bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

Transcriptome analysis of Tartary buckwheat revealed differentially expressed genes in salt pathway

Jin-Nan Song*, Xue-Hua Liu*, Ya-Qi Wang, 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.

* These authors contributed equally to this work.

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

ABSTRACT

Tartary buckwheat (*Fagopyrum tataricum*) is a kind of annually herbaceous crop with higher content of rutin and flavonoids than other crops. Through transcriptome analysis to reveal the differentially expressed genes in salt pathway, the salt tolerance mechanism of Tartary buckwheat was studied. Two varieties and two new strains of Tartary buckwheat were used as experimental materials, and 100 mM was used as concentration of NaCl for treatment. We *de novo* assembled the reads into a transcriptome dataset containing 62,637 unigenes with N50 length of 1596 bp. A total of 205.23 million clean reads were obtained, and 33,495 (53.47%) of them were annotated. GO and KEGG classification suggested that the enrichment of these unigenes was in 47 sub-categories, 25 KOG functional categories and 129 pathways, respectively. By comparisons among 11 groups, we got many differentially expressed genes (DEGs) of transcriptome data in 8 samples. The GO classification showed that many DEGs between samples encoding for biological process, cellular component and molecular function pathway. These processes may be related to salt tolerance mechanism in Tartary buckwheat, which will provide valuable reference genetic information for the study on salt tolerance mechanism in plants and cultivating strong salt-tolerant varieties of Tartary buckwheat for more yield in saline-alkali soil. In addition, the relative expression of *Unigene0009545* and *Unigene0057231* in M1 and that of *Unigene0023096* and *Unigene0060321* in M2 was increased significantly under salt stress. We speculate that these four genes of new strains in Tartary buckwheat may be related to salt tolerance.

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

SUMMARY Heading: Transcriptome analysis of Tartary buckwheat

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

1. Introduction

Buckwheat (*Fagopyrum esculentum* Moench) belongs to Polygonaceae, is a kind of dicotyledon plant. It originates from the Southwest in China at the early time, after that, it spreads to other countries gradually [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 used to be a crop because it contained some nutrients. Common buckwheat (*F. esculentum* Moench) and Tartary buckwheat (*F. tataricum* L. Gaertn.) are used for food [2], both are suitable for daily diet. 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 of 2050 [4]. When crops are exposed to salt stress, the function of crops will be damaged to a certain extent. Therefore, salt stress can restrict crop growth and consequently reduce agricultural yield [5]. In addition, high salinity will increase Na^+ content and destroy the transmission system. More importantly, once plants suffered sodium toxicity, tissues at different levels will be damaged or even destroyed [6]. Under this severe salt stress, a large number of salt-related genes are highly expressed.

By using RNA-seq technology, we can identify almost all RNAs with very small deviation, get more current data and directly measure gene expression level [7]. Compared with other sequence-based methods, such as Sanger sequencing or EST Library of cDNA, this popular technology has many advantages, including more accuracy, simplicity and sensitivity [8]. Through this useful technology, we studied the transcriptome of Tartary buckwheat and established the transcriptome database of salt-related genes in Tartary buckwheat. On the basis of understanding the transcriptome information of Tartary buckwheat, more salt-related components and more comprehensive transcriptome letters were obtained. The aim was to investigate the expression of important genes in Tartary buckwheat under salt stress.

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, and the salt tolerance of M1 and M2 is obviously higher than that of WT1 and WT2. The experimental groups were treated with Hoagland nutrient solution containing 100

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

mM NaCl at the stage of 2 leaves and 1 center leaf, while the control groups were treated with Hoagland nutrient solution without NaCl. Samples were then taken at 0 and 48 hours of processing time. The 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

After got the total RNA, the magnetic beads with Oligo (dT) were used to enrich the eukaryotic mRNA, then added fragmentation buffer to break the mRNA short (200~700 nt). Having gotten these mRNA, we used them to synthesize the first strand cDNA with random hexamers. 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 connected the sequencing connector, and agarose gel electrophoresis was used for fragment size selection. Finally, we put these fragments to do PCR amplification. The libraries were sequenced using the Illumina HiSeq 2000 instrument.

2.3. Data quality control and sequences assembly

To get high quality clean reads, we do the following steps: (1) remove the adaptors, (2) remove the reads which contain more than 10% nucleotides (N), (3) remove the low quality reads (Q-value≤5).

With the help of assembling software called Trinity [9], we assembled the short reads as transcriptome start, then this software linked the reads with some overlaps to longer segments, these segments with no nucleotides (N), afterwards we assembled these segments into a series of sequences, which were defined as unigenes, they were used to react functional annotation and coding sequence (CDS) prediction.

2.4. Functional annotation and classification

Using the BLASTx (evalue < 0.00001), all assembled unigenes were searched against the databases of NCBI Nr, Swiss-Prot, COG and KEGG. KEGG is a database on genome decoding, it can analyze the pathway of gene products in cell and function of these gene products, we can obtain the pathway annotation of unigene according to the KEGG annotation information [10]. GO (Gene Ontology) is a collaborative tool that can solve the collective information [11]. We took the *Fagopyrum tataricum* genes under the GO (http://www.geneontology.org/) classifications, they were divided into 3 classes of type: biological process, cellular component and molecular function. 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 [12]. COG/KOG

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

(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 [13].

2.5. Identification of transcription factor families and SSRs

The identifications of transcription factors (TF) were analyzed by PlantTFcat online tool (http://plantgrn.noble.org/PlantTFcat) [14]. SSRs (Simple Sequence Repeats), one of molecular markers, it is a potent tool for genetic study and marker assisted selection (MAS) in many crops [15,16]. In addition, in this research, we adopted MISA search tool for SSRs screening [17].

2.6. Analysis of differentially expressed genes (DEGs)

To identify differentially expressed genes across samples or groups, the edgeR package (http://www.r-project.org/) was used. 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 verify the correctness of RNA-seq, we utilized the RT-qPCR to get the relative expression level on related genes which we took care of. Based on the quantity of RNA, we got the cDNA by reverse transcription reactions with a kit called PrimeScript RT reagent Kit with gDNA Eraser (Takara Biotech, Dalian, China), the *actin* gene was used as internal controls, which was the *actin* of Tartary buckwheat. The primers were designed by NCBI and were synthesized by Tsingke company. The real time PCR program consisted at 94℃ for 30 s, 40 cycles of denaturing at 95℃ for 10 s, annealing and extension at 58℃ for 20 s. Finally, data processing and statistical analysis were used the $2^{(-\Delta\Delta Ct)}$ method [18] 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 was 62,637, and the total GC percentage was 41.20%. The length of N50 was 1,596 bp, revealing that the assembly quality was not half bad, so we could take use of the assembly results to carry out the next experiments. In addition, we got the length distribution graph of 8 samples summary, the length of most genes was between 200 bp to 299 bp, and the number of genes was about 16,857 (Fig. 1; 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 value of all clean reads was over 97%, and the Q30 value was over 94%. All clean

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

reads had been deposited into the Sequence Read Archive (SRA) under the accession of PRJNA418602, NCBI (https://www.ncbi.nlm.nih.gov/geo/).

3.2. Unigene Functional Annotation

3.2.1 . Four big database annotation statistics

We got the whole of 62,637 unigenes, which were searched against 4 main databases, including Nr, Swissprot, KOG and KEGG, the final annotated results were showed in Table 2. Among the whole unigenes, 33,495 unigenes (53.47%) were annotated, while 29,142 unigenes(46.53%) could not be annotated in any databases. Among the annotated unigenes, 33,158 unigenes (52.94%) were in Nr database, which contained non-redundant protein sequence with entries based on BLAST tool [19], 25,430 unigenes (40.60%) were in Swissprot database, and 21,489 unigenes (34.31%) and 13,620 unigenes (21.74%) were in KOG [18] and KEGG [20] database, respectively. As shown in Fig. 3, 10,865 unigenes could be annotated in four databases, so it gave us a notice that these unigenes might be superbly vital to life activities 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 [21]. 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 has a lot of 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) [22]. 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 a large of different classifications [23]. We may find some

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

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 most number of genes is in "metabolic process", which is assigned with 12,473 (21.00%) unigenes, while that in cellular component is "cell", which is assigned with 9,128 (24.08%) unigenes, and that in molecular function is "catalytic activity", which is assigned with 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 being 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%). The results were better comparing to the SSRs and GC distributions in *Arabidopsis* [24]. We believed that our analysis could provide a solid foundation for the future study (Fig. 7).

3.2.5. Identification of differentially expressed genes

Aimed at 8 samples, we obtained the expression profile data by RNA-seq, some salt-related genes predicted to participate in this salt stress process. We got many DEGs (differentially expressed genes). As shown in Fig. 8, green pillar meaned that relative expressions were down regulated, red pillar meaned that relative expressions were up regulated. For getting an intuitional view between control and NaCl solution treatment, we tidied up the raw data and made Venn diagrams (Fig. 9). There were 42 (0.30%) common unigenes in four groups, which mostly showed that these 42 unigenes were all salt-related. The detailed

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

unigenes were showed in Table S3.

3.2.6. The differences between samples expresses in GO/Pathway

We completed GO term enrichment by Blast2GO for further understanding the biological function of DEGs. Most of up-going DEGs under salt treatment were abundant in 'cell morphogenesis', 'cellular component morphogenesis', 'oxidation-reduction process', 'catalytic activity', 'cellular process', 'metabolic process', 'single-organism process', 'binding' and 'cell part'. And the down-going 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 overrepresented 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) usually played vital roles in growth and development processes of plants. Collecting data indicated that many TF families were involved in abiotic stress responses [25]. Though PlantTFcat online tool, we identified 57 transcription factor families, including 7,208 unigenes. The top of 10 most 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 Real-Time PCR

To validate the facticity of expression profile data obtained by RNA-Seq, we carried out real-time PCR analysis on nine unigenes which were related to salt stress: SOD [26], POD [27], APX [28], SR(salt-related), Na^+/H^+ [29] and SOS (salt overly sensitive) [30]. 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^+ 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/). Also, we got the *Actin* [31] sequence of *F. tataricum* and the primer for RT-qPCR. They were chosen to act qualified by real-time PCR and were selected by

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

RNA-Seq analysis. 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 significantly 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 significantly under salt stress compared with the control, which was increased by 211.48% and 73.08%, respectively. Therefore, we speculate that these four genes of new strains in Tartary buckwheat may be related to salt tolerance.

4. Discussion

There are more than 1 billion hm^2 of saline soil in the world [32]. However, saline soil is an important land resource. In addition, salinization can also reduce crop yields. Therefore, it is urgent to improve crop salt tolerance or reduce saline soil. On the former issue, the urgent research in recent years has focused on the mechanism of salt tolerance in plants [33]. On the latter issue, soil salinization assessment is the premise, which can reveal the basic situation of soil salinization in this area [34]. Prevention and control has become the mainstream method to reduce saline soil. In recent years, transcriptome analysis has become more and more popular, such as *maize (Zea mays* L.) [35] and *Spodoptera litura* [36], which provides a comprehensive survey of gene expression changes and some insights into specific biological processes [37]. 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 planted them in saline soil and obtained two new strains: Chuanqiao NO.1-1 and Chuanqiao NO.2-1. By *de novo* technology study on two varieties and two new strains of Tartary buckwheat, we got 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 [38]. According to other reported transcriptome data, Wu et al. [39] 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*. We totally obtained 62,637 assembly unigenes, 33,495 (53.47%) were annotated in four big databases, our high-throughput RNA-seq yielded more unigenes with slightly shorter N50, and 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. Many genes are related to two stress reaction processes. In addition, the results of KOG

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

241 classification showed that metabolic pathways were overexpressed in terms of 'General function prediction
242 only', 'Posttranslational modification, protein turnover, chaperones' and 'Signal transduction mechanisms'.
243 This provides more important information for the future study on molecular mechanism of salt tolerance in
244 plants.

245 The comparison of gene expression patterns between control and salt treatment group: WT1-CK VS
246 WT1-T, M1-CK VS M1-T, WT2-CK VS WT2-T, M2-CK VS M2-T, there were 4049 DEGs, 4167 DEGs,
247 6752 DEGs and 2239 DEGs, respectively, all containing low and high expression genes. Then, we got 42
248 common DGEs in the above groups. The number of DEGs in each sample was larger than that in
249 *Arabidopsis* [40]. It indicated that these 42 common DEGs may be related to salt treatment.

250 As shown by the results of transcription factor (TF), bHLH [41], NAC [42], MYB [43] and WRKY
251 [44] have the first four gene. That is to say, when exposed to salt stress, the transcription factors dealing
252 with this stress belong to the aforementioned TF families. The results showed that many genes related to
253 salt stress might be found in *F. tataricum*.

254 bHLH is an important transcription factor involved in some processes of environmental stress
255 response [45]. *OrbHLH2* was cloned from *Oryzara fipogon* Griff by Zhou et al. [46], which is highly
256 homologous to ICE1 protein. We obtained 741 DEGs belonging to the bHLH transcription factor family,
257 such as *Unigene0005747* and *Unigene0006105*. NAC protein was unique and numerous to plants, the
258 researches showed that the transcription factor had a number of functions in some ways. A large amount of
259 evidence indicated that NAC transcription factors activated or inhibited target gene expression in biological
260 damage, such as pathogen infection and other biological damage, such as high salt, drought, low
261 temperature, ABA and mechanical damage, etc. Also, this transcription factor was related to salt stress.
262 Ganesan et al. [47] found that *AmNAC1* played an important role in early salt stress response and long-term
263 salt regulation in mangrove plant *Avicennia marina*. A total of 541 DEGs were obtained, such as
264 *Unigene0031578*, *Unigene0033494*, etc, belonging to the NAC transcription factor family. As an important
265 transcription factor, they may play important roles in salt tolerance of *F. tataricum*. MYB transcription
266 factors are related to plant development, secondary metabolism, hormone signal transduction, disease
267 resistance and abiotic stress tolerance [48]. According to the study of Zhang et al. [49] in *Arabidopsis*, high
268 salt stress activates MYB-related gene *AtMYBL* that is highly expressed. *Unigene0027989* and
269 *Unigene0027988* are the most important expression genes. They belong to the MYB transcription factor
270 family. We can obtain many genes that response to MYB transcription factor of *F. tataricum* under salt

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

stress. As an important transcription factor superfamily, WRKY transporters are involved in response to environmental stimuli, such as high salt, low temperature, etc [50]. This transcription factor in *F. tataricum* is highly expressed under salt stress, which has the complexity and importance of stress regulation. *Unigene0043192*, *Unigene0043195* and *Unigene0043300* in *F. tataricum* have high similarity in *Oryza barthii*, and they are highly expressed under salinity stress. In conclusion, bHLH, NAC, MYB and WRKY are the most common transcription factors under salt stress in *F. tataricum*. Therefore, we can conclude that among these four transcription factors, more salt-related genes can be found, which work together to respond to salt stress in *F. tataricum*.

