted of (A) 10 mmol/L AcONH_4 in water, pH 9.8 and (B) ACN. The gradient was as follows: 0.0-1.0 min (90% B), 1.0-15.0 min (30% B), 15.0-18.0 min (30% B), 18.0-19.0 min (90% B), 19.0-29.0 min (90% B), then initial conditions were maintained for 5 min to equilibrate the column. The flow rate was 250 $\mu\text{l}/\text{min}$ and the injection volumes were set to 2 μL . All of the samples plus a QC sample containing equal amount of each sample, were analyzed by negative and positive electrospray ionization in full scan MS mode with mass range of 70-1050 m/z while the rest setting of the MS parameters was the same as for C18. The acquired raw files were processed using Compound Discoverer 2.1 as describe above.

(4) Data analysis. A three-dimensional data matrix, including the metabolite name (putatively identified by UHPLC-QE-HF-MS), sample information (three biological repeats for each sample), and raw abundance (peak area for each putatively identified metabolite) was generated (Supplemental file 1). Raw data were subjected to three categories of normalization: normalization by median, log transformation, and pareto scaling. SIMCA 14.1 software (Umetrics, Umea, Sweden) was used to perform unsupervised principal component analysis (PCA) and supervised orthogonal projection to latent structures-discriminant analysis (OPLS-DA). R statistical environment was used to perform univariate analysis (fold change analysis, t -tests), K-means clustering analysis (Pearson correlation distances, Ward.D clustering algorithm). Time-series clustering analysis was performed using the Mfuzz package to study temporal expression patterns of the differential metabolites. And the differential metabolites were uploaded to MetaboAnalyst 4.0 (<http://www.metaboanalyst.ca/>) platform for Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathway analysis.

Results

Metabolic profiles. UHPLC-QE-HF-MS-based untargeted metabolomic approaches were performed to profile the metabolites present in pericarp of soapberry fruit in 8 different ripening stages. After pre-processing, 1790 features were extracted from the UHPLC-QE-HF-MS on the C18 column in negative mode, and 13000 features (5000 in negative mode and 8000 in positive mode) were extracted on the HILIC column. We couldn't extract metabolites from the chromatogram using the C18 column in positive

mode. Compared with in-house database and the online databases, as well as the literature and commercial standards, 56 metabolites were putatively identified in the negative ion mode on the C18 column, 19 of them were putatively annotated, and 37 metabolites were putatively chemical formula; 265 metabolites were putatively identified in the negative ion mode on the HILIC column, 88 of them were putatively annotated, and 177 metabolites were putatively chemical formula; and 53 metabolites were putatively identified in the positive ion mode on the HILIC column, 28 of them were putatively annotated, and 25 metabolites were putatively chemical formula. Excluding the common metabolites, a total of 111 metabolites were putatively annotated (Supplemental file 2: Table S1), including 34 amino acids and its derivatives, 12 organic acids, 10 fatty acids, 9 amines, 6 flavonoids, 6 nucleotides and its derivatives, 5 alkaloids, 4 carbohydrates, 4 terpenoids, 3 vitamins, 3 phosphorylated intermediates, 2 phenylpropanoids, and 13 other metabolites.

(1) Principal component analysis. We performed unsupervised principal component analysis (PCA) to assess the variations in the 111 metabolites detected across the 24 pericarp samples from soapberry fruit in 8 different development stages. According to the PCA score plot (Fig. 2), QCs were clustered crowdedly, indicating that the experimental method was reliable and the instrument was stable. In addition, all 24 samples were within the 95% confidence regions, and eight groups were separated clearly in the PCA score plot indicating distinct metabolome patterns among them. Furthermore, clear stepwise alterations of soapberry pericarp metabolome were observed during the fruit development process from S1 to S8. Stage S1 was on the right side of the first principal component, and the other 7 stages were all on the other side, suggesting that larger variations were observed within the S1 when compared to the following 7 stages.

(2) Clustering analysis. The relative content of metabolites was reflected by the abundance of metabolites. The abundance of 111 putatively annotated metabolites was analyzed by hierarchical clustering with heat map. According to the difference of metabolite abundance between different stages, the metabolites could be divided into two clusters. Stage S1 and stage S2 were classified into one cluster and the others (stages S3-S8) into another cluster (Fig. 3). And stages S3-S8 were clearly divided into two clusters, S3-S5 and S6-S8. The thermogram was also divided into two clusters by the changes of metabolites with fruit development. Cluster 1 included 38 metabolites (9 amino acids and their derivatives, 5 organic acids, 8 fatty acids, 1 amine, 3 flavonoids, 2 nucleotides and their derivatives, 3 terpenoids, 1 vitamin, 2 phosphorylated intermediates, and 4 other metabolites), and the levels of which were the highest at stage S1, and then gradually decreased. However, cluster 2 could be subdivided into two subgroups, subgroup 1 contained 39 metabolites (9 amino acids and their derivatives, 5 organic acids, 2 fatty acids, 1 amine, 2 flavonoids, 5 alkaloids, 3 carbohydrates, 2 vitamins, 1 phosphorylated intermediate, 2 phenylpropanoids, and 7 other metabolites), these metabolites accumulated gradually with the development of fruit and reached the highest at stage S8; subgroup 2 had 34 metabolites (16 amino acids and their derivatives, 2 organic acids, 7 amines, 1 flavonoid, 4 nucleotides and its derivatives, 1 carbohydrate, 1 terpenoid and 2 other metabolites), and their relative contents were the highest at stage S4.

(3) OPLS-DA analysis. PCA allows for an analysis of which factors are responsible for the largest part of the variation of the data. In contrast, supervised models can be used to understand whether samples can be separated using a specific factor. Therefore, we divided the samples into seven groups according to stages: S1-S2, S2-S3, S3-S4, S4-S5, S5-S6, S6-S7, S7-S8. Next, supervised OPLS-DA models were built. Figures of merit of the internal cross-validation are shown in Supplemental file 2: Table S3. R^2X , R^2Y and Q^2 of the fitting equations were all greater than 0.5, and the differences among the parameters were small (Supplemental file 2: Table S3), indicating that the fitting equations were reliable. According to the OPLS-DA score plots (Supplemental file 2: Fig. S1 A-G), all samples in different groups were located in 95% confidence regions, and could be distinguished by OPLS models, which indicated that there were significant differences between groups; The samples in the same group were concentrated, showing that the samples had good repeatability. It could be seen from the OPLS-DA permutation test graphs (Supplemental file 2: Fig. S2 A-G) that all Q^2 points were lower than the original Q^2 points on the right, and Q^2 was less than 0, and the regression lines of R^2 and Q^2 cross with the abscissa or less than 0, which demonstrated that these OPLS-DA models were reliable and effective²⁵.

