ORIGINAL ARTICLE

Transcriptome Sequencing of Solanum Nigrum Linn Reveals the Genes Related to the Biosynthesis of Flavonoids

Yuan Wu, Wenjing Wang, Song Liu

Institute of Chemical Biology and Nanomedicine, Hunan University, Changsha, 410082

Abstract: In this work we comprehensively compared the transcriptome profiles of ripe fruits (RFs) and unripe fruits (UFs) and identified the genes related to flavonoid biosynthesis in Solanum nigrum Linn. (S. nigrum). Firstly, 118198 unigenes with a N50 length of 1339 bp were de novo assembled. In addition, 527 genes that were differential expression in RFs and UFs have been identified, including 329 down-regulated genes and 198 up-regulated genes. Especially, 8 differential expression genes (DEGs) were identified to be involved in flavonoid biosynthesis. Coincidentally, most of the 8 flavonoid-related genes were highly expressed in RFs. Finally, we found that the differential expression of three genes that were related to the synthesis of dihydroflavonol-4-reductase (DFR), flavonoid 3', 5'-hydroxylase (F35H) and anthocyanin synthase (ANS) was the main reason for differences in flavonoid biosynthesis between RFs and UFs of S. nigrum. This study provides some supporting information for the flavonoid accumulation in RFs. Moreover, this study can improve our understanding of the molecular mechanisms of flavonoid biosynthesis in S. nigrum.

Keywords: Transcristom; Solanum Nigrum; Next-generation Sequencing; Flavonoid; Differential Expression Genes

Introduction

Solanum nigrum is an annual solanaceae herb widely distributed in northeast Asia and can be used as a Chinese herbal medicine. In compendium of materia medica, it is said to be capable of treating carbuncle, swelling and toxin, traumatic injury, removing blood stasis and swelling, and in Chinese medicine, it is said to have diuretic effect. In recent years, with the discovery of anti-cancer active ingredients in Solanum nigrum, the research on Solanum nigrum has gradually become a hot spot, especially in its pharmacological research and application has achieved certain results. Solanum nigrum is found to have anti-cyclical,[2] anti-inflammatory,[3] calming,[4] enhancing resistance, treating allergic reactions, preventing cardiovascular and respiratory diseases and other effects,[1] and can be widely used in medicine. Industry. In addition, Solanum nigrum can also be used in the food industry, can be used as vegetables and fruits, can be used for developing Solanum nigrum juice, fruit wine and jam, and can be used for developing health soft sweets and the like[5].

Various bioactive components can be isolated from Solanum nigrum L., mainly including flavonoids, organic acids, steroids, lignans and polysaccharides[6-7]; Flavonoids are secondary metabolites based on phenylpropane, which are widely distributed in vascular plants and can protect plants from pathogens and ultraviolet radiation[8-9]." In addition, flavonoids have the functions of scavenging free radicals, resisting bacteria, inducing cell apoptosis, resisting cell proliferation, interfering cell signal transduction, enhancing anti-oncogene activity, inhibiting oncogene expression, etc[10-12].

Solanum nigrum fruit has two distinct states, the immature fruit is turquoise, while the mature fruit is dark purple. Unfinished fruits have been proved to be able to down-regulate the expression of multidrug-resistant proteins through the Jak-STAT pathway, thus reducing adriamycin resistance. Therefore, they may be used as chemosensitizers for

doi:10.18063/gse.v4i1.1172

adriamycin in the process of anti-cancer cells[13]. Mature fruits of Solanum nigrum have been proved to inhibit the proliferation of human breast cancer cells and also have free radical scavenging activity[14]. In addition, the contents of xanthones in mature and immature fruits are obviously different[15]. Therefore, this experiment studied the different growth conditions of Solanum nigrum by transcription sequencing differences in gene expression in fruits and in-depth analysis of genes related to flavonoid biosynthesis.

1. Experimental section

1.1 RNA extraction and quality detection

The wild Solanum nigrum seeds used in this experiment came from Jilin Province in northeast China, and were planted in Hunan University Institute of Chemical Biology and Nano Medicine by pot culture after germination. After the fruits are produced, the immature fruits and the mature fruits are picked and subpackaged and stored in -80℃ refrigerator.

RNA and RNA should be extracted on the ultra-clean bench to prevent RNA and contamination. The fruits shall be fully ground by mortar, and the grinding shall be carried out in liquid nitrogen. After full grinding, the kit shall be used to extract total and RNA. NanoDrop 2000, Ultraviolet Spectrophotometer, Thermo, China and Agilent Bioanalyzer 2100, Bio Analyzer, Agilent, China are used to detect the purity and integrity of RNA. The ratio of OD260/OD280 is selected to be between 1.8, 2.2, and 16. RNA samples with integrity greater than 8 for integrity detection[17]. Finally, from mature fruit group and immature fruit component. Don't choose 3 samples as biological repetitions.