Salt stress increased the activities of antioxidant enzyme (SOD, POD, APX) and improved the adaptability of plants to adversity, such as cotton [51] and maize [52]. The relative expression levels of *NHX1* and *SOS1* genes were increased, such as the overexpression of *AeNHX1* gene in *Arabidopsis* plants [53], which could improve the salt tolerance of *Arabidopsis thaliana*; SOS1 could also improve the salt tolerance in *S. pombe* [54]. The corresponding genes were higher expressed in *F. tataricum*. In addition, the expression levels of *FtNHX1* and *FtSOS1* were higher under salt stress. These salt-related genes were found and clarified in *F. tataricum*, which helps us understand the salt tolerance mechanism in plants. 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

Though *de novo*, we generated and annotated a salt-responsive transcriptome of *F. tataricum*. Though RNA-seq, we obtained the DEGs of 8 samples. Genome-wide assay of the transcriptional differences between control and salt treatment group led to the identification of many key regulatory factors in salt reaction mechanism. We also found some genes related to antioxidant enzymes and important gene sequences under salt stress. In addition, we have selected two new salt-tolerant strains, which provide a more reasonable explanation for the study on salt tolerance mechanism. The results will help us to understand the salt tolerance mechanism in plants. The transcriptome data obtained can provide molecular and genomic information for future research.

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

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).

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 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,

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

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)

10 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)

11 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)

12 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)

13 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)

14 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)

15 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)

16 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)

17 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)

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

19 Pruitt KD, Tatusova T, Maglott DR. 2007 NCBI reference sequences (RefSeq): a curated non-redundant sequence

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D61–65. (doi:10.1093/nar/gki025)

20 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)

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

22 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)

23 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)

24 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)

25 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)

26 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)

27 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)

28 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)

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

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

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

32 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)

33 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)

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

391 34 Letey J, Hoffman GJ, Hopmans JW, Grattan SR, Suarez D, Corwin DL, Oster JD, Wu L, Amrhein C. 2011 Evaluation
392 of soil salinity leaching requirement guidelines. *Agric. Water Manage.* **98**, 502–506. (doi:10.1016/j.agwat.2010.08.009)
393 35 Jaiswal P. 2013 Maize metabolic network construction and transcriptome analysis. *Plant Genome* **6**, 1–12.
394 (doi:10.3835/plantgenome2012.09.0025)
395 36 Gu J, Huang LX, Gong YJ, Zheng SC, Liu L, Huang LH, Feng QL. 2013 De novo characterization of transcriptome and
396 gene expression dynamics inepidermis during the larval-pupal metamorphosis of common cutworm. *Insect Biochem.*
397 *Mol. Biol.* **43**, 794–808. (doi:10.1016/j.ibmb.2013.06.001)
398 37 Nagalakshmi U, Waern K, Snyder M. 2010 RNA-Seq: a method for comprehensive transcriptome analysis. *Curr. Protoc.*
399 *Mol. Biol.* **4**, 11–13. (doi:10.1002/0471142727.mb0411s89)
400 38 Makinen V, Salmela L, Ylinen J. 2012 Normalized N50 assembly metric using gap-restricted co-linear chaining. *BMC*
401 *Bioinformatics* **13**(1), 255. (doi:10.1186/1471-2105-13-255)
402 39 Wu Q, Bai X, Zhao W, Xiang D, Wan Y, Yan J, Zou L, Zhao G. 2017 De novo assembly and analysis of Tartary
403 buckwheat (*Fagopyrum tataricum* Gaertn.) transcriptome discloses key regulators involved in salt-stress response.
404 *Genes* **8**(10), 255. (doi:10.3390/genes8100255)
405 40 Taji T, Seki M, Satou M, Sakurai T, Kobayashi M, Ishiyama K, Narusaka Y, Narusaka M, Zhu JK, Shinozaki K. 2004
406 Comparative genomics in salt tolerance between *Arabidopsis* and *Arabidopsis*-related halophyte salt cress using
407 *Arabidopsis* microarray. *Plant Physiol.* **135**(3), 1697–1709. (doi:10.1104/pp.104.039909)
408 41 Satou Y, Imai KS, Levine M, Kohara Y, Rokhsar D, Satoh N. 2003 A genomewide survey of developmentally relevant
409 genes in *Ciona intestinalis* - I. Genes for bHLH transcription factors. *Dev. Genes Evol.* **213**, 213–221.
410 (doi:10.1007/S00427-003-0319-7)
411 42 Olsen AN, Ernst HA, Leggio LL, Skriver K. 2005 NAC transcription factors: structurally distinct, functionally diverse.
412 *Trends Plant Sci.* **10**, 79–87. (doi:10.1016/j.tplants.2004.12.010)
413 43 Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. 2010 MYB transcription factors in *Arabidopsis*.
414 *Trends Plant Sci.* **15**, 573.(doi:10.1016/j.tplants.2010.06.005)
415 44 Eulgem T, Somssich IE. 2007 Networks of WRKY transcription factors in defense signaling. *Curr. Opin. Plant Biol.* **10**,
416 366–371. (doi:10.1016/j.pbi.2007.04.020)
417 45 Song XM, Huang ZN, Duan WK, Ren J, Liu TK, Li Y, Hou XL. 2014 Genome-wide analysis of the bHLH transcription
418 factor family in Chinese cabbage (*Brassica rapa* ssp. pekinensis). *BMC Genomics* **289**, 77.
419 (doi:10.1007/s00438-013-0791-3)
420 46 Zhou J, Li F, Wang JL, Ma Y, Chong K, Xu YY. 2009 Basic helix-loop-helix transcription factor from wild rice

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

(OrbHLH2) improves tolerance to salt- and osmotic stress in *Arabidopsis*. *J. Plant Physiol.* **166**, 1296–1306. (doi:10.1016/j.jplph.2009.02.007)

47 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)

48 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)

49 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)

50 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)

51 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)

52 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)

53 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)

54 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)

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

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 distribution in *Fagopyrum tataricum*.

Fig. 3. Four database annotation of Venn diagrams.

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. DEGs among four groups of Venn diagrams.

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.

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

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

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

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

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

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

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

487 Table S6. TF. Class.

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

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

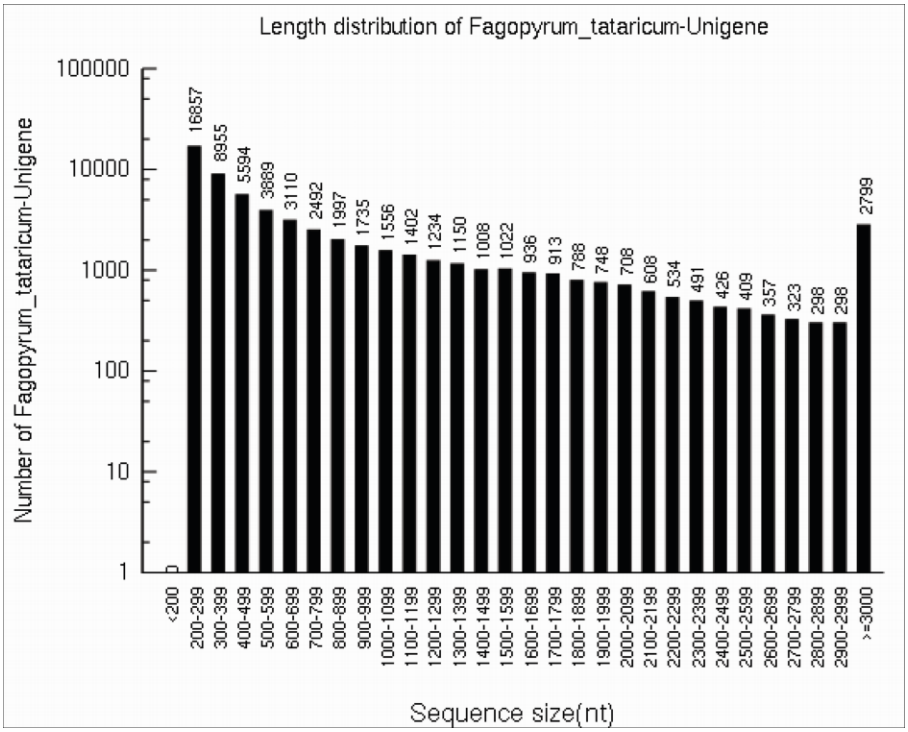


Fig. 1. Length distribution of genes in *Fagopyrum tataricum*.

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

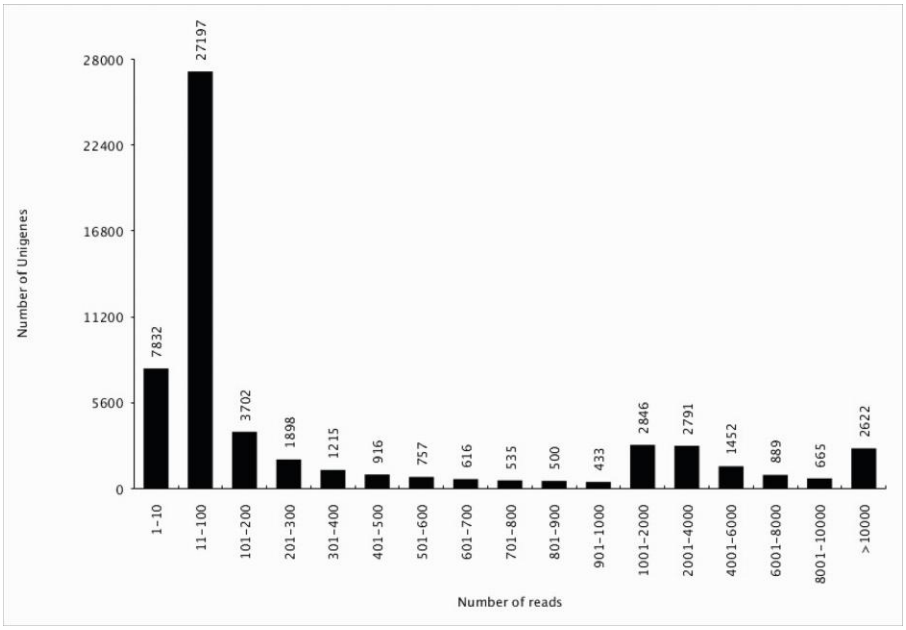


Fig. 2. Number of reads distribution in *Fagopyrum tataricum*.

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

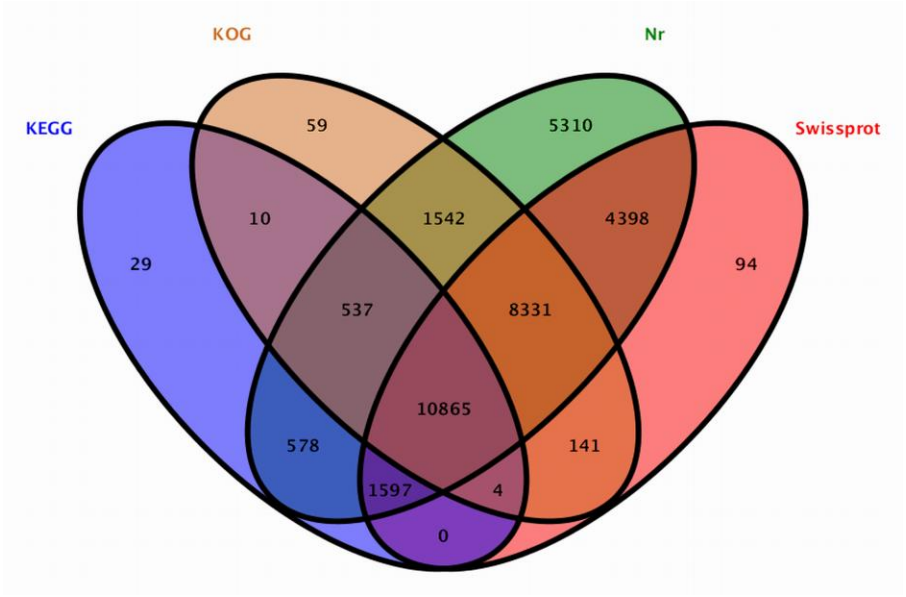


Fig. 3. Four database annotation of Venn diagrams.

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

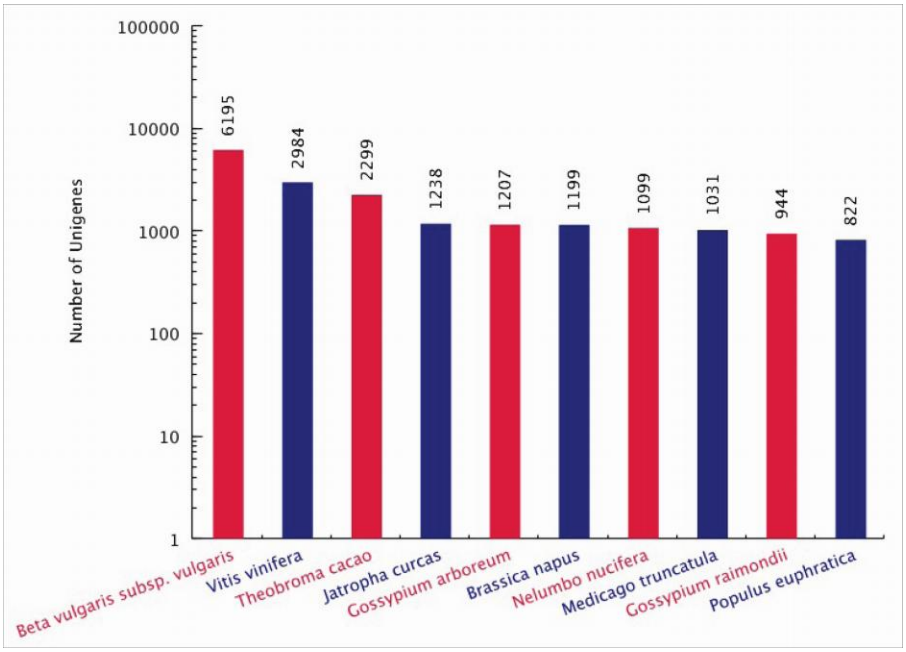
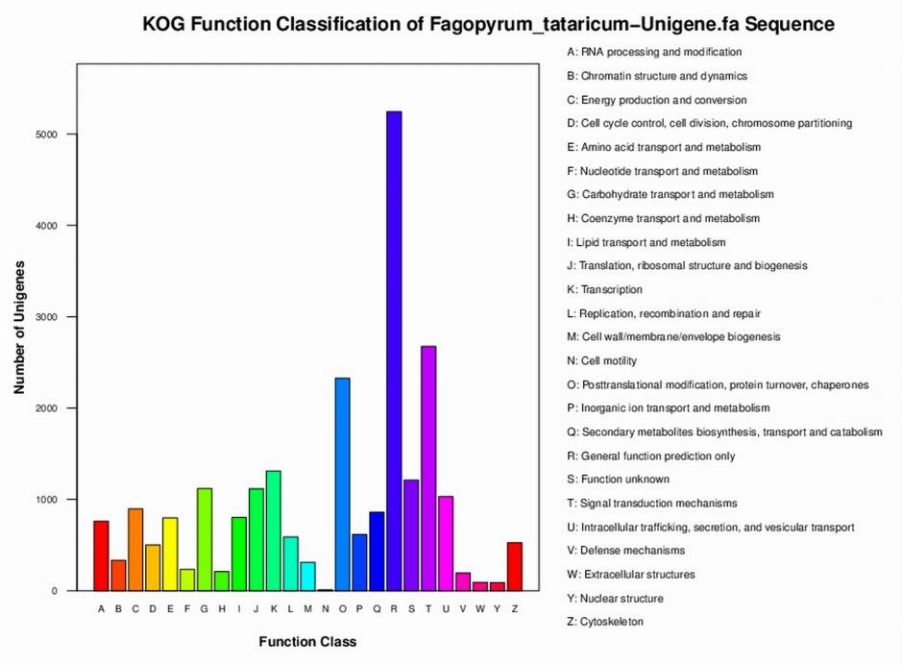


Fig. 4. Distribution statistics of near source varieties (Top 10).