Screening of differential metabolites. The multidimensional analysis and unidimensional analysis were used to identify the differential metabolites with variable importance in projection (VIP) > 1, P -value (t -test) < 0.05 and fold change > 3 or < 0.33. As showed in Table 1, 4, 3, 2, 1, 6, 12 and 1 candidate differential metabolites were screened from S1-S2, S2-S3, S3-S4, S4-S5, S5-S6, S6-S7 and S7-S8, respectively. After removing the common metabolites, 25 differential metabolites were screened, including 8 amino acids and their derivatives (alpha-amino adipic acid, N-acetyl-L-glutamine, N-(beta-glucosyl)-(oxindol-3-yl)acetyl-aspartate, N-undecanoylglycine, DL-histidine, 3-hydroxy-L-glutamic acid, 3-(butylsulfinyl)-L-alanine, L-leucyl-L-valine), 3 flavonoids (quercitrin, leucodelphinidin, rutin), 2 organic acids (protocatechuic acid, 3-dehydroquinic acid), 2 fatty acids (8-hydroxyhexadecanedioic acid, pinolenic acid), 2 nucleotides and their derivatives (uridine-5'-diphosphate-glucose, 8-hydroxy-deoxyguanosine), 1 alkaloids (furamizole), 2 carbohydrates (sucrose, 2-acetamido-2,6-dideoxy-alpha-D-galactopyranose), 2 terpenoids (betulin, agnuside), 1 vitamin (alpha-tocopherol quinone), 1 phosphorylated intermediate (glucose 1-phosphate) and 1 other metabolite (2,3-dinor-8-epi-prostaglandin F2alpha).

Table 1 Differential metabolites between different groups

Compare group	Candidate name	VIP score (component 1)	Fold change (S_n/S_{n+1})	<i>P</i> -value ^a	False discovery rate FDR
S1-S2	Quercitrin	2.20	6.02	2.73E-03	4.33E-02
	Protocatechuic acid	1.49	10.11	1.25E-03	2.57E-02
	Uridine-5'-diphosphate-glucose	1.18	17.10	3.84E-03	4.74E-02
	8-Hydroxyhexadecanedioic acid	1.23	0.19	8.88E-04	2.57E-02
S2-S3	Betulin	1.48	0.17	4.28E-03	4.90E-02
	Leucodelphinidin	1.23	0.27	7.15E-05	3.97E-03
	alpha-Aminoadipic acid	1.31	4.19	9.23E-04	3.41E-02
S3-S4	N-Acetyl-L-glutamine	1.54	0.32	8.82E-04	3.27E-02
	N-(beta-Glucosyl)-(oxindol-3-yl)acetyl-aspartate	1.81	0.21	2.84E-04	1.58E-02
S4-S5	N-Acetyl-L-glutamine	1.16	0.32	8.82E-04	3.25E-02
S5-S6	alpha-Tocopherol quinone	1.46	0.14	2.48E-03	3.93E-02
	8-Hydroxy-deoxyguanosine	1.57	0.10	7.12E-03	7.78E-02
	Agnuside	1.44	0.15	3.36E-03	4.67E-02
	Rutin	1.80	0.07	1.36E-03	2.51E-02
	Sucrose	1.58	0.10	1.20E-04	5.92E-02
	N-Undecanoylglycine	1.3	0.21	1.60E-04	5.92E-03
S6-S7	Glucose 1-phosphate	1.72	17.83	3.11E-03	3.33E-02

Compare group	Candidate name	VIP score (component 1)	Fold change (S_n/S_{n+1})	<i>P</i> -value ^a	False discovery rate FDR
	Furamizole	1.94	40.91	1.57E-10	1.74E-08
	DL-histidine	2.23	133.73	3.27E-03	3.34E-02
	3-Hydroxy-L-glutamic acid	2.03	59.51	3.31E-03	3.34E-02
	3-Dehydroquinic acid	1.97	45.44	2.07E-03	3.28E-02
	Uridine-5'-diphosphate-glucose	2.44	353.18	9.89E-05	3.66E-03
	3-(Butylsulfinyl)-L-alanin	2.09	0.01	1.82E-03	3.28E-02
	Rutin	1.37	6.47	1.34E-03	2.98E-02
	2-Acetamido-2,6-dideoxy-alpha-D-galactopyranose	1.51	0.11	2.24E-04	6.23E-03
	alpha-Tocopherol quinone	1.13	3.63	2.89E-03	3.34E-02
	L-Leucyl-L-valine	1.09	3.23	5.82E-05	3.28E-03
	Pinolenic acid	1.24	4.75	4.76E-03	4.40E-02
S7-S8	2,3-Dinor-8-epi-prostaglandin F2alpha	1.17	0.08	2.07E-03	4.60E-02

^a Statistical differences ($P < 0.05$) were assessed by applying a two-tailed paired Student's *t*-test.

In order to visualize the accumulation characteristics at eight different stages, time-series expression analysis was carried out to reveal the temporal expression pattern of 25 differential metabolites. The 25 differential metabolites were classified into three clusters (Fig. 4A) and their corresponding members and their accumulation level were shown in Fig. 4B. Among the three clusters, clusters 1 showed a similar expression trend where metabolites increased first and then decreased, and reached the highest at stage S3 or S4, included 2-Acetamido-2,6-dideoxy-alpha-D-galactopyranose, 8-hydroxy-deoxyguanosine, betulin, DL-histidine, L-leucyl-L-valine, N-(beta-Glucosyl)-(oxindol-3-yl)acetyl-aspartate. Meanwhile, the metabolites in cluster 2 formed a clear tendency that accumulated to highest levels under stage S1, and then decreased gradually with the development of fruit, including 3-hydroxy-L-glutamic acid, 8-

hydroxyhexadecanedioic acid, agnuside, glucose 1-phosphate, N-undecanoylglycine, pinolenic acid, protocatechuic acid, quercitrin, rutin, uridine-5'-diphosphate-glucose, alpha-amino adipic acid. However, cluster 3 had a completely opposite trend of gradually increasing and the level of metabolites reached the highest at stage S8. These metabolites included 2,3-dinor-8-epi-prostaglandin F2alpha, 3-(butylsulfinyl)-L-alanine, 3-dehydroquinic acid, furamizole, leucodelphinidin, N-acetyl-L-glutamine, sucrose, alpha-tocopherol quinone. The clustering of in three groups clearly indicated the metabolic changes in different fruit development stage.

