Transcriptome analysis reveals salinity responses in four Tartary cultivars

Jin-Nan Song[1], Xue-Hua Liu[1], Ya-Qi Wang[1], Hong-Bing Yang[*]

Key Lab of Plant Biotechnology in Universities of Shandong Province, College of Life Sciences, Qingdao Agricultural University, Qingdao, 266109, China.

Address: No.700 Changcheng Road, Chengyang District, Qingdao, China.

Post Code: 266109.

[1] These authors contributed equally to this work.

[*] Author for correspondence: Hong-Bing Yang. e-mail: hbyang@qau.edu.cn

Running Title: Salt pathway genes in Tartary buckwheat

ABSTRACT

Tartary buckwheat (*Fagopyrum tataricum*) is an annual herbaceous crop with a higher content of rutin and flavonoids than other crops. In order to understand the salt tolerance mechanism of Tartary buckwheat, we conducted a transcriptome sequencing study. Two existing and two new salt-tolerant varieties of Tartary buckwheat were used as experimental materials. These varieties were subjected to 100 mM NaCl for 48 hours treatment at the 3 leaf stage. Leaf RNA was extracted from 8 samples and performed RNA-seq. We *de novo* assembled reads into a transcriptome dataset containing 62,637 unigenes with N50 length of 1596 bp. A total of 205.2 million clean reads were obtained, and 33,495 (53.47%) of them were annotated. GO and KEGG classification showed enrichment of these unigenes in 47 sub-categories, 25 KOG functional categories and 129 pathways, respectively. With 11 contrasts, we obtained many differentially expressed genes (DEGs) of transcriptome data in 8 samples. The relative expression of *Unigene0009545* and *Unigene0057231* in M1 and that of *Unigene0023096* and *Unigene0060321* in M2 was increased significantly under salt stress and may be involved in salt tolerance. These findings will provide valuable reference genetic information for the study on salt tolerance mechanism in plants and assist in cultivating strong salt-tolerant varieties of Tartary buckwheat for greater yield in saline-alkali soil.

Keywords: Tartary buckwheat, transcriptome data, differential expression genes, salt tolerance

1. Introduction

Buckwheat (*Fagopyrum esculentum* Moench) unlike its namesake is a dicot plant, belonging to the family Polygonaceae. It was originally cultivated in Southwest China but has gradually spread to other countries [1]. It has about 20 varieties, most of them are grown in mountain areas, such as Chuanqiao NO.1, which grows in Sichuan Province of China. In some areas, it is cultivated due to its nutrient content. Common buckwheat (*F. esculentum* Moench) and Tartary buckwheat (*F. tataricum* (L.) Gaertn.) are used for food [2], both are suitable for daily consumption. Some studies have shown that Tartary buckwheat is more nutritious than common buckwheat [3].

As a serious global problem, soil salinity is threatening land productivity. It is estimated that 50% of cultivated land will be affected by this problem by the year 2050 [4]. Exposure to salt stress restricts crop growth and consequently reduces agricultural yield [5]. In addition, high salinity increases Na^+ content and destroys the transmission system. More importantly, once plants suffer sodium toxicity, tissues at different levels will be damaged or even destroyed [6]. Under severe salt stress, a large number of salt stress response genes become highly expressed.

Compared with other profiling methods, RNA-seq has the advantage of high transcriptome coverage, precision and sensitivity [7,8]. Tartary buckwheat genome provided many new genes involved in drought and cold stress responses, which was attributed to the expansion of several gene families involved in signal transduction, gene regulation, and membrane transport [9]. With this useful technology, we studied the transcriptome of Tartary buckwheat and established a transcriptome database of salt-related genes with the aim of furthering understanding of salinity responses in this crop.

2. Materials and Methods

2.1. Materials treatment and RNA extraction

Two varieties of Chuanqiao NO.1 (WT1) and Chuanqiao NO.2 (WT2) and two new strains of Chuanqiao NO.1-1 (M1) and Chuanqiao NO.2-1 (M2) were used as experimental materials. All of them belonged to Tartary buckwheat type. M1 and M2 are new strains of WT1 and WT2 selected five years after natural mutagenesis in saline-alkali soil. Thus the salt tolerance of M1 and M2 is reported to be higher than that of WT1 and WT2. The experimental groups were treated with Hoagland nutrient solution containing 100 mM NaCl at the 3 leaf stage, while the control groups were treated with Hoagland nutrient solution without NaCl. Leaf samples were obtained at 0 and 48 hours of treatment time. Total RNA was isolated from each sample using Trizol (Life technologies) reagent following the protocol of the manufacturer, and

the quality and quantity of the RNA were assessed by NanoDrop 2000.

2.2. RNA library construction and sequencing

Magnetic beads with Oligo (dT) were used to enrich eukaryotic mRNA, then added fragmentation buffer to break the mRNA strands to the desired length (200~700 nt). Next, first strand cDNA synthesis was carried out with random hexamer priming. The related buffer, dNTPs, RNase H and DNA polymerase I in one reaction was used to synthesize the second-strand cDNA. With the help of QiaQuick PCR extraction kit for purificating, we added EB buffer solution to the end repair, added poly (A) and ligated the sequencing adaptor, and used agarose gel electrophoresis for fragment size selection. Finally, these fragments underwent PCR amplification using Illumina-compatible primers. The libraries were sequenced using the Illumina HiSeq 2000 instrument to generate 125 bp paired end reads.

2.3. Data quality control and sequence assembly

To get high quality clean reads, we did the following: (1) removed adaptors, (2) removed reads with more than 10% unknown bases (N), (3) removed the low quality reads (Q-value≤5).

We then used Trinity [10] to assemble the short reads into transcripts and subsequently unigenes, followed by functional annotation and coding sequence (CDS) prediction.

2.4. Functional annotation and classification

Using BLASTx (E value < 1e-5), all assembled unigenes were searched against the databases of NCBI Nr, Swiss-Prot, COG and KEGG. KEGG is a database of functional gene pathways, that facilitates pathway annotation of unigene according to its annotation information [11]. GO (Gene Ontology, http://www.geneontology.org/) is another database of gene annotations that provides information on biological process, cellular component and molecular function [12]. Nr (http://www.ncbi.nlm.nih.gov) and SwissProt are two well-known protein databases, and the latter one was strictly filtered to be redundant. Through them we can get more information for function and the highest sequence similarity of protein [13]. COG/KOG (http://www.ncbi.nlm.nih.gov/COG) is a database which can treat the gene products with direct homologous classification. It is constructed based on bacteria, algae and eukaryotic organisms that have the whole genome encoding protein and the system evolution [14].

2.5. Identification of transcription factor families and SSRs

The presence of transcription factors (TF) binding motifs were analyzed with the PlantTFcat online tool (http://plantgrn.noble.org/PlantTFcat) [15]. SSRs (Simple Sequence Repeats), a type of molecular marker, is a potent tool for genetic studies and marker assisted selection (MAS) in many crops [16,17]. In

addition, in this research, we adopted the MISA search tool for SSR screening [18].

2.6. Analysis of differentially expressed genes (DEGs)

To identify differentially expressed genes across samples or groups, the edgeR package was used [19]. We identified genes with a fold change ≥2 and a false discovery rate (FDR) <0.05 in a comparison as significant DEGs. DEGs were then subjected to enrichment analysis of GO functions and KEGG pathways.

In order to validate the RNA-seq results, we utilized RT-qPCR to measure the relative expression level of genes of interest. Based on the quantity of RNA, we synthesized cDNA by reverse transcription reactions with a kit called PrimeScript RT reagent Kit with gDNA Eraser (Takara Biotech, Dalian, China), the Tartary buckwheat *actin* gene was used as an internal control. The primers were designed using NCBI Primer-BLAST and were synthesized by Tsingke company. The real time PCR program consisted at 94 °C for 30 s, 40 cycles of denaturing at 95 °C for 10 s, annealing and extension at 58 °C for 20 s. Finally, data processing and statistical analysis were used the $2^{(-\Delta\Delta Ct)}$ method [20] to analyse the relative expression level of genes.

3. Results

3.1. De novo assembly, classification of clean reads in each sample

As shown in Table 1, the total number of genes identified was 62,637, and the total GC percentage was 41.20%. The length of N50 was 1,596 bp, and longest gene was 16598 bp, revealing that the assembly quality was satisfactory. In addition, we got the length distribution graph of 8 samples summary, the mode genes length was between 200 bp to 299 bp (n=16,857) (Fig. 1), and the number of unigenes between 11 bp to 100 bp was the most (n=27,197) (Fig. 2).

A total of 209,158,434 raw reads were generated. After filtering the reads, we got a total of 205,227,656 clean reads, the Q20 proportion of all clean reads was over 97%, and the Q30 proportion was over 94%. All clean reads have been deposited into the Sequence Read Archive (SRA) under the accession number SRP125065.

3.2. Unigene Functional Annotation

3.2.1 . Gene annotation with Nr, Swiss-Prot, KOG and KEGG

We searched 62,637 unigene sequences against 4 main databases, including Nr, Swiss-Prot, KOG and KEGG, the final annotated results are shown in Table 2. Among the whole unigenes, 33,495 (53.47%) were annotated, while 29,142 (46.53%) were not annotated in any database. Among the annotated unigenes, 33,158 (52.94%) were in Nr database, which contained non-redundant protein sequence with entries based

on BLAST tool [21], 25,430 (40.60%) were in Swiss-Prot database, and 21,489 (34.31%) and 13,620 (21.74%) were in KOG [20] and KEGG [22] database, respectively. As shown in Fig. 3, 10,865 unigenes could be annotated in four databases, suggesting that these unigenes might represent essential genes in Tartary buckwheat.

3.2.2. Distribution statistics of near source varieties analysis

Homologous sequences means some similar sequences or some sequences can be somewhat compared with. If two aligned unigenes had many accordant sequences, they might have similar functions [23]. As shown in Fig. 4, compared with *Fagopyrum tataricum*, the largest number of homologous sequences was *Beta vulgaris* subsp. *vulgaris*, which had 6,195 homologous unigenes, the second was *Vitis vinifera* with 2,984 homologous unigenes, and the third was *Theobroma cacao* with 2,299 homologous unigenes.

