

Transcriptome analysis identifies the key genes responsible for high anthocyanin content in the fruits of *Lycium ruthenicum* Murray

Xuebin Zhu^{1,2,3,*}, Jianmin Li^{3,*}, Yuan Zong^{1,2,*}, Xuemei Sun^{2,*}, Shiming Li¹, Dong Cao^{1,2}, Bo Zhang¹, Wenjie Chen¹ and Baolong Liu^{1,**}

¹Qinghai Province Key Laboratory of Crop Molecular Breeding, Xining 810008, China

²State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, Xining 800010, China

³College of Biologic and Geographic Sciences, Qinghai Normal University, Xining 810008, China

Lycium ruthenicum Murray (black wolfberry) is an important economic plant producing black fruit with high anthocyanin content. The transcriptome of its fruits was compared with that of *Lycium barbarum* L. (wolfberry), which produces orange-red fruits, with its pigmentation depending largely on carotenoids, with the aim of identifying the key genes responsible for anthocyanin accumulation through RNA sequencing. A total of 32.05 and 28.52 Gb clean reads was obtained after filtering in *L. barbarum* and *L. ruthenicum* respectively. Altogether 192,869 unigenes were assembled with an average length of 1064 bp. These unigenes were predicted to encode 152,209 specific proteins with the help of protein databases. Compared with *L. barbarum*, 733,070 genes were upregulated while 25,779 genes appeared downregulated in the fruits of *L. ruthenicum*. The genes of the anthocyanin biosynthesis pathway exhibited more differences between the two species than did those of any other biosynthetic pathway. All structural genes in connection with anthocyanin biosynthesis had higher expression level in *L. ruthenicum* than in *L. barbarum*, except *F3'H* and *3GT*. The downregulation of *F'H* and *3GT* in *L. ruthenicum* would be responsible for the absence of cyanidin and glycosylation in this species. The *MYB* and *bHLH* genes regulating anthocyanin biosynthesis also displayed higher transcript levels in *L. ruthenicum* than in *L. barbarum*, especially the *MYB* transcription factor gene, which should be the reason for the activation of the anthocyanin biosynthesis structural genes. More work should be carried out to isolate the *MYB* and *bHLH* transcription factor genes, and to confirm their functions in producing the pigments found in the black fruits of *L. ruthenicum*.

Keywords: Anthocyanin biosynthesis, key genes, transcriptome analysis, wolfberry.

LYCIUM RUTHENICUM Murray, a traditional Chinese herb mentioned in the Tibetan medical classic *Jing Zhu Ben*

*Contributed equally to this work.

**For correspondence. (e-mail: blliu@nwipb.cas.cn)

Cao, has been used for thousands of years in Northwest China for heart disease, abnormal menstruation and menopause¹. The specific physiological features of drought resistance and salt tolerance also render it an ideal plant for preventing soil desertification and withstanding soil salinity and alkalinity; these are significant features for maintaining the ecosystem and agriculture in remote areas^{1,2}. Anthocyanin content can reach 3.1% (w/w) in the dried fruits of *L. ruthenicum*. Petunidins take up 95% of the total anthocyanin, while delphinidins and malvidins comprise the remaining 5% (ref. 1). Anthocyanins have been suggested to confer antiinflammatory, antimutagenic, anticarcinogenic and antibacterial properties as well as hepatotoxicity and the induction of apoptosis in humans³⁻⁵. Pharmacological experiments have proved that the fruits have antifatigue, antihypoglycaemic and antioxidant properties^{1,6,7}. Though the chemical components and pharmacological properties have been studied by some researchers, the molecular basis of high anthocyanin content in the fruits of *L. ruthenicum* remains unclear.

The anthocyanin biosynthesis pathway has been studied in terms of model plants for decades. Anthocyanin biosynthesis involves first producing *p*-coumaric acyl-CoA from phenylalanine in three enzymatic steps. One molecule of *p*-coumaric acyl-CoA plus three molecules of propylene acyl-CoA in the presence of chalcone synthase (CHS) generate chalcone. Under the catalysis of chalcone isomerase (CHI), chalcone is quickly isomerized to naringenin (flavanone), which in turn is converted into dihydrokaempferol by flavonoid-3-hydroxylase (F3H). Then dihydrokaempferol is converted by two enzymes – flavonoid-3'-hydroxylase (F3'H) and flavonoid-3',5'-hydroxylase (F3'5'H) – to dihydroquercetin and dihydromyricetin, independently. F3H, F3'H and F3'5'H belong to the cytochrome P450 superfamily. The products of these three hydroxylases, F3H, F3'H and F3'5'H, are the direct precursors in the synthesis of anthocyanin. Leucoanthocyanidin dioxygenase (LDOX) catalyses the conversion of colourless leucoanthocyanidins to coloured anthocyanidins, and hence it is transformed from a white

to a coloured flower^{8–10}. Then, different glycosylation enzymes begin to operate, generating diversified anthocyanin. In general, the anthocyanin biosynthesis pathway is regulated by V-myb avian myeloblastosis viral oncogene homolog (MYB), basic helix–loop–helix (bHLH) and WD40 proteins^{10,11}.