KEGG pathway analysis of differential metabolites. To better elucidate the biological functions of differential metabolites, a pathway analysis was performed by comparing the metabolites with the KEGG reference pathway. As expected, these metabolites were involved in 13 different pathways (Supplemental file 2: Table. S3). Fig. 5 showed the influence factors of metabolic pathway, according to the impact or $-\log(P)$ value, the most relevant pathways were selected, including flavonoid biosynthesis; histidine metabolism; aminoacyl-tRNA biosynthesis; flavone and flavonol biosynthesis; and phenylalanine, tyrosine and tryptophan biosynthesis.

Discussion

Metabonomic profiling of soapberry pericarp. LC-MS is currently the most widely used means of determining metabolic phenotypes via both untargeted and targeted analysis²⁶. Metabolome profiling studies are often performed using reversed phase (RP) chromatography (particularly C18) because of its robust and reproducible separation characteristics and the coverage of a wide range of metabolites. Nevertheless, most biological matrices contain plenty polar metabolites, which cannot be retained on RP stationary phases. Correspondingly, hydrophilic interaction liquid chromatography (HILIC) has recently appeared to be the best LC approach for complementing RP chromatography in the field of metabolomics for separation of polar compounds^{27,28}. RP-based methods are used for medium to nonpolar metabolites, while HILIC is employed for the more polar metabolites that are not well retained in RP systems²⁹. In recent years, the applications of HILIC for non-targeted metabolomics are increasing^{30,31}. In the present study, we conducted a non-targeted metabolomic study based on UHPLC-QE-HF-MS with both C18 column and HILIC column in pericarp of soapberry fruit in 8 different ripening stages. After pre-processing, 1790 features in negative mode were extracted on C18 column, and 13000 features (5000 in negative mode and 8000 in positive mode) were extracted on HILIC column, which showed that the coverage of metabolites with HILIC column was higher than that with C18 column. This result suggested that the HILIC column was better than the C18 column in non-targeted metabolomics analysis of the soapberry pericarp. This may be due to the fact that there are more polar compounds in the pericarp, as we putatively identified 34 amino acids and their derivatives, 12 organic acids, 10 fatty acids, 9 amines, 6 flavonoids, 6 nucleotides and its derivatives, 5 alkaloids, 4 carbohydrates, 4 terpenoids, and 21 other metabolites. Besides, HILIC can be conveniently coupled to MS, especially in the electrospray ionization (ESI) mode, and HILIC mobile phases are very compatible and give high sensitivity¹¹. Previous studies demonstrated that there are many triterpenoid saponins in the pericarp of soapberry^{15,17,32}. Unfortunately,

we did not find triterpenoid saponins in our non-targeted metabolomics data. The reason for this result may be that there are few reports about soapberry saponins, lack of MS spectra information in various online databases, and authentic standards are not commercially available. In addition, the difference of extraction methods may also be the reason for the absence of saponins. Actually, none of the currently available separation modes in LC offers the ability to monitor the whole types of metabolites encountered in metabotyping studies²⁵. In the future, the comprehensive metabolome coverage might be obtained by using the serial coupling of RPLC and HILIC to study the non-targeted metabolomics of soapberry pericarp and continuously optimizing samples preparation, chromatographic conditions and MS parameters^{26,29,33}.

Our results indicated that no matter PCA score plot or OPLS-DA score plots, groups were separated clearly among the pericarps in different stages of soapberry fruit development, which demonstrated that the data in different stages were significantly different, and the metabolic spectrum had some changes in the process of fruit development. Metabolites showed mostly 3 different patterns of accumulation in pericarp at 8 different development stages. There were 38 metabolites belong to pre-accumulation pattern, including 9 amino acids and its derivatives, 5 organic acids, 8 fatty acids, 3 flavonoids, 3 terpenoids, 2 nucleotides and their derivatives, 2 phosphorylated intermediates, and 6 other metabolites. These metabolites were highest in S1 or S2 and then decreased gradually with fruit development. And the pericarp of stage S3-S5 contained higher concentrations of these compounds (including 16 amino acids and their derivatives, 2 organic acids, 7 amines, 4 nucleotides and its derivatives, and 5 other metabolites), which belong to mid-accumulation pattern. Other 39 metabolites (9 amino acids and their derivatives, 5 organic acids, 2 fatty acids, 2 flavonoids, 5 alkaloids, 3 carbohydrates, 2 vitamins, 2 phenylpropanoids, and 9 other metabolites) were the post-accumulation pattern, showing a trend of gradual increase and reached the highest at stage S8. It can be seen that there may be a close relationship between metabolites and fruit growth, and further studies are needed to elucidate this relationship.

Biological analysis of differential metabolites. Our non-targeted metabolomic revealed 25 differential metabolites of the soapberry pericarp in fruit development, including 8 amino acids and their derivatives, 3 flavonoids, 2 organic acids, 2 fatty acids, 2 nucleotides and their derivatives, 1 alkaloid, 2 carbohydrates, 2 terpenoids, 1 vitamin, 1 phosphorylated intermediate and 1 other metabolite. Surprisingly, the number of differential metabolites between S6 and S7 was the most (12), while those between S4 and S5, S7 and S8 were the least (1). Among these 12 differential metabolites, 10 were down-regulated, and the maximum folds change was 353.18 (uridine-5'-diphosphate glucose). This indicated that the period from fruit turning color stage (S6) to fruit near ripening stage (S7) is the key period for the change of metabolites of the soapberry pericarp in fruit development.

In this study, the pathway analysis of differential metabolites by using MetaboAnalyst was able to link with 13 metabolic and biosynthesis pathways, and the most relevant pathways were flavonoid biosynthesis; histidine metabolism; aminoacyl-tRNA biosynthesis; flavone and flavonol biosynthesis; and phenylalanine, tyrosine and tryptophan biosynthesis. There were three differential metabolites (i.e.,

quercitrin, leucodelphinidin, rutin) involved in flavonoid biosynthesis and flavone and flavonol biosynthesis. Quercitrin is a quercetin O-glycoside that is quercetin substituted by an alpha-L-rhamnosyl moiety at position 3 via a glycosidic linkage. It is an antioxidant, antileishmanial agent, and a plant metabolite. We showed that it had high contents in the initial pericarps, and was decreased within stages S2-S3, and reached the maximum at stage S4, then decreased rapidly and remained low contents during stages S5-S8. Other flavonoids, rutin and leucodelphinidin were also found changed during the fruit development. Rutin was accumulated highest at stage S1, and then declined continually. Conversely, the level of leucodelphinidin remained low stably within stage S1-S7, and elevated dramatically at stage S8. Rutin, as a metabolite and an antioxidant, is known to have a variety of biological activities including antiallergic, anti-inflammatory, antiproliferative, and anticarcinogenic properties. Leucodelphinidin is a metabolism in the plant. Previously, researchers have studied the flavonoids from leaves and stem bark of soapberry^{34,35}. Now, we also found several flavonoids in the soapberry pericarp, and three of them have obviously changed in the process of fruit development, suggesting that flavonoids play an important role in the soapberry. Based on these, the distribution, components, bioactivity and biosynthesis of flavonoids of soapberry will be investigated in future work.