1.2 cDNA, construction of library and quality detection

First, we need to isolate mRNA from the total RNA and purify it. The most prominent feature of mRNA and eukaryotic cell is that it has polyadenylic acid tail at its 5 and 5 ends, namely Poly (A), which can be used to separate mRNA from total and RNA. In this experiment, nebnext poly (a) mRNA magnetic isolation module (neb, usa) was used to separate and purify mrna. Subsequently, the construction of cDNA and library was carried out using nebnext ultra ii RNA library prepkit (neb, USA). During this period, DNA is fragmented and a linker sequence (Index) is added at both ends to distinguish the sample source of DNA. Finally, biological analyzer and Qubit 3.0 (Thermo, China) were used to detect the fragment size and concentration of cDNA and library. The length of the fragment is required to be between 300~500 bp.

1.3 cDNA library sequencing

CDNA libraries need to be pretreated before they are put on the computer. Each library is diluted to 4 nM accurately and treated with NaOH solution to unlock the double-stranded structure. The amount of NaOH and solution added needs to be strictly controlled. If the concentration of NaOH and solution added is too low, the DNA and double chain structure cannot be opened. If the concentration is too high, the whole reaction system will be destroyed. Finally, the processed samples are collected and sequenced with Illumina NextSeq 500, Sequencing Platform (Illumina, USA) and Sequencing.

In the second generation sequencing process, each base detected will determine the corresponding quality value (Q), which is an important indicator to measure the sequencing accuracy. Q30 means the probability of wrong recognition is 0.1%, Q and the higher the value means the higher the positive rate of base recognition, the better the sequencing quality. The sequencing data can be used when Q30 is greater than 75% Q30 and greater than 92.5% of this sequencing indicate that the sequencing quality is relatively high and can be used for subsequent analysis.

1.4 Data processing

The data obtained from the sequencer are bcl and format. First, software bcl2fastq converts bcl format into fastq grid format. At the same time bcl2fastq software can separate samples according to the linker sequence of each sample[18-19]. FastQC[20] and TRIMMATIC[21] software are used to detect and filter the quality of data. Some low-quality sequences and linker sequences are screened out. The data obtained after the above operations are clean data ,and clean data can be assembled from scratch (de novo) using Trinity and software. Trinity software includes Inchworm, Chrysalis and Butterfly which are three independent modules that deal with large-scale RNA-seq, reads and Data in sequence.

BUSCO is used to evaluate the quality of assembly. The principle of the software is to compare the splicing sequence with the conserved gene set of terrestrial plants and 24.

After the assembly is completed, gene expression quantification, functional annotation, differential gene analysis, etc. can be carried out. The comparison software used in gene expression quantification is Bowtie 2 and the quantification software is RSEM .Both of these two software can be used in combination with Trinity. The expression amount of each gene in the sample can be obtained by comparing the clean reads of each sample with the assembled unigene and calculating. Functional annotation is an important part of transcriptome data analysis. In this paper, software TRINOTE is used to assist in the completion, TRINOTE can compare the sequence of experimental samples with SWISS PROT, PFAM, EGGNOG, GENE Ontology (GO) and KEGG and 5 databases[25]. In this experiment, FDR is less than 0.001 and absolute value of logFC is less than 2 to screen differential genes.

2. Results and discussions

2.1 Transcriptome splicing

Through filtering and splicing the original data, this experiment has obtained 160544 ,articles ,trans-scripts, 118198, articles and Unigenes. Among them, Transcript has an average length of 825 bp, N50, 1313 bp, Unigene, 794 bp, N50, 1339 bp. N50 in transcriptome sequencing is one of the standards to measure the assembly quality of transcripts. According to the length statistics of 118198 articles Unigenes, their lengths are mostly distributed in 200~300 bp. The longer the sequence, the more valuable it is for reference. the length distribution of unigenes is shown in Figure 1.