3.2.3. KOG function and gene ontology (GO) classification of unigenes

The KOG database is a protein pool which contains many secondary databases, it comprises sequences classified into functional categories, and it has 78,096 protein databases, where 24,154 proteins from *Arabidopsis thaliana* (Ath), 17,101 from *Caenorhabditis elegans* (Cel), 10,517 from *Drosophila melanogaster* (Dme) and 26,324 from *Homo sapiens* (Hsa) [24]. Among the 62,637 unigenes in the final set, 21,489 unigenes were annotated and classified into 25 KOG categories, the largest group of unigenes (5,243 members, 21.97%) was assigned to the R class (General function prediction only); T class (Signal transduction mechanisms) accounted for 2,675 unigenes (11.21%); O class (Posttranslational modification, protein turnover, chaperones) accounted for 2,326 unigenes (9.75%); with the number of 9 unigenes (0.038%), N class (Cell motility) was the smallest (Fig. 5; Table S1). It is known that different varieties have different conditions so that they can yield different classifications [25]. We may find some useful genes in V class (defense mechanisms) of salt-stress buckwheat. This will be of value for future studies on salt-stress mechanism in *Fagopyrum tataricum* (L.) Gaertn.

With the help of BLAST2 GO software, we completed the search of GO database and facilitation of global analysis of gene expressions. GO classification was carried on the annotated unigenes which showed that mainly involved three biological functions: 59,409 (94.85%) unigenes annotated to “biological process”, 37,902 (60.51%) unigenes annotated to “cell component”, 21,526 (34.37%) unigenes annotated to “molecular function” (Table S2).

Using the GO function classification, the global unigenes in 3 biological function categories were divided into 47 biological function subgroups. Among the global GO classifications, the largest fraction of

genes is in “metabolic process”, which is assigned with 12,473 (21.00%) unigenes, while for cellular component, “cell”, was 9,128 (24.08%) unigenes, and for molecular function, “catalytic activity”, was assigned 10,560 unigenes (49.06%) (Fig. 6; Table S2).

3.2.4. SSR analysis of *F. tataricum*

We identified 3,355 SSRs obtained from 62,637 sequences totaling 56,788,193 bp. A total of 3,025 sequences contained SSRs, in which 289 sequences contained more than a single SSR, and 121 SSRs exhibited in compound formation. The most abundant SSR type was tri-nucleotide (1588, 47.33%), followed by di-nucleotide (1234, 36.78%), tetra-nucleotide (294, 8.76%), hexa-nucleotide (144, 4.29%) and penta-nucleotide (95, 2.84%) (Table 3).

The repeat counts range from 4 to 15+. SSRs with 5 repeats were the most abundant, followed by those with 6 repeats. Among the whole SSRs, AT/AT was the most abundant motif (20.10%), followed by AAG/CTT (14.10%), and the third was AG/CT (13.40%) (Fig. 7). These results were consistent with the SSRs and GC distributions in *Arabidopsis* [26] and provide a solid foundation for future SSR marker development.

3.2.5. Identification of differentially expressed genes

In order to identify genes that exhibit altered expression in the varieties (WT1, WT2, M1 and M2) and treatment groups (CK and T), we undertook differential expression analysis. Many DEGs (differentially expressed genes) were identified in these 11 contrasts conducted (Fig. 8). The DEGs altered by salinity stress in the four cultivars were intersected with a Venn diagram (Fig. 9). There were 42 unigenes common to the four varieties (detailed in Table S3).

3.2.6. The differences between samples expresses in GO/Pathway

We completed GO term enrichment by Blast2GO to further understand the biological function of DEGs. Up-regulated DEGs under salt treatment were enriched in ‘cell morphogenesis’, ‘cellular component morphogenesis’, ‘oxidation-reduction process’, ‘catalytic activity’, ‘cellular process’, ‘metabolic process’, ‘single-organism process’, ‘binding’ and ‘cell part’. Down-regulated DEGs under salt treated conditions were ‘growth’, ‘localization’, ‘cellular component organization or biogenesis’, ‘catalytic activity’, ‘membrane part’ and ‘organelle part’. These results suggest that the biological functions of Tartary buckwheat may be activated by salt stress (Fig. 10 A-D; Table S4 A-D).

For further predicting which pathways these DEGs may be involved in, we conducted KEGG enrichment analysis. The consequences revealed that the DEGs were over-represented in pathway terms

'Plant hormone signal transduction', 'Biosynthesis of secondary metabolites' and 'Metabolic pathways' (Fig. 11 A-D; Table S5 A-D).

3.2.7. Identification of transcription factor families

Transcription factors (TF) play vital roles in plant development and abiotic stress responses [27]. With the PlantTFcat online tool, we identified 57 transcription factor families, including 7,208 unigenes. The largest families were bHLH (741, 10.28%), NAC (541, 7.51%), MYB (430, 5.97%), WRKY (372, 5.16%), B3 (360, 4.99%), FAR1 (320, 4.44%), ERF (314, 4.36%), C3H (301, 4.18%), bZIP (296, 4.11%) and C2H2 (288, 4.00%) (Fig. 12; Table S6).

3.2.8. Verification of differentially expressed genes by RT-qPCR

To validate RNA expression changes measured by RNA-Seq, we carried out RT-qPCR analysis on nine unigenes which were altered by salt stress: SOD [28], POD [29], APX [30], SR(salt-related), Na^+/H^+ [31] and SOS (salt overly sensitive) [32]. The SOD-related genes contained *Unigene0023096* and *Unigene0017446*, the POD-related genes contained *Unigene0011757* and *Unigene0057231*, the APX-related genes contained *Unigene0029688* and *Unigene0060321*, the SR (salt-related) gene contained *Unigene0009545*, the Na^+/H^+ antiport gene was *FtNHX1* of *Unigene0036375*, SOS (salt overly sensitive) gene was *Unigene0038654*. We got each primer (Table 4) from NCBI primer tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). *F. tataricum Actin* [33] was used as an endogenous control. The results showed that the expression data from RNA-seq were accurate and reliable (Table S7). In addition, Table S7 showed that the relative expression of *Unigene0009545* and *Unigene0057231* in M1 was increased obviously under salt stress compared with the control, which was increased by 99.74% and 294.61%, respectively, while that of *Unigene0023096* and *Unigene0060321* in M2 was increased obviously under salt stress compared with the control, which was increased by 211.48% and 73.08%, respectively. Taken together, we have validated 9 genes that were identified with RNA-seq with RT-qPCR and suggest these genes may be involved in salt tolerance.

4. Discussion

There are more than 1 billion hm^2 of saline soil in the world [34]. Saline soil is an important land management problem and reduces crop yields. Therefore, it is urgent to improve crop salt tolerance or reduce saline soil. On the former issue, the research in recent years has focused on the mechanism of salt tolerance in plants [35]. On the latter issue, a soil salinization assessment provides an indication of the severity of soil salinization in an area [36]. The use of prevention and control has become the mainstream

method to reduce saline soil. In recent years, transcriptome analysis has become widely used, for example in maize (*Zea mays* L.) [37] and *Spodoptera litura* [38], to provides a comprehensive survey of gene expression changes and some insights into specific biological processes [39]. On this basis, we carried out transcriptome analysis to understand the mechanism of salt response in Tartary buckwheat.

We previously found that Chuanqiao NO.1 was more salt tolerant than Chuanqiao NO.2 , then cultivated them in saline soil for five years and obtained two new strains: Chuanqiao NO.1-1 and Chuanqiao NO.2-1. Using *de novo* transcriptome sequencing assembly on these four cultivars, we obtained 62,637 unigenes with N50 length of 1596 bp and total length of 56,788,193 bases. Assembly quality can be evaluated by N50 value [40]. According to previously reported transcriptome data, Wu et al. [41] found that the salt stress response transcriptome yielded 57,921 unigenes with N50 length of 1,400 bp and a total length of 44.5 Mb bases in *F. tataricum*. In total, we obtained 62,637 assembly unigenes, and of these 33,495 (53.47%) were annotated in four large databases, our high-throughput RNA-seq yielded more unigenes but with slightly shorter N50. The largest number of annotated unigenes was the 'biological process' group, with the number of 59,409 (94.85%). Also, 'metabolic process', 'cellular process', 'single-organism process' and 'biological regulation' were the top four GO sub-categories, and 'metabolic process' and 'biological regulation' were overrepresented under salt stress. In addition, the results of KOG classification showed that metabolic pathways were over-expressed in terms of 'General function prediction only', 'Post-translational modification, protein turnover, chaperones' and 'Signal transduction mechanisms'. This provides more important information for the future study on molecular mechanism of salt tolerance in plants.

The comparison of gene expression patterns between control and salt treatment group: WT1-CK VS WT1-T, M1-CK VS M1-T, WT2-CK VS WT2-T, M2-CK VS M2-T, there were 4049 DEGs, 4167 DEGs, 6752 DEGs and 2239 DEGs, respectively, all containing down and up-regulated genes. There were 42 common DGEs in the four cultivars. The number of DEGs in each variety was larger than that reported for *Arabidopsis* [42]. It indicated that these 42 common DEGs may be involved in salinity response.

