biologia plantarum
an international journal for experimental botany

BIOLOGIA PLANTARUM (2023) 67: 271-284 DOI: 10.32615/bp.2023.031

Screening of transcription factors related to flower and fruit development by differential gene analysis in *Lycium* species

J. ZHAO[+], K. LI[+], Z.H. ZHANG, Y.Z. XU, D.W. CHEN, and K. SUN[*]

College of Life Science, Northwest Normal University, Lanzhou 730070, P.R. China

[*]Corresponding author: E-mail: kunsun@nwnu.edu.cn

Abstract

Lycium barbarum Thunb. and *Lycium ruthenicum* Murray (wolfberries) have been utilized as traditional medicinal and nutritional plants in China for centuries. Much research has been focused on their high quality, yet the molecular mechanisms underlying morphological differences remain unclear. In this study, a comparative analysis of morphological and cytological characteristics indicated that significant differences existed. Meanwhile, transcriptomic analyses of the flower and fruit were performed at different developmental stages, and a total of 54 795 differentially expressed genes (DEGs) were screened. Gene Ontology and Kyoto Encyclopedia of Genes and Genomes analyses showed that these DEGs were significantly enriched in substance metabolism, catalytic activity, single organism process, starch and sucrose metabolism, carotenoid biosynthesis, amino sugar and nucleotide sugar metabolism, phenylpropanoid biosynthesis, and other pathways. Based on these significantly enriched pathways, the ratio between nonsynonymous and synonymous substitution rates (Ka/Ks), and numerous studies related to flower and fruit development, we preliminarily screened eight transcription factor families related to flower and fruit development and counted the number of potential transcription factor genes. These candidate genes could provide a basis for future functional verification, helping to further research on the molecular mechanism of morphological differences in the two *Lycium* species.

Keywords: flower and fruit development, *Lycium barbarum*, *Lycium ruthenicum*, transcriptome analysis, transcription factor, wolfberry.

Introduction

Lycium barbarum and *Lycium ruthenicum* belong to the *Lycium* genus within the *Solanaceae* family. These plants have been recognized for their traditional medicinal and edible properties in China for centuries, playing crucial roles in protecting liver function, lowering blood sugar and lipids content, enhancing immunity, and combating cancer (Zheng *et al*. 2011, Tian *et al*. 2022). Currently, research on the two kinds of wolfberry mainly focuses on the analysis of their nutritional components and pharmacological studies, little research on their morphological differences can be found, which necessitates a deeper understanding of the molecular mechanisms regulating flower and fruit development.

The process of flower and fruit development in plants is regulated by numerous genes (Robles and Pelaz 2005, Giovannoni 2007, Karlova *et al*. 2014, Lopez-Ortiz *et al*. 2021). Many studies have found that some special transcription factors could precisely regulate the fruit morphology, such as *fw2.2* (Beauchet *et al*. 2021), *fw11.3* (Huang and van der Knaap 2011), *WUS* (van der Graaff

Received 21 April 2023, *last revision* 18 August 2023, *accepted* 25 August 2023.
Abbreviations: DEGs - differentially expressed genes; GO - Gene Ontology; KEGG - Kyoto Encyclopedia of Genes and Genomes; TF - transcription factor.
Acknowledgments: We thank Lili Wang and Xingnan Du for providing support in the preliminary work. This research was supported by the National Natural Science Foundation of China Regional Science Foundation Project (32060060), the Gansu Provincial Science and Technology Program Project (21JR1RA231), and the Young Teachers' Scientific Research Ability Improvement Plan of Northwest Normal University (NWNU-LKQN2021-30).
[+]These authors contributed equally.
Conflict of interest: The authors declare that they have no conflict of interest.
Data availability statement: Transcriptome sequencing data are available in the SRA database of the National Center for Biotechnology Information (NCBI) under the accession number PRJNA941957.

et al. 2009), *OVATE* (Liu *et al*. 2002), *SUN* (Xiao *et al*. 2008), *FAS* (Cong *et al*. 2008), *LC* (Rodríguez *et al*. 2011), *POS1* (Wang *et al.* 2014).

In addition to the above special transcription factor family genes, there are also many large TF family genes playing an important and broad-spectrum role in flower and fruit development. Firstly, C2H2 transcription factors are involved in regulating flower organs, fruit ripening, and softening (Jiao *et al*. 2020, Lai *et al*. 2022). MYB transcription factors play a crucial role in anther and pollen development, and some have significant effects on flowering time and flower color (Ferrario *et al*. 2006, Ballester *et al*. 2010, Dubos *et al*. 2010). Transcription factors bHLH have been reported to regulate flower color and fruit ripening time, such as *AmbHLH1/2* and *CmbHLH32* (Albert *et al*. 2021, Tan *et al*. 2021). Transcription factors bZIP have been found to influence flower meristem and regulate fruit growth and ripening, such as *SlPAN* and *SlTGA2.2* (Lemaire-Chamley *et al*. 2022, Zhang *et al*. 2022). WRKY transcription factors are also involved in affecting flowering and pollen development (Zhang *et al*. 2018, Wang *et al*. 2019). The B3 transcription factor family plays an important role in the entire life process of plant growth and development, including flowering induction (Ruan *et al*. 2021). Lastly, NAC transcription factors are thought to be involved in the regulation of fruit ripening (Nieuwenhuizen *et al*. 2021, Liu *et al*. 2022).

Nowadays, transcriptome sequencing has been widely recognized as an effective method of analyzing complicated pathways and gene expression networks. Here, we selected *L. barbarum* and *L. ruthenicum* flowers and fruits from different growth and development stages as materials, and conducted transcriptome analysis to preliminarily identify significantly enriched pathways and differentially expressed transcription factor genes. This information could provide a reference for further research on the transcription factors related to flower and fruit development in these two wolfberry species.

Materials and methods

Plant materials: Flowers and fruits of *Lycium barbarum* Thumb. and *L. ruthenicum* Murray at different growth stages were collected from Caiqi Township, Minqin County, Wuwei City (E102.748460, N38.221910), located in the northwest of Gansu province, during June and July. Flowers and fruits from the two wolfberry species were collected at 15 developmental stages each (Tables 1,2 Suppl.). After collection, they were immediately frozen in liquid nitrogen and stored at -80°C. For *L. barbarum*, RNA samples from its 15 developmental stages were equally mixed, and three biological replicates were named A-1, A-2, and A-3, respectively. Similarly, for *L. ruthenicum*, RNA samples from the 15 developmental stages were equally mixed, and three biological replicates were named B-1, B-2, and B-3, respectively.

Morphological and cytological characteristics analysis: In the flower and fruit morphology of *L. barbarum* and *L. ruthenicum*, collected from different stages, existed significant differences, especially in fruits. Fruit longitudinal and transverse diameters were measured by using a digital vernier caliper (*LR44 AG13*, Hengliang, China), and the mean values of the fruit diameter were calculated. The fruits were fixed in FAA fixative (70% ethanol: formalin: acetic acid, 18:1:1) for 24 h, and then dehydrated with a series of ethanol concentrations (75, 85, 95, 100%). Subsequently, the dehydrated fruits were transferred to a mixture (xylene: ethanol; 1:1) for 30 min and xylene for 1 h, respectively, and embedded in paraffin. Then, the longitudinal and transverse sections of fruits were cut into slices with 8 μm thickness by using a rotary microtome (*Leica RT2235*, Barcelona, Spain), and the slices were stained. Finally, the well-stained longitudinal and transverse sections of fruits were sealed and selected for photographing.

RNA quantification and qualification: RNA samples from the 15 developmental stages were extracted using a total RNA extraction kit (*DP441*, *Tiangen*, China), and evenly mixed. A total of 3 μg mixed RNA sample was used as input material for RNA sample preparations. RNA degradation and contamination were monitored on 1% agarose gels. RNA purity was checked using the *NanoPhotometer*® spectrophotometer (*IMPLEN*, München, Germany). RNA concentration was measured using the *Qubit*® RNA assay kit in *Qubit*® *2.0* Fluorometer (*Life Technologies*, USA). RNA integrity was assessed using the RNA *Nano 6000* assay kit of the *Agilent Bioanalyzer 2100* system (*Agilent Technologies*, Santa Clara, CA, USA).

Table 1. The summary of data output quality. A-1, A-2, and A-3 represent three biological replicates of mixed RNA samples of *L. barbarum*. B-1, B-2, and B-3 represent three biological replicates of mixed RNA samples of *L. ruthenicum*.