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

588



589

590

591 Fig. 5. KOG function classification of *Fagopyrum tataricum* unigenes (Note: 21,489 unigenes are

592 annotated and classified into 25 KOG categories from the final set of 62,637 unigenes).

593

594

595

596

597

598

599

600

601

602

603

604

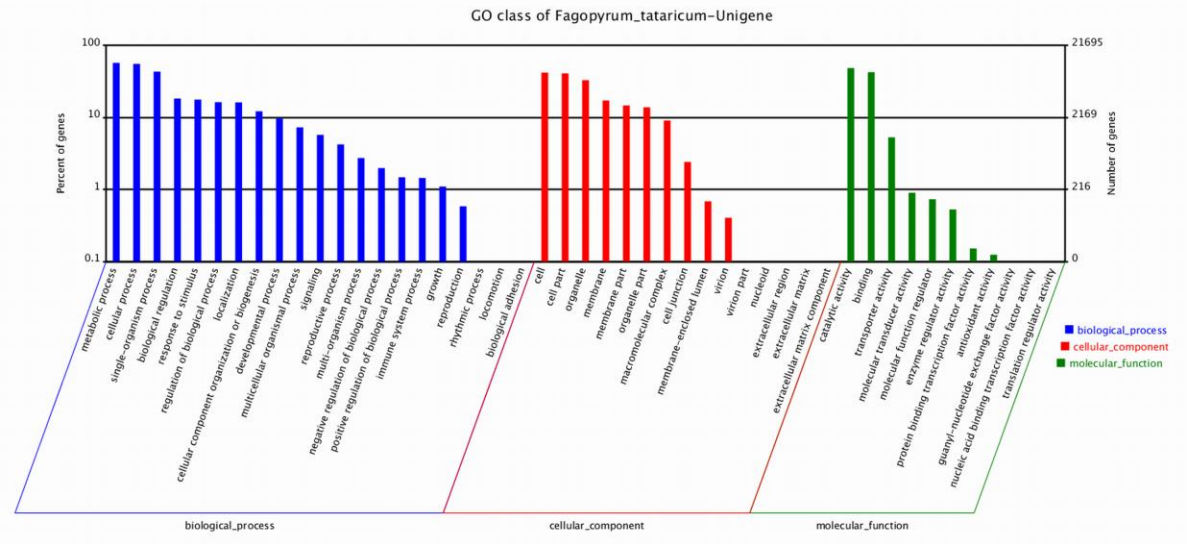
605

606

607

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

608



609

610

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

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

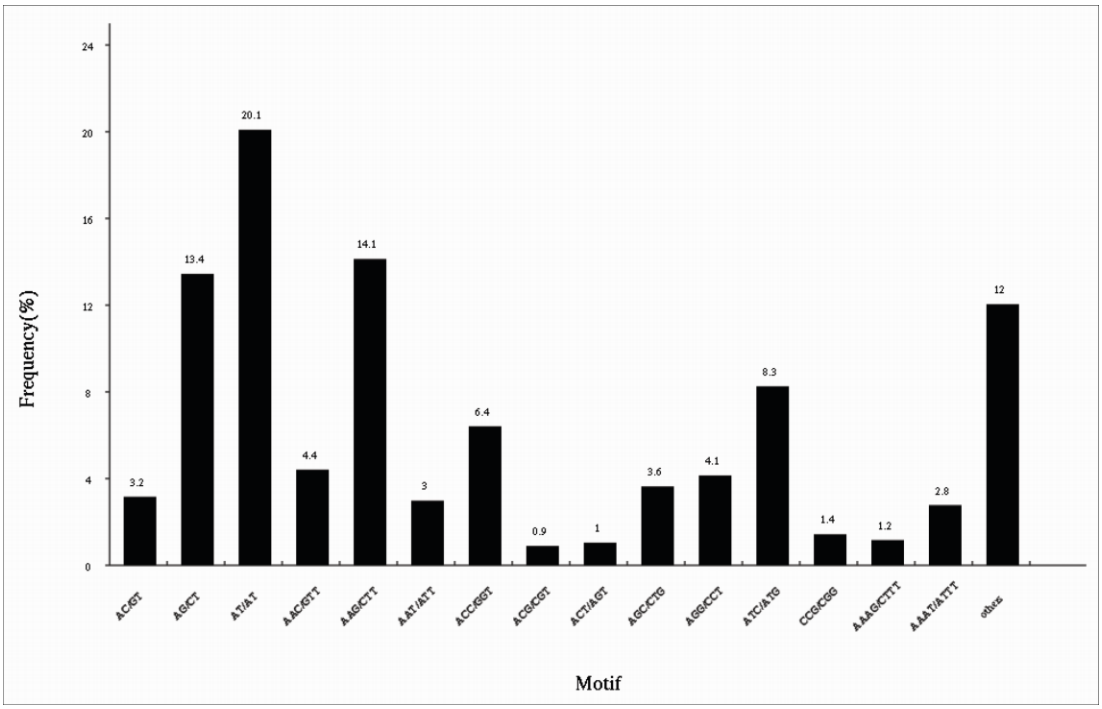


Fig. 7. Distribution of SSR motifs in *Fagopyrum tataricum.*

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

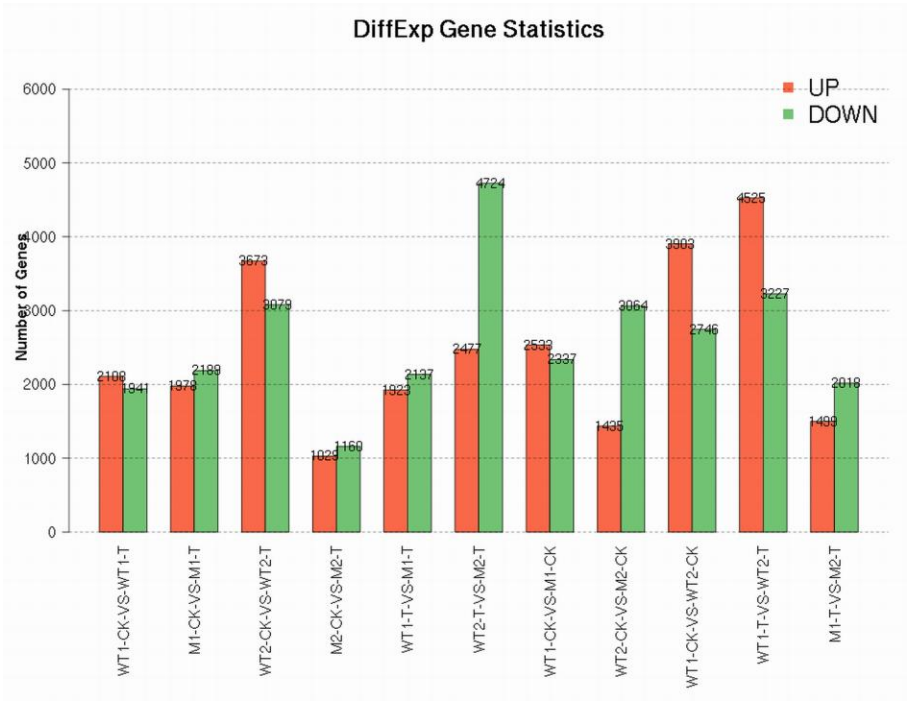


Fig. 8. Different expression of genes among samples.

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

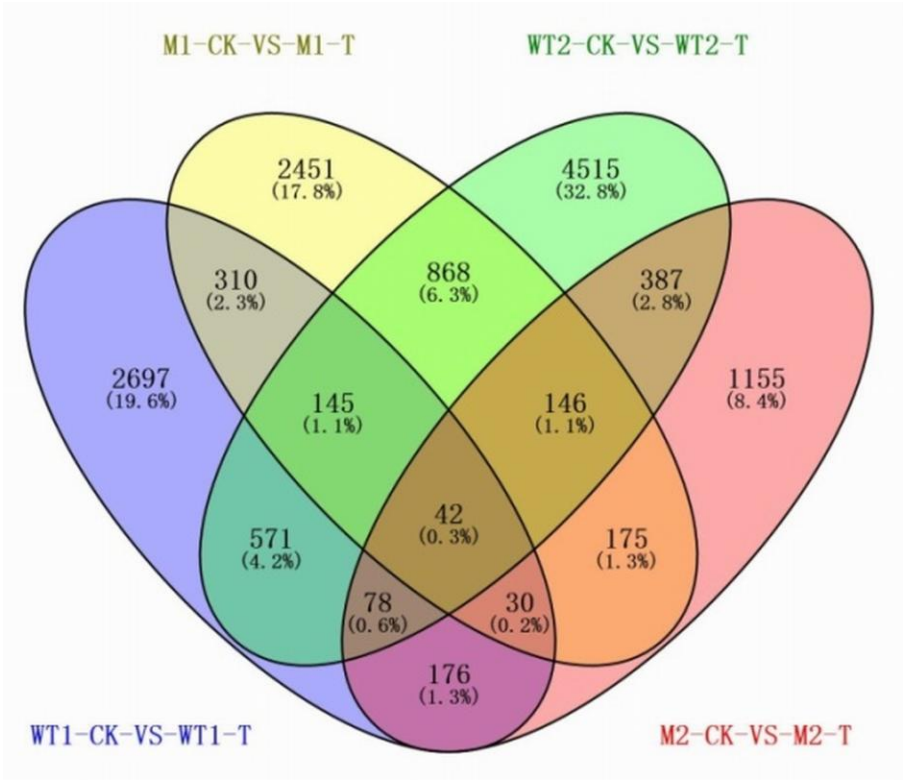
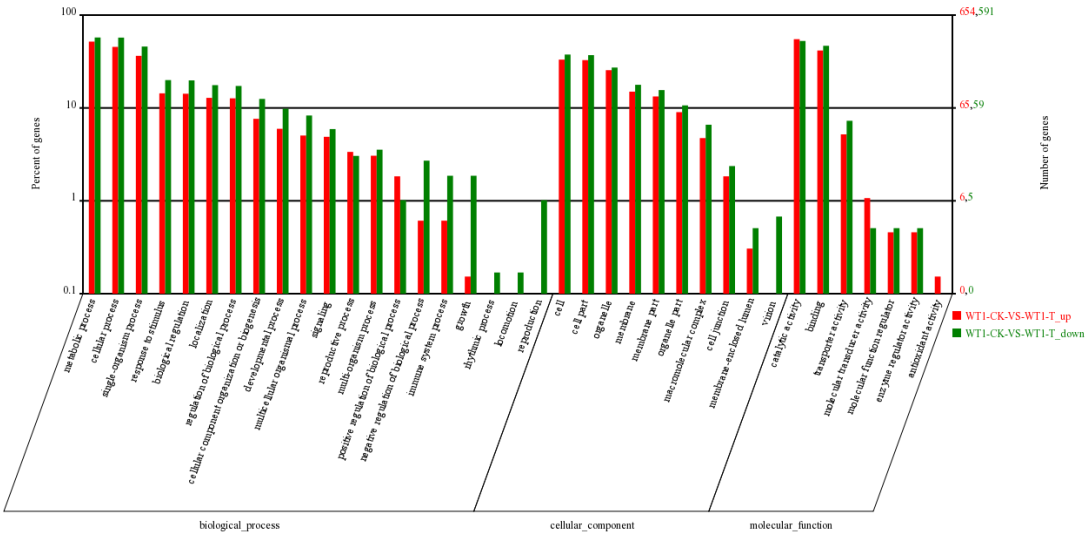


Fig. 9. DEGs among four groups of Venn diagrams.

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

689



690

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

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

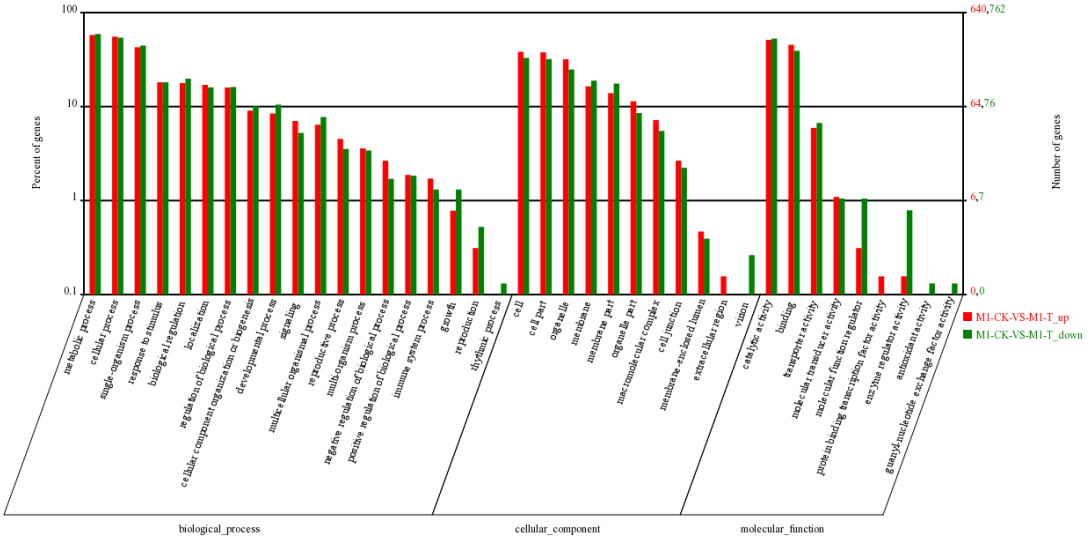


Fig. 10B. GO classification of DEGs between M1-CK and M1-T.

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

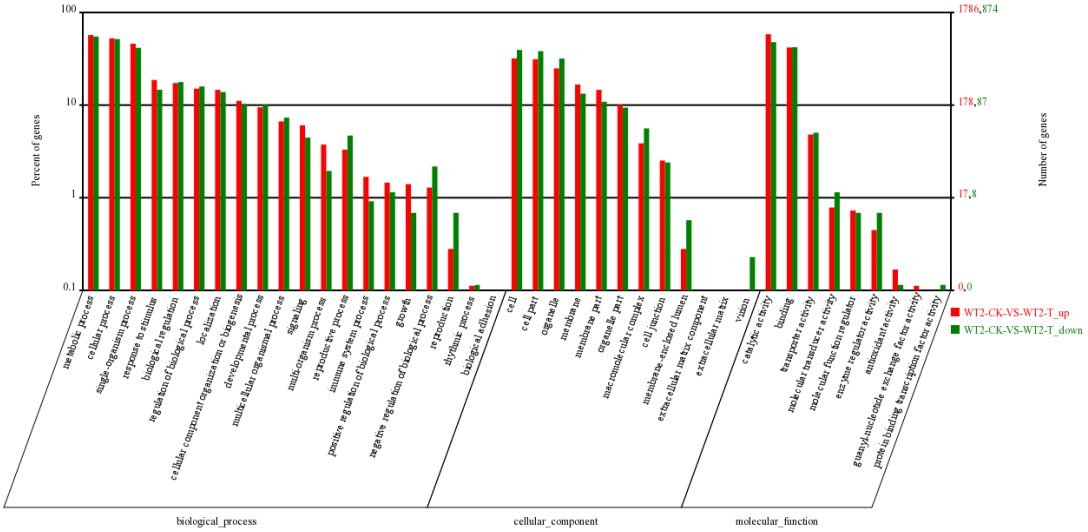


Fig. 10C. GO classification of DEGs between WT2-CK and WT2-T.