Our TF classification results identified that the bHLH [43], NAC [44], MYB [45] and WRKY [46] families are the four largest. It is possible that some of these TFs may be involved in salt response.

bHLH is an important transcription factor involved in environmental stress response [47]. *OrbHLH2* was cloned from *Oryzara fipogon* Griff by Zhou et al. [48], which is highly homologous to ICE1 protein. We obtained 741 DEGs belonging to the bHLH transcription factor family, such as *Unigene0005747* and

Unigene0006105. NAC is a TF class unique to plants and is involved in response to both biotic and abiotic stresses including low temperature, mechanical damage, drought and salinity. Ganesan et al. [49] found that *AmNAC1* played an important role in early salt stress response and long-term salt regulation in mangrove plant *Avicennia marina*. A total of 541 DEGs belonging to the NAC transcription factor family were obtained, such as *Unigene0031578*, *Unigene0033494*, etc. As an important transcription factor family, they may play important roles in salt tolerance of *F. tataricum*. MYB transcription factors contribute to plant development, secondary metabolism, hormone signal transduction, disease resistance and abiotic stress tolerance [50]. According to the study of Zhang et al. [51] in *Arabidopsis*, high salt stress activates the expression of MYB-related gene *AtMYBL*. *Unigene0027989* and *Unigene0027988* belong to the MYB transcription factor family, and are likely key regulators of salinity stress response in *F. tataricum*. We can obtain many genes that response to MYB transcription factor of *F. tataricum* under salt stress. As an important transcription factor superfamily, WRKY transporters are involved in response to environmental stimuli, such as high salt, low temperature, etc [52]. This transcription factor family in *F. tataricum* is highly up-regulate under salt stress, which has the complexity and importance of stress regulation. *Unigene0043192*, *Unigene0043195* and *Unigene0043300* in *F. tataricum* have high sequence similarity to homologues in *Oryza barthii*, and they are highly expressed under salinity stress. In conclusion, bHLH, NAC, MYB and WRKY are the largest transcription factors families under salt stress in *F. tataricum*. Among these four transcription factors families, more salt-related genes will likely be found.

Salt stress increases the activities of antioxidant enzyme (SOD, POD, APX) and improves the adaptability of plants to adversity, such as cotton [53] and maize [54]. Under salinity, expression levels of *NHX1* and *SOS1* genes increase, for example *AeNHX1* in *Arabidopsis* [55], which likely improves the salt tolerance of *Arabidopsis thaliana*; and SOS1 that improves salt tolerance in *S. pombe* [56]. The corresponding genes were up-regulated in *F. tataricum*. In addition, the expression levels of *FtNHX1* and *FtSOS1* were higher under salt stress. These key salinity-response associated genes were found and clarified in *F. tataricum*. The salt tolerance of new strains Chuanqiao No.1-1 and Chuanqiao No.2-1 was higher than that of the original varieties. Under salt stress, the relative expression of *Unigene0009545* and *Unigene0057231* in Chuanqiao No.1-1 and that of *Unigene0023096* and *Unigene0060321* in Chuanqiao No.2-1 was increased significantly. The high salt tolerance of new strains in Tartary buckwheat may be related to the high expression of these four genes.

5. Conclusion

With *de novo* sequence assembly, we generated and annotated a salt-responsive transcriptome of *F. tataricum*. Though RNA-seq, we obtained expression profiles that highlighted differences between varieties and response to salinity. Comparison of control and salt treatment groups led to the identification of many key regulatory factors associated with salinity response. We also found some genes related to antioxidant enzymes and important transcriptional regulators under salt stress. In addition, we have selected two new salt-tolerant strains, which provide a reasonable basis for the study on salt tolerance in *F. tataricum*. The data and results will contribute to the understanding of salt tolerance mechanisms in plants and provide molecular and genomic information for future research.

Contribution

Song Jin-Nan analyzed the relative expression level of genes and drafted the manuscript; Liu Xue-Hua analyzed the transcriptome data of Tartary buckwheat; Wang Ya-Qi cultivated the plant and treated; Yang Hong-Bing designed the study and helped draft the manuscript. All authors gave final approval for publication.

Acknowledgements

The authors would like to acknowledge the financial support from the National Natural Science Foundation of China (31371552), and acknowledge Dr. Mark Ziemann of Deakin University in Australia for language editing.

Appendix A. Supplementary data

Raw reads have been deposited to the NCBI Sequence Read Archive under the accession ID SRP125065: https://www.ncbi.nlm.nih.gov/sra/?term=SRP125065.

References

1 Fabjan N, Rode J, Košir IJ, Wang Z, Zhang Z, Kreft I. 2003 Tartary buckwheat (*Fagopyrum tataricum* Gaertn.) as a source of dietary rutin and quercitrin. *J. Agric. Food Chem.* **51**, 6452–6455. (doi:10.1021/jf034543e)

2 Bonafaccia G, Galli V, Francisci R, Mair V, Skrabanja V, Kreft I. 2000 Characteristics of spelt wheat products and nutritional value of spelt wheat-based bread. *Food Chem.* **68**, 437–441. (doi:10.1016/S0308-8146(99)00215-0)

3 Bonafaccia G, Gambelli L, Fabjan N, Kreft I. 2003 Trace elements in flour and bran from common and Tartary buckwheat. *Food Chem.* **83**, 1–5. (doi:10.1016/s0308-8146(03)00228-0)

4 Butcher K, Wick AF, Desutter T, Chatterjee A, Harmon J. 2016 Soil salinity: a threat to global food security. *Agron. J.* **108**, 2189–2200. (doi:10.2134/agronj2016.06.0368)

5 Li D, Su Z, Dong J, Wang T. 2009 An expression database for roots of the model legume *Medicago truncatula* under

salt stress. *BMC Genomics* **10**, 517. (doi:10.1186/1471-2164-10-517)

6 Tester M, Davenport R. 2003 Na^+ tolerance and Na^+ transport in higher plants. *Ann. Bot.* **91**, 503–527. (doi:10.1093/aob/mcg058)

7 Trapnell C, Pachter L, Salzberg SL. 2009 TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111. (doi:10.1093/bioinformatics/btp120)

8 Wang Z, Gerstein M, Snyder M. 2009 RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63. (doi:10.1038/nrg2484)

9 Zhang L, Li X, Ma B, Gao Q, Du H, Han Y, Li Y, Cao Y, Qi M, Zhu Y, Lu H, Ma M, Liu L, Zhou J, Nan C, Qin Y, Wang J, Cui L, Liu H, Liang C, Qiao Z. 2017 The Tartary buckwheat genome provides insights into rutin biosynthesis and abiotic stress tolerance. *Mol. Plant* **10**, 1224–1237. (doi: 10.1016/j.molp.2017.08.013)

10 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652. (doi:10.1038/nbt.1883)

11 Kanehisa M, Goto S. 2000 KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. (doi:10.1093/nar/28.1.27)

12 Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, Sethuraman A, Theesfeld CL, Botstein D, Dolinski K, Feierbach B, Berardini T, Mundodi S, Rhee SY, Apweiler R, Barrell D, Camon E, Dimmer E, Lee V, Chisholm R, Gaudet P, Kibbe W, Kishore R, Schwarz EM, Sternberg P, Gwinn M, Hannick L, Wortman J, Berriman M, Wood V, de la Cruz N, Tonellato P, Jaiswal P, Seigfried T, White R, Gene Ontology C. 2004 The gene ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32**, D258–261. (doi:10.1093/nar/gkh036)

13 Louie B, Higdon R, Kolker E. 2009 A statistical model of protein sequence similarity and function similarity reveals overly-specific function predictions. *PLoS One* **4**(10), e7546. (doi:10.1371/journal.pone.0007546)

14 Liu F, Sun X, Wang W, Liang Z, Wang F. 2014 De novo transcriptome analysis-gained insights into physiological and metabolic characteristics of *Sargassum thunbergii* (Fucales, Phaeophyceae). *J. Appl. Phycol.* **26**, 1519–1526. (doi:10.1007/s10811-013-0140-2)

15 Dai X, Sinharoy S, Udvardi M, Zhao PX. 2013 PlantTFcat: an online plant transcription factor and transcriptional

regulator categorization and analysis tool. *BMC Bioinformatics* **14**, 321. (doi:10.1186/1471-2105-14-321)

16 Wei W, Qi X, Wang L, Zhang Y, Hua W, Li D, Lv H, Zhang X. 2011 Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. *BMC Genomics* **12**, 451. (doi:10.1186/1471-2164-12-451)

17 Yang Z, Chen Z, Peng Z, Yu Y, Liao M, Wei S. 2017 Development of a high-density linkage map and mapping of the three-pistil gene (Pis1) in wheat using GBS markers. *BMC Genomics* **18**, 567. (doi:10.1186/s12864-017-3960-7)

18 Thiel T, Michalek W, Varshney RK, Graner A. 2003 Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **106**, 411–422. (doi:10.1007/s00122-002-1031-0)

19 Robinson MD, McCarthy DJ, Smyth GK. 2010 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**(1), 139–140. (doi: 10.1093/bioinformatics/btp616)

20 Livak KJ, Schmittgen TD. 2001 Analysis of relative gene expression data using real-time quantitative PCR and the 2-$^{\triangle\triangle}$CT method. *Methods* **25**, 402–408. (doi:10.1006/meth.2001.1262)

21 Pruitt KD, Tatusova T, Maglott DR. 2007 NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D61–65. (doi:10.1093/nar/gki025)

22 Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2012 KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**, D109–114. (doi:10.1093/nar/gkr988)

23 Thim L. 1988 A surprising sequence homology. *Biochem. J.* **253**, 309. (doi:10.1042/bj2530309a)

24 Mudado MA, Ortega JM. 2006 A picture of gene sampling/expression in model organisms using ESTs and KOG proteins. *Genet. Mol. Res.* **5**, 242–253. (doi:10.1590/S1415-47572006000200033)

25 Zhu Y, Wang X, Huang L, Lin C, Zhang X, Xu W, Peng J, Li Z, Yan H, Luo F, Wang X, Yao L, Peng D. 2017 Transcriptomic identification of drought-related genes and SSR markers in sudan grass based on RNA-Seq. *Front. Plant Sci.* **8**, 687. (doi:10.3389/fpls.2017.00687)

26 Lawson MJ, Zhang L. 2006 Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. *Genom. Biol.* **7**, R14. (doi:10.1186/gb-2006-7-2-r14)

27 Nakashima K, Yamaguchi-Shinozaki K. 2009 Transcriptional regulatory networks in response to abiotic stresses in *Arabidopsis* and grasses. *Plant Physiol.* **149**, 88–95. (doi:10.1104/pp.108.129791)

28 Askira Y, Rubin B, Rabinowitch HD. 1991 Differential response to the herbicidal activity of delta-aminolevulinic acid in plants with high and low SOD activity. *Free Radic. Res. Commun.* **13**(1), 837–843. (doi:10.3109/10715769109145865)

