Zhu et al., Vegetos 2018, 31:2
DOI: 10.4172/2229-4473.1000410

Vegetos- An International Journal of Plant Research

Research Article

A SCITECHNOL JOURNAL

Analysis of Transcriptome and Differentially Expressed Genes Involved in Antioxidant Enzymes and Flavonoids Biosynthesis against UV-B Stress in the Desert Plant, *Reaumuria Soongorica*

Ruiqing Zhu*, Hui Yang and Zhiye Wang

Abstract

Reaumuria soongorica (Pall.) Maxim., a typical species of desert plant, presents excellent tolerability to adverse environment. Until yet, little is known about the molecular mechanisms of stress tolerance in *R. soongorica*. Herein, we used the RNA-seq to study the transcriptome of *R. soongorica* leaves. The differentially expressed genes (DEGs) involved in antioxidant enzymes and flavonoids under UV-B treatment were examined. *De novo* assembly produced 66,117 unigenes with an average length of 722 bp. Based on sequence similarity with known proteins, 38,762 (58.62%) genes had known homologs in protein databases. Out of these annotated unigenes, 29,460 and 13,582 unigenes had at least one gene ontology classification or were assigned to clusters of orthologous group, respectively. Searching against the Kyoto Encyclopedia of Genes and Genomes Pathway database (KEGG) showed that 21,202 (32.06%) unigenes were classified to 128 KEGG pathways. In addition, 7,711 simple sequence repeats (SSRs) and 216,851 single nucleotide polymorphisms (SNPs) were identified as potential molecular markers. The DEGs analysis between control and UV-B treated plants showed that the DEGs involved in antioxidant enzymes system were down-regulated. However, the DEGs involved in flavonoid pathway were up-regulated under UV-B stress. The transcriptome of *R. soongorica* prepared a foundation for further studies on gene interaction and regulation. Thousands of SSRs and SNPs markers identified can facilitate studies on genetic variation in *R. soongorica*. Additionally, the analysis of DEGs implied that *R. soongorica* resists to UV-B stress by up-regulation of flavonoids biosynthesis.

Keywords

Reaumuria soongorica; Transcriptome; UV-B; Differentially expressed gene

Introduction

Reaumuria soongorica (Pall.) Maxim., an extreme xerophytic semi-shrub, is a typical constructive and dominant species of desert vegetation in China. It forms the vast and zonal landscape with wide spread from the western Erdos, Alashan, Hexi Corridor, Beishan, Qaidam Basin, Gashun Gobi to Tarim Basin and Junggar Basin in China [1]. It serves in dune fixation, which plays an important role in maintaining the stability and continuity of the desert ecosystem [2]. Thus, researches focusing on its ecology, morphology, physiology of *R. soongorica*, such as spatial pattern, eco-adaptability of seed germination, stable carbon isotope, water-use strategy were performed [1-4]. Notably, the habitat of *R. soongorica* is characterized by intensive ultraviolet, high salinity, low ability of water and nutrient and extreme temperature fluctuations [5]. Attributed to the special habitat, *R. soongorica* exhibits excellent tolerability to severe environments [6]. Studies on the protection mechanism for photosynthetic properties, activity of antioxidant enzyme and metabolite changes under stress were conducted [4,6,7]. However, relative fewer studies have been carried out on the molecular level, especially the molecular mechanisms of stress response in *R. soongorica*. A preliminary genome survey, including genome size, chromosome number, karyotype, chromosomal localization of 45S rDNA loci was conducted with *R. soongorica* [8]. The genetic structure and differentiation of *R. soongorica* were assessed by the random amplified polymorphic DNA (RAPD) markers method, and it concluded that the ecological factors have tremendous impacts on the adaptive evolution of *R. soongorica* [9]. Then the conclusion was verified again by another research using inter-simple sequence repeat (ISSR) method [10]. The mitogen-activated protein kinase gene (*RsMPK2*) has been isolated and its participation as possible mediators under stress treatment was investigated [11].

Until now, the genome of *R. soongorica* has not been sequenced, and the absence of genomic information prevents many valuable experimental approaches from being applied to *R. soongorica*. Whereas, transcriptome sequencing is an effective way for gene discovery, expression patterns, development of simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs) markers, and researches about functional, comparative and evolutionary genomics [12]. The transcriptome of *R. soongorica* originated from flowers and leaves of mature plants and whole seedlings of two-week old was sequenced; however, the transcripts' sequences, such valuable information as SNPs and SSRs markers were not reported [13]. Furthermore, detecting differentially expressed genes (DEGs) is one of the major goals in the statistical analysis of RNA-seq data. The identification of DEGs helps to elucidate the function of genes when plant responds to different stress [14].

As the increasing of UV-B radiation at the earth's surface which led by the depletion of stratospheric ozone, the influence of UV-B on plant has attracted many scientists [15]. UV-B is expected to have detrimental impacts on terrestrial organisms and ecosystems [16]. Most studies have found that UV-B radiation harms the morphological, physiological, metabolic and reproductive properties of plants. Exposure of higher plants to UV-B radiation can induce oxidative damage to plant [17]. In order to overcome oxidative stress, plants have developed the antioxidant system to scavenge the ROS. Activities of antioxidant enzymes such as superoxide dismutase (SOD), peroxidase (POD), ascorbate peroxidase (APX) and catalase (CAT) were increased after stress treatment in *R. soongorica* [18]. When treated with UV-B radiation, *Nicotiana plumbaginifolia* showed

***Corresponding author:** Ruiqing Zhu, Institute of Biology, Gansu Academy of Sciences, 197 Dingxi South Road, Lanzhou 730000, China, E-mail: 153380101@qq.com

Received: April 09, 2018 **Accepted:** April 26, 2018 **Published:** May 03, 2018

rapid increase in transcripts of antioxidant enzymes [19]. Similarly, UV-B boosts the accumulation of flavonoids, a class of low-weight polyphenols which perform various biological functions in plants [20]. Studies focusing on the UV-B protection role of flavonoids were conducted by many scientists. For example, Burchard et al. [21] reported that flavonoids are important UV-B shielding compounds due to their absorbance in this wavelength region and the significant increment of concentrations in epidermal tissues after UV-B exposure. It is also reported that flavonoids serve as an important antioxidant in plant because they have the capability to directly scavenge ROS induced by UV-B irradiation [22]. Additionally, flavonoids have been reported to signaling molecules that could act at intracellular signaling cascades to respond to biotic and abiotic factors [23]. In previous studies, we found the anthocyanin and total flavonoids accumulated and their antioxidant abilities were enhanced significantly under UV-B radiation treatment. The flavonone 3-hydroxylase (*RsF3H*) was isolated and the real-time PCR showed the transcription of *RsF3H* was up-regulated by UV-B. Besides, the activities of key enzymes, phenylalanine ammonia-lyase (PAL), chalcone isomerase (CHI), F3H and dihydroflavonol 4-reductase (DFR), which involved in flavonoid biosynthesis pathway were increased under stress.

In this study, we generated the transcriptome using RNA-seq to provide sequence resources and genetic information to explore the stress tolerance mechanism of *R. soongorica*; the DEGs of UV-B-stressed and control plants were detected to identify the genes and pathways showing changes. We expected the transcriptome and DEGs analysis will provide genetic resources for exploring the UV-B tolerance mechanisms in *R. soongorica*.

Materials and Methods

Plant material and treatments

Seeds of *R. soongorica* were collected from Jiuzhoutai of Lanzhou City, Gansu, China (36°57′N, 103°42′E, and 1432 m elevation). The seeds were planted in the breeding base of Gansu Academy of Sciences. When the plants grew up to two years old, two groups were established: control (C) and UV-B (T) treatment. The control samples were maintained with the natural conditions as before. For the UV-B radiation treatment, plants were exposed to supplementary UV-B radiation of 0.2 Wm^{-2} using hanging UV-B lamps (UVB-313, 40W, Chenchen Lighting and Electronics Company, China) at appropriate height above the plants. The plants were treated for 6 hours, from 9:00 to 15:00 local time. Leaves from each treatment were collected and immediately stored in liquid nitrogen for subsequent use. Each treatment had three replicates. Leaves from each treatment were collected and then immediately stored in liquid nitrogen for later use.

