Genetics and Molecular Research

Transcriptome analysis and anthocyanin-related genes in red leaf lettuce

Y.Z. Zhang*, S.Z. Xu*, Y.W. Cheng, H.Y. Ya and J.M. Han

Life Science Department, Luoyang Normal University, Luoyang, China

*These authors contributed equally to this study.
Corresponding author: J.M. Han
E-mail: lyhjm2000@163.com

Genet. Mol. Res. 15 (1): gmr.15017023
Received June 17, 2015
Accepted October 23, 2015
Published January 29, 2016
DOI http://dx.doi.org/10.4238/gmr.15017023

ABSTRACT. This study aimed to analyze the transcriptome profile of red lettuce and identify the genes involved in anthocyanin accumulation. Red leaf lettuce is a popular vegetable and popular due to its high anthocyanin content. However, there is limited information available about the genes involved in anthocyanin biosynthesis in this species. In this study, transcriptomes of 15-day-old seedlings and 40-day-old red lettuce leaves were analyzed using an Illuminia Hiseq™ 2500 platform. A total of 10.6 GB clean data were obtained and *de novo* assembled into 83,333 unigenes with an N50 of 1067. After annotation against public databases, 51,850 unigene sequences were identified, among which 46,087 were annotated in the NCBI non-redundant protein database, and 41,752 were annotated in the Swiss-Prot database. A total of 9125 unigenes were mapped into 163 pathways using the Kyoto Encyclopedia of Genes and Genomes database. Thirty-four structural genes were found to cover the main steps of the anthocyanin pathway, including chalcone synthase, chalcone isomerase, flavanone 3-hydroxylase, flavonoid 3'-hydroxylase, flavonoid 3',5'-hydroxylase, dihydroflavonol 4-reductase, and anthocyanidin synthase. Seven MYB, three bHLH, and two WD40 genes, considered anthocyanin regulatory genes, were also identified. In addition, 3607 simple sequence repeat (SSR) markers were identified from 2916 unigenes. This

research uncovered the transcriptomic characteristics of red leaf lettuce seedlings and mature plants. The identified candidate genes related to anthocyanin biosynthesis and the detected SSRs provide useful tools for future molecular breeding studies.

Key words: Transcriptome; Anthocyanin; Red leaf lettuce; Simple sequence repeat

INTRODUCTION

Lettuce (*Lactuca sativa* L.) is a leafy vegetable belonging to the Compositae (Asteraceae) family. It is popular worldwide due to its healthy properties and is primarily consumed fresh or in mixed vegetable salad (Doğan and Salman, 2007; Baslam et al., 2013). Most cultivated lettuce varieties have green color, but a small group displays red color due to its high anthocyanin accumulation. Anthocyanins are a subgroup of flavonoids mainly responsible for red, purple, and blue pigments in plants. A previous research revealed that anthocyanin may prevent the onset of chronic diseases like cancer or cardiovascular diseases when it is part of a regular healthy diet (Scalbert and Williamson, 2000; Knekt et al., 2002). Colored vegetables and fruits gained, since then, increased attention.

The anthocyanin biosynthesis pathway is conserved in plants. Plants, such as *Arabidopsis*, maize, and petunia, are model plants for studies of anthocyanin biosynthesis and enzymes involved in this pathway. There are six main types of enzymes in the anthocyanin biosynthesis pathway: chalcone synthase (CHS), chalcone isomerase (CHI), flavanone 3-hydroxylase (F3H), flavonoid 3'-hydroxylase (F3'H), dihydroflavonol 4-reductase (DFR), and anthocyanidin synthase (ANS). These enzymes synthesize anthocyanins with different colors (Winkel-Shirley, 2001). The anthocyanin pathway is regulated by transcription factors from the MYB, bHLH, and WD40 families. These factors form a complex that binds to the anthocyanin structural gene promoter to control its transcription (Grotewold et al., 2000; Gonzalez et al., 2008; Schaart et al., 2013). In many plants, the mutation of anthocyanin-related genes usually causes color variation, such as in *Vitis vinifera* (Kobayashi et al., 2004), *Dahlia* (Ohno et al., 2011), and *Tulipa fosteriana* (Toda et al., 2002; Yuan et al., 2014). Uncovering anthocyanin-related genes using a traditional method for plants without a reference genome is time consuming. Recently, researchers have found out that high-throughput sequencing technology is useful for identifying genes responsible for plant color. This technology is also preferably used to investigate gene expression in red leaf lettuce.