In the current study, we obtained 8 differential amino acids and their derivatives. Among them, DL-histidine is involved in histidine metabolism and aminoacyl-tRNA biosynthesis pathway. Alpha-Aminoadipic acid is a metabolite in the principal biochemical pathway of lysine, which can antagonize neuroexcitatory activity modulated by the glutamate receptor N-methyl-D-aspartate. Additionally, other amino acids and their derivatives, N-acetyl-L-glutamine, N-Undecanoylglycine, L-leucyl-L-valine, have a role as metabolite. 3-Hydroxy-L-glutamic acid is a non-proteinogenic L-alpha-amino acid (<https://pubchem.ncbi.nlm.nih.gov/>).

The results of the present experiment showed that 3-dehydroquinic acid and protocatechuic acid participate in the pathway of phenylalanine, tyrosine and tryptophan biosynthesis. The level of protocatechuic acid was decreased gradually during the stages S1-S8. Recent studies indicated that protocatechuic acid could be used as a protective agent against cardiovascular diseases and neoplasms. The mechanism of its action is mostly associated with antioxidant activity³⁶.

Some terpenoids, such as agnuside and betulin, have also been found in our study. The relative content of agnuside showed a trend of decrease-increase-decrease, and reached the highest in stage S6. And agnuside has an anti-arthritis activity³⁷. However, betulin showed a trend of increase-decrease-increase and had higher content in stage S3 and S4. Betulin exhibits a wide spectrum of biological and pharmacological properties, such as anti-HIV, anti-inflammatory, and anti-cancer³⁸. This triterpene is also a precursor of triterpenoid saponins of soapberry. As we know, a lupine-type triterpenoid saponin isolated from the soapberry pericarp by Hu et al.¹⁷, betulinic acid 3-O-β-D-xylopyranosyl-(1→3)-α-L-rhamnopyranosyl-(1→2)-α-L-arabinopyranoside, is derived from betulin. Although triterpenoid saponins have not been identified in this study, we might use targeted metabolomics to study triterpenoid saponins in the future.

We found that uridine-5'-diphosphate glucose, 8-hydroxyhexadecanedioic acid, pinolenic acid and glucose 1-phosphate were the highest in stage S1, and then gradually decreased with fruit development. Uridine-5'-diphosphate glucose, a nucleotide derivative, acts intracellularly as an important intermediate in several different metabolic pathways and biosynthetic reactions, including the biosynthesis of polysaccharides such as starch and glycogen, lipopolysaccharides, and glycosphingolipids³⁹. 8-Hydroxyhexadecanedioic acid is a constituent of various plant cutins. Pinolenic acid is an octadecatrienoic acid, which has a role as a plant metabolite and an antineoplastic agent. However, glucose 1-phosphate is a fundamental metabolite (<https://pubchem.ncbi.nlm.nih.gov/>).

On the contrary, several bioactive and valuable metabolites (e.g., furamizole, alpha-tocopherol quinone and sucrose) were accumulated and remained higher in the late stage of fruit ripening. Furamizole is a nitrofur derivative that has strong antibacterial activity^{40,41}. Alpha-Tocopherol quinone, a vitamin E oxidation product, can decrease androgen receptor protein and transcript levels in prostate cancer cells²⁹. Sucrose is widespread in seed, leaves, fruit, flowers and roots of plants, where it functions as an energy store for metabolism and as a carbon source for biosynthesis. In addition to its use as a sweetener, sucrose is used in food products as a preservative, antioxidant, moisture control agent, stabilizer and thickening agent.

We also found some harmful compounds in our metabolome data, such as 5'-N-ethylcarboxamidoadenosine, methyl dopa, nitrilotriacetic acid and vorinostat, which are health hazard if swallowed. However, these compounds also have biological and pharmacological activities. For instance, 5'-N-ethylcarboxamidoadenosine is an antineoplastic agent and a vasodilator agent; methyl dopa and vorinostat have antihypertensive activity and antineoplastic activity, respectively. Besides, there were L-(+)-tartaric acid and 2-acrylamido-2-methyl-1-propane sulfonic acid in the pericarp of soapberry, which are corrosive. L-(+)-tartaric acid is often used as a food additive serving as antioxidants. But in high doses, this agent acts as a muscle toxin by inhibiting the production of malic acid, which could cause paralysis and maybe death. 2-Acrylamido-2-methyl-1-propane sulfonic acid is a Food and Drug Administration approved for use in polymer components of food-contact paper and board adhesive, but it also can cause severe skin burns and eye damage if used improperly (<https://pubchem.ncbi.nlm.nih.gov/>). These results verified the theory that “the pericarp of soapberry can be used as medicine, but have a small toxicity on people” recorded in the Compendium of Materia Medica, but the specific mechanism needs further study. Therefore, these factors should be taken into account in the future when using the pericarp.

According to our investigation on the fruit morphology, the soapberry fruit grows rapidly in the early stage, which can grow to the maximum size at fruit turning color stage (S6), and the pericarp volume is also the largest at this time⁴². Although many bioactive metabolites were accumulated in the pericarp at initial fruit stage (S1) and fruit expanding stage (S4), the first five stages were not suitable for fruit harvesting considering the maximization of economic benefits. However, in the later stage of fruit development, the level of several bioactive metabolites was low at S7 stage, but rapidly accumulated at stage S8, so stage S8 was the most suitable stage for fruit harvesting, and the pericarp collected at this time had the highest utilization value.