RNA isolation and quality controls

The total RNA of each sample was isolated using a CTAB-based method and purified with the DNase I treatment. The quality and quantity of each RNA sample was assessed using Nano Drop 2000 Spectrophotometer (Nano Drop, Wilmington, DE). Only the RNA samples with 260/280 ratio from 1.9 to 2.1, 260/230 ratio from 2.0 to 2.5 and RIN (RNA integrity number) more than 8.0, were used for the analysis. A total of 20 μg of RNA was equally pooled from the two groups for cDNA library preparation. The construction of cDNA library was performed at the Beijing Genomics Institute (BGI), Shenzhen, China. Briefly, magnetic beads with oligo (dT) were used to isolate mRNA. Mixed with the fragmentation buffer, the mRNA was fragmented into short fragments. Then cDNA was synthesized using the mRNA fragments as templates. Short fragments were purified and resolved with EB buffer for end reparation and single nucleotide A (adenine) addition. After that, the short fragments were connected with adapters. The suitable fragments were selected for the PCR amplification as templates. Real-Time PCR System was used in quantification and qualification of the sample library. At last, the library could be sequenced using Illumina HiSeq™ 2000 platform.

Data filtering and de novo assembly

Reads were filtered and the clean reads were assembled using Trinity [24]. The longest assembled sequences (called contigs) were used as the core sequences of transcripts. Then the reads were mapped back to contigs; with paired-end reads it was able to detect contigs from the same transcript as well as the distances between these contigs. Finally, sequences cannot be extended on either end were obtained. Such sequences are defined as unigenes. Then by clustering, the unigenes will be divided to clusters (prefix is "*CL*") and unigenes (prefix is "*Unigene*").

Functional annotation

BLASTx alignment between the assembled sequences and protein databases like NCBI non-redundant protein (Nr) database, NCBI non-redundant nucleotide sequence (Nt) database (ncbi.nlm.nih.gov), and Swiss-Prot protein database (expasy.ch/sprot), the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (genome.jp/kegg), and the Cluster of Orthologous Groups (COG) (ncbi.nlm.nih.gov/COG) was performed with E-value $<10^{-5}$. If the results of different databases conflicted with each other, a priority order of Nr, Swiss-Prot, KEGG and COG should be followed when deciding the sequence direction of unigenes. A Perl script known as MISA (pgrc.ipk-gatersleben.de/misa/) was used to identify microsatellites in the unigene sequences. SNPs were detected using SOAPsnp after unigene sequences assembled [25].

DEGs analysis

RNAs from control (C) and UV-B treated (T) were obtained, respectively. For each group, 8 μg of RNA were used for Illumina DEG tag profiling. Sequence tag preparation was performed with the Illumina DEG tag profiling kit according to the manufacturer's protocol. The transcriptomic sequences of *R. soongorica* were generated in this study; suffix is "_3", (ncbi.nlm.nih.gov/bioproject/241148) and Shi et al. [13] (suffix is "_A") (www.ebi.ac.uk/arrayexpress/experiments/EMTAB-1543/) were used as the reference database. All clean tags were mapped to the *R. soongorica* reference sequence, and no more than one nucleotide mismatch was allowed. The clean tags mapped to reference sequences from multiple genes were filtered. The remained clean tags were designed as perfect clean tags. The number of perfect clean tags for each gene was calculated and then normalized in transcripts per million clean tags (TPM). DEGs were defined by using IDEG6 [26], with a relative change threshold of 2-fold ($P<0.005$, FDR ≤ 0.001). For hierarchical clustering analysis, the software Cluster 2.20 was used. Functional annotation analysis of DEGs was performed by the DAVID [27] web tools. The available raw data were then deposited in the NCBI Gene Expression Omnibus database (ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE109723).

Result

RNA-seq and *de novo* assembly

In this study, we prepared the mixed RNAs from the control and UV-B treatment at equivalent amount. A total of 117,425,498 raw

reads were generated. After adaptors, primer sequences, poly-A tails and low-quality sequences were removed, 102,422,382 clean reads were obtained with 96.26% Q20 bases (base quality more than 20). An assembler, Trinity (trinityrnaseq.sourceforge.net) was employed for *de novo* assembly. A total of 146,922 contigs were assembled with an average length of 312 bp. Finally, the *de novo* assembly yielded 66,117 unigenes with an average length of 722 bp, which contains 23,111 distinct clusters and 43,006 distinct singletons. The length of assembled unigenes ranged from 200 to 9,408 bp. There were 30,435 unigenes (46.02%) with length varying from 200 to 400 bp, 20,150 unigenes (30.48%) in the length range of 401 to 1,000 bp, and 15,532 unigenes (23.50%) with length more than 1,000 bp (Figure 1).

Functional annotation of unigenes

For validation and annotation of assembled unigenes, sequence similarity search was conducted against the NCBI Nr database, Nt database and the Swiss-Prot protein database using BLASTx algorithm with an E-value threshold of 10^{-5}. The results showed that out of 66,117 unigenes, 37,242 (56.33%) showed significant similarity to known proteins in Nr database (Table 1). A total of 27,633 unigenes (41.79%) were matched to Nt database. 22,676 (46.21%) had BLAST hits in Swiss-Prot database. Our results also showed that 25,364 (71.08%) of unigenes over 500 bp in length had BLAST matches, whereas only 5,073 (25.32%) of unigenes shorter than 300 bp did. The E-value distribution of the top hits in the Nr database revealed that 51.82% of the sequences with homologs showed significant homology (less than 1.0E-45), and 18.6% of the sequences with greater than 80% similarity were found (Figure 2A and Figure 2B). The species distribution of transcript set of *R. soongorica* was also analyzed. The result showed of *R. soongorica* transcripts have maximum similarity with that of *Vitis vinifera* unigenes (42.8%) followed by *Ricinus communis* (14.0%), *Populua trichocarpa* (11.6%) and *Glycine max* (7.9%, Figure 2C).

To identify functional categories among the 66,117 unigenes, the GO analysis was performed. 29,460 unigenes were assigned to the three main GO classifications, and the biological process made up the majority (121,791, 48.49%) followed by cellular component (95,918, 38.19%) and molecular function (33,451, 13.32%) (Figure 3A). Interestingly, 9,833 and 171 unigenes were annotated to the "*response to stimulus*" and "*antioxidant activity*" terms, respectively.

The COG database is a database where the orthologous gene products were classified. Out of 37,242 Nr hits, 13,582 sequences were assigned to the COG classifications (Figure 3B). Among the 25 COG categories, the cluster for general function prediction only (4,276, 34.80%) represented the largest group, followed by transcription (2,250, 16.57%), translation, ribosomal structure and biogenesis (2,094, 15.42%), replication, recombination and repair (2,062, 15.18%). In addition, 619 unigenes were assigned to secondary metabolites biosynthesis, transport and catabolism, and 363 unigenes were assigned to defense mechanisms.

To further investigate the biological function and interactions of unigenes, a pathway-based analysis was performed in the KEGG database. Based on a comparison against the KEGG database using BLASTx with an E-value $<10^{-5}$, out of the 66,117 unigenes, 21,202 (31.14%) had significant matches in the database and were assigned to 128 KEGG pathways. There was a large number of unigenes restricted to only a single one pathway, such as metabolism pathway (4,669 unigenes), biosynthesis of secondary metabolites (2,292 unigenes) and plant hormone signal transduction (988 unigenes). The top 14 pathways with the largest number of sequences are shown in Figure 4A, and the greatest number of transcripts was found in the metabolic pathways. In the secondary metabolism, 1943 unigenes were classified into 23 subcategories, and most of them were mapped to phenylpropanoid biosynthesis, flavonoid biosynthesis, stilbenoid, diarylheptanoid and gingerol biosynthesis (Figure 4B).

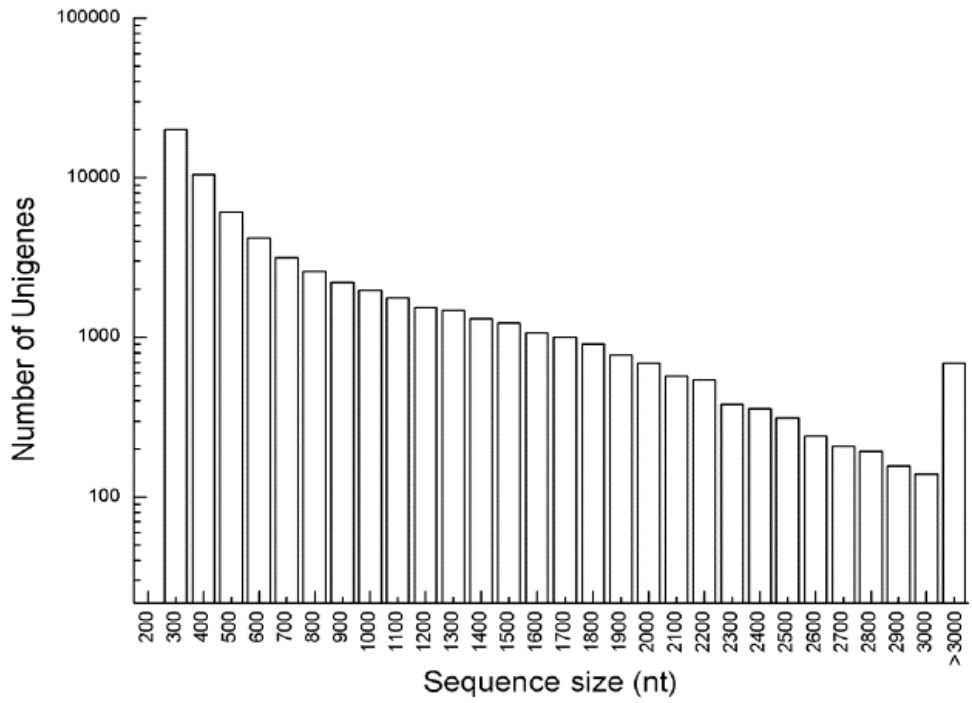

Figure 1: Length distributions of unigenes assembled by a multiple k-mer method.