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

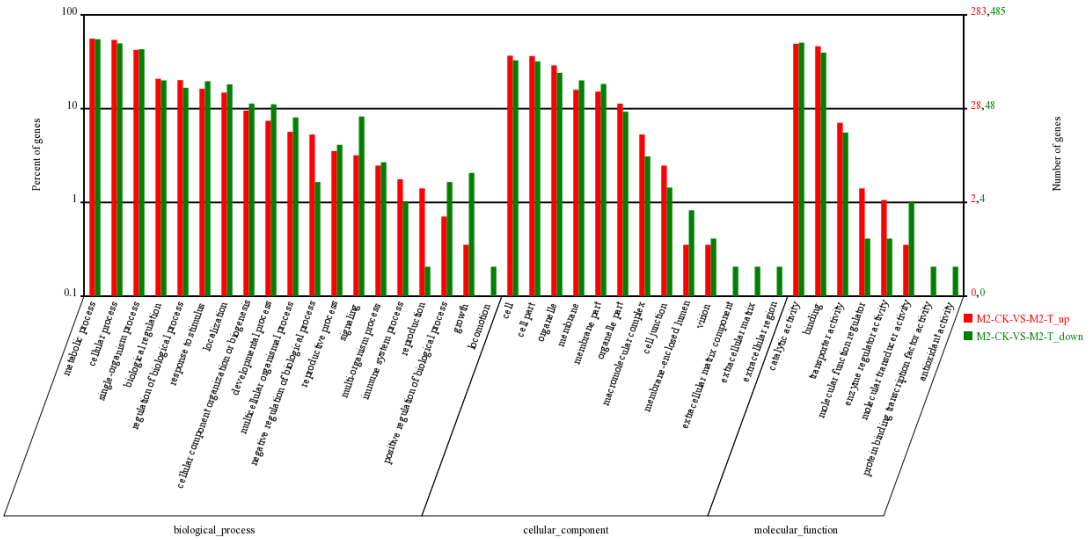


Fig. 10D. GO classification of DEGs between M2-CK and M2-T.

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

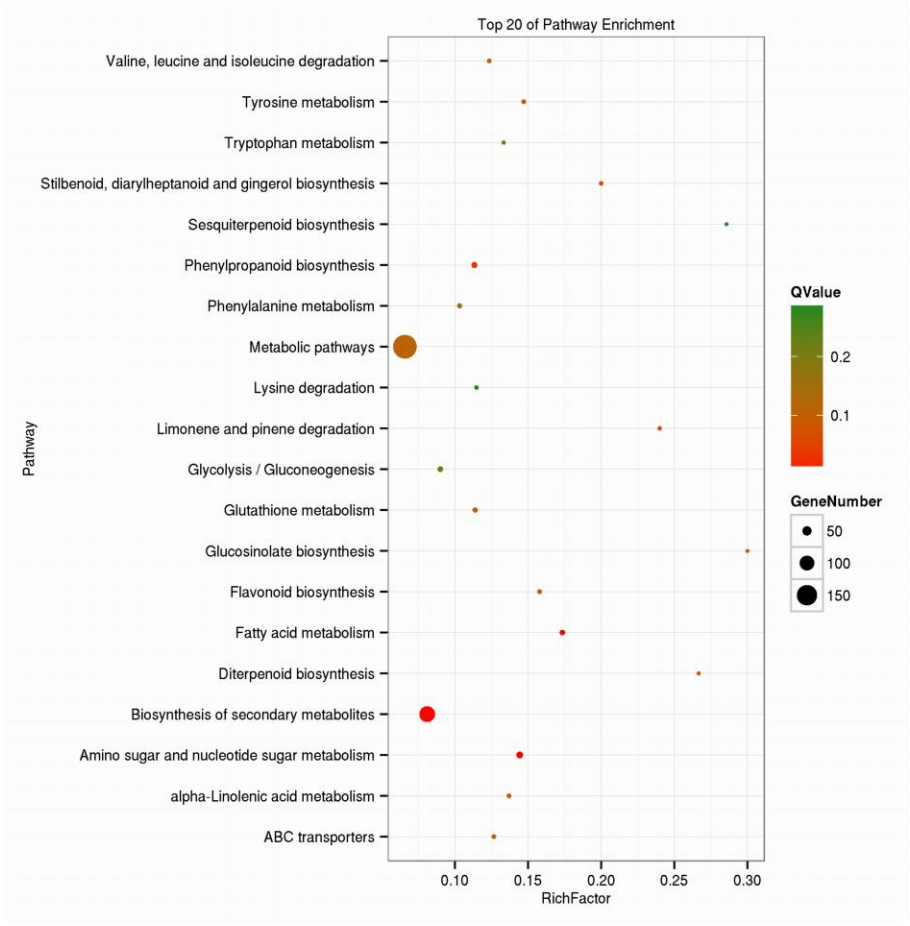


Fig. 11A. KEGG classification of DEGs between WT1-CK and WT1-T.

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

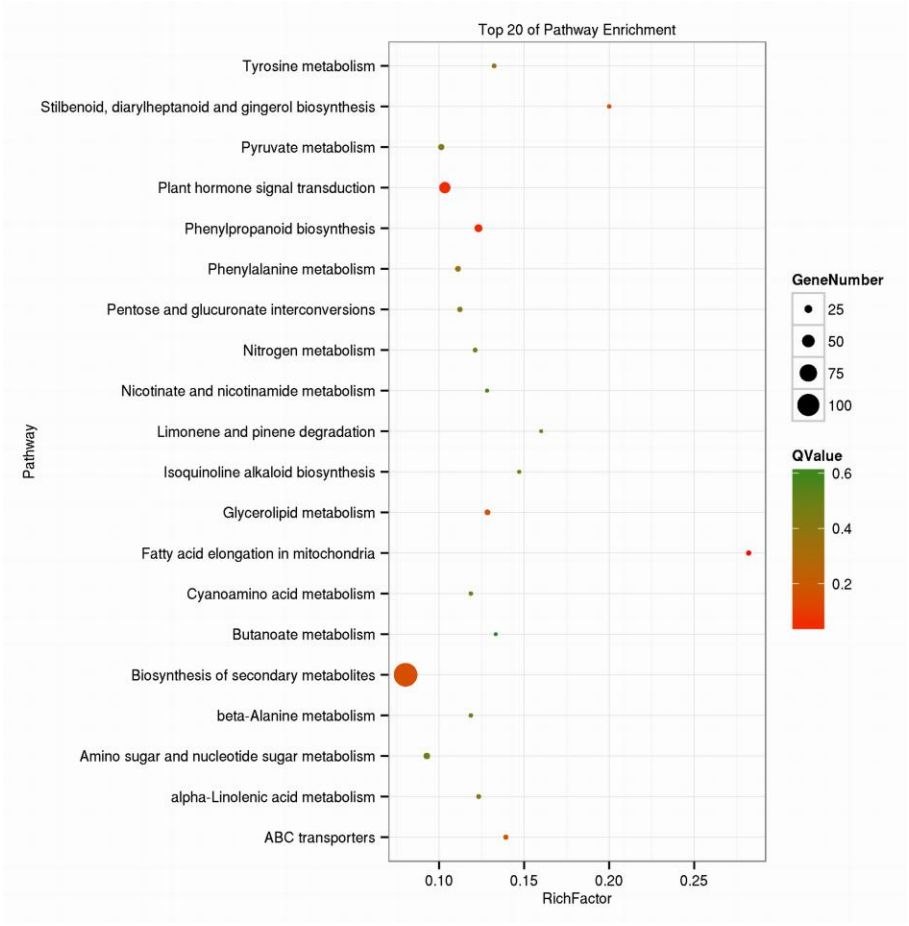


Fig. 11B. KEGG classification of DEGs between M1-CK and M1-T.

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

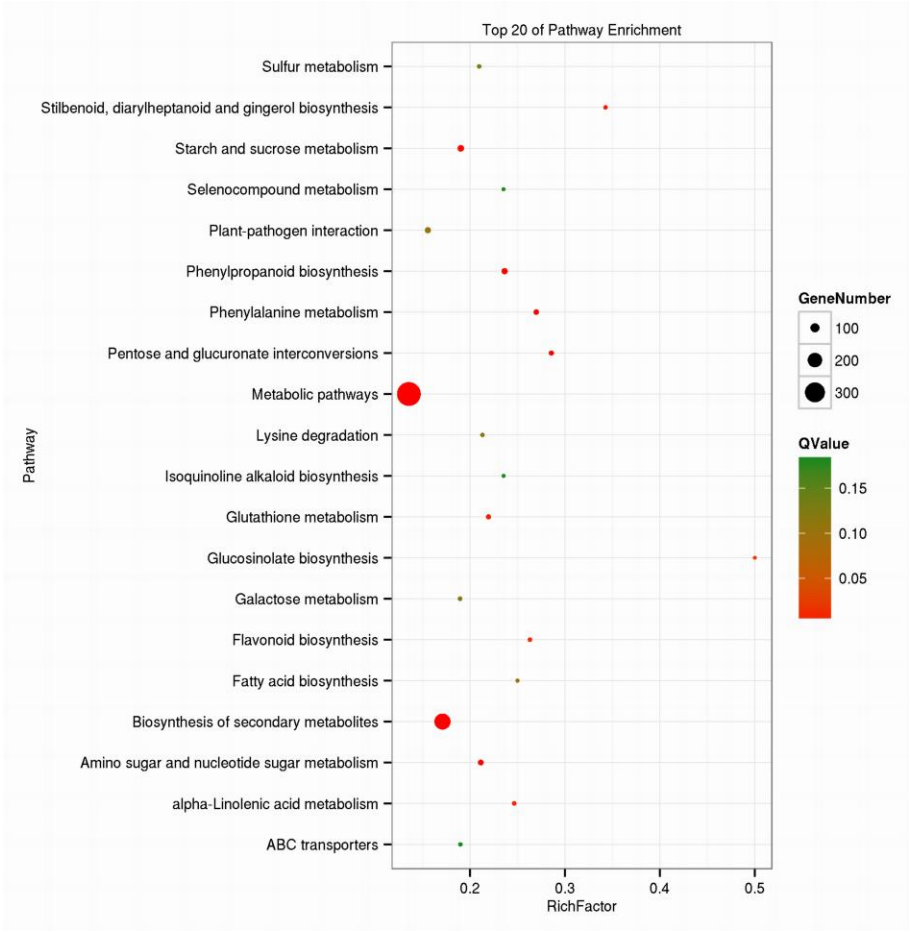


Fig. 11C. KEGG classification of DEGs between WT2-CK and WT2-T.

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

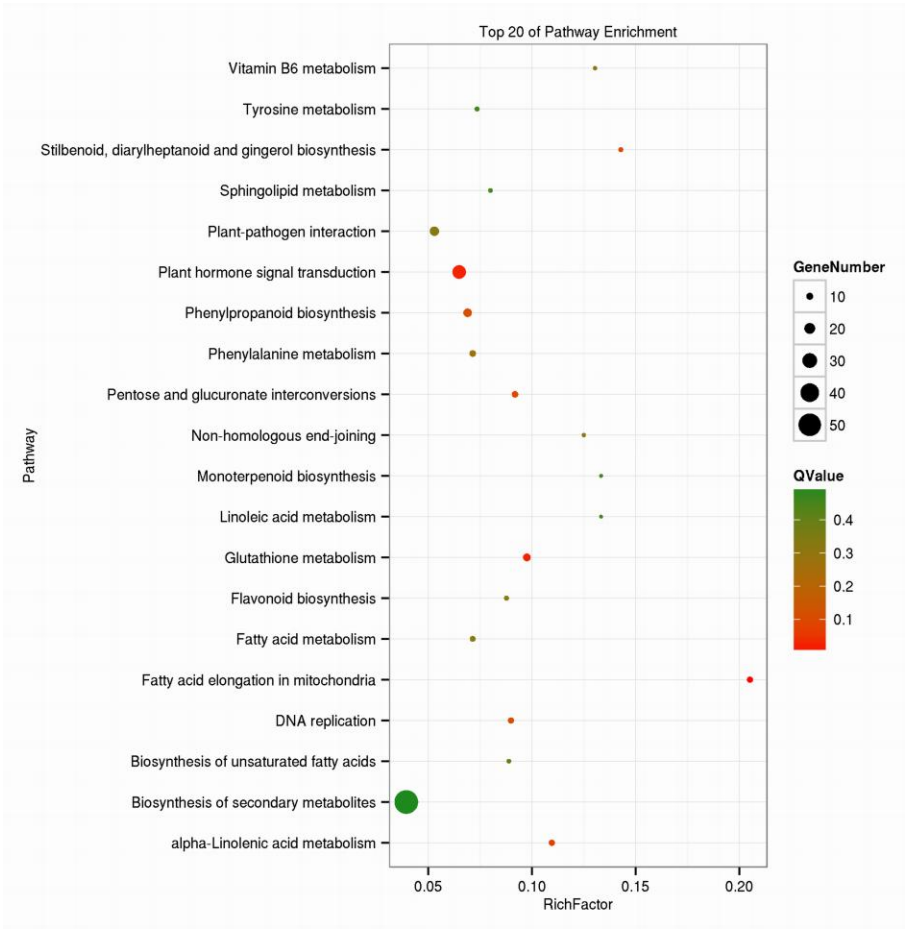
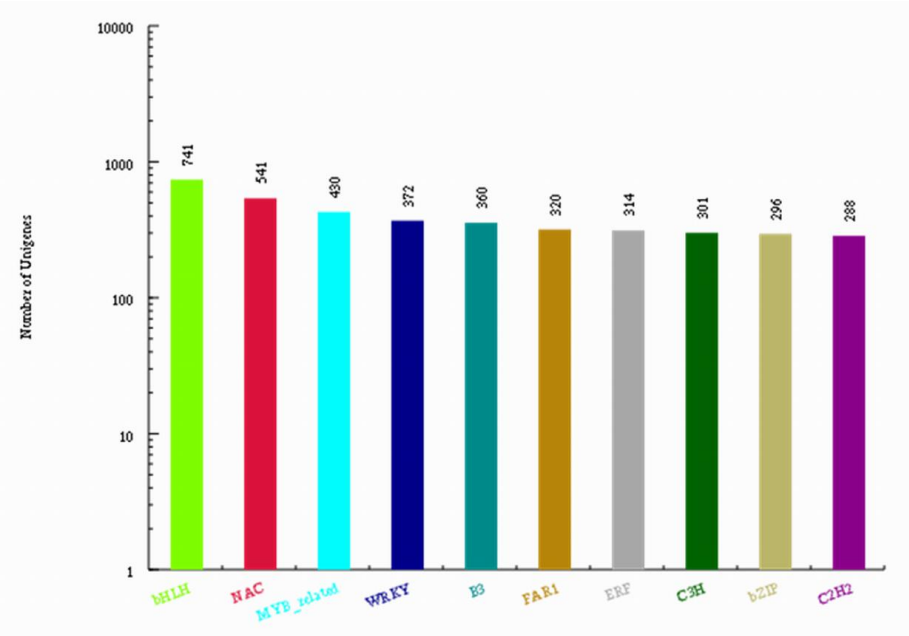


Fig. 11D. KEGG classification of DEGs between M2-CK and M2-T.