Conclusion

A non-targeted UHPLC-QE-HF-MS approach was firstly established in this study to investigate the metabolite profile of the soapberry pericarp at eight different fruit development stages. The metabolome coverage of HILIC column was higher than that of C18 column, suggesting that HILIC may be more suitable for non-targeted metabolomics analysis of the soapberry pericarp. A total of 111 metabolites were putatively identified, and these metabolites showed three accumulation patterns with fruit development, namely, pre-accumulation, mid-accumulation and post-accumulation. Twenty-five metabolites were found to be significantly different between two adjacent stages, including 8 amino acids and their derivatives, 3 flavonoids, 2 organic acids, 2 fatty acids, 2 nucleotides and their derivatives, 1 alkaloid, 2 carbohydrates, 2 terpenoids, 1 vitamin, 1 phosphorylated intermediate and 1 other metabolite. These differential metabolites were involved in 13 KEGG pathways, and the most relevant pathways were flavonoid biosynthesis; histidine metabolism; aminoacyl-tRNA biosynthesis; flavone and flavonol biosynthesis; and phenylalanine, tyrosine and tryptophan biosynthesis. S8 stage (fruit ripening stage) is the most suitable stage for fruit harvesting to utilize the pericarp, during which the accumulation of many bioactive and valuable metabolites (e.g., furamizole, alpha-tocopherol quinone, sucrose) in the pericarp was highest. This study offers a comprehensive view of the variations of the soapberry pericarp metabolome during fruit development, will be beneficial to guide the harvesting, processing and application, and pave the way for further studies on the biosynthesis of the main metabolites of the soapberry pericarp.

Declarations

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 32071793), the Special Foundation for National Science and Technology Basic Research Program of China (No. 2019FY100803) and Yuanhua Forestry Biotechnology Co., Ltd. We thank the Biotechnology Resource Center Metabolomics Facility, Institute of Biotechnology, Cornell University for technical support in non-targeted metabolomics and data processing. We are also grateful to Yulong Wei and Shangwei Huang for providing some help in freeze-drying of experimental samples. We gratefully acknowledge the assistance of Jian Ming, Lixian Wang, Xue Rui and Shuijing Luo in this study.

Author contributions

Y.X. and Y.G. were responsible for the conception of experimental ideas, the design of experimental methods and the writing of the first draft. G.Z., S.G. and D.Z. collected the samples. Y.X., X.W. and J.L. conducted experiments and performed data analyses. Z.C. and L.J. supervised and led the experiment, and completed the review and revision of the manuscript.

Competing interests

The authors declare no competing interests.

References

- 1 Fiehn, O. Metabolomics—the link between genotypes and phenotypes. *Plant Mol. Biol.***48**, 155-171 (2002).
- 2 Mamas, M., Dunn, W. B., Neyses, L. & Goodacre, R. The role of metabolites and metabolomics in clinically applicable biomarkers of disease. *Arch. Toxicol.***85**, 5-17 (2011).
- 3 De Vos, R. C. *et al.* Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat. protoc.***2**, 778-791 (2007).
- 4 Toffali, K. *et al.* Novel aspects of grape berry ripening and post-harvest withering revealed by untargeted LC-ESI-MS metabolomics analysis. *Metabolomics***7**, 424-436 (2011).
- 5 Tan, J. *et al.* Study of the dynamic changes in the non-volatile chemical constituents of black tea during fermentation processing by a non-targeted metabolomics approach. *Food Res. Int.***79**, 106-113 (2016).
- 6 Chen, S. *et al.* Non-targeted metabolomics analysis reveals dynamic changes of volatile and non-volatile metabolites during oolong tea manufacture. *Food Res. Int.***128**, 108778 (2020).
- 7 Wu, L. *et al.* Understanding the formation mechanism of oolong tea characteristic non-volatile chemical constituents during manufacturing processes by using integrated widely-targeted metabolome and DIA proteome analysis. *Food Chem.***310**, 125941 (2020).
- 8 Garcia, C. J., García-Villalba, R., Garrido, Y., Gil, M. I. & Tomás-Barberán, F. A. J. M. Untargeted metabolomics approach using UPLC-ESI-QTOF-MS to explore the metabolome of fresh-cut iceberg lettuce. *Metabolomics***12**, 138 (2016).
- 9 Pérez-del Palacio, J. *et al.* Metabolomic analysis of *Lavandula dentata* L. and *Lavandula stoechas* L. extracts by LC-QTOF/MS experiments and multivariate analysis techniques as a chemotaxonomical tool. *Plant Biosyst.***154**, 231-240 (2019).
- 10 Hu, C. *et al.* Metabolic variation between japonica and indica rice cultivars as revealed by non-targeted metabolomics. *Sci. Rep.***4**, 5067 (2014).
- 11 Buszewski, B. & Noga, S. Hydrophilic interaction liquid chromatography (HILIC)—a powerful separation technique. *Anal. Bioanal. Chem.***402**, 231-247 (2012).
- 12 Gao, Y. *et al.* Canopy characteristics and light distribution in *Sapindus mukorossi* Gaertn. are influenced by crown architecture manipulation in the hilly terrain of Southeast China. *Sci. Hortic.***240**, 11-

22 (2018).

13 Li, S. *Compendium of Materia Medica*. (People's Medical Publishing House, 1975).

14 Peng, Q. *et al.* Protective effects of *Sapindus mukorossi* Gaertn against fatty liver disease induced by high fat diet in rats. *Biochem. Bioph. Res. Co.***450**, 685-691 (2014).

15 Zhang, X. *et al.* Two new glycosides isolated from *Sapindus mukorossi* fruits: effects on cell apoptosis and caspase-3 activation in human lung carcinoma cells. *Nat. Prod. Res.***30**, 1459-1463 (2016).

16 Upadhyay, A. & Singh, D. K. Inhibition kinetics of certain enzymes in the nervous tissue of vector snail *Lymnaea acuminata* by active molluscicidal components of *Sapindus mukorossi* and *Terminalia chebula*. *Chemosphere***85**, 1095-1100 (2011).

17 Hu, Q. *et al.* Triterpenoid saponins from the pulp of *Sapindus mukorossi* and their antifungal activities. *Phytochemistry***147**, 1-8 (2018).

18 Xu, Y., Jia, L., Chen, Z. & Gao, Y. Advances on triterpenoid saponin of *Sapindus mukorossi*. *Chemistry***81**, 1078-1088 (2018).

19 Basu, A., Basu, S., Bandyopadhyay, S. & Chowdhury, R. Optimization of evaporative extraction of natural emulsifier cum surfactant from *Sapindus mukorossi*—Characterization and cost analysis. *Ind. Crop. Prod.***77**, 920-931 (2015).

20 Linde, H. Über Inhaltsstoffe der perikarprien von *Sapindus mukorossi* Gaertn. *Arch. Pharm.***312**, 416-425 (1979).

21 Kasai, R. *et al.* Acyclic sesquiterpene oligoglycosides from pericarps of *Sapindus mukorossi*. *Phytochemistry***25(4)**, 871-876 (1986).

22 Kuo, Y.-H. *et al.* New dammarane-type saponins from the galls of *Spindus mukorossi*. *J. agr. food chem.***53**, 4722-4727 (2005).

23 Zhang, X. M. *et al.* A new triterpenoid saponin and an oligosaccharide isolated from the fruits of *Sapindus mukorossi*. *Nat. Prod. Res.***28**, 1058-1064 (2014).