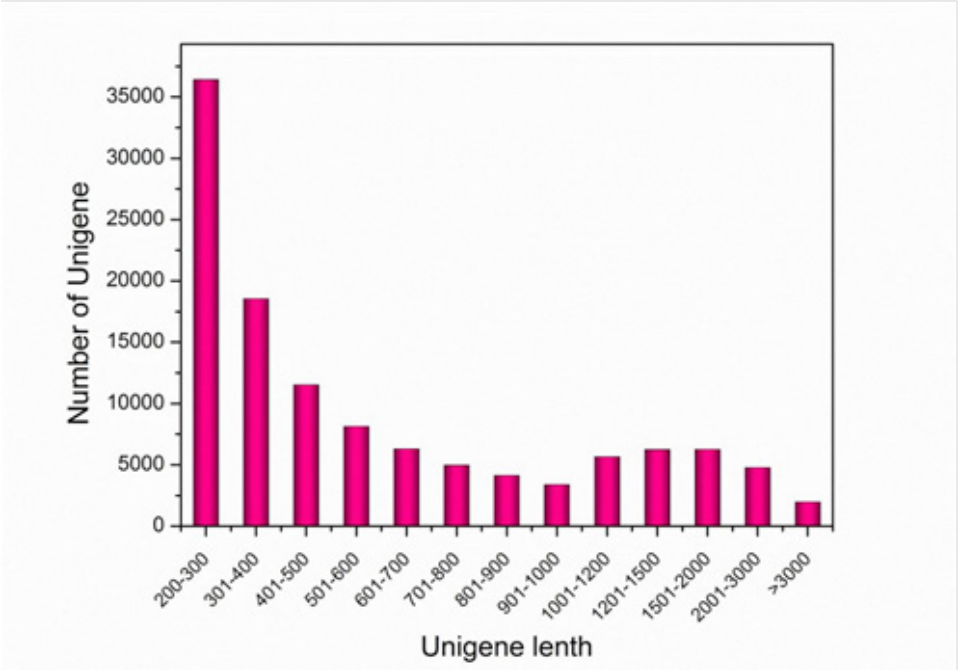

Figure 1. Length distribution of Unigene.

2.2 Transcript bssembly quality assessment

The reference genomes selected in the BUCSO and BU CSO experiments are the genomes of terrestrial plants, and there are 1440 in the genome database. A total of 1065 and 1065 complete genes in the sample were found in the gene set through comparison, accounting for about the total conservation. 74.4% of the gene set, including 1040, single copy genes and 25 multiple copy genes. " In addition, there are 231, 16.04%, 144, no genes found at all, 10%, and the comparison results are shown in figure 1. It is generally believed that when the aligned similar sequences reach 70% ,the splicing result is considered good, which shows that the assembly result of this transcriptome sequencing is good and can be used for subsequent analysis.

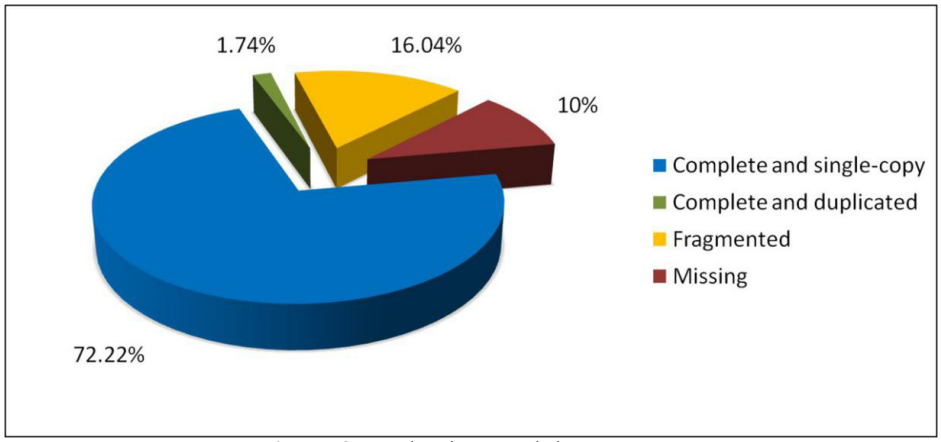

Figure 2. Evaluation result by BUSCO.

2.3 Differential gene expression analysis

In this experiment, FDR < 0.001 and /log FC/ > 2 thresholds were used to determine whether genes were differentially expressed in immature and mature fruits of Solanum nigrum. After analysis, it was found that there were 527 and 527 genes with significant differences in expression between the two groups of samples, of which 329 genes expressed in mature fruits of Solanum nigrum were lower and higher in immature fruits, while the other 198 genes expressed in immature fruits. The expression of stripe gene is higher in mature fruits of Solanum nigrum and lower in immature fruits. In Figure 3, MA map (a) and volcano map (b) of gene expression can be seen. red indicates gene distribution with higher expression in mature fruits, blue points indicate gene distribution with lower expression in mature fruits, and positions closer to both sides indicate greater difference.