29 Shah K, Nahakpam S. 2012 Heat exposure alters the expression of SOD, POD, APX and CAT isozymes and mitigates low cadmium toxicity in seedlings of sensitive and tolerant rice cultivars. *Plant Physiol. Biochem.* **57**(3), 106–113. (doi:10.1016/j.plaphy.2012.05.007)

30 Büyük İ, Aydın SS, Duman DC, Aras S. 2014 Expression analysis of APX and CAT genes in eggplants subjected to Cu^{+2} and Zn^{+2} heavy metals. *New Biotech.* **31**, S138. (doi:10.1016/j.nbt.2014.05.1956)

31 Krulwich TA. 1983 Na^{+}/H^{+} antiporters. Biochim. Biophys. *Acta* **726**, 245–264. (doi:10.1016/0304-4173(83)90011-3)

32 Little JW, Mount DW, 1982 The SOS regulatory system of *Escherichia coli*. *Cell* **29**, 11–22. (doi:10.1016/0092-8674(82)90085-X)

33 Mitchison TJ, Cramer LP. 1996 Actin-based cell motility and cell locomotion. *Cell* **84**, 371–379. (doi:10.1016/S0092-8674(00)81281-7)

34 Metternicht GI, Zinck JA. 2003 Remote sensing of soil salinity: potentials and constraints. *Remote Sens. Environ.* **85**, 1–20. (doi:10.1016/S0034-4257(02)00188-8)

35 Belver A, Olias R, Huertas R, Rodriguez-Rosales MP. 2012 Involvement of SlSOS2 in tomato salt tolerance. *Bioengineered* **3**, 298–302. (doi:10.4161/bioe.20796)

36 Letey J, Hoffman GJ, Hopmans JW, Grattan SR, Suarez D, Corwin DL, Oster JD, Wu L, Amrhein C. 2011 Evaluation of soil salinity leaching requirement guidelines. *Agric. Water Manage.* **98**, 502–506. (doi:10.1016/j.agwat.2010.08.009)

37 Jaiswal P. 2013 Maize metabolic network construction and transcriptome analysis. *Plant Genome* **6**, 1–12. (doi:10.3835/plantgenome2012.09.0025)

38 Gu J, Huang LX, Gong YJ, Zheng SC, Liu L, Huang LH, Feng QL. 2013 De novo characterization of transcriptome and gene expression dynamics inepidermis during the larval-pupal metamorphosis of common cutworm. *Insect Biochem. Mol. Biol.* **43**, 794–808. (doi:10.1016/j.ibmb.2013.06.001)

39 Nagalakshmi U, Waern K, Snyder M. 2010 RNA-Seq: a method for comprehensive transcriptome analysis. *Curr. Protoc. Mol. Biol.* **4**, 11–13. (doi:10.1002/0471142727.mb0411s89)

40 Makinen V, Salmela L, Ylinen J. 2012 Normalized N50 assembly metric using gap-restricted co-linear chaining. *BMC Bioinformatics* **13**(1), 255. (doi:10.1186/1471-2105-13-255)

41 Wu Q, Bai X, Zhao W, Xiang D, Wan Y, Yan J, Zou L, Zhao G. 2017 De novo assembly and analysis of Tartary buckwheat (*Fagopyrum tataricum* Gaertn.) transcriptome discloses key regulators involved in salt-stress response. *Genes* **8**(10), 255. (doi:10.3390/genes8100255)

42 Taji T, Seki M, Satou M, Sakurai T, Kobayashi M, Ishiyama K, Narusaka Y, Narusaka M, Zhu JK, Shinozaki K. 2004 Comparative genomics in salt tolerance between *Arabidopsis* and *Arabidopsis*-related halophyte salt cress using

Arabidopsis microarray. *Plant Physiol.* **135**(3), 1697–1709. (doi:10.1104/pp.104.039909)

43 Satou Y, Imai KS, Levine M, Kohara Y, Rokhsar D, Satoh N. 2003 A genomewide survey of developmentally relevant genes in *Ciona intestinalis* - I. Genes for bHLH transcription factors. *Dev. Genes Evol.* **213**, 213–221. (doi:10.1007/S00427-003-0319-7)

44 Olsen AN, Ernst HA, Leggio LL, Skriver K. 2005 NAC transcription factors: structurally distinct, functionally diverse. *Trends Plant Sci.* **10**, 79–87. (doi:10.1016/j.tplants.2004.12.010)

45 Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. 2010 MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* **15**, 573.(doi:10.1016/j.tplants.2010.06.005)

46 Eulgem T, Somssich IE. 2007 Networks of WRKY transcription factors in defense signaling. *Curr. Opin. Plant Biol.* **10**, 366–371. (doi:10.1016/j.pbi.2007.04.020)

47 Song XM, Huang ZN, Duan WK, Ren J, Liu TK, Li Y, Hou XL. 2014 Genome-wide analysis of the bHLH transcription factor family in Chinese cabbage (*Brassica rapa* ssp. pekinensis). *BMC Genomics* **289**, 77. (doi:10.1007/s00438-013-0791-3)

48 Zhou J, Li F, Wang JL, Ma Y, Chong K, Xu YY. 2009 Basic helix-loop-helix transcription factor from wild rice (OrbHLH2) improves tolerance to salt- and osmotic stress in *Arabidopsis*. *J. Plant Physiol.* **166**, 1296–1306. (doi:10.1016/j.jplph.2009.02.007)

49 Ganesan G, Sankararamasubramanian HM, Narayanan JM, Sivaprakash KR, Parida A. 2008 Transcript level characterization of a cDNA encoding stress regulated NAC transcription factor in the mangrove plant *Avicennia marina*. *Plant Physiol. Biochem.* **46**, 928–934. (doi:10.1016/j.plaphy.2008.05.002)

50 Allan AC, Hellens RP, Laing WA. 2008 MYB transcription factors that colour our fruit. *Trends Plant Sci.* **13**(3), 99–102. (doi:10.1016/j.tplants.2007.11.012)

51 Zhang X, Ju HW, Chung MS, Huang P, Ahn SJ, Kim CS. 2011 The R-R-type MYB-like transcription factor, AtMYBL, is involved in promoting leaf senescence and modulates an abiotic stress response in *Arabidopsis*. *Plant Cell Physiol.* **52**, 138–148. (doi:10.1093/pcp/pcq180)

52 Zhang C, Wang D, Yang C, Kong N, Zheng S, Peng Z, Nan Y, Nie T, Wang R, Ma H. 2017 Genome-wide identification of the potato WRKY transcription factor family. *PLoS One* **12**, e0181573. (doi:10.1371/journal.pone.0181573)

53 Gossett DR, Millhollon EP, Lucas MC, Banks SW, Marney MM. 1994 The effects of NaCl on antioxidant enzyme activities in callus tissue of salt-tolerant and salt-sensitive cotton cultivars (*Gossypium hirsutum* L.). *Plant Cell Rep.* **13**, 498–503. (doi:10.1007/bf00232944)

54 Azooz MM, Ismail AM, Elhamd MFA. 2009 Growth, lipid peroxidation and antioxidant enzyme activities as a selection

criterion for the salt tolerance of maize cultivars grown under salinity stress. *Int. J. Agric. Biol.* **11**, 572–577. (doi:10.15439/2014F360)

55 Qiao WH, Zhao XY, Li W, Luo Y, Zhang XS. 2007 Overexpression of AeNHX1, a root-specific vacuolar Na^+/H^+ antiporter from *Agropyron elongatum*, confers salt tolerance to *Arabidopsis* and *Festuca* plants. *Plant Cell Rep.* **26**, 1663–1672. (doi:10.1007/s00299-007-0354-3)

56 Ullah A, Dutta D, Fliegel L. 2016 Expression and characterization of the SOS1 *Arabidopsis* salt tolerance protein. *Mol. Cell Biochem.* **415**, 133–143. (doi:10.1007/s11010-016-2685-2)

Legends

Table 1. Characteristics of unigenes.

Table 2. Major characteristics in *de novo* assembled data of 8 samples.

Table 3. Type statistics of SSRs.

Table 4. Primer names and sequences of RT-qPCR.

Fig. 1. Length distribution of genes in *Fagopyrum tataricum*.

Fig. 2. Number of reads per unigene distribution in *Fagopyrum tataricum*.

Fig. 3. Four database annotation intersection with Venn diagram.

Fig. 4. Distribution statistics of near source varieties (Top 10).

Fig. 5. KOG function classification of *Fagopyrum tataricum* unigenes (Note: 21,489 unigenes are annotated and classified into 25 KOG categories from the final set of 62,637 unigenes).

Fig. 6. GO classification of unigenes in *Fagopyrum tataricum*.

Fig. 7. Distribution of SSR motifs in *Fagopyrum tataricum*.

Fig. 8. Different expression of genes among samples.

Fig. 9. Intersection of DEGs of four varieties Venn diagram.

Fig. 10A. GO classification of DEGs between WT1-CK and WT1-T.

Fig. 10B. GO classification of DEGs between M1-CK and M1-T.

Fig. 10C. GO classification of DEGs between WT2-CK and WT2-T.

Fig. 10D. GO classification of DEGs between M2-CK and M2-T.

Fig. 11A. KEGG classification of DEGs between WT1-CK and WT1-T.

Fig. 11B. KEGG classification of DEGs between M1-CK and M1-T.

Fig. 11C. KEGG classification of DEGs between WT2-CK and WT2-T.

Fig. 11D. KEGG classification of DEGs between M2-CK and M2-T.

Fig. 12. The top 10 numbers of TF classification statistics.

Supplements

Table S1. *Fagopyrum tataricum*-Unigene.fa.kog.class.annot.

Table S2. *Fagopyrum tataricum*-Unigene.fa.GO2gene.

Table S3. Common DEGs of four groups under salt stress.

Table S4A. Gene Ontology of DEGs between WT1-CK and WT1-T.

Table S4B. Gene Ontology of DEGs between M1-CK and M1-T.

Table S4C. Gene Ontology of DEGs between WT2-CK and WT2-T.

Table S4D. Gene Ontology of DEGs between M2-CK and M2-T.