In this species, cyanidin 3-glucoside is the pigment primarily responsible for the red color. Park et al. (2007) isolated five fragments of the anthocyanin synthase gene, whose transcription coincided with plant color. However, the molecular mechanism of anthocyanin accumulation in lettuce is still unclear and gene transcription has a poor outcome. This study aimed to analyze the transcriptome profile of red lettuce and identify the genes involved in anthocyanin accumulation. The findings of this study are expected to help uncovering the transcriptomic characteristics of red leaf lettuce and contribute with important information for future molecular research.

MATERIAL AND METHODS

Plant materials and RNA extraction

L. sativa L. var. *capitata* was grown in the field in Luoyang (China). A total of 30 seedlings

from a 15-day-old plant and young leaves from thirty 40-day-old plants were collected, frozen immediately in liquid nitrogen, and stored at -80°C. Total RNA was extracted using the RNAprep pure Plant Kit (Tiangen, China) according to the manufacturer protocol and treated with DNase to prevent genomic DNA contamination. RNA quality was verified on a Bioanalyzer 2100 (Agilent, CA, USA).

cDNA library construction and Illumina sequencing

Illumina sequencing (USA) analysis was performed at Biomarker Technologies Co., Ltd. (Beijing, China), following manufacturer protocols. Briefly, mRNAs were isolated using oligo (dT) magnetic beads and fragmented into small pieces using a fragmentation buffer. First strand cDNA was reverse transcribed by random hexamers, followed by second strand cDNA synthesis with RNase H and DNA polymerase I. The double-stranded cDNA was subjected to an end repair process and the addition of single nucleotide A (adenine). Finally, paired-end adapters were ligated to the ends of the cDNA fragments. Suitable fragments were selected as templates for polymerase chain reaction amplification. cDNA libraries were sequenced using an Illumina HiSeq2500 platform. Sequence data were deposited in the NCBI database under accession Nos. SRR2057015 and SRR2057017.

De novo transcriptome assembly

Raw reads were cleaned by removing adapter sequences, low-quality reads in which the percentage of unknown bases (N) is >10%, and reads with more than 20% Q < 20 bases (base quality lower than 20). The remaining clean reads from the 15-day-old seedling library and the 40-day-old leaf library were assembled. *De novo* assembly was carried out using the Trinity software. To eliminate redundant sequences, all transcripts were clustered and the longest transcript in each cluster was selected as unigene.

Gene annotation and classification

The assembled unigene sequences were used to search public databases, including the NCBI non-redundant protein (NR), Swiss-Prot, TrEMBL, and Pfam using BLAST with an E-value lower than 10^{-5}. Based on NR annotation, the Blast2GO software was used to get the gene ontology (GO) annotation according to molecular function, biological process, and cellular components. The WEGO software (http://wego.genomics.org.cn/cgi-bin/wego/index.pl) was then used to perform a GO functional classification. All unigene sequences were also aligned to the KOG (eukaryotic orthologous groups) database to predict and classify possible functions. Pathway assignments were carried out based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database.

Simple sequence repeat (SSR) development

MISA software was used to detect SSRs. The parameters were adjusted for identifying perfect di-, tri-, tetra-, penta-, and hexa-nucleotide motifs with a minimum of 6, 5, 5, 5, and 5 repeats, respectively. Only unigenes that were longer than 1 kb were included in the SSR detection.

RESULTS

Sequencing and *de novo* assembly

cDNA libraries of seedlings and leaves were sequenced. After cleaning the data, 23,047,704 (5.8 G data) and 22,951,504 (5.78 G data) clean reads were obtained from, respectively, seedling and leaf libraries. All clean reads from the 2 libraries were assembled, generating 139,269 transcripts with an N50 of 1779 and 83,335 unigenes with an N50 of 1067. The length distribution of unigenes is shown in Figure 1. A total of 53,054 unigene sequences (63.66%) had a length between 200 and 500 nucleotides (nt), 15,065 unigenes (18.08%) were between 500 and 1000 nt in length, and 15,216 unigenes (18.25%) were longer than 1000 nt.