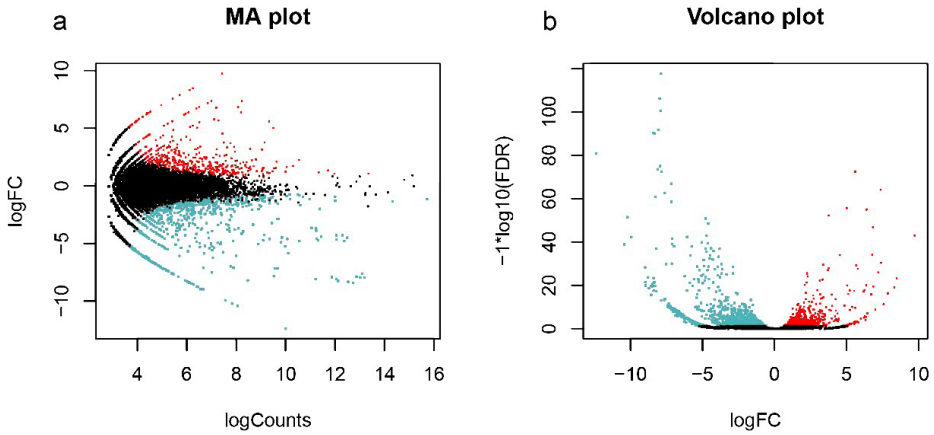

Figure 3. Ma plot (a) and (b) volcano plot of degs.

2.4 Differential gene enrichment analysis

In this experiment, we found that there are 275 and 275 differentially expressed genes are enriched into Biological process, biological process, Cellular component and Molecular function. There are 245 and 242 differentially expressed genes, which are the first of the three categories, followed by the category of biological processes. The last is the cellular component, with 204 and the genes enriched into this category. Subsets, the most enriched subset is “intricately”; The category of molecular function includes 8 subsets, and the most enriched subset is “catalytic activity”. The category of biological processes includes 11 subsets, and the most enriched subset is “metabolic process”. Details can be seen in the Figure 4 .