Sample	Raw reads	Clean reads	Clean bases	Error [%]	Q20 [%]	Q30 [%]	GC [%]
A-1	68 442 114	65 104 854	9.77 G	0.01	97.77	94.33	42.54
A-2	58 755 678	55 264 280	8.29 G	0.01	97.82	94.41	42.48
A-3	68 244 328	64 814 950	9.72 G	0.01	97.83	94.47	42.43
B-1	72 042 116	70 163 714	10.52 G	0.01	98.34	95.83	42.15
B-2	56 879 448	54 184 334	8.13 G	0.01	97.71	94.2	42.05
B-3	59 641 720	56 522 544	8.48 G	0.01	97.74	94.25	42.12

Table 2. A statistical table of transcription factors selected from differential expression genes of *L. barbarum* and *L. ruthenicum*. The number of up types represents the number of highly expressed genes in *L. barbarum* and the number of down types represents the number of highly expressed genes in *L. ruthenicum*.

TF family	Up	Down
C2H2	5	4
MYB	6	21
bHLH	1	14
B3	10	12
MADS	12	8
WRKY	4	13
NAC	19	8
bZIP	4	4

Library preparation and transcriptome sequencing: The *NEBNext® Ultra™ RNA Library Prep Kit* for *Illumina®* (*New England Biolabs*, Ipswich, MA, USA) was used to generate sequencing libraries following the manufacturer's recommendations. Library fragments were purified with the *AMPure XP* system (*Beckman Coulter*, Beverly, USA) to select cDNA fragments of 150~200 bp in length. Size-selected and adaptor-ligated cDNA fragments were generated using 3 µl *USER Enzyme* (*New England Biolabs*, Ipswich, MA, USA) at 37°C for 15 min followed by 5 min at 95°C before PCR. Then PCR was performed with *Phusion High-Fidelity* DNA polymerase, Universal PCR primers, and Index (X) Primer. PCR products were purified (*AMPure XP* system) and library quality was assessed on the *Agilent Bioanalyzer 2100* system (*Agilent Technologies*). The clustering of the index-coded samples was performed on a *cBot Cluster Generation* System by using *TruSeq PE Cluster Kit v3-cBot-HS* (*Illumina*) according to the manufacturer's instructions. After cluster generation, library preparations were sequenced on an *Illumina Hiseq 2500* platform (*Illumina*), and paired-end reads were generated.

Sequencing quality assessment: Raw data (raw reads) in fastq format were initially processed using in-house Perl scripts, and clean data (clean reads) were obtained by removing reads containing adapters, reads containing ploy-N, and low-quality reads from raw data. Simultaneously, Q20, Q30, GC-content, and sequence duplication levels of the clean data were calculated. All downstream analyses were based on high-quality clean data. Upon completing sequencing, the rationality of sample selection and the reliability of sequencing results needed to be verified through the overall quality assessment of RNA-seq. The quality assessment method involved using the Pearson correlation coefficient of gene expressions among different samples, aided by the *R package* (www.r-project.org/).

Transcriptome assembly and function annotation: Transcriptome assembly was accomplished based on the left.fq and right.fq using *Trinity* software (*version r20140413p1*), with min_kmer_cov set to 2 by default and all other parameters set to default (Grabherr *et al*. 2011). The online *Blastx* tool was used to compare unigenes with several databases, including the *NCBI* non-redundant protein sequence database (*NR*, https://www.ncbi.nlm.nih.gov/nuccore/), *NCBI* non-redundant nucleotide sequence database (*NT*, http://www.ncbi.nlm.nih.gov/protein/), Protein family (*InterPro*, https://www.ebi.ac.uk/interpro/) (Finn *et al*. 2014), Clusters of Orthologous Groups of proteins (*COG*, https://www.ncbi.nlm.nih.gov/research/COG) (Tatusov *et al*. 2000), a manually annotated and reviewed protein sequence database (*Swiss-Prot*, http://www.uniprot.org/) (Bairoch and Apweiler 2000), *KEGG* Ortholog database (*KO*, http://www.genome.jp/kegg/ko.html), and Gene Ontology (*GO*, http://www.geneontology.org/) (Gene Ontology Consortium 2004).

Identification of differential expression genes (DEGs): Clean data from *L. barbarum* and *L. ruthenicum* were separately mapped back onto the assembled transcriptome. The readcount for each gene was obtained from the mapping results. The input data of gene differential expression is the readcount data obtained from gene expression analysis. Gene expressions were estimated by using *RSEM* (*v. 1.3.3*) with bowtie2 parameter mismatch set to 0 (Li and Dewey 2011). Genes with an adjusted *P*-value (padj) < 0.05, as determined by the *DESeq2 R package* (*v. 1.10.1*), were designated as differentially expressed (Anders and Huber 2010). The padj was generated using the Benjamini and Hochberg (1995) method.

***GO* enrichment analysis and *KEGG* enrichment analysis:** The *clusterProfiler R package* was employed for both *GO* enrichment analysis and *KEGG* enrichment analysis of differentially expressed genes. The *GO* enrichment analysis initially mapped all DEGs to various terms in the *GO* database, calculated the number of genes per term, and then identified significant enrichment of the DEGs compared to the entire genomic background. To better study the function of differential genes, enrichment analysis of all DEGs in each combination was performed, as well as enrichment analysis of DEGs in each combination according to up-regulation or down-regulation, respectively. *KEGG* enrichment analysis (https://www.genome.jp/kegg/pathway.html) was based on the *KEGG* Pathway (Kanehisa *et al*. 2008), and hypergeometric testing was utilized to identify pathways with significant DEGs enrichment relative to all annotated genes.

Ka/Ks analysis: Comparative transcriptome analysis was used for Ka/Ks analysis. Linear homologous genes were identified using *OrthoMCL* (Li *et al*. 2003, Pei *et al*. 2020). *PAML codeml* was utilized to determine which of these one-to-one linear homologous genes belonged to synonymous substitution and which belonged to nonsynonymous substitution. Ka/Ks calculation was performed using the *PAML* package with default settings (Xu and Yang 2013). In genetics, Ka/Ks or dN/dS represents the ratio between nonsynonymous substitution

rate (Ka) and synonymous substitution rate (Ks). This ratio can be used to determine whether there is selective pressure on a protein-coding gene. *KOBAS* software was employed to test the statistical enrichment of divergent and conserved gene orthologous groups in *KEGG* pathways (Mao *et al*. 2005).

Screening of differentially expressed TFs related to flower and fruit development: In this study, candidate transcription factor family genes were selected from the seven functional annotation databases based on differentially expressed genes and identified by the online tool *iTAK* (http://itak.feilab.net/cgi-bin/itak/index.cgi) (Zheng *et al*. 2016). Since DEGs were screened based on the condition of *P*-adjusted < 0.05, the number of candidate transcription factor family genes was relatively large, and the expressions of some DEGs were not significant. Generally, $\log_2$(fold change) ≥ 2 and $\log_2$(fold change) ≤ -2 were defined as significant upregulation and downregulation between the two groups, respectively (Singh *et al*. 2014, Cao *et al*. 2021).

Results

Morphological and cytological characteristics analysis: From the 15 stages of flower and fruit development of *L. barbarum* and *L. ruthenicum*, there are significant differences in their morphology, especially of the fruit, including color, size, and volume (Fig. 1). Therefore, we conducted morphological and cytological comparative analyses on the differences between the two kinds of fruits.

Firstly, we conducted longitudinal and transverse diameter measurements of these two types of fruits. By measuring the longitudinal and transverse diameters of the fruits, it was not difficult to find that there was no significant difference in size between them in the early stages of development. However, there were significant differences in size in the later stages of development, and the longitudinal diameter growth of *L. barbarum* was always greater than that of *L. ruthenicum*. On the contrary, the transverse diameter of *L. ruthenicum* was always greater than that of *L. barbarum*. These differences led to significant differences in shape between the two, with the former presenting an ellipsoidal shape, while the latter was approximately spherical (Fig. 2*A*,*B*).

The process of fruit growth and development was actually the process of fruit cell division and extension. Based on the transversal and longitudinal section slices of two types of fruits, the main stages of fruit cell division and extension were analyzed. Randomly were selected 4 stages (G1, G3, G5, G7) from the 10 developmental stages of fruits of the *L. barbarum* and *L. ruthenicum*, respectively, and paraffin sectioning was conducted at two angles: transversal and longitudinal (Fig. 2*C*). It can be concluded that the growth and development of *L. barbarum* in the early stage was mainly due to cell division, which increases the number and density of cells. In the later stage, cell extension was mainly responsible, leading to the expansion of fruit volume. For the fruit of *L. ruthenicum* (Fig. 2*D*), it can be seen that in the early stage of growth and development, fruit cells mainly grew through cell division, but the change in fruit size was not significant, and the cell density inside the fruit gradually increased. Conversely, the main reason for the expansion of the fruit in the later stage was due to the extension of the cells.