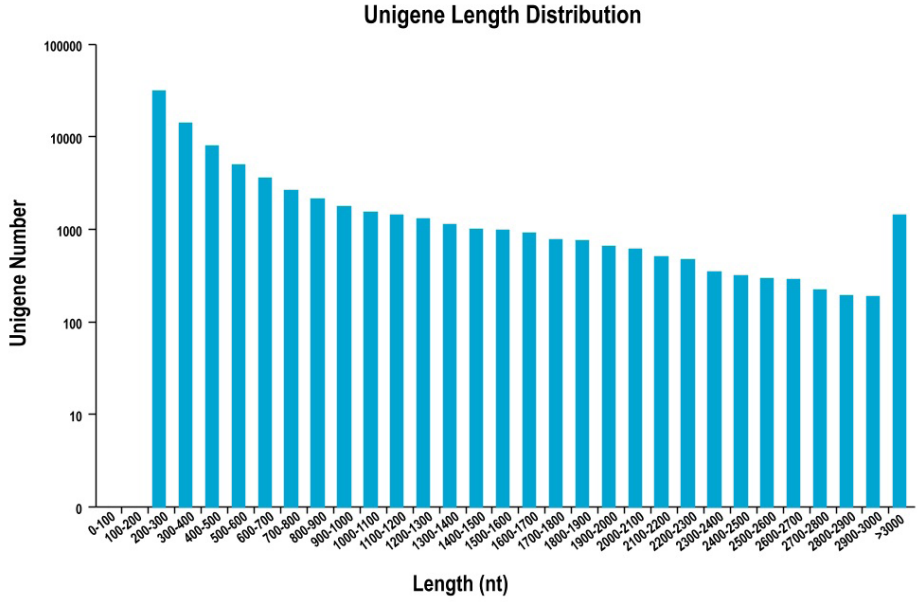

Figure 1. Length distribution of *de novo* assembled unigenes.

Sequence annotation

Several complementary approaches were used to annotate the assembled sequences. The overall functional annotation is shown in Table 1. A total of 54,868 unigenes (65.8%) had significant hits in the public database. Among them, 53,991 (64.8%) had significant matches in the NR database and 37,966 (45.6%) had similarity to proteins in the Swiss-Prot database. Based on NR annotation, GO analyses were carried out and 34,215 unigenes were assigned to 3 GO categories, that is, biological processes, molecular functions, and cellular components (Figure 2). In the cellular component category, “cell part” and “cell” were the two most represented GO terms. In the molecular functions category, “catalytic activity” and “binding” were the most represented GO terms. In the biological process category, “metabolic process”, “cellular process”, and “single-organism process” were the most represented terms.

Table 1. Overview of unigene annotations in public databases.

Annotated databases	Unigene	≥300 nt	≥1000 nt
nr	53,991	38,086	14,313
Swiss-Prot	37,966	27,790	11,226
Pfam	32,978	26,296	12,376
GO	34,215	23,239	8,264
COG	16,850	13,105	6,164
KOG	32,724	22,841	8,938
KEGG	14,334	9,920	3,562
All	54,868	38,494	14,341

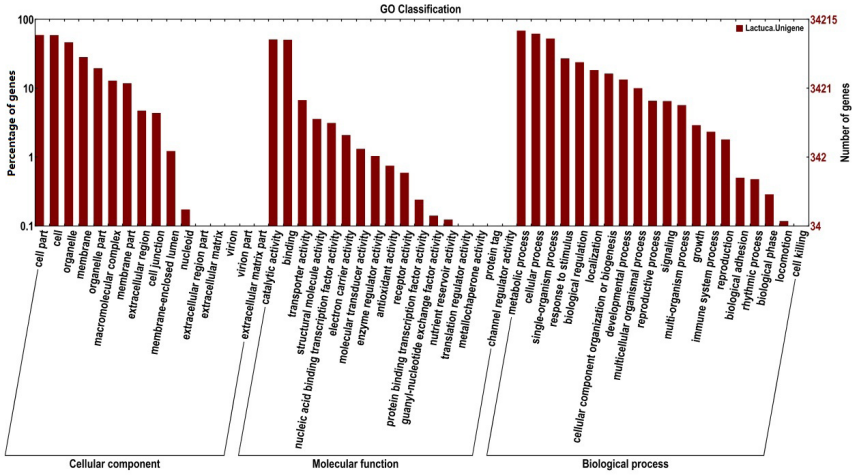

Figure 2. Functional annotation of assembled sequences based on gene ontology categorization.

Unigene sequences were searched against the KOG database and 32,724 unigenes were annotated into 25 categories (Figure 3). Among these, the ‘‘general function prediction only’’ category represented the largest group (6622 unigenes), followed by the ‘‘post-translational modification, protein turnover, chaperones’’ (3963 unigenes) and the ‘‘signal transduction mechanisms’’ (3337 unigenes) categories.

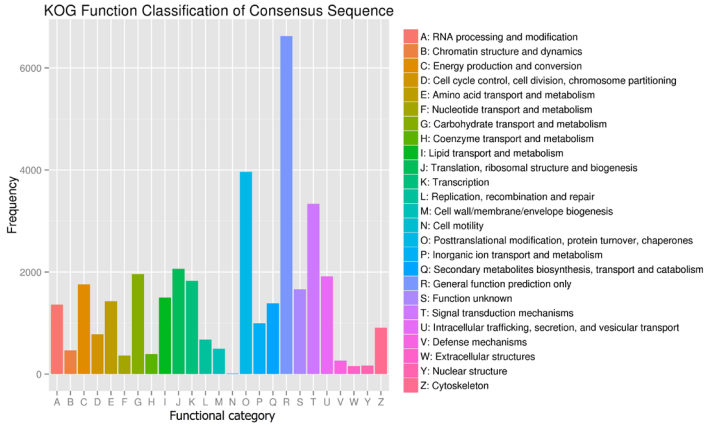

Figure 3. Clusters of eukaryotic orthologous groups (KOG) classification.