Table 1: Summary of sequence annotation for *R. soongorica*.

Database	Nr	Nt	Swiss-Prot	GO	COG	KEGG	Total
Annoted numbers	37,242	27,633	22,676	29,406	13,582	21,202	38,762
Annotated percentage	56.33%	41.79%	46.21%	44.48%	20.54%	32.07%	58.63%

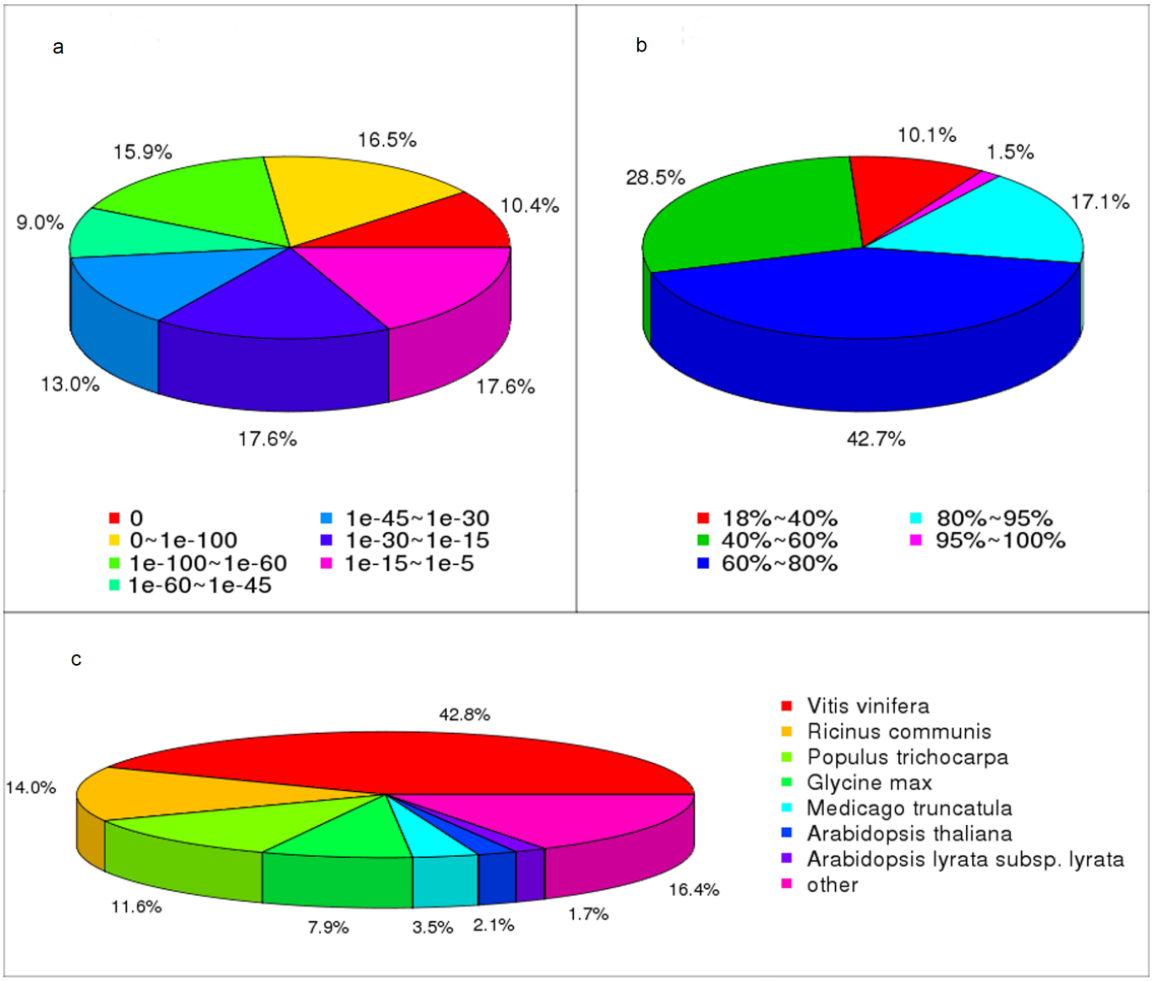

Figure 2: Characteristics of similarity search of unigenes against Nr database.

Marker identification

A total of 7,711 potential SSRs were identified in 6,791 unigenes, of which, 748 sequences contained more than one SSR, and 269 SSRs were present in compound form. The frequency, type and distribution of the potential 7,711 SSRs were also analyzed. The compilation of all SSRs revealed that, on the average, one SSR can be found every 3.35 kb in unigenes, and the frequency of SSR was 7.78%. Among the 4,114 SSRs, monomers were the most abundant repeat motif accounting for 4,059 (52.6%) of the SSRs, followed by trimers and dimers (1,717, 22.20%; 1,551, 20.11%, respectively). Hexamers were the least common motifs with 111 (1.4%) of the SSRs (Table 2). SSRs lengths were mostly distributed from 12 to 20 bp, accounting for 83.79% of total SSRs, followed by 21–30 bp length range (701 SSRs, 9.09%). There were 278 SSRs with length larger than 30 bp. Within the searched SSRs, 280 motif sequence types were identified. The A/T mono-nucleotide repeat motif was the most abundant motif detected in our SSRs (3,999, 51.86%). AG/CT was the most abundant di-nucleotide repeat motif detected in our SSRs (827, 10.72%). (Figure 5).

In our study, transcriptome sequencing data were generated from various treated individuals, and thus provided an opportunity to investigate the frequency of SNPs in transcribed regions. As a result, we identified a total of 216,851 putative SNPs in 38,868 unigenes (58.79% of the total unigenes), and the average density is 0.26/kb. Of all the SNPs, 136,201 (62.81%) and 80,650 (37.19%) were transitions and transversions, respectively. Transitions A/G and C/T were the two most abundant SNPs and accounted for 31.73 and 31.07% of all SNPs, respectively. Each of the four transversions (A/C, A/T, G/C, and G/T) was accounting for approximately 10% of all SNPs.

Genes and DEGs involved in antioxidant enzyme system

The GO analysis of *R. soongorica* transcriptome showed that 171 unigenes were annotated to the "*antioxidant activity*" term (Fig. 2A). The DEG analysis revealed that 2,160 unigenes were defined as