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.

841



842

843

844 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%

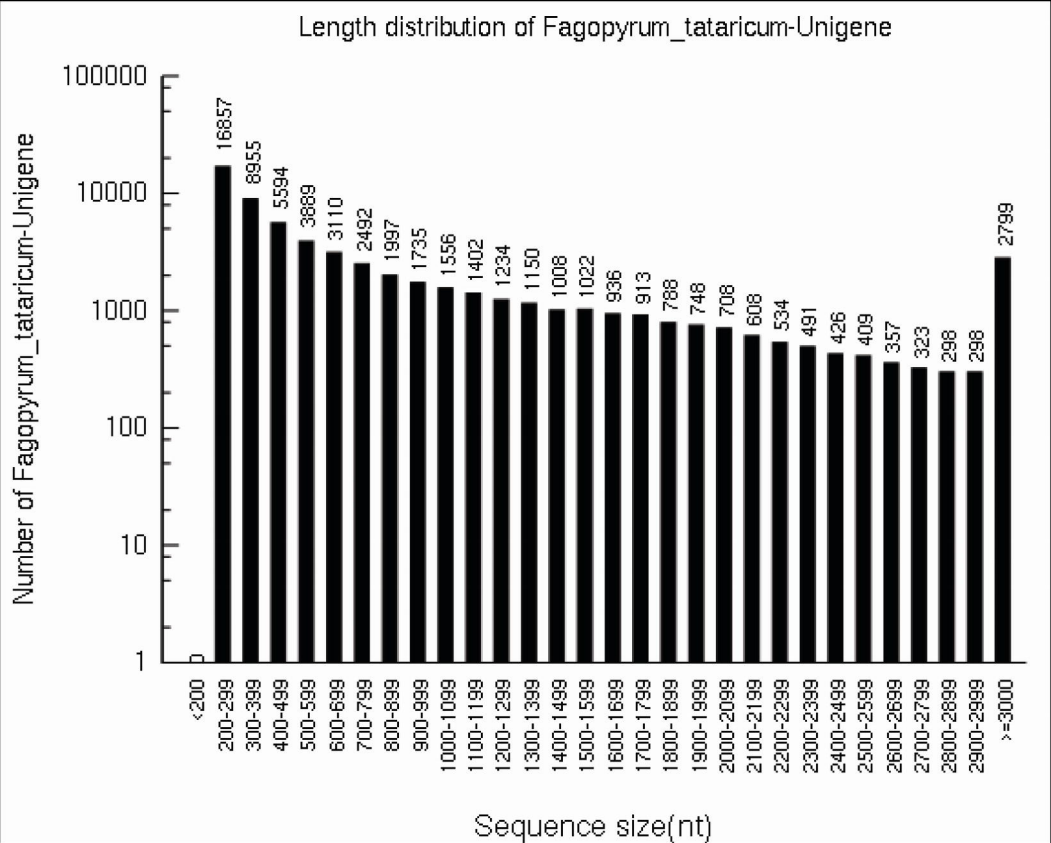
Notes: Q20 (%) and Q30 (%) are the percentages of reads with Phred scores over than 20 and 30, respectively. GC content (%) means G+C bases as the percentage of total bases; WT1-CK, M1-CK, WT2-CK, M2-CK and WT1-T , M1-T, WT2-T, M2-T represent control and salt-treated samples respectively.

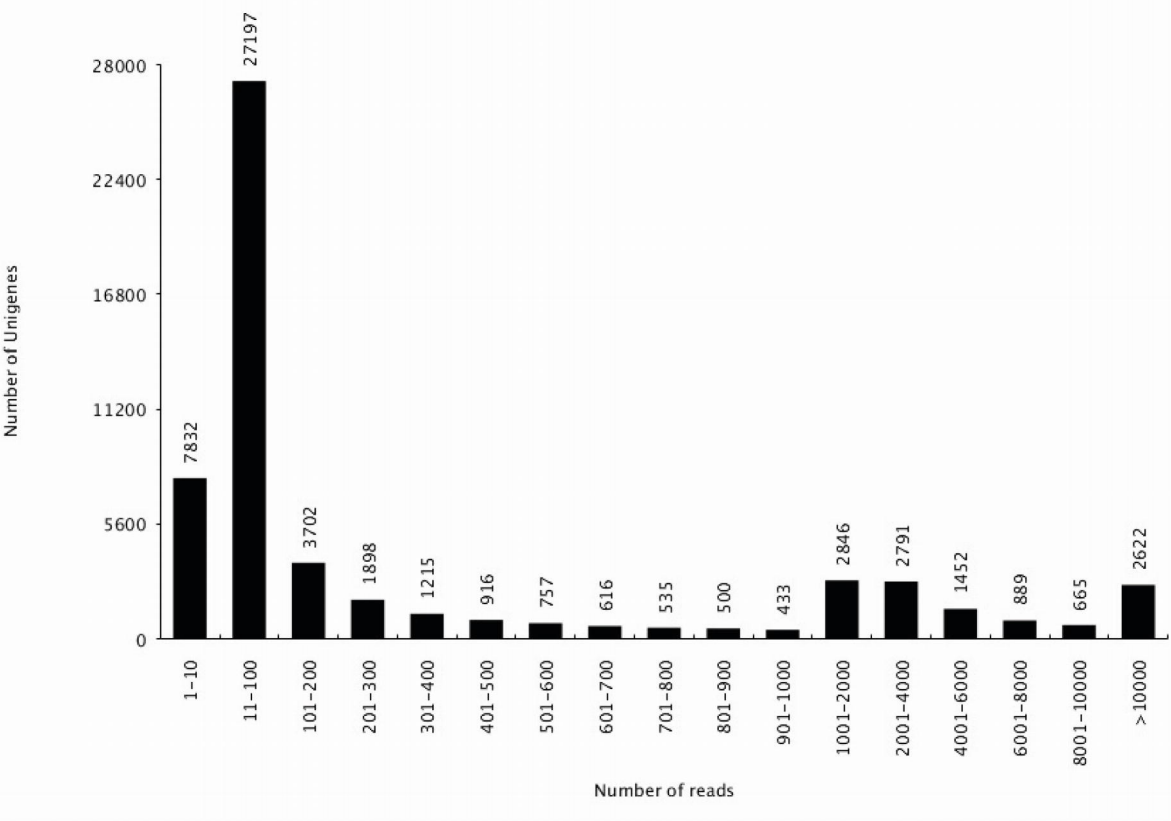
Table 3. Type statistics of SSRs.

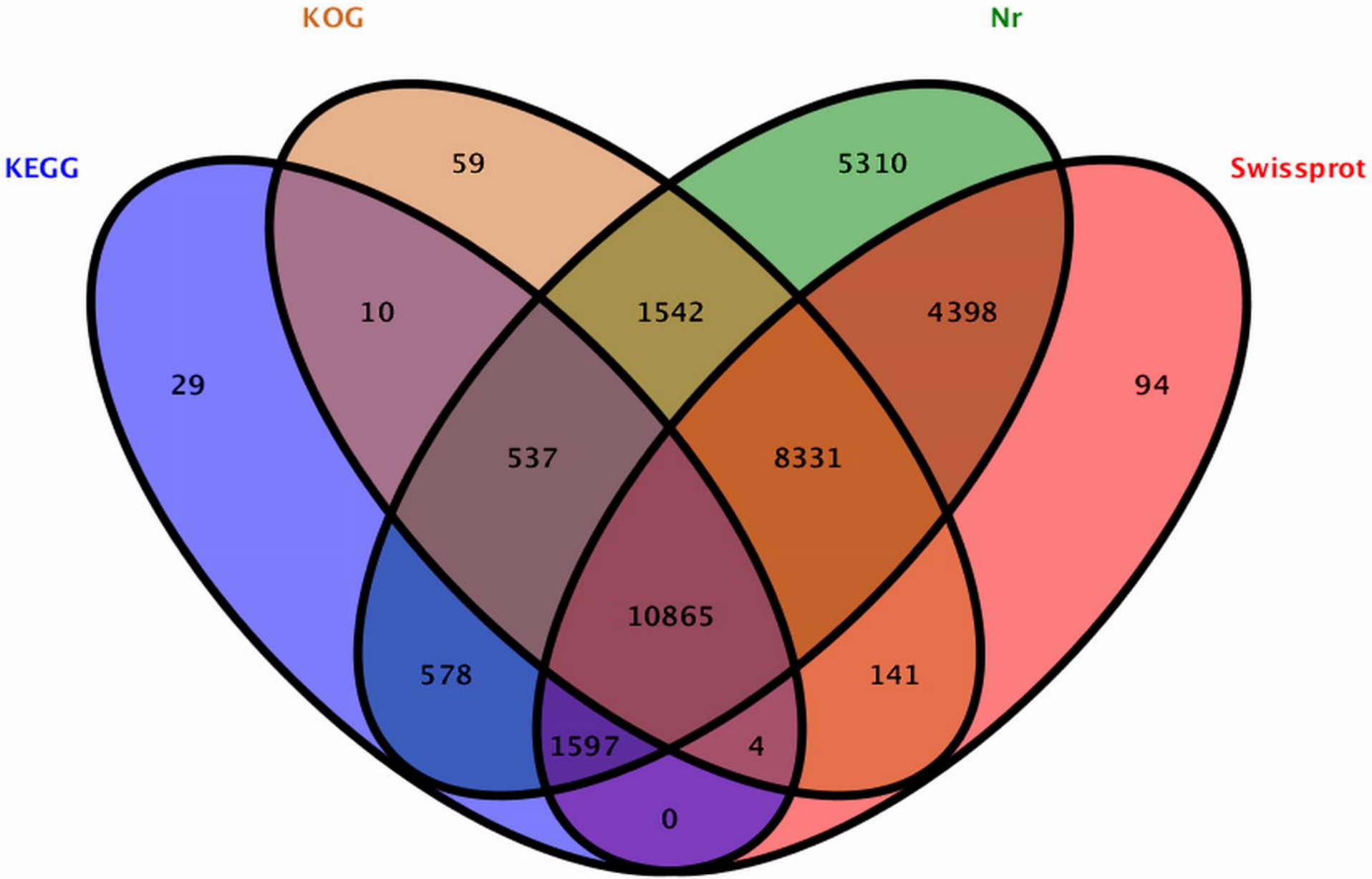
Stat Item	Num
Total number of sequences examined	62637
Total size of examined sequences (bp)	56788193
Total number of identified SSRs	3355
Number of SSR containing sequences	3025
Number of sequences containing more than 1 SSR	289
Number of SSRs present in compound formation	121
Di-nucleotide	1234
Tri-nucleotide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

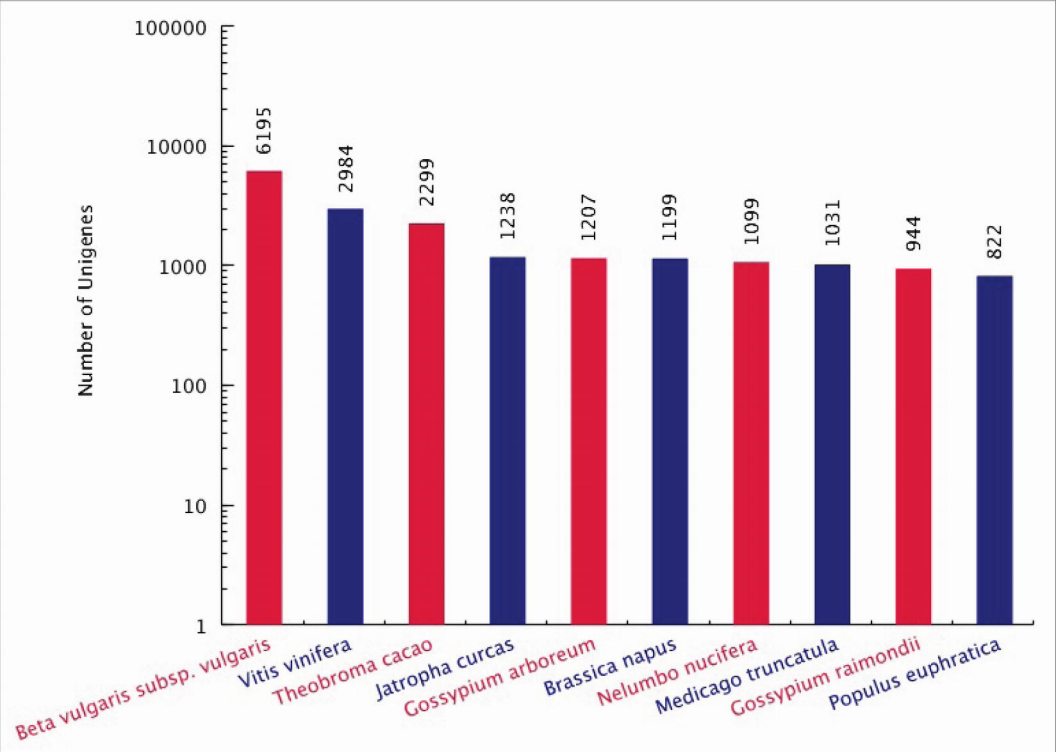
Table 4. RT-qPCR primer names and primer sequences.