Table S5A. Pathway annotation of DEGs between WT1-CK and WT1-T.

Table S5B. Pathway annotation of DEGs between M1-CK and M1-T.

Table S5C. Pathway annotation of DEGs between WT2-CK and WT2-T.

Table S5D. Pathway annotation of DEGs between M2-CK and M2-T.

Table S6. TF. Class.

Table S7. Validation of RNA sequence (RNA-Seq) expression profiles by real-time PCR.

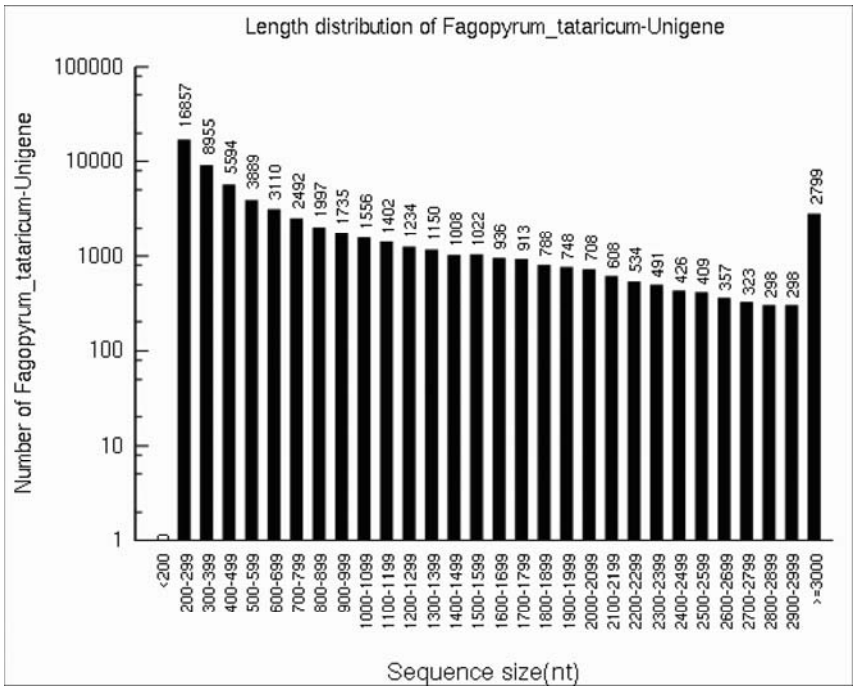

Fig. 1. Length distribution of genes in *Fagopyrum tataricum*.

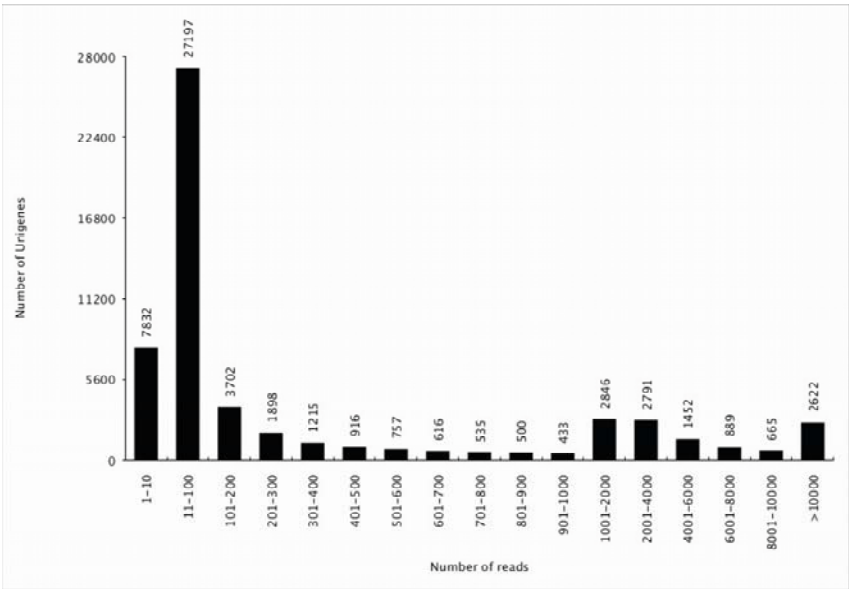

Fig. 2. Number of reads per unigene distribution in *Fagopyrum tataricum*.

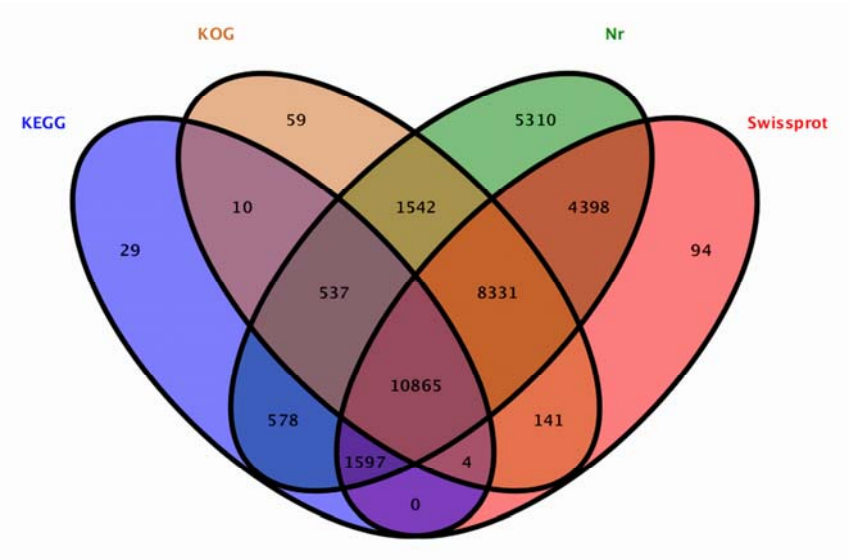

Fig. 3. Four database annotation intersection with Venn diagram.

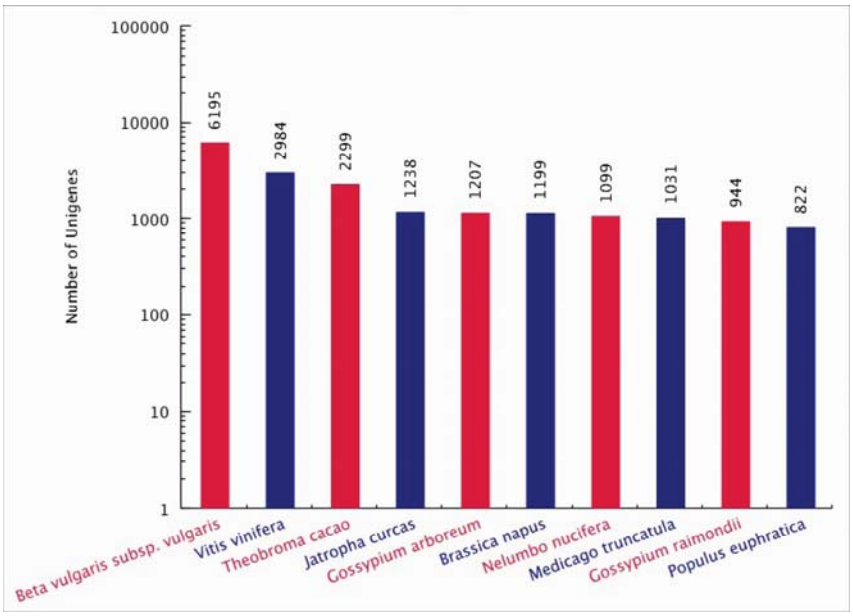

Fig. 4. Distribution statistics of near source varieties (Top 10).

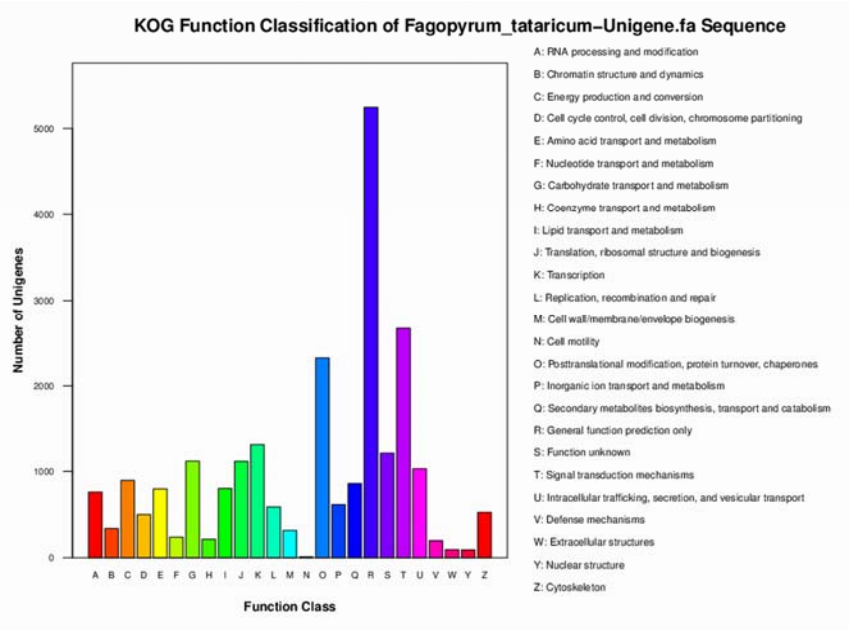

Fig. 5. KOG function classification of *Fagopyrum tataricum* unigenes (Note: 21,489 unigenes are annotated and classified into 25 KOG categories from the final set of 62,637 unigenes).