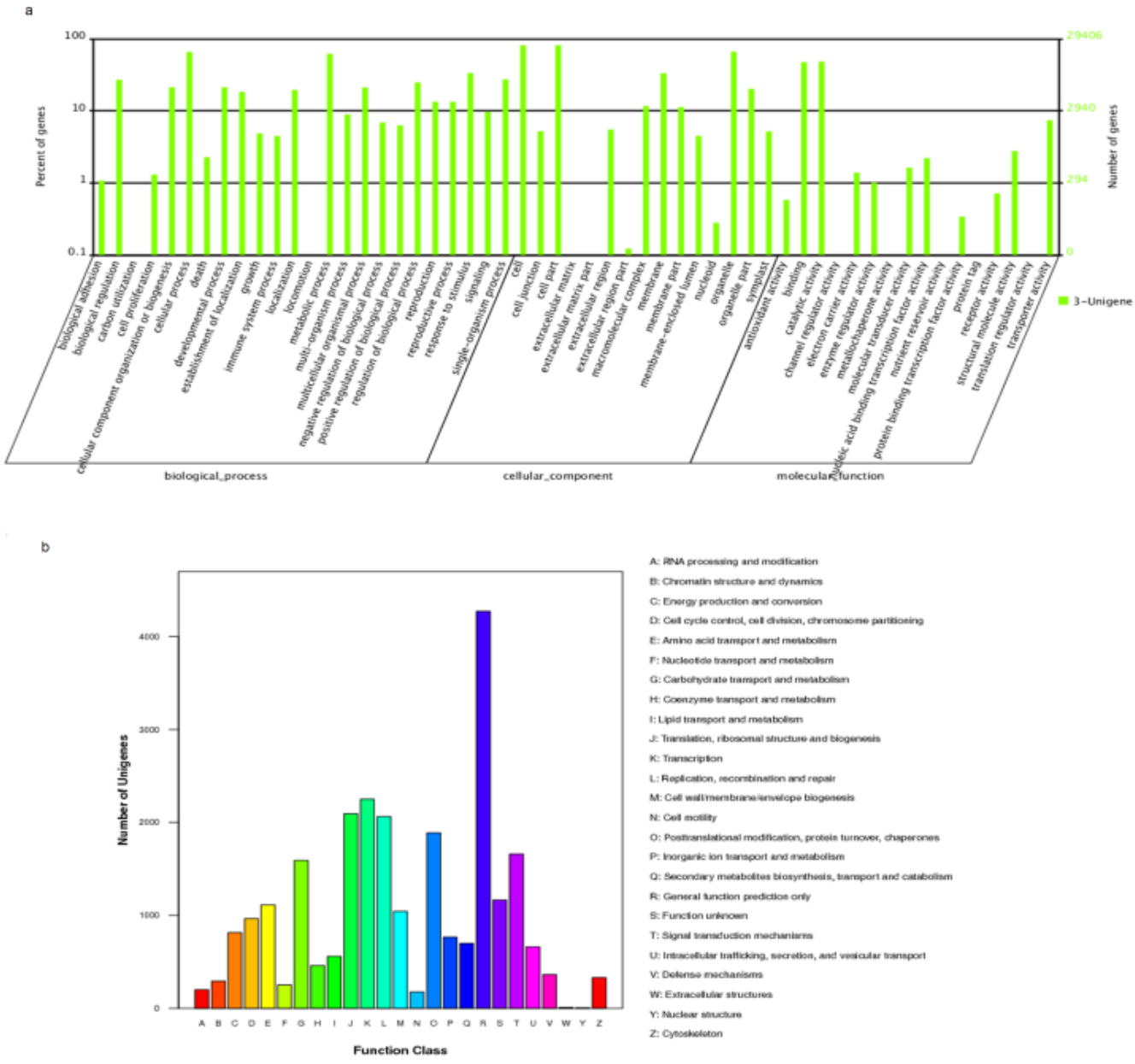

Figure 3: Annotation of unigenes to GO and COG.

DEGs using the thresholds of false discovery rate (FDR) ≤ 0.001 and |log2Ratio| ≥ 1, including 564 down-regulated and 1,596 up-regulated (Figure 6). Out of the annotated DEGs, 22 DEGs were predicted to encode antioxidant enzymes (Table 3). There were 3,11,8 unigenes categorized into the GSH-ascorbate cycle, the glutathione peroxidase (GPX) pathway, and the peroxiredoxin/thioredoxin (PrxR/Trx) pathway, respectively (Figure 7A). We detected 7 genes encoding POD, in which only one was up-regulated by 5-fold and the other six were down-regulated by more than 3-fold under UV-B treatment (Figure 7A). Among the down-regulated DEGs encoding POD, the DEG showed greatest changes was down-regulated by 8-fold under UV-B stress. In the GSH-ascorbate cycle, the DEGs encoding APX, glutaredoxin (GLR) were all down-regulated by approximately 4-fold under UV-B treatment. In the GPX pathway, there were three and one DEGs encoding GST and GPX respectively, and all the unigenes were down-regulated by more than 2.5-fold. One of the DEG encoding GST was down-regulated by approximately 7-fold than that in control. The DEGs encoding Trx and PrxR categoried in the PrxR/Trx pathway also showed approximately 3-fold of down-regulation under UV-B treatment. The two DEGs encoding SOD showed down-regulation, one of which was down-regulated by approximately 285-fold after UV-B radiation. Overall, the boxplot analysis showed that the DEGs involved in antioxidant enzyme system were not as abundant in UV-B group as that in control (Figure 7B).

Genes and DEGs involved in flavonoid biosynthesis

Biosynthesis of flavonoid is participated by a great number of enzymes (Figure 8). Based on the KEGG pathway assignment of *R.*

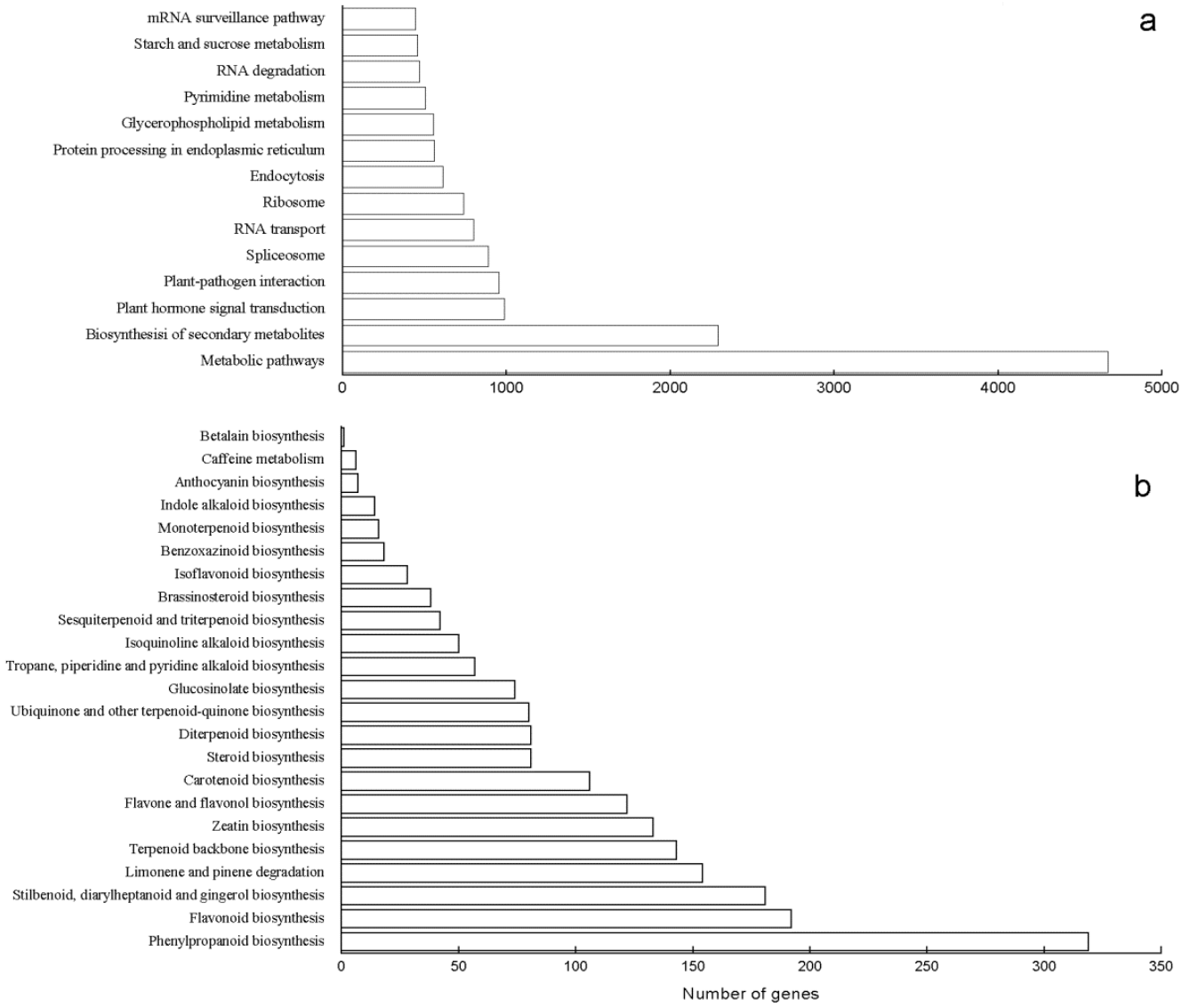

Figure 4: Pathway assignment based on KEGG.

Table 2: Summary of SSRs searching results.