Comparative analysis of anthocyanin accumulation has been carried out in *L. ruthenicum* fruits through real-time polymerase chain reaction (RT-PCR), but genes in relation to anthocyanin biosynthesis were assembled from the transcriptome database of a leaf¹², which implies that some genes expressed specifically in the fruits would be missed. Moreover, *Lycium barbarum* L. and *L. ruthenicum* belong to different species of the genus. Therefore, it is necessary to make a comparison between the sequences of structural genes and regulatory factors in the fruits of *L. barbarum* and *L. ruthenicum*, and then compare their transcript levels. Next-generation sequencing technology (NGS) has proved to be a powerful and economical tool for genome sequencing, genome resequencing, mRNA expression profiling and DNA methylation analysis. Currently, *de novo* transcriptome sequencing has been widely employed in light of functional gene discovery in non-model organisms for which reference genome information was lacking¹³. In addition, this approach has also made possible the analysis of gene expression profiling with the quantification of short cDNA reads^{14–16}. In the present study, RNA sequencing (RNA-seq) using NGS has been employed to compare the transcriptome differences in *L. ruthenicum* and *L. barbarum* fruits so as to identify the key genes responsible for high anthocyanin content in *L. ruthenicum* fruits.

Methods

Plant material

L. barbarum cv. Ningqi 7 and *L. ruthenicum* cv. LMH1 are the type cultivars widely planted in Qinghai, China. The two cultivars were maintained as germplasm resources in the Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, Qinghai, China. The fruits were collected 10 days after flowering (Figure 1) and



Figure 1. Fruits of *Lycium barbarum* Ningqi 7 (a) and *L. ruthenicum* LMH 1 (b) 10 days after flowering.

stored at -80°C prior to mRNA extraction. Two mRNA samples were extracted from the fruits of each species. Total RNA was extracted from the fruits employing the Tiangen RNAprep Pure Plant Kit (Tiangen, China) in accordance with the manufacturer's protocol. Electrophoresis on a 1.0% (w/v) agarose gel checked the quality of total RNA, while the concentration of total RNA was determined by NanoDrop (Thermo Scientific, Wilmington, DE, USA).

Illumina sequencing

The four cDNA libraries (two from each of the two *Lycium* species) of the LMH1 and Ningqi 7 fruits were prepared according to the manufacturer's instructions for mRNA-Seq sample preparation (Illumina Inc, San Diego, CA, USA). The cDNA library products were analysed using Illumina paired-end sequencing technology with read lengths of 100 bp, and a Illumina HiSeq 2000 instrument (BGI Technologies Co, Ltd, Shenzhen, China).

Sequence data filtering, *de novo* assembly and annotation

The original data from the images were assembled with Illumina GA Pipeline 1.6 by removing low-quality reads (reads with unknown sequences *N*, adaptor sequence fragments and empty reads). Next, *de novo* assembly of the transcriptome into unique-sequence unigenes was processed using Trinity software, a short-read assembly program¹⁷. In short, Trinity integrates reads with certain lengths of overlap to form longer fragments or contigs in the first place. Then the reads are mapped back to contigs, and with paired-end reads, it becomes possible to detect contigs from the same transcript as well as the distance between these contigs. Finally, Trinity connects the contigs, and generates sequences that can be extended on either end. Such sequences are defined as unigenes¹⁷. Later, a Basic Local Alignment Search Tool (BLAST) BLASTx alignment (*e*-value < 0.00001) was carried out on the unigenes from protein databases, involving NCBI non-redundant protein(Nr), NCBI (National Center for Biotechnology Information) nucleotide sequence (Nt), Swiss-Prot, Kyoto Encyclopaedia of Genes and Genomes (KEGG), Cluster of Orthologous Groups of proteins (COG), Interpro and Gene Ontology (GO). Also, the best alignments were employed to determine the sequence direction of the unigenes. If the results from the different databases showed a conflict between one another, an order with priority of Nr, Swiss-Prot, KEGG and COG was followed to decide the sequence direction of the unigenes. When a unigene was not involved in any of the above-mentioned databases, the software ESTScan was employed to calculate its coding regions and decide its sequence direction¹⁸. In the final step, with Nr annotation,

Table 1. Details of sequencing, filtering and assembling

Sample	Sequencing and filtering				Unigene assembling			
	Total raw reads (Mb)	Total clean reads (Mb)	Total clean bases (Gb)	Clean reads Q20 (%)	Total number	Total length	Mean length	N50
Black1	116.78	116.74	17.51	97.67	100,553	99,760,831	992	1733
Black2	96.95	96.91	14.54	97.81	120,811	110,271,124	912	1584
Red1	87.59	87.56	13.13	97.65	77,156	72,271,150	936	1581
Red2	102.62	102.59	15.39	97.75	86,031	80,804,362	939	1606
Total	403.94	403.80	60.57	97.72	192,869	205,220,696	1064	1831

Table 2. Annotation summary of predicted *Lycium* protein numbers from various databases

Parameters	Total	Nr	Nt	Swiss-Prot	KEGG	COG	Interpro	GO	Overall
Number	192,869	101,079	133,750	63,305	72,719	40,270	62,303	18,541	152,209
Percentage	100	52.41	69.35	32.82	37.70	20.88	32.30	9.61	78.92

the Blast2GO program was used to obtain the GO and KEGG annotations for the unigenes¹⁹. After the GO annotation was obtained for each unigene, WEGO software was used to categorize them in light of their functions and to determine the distribution of gene functions in the species at the macro level²⁰. A Venn diagram of the annotation proteins was drawn using the on-line software Venny (<http://bioinfogp.cnb.csic.es/tools/venny/index.html>).