24 Sachla, A. J. & Helmann, J. D. A bacterial checkpoint protein for ribosome assembly moonlights as an essential metabolite-proofreading enzyme. *Nat. Commun.***10**, 1526 (2019).

25 Boulesteix, A.-L. & Strimmer, K. Partial least squares: A versatile tool for the analysis of high-dimensional genomic data. *Brief. Bioinform.***8**, 32-44 (2007).

26 Gika, H., Virgiliou, C., Theodoridis, G., Plumb, R. S. & Wilson, I. D. Untargeted LC/MS-based metabolic phenotyping (metabonomics/metabolomics): The state of the art. *J. Chromatogr. B***1117**, 136-147 (2019).

- 27 Nemutlu, E. & Kir, S. *HILIC based LC/MS for metabolite analysis*. (Novel Developments in Pharmaceutical and Biomedical Analysis, 2018).
- 28 Sillner, N., Walker, A., Harrieder, E. M., Schmitt-Kopplin, P. & Witting, M. Development and application of a HILIC UHPLC-MS method for polar fecal metabolome profiling. *J. Chromatogr. B* **1109**, 142-148 (2019).
- 29 Theodoridis, G. A., Gika, H. G., Plumb, R. & Wilson, I. D. *Proteomic and Metabolomic Approaches to Biomarker Discovery* (Academic Press, 2020).
- 30 Spagou, K. *et al.* HILIC-UPLC-MS for exploratory urinary metabolic profiling in toxicological studies. *Anal. Chem.* **83**, 382-390 (2011).
- 31 Arapitsas, P. *et al.* Studying the effect of storage conditions on the metabolite content of red wine using HILIC LC-MS based metabolomics. *Food Chem.* **197** 1331-1340 (2016).
- 32 Huang, H. C. *et al.* Triterpenoid saponins from the fruits and galls of *Sapindus mukorossi*. *Phytochemistry* **69**, 1609-1616 (2008).
- 33 Wahman, R., Grassmann, J., Schröder, P. & Letzel, T. Plant metabolomic workflows using reversed-phase LC and HILIC with ESI-TOF-MS. *LC GC. Europe* **37**, 8-15 (2019).
- 34 Zikova, N. I. & Krivenchuk, P. E. Chemical study of flavonoids from the leaves of *Sapindus Mukorossi* Gaerth. *Farmatsevtichnyĭ Zhurnal* **25**, 43-45 (1970).
- 35 Shah, M., Parveen, Z. & Khan, M. R. Evaluation of antioxidant, anti-inflammatory, analgesic and antipyretic activities of the stem bark of *Sapindus mukorossi*. *BMC Complem. Altern. M.* **17**, 526 (2017).
- 36 Szumiło, J. Protocatechuic acid in cancer prevention. *Postep. Hig. Med. Dosw.* **59**, 608-615 (2005).
- 37 Pandey, A., Bani, S., Satti, N. K., Gupta, B. D. & Suri, K. A. Anti-arthritic activity of agnuside mediated through the down-regulation of inflammatory mediators and cytokines. *Inflamm. Res.* **61**, 293-304 (2012).
- 38 Hordyjewska, A., Ostapiuk, A., Horecka, A. & Kurzepa, J. Betulin and betulinic acid: triterpenoids derivatives with a powerful biological potential. *Phytochem. Rev.* **18**, 929-951 (2019).
- 39 Ralevic, V. UDP-Glucose. *Xpharm Compr. Pharmacology Ref.* **80**, 1-4 (2008).
- 40 Kishima, M. *et al.* Sensitivities in vitro to antimicrobial drugs of bovine mycoplasmas isolated from respiratory and genital tracts. *Natl. Inst. Anim. Health Q* **18**, 18-26 (1978).
- 41 Kishima, M. & Hashimoto, K. In vitro sensitivities to antimicrobial drugs of Ureaplasmas isolated from the bovine respiratory tract, genital tract and eye. *Res. Vet. Sci.* **27**, 218-222 (1979).
- 42 Zhao, G. *et al.* The phenological growth stages of *Sapindus mukorossi* according to BBCH scale. *Forests* **10**, 462 (2019).