Searching item	Numbers
Total number of sequences examined	66,117
Total size of examined sequences (bp)	47,730,030
Total number of identified SSRs	7,711
Number of SSRs containing sequences	6,791
Number of sequences containing more than 1 SSR	748
Number of SSRs present in compound formation	269
Monomers	4,059
Dimers	1,551
Trimers	1,712
Quadmers	120
Pentamers	158
Hexamers	111

soongorica transcriptome, we found a total of 129 unigenes encoding the 10 key enzymes involves in the flavonoids backbone synthesis (Table 4). In most cases, more than one unigene was annotated as the same enzyme. Such unigenes may represent different fragments of a single transcript, different members of a gene family, or both.

In the pathway enrichment analysis, eight unigenes encoding five key enzymes were identified as DEGs (Table 5). The DEGs encoding chalcone synthase (CHS) and flavonoid 3-O-glycosyltransferase (FGT) were up-regulated by 7-fold and 239-fold respectively under UV-B stress (Figure 9A). The DEG encoding leucoanthocyanidin reductase (LAR) showed 2-fold of down-regulated transcription by UV-B treatment. Out the three DEGs encoding DFR, one showed increasing transcription by 5-fold and two exhibited approximately 3-fold decreasing transcription in UV-B treated group. Among the two DEGs encoding F3'H, one was down-regulated by 2.8-fold and the other was up-regulated by 5-fold under UV-B stress. The boxplot analysis showed the DEGs involved in flavonoid biosynthesis were

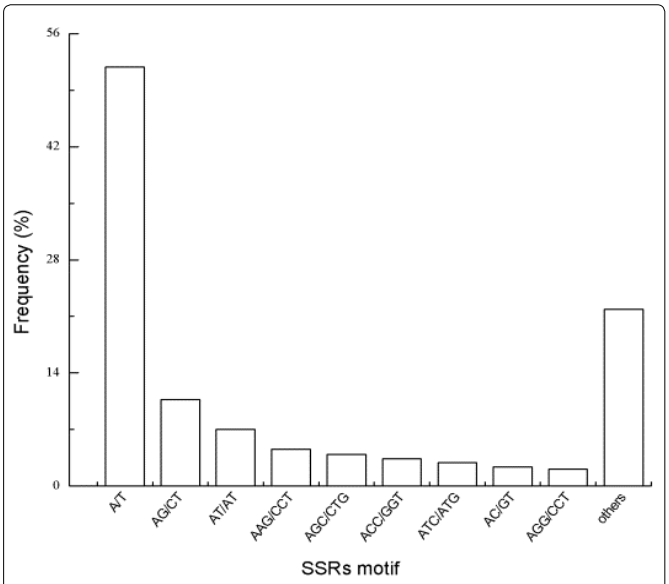

Figure 5: Frequency distribution of SSRs based on motif sequence types.

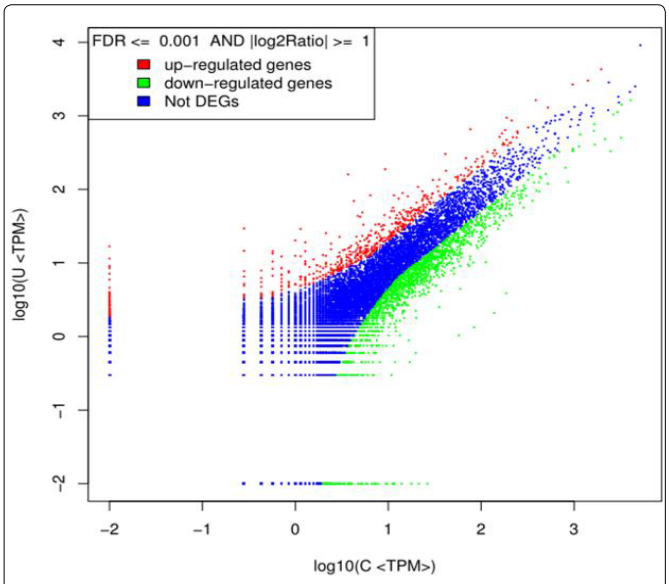

Figure 6: Identification of DEGs between C andT.

Table 3: The DEGs involved in antioxidant enzymes system.

Gene name	Gene ID	Fold change
GLR	Unigene45877_A	-1.34
APX	CL1969.Contig1_3	-2.11
	CL8055.Contig1_A	-1.85
GST	Unigene10576_A	-2.75
	Unigene9818_A	-2.06
	CL163.Contig2_A	-1.38
GPX	Unigene43594_A	-1.60
POD	CL8620.Contig1_A	2.15
	CL7207.Contig1_3	-3.13
	CL1969.Contig1_3	-2.11
	CL8055.Contig1_A	-1.85
	CL4242.Contig2_3	-1.47
Trx	Unigene46454_A	-2.20
	Unigene2197_A	-1.70
	CL9335.Contig1_3	-1.40
	Unigene43099_A	-1.84
	Unigene19525_A	-1.38
	Unigene30373_3	-1.03
PrxR	Unigene19069_A	-1.63
	Unigene26941_3	-1.63
SOD	Unigene1054_A	-8.15
	Unigene19255_A	-1.53

significant abundant in UV-B-treated group than that in control (Figure 9B).

Discussion

During the past several years, RNA-seq has become a tremendous approach for high-throughput sequencing on a genome-wide scale in model and non-model organisms [28]. In addition to its relatively low cost and great improvement of efficiency and speed, RNA-seq can eliminate the bacterial cloning step that can bias the composition of the cDNA library. Recently, there is great interest in utilizing the RNA-seq technology for transcriptome sequencing and assembly for model and non-model organisms, such as zebrafish [29], *Aspergillus oryzae*, whitefly, sweetpotato, *Scabiosa columbaria* and *Anopheles* [30-32]. Consistent with these reports previously, our results also revealed that RNA-seq reads can be assembled efficiently and utilized for gene discovery, SNPs and SSRs development in *R. soongorica*. In present study, a total of 102,422,382 clean reads were generated. These sequences were then assembled into unigenes, with 722 bp in average length and 1,145 bp in N50 length. While, Shi et al. [13] reported the average and N50 length of unigenes were 677 and 1,109 bp, respectively. The results of our study demonstrated that we assembled the sequences with higher quality because the length of sequences assmbled is a criterion for assembly success [24]. Furthermore, we got a greater percentage of annotated unigenes (58.63%) as is suggested

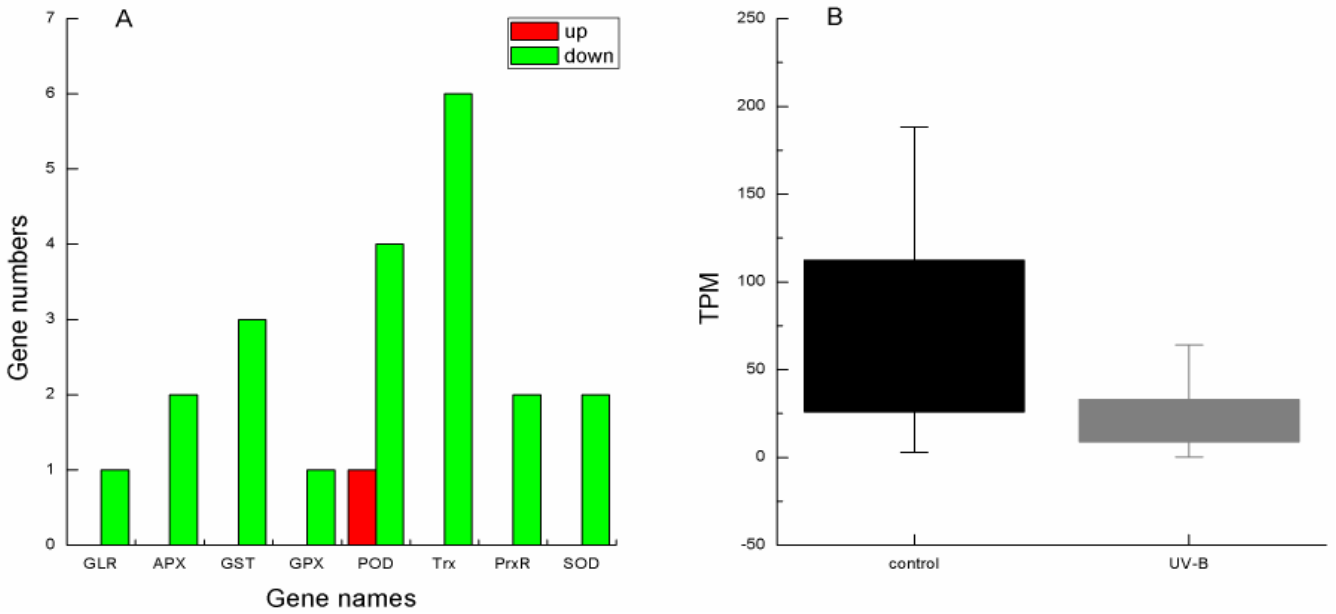

Figure 7: DEGs involved in antioxidant enzymes system under control and UV-B treatment.