Pathway assignment based on the KEGG classification system

Overall, 12,082 annotated unigene sequences were assigned into 119 KEGG pathways (**Table S1**). Among these, three pathways were found to be related to plant color: the "flavonoid biosynthesis" (34 unigenes; ko00941), the "flavone and flavonol biosynthesis" (9 unigenes; ko00944), and the "carotenoid biosynthesis" (87 unigenes; ko00906) pathways. Eight glucoside metabolic pathways were found, among which the "glycolysis/gluconeogenesis" (ko00010), the "pentose phosphate pathway" (ko00030), the "fructose and mannose metabolism" (ko00051), the "galactose metabolism" (ko00052), the "starch and sucrose metabolism" (ko00500), and the "amino sugar and nucleotide sugar metabolism" (ko00520) pathways. This shows that glucoside is an important substrate for anthocyanin. In addition, many genes were assigned into the "healthy properties biosynthesis" pathway, including "thiamine metabolism" (ko00730), "riboflavin metabolism" (ko00740), "vitamin B6 metabolism" (ko00750), and "biotin metabolism" (ko00780).

Prediction of transcription factors

Transcription factors play a key role in regulating secondary metabolism by controlling gene expression. Based on gene annotation results, 4454 unigenes were found to display a high homology with known transcription factors belonging to 51 families (Table 2). Among these, 175 and 524 unigenes belonged to MYB and bHLH families, respectively.

Table 2. Transcription factor families identified in red leaf lettuce transcriptome.

Transcription factor family	Unigene number	Transcription factor family	Unigene number	Transcription factor family	Unigene number
Dof	869	bZIP	142	TALE	66
bHLH	524	NAC	135	G2-like	65
WRKY	366	related	118	CO-like	58
C2H2	270	GRAS	115	CAMTA	53
C3H	224	ARF	98	B3	52
MYB	175	HD-ZIP	84	GATA	51
ERF	174	Trihelix	81	Others	1342

Discovery of putative anthocyanin-related genes

In model plants, the anthocyanin pathway is usually divided into "early biosynthetic genes (EBGs)" and "late biosynthetic genes (LBGs)". EBGs include CHS, CHI, F3H, F3'H, and F3'5'H, whereas LBGs include DFR and ANS. In this study, 34 genes encoding enzymes of the anthocyanin biosynthesis pathway were identified (Table 3). Among these, eight unigenes belonged to the CHS family, three belonged to the CHI family, eight belonged to the F3H family, four to the F3'H family, and three to the F3'5'H family. These resulted in the biosynthesis of cyandidin, pelargonidin, and delphinidin. In the LBG pathway, five unigenes belonged to the DFR family and three to ANS family. This showed that unigenes from this pathway are involved in flavonoid biosynthesis.

Previous studies revealed that the anthocyanin pathway was mainly regulated by transcription factors from the MYB, bHLH, and WD40 families (Gonzalez et al., 2008; Schaart et al., 2013). In this study, 12 candidate transcription factors were found to regulate anthocyanin biosynthesis (Table 3), seven of which belonged to the MYB family, three to the bHLH family, and two to the WD40 family. Two MYB genes showed high similarity to AtMYB12, which mainly regulates EBGs (Stracke et al., 2007; Stracke et al., 2010). In *Arabidopsis*, five unigenes showed high similarity to four MYB genes (MYB75, MYB90, MYB113, and MYB114), mainly responsible

for regulating LBGs (Baudry et al., 2004; Cominelli et al., 2008). In addition, two unigenes from the bHLH family showed high similarity to EGL1, whereas one showed high similarity to GL3. EGL1 and GL3 are known to regulate anthocyanin biosynthesis in seedlings and mature plants of *Arabidopsis* (Payne et al., 2000; Feyissa et al., 2009). In the same species, two unigenes from the WD40 family showed similarity to TTG1 (Walker et al., 1999). The predicted regulatory mechanism of anthocyanin in red leaf lettuce is shown in Figure 4.

Table 3. Anthocyanin biosynthetic and regulatory genes in red leaf lettuce.