Analysis of differential gene expression

To map the differential gene expression (DGE) reads, the sequenced raw data were filtered to remove low-quality reads. For annotation, clean read sequences were mapped to the transcriptome reference database, with no more than a two-nucleotide mismatch allowed^{17,21}. Gene coverage refers to the percentage of a gene covered by reads. This value is identical to the ratio of base number in a gene covered by unique mapping reads to the total base number of that gene. Concerning gene expression analysis, the number of expressed sequences was calculated and normalized to the number of reads per Kb per million reads (RPKM).

To compare the distinctions in gene expression, the read frequency in each DGE library was analysed using a statistical approach according to the method of Audic and Claverie²². We used a false discovery rate (FDR) <0.001 as well as an absolute value of the log 2 ratio >1 as threshold to judge the statistical significance of the gene expression differences²³. Next, the genes expressed differentially were subjected to GO and KEGG Ontology (KO) enrichment analysis²⁴. The genes related to anthocyanin biosynthesis in KEGG pathways were collected and aligned to the unigenes from a transcriptome mixture of the fruits from Ningqi 7 and LMH 1, by means of the BLASTX algorithm with an *E*-value of < 1e⁻⁵. The selected genes were: phenylalanine ammonia-lyase (*PAL*),

cinnamate 4-hydroxylase (*C4H*), 4-hydroxycinnamoyl-CoA ligase (*C4L*), *CHS*, *CHI*, *F3H*, *F3'H*, *F3'5'H*, *DFR*, *LDOX*, UDP-glucose flavonoid 3-glucosyltransferase (*3GT*), *MYB* and *bHLH*.

Results

Sequencing and sequence assembly

RNA-seq enables rapid access to all transcripts from biological tissues, in a way which is efficient and convenient. After sequencing with Illumina HiSeq 2500, 403.94 million reads were obtained in the original data from the four samples and 60.57 GB (Phred quality score Q20 > 97.72%) of valid data was acquired after removing low-quality reads. Trinity software was used to link the valid data to 192,869 unigenes with a total of 205,220,696 bp. The average length of each gene was 1064 bp and the length of *N50* was 1831 (Table 1).

Summary of annotation results

BlastX software was employed to annotate the 192,869 assembly unigenes employing Nr, Nt, Swiss-Prot, KEGG, COG, Interpro and GO protein databases, which predicted 101,079, 133,750, 63,305, 72,719, 40,270, 62,303 and 18,541 proteins respectively (Table 2). A total of 152,209 proteins were obtained after removing the repeated ones.

Among the predicted *L. ruthenicum* proteins, 27.31% had the highest homology to *Solanum tuberosum*, 18.83% to *Nicotiana glauca*, and 16.40% to *Nicotiana glauca*, while 12.12% had the highest homology to *Solanum lycopersicum* and 25.35% had the closest genetic relationship to other species (Figure 2). Both *L. barbarum* and *L. ruthenicum* are part of the Solanaceae

family to which potato, tobacco and tomato also belong, and the predicted proteins were highly homologous to these species.

The number of common proteins existing in both Black 1 and Black 2 (LMH 1) and Red 1 and Red 2 (Ningqi 7) was 52,143, accounting for almost one-third of the 152,209 *Lycium* proteins. There were 25,891 species-specific genes in *L. ruthenicum*, compared to 16,636 expressed in *L. barbarum* (Figure 3). These *Lycium*-specific genes should include those determining the production of different pigments in fruits of *L. ruthenicum* and *L. barbarum*.

Analysis of differentially expressed genes

To identify the differentially expressed unigenes between Ningqi 7 and LMH 1, putative differentially expressed unigenes were identified on basis of the fragments per kilobase million (FPKM) values calculated from the read counts mapped onto the reference transcriptome. A total of 58,849 unigenes were differentially presented between

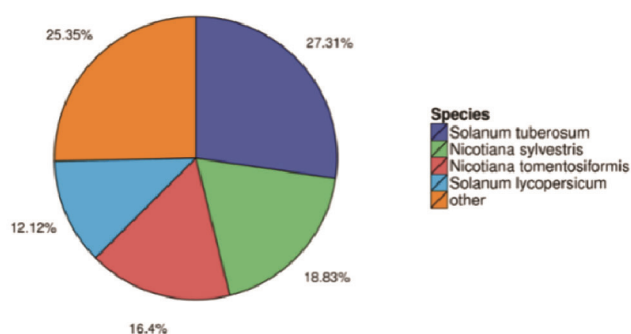


Figure 2. Percentage homology between predicted proteins of *L. ruthenicum* and those of other species.