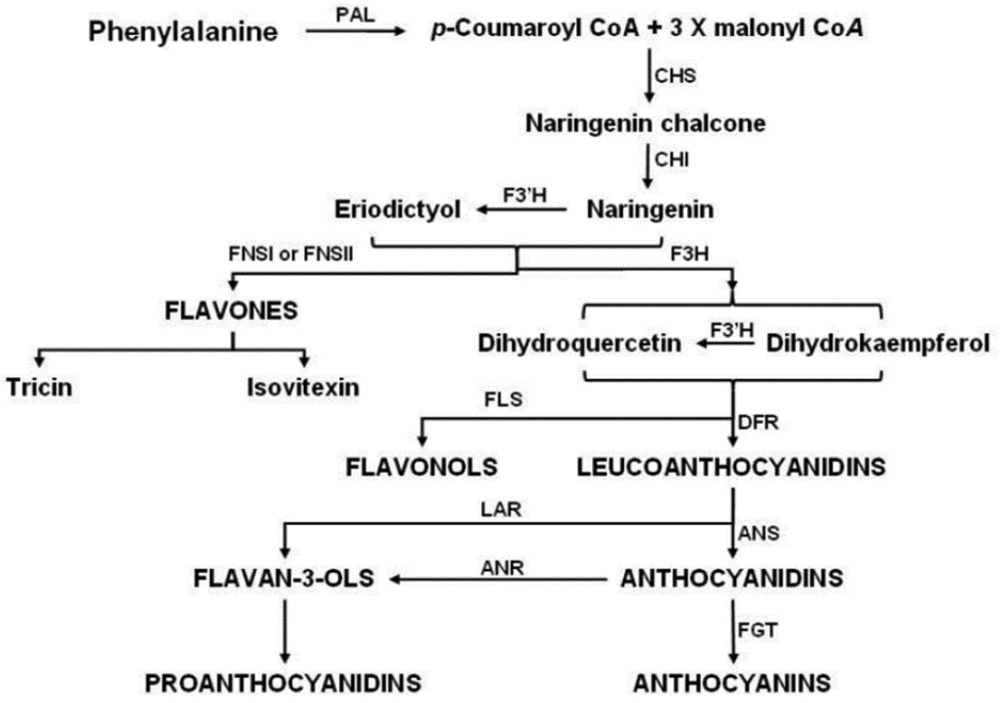

Figure 8: Pathway of flavonoids biosynthesis.

that the longer unigenes were more likely to have BLAST matches in the protein databases [33]. In contrast, only 55.17% unigenes were annotated to proteins database in the previous *R. soongorica* transcriptome, and the result suggested that our transcriptome has a higher quality.

A large number of unigenes were assigned to a wide range of GO categories and COG classifications, implying that our RNA-seq data represented a wide diversity of transcripts. The GO analysis showed that 171 unigenes were annotated to the antioxidant activity category which provided sequence resources for further studies on the antioxidant mechanisms in *R. soongorica*. Based on the KEGG annotation, the well represented pathways were metabolic pathways, biosynthesis of secondary metabolites, plant hormone signal transduction and plant-pathogen interaction (Figure 4A). In the secondary metabolism, 2,493 unigenes were classified into 23 different

Table 4: The unigenes involved in flavonoid biosynthesis pathway.

Gene Name	Unique sequence
CHS	Unigene19885_3, Unigene26057_3, Unigene26115_3, Unigene33191_3, Unigene758_3
CHI	CL381.Contig1_3, Unigene1822_3, Unigene26426_3
F3H	CL1892.Contig1_3, CL2300.Contig1_3, CL2300.Contig2_3, CL6265.Contig1_3, CL8207.Contig1_3, CL8207.Contig2_3, Unigene16743_3, Unigene30508_3, Unigene30668_3, Unigene30903_3
F3'H	CL10160.Contig1_3, L10586.Contig1_3, CL10586.Contig2_3, CL1068.Contig1_3, CL10910.Contig2_3, L10910.Contig3_3, CL113. Contig1_3, CL113.Contig2_3, CL2370.Contig1_3, CL2370.Contig2_3, CL2502.Contig2_3, CL3950.Contig1_3, CL3950.Contig2_3, CL4272. Contig10_3, CL4272.Contig2_3, CL4272.Contig3_3, CL4272.Contig4_3, CL4272.Contig5_3, CL4272.Contig7_3, CL4272.Contig8_3, CL4272.Contig9_3, CL50.Contig1_3, CL50.Contig2_3, CL50.Contig3_3, CL50.Contig4_3, CL50.Contig5_3, CL950.Contig10_3, CL950. Contig11_3, CL950.Contig12_3, CL950.Contig1_3, CL950.Contig2_3, CL950.Contig4_3, CL950.Contig5_3, CL950.Contig6_3, CL950. Contig7_3, CL950.Contig8_3, CL950.Contig9_3, Unigene10169_3, Unigene11480_3, Unigene13307_3, Unigene1473_3, Unigene15311_3, Unigene15475_3, Unigene18345_3, Unigene19261_3, Unigene25117_3, Unigene25862_3, Unigene31096_3, Unigene32504_3, Unigene32704_3, Unigene5873_3, Unigene6188_3, Unigene8261_3, Unigene9162_3
FLS	CL1418.Contig2_3, CL1418.Contig3_3, CL1418.Contig4_3, CL2976.Contig1_3, CL2976.Contig2_3, CL2976.Contig3_3, CL2976.Contig4_3, CL328.Contig1_3, CL4230.Contig1_3, CL4230.Contig2_3, CL4727.Contig1_3, Unigene22161_3, Unigene25584_3, Unigene27802_3, Unigene29997_3, Unigene30872_3, Unigene31018_3, Unigene31715_3, Unigene39041_3
DFR	Unigene16014_3, Unigene20214_3, Unigene23253_3, Unigene23254_3, Unigene26391_3
ANS	CL2976.Contig1_3, CL2976.Contig2_3, CL328.Contig1_3, CL328.Contig2_3, CL4230.Contig1_3, CL4727.Contig1_3, CL6265.Contig1_3, Unigene29996_3, Unigene29997_3, Unigene31018_3, Unigene35561_3, Unigene39041_3, Unigene4014_3, Unigene4392_3
ANR	CL7841.Contig1_3, CL7841.Contig2_3, Unigene19339_3, Unigene26391_3
FGT	CL2005.Contig1_3, CL2005.Contig2_3, CL6983.Contig1_3, Unigene13961_3, Unigene35785_3, Unigene35875_3
LAR	CL10202.Contig1_3, CL1097.Contig1_3, CL1097.Contig2_3, CL1097.Contig3_3, CL5065.Contig1_3, Unigene29022_3, Unigene29720_3, Unigene30941_3, Unigene31395_3

Table 5: The DEGs involved in flavonoid biosynthesis pathway.

Gene name	Gene ID	Fold change
CHS	Unigene10312_A	2.67
DFR	CL6414.Contig1_A	-1.69
	Unigene29399_3	-1.51
	Unigene19573_A	2.39
F3H	CL113.Contig1_3	2.23
F3'5'H	CL10160.Contig1_3	-1.49
LAR	Unigene19830_A	-1.13
FGT	Unigene46989_A	7.90