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

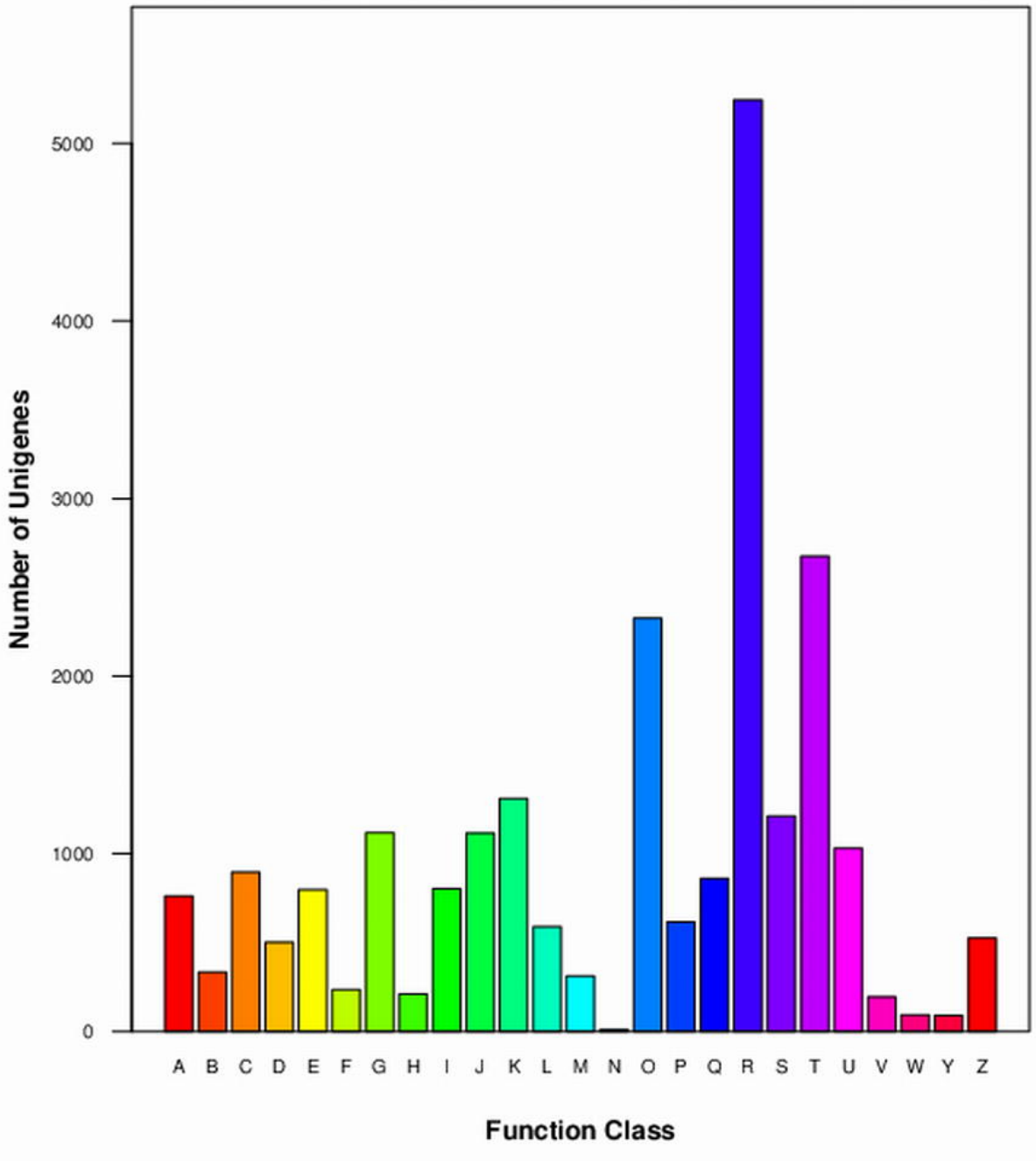

Length distribution of Fagopyrum_tataricum-Unigene
Number of Fagopyrum_tataricum-Unigene
100000
10000
1000
100
10
1
16857
8955
5594
3889
3110
2492
1997
1735
1556
1402
1234
1150
1008
1022
936
913
788
748
708
608
534
491
426
409
357
323
298
298
2799
<200
200-299
300-399
400-499
500-599
600-699
700-799
800-899
900-999
1000-1099
1100-1199
1200-1299
1300-1399
1400-1499
1500-1599
1600-1699
1700-1799
1800-1899
1900-1999
2000-2099
2100-2199
2200-2299
2300-2399
2400-2499
2500-2599
2600-2699
2700-2799
2800-2899
2900-2999
>=3000
Sequence size(nt)


Number of Unigenes
28000
22400
16800
11200
5600
0
7832
27197
3702
1898
1215
916
757
616
535
500
433
2846
2791
1452
889
665
2622
1-10
11-100
101-200
201-300
301-400
401-500
501-600
601-700
701-800
801-900
901-1000
1001-2000
2001-4000
4001-6000
6001-8000
8001-10000
>10000
Number of reads


KOG
Nr
KEGG
Swissprot
59
5310
10
1542
4398
29
94
537
8331
10865
578
141
1597
4
0


Number of Unigenes
100000
10000
1000
100
10
1
6195
2984
2299
1238
1207
1199
1099
1031
944
822
Beta vulgaris subsp. vulgaris
Vitis vinifera
Theobroma cacao
Jatropha curcas
Gossypium arboreum
Brassica napus
Nelumbo nucifera
Medicago truncatula
Gossypium raimondii
Populus euphratica

KOG Function Classification of Fagopyrum_tataricum–Unigene.fa Sequence



A: RNA processing and modification

B: Chromatin structure and dynamics

C: Energy production and conversion

D: Cell cycle control, cell division, chromosome partitioning

E: Amino acid transport and metabolism

F: Nucleotide transport and metabolism

G: Carbohydrate transport and metabolism

H: Coenzyme transport and metabolism

I: Lipid transport and metabolism

J: Translation, ribosomal structure and biogenesis

K: Transcription

L: Replication, recombination and repair

M: Cell wall/membrane/envelope biogenesis

N: Cell motility

O: Posttranslational modification, protein turnover, chaperones

P: Inorganic ion transport and metabolism

Q: Secondary metabolites biosynthesis, transport and catabolism

R: General function prediction only

S: Function unknown

T: Signal transduction mechanisms

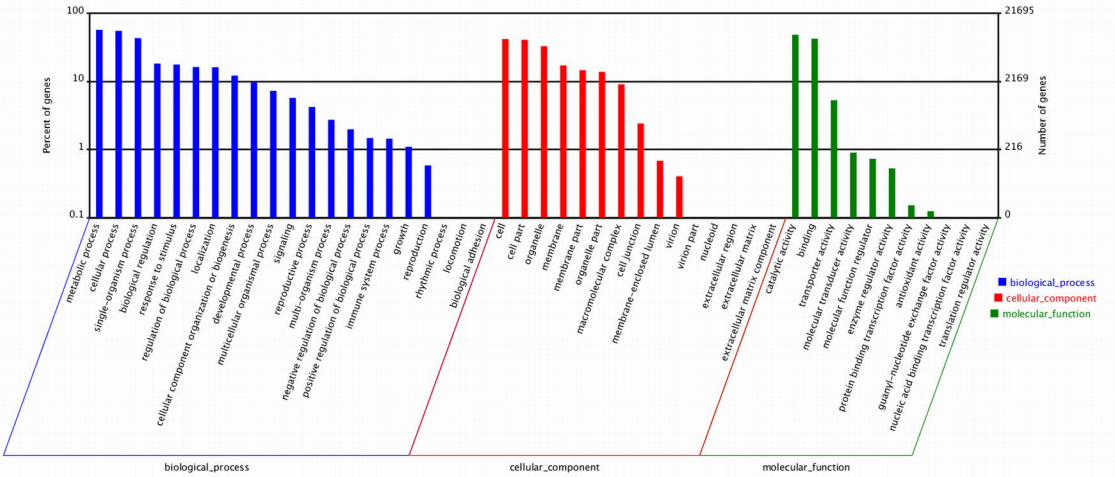
U: Intracellular trafficking, secretion, and vesicular transport