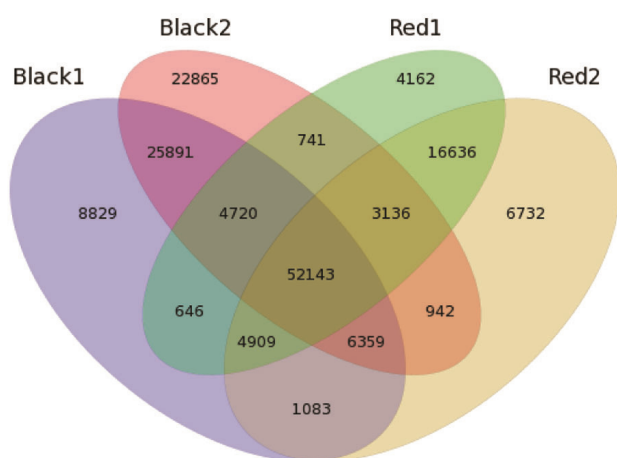


Figure 3. Venn diagram of the annotation proteins from *L. ruthenicum* and *L. barbarum* fruits. Black 1 and Black 2 are the fruits of LMH 1, while Red 1 and Red 2 are the fruits of Ningqi 7.

Ningqi 7 and LMH 1, according to the comparison of expression levels with $FDR \leq 0.001$ and $|\log_2 \text{ratio}| \geq 1$ (Figure 4). The highest \log_2 ratio value reached 14.58, while the lowest value was -13.16 . Using Ningqi 7 as a reference, 33,070 upregulated unigenes (with higher levels of expression in LMH 1) and 25,779 downregulated unigenes (with higher levels of expression in Ningqi 7) were identified. Significantly more unigenes were upregulated in LMH 1 than in Ningqi 7.

The genes were classified into three classes. Red genes were upregulated if gene expression of LMH 1 was higher than that in Ningqi 7. Blue genes were downregulated if gene expression of Ningqi 7 was higher than that in LMH 1. Black genes were not differentially expressed. The horizontal coordinates in Figure 4 refer to the average FPKM value (A) and the vertical coordinates refer to the \log_2 ratio (M).

GO functional classification of differentially expressed unigenes

GO analysis assigned 58,849 differential expression unigenes (DEGs) to 53 subcategories. Among them, 20,520 DEGs were involved in biological processes; 19,734 DEGs were relevant to cellular components and 10,358 unigenes were grouped under molecular functions. Within the category of 'biological process', most DEGs were related to 'metabolic processes' (5125), 'cellular processes' (4952) and 'single-tissue processes' (3310). Within the category of 'cellular component', most DEGs were enriched in the subcategories 'cell process' (4353), 'cell components' (4316) and organelles (3198). Within the category of 'molecular function', most parts of the DEGs were involved in 'binding' (4374) and 'catalytic activity' (4182) (Figure 5).

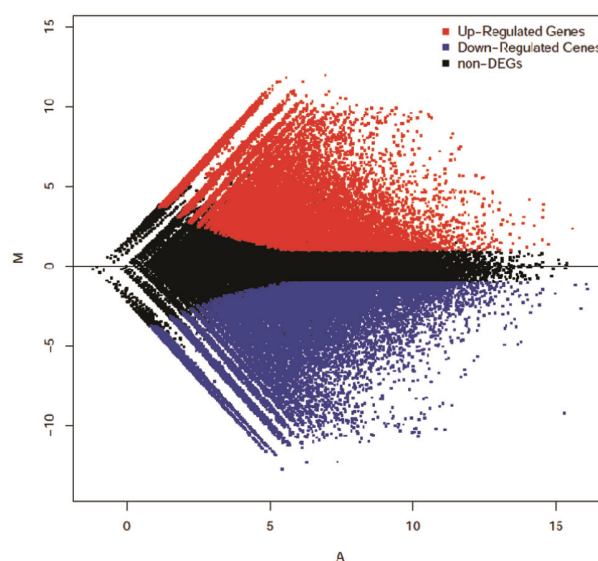


Figure 4. Differentially expressed genes between Ningqi 7 and LMH 1.