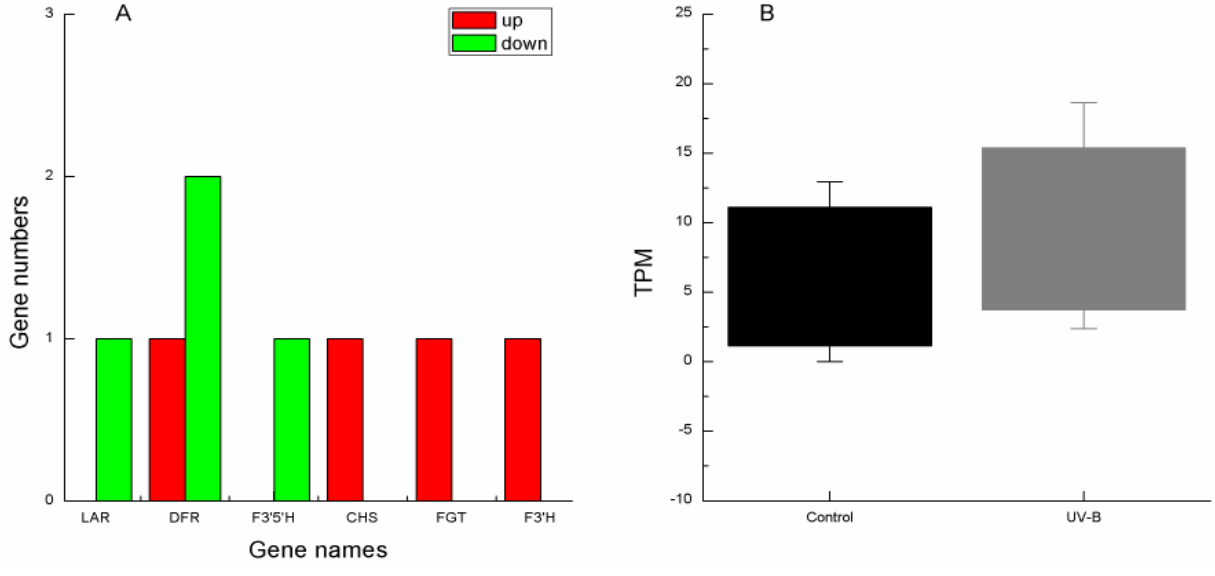

Figure 9: DEGs involved in flavonoid biosynthesis pathway under control and UV-B treatment.

subcategories (Figure 4B). A total of 350 unigenes were annotated to the flavonoids biosynthesis, including flavonoid biosynthesis (192), flavone and flavonol biosynthesis (122), isoflavonoid biosynthesis (29) and anthocyanin biosynthesis (7). This is as expected because in previous studies flavonoid were demonstrated to involved in diverse stress responses in plants, in addition to have functions in growth, development, reproduction [34].

SSRs are very common in natural populations and provide valuable markers for genetic mapping as well as population genetic studies [35]. Li et al. [36] reviewed that SSRs have different putative functions: SSR variations in 5'-UTRs could regulate gene expression by affecting transcription and translation; SSR expansions in the 3'-UTRs cause transcription slippage and produce expanded mRNA; intronic SSRs can affect gene transcription, mRNA splicing, or exportation to cytoplasm; SSRs within genes should be subjected to stronger selective pressure than other genomic regions. We located 7,711 potential SSRs in our unigene data set. The number of tri-nucleotide SSRs was much higher than di-nucleotide SSRs. This result is consistent with Garg et al. [37], who also found the tri-nucleotide SSRs was much abundant than di-nucleotide SSRs. However, recently the larger number of di-nucleotide SSRs than tri-nucleotide SSRs has been reported in pigeonpea [38]. The difference of frequency and distribution of SSRs might be related with various factors such as size of data set, tools and criteria used [39].

SNPs markers are of great importance for understanding the genetic variation and identification of quantitative trait locus in molecular breeding applications. Until now, Xu et al. [9] detected only 69 polymorphic loci by the RAPD markers method. We discover 216,851 putative SNPs due to the deep coverage produced by RNA-seq and different treatments of RNA samples. Our results clearly demonstrate that the RNA-seq technology is a cost-effective way to produce most extensive gene-associated SNPs from transcriptome data. A low density of SNPs (0.26 SNPs per kb) in the unigenes was observed in our study. Similarly, Wang et al. [35] detected an average density of 0.2 SNPs per kb. However, this SNPs density in the *R. soongorica* transcriptome is extremely lower than that reported by Lijavetzky et al. [40], who observed an average density of 15.6 SNPs per kb in grapevine. A low density of SNPs in the *R. soongorica* transcriptome could be partially attributed to a small sample of *R. soongorica* varieties used in this study [35]. The large number SNPs we identified represent a large resource of molecular markers that can be utilized for the detection of functional variation and the effect of selection [41]. Of course, the validation of the SNPs needed to be verified using PCR and Sanger sequencing.

UV-B radiation induce the generation of reactive oxygen species (ROS), including superoxide radical ($O_2^{\cdot-}$), hydroxyl radical ($OH^{\cdot}$), hydrogen peroxide (H_2O_2) and singlet oxygen ($^1O_2$) in plants [42]. The ROS would cause oxidative stress to plant [17]. Cheruth et al. [43] reviewed that plants have developed antioxidant enzymes, such as SOD, POD and APX to scavenge ROS induced by stress. In this study, the enrichment of DEGs encoding antioxidant enzyme decreased after 12 h of UV-B radiation. This founding was similar to the results of Radyukina et al. [44], who reported antioxidant enzymes activities were inhibited under UV-B stress. Furthermore, our results are in agreement with Zu et al. [45], who concluded that the antioxidant enzymes system may be insufficient to defend the oxidative damage induced by UV-B radiation. However, an opposite result was obtained by Willekens et al. [19], who found increments of mRNA transcripts of antioxidant enzymes in responses to enhanced UV-B flux. The discrepancy between our outcomes and those of Willekens et al. [19] might suggest that *R. soongorica* has different ROS scavenging mechanisms to other plants.

On the other hand, increased expression of gene involved flavonoid pathway was detected in our previous study on *R. soongorica*. In our dataset, the DEGs encoding key enzymes involved in flavonoid biosynthesis pathway were up-regulated, especially the DEGs which encode CHS and FGT. These findings were similar to the results of other studies on UV-B treated plants [20,46,47]. CHS catalyzes the first committed step in the biosynthesis of flavonoid [48]. FGT catalyzes the last step in the anthocyanin biosynthesis pathway [49]. As a result, there may be accumulations in flavonoids, especially anthocynanin. The effects of up-regulated flavonoids may be interpreted from three aspects. Firstly, flavonoids may play important UV-B shielding role attributed to their absorbance in this wavelength region [21]. Secondly, Agati et al. [22] reported that flavonoids have ability to directly scavenge the ROS induced by the penetration of UV-B into leaf. Furthermore, flavonoids may serve as signaling molecules to act at signaling cascades, consequently, affect cellular function or modulate gene expression to defend to UV-B stress [23]. Therefore, the up-regulation of flavonoid pathway may be an important factor in the UV-B tolerance mechanisms of *R. soongorica*.

In conclusion, the transcriptome was obtained in the study, and it provides genetic foundation for molecular mechanisms of stress tolerance in *R. soongorica*. The identification of SSRs and SNPs markers can facilitate further studies on genetic variation in *R. soongorica*. The DEGs analysis implied that the flavonoid biosynthesis pathways play an important role in resisting to UV-B stress. These results will prompt studies on the molecular mechanisms of UV-B resistance in *R. soongorica*, and possibly in other species that survive in areas of intense UV-B radiation.

Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (Grant No. 31600332), Natural Science Foundation of Gansu, China (1508RJZA056), Western Light Talent Training Program of Chinese Academy of Science and Youth science and technology innovation fund project of Gansu Academy of Science (2017 QN-07).

References

1. Ma JY, Chen T, Qiang WY, Wang G (2005) Correlations between foliar stable carbon isotope composition and environmental factors in desert plant *Reaumuria soongorica* (Pall.) Maxim. J Intergr Plant Biol 47: 1065-1073.
2. Ma MH, Kong LS (1998) The bio-ecological characteristics of *Reaumuria soongorica* on the border of oasis at Hutubi, Xinjiang. Acta Phyto Sin 22: 237-244
3. Zeng YJ, Wang YR, Zhang BL, Li BE (2000) Eco-adaptability studies of seed germination in species of *Reaumuria soongorica* and Oxytropis aciphylla. Acta Prata Sin 9: 36-42.
4. Xu H, Li Y (2006) Water-use strategy of three central Asian desert shrubs and their responses to rain pulse events. Plant Soil 285: 5-17.
5. Ma JY, Zhou BC, Xia DS, Zhang HW, Jiang XL (2007) Relationships between environmental factors and chlorophyll, proline accumulation in desert plant *Reaumuria soongorica*. Acta Bot Borea Occident Sin 27: 769-775.
6. Bai J, Gong CM, Chen K, Kang HM, Wang G (2009) Examination of antioxidative system's responses in the different phases of drought stress and during recovery in desert plant *Reaumuria soongorica* (Pall.) Maxim. J Plant Biol 52: 417-425.
7. Liu YB, Zhang TG, Wang J (2008) Photosynthesis and metabolite leaves levels in dehydrating leaves of *Reaumuria soongorica*. Acta Biol Cracov Bot 50: 19-26.
8. Wang XH, Zhang T, Wen ZN, Xiao HL, Yang ZJ, et al. (2011) The chromosome number, karyotype and genome size of the desert plant diploid *Reaumuria soongorica* (Pall.) Maxim. Plant Cell Rep 30: 955-964.