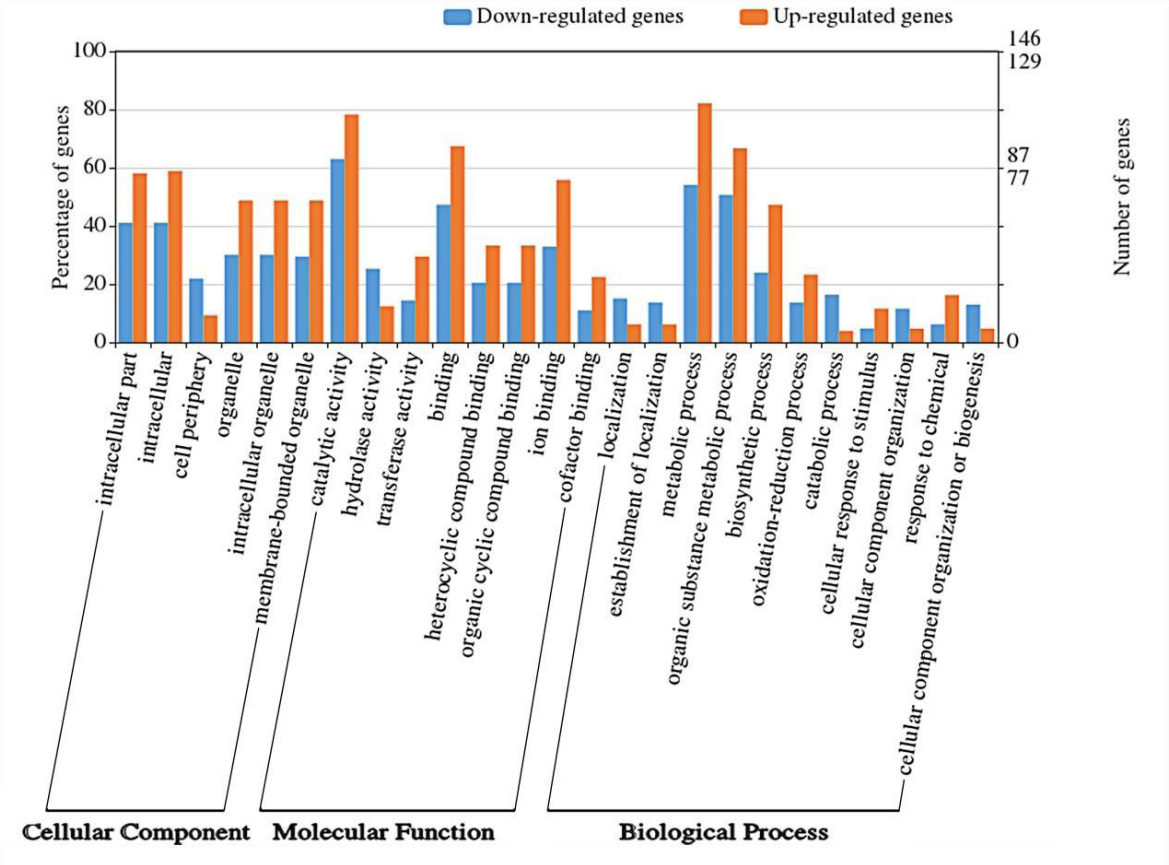

Figure 4. GO classification of DEGs.

2.5 Correlation analysis of flavonoid metabolic pathway

The differential genes were analyzed by combining with KEGG pathway, and it was found that there were 8 unigene directly related to flavonoid biosynthesis (path): ko 00941 pathway in the transcriptome of Solanum nigrum fruit. these genes can regulate shikimate hydroxycinnamyl transferase (HCT), dihydroflavonol -4- reductase (DFR), flavonoid 3 ,5-hydroxylase (F35H), chalone synthase (CHS), and caffeoycoenzyme Synthesis of A-O- Methyltransferase (CCoAOMT) and Anthocyanin Synthase (ANS). Except for one, HCT and gene, the other, 7 and 7 genes have higher expression in mature fruits. The immature fruits of Solanum nigrum contain p-Coumaric acid, p-coumaric acid, Naringenin, luteolin, Quercetin, but Delphinidin, Pelargonidin and other substances are only detected in mature fruits. The final reason for this difference is the differential expression of related genes. Visible from Figure 5, the role of these differential genes in flavonoid biosynthesis pathway. Under the action of HCT ,CHS and HCT ,-coumaric coenzyme, A caffeoyl-CoA and naringenin chalone[26-27] ,respectively. CHS is the first speed-limiting enzyme in flavonoid synthesis pathway, and CHS in the first plant was found in 1983 and year in parsley. It can combine 3 molecules of malonyl, Co A and 1 molecules of coumaric acid -CO A or cinnamoyl-CO A to form a naringenin chalcone[28] with C13 structure. CHS can convert caffeoyl coenzyme A into pentahydroxychalcone (2,3,4,4,6-Pentahydroxychalcone) and pentahydroxychalcone is the substrate[29] for the formation of Eriodictyol. Eriodictyol can form Dihydroquercetin, dihydroquercetin, dihydroflavones, Luteolin and Homoeriodictyol respectively, in which F35H catalyzing dihydroflavones has significantly higher content in mature fruits[30-31]. Naringenin, which is generated from naringenin chalcone in the pathway, can generate dihydroflavonoids through intramolecular cyclization reaction. dihydroflavonoids are the main precursors of other flavonoids, and can generate flavonols, flavones, flavanols, isoflavones, anthocyanidins, etc. through different branch synthesis pathways. Quercetin can be formed through Kaempferol, F35H, catalysis, or Dihydroquercetin, conversion[33]. ANS must be involved in the synthesis of Delphinidin, cyanidin and Pelargonidin[34].

2.6 Genes involved in flavonoid transport

Regulating the expression of flavonoid biosynthesis genes is one of the most famous regulatory systems in plants. Flavonoids are synthesized in cytosol and transported to vacuoles for storage or to other destinations[35]. Like many

other secondary metabolites at the cellular level, flavonoids need to be properly transported and stored in different compartments, mainly in vacuoles and cell walls[36-38]. Transport of flavonoids from the cytoplasm of endoplasmic reticulum (ER) to vacuole can occur through three basic mechanisms Membrane transporters, GST, Mediated Transport or Vesicle Transport Mediated Transport[39]. Multidrug and Toxic Compound Exclusion Transporter (MATE), ATP, Binding Cassette Transporter, Multidrug Resistance Related Protein ,(MRP) ,Glutathione, s-Transferase, (GST), (VSR), Soluble, N-Ethylmaleimide Sensitive Factor Attachment Protein Receptor, (SNARE), H+-PPase, Yihe, H+-ATPase play an important role in the isolation and transport of flavonoids in vacuoles[40].