Gene type	Gene name	Gene number
Early biosynthesis gene	Chalcone synthase	8
	Chalcone isomerase	3
	Flavanone 3-hydroxylase	8
	Flavonoid 3'-hydroxylase	4
	Flavonoid 3'5'-hydroxylase	3
Late bisynthesis gene	Dihydroflavonol-4-reductase	5
	Anthocyanidin synthase/ leucoanthocyanidin dioxygenase	3
MYB regulate gene	AtMYB12-like gene	2
	AtMYB75-like gene	2
	AtMYB90-like gene	1
	AtMYB113-like gene	1
	AtMYB114-like gene	1
bHLH regulate gene	AtEGL1-like gene	2
	DvIVS-like gene	1
WD40 regulate gene	TRANSPARENT TESTA GLABRA 1-like gene (*Arabidopsis*)	2

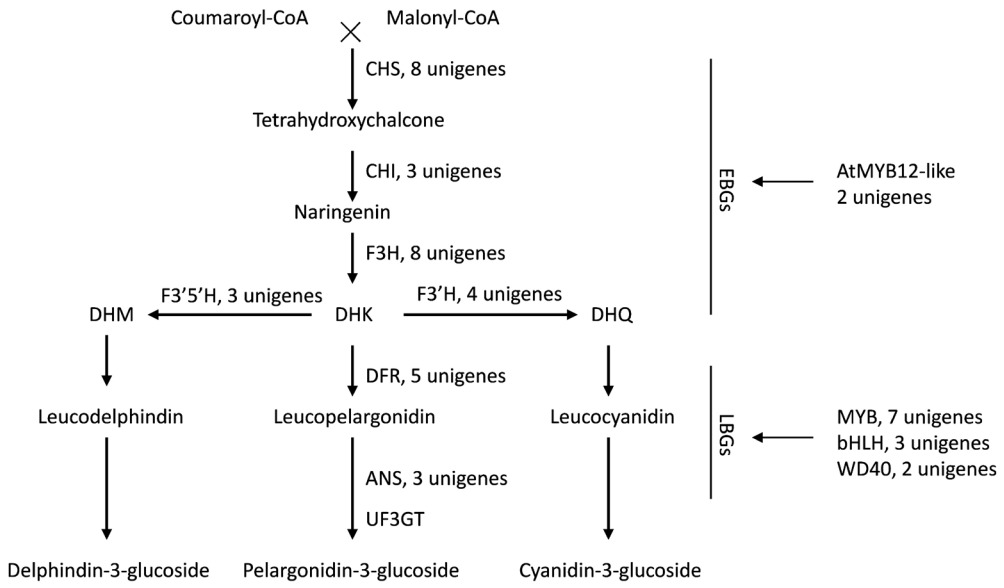

Figure 4. Simplified pathways for flavonoid biosynthesis. The number of candidate unigenes resulting from this study is shown after some enzymes and transcription factors.

SSR discovery

In total, 3607 SSRs were obtained from 2916 unigenes. A total of 569 unigene sequences contained more than one SSR and 142 SSRs were present in compound forms. As shown in Table 4, the tri-nucleotide repeat motif was the most abundant motif, accounting for 61.5% of the SSRs, followed by the di-nucleotide (34.5%), the tetra-nucleotide (1.32.6%), the penta-nucleotide (0.55%), and the hexa-nucleotide (0.47%) repeat motifs. The frequency of classified SSR repeat motifs is shown in Figure 5: the AG/CT motif was the most abundant (726, 20.1%), followed by the AG/CT (503, 13.9%), the AT/TA (453, 12.6%), the AAC/GTT (427, 11.8%), the AAG/CTT (302, 8.4%), the AAT/ATT (296,8.2%), the ACC/GGT (235, 6.5%), the ACG/CGT (158, 4.4%), and the ACT/AGT (138, 3.8%) motifs. The frequency of the remaining repeat motifs accounted for 22.6%.

Table 4. Distribution of identified SSRs using the MISA software.

Motif	Repeat time									Total	%
	5	6	7	8	9	10	11	12	17		
Di-	-	397	236	173	161	198	86	6	-	1257	34.8
Tri-	1232	659	306	18	2	-	-	-	1	2218	61.5
Tetra-	75	18	-	1	1	-	-	-	-	95	2.6
Penta-	18	2	-	-	-	-	-	-	-	20	0.55
Hexa-	11	4	1	1	-	-	-	-	-	17	0.47