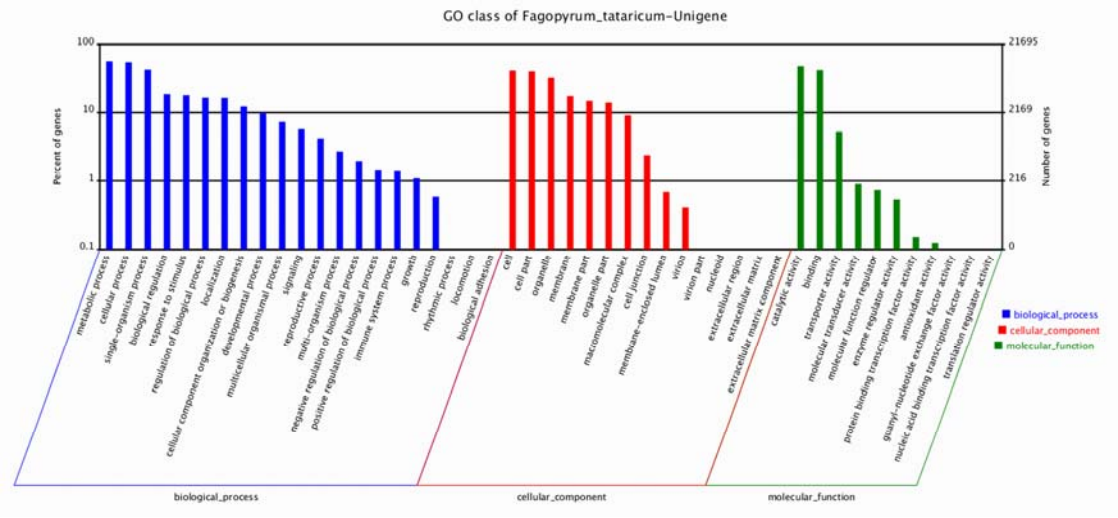

Fig. 6. GO classification of unigenes in *Fagopyrum tataricum*.

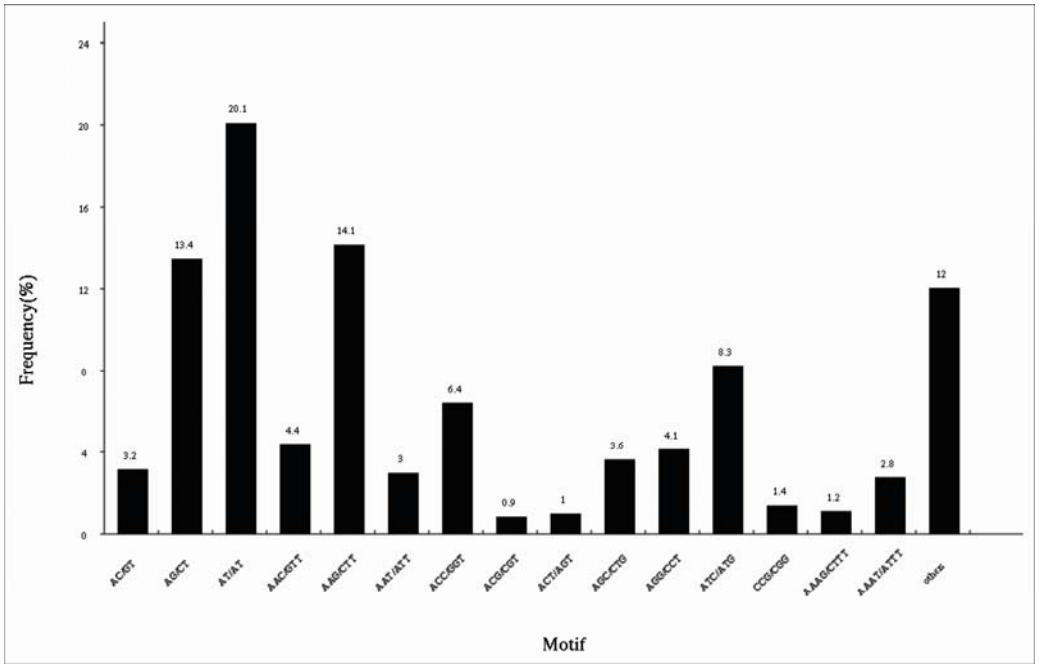

Fig. 7. Distribution of SSR motifs in *Fagopyrum tataricum*.

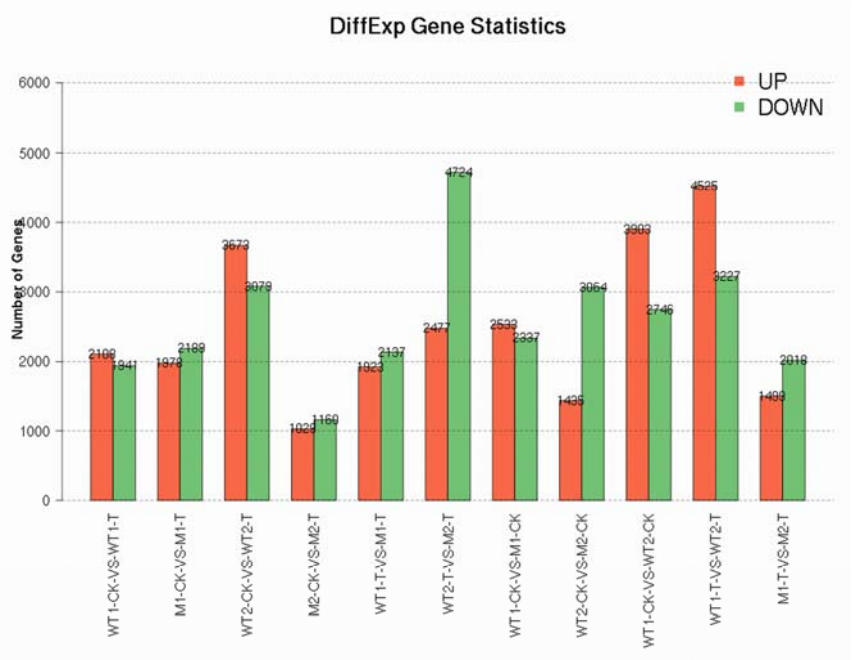

Fig. 8. Different expression of genes among samples.

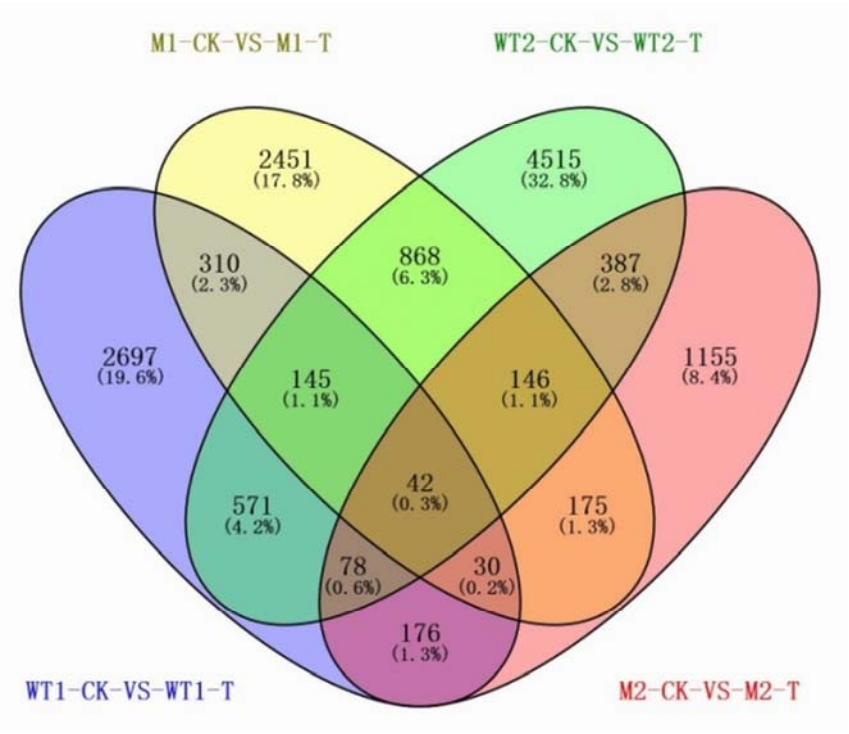

Fig. 9. Intersection of DEGs of four varieties Venn diagram.

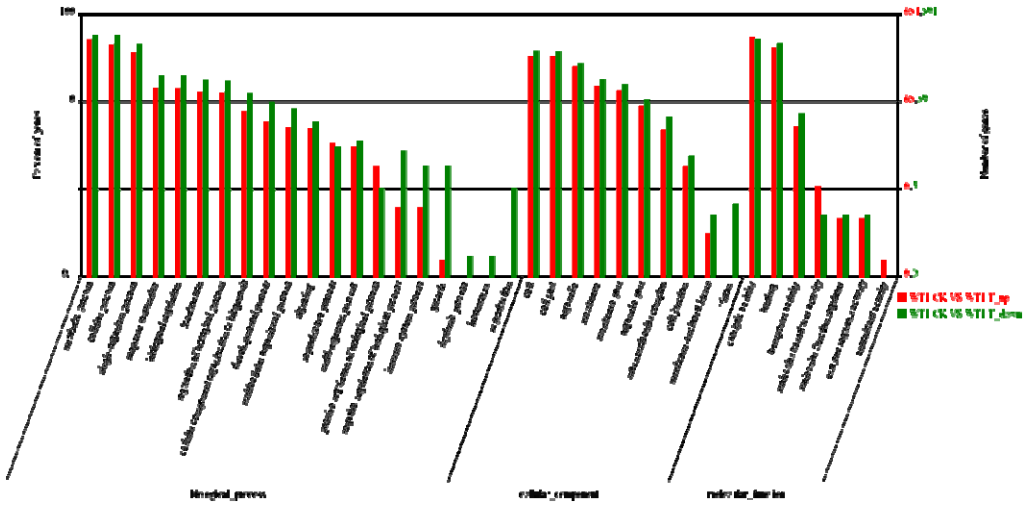

Fig. 10A. GO classification of DEGs between WT1-CK and WT1-T.