To further explore the genetic factors and molecular mechanisms of the morphological variation of the flowers and fruits of the two kinds of *Lycium* species, we conducted transcriptome analysis on flowers and fruits of the two kinds of *Lycium* species, to preliminarily screen the candidate genes related to the development of the two kinds of flowers and fruits. Although the *Lycium* species genome has been sequenced, its annotation information has not yet been released. We selected the second generation of *de-novo* assembly transcriptome for subsequent analysis.

Sequencing quality assessment: In general, the error rate for a single base position in the measured data should not exceed 1%. The sequence error rate distribution in this experiment was entirely within the acceptable range of sequencing error rate (Fig. 3*A*). The GC content distribution test is an effective method to detect whether AT and GC separation exists in sequencing data. In theory, the number of base pairs should correspond to the total sequencing result. Except for the first 6 bp, the base content distribution of other reads at each location was stable without AT or GC separation, and the GC content was maintained within the normal range, at approximately 42% (Fig. 3*B*). Additionally, both Q20 and Q30 were above 97.7 and 94.2%, respectively (Table 1). These results indicated that the quality of transcriptome sequencing and the accuracy of the obtained reads were both very high in this experiment.

Sequencing data filtering and assembly: The *Illumina HiSeq 2500* sequencing platform was used to perform transcriptome sequencing of flowers and fruits at different stages of growth and maturity. The sequencing quality was relatively high, and the data volume was abundant. A total of 195 442 120 (3 biological replicates, including A-1, A-2, A-3) and 188 563 284 (3 biological replicates, including B-1, B-2, B-3) raw reads were obtained from *L. barbarum* and *L. ruthenicum* sequencing, respectively. Similarly, a total of 185 184 084 (including A-1, A-2, A-3) and 180 870 592 (including B-1, B-2, B-3) high-quality clean reads were obtained for *L. barbarum* and *L. ruthenicum*, respectively, after removing low-quality reads and other reads with connectors. Only clean reads were used in the following analysis. The clean reads data have been deposited in the *NCBI Sequence Read Archive* (http://www.ncbi.nlm.nih.gov/sra/) and the SRA accession number is PRJNA941957.

After assembling these reads using *Trinity* software (*version r20140413p1*), a total of 551 009 transcripts and 221 105 unigenes were obtained, with N50 values of 1 102 and 906, respectively, and average lengths of 745 bp and 637 bp, respectively. In general, the number of assembled long fragments and the assembly quality are proportional

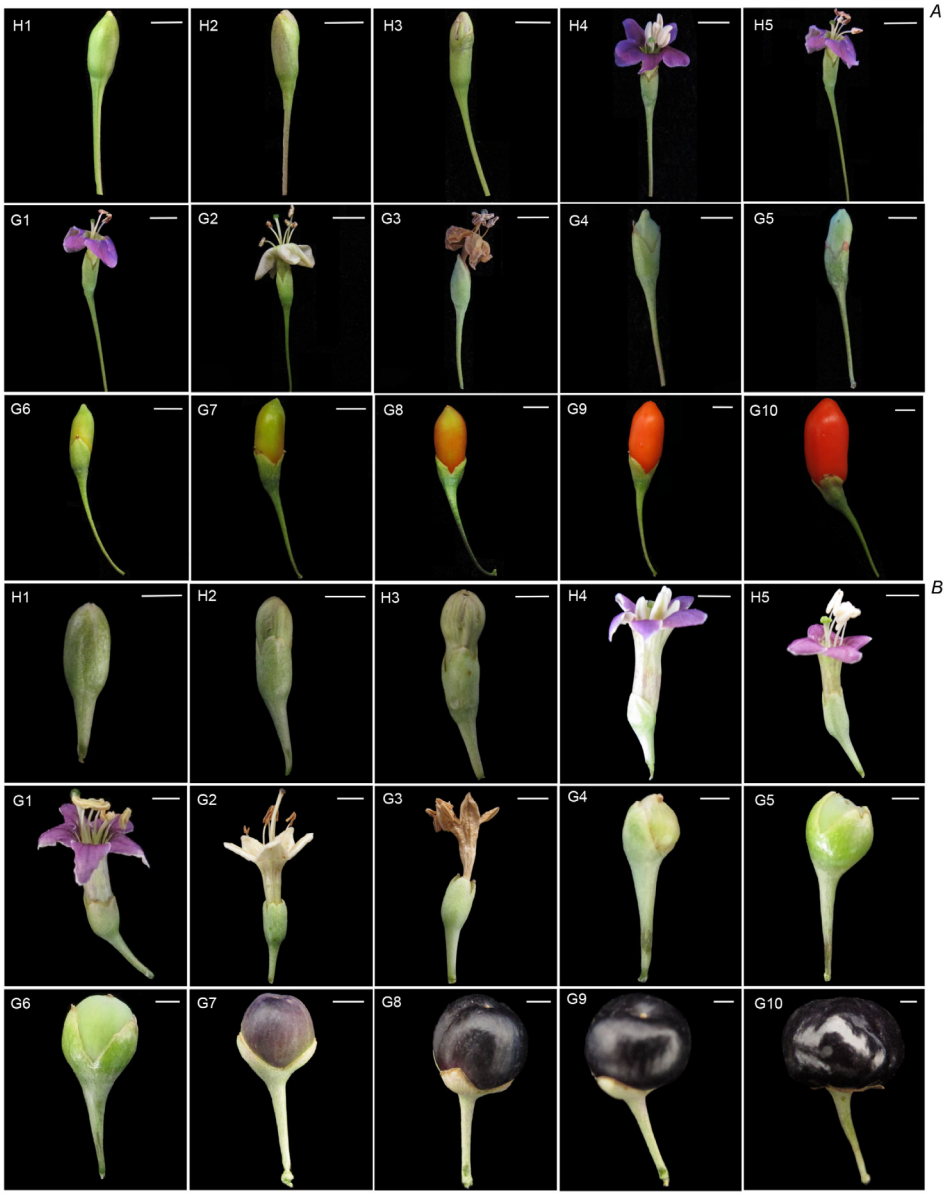

Fig. 1. Experimental materials and the specific sampling standards. *A* - The specific sampling standards for *L. barbarum*. *B* - The specific sampling standards for *L. ruthenicum*. H1 - H5 represent five developmental stages of flowers, G1 - G10 represent ten developmental stages of fruits. *Scale bar* = 2.5 mm.

to the N50 value. The quantity and distribution of transcripts and unigenes revealed that transcripts were predominantly 200 - 1 000 bp long, accounting for 83.34% of the transcripts (Tables 3,4 Suppl.). There were 285 932 and 135 410 transcripts and unigenes, respectively, in the 200 - 500 bp range, accounting for 51.89 and 61.24%, respectively. The number of 500 - 1 000 bp sequences was 138 279 and 48 878, with percentages of 25 and 22.1%. The histogram more intuitively reflects the relationship between spliced transcripts and the frequency of the unigenes length distribution (Fig. 3*C*). Most transcripts and unigenes were found to be 200 - 1 000 bp in length, while only 16.65% of transcripts were longer than 1 000 bp. These results indicated that the transcriptome library in this study is of good quality and its length can meet the requirements of basic transcriptome analysis, making it suitable for subsequent analysis and information mining.