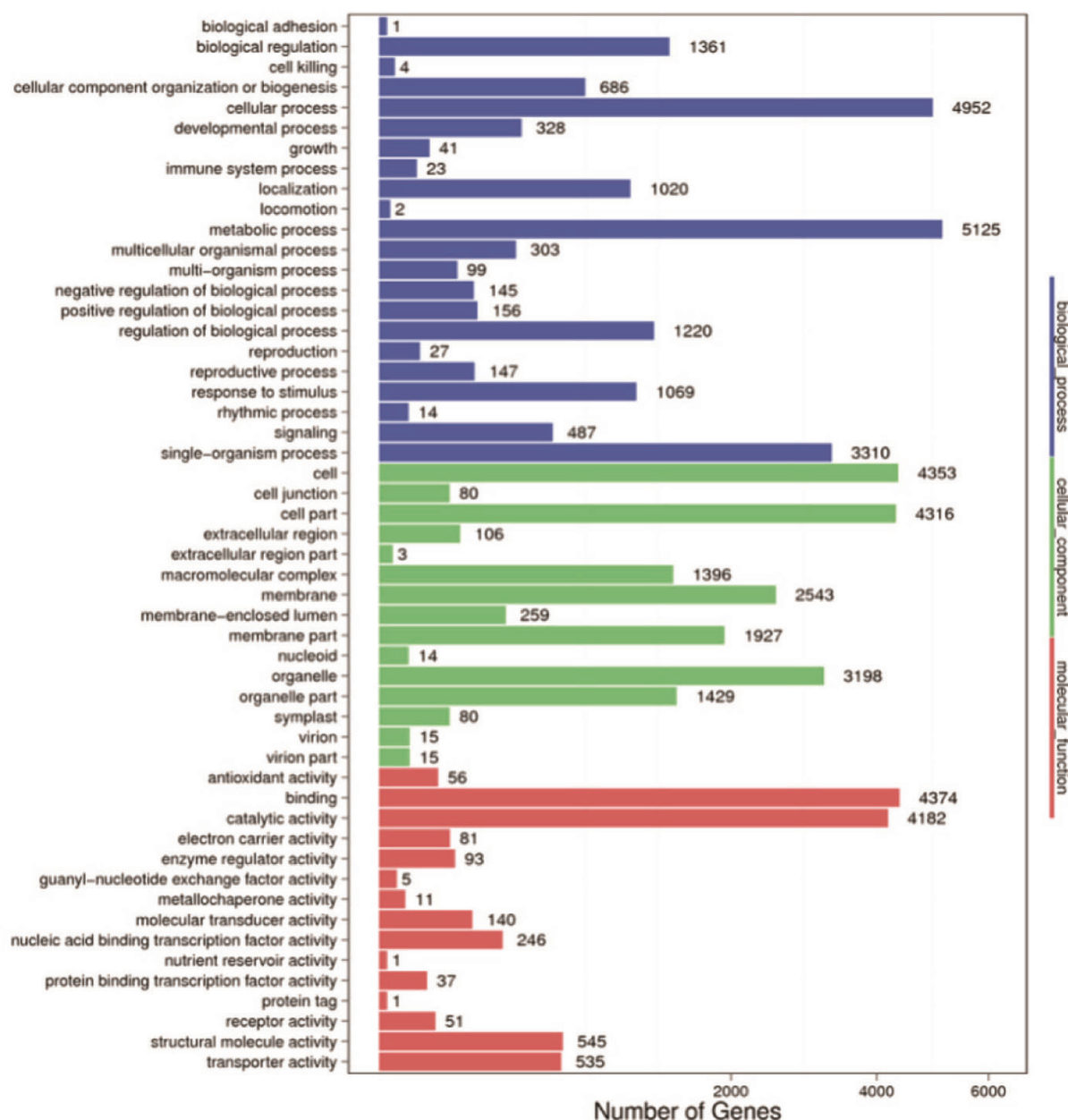


Figure 5. Gene ontology classifications of DEGs.

Unigenes were divided into three categories, namely cellular components, molecular functions and biological processes.

KEGG pathway mapping of differential expression unigenes

A total of 72,719 DEGs were annotated and 36,607 were mapped onto 136 pathways in the KEGG database. Majority of the DEGs were involved in metabolic pathways (8151), biosynthesis of secondary metabolites

(4525), plant-pathogen interaction (1697) and RNA transport (1693). However, these pathways contain many unigenes, which indicates that they do not exhibit the largest number of changes. According to the *Q* value, the most differentially expressed pathway was plant hormone signal transduction (1157 DEGs/2007 unigenes), followed by circadian rhythm-plant (332/559), flavonoid biosynthesis (220/367), anthocyanin biosynthesis (51/71) and carotenoid biosynthesis (214/359) (Supplementary Table 1). Flavonoid, anthocyanin and carotenoid biosynthesis were listed as the pathways with the largest number of changes.

Table 3. Difference in the expression of structural and regulatory genes of the anthocyanin biosynthesis pathway