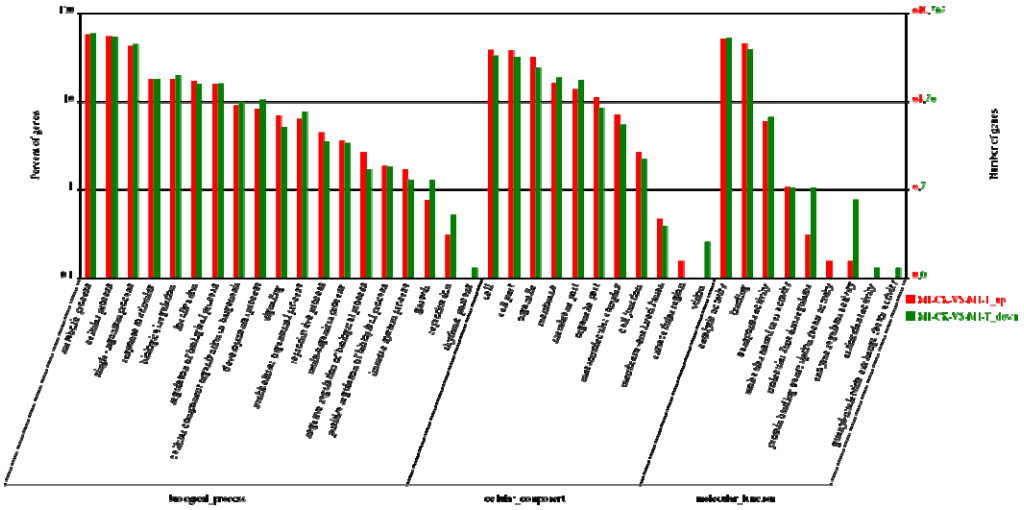

Fig. 10B. GO classification of DEGs between M1-CK and M1-T.

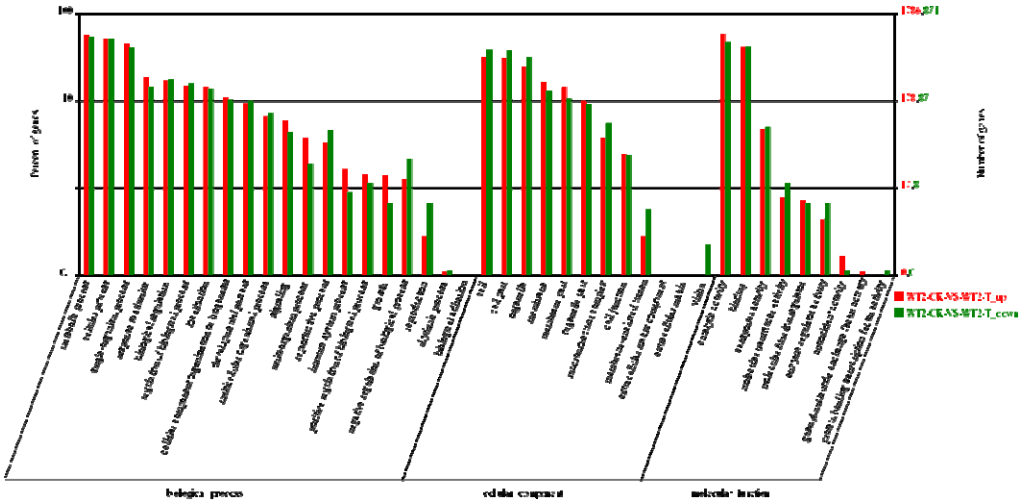

Fig. 10C. GO classification of DEGs between WT2-CK and WT2-T.

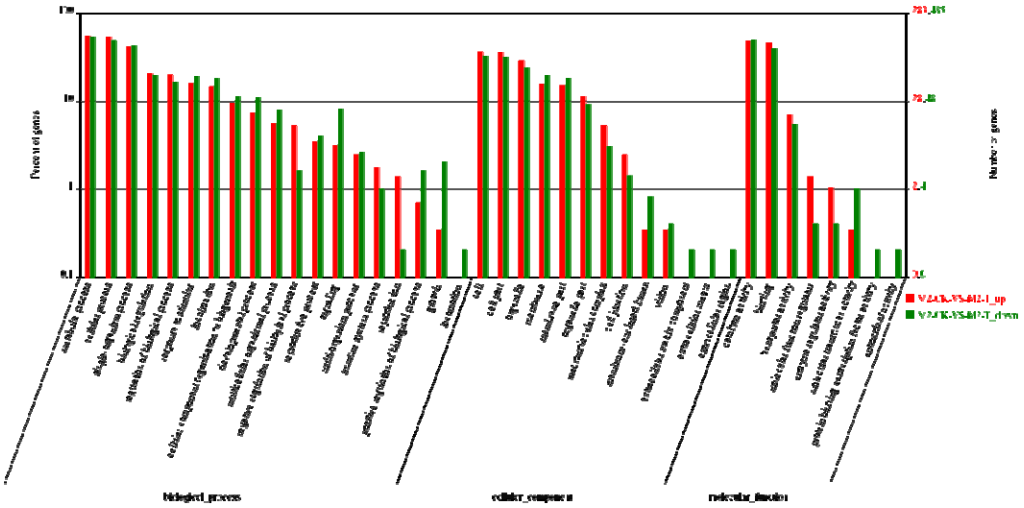

Fig. 10D. GO classification of DEGs between M2-CK and M2-T.

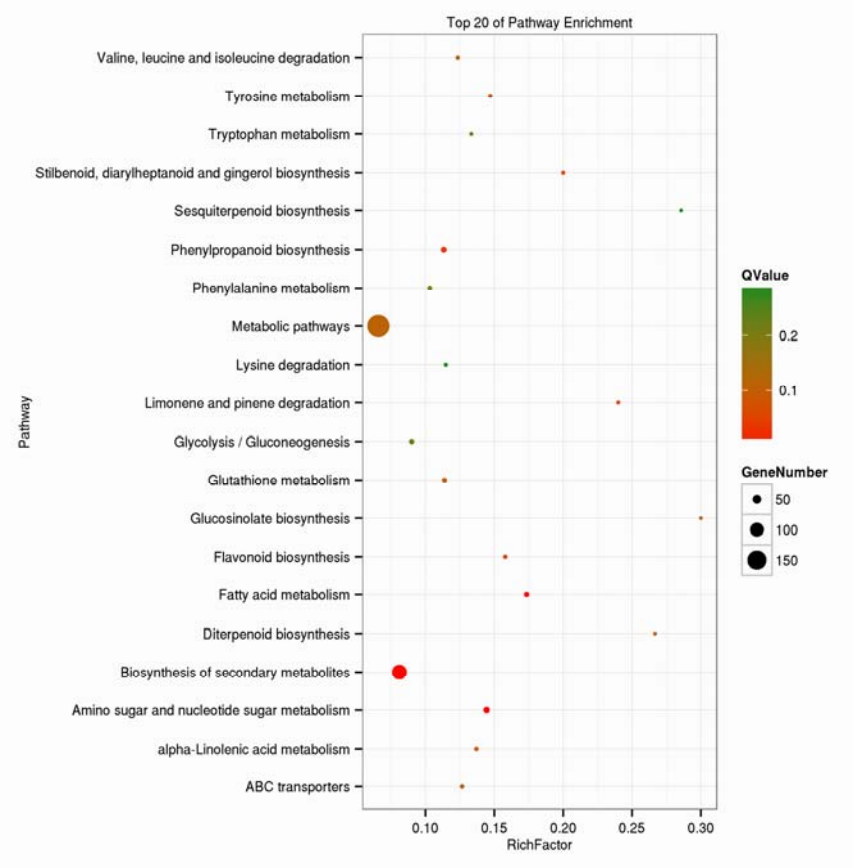

Fig. 11A. KEGG classification of DEGs between WT1-CK and WT1-T.

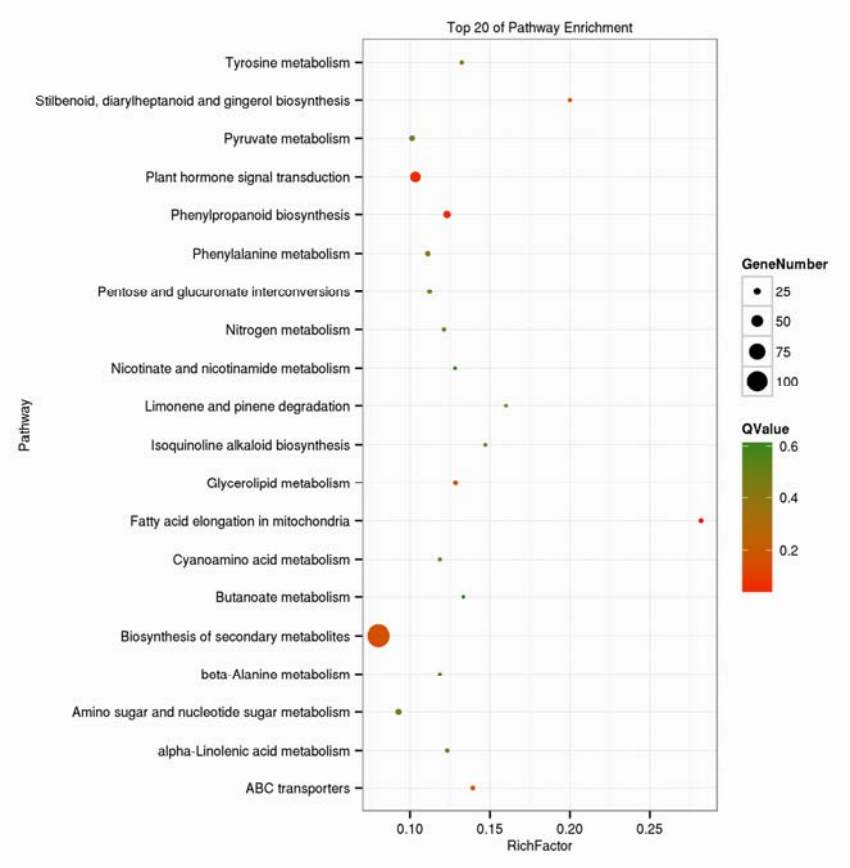

Fig. 11B. KEGG classification of DEGs between M1-CK and M1-T.