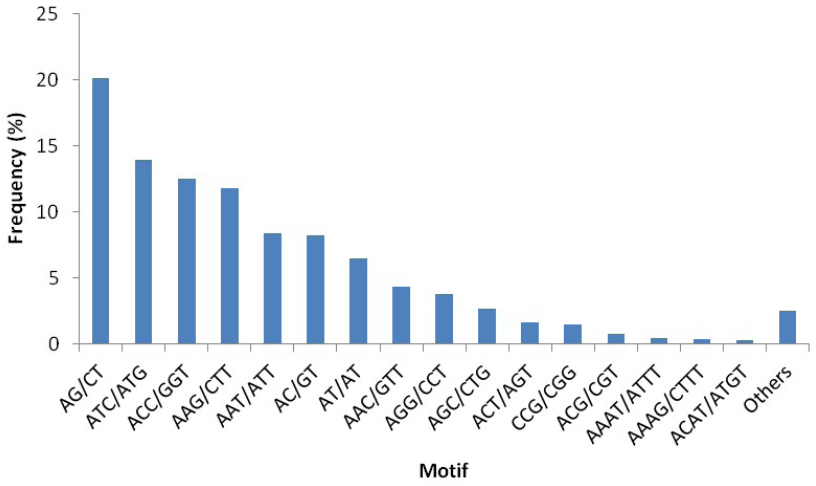

Figure 5. Frequency distribution of SSRs based on motif types. The AG/CT dinucleotide repeat motif was the most abundant detected motif.

DISCUSSION

Next generation sequencing technology is a reliable and efficient tool for transcriptome analysis in organisms without reference genome. A series of transcriptomes of colored plant tissues have been sequenced using high-throughput sequencing technology, leading to the identification of many anthocyanin-related genes, as has been done in grape hyacinth (Lou et al., 2014), *Camellia*

chekiangoleosa (Wang et al., 2014), and in the pericarp of *Litchi chinensis* (Lai et al., 2015). In this study, the transcriptome of red leaf lettuce was sequenced using an Illumina HiSeq 2500 platform. A total of 10.6 GB of data was generated and assembled into 83,335 unigenes, 54,868 of which were annotated in public databases. The anthocyanin pathway was reconstructed and a total of 12 regulatory genes and 34 anthocyanin structural genes were found. In addition, 3607 SSRs were identified in 2916 unigenes larger than 1000 bp. This study provides valuable information for future molecular research on red leaf lettuce. The reconstruction of the anthocyanin pathway and its regulatory mechanisms explained the red color of some lettuce plants. Anthocyanin-related genes may also be used to genetically modify lettuce.

In total, 12,082 unigenes that had hits in public databases were assigned to different pathways in the KEGG classification system. Two pathways, the “Flavonoid biosynthesis” (ko00941) pathway and the “Flavone and flavonol biosynthesis” (ko00944) pathway, involved in the red color of the lettuce, and eight glucoside metabolic pathways related to synthetic substrates for anthocyanin were identified. In many plants, flavonoid pathways were found in colored tissues and were mainly responsible for color variation, as seen in *Magnolia sprengeri* (Shi et al., 2014), peach flowers (Chen et al., 2014), and the pericarp of *L. chinensis* (Lai et al., 2015). These data indicate that the sequencing and assembly were successfully performed, covering many metabolic pathways. Annotations were valuable for understanding anthocyanin accumulation and are important for further studies of functional genes.

The anthocyanin pathway is important for the synthesis of plant pigments. In some plants, enzymes in the anthocyanin pathway are encoded by a multigene family, e.g., six CHS were identified in *Petunia hybrida* (Koes et al., 1989), and four DFR and four ANS were found in *Lilium* (Zhang et al., 2015). This study discovered 34 unigenes belonging to seven anthocyanin enzyme families, including 26 genes involved in EBGs and eight involved in LBGs. These genes cover the whole anthocyanin pathway. Moreover, a total of 12 anthocyanin-regulated genes, belonging to the MYB, the bHLH, and the WD40 families were identified. In plants, the MYB-bHLH-WD40 complex is critical for activating the anthocyanin biosynthesis pathway. The mutation of these regulated genes usually causes color variation in apple (Espley et al., 2007), purple cauliflower (Chiu et al., 2010), *Dahlia* (Ohno et al., 2011), and blood oranges (Butelli et al., 2012). Anthocyanin-regulated genes were found to be useful for increasing food quality. In tomatoes, two snapdragon anthocyanin-regulated genes were expressed, and their anthocyanin content was comparable to that in blackberries and blueberries (Butelli et al., 2008). In conclusion, the anthocyanin pathway and regulatory network in red leaf lettuce described in this study (Figure 5) may help understanding color-regulating mechanisms and molecular breeding in colored crops.

SSR markers are co-dominant and multi-allelic. Compared with others markers, SSR markers have many advantages such as simplicity, abundance, effectiveness, and extensive genomic coverage (Powell et al., 1996). The availability of a high-density SSR map is valuable as a public resource for many genetic studies. Based on the EST and genomic sequences, 61 EST-SSR markers and 97 genomic SSR markers have been identified (Simko, 2009; Rauscher and Simko, 2013). However, information on SSR markers is still limited, and it is, therefore, important to develop more SSR markers. In this study, a total of 3607 EST-SSR markers from the transcriptome of red leaf lettuce were identified, in which tri-nucleotide repeat motifs were the most frequent SSR motifs. The newly developed genomic SSR markers enriched the pool of previously developed markers and provide useful tools for cultivar fingerprinting, gene mapping, and selection of desirable genotypes in lettuce breeding programs.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the National Natural Science Foundation of China (#31400602 and #U1204307) and the Cultivation Foundation of Luoyang Normal University (#2013-PYJJ-001).