Gene function annotation: Out of 221 105 unigenes, 133 471 unigenes were annotated, accounting for 60.36% of the total, as per the results of comparing all unigenes with seven databases using *BLAST* (Fig. 3*E*). Among them, the number of annotated unigenes in the *NT* database was the largest, with 106 522, accounting for 48.17% of the total. The number of annotated unigenes in the *KOG* database was 16 168 (7.31%). The number of annotated unigenes in the *GO* database was 53 735 (24.3%). The number of annotated unigenes in other databases ranged between 50 000 and 110 000. However, there are still 87 634 (39.63%) unannotated unigenes with

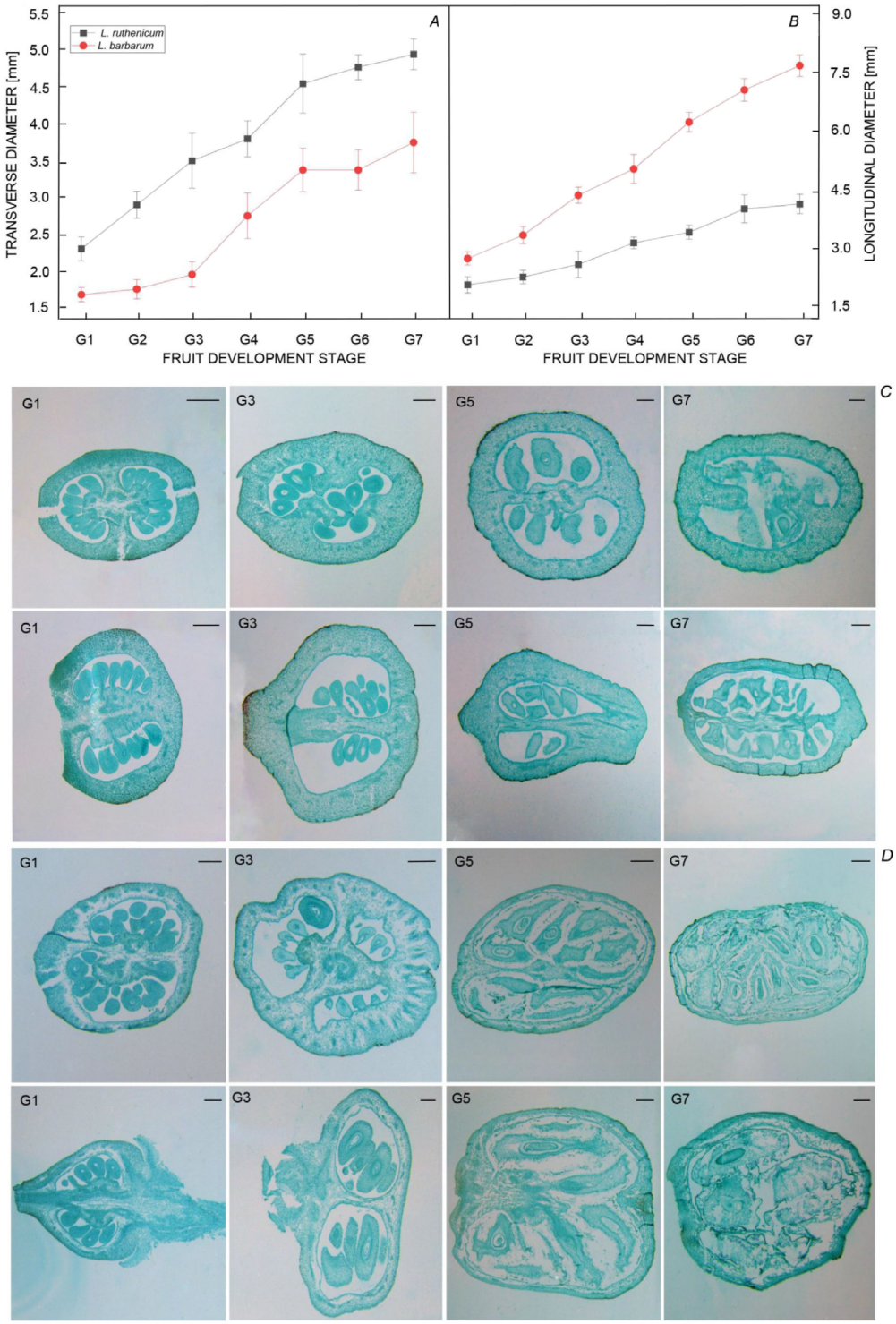

Fig. 2. Comparison of fruit morphology and cytology between two *Lycium* species at different developmental stages. *A* - The transverse diameters of two *Lycium* species. *B* - The longitudinal diameters of two *Lycium* species. *C* - Transversal and longitudinal section slices of *L. barbarum*. *D* - Transversal and longitudinal section slices of *L. ruthenicum*. *Scale bar* = 450 μm.

unknown functions, which can be considered as new genes. A Venn diagram of the gene annotation results was generated using five databases selected from the seven database annotation results (Fig. 3*D*).

Sample correlation test: The reliability of this experiment is relatively high, as demonstrated by the analysis of the correlation of gene expression levels between samples (Fig. 4*A*). The correlation coefficient of the same sample is

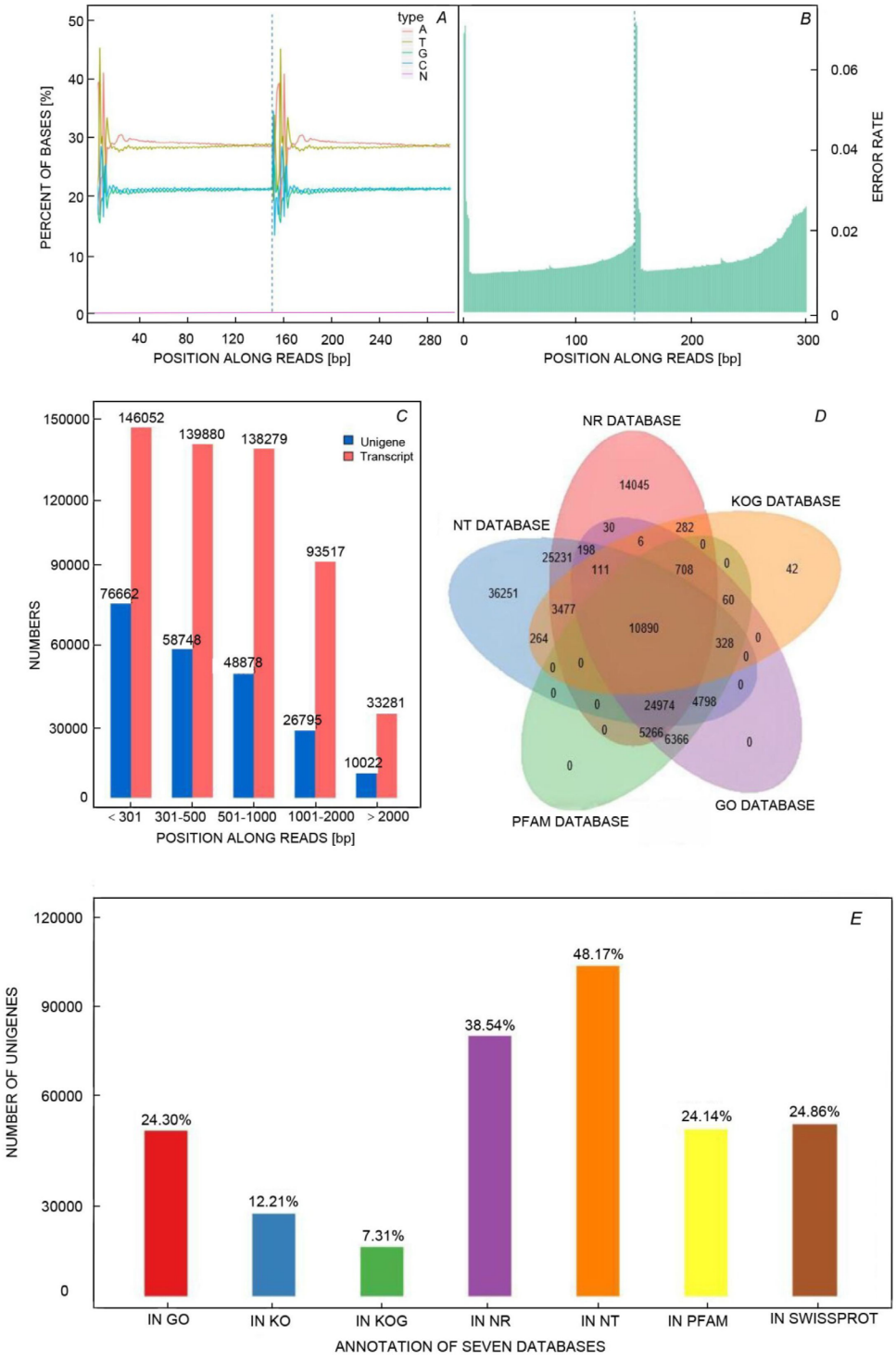

Fig. 3. Sequencing quality assessment, sequencing data filtering and assembly, and annotation of gene function. *A* - Distribution of sequencing error rate in this experiment, the three replicates of *L. barbarum* were named A-1, A-2, and A-3. *B* - Base content distribution map of this experiment. *C* - Gene transcript/gene sequence length distribution diagram. *D* - Venn graph of gene annotation result produced by five databases selected from seven database annotation results. *E* - The results of comparing all unigenes with 7 large databases using the *BLASTx*.