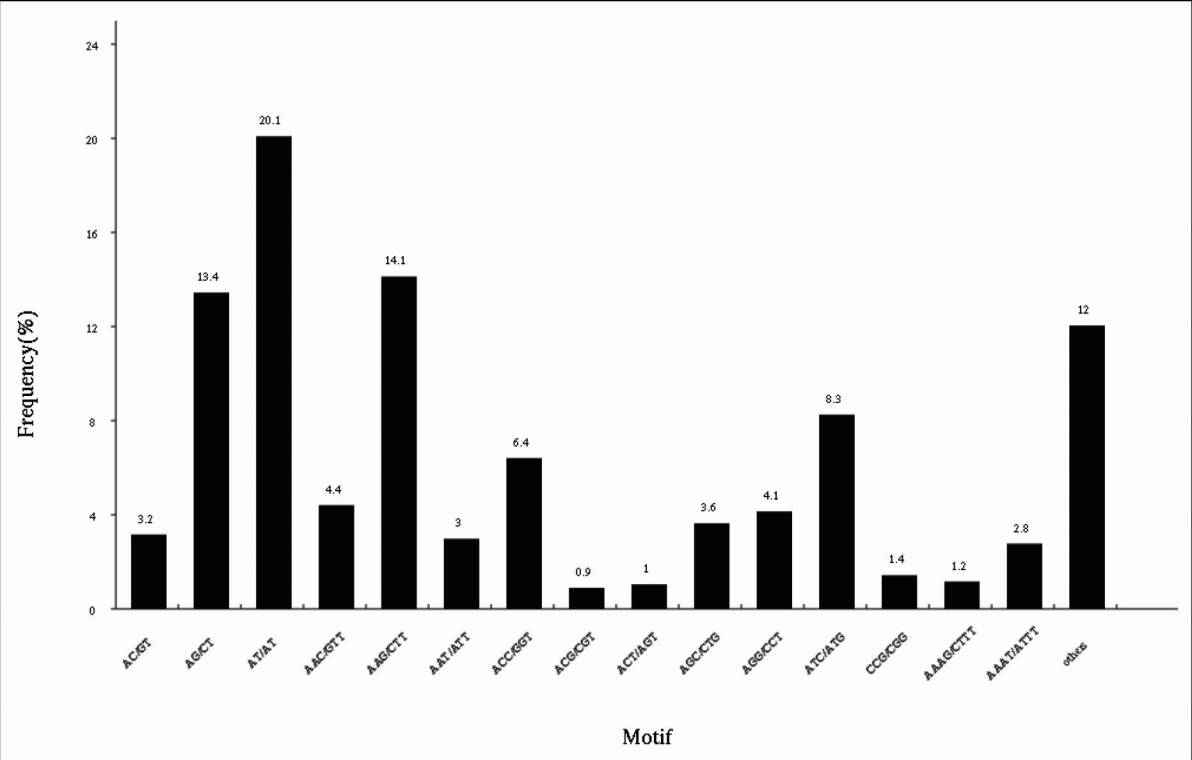
V: Defense mechanisms

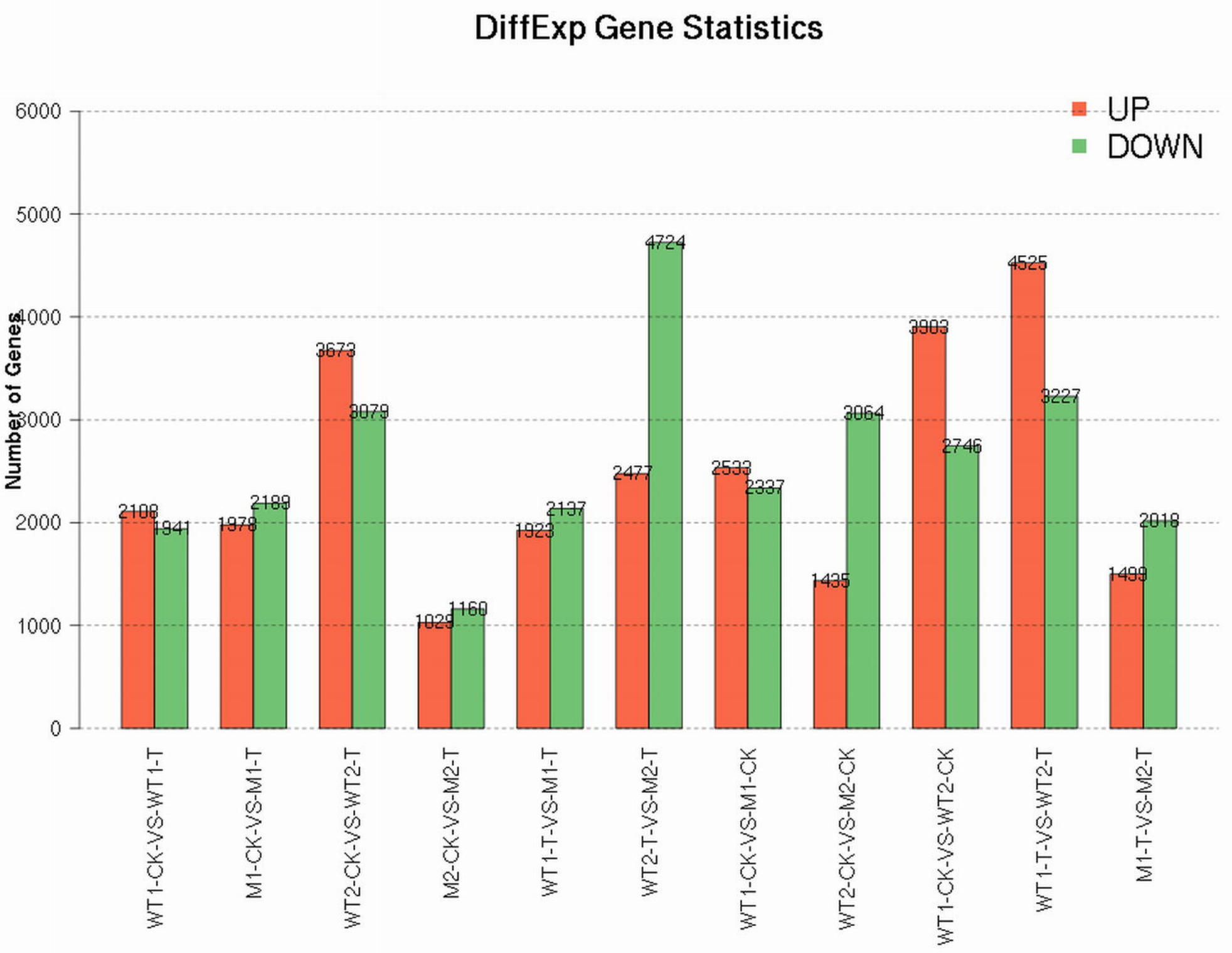
W: Extracellular structures

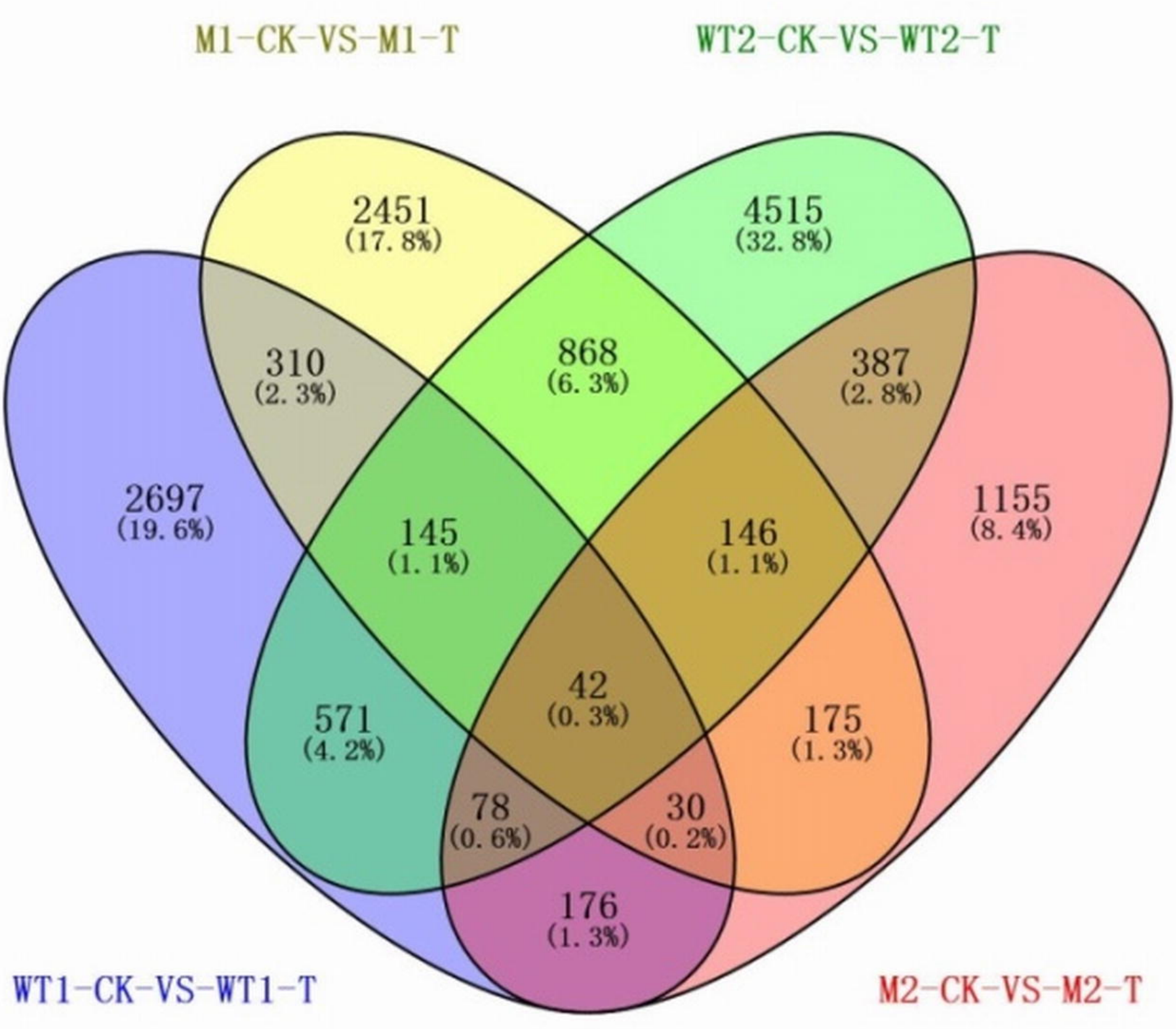
Y: Nuclear structure

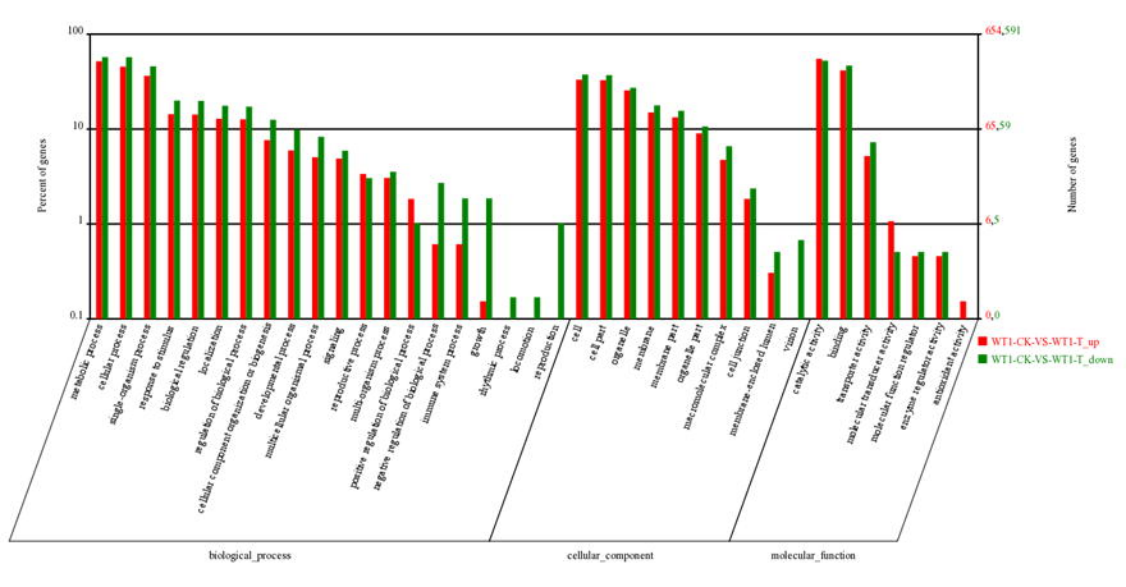
Z: Cytoskeleton

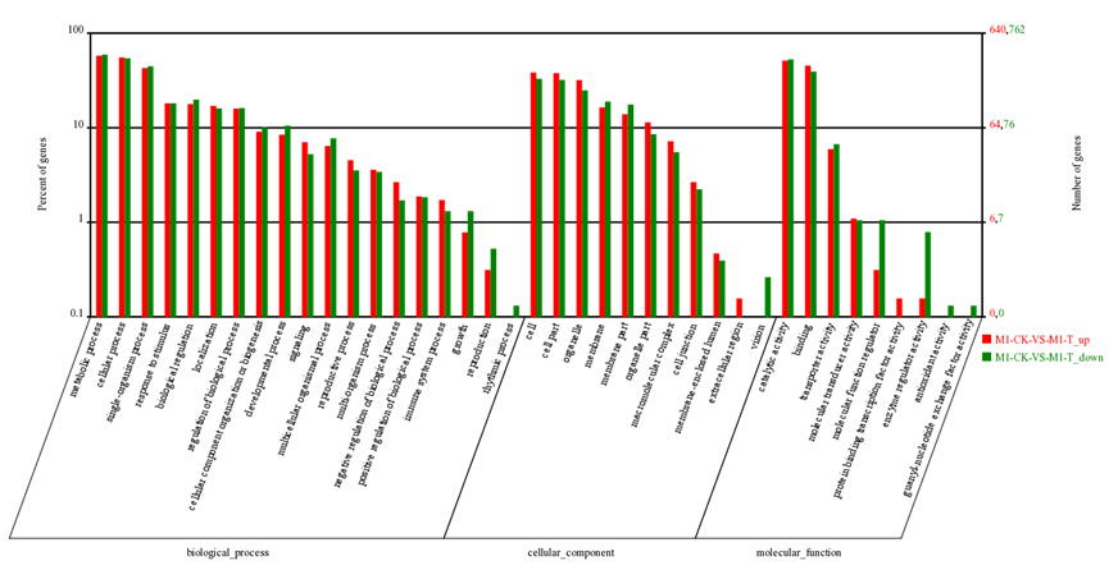

GO class of Fagopyrum_tataricum-Unigene
Percent of genes
Number of genes
100
10
1
0.1
21695
2169
216
0
metabolic process
cellular process
single-organism process
biological regulation
response to stimulus
regulation of biological process
cellular component organization or biogenesis
localization
developmental process
multicellular organismal process
signaling
reproductive process
multi-organism process
negative regulation of biological process
positive regulation of biological process
immune system process
growth
reproduction
rhythmic process
locomotion
biological adhesion
cell
cell part
organelle
membrane
membrane part
organelle part
macromolecular complex
cell junction
membrane-enclosed lumen
virion
virion part
nucleoid
extracellular region
extracellular matrix
extracellular matrix component
catalytic activity
binding
transporter activity
molecular transducer activity
molecular function regulator
enzyme regulator activity
protein binding transcription factor activity
antioxidant activity
guanyl-nucleotide exchange factor activity
nucleic acid binding transcription factor activity
translation regulator activity
biological_process
cellular_component
molecular_function
biological_process
cellular_component
molecular_function

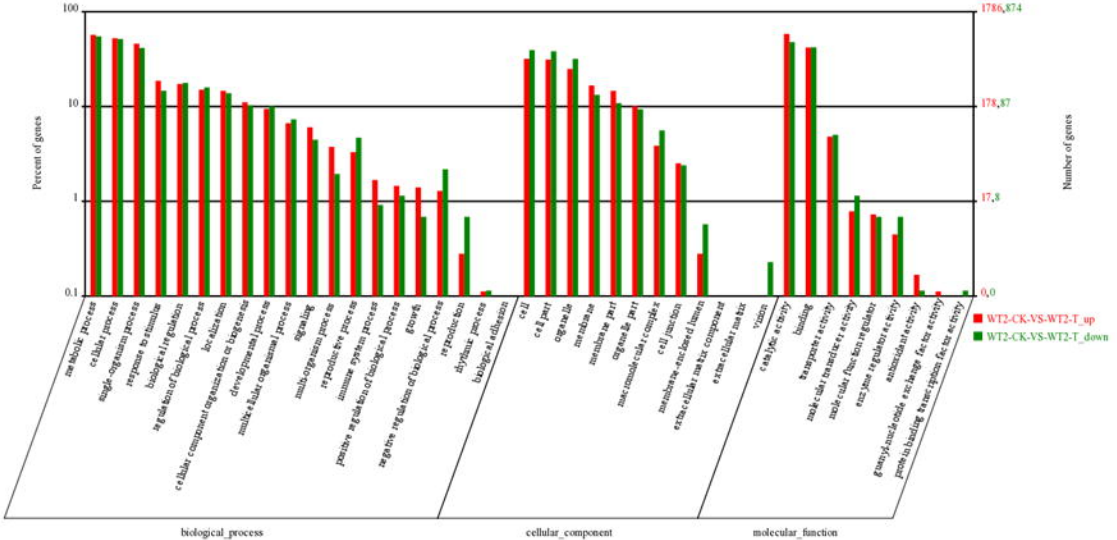

Frequency(%)
24
20
16
12
8
4
0
3.2
13.4
20.1
4.4
14.1
3
6.4
0.9
1
3.6
4.1
8.3
1.4
1.2
2.8
12
AC/GT
AG/CT
AT/AT
AAC/GTT
AAG/CTT
AAT/ATT
ACC/GGT
ACG/CGT
ACT/AGT
AGC/CTG
AGG/CCT
ATC/ATG
CCG/CGG
AAAG/CTTT
AAAT/ATTT
others
Motif

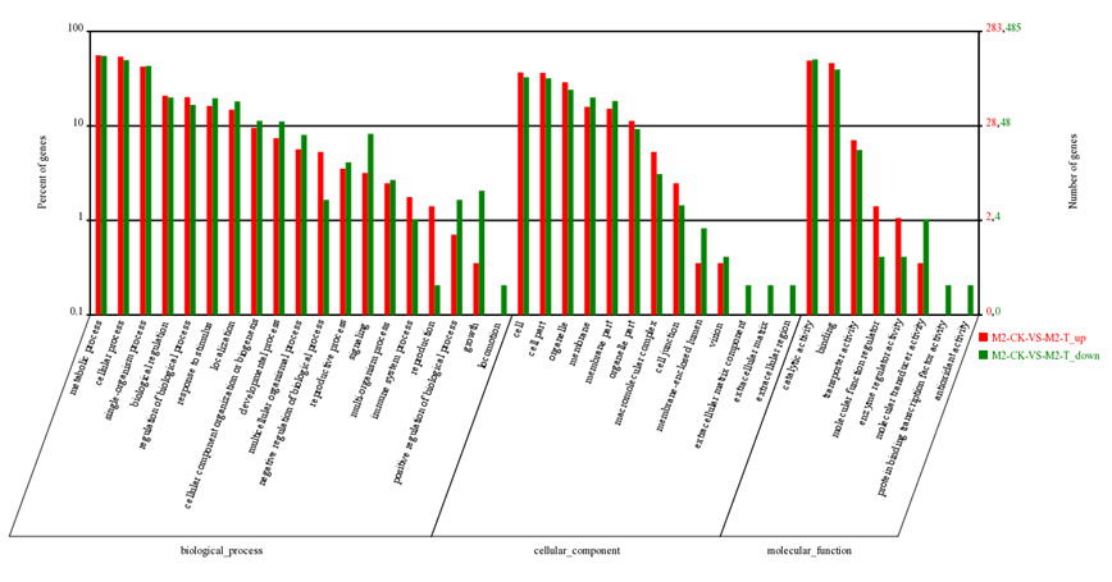

DiffExp Gene Statistics
UP
DOWN
Number of Genes
6000
5000
4000
3000
2000
1000
0
2108
1941
1978
2189
3673
3079
1029
1160
1923
2137
2477
4724
2533
2337
1435
3064
3903
2746
4525
3227
1498
2018
WT1-CK-VS-WT1-T
M1-CK-VS-M1-T
WT2-CK-VS-WT2-T
M2-CK-VS-M2-T
WT1-T-VS-M1-T
WT2-T-VS-M2-T
WT1-CK-VS-M1-CK
WT2-CK-VS-M2-CK
WT1-CK-VS-WT2-CK
WT1-T-VS-WT2-T
M1-T-VS-M2-T

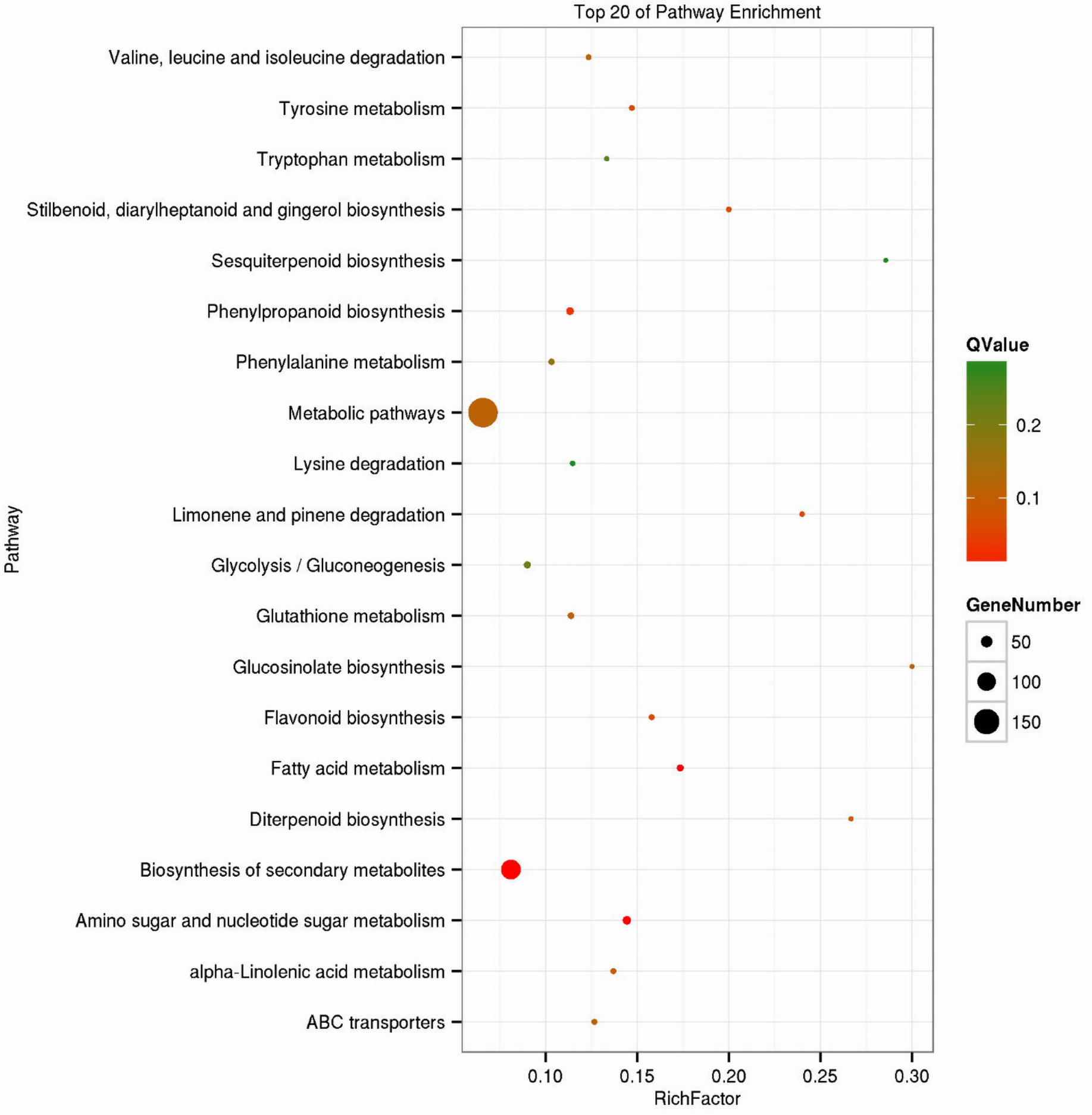

M1-CK-VS-M1-T
WT2-CK-VS-WT2-T
2451
(17.8%)
4515
(32.8%)
310
(2.3%)
868
(6.3%)
387
(2.8%)
2697
(19.6%)
145
(1.1%)
146
(1.1%)
1155
(8.4%)
42
(0.3%)
571
(4.2%)
175
(1.3%)
78
(0.6%)
30
(0.2%)
176
(1.3%)
WT1-CK-VS-WT1-T
M2-CK-VS-M2-T