In this study, we found the formation of 226, strip, unigene, codeable, ATP, binding cassette transporter, 68, strip, unigene, coding, GST, formation, 11, strip, unigene, coding, H+ -ATPase, formation by examining annotation information. These compounds are responsible for the isolation of flavonoid molecules and can coexist in plant cells. H+ -ATPases Establishing proton gradient between cytosol and vacuole (or cell wall) is the main driving force for transporting some flavonoids, especially anthocyanin, into vacuole[41]. In addition, the formation of 12, article, unigene, code, VSR, 2, article, unigene, code, SNARE, formation of 2, article, unigene, code, H+-PPase was also found. " Among the above genes, there are 5 strips encoding membrane transporters unigene and 1 strips encoding ABC transporters unigene which are significantly up-regulated in mature fruits. The high expression of these genes may be in mature fruits of Solanum nigrum. The accumulation of flavonoids has an important impact.

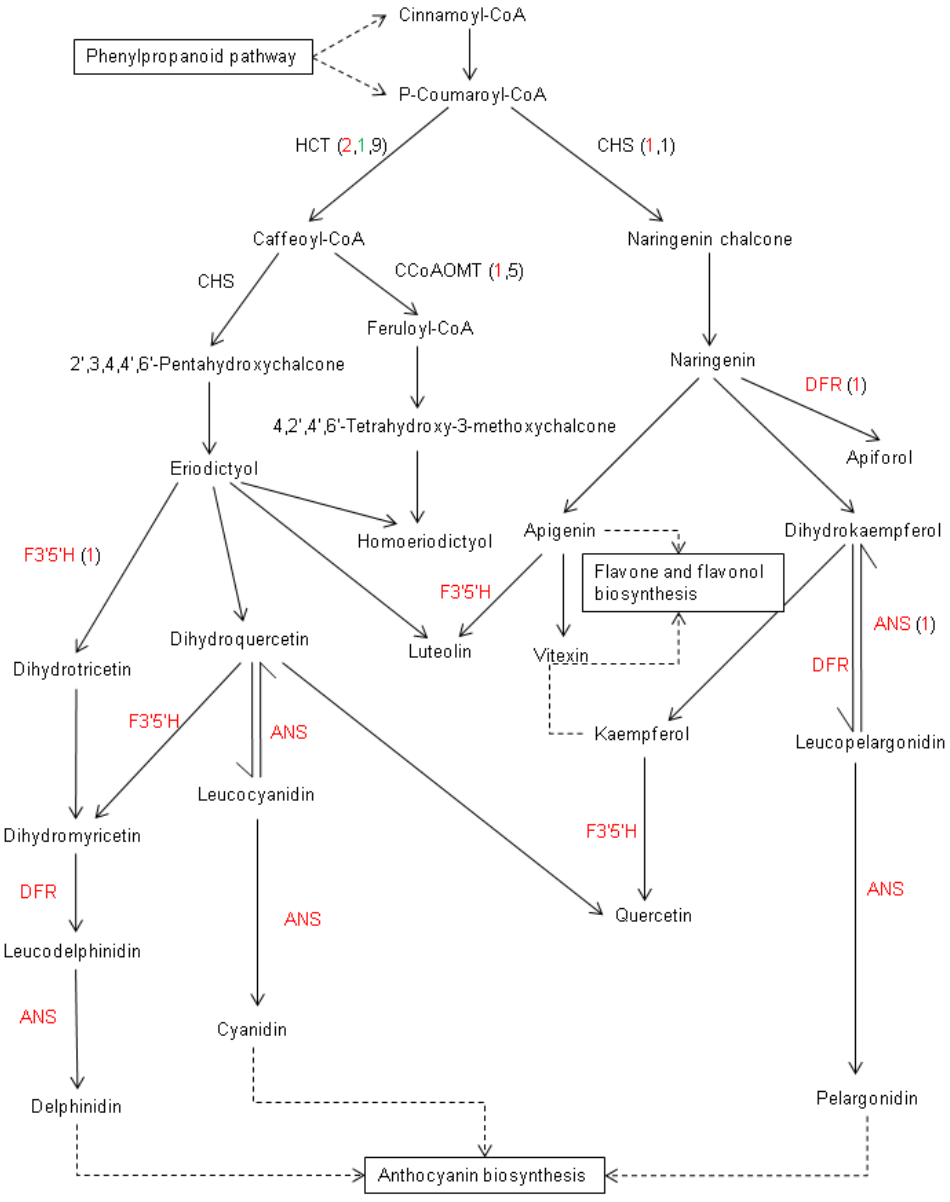

Figure 5. The PuTATIVE sketch map of the synthesis pathway of flavonoid in fruits of S.nigrum Note: The numbers in

brackets indicate the number of genes that regulate the protein synthesis, the number of genes that express up-regulation in mature fruits in red, the number of genes that express down-regulation in mature fruits in green, the number of genes that have no difference in black, and each protein is only annotated once.