REFERENCES

Baslam M, Morales F, Garmendia I, Goicoechea N (2013). Nutritional quality of outer and inner leaves of green and red pigmented lettuces (*Lactuca sativa* L.) consumed as salads. *Sci. Hortic.* 151: 103-111. http://dx.doi.org/10.1016/j.scienta.2012.12.023

Baudry A, Heim MA, Dubreucq B, Caboche M, et al. (2004). TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana. Plant J.* 39: 366-380. http://dx.doi.org/10.1111/j.1365-313X.2004.02138.x

Butelli E, Titta L, Giorgio M, Mock HP, et al. (2008). Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat. Biotechnol.* 26: 1301-1308. http://dx.doi.org/10.1038/nbt.1506

Butelli E, Licciardello C, Zhang Y, Liu J, et al. (2012). Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell.* 24: 1242-1255. http://dx.doi.org/10.1105/tpc.111.095232

Chen Y, Mao Y, Liu H, Yu F, et al. (2014). Transcriptome analysis of differentially expressed genes relevant to variegation in peach flowers. *PloS One* 9: e90842. http://dx.doi.org/10.1371/journal.pone.0090842

Chiu LW, Zhou X, Burke S, Wu X, et al. (2010). The purple cauliflower arises from activation of a MYB transcription factor. *Plant Physiol.* 154: 1470-1480. http://dx.doi.org/10.1104/pp.110.164160

Cominelli E, Gusmaroli G, Allegra D, Galbiati M, et al. (2008). Expression analysis of anthocyanin regulatory genes in response to different light qualities in *Arabidopsis thaliana*. *J. Plant Physiol.* 165: 886-894. http://dx.doi.org/10.1016/j.jplph.2007.06.010

Doğan S and Salman Ü (2007). Partial characterization of lettuce (*Lactuca sativa* L.) polyphenol oxidase. *Eur. Food Res. Technol.* 226: 93-103. http://dx.doi.org/10.1007/s00217-006-0513-8

Espley RV, Hellens RP, Putterill J, Stevenson DE, et al. (2007). Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *Plant J.* 49: 414-427. http://dx.doi.org/10.1111/j.1365-313X.2006.02964.x

Feyissa DN, Løvdal T, Olsen KM, Slimestad R, et al. (2009). The endogenous GL3, but not EGL3, gene is necessary for anthocyanin accumulation as induced by nitrogen depletion in *Arabidopsis* rosette stage leaves. *Planta* 230: 747-754. http://dx.doi.org/10.1007/s00425-009-0978-3

Gonzalez A, Zhao M, Leavitt JM, Lloyd AM (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *Plant J.* 53: 814-827. http://dx.doi.org/10.1111/j.1365-313X.2007.03373.x

Grotewold E, Sainz MB, Tagliani L, Hernandez JM, et al. (2000). Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R. *Proc. Natl. Acad. Sci.* 97: 13579-13584. http://dx.doi.org/10.1073/pnas.250379897

Knekt P, Kumpulainen J, Järvinen R, Rissanen H, et al. (2002). Flavonoid intake and risk of chronic diseases. *Am. J. Clin. Nutr.* 76: 560-568.

Kobayashi S, Goto-Yamamoto N, Hirochika H (2004). Retrotransposon-induced mutations in grape skin color. *Science* 304: 982. http://dx.doi.org/10.1126/science.1095011

Koes RE, Spelt CE, van den Elzen PJ, Mol JN (1989). Cloning and molecular characterization of the chalcone synthase multigene family of *Petunia hybrida*. *Gene* 81: 245-257. http://dx.doi.org/10.1016/0378-1119(89)90185-6

Lai B, Hu B, Qin YH, Zhao JT, Wang HC, Hu GB (2015) Transcriptomic analysis of Litchi chinensis pericarp during maturation with a focus on chlorophyll degradation and flavonoid biosynthesis. *BMC Genomics* 16:225. http://dx.doi.org/10.1186/s12864-015-1433-4

Lou Q, Liu Y, Qi Y, Jiao S, et al. (2014). Transcriptome sequencing and metabolite analysis reveals the role of delphinidin metabolism in flower colour in grape hyacinth. *J. Exp. Bot.* 65: 3157-3164. http://dx.doi.org/10.1093/jxb/eru168