9. Xu L, Wang YL, Wang XM, Zhang LJ, Yue M, et al. (2003) Genetic structure of *Reaumuria soongorica* population in Fukang desert, Xinjiang and its relationship with ecological factors. Acta Bot Sin 45: 787-794.

10. Qian ZQ, Li X, Wang YL, Yang Y, Zhao GF (2008) Ecological genetics of *Reaumuria soongorica* (Pall.) Maxim. population in the oasis-desert ecotone in Fukang, Xinjiang, and its implications for molecular evolution. Biochem Systemat Ecol 36: 593-601.

11. Liu YB, Li XR, Tan HJ, Liu ML, Zhao X, et al. (2010) Molecular characterization of *RsMPK2*, a C1 subgroup mitogen-activated protein kinase in the desert plant *Reaumuria soongorica*. Plant Physiol Bioch 48: 836-844.

12. Der JP, Baker MS, Wickett NJ, dePamphilis CW, Wolf PG (2011) *De novo* characterization of the gametophyte transcriptome in bracken fern, *Pteridium aquilinum*. BMC Genomics 12: 99-113.

13. Shi Y, Yan X, Zhao PS, Yin HX, Zhao X, et al. (2013) Transcriptomic Analysis of a Tertiary Relict Plant, Extreme Xerophyte *Reaumuria soongorica* to Identify Genes Related to Drought Adaptation. PLoS one 8: e63993.

14. Kvam VM, Liu P, Si Y (2012) A comparison of statistical methods for detecting differentially expressed genes from RNA-seq Data. Botany 99: 248-256.

15. Flint SD, Ryel RJ, Caldwell MM (2003) Ecosystem UV-B experiments in terrestrial communities: a review of recent findings and methodologies. Agr Forest Meteorol 120: 177-189.

16. Jansen MAK, Gaba V, Greenberg BM (1998) Higher plants and UV-B radiation: balancing damage, repair and acclimation. Trends Plant SCI 3: 131-135.

17. Rowe LA, Degtyareva N, Doetsch PW (2008) DNA damage-induced reactive oxygen species (ROS) stress response in *Saccharomyces cerevisiae*. Free Radical Biol Med 45: 1167-1177.

18. Liu YB, Liu ML, Cao B (2011) Effects of drought stress on antioxidant enzyme, photosynthetic pigment and flavonoid pathway in two desert shrubs. SCI Cold Arid Regions 3: 332-338.

19. Willekens H, Camp WV, Montagu MV, Inzé D, Langebartels C, et al. (1994) Ozone, sulfur dioxide and ultraviolet B have similar effects on mRNA accumulation of antioxidant genes in *Nicotiana plumbaginifolia* L. Plant Physiol 106: 1007-1014.

20. Tossi V, Amenta M, Lamattina L, Cassia R (2011) Niric oxide enhances plant ultraviolet-B protection up-regulating gene expression of the phenylpropanoid biosynthetic pathway. Plant Cell Environ 34: 909-921.

21. Burchard P, Bilger W, Weissenböck G (2000) Contribution of hydroxycinnamates and flavonoids to epidermal shielding of UV-A and UV-B radiation in developing rye primary leaves as assessed by ultraviolet-induced chlorophyll fluorescence measurements. Plant Cell Environ 23: 1373-1380.

22. Agati G, Biricolti S, Guidi L, Ferrni F, Fini A, et al. (2011) The biosynthesis of flavonoids is enhanced similarly by UV radiation and root zone salinity in L. *vulgare* leaves. J Plant Physiol 168: 204-212.

23. Williams RJ, Spencer JPE, Rice-Evans C (2004) Flavonoids: antioxidants or signaling molecules? Free Radical Biol Med 26: 838-849.

24. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat Biotechnol 29: 644-652.

25. Li RQ, Li YR, Fang XD, Yang HM, Wang J, et al. (2009) SNP detection for massively parallel whole-genome resequencing. Genome Res 19: 1124-1132.

26. Romualdi C, Bortoluzzi S, D'Alessi F, Danieli GA (2003) IDEG6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments. Physiol Genomics 12: 159-162.

27. Huang DW, Sherman BT, Lempicki RA (2008) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44-57.

28. Bräutigam A, Gowik U (2010) What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. Plant Biol 12: 831-841.

29. Hegedus Z, Zakrzewska A, Agoston VC, Ordas A, Racz P, et al. (2009) Deep sequencing of the zebrafish transcriptome response to mycobacterium infection. Mol Immunol 46: 2918-2930.

30. Wang B, Guo G, Wang C, Lin Y, Wang X, et al. (2010) Survey of the transcriptome of Aspergillus oryzae via massively parallel mRNA sequencing. Nucleic Acids Res 38: 5075-5087.

31. Angeloni F, Wangemaker CAM, Jetten MSM, Opdencamp HJM, Janssen-Megens EM, et al. (2011) *De novo* transcriptome characterization and development of genomic tools for *Scabiosa columbaria* L. using next-generation sequencing techniques. Mol Ecol Resour 11: 662-674.

32. Crawford JE, Guelbeogo WM, Sanou A, Traoré A, Vernick KD, et al. (2010) *De novo* transcriptome sequencing in *Anopheles funestus* using illumine RNA-seq technology. PLoS one 5: e14202.

33. Parchman TL, Geist KS, Grahnen JA, Benkman CW, Buerkle CA (2010) Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. BMC Genomics 11: 180-192.

34. Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol 126: 485-493.

35. Wang L, Zhao S, Gu S, Zhou Y, Zhou H, et al. (2013) Deep RNA-seq uncover the peach transcriptome landscape. Plant Mol Biol.

36. Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. Mol Biol Evol 21: 991-1007.

37. Garg R, Patel RK, Tyagi AK, Jain M (2011) *De novo* assembly of Chickpea transcriptome using short reads for gene discovery and marker identification. DNA Res 18: 53-63.

38. Raju NL, Gnanesh BN, Lekha P, Jayashree B, Pande S, et al. (2010) The first set of EST resource for gene discovery and marker development in pigeonpea (*Cajanus cajan* L.). BMC Plant Biol 10: 45-66.

39. Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23: 48-55.

40. Lijavetzky D, Cabezas JA, Ibáñez A, Rodríguez V, Martínez-Zapater JM (2007) High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology. BMC Genomics 8: 424-433.

41. Meyer E, Aglyamova GV, Wang S, Buchanan-Carter J, Abrego D, et al. (2009) Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFlx. BMC Genomics 10: 219-236.

42. He YY, Häder DP (2002) Involvement of reactive oxygen species in the UV-B damage to the *cyanobacterium Anabaena* sp. J Photoch Photobio B 66: 73-80.

43. Jaleel CA, Riadh K, Gopi R, Manivannan P, Inès J, et al. (2009) Antioxidant defense responses: physiological plasticity in higher plants under abiotic constraints. Acta Physiol Plant 31: 427-436.

44. Radyukina NL, Shashukova AV, Makarova SS, Kuznetsov VI V (2011) Exogenous proline modifies differential expression of superoxide dismutase genes in UV-B-irradiated Salvia officinalis plants. Russ J Plant Physiol 58: 51-59.

45. Zu YG, Wei XX, Yu JH, Li DW, Pang HH, et al. (2011) Responses in the physiology and biochemistry of Korean pine (Pinus Koraiensis) under supplementary UV-B radiation. Photosynthetica 49: 448-458.

46. Park JS, Choung MG, Kim JB, Hahn BS, Kim JB, et al. (2007) Genes up-regulated during red coloration in UV-B irradiated lettuce leaves. Plant Cell Rep 26: 507-516.

47. Kusano M, Tohge T, Fukushima A, Kobayashi M, Hayashi N, et al. (2011) Metabolomics reveals comprehensive reprogramming involving two independent metabolic responses of Arabidopsis to UV-B light. Plant J 67: 354-369.

48. Herrmann A, Schulz W, Hahlbrock K (1988) Two alleles of the single-copy chalcone synthase gene in parsley differ by a transposon-like element. Mol Gen Genet 212: 93-98.

49. Singh S, Vishwakarma RK, Kumar RJS, Ruby PDS, Khan BM (2013) Functional characterization of a flavonoid glycosyltransferase gene from *Withania somnifera* (Ashwagandha). Appl biochem Biotechnol 170: 729-741.

Author Affiliations

Top

Institute of Biology, Gansu Academy of Sciences, Lanzhou, 730000, China