Figures

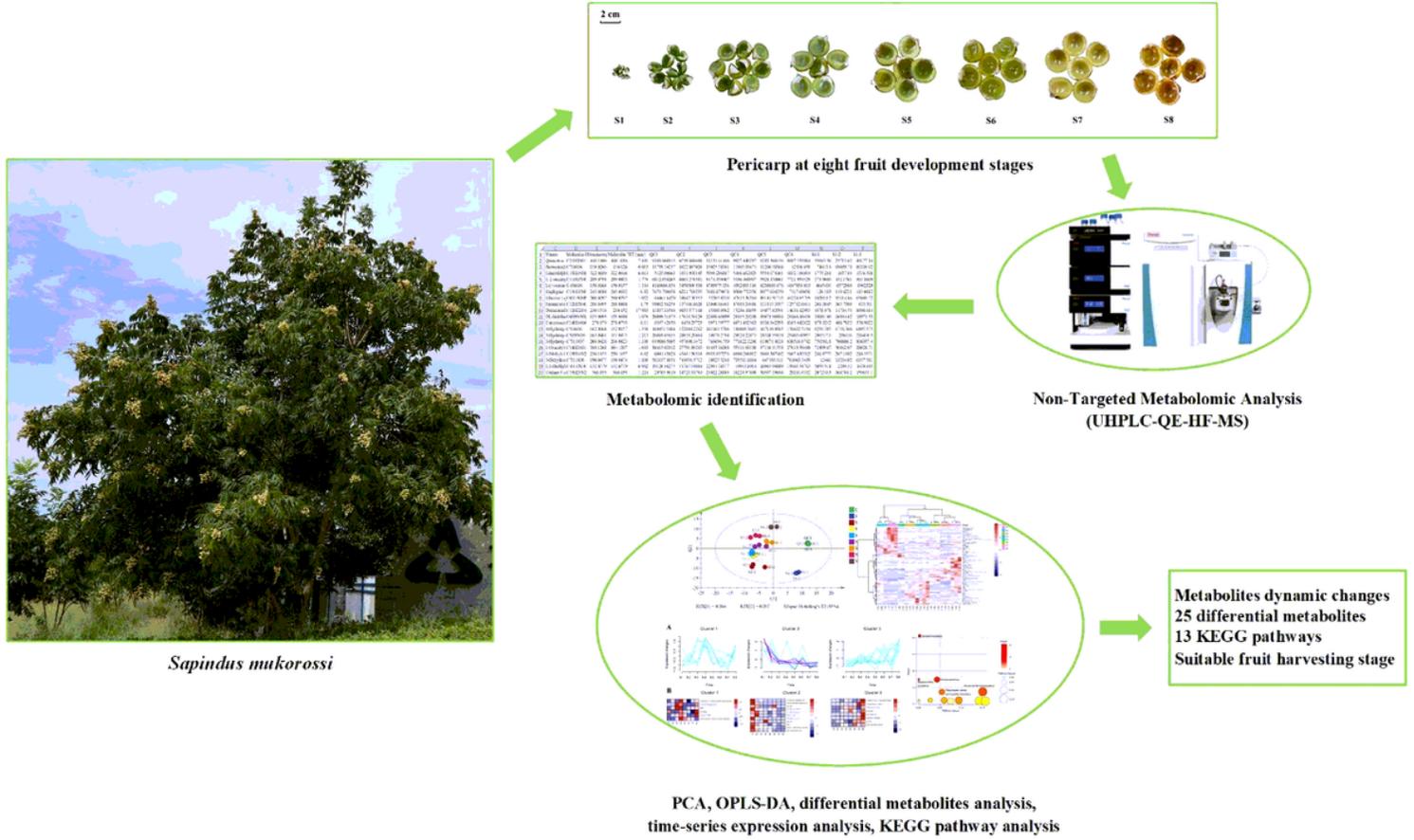


Figure 1

Flow chart of soapberry sample collection and analysis

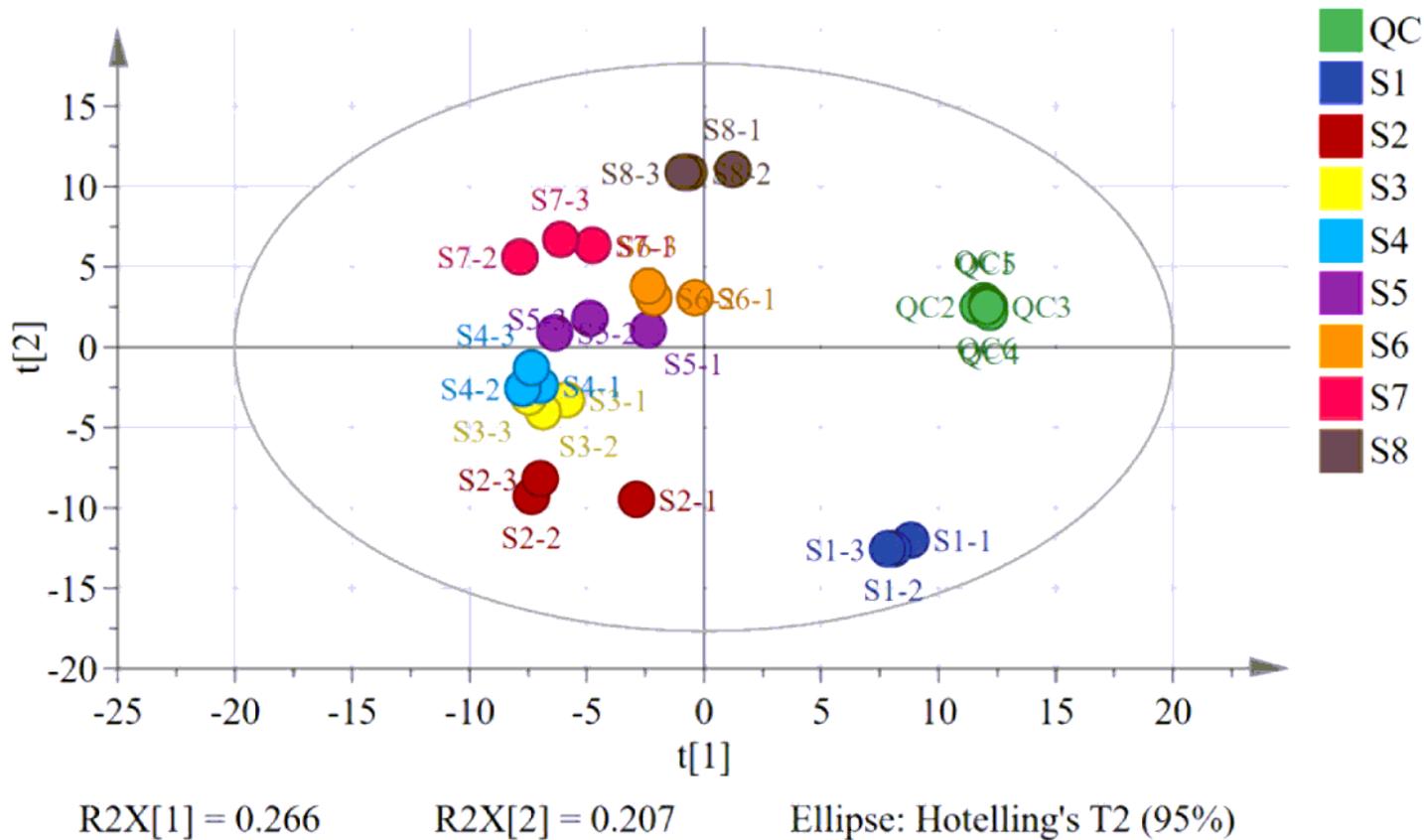


Figure 2

PCA scores plot of the all metabolites detected across the 24 pericarp samples from soapberry fruit in 8 different development stages.