3. Conclusion

Transcriptome sequencing and analysis were carried out on immature and mature fruits of Solanum nigrum. 118198 genes were assembled, including 527 differentially expressed genes, 8 of which were related to flavonoid biosynthesis. Combined with metabolic pathway discovery, DFR, F35H, and ANS are the three most important basic factors that cause the difference of flavonoids content in the fruits of Solanum nigrum in two different growth states, and they are all highly expressed in mature fruits. Together with HCT, CHS, and CCoAOMT, they accelerate the synthesis of anthocyanins in mature fruits. This experiment explains the reason why there are many kinds of flavonoids in the mature fruits of Solanum nigrum at the gene level.

References

1. Wen Zhihua. Development and Utilization of Wild Resource Solanum nigrum[J] Yunnan Agricultural Science and Technology, 2006(1):56-58.
2. Yuan H L, Liu X L, Liu Y J. Solanum Nigrum Polysaccharide (SNL) Extract Effects in Transplanted Tumor- bearing Mice - Erythrocyte Membrane Fluidity and Blocking of Functions[J]. Asian Pacific journal of cancer prevention: APJCP, 2014, 15(23):10469-10473.
3. Wang Y, Xiang L, Yi X, *et al.* Potential Anti-inflammatory Steroidal Saponins from the Berries of Solanum nigrum L. (European Black Nightshade)[J]. Journal of Agricultural & Food Chemistry, 2017, 65(21):4262-4272.
4. Taherpour A, Khodaei M M, Ameen B A H, *et al.* Chemical composition analysis of the essential oil of Solanumn nigrum L. by HS/ SPME method and calculation of the biochemical coefficients of the components[J]. Arabian Journal of Chemistry, 2013:S1878535213002761.
5. Huang Hua, Zhou Jianhua, New Development and Application of Solanum nigrum[J] Food Industry Science and Technology, 2009, 30(1).
6. Chen R, Feng L, Li H D, *et al.* Two novel oligosaccharides from Solanum nigrum[J]. Carbohydrate Research, 2009, 344(13):1775-1777.
7. He Jun, Zhou Chang Ji, Ma Bingzhi, *et al.* Research Progress on Chemical Constituents and Antitumor Pharmacological Activities of Solanum nigrum [J]. China Pharmacy, 2015, 26(31):4433-4436.
8. Hichri, Imène, Barrieu, François, Bogs J, *et al.* Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway[J]. Journal of Experimental Botany, 2011, 62(8):2465.
9. Meer I M V D, Stam M E, Tunen A J V, *et al.* Antisense Inhibition of Flavonoid Biosynthesis in Petunia Anthers Results in Male Sterility[J]. The Plant Cell, 1992, 4(3):253-262.
10. Cushnie T P T, Lamb A J. Antimicrobial activity of flavonoids[J]. Int J Antimicrob Agents, 2005, 26(5):343- 356.
11. Zhang Haifeng, Zhang Shaoliang, Zhang Chunlei, *et al.* Research Progress in Flavonoids with Anti-free Radical and UV Protective Effects [J] Daily Chemical Industry, 2008, 38(1):54-57.
12. Huang Huayi, Zha Xiliang, Research Progress on Antitumor Effects of Flavonoids [J] Chinese Journal of New Drugs and Clinical Medicine, 2002, 21(7).
13. Jagadeeshan S, David D, Jisha S, *et al.* Solanum nigrum Unripe fruit fraction attenuates Adriamycin resistance by down-regulating multi-drug resistance protein (Mdr) -1 through Jak-STAT pathway[J]. BMC Complementary and Alternative Medicine, 2017, 17(1):370.
14. Son Y O, Kim J, Lim J C, *et al.* Ripe fruit of Solanum nigrum L. inhibits cell growth and induces apoptosis in MCF-7 cells[J]. Food and chemical toxicology, 2003, 41(10): 1421-1428
15. Huang H C, Syu K Y, Lin J K. Chemical composition of Solanum nigrum linn extract and induction of autophagy by leaf water extract and its major flavonoids in AU565 breast cancer cells[J]. J Agric Food Chem, 2010, 58(15): 8699-8708.
16. Li Y, Wang F, Jiang Q, *et al.* Identification of SSRs and differentially expressed genes in two cultivars of celery