Gene	KO	Gene ID	Length (bp)	Red-expression	Black-expression	log 2 ratio (Black/Red)	P-value	FDR	Up/down regulation (blue/white)
<i>PAL</i>	K10775	CL6881.Contig1	2708	131.51	563.29	1.590793	0.000214537	0.000887349	Up
		CL6881.Contig2	2723	120.73	761.2	2.128302	6.45E-07	4.07E-06	Up
<i>C4H</i>	K00487	Unigene100773	2220	0	40.31	7.297634	0.00091533	0.003277126	Up
		Unigene45017	2290	0	349.81	10.10562	1.16E-07	8.11E-07	Up
<i>C4L</i>	K01904	CL8771.Contig5	2073	13.28	178.18	3.005863	6.83E-09	5.45E-08	Up
		CL8771.Contig2	2406	0	105.61	8.544641	2.14E-05	0.000107253	Up
		CL8771.Contig3	1982	20.24	496.15	4.092271	1.78E-08	1.36E-07	Up
		Unigene37862	2578	82.39	3499.92	4.634542	3.16E-27	7.86E-26	Up
		CL8771.Contig1	2213	580.43	2640.49	1.590103	5.22E-06	2.89E-05	Up
		CL8771.Contig4	2230	600.23	2589.4	1.502139	3.90E-05	0.000186296	Up
		CL19587.Contig2	2629	1.5	142.43	5.803996	7.61E-10	6.71E-09	Up
		CL19587.Contig3	2076	1	157.07	6.469835	2.85E-10	2.61E-09	Up
<i>CHS</i>	K00660	CL7019.Contig1	1537	8.55	8401.06	9.173604	2.84E-82	3.98E-80	Up
		CL7019.Contig2	4647	110.7	64330.32	8.31399	2.15E-112	5.87E-110	Up
<i>CHI</i>	K01859	Unigene499	1193	27.65	630.65	3.740279	1.60E-21	3.03E-20	Up
<i>F3H</i>	K00475	Unigene37743	1561	709.3	29554.01	4.603313	2.33E-51	1.40E-49	Up
<i>F3'H</i>	K05280	–	–	–	–	–	–	–	–
<i>F3'5'H</i>	K13083	Unigene36915	2000	26.7	32138.07	9.360939	2.69E-117	8.37E-115	Up
<i>DFR</i>	K13082	Unigene36873	1489	25.29	23348.23	8.97309	8.76E-104	2.01E-101	Up
<i>LDOX</i>	K05277	CL17790.Contig1	1468	14.96	13533.64	8.990812	1.23E-116	3.76E-114	Up
		CL17790.Contig4	1657	45.8	28223.72	8.446119	3.35E-142	1.55E-139	Up
		CL17790.Contig2	2764	496.36	21001.04	4.618735	9.06E-56	6.21E-54	Up
		CL17790.Contig3	1626	14.91	3771.26	7.283609	2.56E-120	8.38E-118	Up
<i>3GT</i>	K12930	–	–	–	–	–	–	–	
<i>MYB</i>	K09422	CL10341.Contig1	1844	8.64	3919.22	8.123761	4.96E-76	6.00E-74	Up
		CL24514.Contig5	1089	302.44	1132.66	1.186741	1.05E-06	6.42E-06	Up
		CL10341.Contig3	1010	8.19	3958.99	8.123849	7.14E-85	1.07E-82	Up
		CL10341.Contig2	2338	0	2375.29	12.40634	8.42E-13	9.41E-12	Up
<i>MYC</i>	K13422	CL22487.Contig1	2727	2346.4	1359.92	-1.44669	1.31E-11	1.34E-10	Down
		CL15739.Contig1	2435	0	360.92	9.976324	1.32E-07	9.12E-07	Up
		CL15739.Contig2	2391	184.51	1.79	-7.2648	1.55E-16	2.23E-15	Down
		CL15739.Contig3	2388	5.91	166.78	3.83313	7.40E-06	4.00E-05	Up

Note: Black was the fruits of *L. ruthenicum* LMH 1 and red are the fruits of *L. barbarum* Ningqi7.

Expression profile of genes associated with anthocyanin biosynthesis

In the transcriptomes from Ningqi 7 and LMH 1, 31 unigenes were found to be associated with anthocyanin biosynthesis. Though 13 genes relevant to anthocyanin biosynthesis were employed in the Blastx analysis, only *PAL*, *C4H*, *C4L*, *CHS*, *CHI*, *F3H*, *F3'5'H*, *DFR*, *LDOX*, *MYB* and *bHLH* have homologous genes in the assembly unigene database (Table 3). Genes homologous to *F3'H* and *3GT* were not found in the unigene database from Ningqi 7 and LMH 1. All the anthocyanin biosynthesis structural genes had higher expression levels in LMH 1 than in Ningqi 7 (Table 3), which indicates that anthocyanin biosynthesis was higher in LMH 1. *MYB* was expressed more in LMH1 than in Ningqi 7, with the highest log2 ratio value 12.40634 in CL10341.Contig2 (Table 3). For *bHLH*, two unigenes showed higher expression levels in LMH 1 than Ningqi 7, while two unigenes displayed higher expression levels in Ningqi 7 than in LMH1.

Discussion

L. ruthenicum and *L. barbarum* belong to the *Lycium* genus of the Solanaceae family, which shows that they are closely related. As a result, one-third of the transcripts of genes from ripe fruits are identical with Ningqi 7 and LMH 1. Thus RNA-seq is appropriate for comparison of different expressions of species-specific genes during fruit development of *L. barbarum* and *L. ruthenicum*. Generally, *L. ruthenicum* accumulates higher anthocyanin content in the fruits, while *L. barbarum* contains higher levels of carotenoids (primarily astaxanthin, beta-carotene and zeaxanthin).

In transcriptome analysis, the KEGG pathways, viz. plant hormone signal transduction, circadian rhythm-plant, flavonoid biosynthesis, anthocyanin biosynthesis, and carotenoid biosynthesis showed the greatest differences between the two *Lycium* species. The plant hormone signal transduction pathway covers the signal transduction pathways of jasmonate, salicylate, auxin,

brassinosteroid, cytokinin, gibberellin, abscisic acid and ethylene. These plant hormones are believed to regulate fruit development and shape^{25–31}. The circadian rhythm–plant pathway belongs to environmental adaptation. The two pathways are not directly related to secondary metabolism. It is possible that they can indirectly regulate the anthocyanin or carotenoid biosynthesis. Anthocyanins belong to the flavonoid family of compounds. The high expression level in the flavonoid and anthocyanin biosynthesis pathway should be the reason for high anthocyanin content in the fruits of *L. ruthenicum* cv. LMH 1. Changes in the carotenoid biosynthesis pathway are responsible for the high carotenoid content in fruits of *L. barbarum* cv. Ningqi 7.