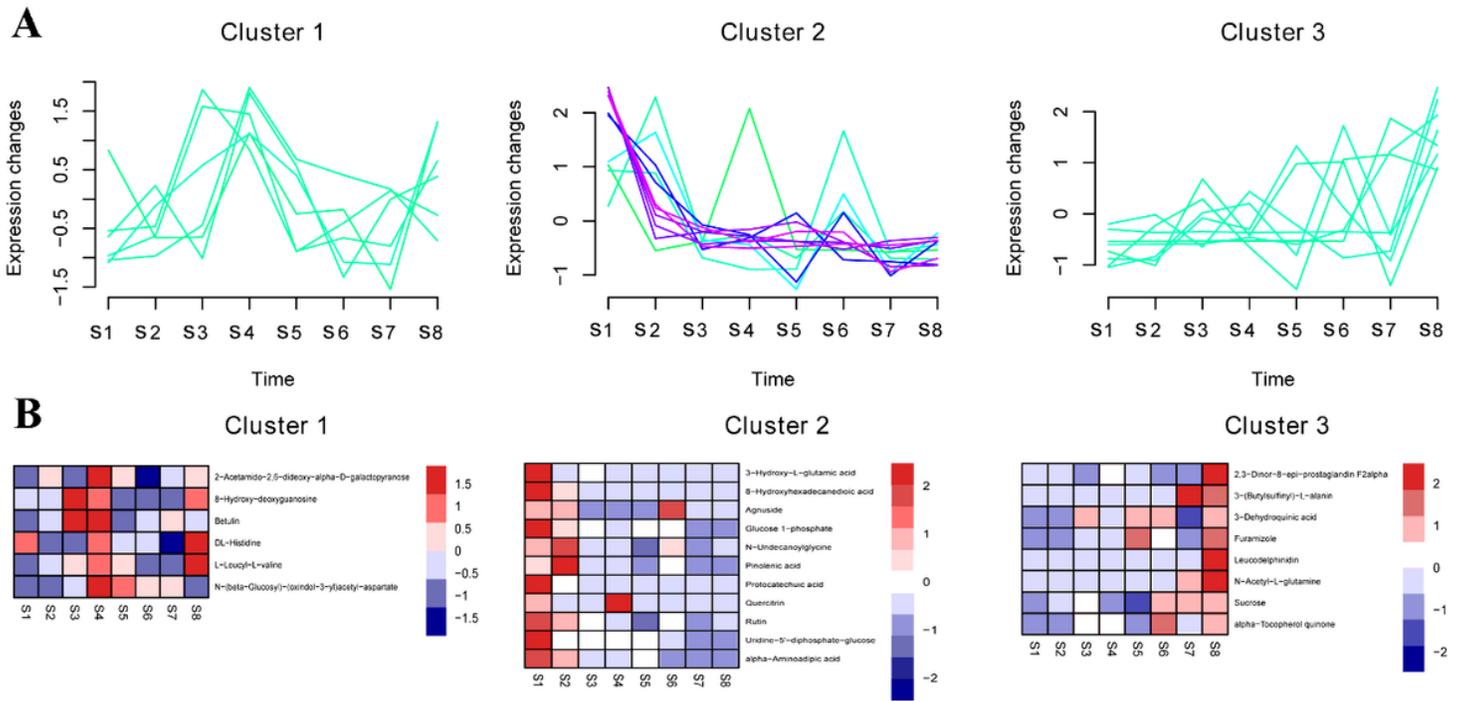


Figure 4

Changes of the differential metabolites in eight fruit development stages of the pericarps. (A) Temporal expression patterns of differential metabolites in the eight fruit development stages of the pericarps. All differential metabolites were classified into three clusters based on their expression patterns using the Mfuzz R package. (B) Heatmap of differential metabolites expression in cluster 1, cluster 2 and cluster 3.

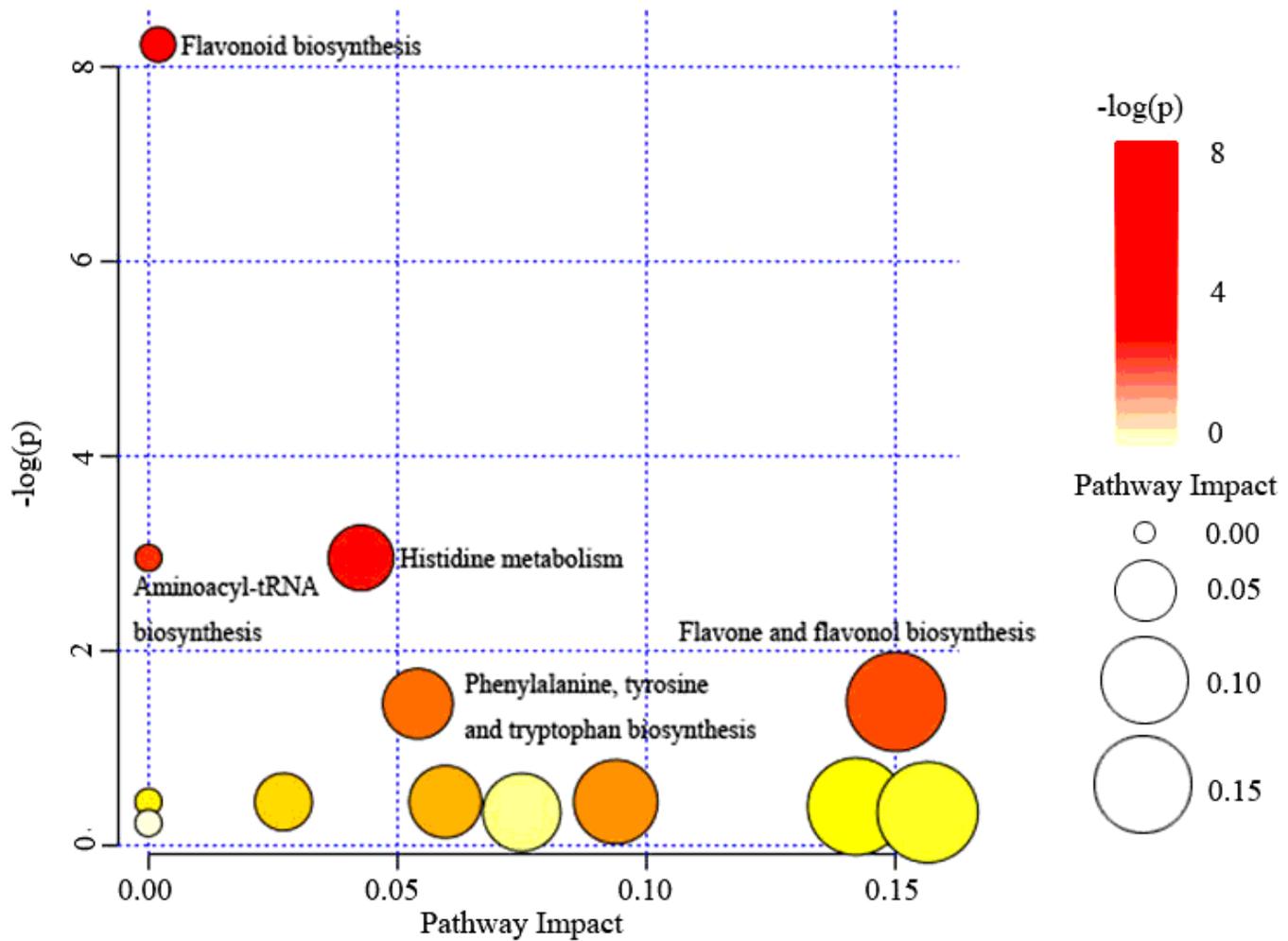


Figure 5

KEGG metabolic pathway classification of the differential metabolites.

Supplementary Files

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

- [Supplementalfile1.xlsx](#)
- [Supplementalfile2.docx](#)