1, which is consistent with common sense. The correlation between biological replicates A-1, A-2, and A-3 of *L. barbarum* was relatively high, with the correlation coefficient being about 0.84. The correlation coefficient of biological replicates B-1, B-2, and B-3 in *L. ruthenicum* was about 0.80. This indicates high homogeneity between

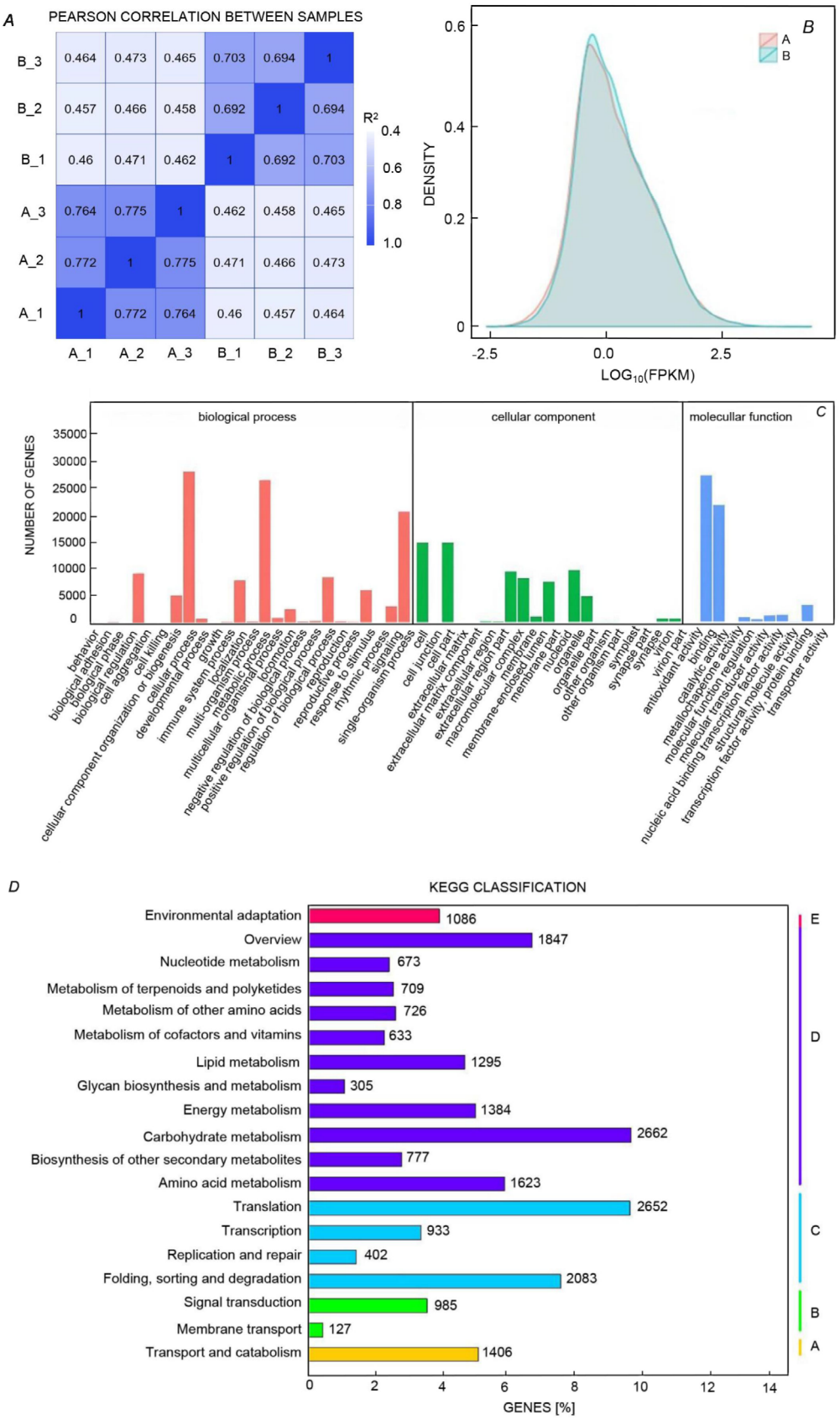

Fig. 4. Annotation of gene function and gene expression level analysis. *A* - Correlation analysis of gene expression between samples. *B* - FPKM density distribution diagram. *C* - *GO* function classification diagram. *D* - *KEGG* classification diagram.

samples. For non-biological replicates, the correlation coefficient between *L. barbarum* and *L. ruthenicum* is about 0.46, indicating that there are many differentially expressed genes between *L. barbarum* and *L. ruthenicum*. These results further suggest that the transcriptome information obtained in this experiment can be used to screen DEGs.

Differentially expressed genes analysis: The mapping rate was above 64% after aligning the clean reads of each sample to the reference genome (Table 5 Suppl.). Density distribution maps and box plots of FPKM for two types of wolfberry genes were generated to examine the overall distribution of FPKM for *L. barbarum* and *L. ruthenicum*. The gene expressions of *L. barbarum* and *L. ruthenicum* were not significantly different, with *L. ruthenicum* being slightly higher than *L. barbarum* (Fig. 4*B*). A total of 54 795 DEGs were detected in this study. Of these DEGs, compared with *L. ruthenicum*, 26 396 genes were upregulated, and 28 399 genes were downregulated in *L. barbarum*.

***GO* classification:** A total of 53 735 unigenes were annotated in 56 functional groups within the three main functional categories of the *GO* database (Fig. 4*C*). There were 25 functional groups related to biological processes, with the main biological processes annotated including cell process, metabolic process, single-organism process, biological regulation, stress response, and development process, among others. A large number of unigenes were annotated in the cell process, metabolic process, and single-organism process. The cellular component can be divided into 21 functional groups, primarily including: cells, cell conduction, organelles, cell membrane, with the largest number of unigenes annotated in cells. Molecular function can be divided into 10 functional groups, including binding activity, catalytic activity, transport activity, and nucleic acid binding transcription factor activity, among others. These unigenes in the transcriptional sequences of wolfberry flowers and fruits play a crucial regulatory role in the study of wolfberry gene function.

***KEGG* classification:** To identify highly active metabolic pathways in *L. barbarum* and *L. ruthenicum*, *KEGG* metabolic pathways annotated by genes were further classified. In the *KEGG* database; a total of 22 308 unigenes were found to be involved in 19 pathways, which can be divided into 5 categories: metabolism (12 534), genetic information regulation (6 070), environmental information regulation (1 112), organismal systems (1 086), and cellular processes (1 406). Among these unigenes, the largest number was involved in glucose metabolism, with a total of 2 662 unigenes. Second, there were 2 652 unigenes in genetic information regulation-related translation (Fig. 4*D*). There were 1 406 unigenes in the cellular processes category. Environmental adaptability in the organism systems category had 1 086 unigenes.

***GO* and *KEGG* enrichment analysis of DEGs:** The volcano map intuitively displays the degree of difference and significance of gene expressions between *L. barbarum* and *L. ruthenicum* (Fig. 5*A*), with scattered dots representing genes. Blue dots represent genes with no significant differences, and red dots represent the upregulated genes with significant differences, indicating higher expression of these genes in *L. barbarum*. Green dots represent downregulated genes with significant differences, indicating higher expression of these genes in *L. ruthenicum*. The Venn diagram of gene expression intuitively displays the number of common and specifically expressed genes between the two groups (Fig. 5*B*). There were 71 084 genes shared by *L. barbarum* and *L. ruthenicum*. Meanwhile, 63 579 genes were unique to *L. barbarum* and 49 878 genes were unique to *L. ruthenicum*.

Through *GO* enrichment analysis, it was found that the number of DEGs was most significant in the three functions of catalytic activity, metabolic process, and single-organism process. The metabolic process accounts for a large proportion, with 11 819 different genes (21.57%), 5 895 upregulated genes, and 5 924 downregulated genes. There were 9 997 different genes in the catalytic activity category (18.24%), 4 885 upregulated genes, and 5 122 downregulated genes. In the single-organism process, there were 8 943 DEGs (16.32%). There were 5 261 DEGs in the metabolic process of a single tissue, 3 814 DEGs in transferase activity, and 3 702 DEGs in hydrolytic enzyme activity, accounting for 9.6, 6.96 and 6.76%, respectively. Other functional categories were relatively low, accounting for only 20.55% of the total (Fig. 5*C*).