Previous studies have suggested the presence of malvidin, petunidin and delphinidin in the fruits of *L. ruthenicum*, but not cyanidin. Petunidins comprised 95% of the total anthocyanin in the fruits of *L. ruthenicum*, while delphinidin and malvidin comprised the remaining 5% (ref. 1). In the anthocyanin biosynthesis process, F3H catalyses flavanone to dihydrokaempferol. The dihydrokaempferol is hydroxylated by F3'H to produce dihydroquercetin, which is the precursor of cyanidin. The dihydrokaempferol could also be hydroxylated, by F3'5'H, to form dihydromyricetin, the precursor of delphinidin, petunidin and malvidin¹². F3'H has not been detected in the assembly unigenes, while the F3'5'H presented a high expression level (32138.07 RPKM) in the transcriptome analysis, which could be the reason for the high petunidin content but lack of cyanidin in the fruit of *L. ruthenicum* cv. LMH (ref. 1). Moreover, most of the anthocyanins in *L. ruthenicum* are acylated by coumaric acid¹. Consequently, the 3GT gene, which achieves acylation of anthocyanin by glucosides, could not be detected in the assembly unigenes database.

All structural genes of anthocyanin biosynthesis had higher transcription rates in LMH 1 than in Ningqi 7 in transcriptome analysis. It is well known that the MYB and bHLH genes are regulators of the entire anthocyanin biosynthesis pathway, and their transcription can lead to the wholesale expression of anthocyanin biosynthesis structural genes^{10,11}. The low level of transcripts of MYB and bHLH in *L. barbarum* could explain the low expression of structural genes associated with anthocyanin biosynthesis in Ningqi 7. The functional loss of MYB or bHLH can impede the whole anthocyanin biosynthesis pathway, and cause pale fruit phenotypes in apple, peach, and tomato^{32–34}. In this case, more experiments are necessary to decide the key regulatory gene for high anthocyanin content in *L. ruthenicum*.

Conclusion

The transcriptome of its fruits was compared with that of *Lycium barbarum* L., wolfberry. All structural genes re-

lated to anthocyanin biosynthesis had higher expression level in *L. ruthenicum* than in *L. barbarum*, except F3'H and 3GT. The MYB and bHLH genes regulating anthocyanin biosynthesis also had higher transcript levels in *L. ruthenicum* than in *L. barbarum*, especially the MYB transcription factor gene. Therefore, the low level of transcripts of MYB and bHLH in *L. barbarum* should be the reason for the activation of the anthocyanin biosynthesis structural genes.

Competing interests: The authors declare that they have no competitive interests.

- Zheng, J. *et al.*, Anthocyanins composition and antioxidant activity of wild *Lycium ruthenicum* Murr. from Qinghai–Tibet Plateau. *Food Chem.*, 2011, **126**(3), 859–865.
- Zhang, H., Xia, L.-I. and Wang, J., The structure characteristic of the plant community in the lower reaches of Tarim River. *Ecol. Environ.*, 2007, **4**, 1219–1224.
- Bowen-Forbes, C.-S., Zhang, Y.-J. and Nair, M.-G., Anthocyanin content, antioxidant, anti-inflammatory and anticancer properties of blackberry and raspberry fruits. *J. Food Compos. Anal.*, 2010, **23**(6), 554–560.
- Wang, L.-S. and Stoner, G.-D., Anthocyanins and their role in cancer prevention. *Cancer Lett.*, 2008, **269**(2), 281–290.
- Mazza, G., Bioactivity, absorption and metabolism of anthocyanins. In Proceedings of the First International Symposium on Human Health Effects of Fruits and Vegetables, 2007, pp. 117–125.
- Feng, W., En-Peng, H.-E. and Chen, X.-Q., Study on the effect of *Lycium ruthenicum* Murray fruit polysaccharide on the athletic ability of mice and the dose-effect. *Arid Zone Res.*, 2010, **26**(4), 586–590.
- Wang, J.-H., Chen, X.-Q. and Zhang, W.-J., Study on hypoglycemic function of polysaccharides from *Lycium ruthenicum* Murr. fruit and its mechanism. *Food Sci.*, 2009, **30**, 244–248.
- Zhang, Y., Butelli, E. and Martin, C., Engineering anthocyanin biosynthesis in plants. *Curr. Opin. Plant Biol.*, 2014, **19**, 81–90.
- Saito, K. *et al.*, The flavonoid biosynthetic pathway in *Arabidopsis*: structural and genetic diversity. *Plant Physiol. Biochem.*, 2013, **72**, 21–34.
- Nishihara, M. and Nakatsuka, T., Genetic engineering of novel flower colors in floricultural plants: recent advances via transgenic approaches. *Methods Mol. Biol.*, 2010, **589**, 325–334.
- Xu, W., Dubos, C. and Lepiniec, L., Transcriptional control of flavonoid biosynthesis by MYB–bHLH–WDR complexes. *Trends Plant Sci.*, 2015, **20**(3), 176–185.
- Zeng, S. *et al.*, Comparative analysis of anthocyanin biosynthesis during fruit development in two *Lycium* species. *Physiol. Plant.*, 2014, **150**(4), 505–516.
- Liu, D. *et al.*, Transcriptome analysis of purple pericarps in common wheat (*Triticum aestivum* L.). *PLoS ONE*, 2016, **11**(5), e0155428.
- Zhang, N. *et al.*, Transcriptome characterization and sequencing-based identification of drought-responsive genes in potato. *Mol. Biol. Rep.*, 2014, **41**(1), 505–517.
- Dorn, K.-M., Fankhauser, J.-D., Wyse, D.-L. and Marks, M.-D., *De novo* assembly of the pennycress (*Thlaspi arvense*) transcriptome provides tools for the development of a winter cover crop and biodiesel feedstock. *Plant J.*, 2013, **75**(6), 1028–1038.
- Firon, N. *et al.*, Transcriptional profiling of sweet potato (*Ipomoea batatas*) roots indicates down-regulation of lignin biosynthesis and