(Apium graveolens L.) by deep transcriptome sequencing[J]. Hortic Res, 2014, 1(10):10.
17. Mueller O M, Schroeder A, Lightfoot S, *et al.* RNA integrity number (RIN) - towards standardization of RNA quality assessment[J]. Cancer Research, 2004, 64.
18. Lambert D, Carrillo C D, Koziol A G, *et al.* GeneSippr: A Rapid Whole-Genome Approach for the Identification and Characterization of Foodborne Pathogens such as Priority Shiga Toxigenic Escherichia coli[J]. PLoS One, 2015, 10(4): e0122928.
19. Labreche K, Simeonova I, Kamoun A, *et al.* TCF12 is mutated in anaplastic oligodendroglioma[J]. Nature Communications, 2016, 6: 7207.
20. Wu A R, Neff NF, Kalisky T, *et al.* Quantitative assessment of single-cell RNA-sequencing methods[J]. Nature Methods, 2014, 11(1): 41-46.
21. Bolger A M, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data[J]. Bioinformatics, 2014, 30(15): 2114-2120.
22. Haas B J, Alexie P, Moran Y, *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis[J]. Nature Protocols, 2013, 8(8): 1494-1512.
23. Grabherr M G, Haas B J, Moran Y, *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome[J]. Nature Biotechnology, 2011, 29(7): 644.
24. Sim O F A, Waterhouse R M, Panagiotis I, *et al.* BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs[J]. Bioinformatics, 2015, 31(19): 3210-3212.
25. Gentleman R . Bioconductor: open software development for computational biology and bioinformatics[J]. Genome biology, 2004, 5.
26. Laurent H, Stephane M, Francoise M, *et al.* Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism[J]. Journal of Biological Chemistry, 2003, 278(1): 95-103.
27. Hoffmann L, Besseau S, Geoffroy P, *et al.* Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis[J]. Plant Cell, 2004, 16(6): 1446-1465.
28. Kang Yalan, Pei Jin, Liu Wei, *et al.* Cloning, Bioinformatics Analysis and Expression of Safflower Chalcone Synthase Gene [J] Chinese Herbal Medicine, 2014, 45(16).
29. Peters A, Schneider-Poetsch H A W, Schwarz H, *et al.* Biochemical and Immunological Characterization of Chalcone Synthase from Rye Leaves[J]. Journal of Plant Physiology, 1988, 133(2): 178-182.
30. Toguri T, Umemoto N, Kobayashi O, *et al.* Activation of anthocyanin synthesis genes by white light in eggplant hypocotyl tissues, and identification of an inducible P-450 cDNA[J]. Plant Mol Biol, 1993, 23(5): 933-946.
31. Martens S, Forkmann G. Cloning and expression of flavone synthase II from Gerbera hybrids[J]. The Plant journal: for cell and molecular biology, 1999, 20(5): 611-618.
32. Ferreyra M L F, Rius S P, Casati P. Flavonoids: biosynthesis, biological functions, and biotechnological applications[J]. Frontiers in Plant Science, 2012, 3(222): 222.
33. Wellmann F, Lukacin R, Moriguchi T, *et al.* Functional expression and mutational analysis of flavonol synthase from Citrus unshiu[J]. European journal of biochemistry, 2002, 269(16): 4134-4142.
34. Holton T A , Cornish E C. Genetics and Biochemistry of Anthocyanin Biosynthesis[J]. Plant Cell, 1995, 7(7):1071-1083.
35. Petrussa E, Braidot E, Zancani M, *et al.* Plant flavonoids--biosynthesis, transport and involvement in stress responses[J]. International journal of molecular sciences, 2013, 14(7): 14950-14973.
36. Markham K R, Gould K S, Winefield C S, *et al.* Anthocyanic vacuolar inclusions--their nature and significance in flower colouration[J]. Phytochemistry, 2000, 55(4): 327-336.
37. Séverine G, Cédric S, Laurence G. Composition and cellular localization of tannins in Cabernet Sauvignon skins during growth[J]. J Agric Food Chem, 2006, 54(25): 9465-9471.
38. Zhao J. Flavonoid transport mechanisms: how to go, and with whom[J]. Trends in plant science, 2015, 20(9): 576-585.
39. Wu Y, Guo J, Zhou Q, *et al.* De novo transcriptome analysis revealed genes involved in flavonoid biosynthesis,

transport and regulation in Ginkgo biloba[J]. Industrial Crops and Products, 2018, 124: 226-235.
40. Zhang F, Gonzalez A, Zhao M, *et al.* A network of redundant bHLH proteins functions in all TTG1-dependent pathways of Arabidopsis[J]. Development, 2003, 130(20): 4859-4869.
41. Sotiropoulos A, Gineitis D, Copeland J, *et al.* Signal-Regulated Activation of Serum Response Factor Is Mediated by Changes in Actin Dynamics[J]. Cell, 1999, 98(2):159-169.