The 20 most significantly enriched pathways are shown in the *KEGG* enrichment distribution plot. The q-value is represented by the color of the dots, with red dots indicating smaller q-values. The DEGs are mainly concentrated in starch and sucrose metabolism, amino acid and nucleotide metabolism, phenylpropanoid biosynthesis, plant hormone signal transduction, pentose and glucuronate interconversions, and glycerophospholipid metabolism pathways. The degree of enrichment in the starch and sucrose metabolic pathway was the highest among DEGs, indicating active gene expression in this pathway (Fig. 5*D*). These results can provide references for studying developmental differences between *L. barbarum* and *L. ruthenicum*.

Ka/Ks analysis: From the distribution of Ka and Ks, we observed that conserved orthologs were widely distributed, while the number of divergent orthologs was much smaller and more concentrated (Fig. 6*A*). The *KEGG* enrichment analysis of conserved orthologous genes indicated that these genes were involved in cutin, suberin, and wax biosynthesis, cysteine and methionine metabolism, photosynthesis antenna proteins, phagosome, and other processes (Fig. 6*B*). In contrast, the analysis of divergent orthologous genes revealed that they were mainly involved in plant-pathogen interactions, protein export, and diterpenoid biosynthesis (Fig. 6*C*).

Screening of differential TFs related to flower and fruit development: Among the differentially expressed

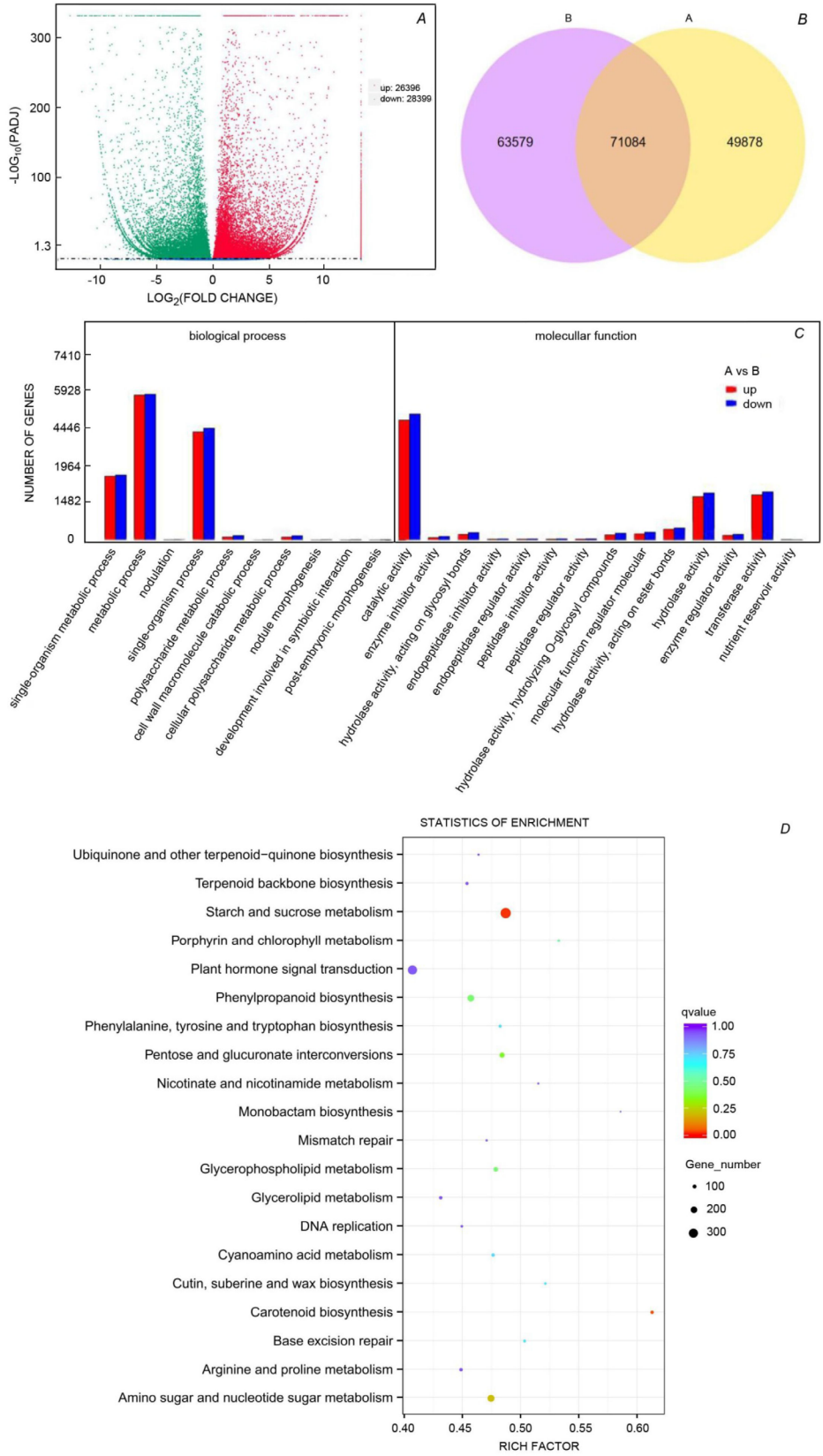

Fig. 5. Screening and analysis of DEGs. *A* - Volcano map of differential gene screening. *B* - Venn diagram of gene expression. *C* - Histogram of upregulated and downregulated gene classification after enrichment of DEGs by *GO*. *D* - Pathway enrichment analysis of DEGs. The *color* of the point represents *P*, and the *size of the point* represents the number of enriched DEGs.

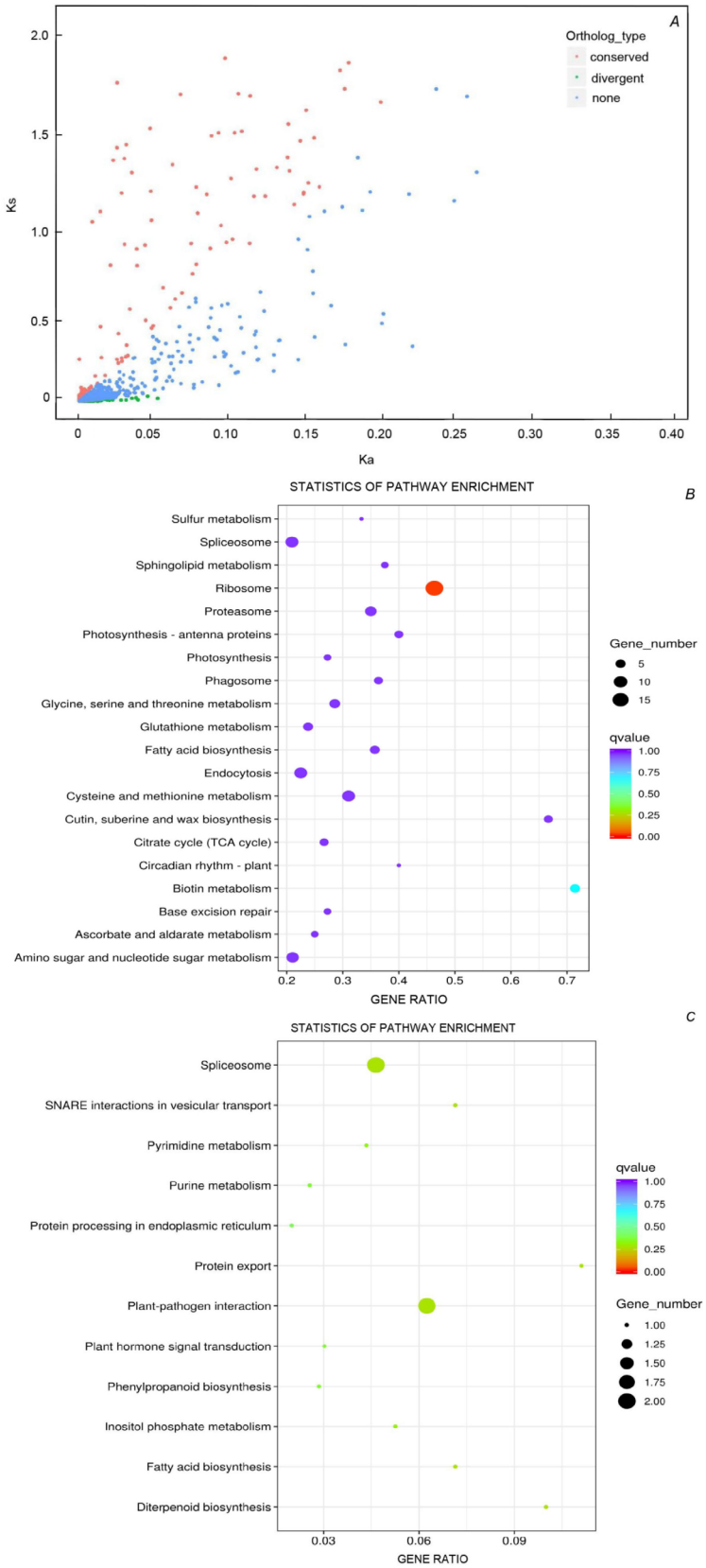

Fig. 6. Ka/Ks analysis. *A* - The distribution of Ka/Ks of orthologous genes in *L. barbarum* and *L. ruthenicum*. The *green dots* represent Ka/Ks > 1, the blue dots represent 0.1 ≤ Ka/Ks ≤ 1, and the *red dots* represent Ka/Ks < 0.1 *B* - The *KEGG* enrichment analysis of conserved orthologous genes (Ka/Ks < 0.1). *C* - The *KEGG* enrichment analysis of divergent orthologous genes (Ka/Ks > 1).