- up-regulation of starch biosynthesis at an early stage of storage root formation. *BMC Genomics*, 2013, **14**(1), 460.
17. Grabherr, M.-G. *et al.*, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnol.*, 2011, **29**(7), 644–652.
 18. Iseli, C., Jongeneel, C.-V. and Bucher, P., ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. In ISMB, 1999, pp. 138–148.
 19. Conesa, A., Götts, S., Garcíagómez, J.-M., Terol, J., Talón, M. and Robles, M., Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 2005, **21**(18), 3674–3676.
 20. Ye, J. *et al.*, WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* (Web Server issue), 2006, **34**, W293.
 21. Mortazavi, A., Williams, B.-A., Mccue, K., Schaeffer, L. and Wold, B., Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, 2008, **5**(7), 621.
 22. Audic, S. and Claverie, J.-M., The significance of digital gene expression profiles. *Genome Res.*, 1997, **7**(10), 986–995.
 23. Benjamini, Y. and Yekutieli, D., The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.*, 2001, **29**(4), 1165–1188.
 24. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. and Hattori, M., The KEGG resource for deciphering the genome. *Nucleic Acids Res.* (Database issue), 2004, **32**, 277–280.
 25. Katsir, L., Chung, H.-S., Koo, A. J.-K. and Howe, G.-A., Jasmonate signaling: a conserved mechanism of hormone sensing. *Curr. Opin. Plant Biol.*, 2008, **11**(4), 428–435.
 26. Johnson, C., Boden, E. and Arias, J., Salicylic acid and NPR1 induce the recruitment of trans-activating TGA factors to a defense gene promoter in *Arabidopsis*. *Plant Cell*, 2003, **15**(8), 1846–1858.
 27. Kim, T. W. and Guan, S.-Y., Brassinosteroid signal transduction from cell-surface receptor kinases to nuclear transcription factors. *Nature Cell Biol.*, 2009, **11**(10), 1254–1260.
 28. Kim, T. W. and Wang, Z.-Y., Brassinosteroid signal transduction from receptor kinases to transcription factors. *Annu. Rev. Plant Biol.*, 2010, **61**, 681–704.
 29. Chapman, E.-J. and Estelle, M., Mechanism of auxin-regulated gene expression in plants. *Annu. Rev. Genet.*, 2009, **43**, 265–285.
 30. Mason, M. G. *et al.*, Multiple type-B response regulators mediate cytokinin signal transduction in *Arabidopsis*. *Plant Cell*, 2005, **17**(11), 3007–3018.
 31. Chen, Y.-F., Etheridge, N. and Schaller, G.-E., Ethylene signal transduction. *Ann. Bot.*, 2005, **95**(6), 901–915.
 32. Qiu, Z., Wang, X., Gao, J., Guo, Y., Huang, Z. and Du, Y., The tomato Hoffman's anthocyaninless gene encodes a bHLH transcription factor involved in anthocyanin biosynthesis that is developmentally regulated and induced by low temperatures. *PLoS ONE*, 2016, **11**(3), e0151067.
 33. Tuan, P. A. *et al.*, The crucial role of PpMYB10.1 in anthocyanin accumulation in peach and relationships between its allelic type and skin color phenotype. *BMC Plant Biol.*, 2015, **15**(1).
 34. Takos, A.-M., Jaffé, F.-W., Jacob, S.-R., Bogs, J., Robinson, S.-P. and Walker, A.-R., Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. *Plant Physiol.*, 2006, **142**(3), 1216–1232.

ACKNOWLEDGEMENTS. This research was financially supported by the Qinghai Science and Technology Department (2018-NK-133, 2018-ZJ-T08, 2019-ZJ-982Q), West Light Talent Program of the Chinese Academy of Sciences, the Open Project of State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University (2018-KF-06), China.

Received 21 October 2017; accepted 4 October 2018

doi: 10.18520/cs/v116/i2/256-263