Ohno S, Hosokawa M, Hoshino A, Kitamura Y, et al. (2011). A bHLH transcription factor, DvIVS, is involved in regulation of

anthocyanin synthesis in dahlia (*Dahlia variabilis*). *J. Exp. Bot.* 62: 5105-5116. http://dx.doi.org/10.1093/jxb/err216

Park JS, Choung MG, Kim JB, Hahn BS, et al. (2007). Genes up-regulated during red coloration in UV-B irradiated lettuce leaves. *Plant Cell Rep.* 26: 507-516. http://dx.doi.org/10.1007/s00299-006-0255-x

Payne CT, Zhang F, Lloyd AM (2000). GL3 encodes a bHLH protein that regulates trichome development in *Arabidopsis* through interaction with GL1 and TTG1. *Genetics* 156: 1349-1362.

Powell W, Morgante M, Andre C, Hanafey M, et al. (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol. Breeding* 2: 225-238. http://dx.doi.org/10.1007/BF00564200

Rauscher G, Simko I (2013). Development of genomic SSR markers for fingerprinting lettuce (*Lactuca sativa* L.) cultivars and mapping genes. *BMC Plant Biol.* 13: 11. http://dx.doi.org/10.1186/1471-2229-13-11

Scalbert A, Williamson G (2000). Dietary intake and bioavailability of polyphenols. *J. Nutr.* 130: 2073S-2085S.

Schaart JG, Dubos C, Romero De La Fuente I, van Houwelingen AM, et al. (2013). Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (Fragaria × ananassa) fruits. *New Phytol.* 197: 454-467. http://dx.doi.org/10.1111/nph.12017

Shi SG, Yang M, Zhang M, Wang P, et al. (2014). Genome-wide transcriptome analysis of genes involved in flavonoid biosynthesis between red and white strains of *Magnolia sprengeri* pamp. *BMC Genomics* 15: 706. http://dx.doi.org/10.1186/1471-2164-15-706

Simko I (2009). Development of EST-SSR markers for the study of population structure in lettuce (*Lactuca sativa* L.). *J. Hered.* 100: 256-262. http://dx.doi.org/10.1093/jhered/esn072

Stracke R, Ishihara H, Huep G, Barsch A, Mehrtens F, Niehaus K, Weisshaar B (2007) Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the Arabidopsis thaliana seedling. *Plant J.* 50:660-677. http://dx.doi.org/10.1111/j.1365-313X.2007.03078.x

Stracke R, Jahns O, Keck M, Tohge T, Niehaus K, Fernie AR, Weisshaar B (2010) Analysis of PRODUCTION OF FLAVONOL GLYCOSIDES-dependent flavonol glycoside accumulation in Arabidopsis thaliana plants reveals MYB11-, MYB12- and MYB111-independent flavonol glycoside accumulation. *New Phytol.* 188:985-1000. http://dx.doi.org/10.1111/j.1469-8137.2010.03421.x

Toda K, Yang D, Yamanaka N, Watanabe S, et al. (2002). A single-base deletion in soybean flavonoid 3'-hydroxylase gene is associated with gray pubescence color. *Plant Mol. Biol.* 50: 187-196. http://dx.doi.org/10.1023/A:1016087221334

Walker AR, Davison PA, Bolognesi-Winfield AC, James CM, et al. (1999). The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in *Arabidopsis*, encodes a WD40 repeat protein. *Plant Cell* 11: 1337-1350. http://dx.doi.org/10.1105/tpc.11.7.1337

Wang ZW, Jiang C, Wen Q, Wang N, et al. (2014). Deep sequencing of the *Camellia chekiangoleosa* transcriptome revealed candidate genes for anthocyanin biosynthesis. *Gene* 538: 1-7. http://dx.doi.org/10.1016/j.gene.2014.01.035

Winkel-Shirley B (2001). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.* 126: 485-493. http://dx.doi.org/10.1104/pp.126.2.485

Yuan Y, Ma X, Tang D, Shi Y (2014). Comparison of anthocyanin components, expression of anthocyanin biosynthetic structural genes, and TfF3¢H1 sequences between *Tulipa fosteriana* 'Albert heijn' and its reddish sport. *Sci. Hortic.* 175: 16-26. http://dx.doi.org/10.1016/j.scienta.2014.05.032

Zhang MF, Jiang LM, Zhang DM, Jia GX (2015). De novo transcriptome characterization of *Lilium* 'Sorbonne' and key enzymes related to the flavonoid biosynthesis. *Mol. Genet. Genomics* 290: 399-412. http://dx.doi.org/10.1007/s00438-014-0919-0

Supplementary material

Table S1. KEGG pathway of unigenes.

http://www.geneticsmr.com/year2016/vol15-1/pdf/gmr7023_supplementary.xls