genes, potential transcription factor genes related to fruit morphology were selected, including 3 *SUNs* (*TRINITY_DN45833_c1_g2*, *TRINITY_DN58828_c1_g2*, *TRINITY_DN58828_c1_g3*), 7 *WUSs* (*TRINITY_DN57651_c4_g2*, *TRINITY_DN44915_c0_g1*, *TRINITY_DN42012_c2_g5*, *TRINITY_DN41705_c3_g1*, *TRINITY_DN36244_c0_g1*, *TRINITY_DN43365_c6_g1*, *TRINITY_DN49839_c1_g2*), and 2 *POS1s* (*TRINITY_DN52120_c0_g3*, *TRINITY_DN57785_c3_g1*). Subsequently, the large transcription factor family genes related to flower and fruit development were screened (Table 2, Tables 6-8 Suppl.).

Discussion

Morphological and cytological characteristics: In this study, the two *Lycium* species fruit's transverse and longitudinal diameters were measured and compared. The significant differences in fruit morphology may mainly be due to cytological reasons. The results of the paraffin section showed the changes in cell division and cell extension during the growth and development of two *Lycium* species at different stages, mastering the laws of fruit development and understanding the impact of cell division and cell extension on fruit development. The main stages of cell division and cell extension are not the same. Cell division generally plays a major role in the flowering and young fruit stages, while cell extension mainly plays a role in the middle and late stages of fruit development, thereby affecting the fruit development process. There are many genes regulating related processes in fruits. Due to the different processes regulated by genes, their high or low expression completely regulates the growth and development of fruits, resulting in significant differences in fruit morphology.

***GO* enrichment and *KEGG* enrichment analysis of DEGs:** In this study, the *GO* enrichment analysis results primarily focus on metabolic processes, catalytic activity, and single organism processes. The functional annotation results of the *KEGG* database indicate that DEGs are mainly concentrated in metabolism, genetic information processing, environmental information processing, organismal systems, and cell processes. Notably, compared to *L. ruthenicum*, carotenoid biosynthesis pathways are significantly enriched in *L. barbarum*, this observation can help explain why the fruit of *L. barbarum* has a bright orange-red color.

Screening of differential TFs related to flower and fruit development: In addition to the significant differential expressions of the above eight transcription factor families, other transcription factors may play essential roles in flower and fruit development. Whirly (WHY) is a relatively small transcription factor family discovered in plants in recent years. Its family members are closely associated with the regulation of plant leaf aging and flowering. In transgenic barley plants lacking the DNA/RNA binding protein WHIRLY1, the rate of leaf senescence is slowed down (Kucharewicz *et al*. 2017). Some researchers found that WHIRLY1 can recruit the histone deacetylase HDA15 to repress leaf senescence and flowering in *Arabidopsis* (Huang *et al*. 2022). We screened a WHIRLY family gene TRINITY_DN43986_c0_g4 highly expressed in *L. barbarum* and TRINITY_DN56094_c5_g1 highly expressed in *L. ruthenicum* from the obtained transcriptome data. Some studies on the RAV family's control of flowering function show that it can inhibit and delay flowering and play a certain role in responding to pathogen infections and abiotic stresses (Matías-Hernández *et al*. 2014). The zinc finger protein of vascular plants, VOZ has been found to promote flowering with VOZ1 and VOZ2 (Yasui *et al*. 2012). From the DEGs, we screened two RAV family genes (TRINITY_DN62726_c0_g1, TRINITY_DN41310_c2_g2) and two VOZ family genes (TRINITY_DN43000_c0_g3, TRINITY_DN54011_c1_g1), all of which exhibit high expression in *L. barbarum*.

In addition, LFY transcription factors in dicotyledonous plants play a crucial role in the transition from vegetative to reproductive development and have been proven to be expressed in both male and female floral meristems (Dornelas and Rodriguez 2005). Previous studies have found that *AGAMOUS* and *APETALA1* are flower homologous genes that directly affect the downstream of LFY (Mandel *et al*. 1992). LFY works in conjunction with the F-box protein named Unusual Floral Organs, which can produce ectopic floral organs (Risseeuw *et al*. 2013). We screened an LFY family gene in transcriptome data, and the gene TRINITY_DN40495_c2_g1 was highly expressed in *L. barbarum*. Studies on the function of the EIL transcription factor family showed that the *McEIL2* gene was involved in the regulatory process of fruit ripening and softening (Zhu *et al*. 2021). *EIL1* has been overexpressed in the ethylene-insensitive non-ripening *Nr* mutant of tomato, causing the mutant to resemble wild-type plants in phenotype (Chen *et al*. 2004). We screened an EIL family gene in transcriptome data and the gene TRINITY_DN46854_c1_g3 was highly expressed in *L. ruthenicum*. Due to the limited number of these genes, we did not spend much time and effort selecting one or two genes from hundreds of DEGs.

Conclusions

GO and *KEGG* analysis revealed that 56 *GO* terms and 19 biological pathways were significantly enriched. Using seven gene function annotation databases and the plant transcription factor database, we screened eight highly expressed transcription factor families, including MYB, MADS-box, B3, bHLH, bZIP, WRKY, NAC, and C2H2, related to flower and fruit development, and preliminarily counted the screened potential transcription factor genes.

Although the wolfberry genome has been sequenced, its specific annotation information has not been released, limiting the mining and understanding of key genes that regulate flower and fruit development. In this study, we

preliminarily screened some potential transcription factors related to flower and fruit development from DEGs. However, more precise and comprehensive research still needs to be further explored.

References

Albert N.W., Butelli E., Moss S.M.A. *et al.*: Discrete bHLH transcription factors play functionally overlapping roles in pigmentation patterning in flowers of *Antirrhinum majus*. - New Phytol. **231**: 849-863, 2021.

Anders S., Huber W.: Differential expression analysis for sequence count data. - Genome Biol. **11**: R106, 2010.

Bairoch A., Apweiler R.: The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. - Nucleic Acids Res. **28**: 45-48, 2000.

Ballester A.-R., Molthoff J., de Vos R. *et al.*: Biochemical and molecular analysis of pink tomatoes: deregulated expression of the gene encoding transcription factor SlMYB12 leads to pink tomato fruit color. - Plant Physiol. **152**: 71-84, 2010.

Beauchet A., Gévaudant F., Gonzalez N., Chevalier C.: In search of the still unknown function of FW2.2/CELL NUMBER REGULATOR, a major regulator of fruit size in tomato. - J. Exp. Bot. **72**: 5300-5311, 2021.

Benjamini Y., Hochberg Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. - J. R. Stat. Soc. B **57**: 289-300, 1995.

Cao H., Chen J., Lai X. *et al.*: Circular RNA expression profile in human primary multiple intracranial aneurysms. - Exp. Ther. Med. **21**: 239, 2021.

Chen G.P., Alexander L., Grierson D.: Constitutive expression of EIL-like transcription factor partially restores ripening in the ethylene-insensitive *Nr* tomato mutant. - J. Exp. Bot. **55**: 1491-1497, 2004.

Cong B., Barrero L.S., Tanksley S.D.: Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication. - Nat. Genet. **40**: 800-804, 2008.

Dornelas M.C., Rodriguez A.P.M.: The rubber tree (*Hevea brasiliensis* Muell. Arg.) homologue of the *LEAFY/FLORICAULA* gene is preferentially expressed in both male and female floral meristems. - J. Exp. Bot. **56**: 1965-1974, 2005.