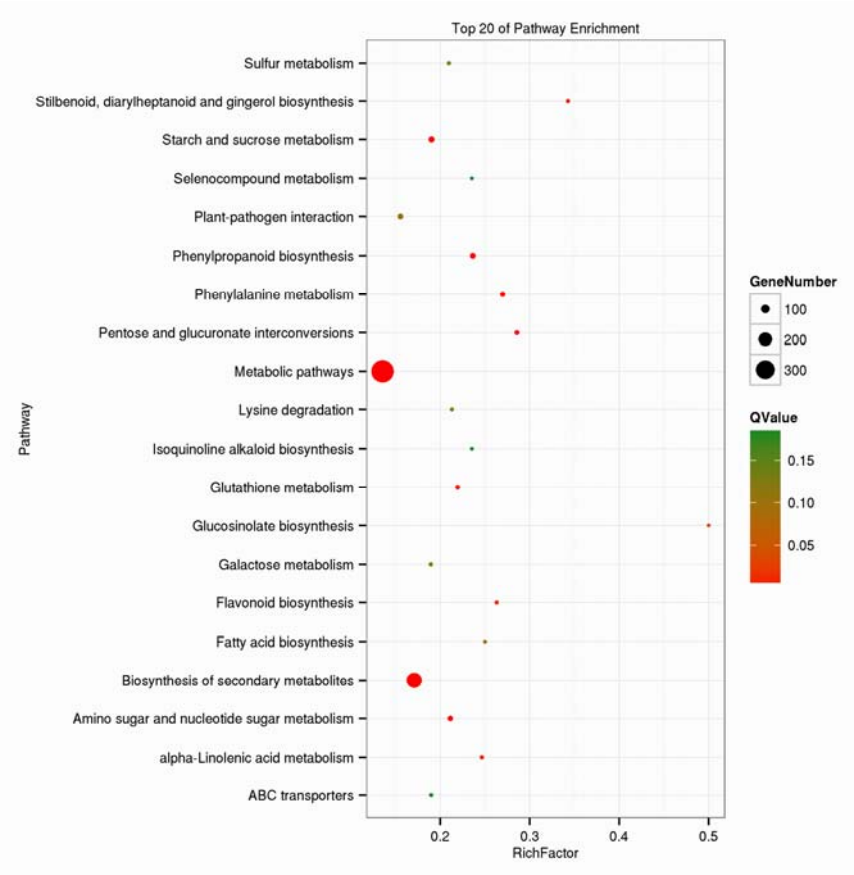

Fig. 11C. KEGG classification of DEGs between WT2-CK and WT2-T.

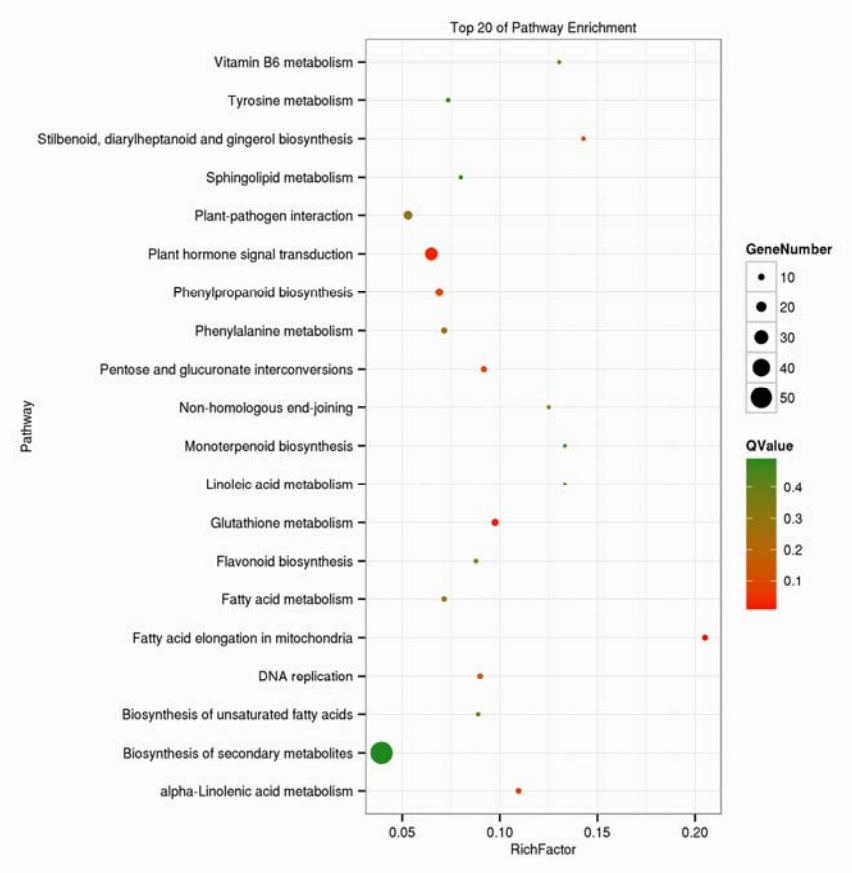

Fig. 11D. KEGG classification of DEGs between M2-CK and M2-T.

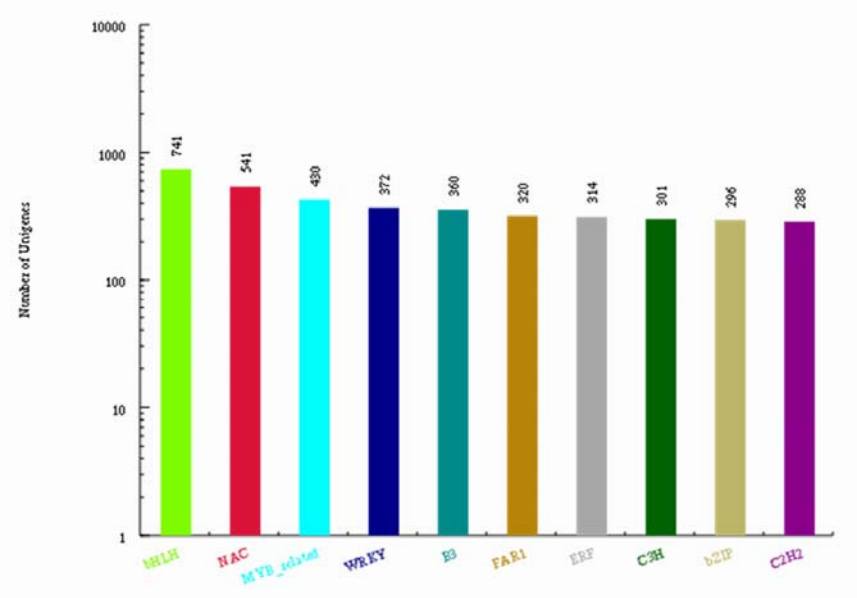

Fig. 12. The top 10 numbers of TF classification statistics.

Table 1. Characteristics of unigenes.

Item	Number
Genes Num	62637
GC percentage	41.20%
N50	1596
Max length	16598
Min length	201
Average length	906.62
Total assembled bases	56788193

Table 2. Major Characteristics of *de novo* assembled data in 8 samples.

Sample	Before Filter Reads Num	After Filter Reads Num (%)	Before Filter Data(bp)	After Filter Data(bp)	GC content (%)	Q20 (%)	Q30 (%)
WT1-CK	31341726	30797606(98.26%)	3917715750	3849700750	46.60%	97.95%	94.68%
WT1-T	26652830	26065952(97.8%)	3331603750	3258244000	46.83%	97.89%	94.57%
M1-CK	25401950	24813440(97.68%)	3175243750	3101680000	47.05%	97.96%	94.72%
M1-T	21751536	21182362(97.82%)	2718942000	2647795250	46.94%	97.96%	94.70%
WT2-CK	26733004	26150598(97.82%)	3341625500	3268824750	46.99%	97.77%	94.27%
WT2-T	27379842	26838652(98.02%)	3422480250	3354831500	47.61%	97.87%	94.48%
M2-CK	22927220	22524706(98.24%)	2865902500	2815588250	47.22%	97.88%	94.52%
M2-T	27330326	26854340(98.26%)	3416290750	3356792500	47.20%	97.84%	94.40%

Notes: Q20 (%) and Q30 (%) are the percentages of reads with Phred scores over than 20 and 30, respectively. GC content (%) means G+C bases as the percentage of total bases; WT1-CK, M1-CK, WT2-CK, M2-CK and WT1-T , M1-T, WT2-T, M2-T represent control and salt-treated samples respectively.

Table 3. Type statistics of SSRs.

Stat Item	Num
Total number of sequences examined	62637
Total size of examined sequences (bp)	56788193
Total number of identified SSRs	3355
Number of SSR containing sequences	3025
Number of sequences containing more than 1 SSR	289
Number of SSRs present in compound formation	121
Di-nucleotide	1234
Tri-nucleotide	1588
Tetra-nucleotide	294
Penta-nucleotide	95
Hexa-nucleotide	144

Table 4. RT-qPCR primer names and primer sequences.

Primer Name	Primer Sequences (5'-3')
Actin-F	GCTGGATTTGCTGGAGATGATGC
Actin-R	CTTCTCCATGTCATCCCAGTTGCT
23096-F	GTGGAAGAAGGAGAAAACAAATTCA
23096-R	GAATCTTTCCCGTCTCCGGC
17446-F	CGAAAGGAAACTCTGGAGGACA
17446-R	CTTGTGTTGCTGATGTTGGGA
11757-F	GGCCCAAAGTGTCATCGGA
11757-R	GAGGTCAGACATGATTGCTCCA
57231-F	AAAGGGATGAAGTGAGCTAGGAAA
57231-R	GATCTCTCCCTTGACCGGCT
29688-F	GCTTCTCTTGAGCTTTGCTGT
29688-R	TCTGTTGGGGAACACCGAGA
60321-F	GCCGTTGATCTTGTTCTGGGT
60321-R	CCAACTTTGGGTCCGGTTTG
9545-F	CAGAAGAGCCAGAAACTCGGAA
9545-R	CCCAATTAGGTCTGCTTCTGC
36375-F	CGTTGCTAGGACGCAATGTTCCA
36375-R	ACAGTCCACGTCGGATGCCTTAT
38654-F	CCTTACACCGTAGCTTTGCTC
38654-R	CCGGAAGAAACACAGCCAACA