Dubos C., Stracke R., Grotewold E. *et al.*: MYB transcription factors in *Arabidopsis*. - Trends Plant Sci. **15**: 573-581, 2010.

Ferrario S., Shchennikova A.V., Franken J. *et al.*: Control of floral meristem determinacy in petunia by MADS-box transcription factors. - Plant Physiol. **140**: 890-898, 2006.

Finn R.D., Bateman A., Clements J. *et al.*: Pfam: the protein families database. - Nucleic Acids Res. **42**: D222-D230, 2014.

Gene Ontology Consortium: The Gene Ontology (GO) database and informatics resource. - Nucleic Acids Res. **32**: D258-D261, 2004.

Giovannoni J.J.: Fruit ripening mutants yield insights into ripening control. - Curr. Opin. Plant Biol. **10**: 283-289, 2007.

Grabherr M.G., Haas B.J., Yassour M. *et al.*: Full-length transcriptome assembly from RNA-Seq data without a reference genome. - Nat. Biotechnol. **29**: 644-652, 2011.

Huang D.M., Lan W., Ma W.B. *et al.*: WHIRLY1 recruits the histone deacetylase HDA15 repressing leaf senescence and flowering in *Arabidopsis*. - J. Integr. Plant Biol. **64**: 1411-1429, 2022.

Huang Z.J., van der Knaap E.: Tomato *fruit weight 11.3* maps close to *fasciated* on the bottom of chromosome 11. - Theor. Appl. Genet. **123**: 465-474, 2011.

Jiao Z.C., Wang L.P., Du H. *et al.*: Genome-wide study of C2H2 zinc finger gene family in *Medicago truncatula*. - BMC Plant Biol. **20**: 401, 2020.

Kanehisa M., Araki M., Goto S. *et al.*: KEGG for linking genomes to life and the environment. - Nucleic Acids Res. **36**: D480-D484, 2008.

Karlova R., Chapman N., David K. *et al.*: Transcriptional control of fleshy fruit development and ripening. - J. Exp. Bot. **65**: 4527-4541, 2014.

Kucharewicz W., Distelfeld A., Bilger W. *et al.*: Acceleration of leaf senescence is slowed down in transgenic barley plants deficient in the DNA/RNA-binding protein WHIRLY1. - J. Exp. Bot. **68**: 983-996, 2017.

Lai X.H., Zhu X.Y., Chen H.C. *et al.*: The MaC2H2-like zinc finger protein is involved in ripening and ripening disorders caused by chilling stress *via* the regulation of softening-related genes in 'Fenjiao' banana. - Postharvest Biol. Tec. **186**: 111817, 2022.

Lemaire-Chamley M., Koutouan C., Jorly J. *et al.*: A chimeric TGA repressor slows down fruit maturation and ripening in tomato. - Plant Cell Physiol. **63**: 120-134, 2022.

Li B., Dewey C.N.: RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. - BMC Bioinformatics **12**: 323, 2011.

Li L., Stoeckert C.J., Roos D.S.: OrthoMCL: identification of ortholog groups for eukaryotic genomes. - Genome Res. **13**: 2178-2189, 2003.

Liu G.-S., Li H.-L., Grierson D., Fu D.-Q.: NAC transcription factor family regulation of fruit ripening and quality: a review. - Cells **11**: 525, 2022.

Liu J., Van Eck J., Cong B., Tanksley S.D.: A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. - PNAS **99**: 13302-13306, 2002.

Lopez-Ortiz C., Peña-Garcia Y., Bhandari M. *et al.*: Identification of miRNAs and their targets involved in flower and fruit development across domesticated and wild *Capsicum* species. - Int. J. Mol. Sci. **22**: 4866, 2021.

Mandel M.A., Gustafson-Brown C., Savidge B., Yanofsky M.F.: Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. - Nature **360**: 273-277, 1992.

Mao X., Cai T., Olyarchuk J.G., Wei L.: Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. - Bioinformatics **21**: 3787-3793, 2005.

Matías-Hernández L., Aguilar-Jaramillo A.E., Marín-González E. *et al.*: *RAV* genes: regulation of floral induction and beyond. - Ann. Bot.-London **114**: 1459-1470, 2014.

Nieuwenhuizen N.J., Chen X.Y., Pellan M. *et al.*: Regulation of wound ethylene biosynthesis by NAC transcription factors in kiwifruit. - BMC Plant Biol. **21**: 411, 2021.

Pei M.-S., Cao S.-H., Wu L. *et al.*: Comparative transcriptome analyses of fruit development among pears, peaches, and strawberries provide new insights into single sigmoid patterns. - BMC Plant Biol. **20**: 108, 2020.

Risseeuw E., Venglat P., Xiang D.Q. *et al.*: An activated form of UFO alters leaf development and produces ectopic floral and inflorescence meristems. - PLoS ONE **8**: e83807, 2013.

Robles P., Pelaz S.: Flower and fruit development in *Arabidopsis thaliana*. - Int. J. Dev. Biol. **49**: 633-643, 2005.

Rodríguez G.R., Muños S., Anderson C. *et al.*: Distribution of *SUN*, *OVATE*, *LC*, and *FAS* in the tomato germplasm and the relationship to fruit shape diversity. - Plant Physiol. **156**: 275-285, 2011.

Ruan C.C., Chen Z., Hu F.C. *et al.*: Genome-wide characterization and expression profiling of B3 superfamily during ethylene-

induced flowering in pineapple (*Ananas comosus* L.). - BMC Genomics **22**: 561, 2021.
Singh S.P., Jeet R., Kumar J. *et al.*: Comparative transcriptional profiling of two wheat genotypes, with contrasting levels of minerals in grains, shows expression differences during grain filling. - PLoS ONE **9**: e111718, 2014.
Tan C., Qiao H.L., Ma M. *et al.*: Genome-wide identification and characterization of melon *bHLH* transcription factors in regulation of fruit development. - Plants-Basel **10**: 2721, 2021.
Tatusov R.L., Galperin M.Y., Natale D.A., Koonin E.V.: The COG database: a tool for genome-scale analysis of protein functions and evolution. - Nucleic Acids Res. **28**: 33-36, 2000.
Tian Y.L., Xia T., Qiang X. *et al.*: Nutrition, bioactive components, and hepatoprotective activity of fruit vinegar produced from Ningxia wolfberry. - Molecules **27**: 4422, 2022.
van der Graaff E., Laux T., Rensing S.A.: The WUS homeobox-containing (WOX) protein family. - Genome Biol. **10**: 248, 2009.
Wang L., He L., Li J. *et al.*: Regulatory change at *Physalis Organ Size 1* correlates to natural variation in tomatillo reproductive organ size. - Nat. Commun. **5**: 4271, 2014.
Wang Y., Li Y., He S.-P. *et al.*: A cotton (*Gossypium hirsutum*) WRKY transcription factor (GhWRKY22) participates in regulating anther/pollen development. - Plant Physiol. Biochem. **141**: 231-239, 2019.
Xiao H., Jiang N., Schaffner E. *et al.*: A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. - Science **319**: 1527-1530, 2008.
Xu B., Yang Z.: PAMLX: a graphical user interface for PAML. - Mol. Biol. Evol. **30**: 2723-2724, 2013.
Yasui Y., Mukougawa K., Uemoto M. *et al.*: The phytochrome-interacting VASCULAR PLANT ONE-ZINC FINGER1 and VOZ2 redundantly regulate flowering in *Arabidopsis*. - Plant Cell **24**: 3248-3263, 2012.
Zhang L.P., Chen L.G., Yu D.Q.: Transcription factor WRKY75 interacts with DELLA proteins to affect flowering. - Plant Physiol. **176**: 790-803, 2018.
Zhang Y., Zhang A., Yang W. *et al.*: Transcriptome analysis and screening of genes associated with flower size in tomato (*Solanum lycopersicum*). - Int. J. Mol. Sci. **23**: 15624, 2022.
Zheng J., Ding C.X., Wang L.S. *et al.*: Anthocyanins composition and antioxidant activity of wild *Lycium ruthenicum* Murr. from Qinghai-Tibet Plateau. - Food Chem. **126**: 859-865, 2011.
Zheng Y., Jiao C., Sun H. *et al.*: iTAK: A program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. - Mol. Plant **9**: 1667-1670, 2016.
Zhu L.S., Shan W., Wu C.J. *et al.*: Ethylene-induced banana starch degradation mediated by an ethylene signaling component *MaEIL2*. - Postharvest Biol. Tec. **81**: 111648, 2021.