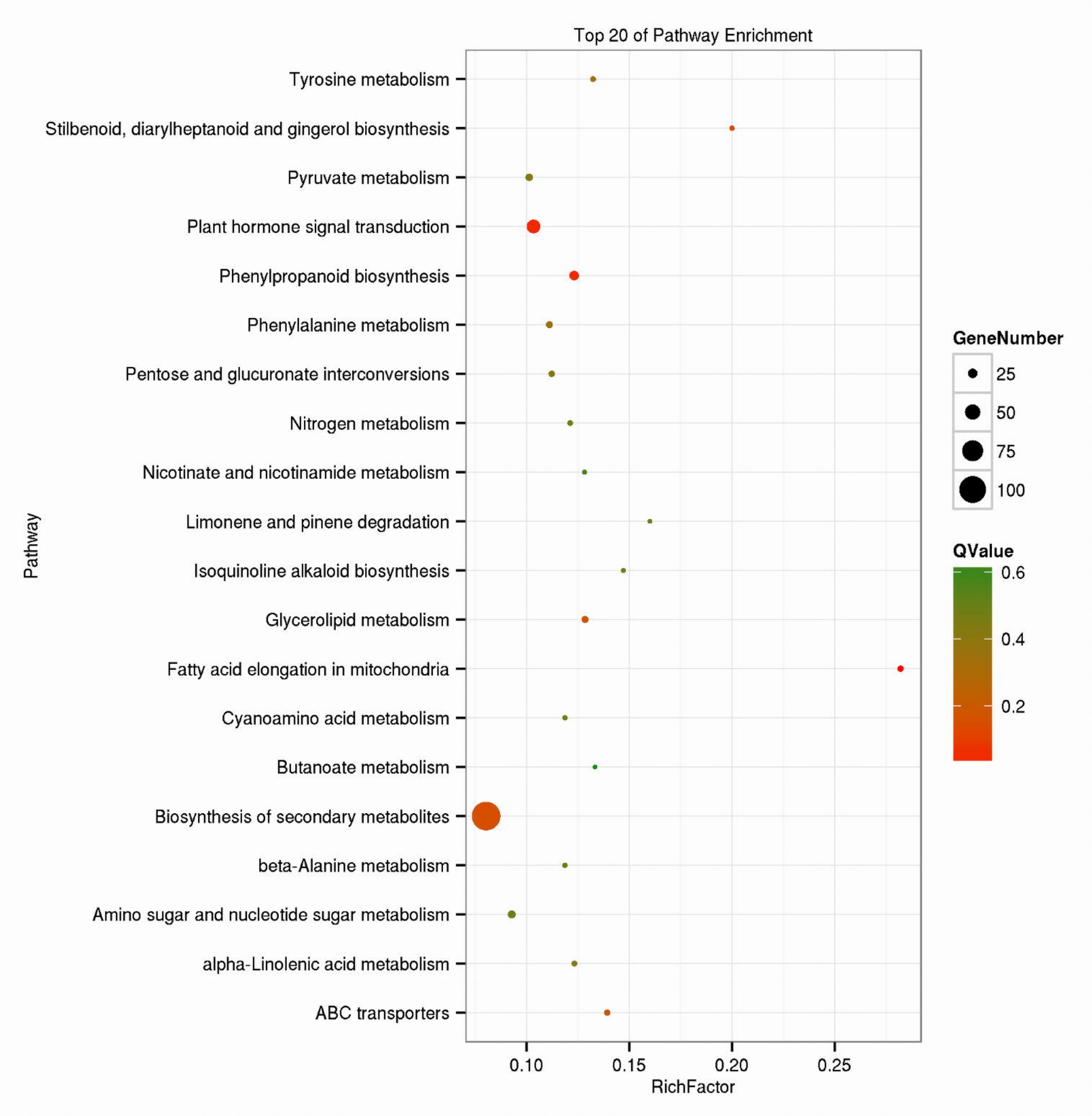

Percent of genes
100
10
1
0.1
654,591
65,59
6,5
0,0
Number of genes
WT1-CK-VS-WT1-T_up
WT1-CK-VS-WT1-T_down
metabolic process
cellular process
single-organism process
response to stimulus
biological regulation
localization
regulation of biological process
cellular component organization or biogenesis
developmental process
multicellular organismal process
signaling
reproductive process
multi-organism process
positive regulation of biological process
negative regulation of biological process
immune system process
growth
rhythmic process
locomotion
reproduction
cell
cell part
organelle
membrane
membrane part
organelle part
macromolecular complex
cell junction
membrane-enclosed lumen
virion
catalytic activity
binding
transporter activity
molecular transducer activity
molecular function regulator
structural molecule activity
antioxidant activity
biological_process
cellular_component
molecular_function

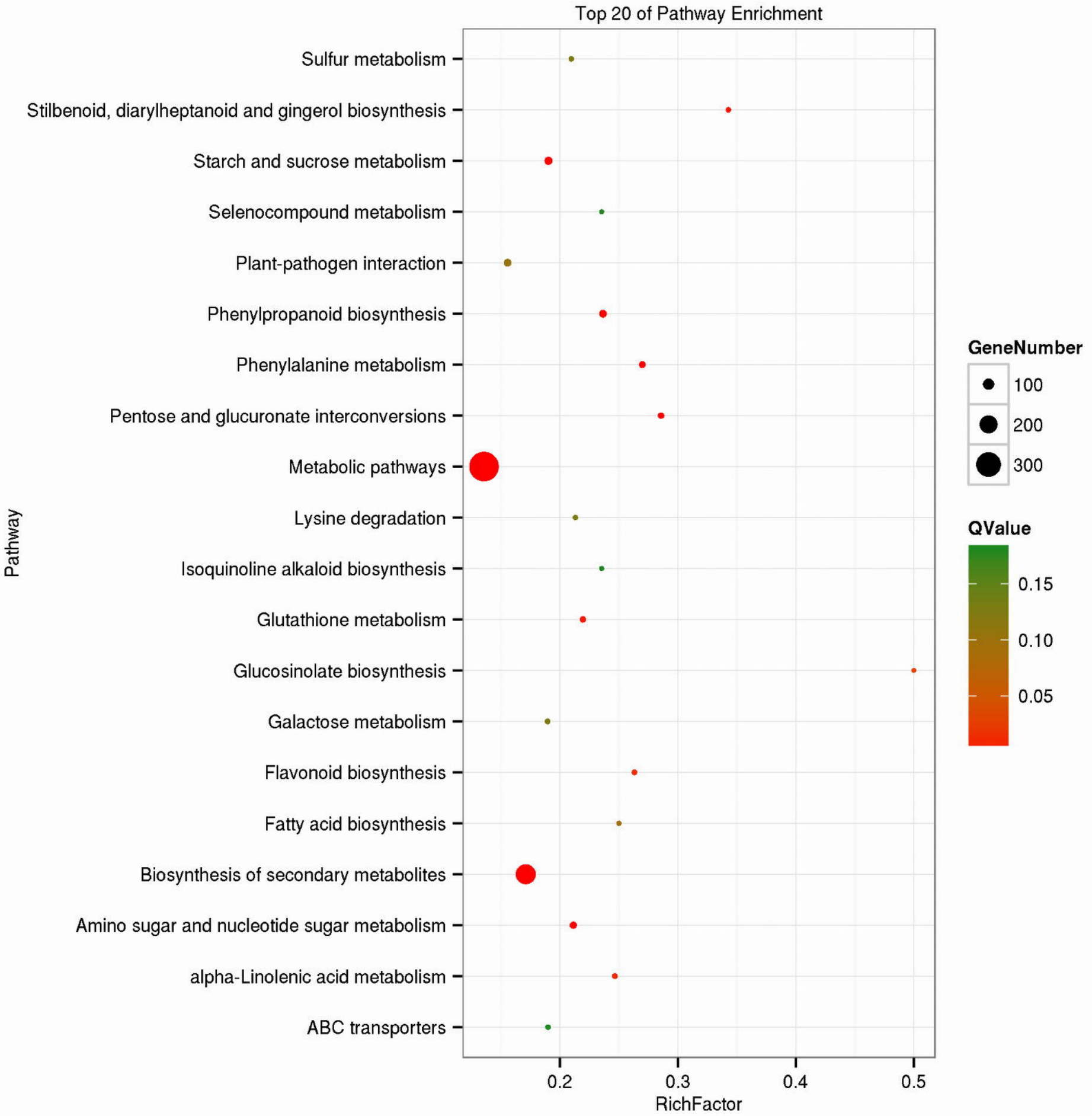

Percent of genes
100
10
1
0.1
Number of genes
640,762
64,76
6,7
0,0
M1-CK-VS-M1-T_up
M1-CK-VS-M1-T_down
multicellular organism process
cellular component organization or biogenesis
developmental process
signaling
multicellular organismal process
reproductive process
multiorganism process
negative regulation of biological process
positive regulation of biological process
immune system process
growth
reproduction
rhythmic process
cell
cell part
organelle
membrane
membrane part
organelle part
macromolecular complex
cell junction
membrane-enclosed lumen
extracellular region
virion
catalytic activity
binding
transporter activity
molecular transducer activity
molecular function regulator
protein binding transcription factor activity
enzyme regulator activity
antioxidant activity
translation regulator activity
nucleic acid binding transcription factor activity
biological_process
cellular_component
molecular_function

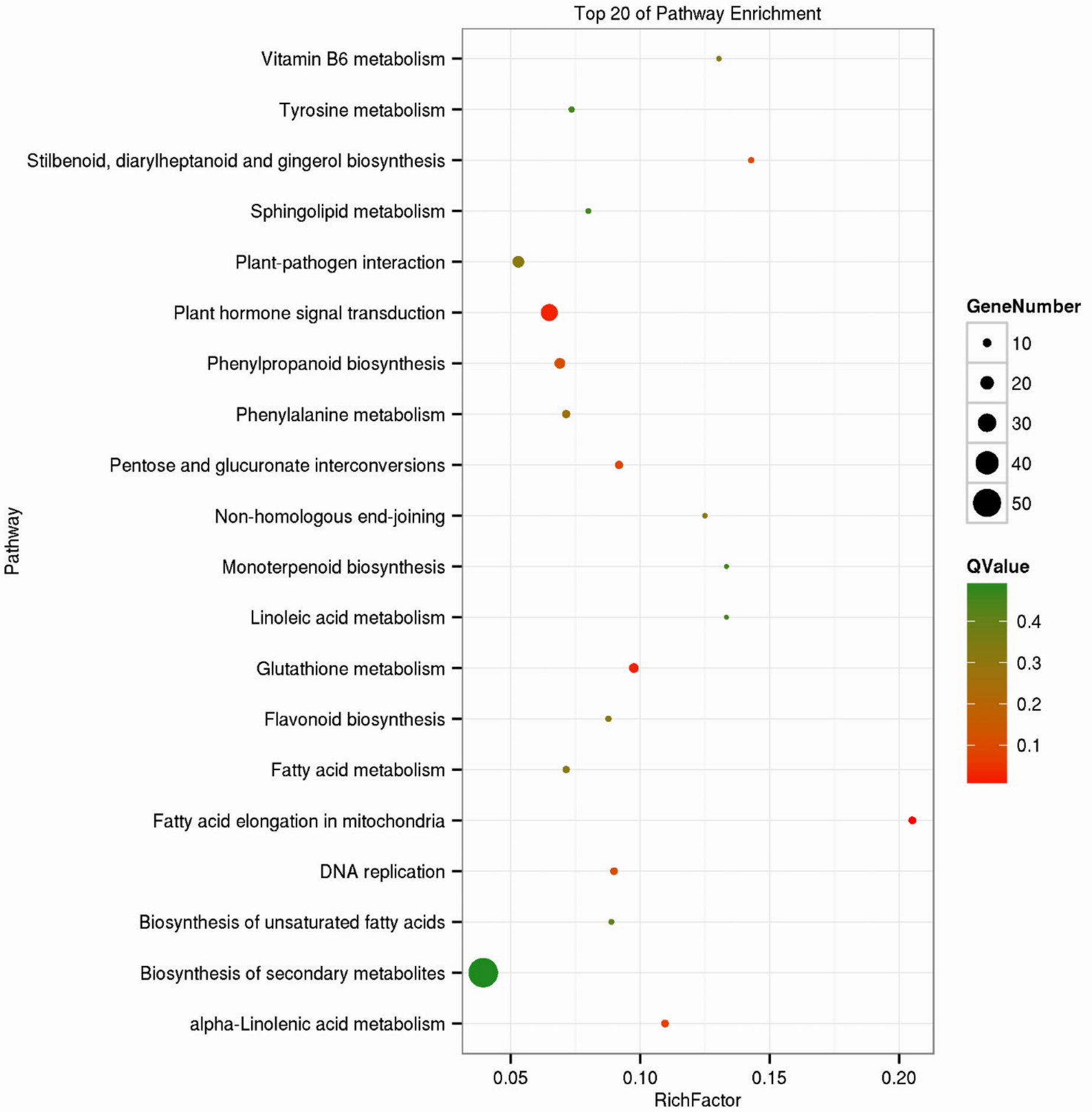

Percent of genes
100
10
1
0.1
Number of genes
1786,874
178,87
17,8
0,0
WT2-CK-VS-WT2-T_up
WT2-CK-VS-WT2-T_down
metabolic process
cellular process
single-organism process
response to stimulus
biological regulation
regulation of biological process
localization
cellular component organization or biogenesis
developmental process
multicellular organismal process
signaling
multi-organism process
reproductive process
immune system process
positive regulation of biological process
growth
negative regulation of biological process
reproduction
detoxification
biological adhesion
cell
cell part
organelle
membrane
membrane part
organelle part
macromolecular complex
cell junction
membrane-enclosed lumen
extracellular matrix component
extracellular matrix
virion
catalytic activity
binding
transporter activity
molecular transducer activity
nucleic acid binding transcription factor activity
enzyme regulator activity
structural molecule activity
antioxidant activity
guanyl-nucleotide exchange factor activity
protein binding transcription factor activity
biological_process
cellular_component
molecular_function


Percent of genes
100
10
1
0.1
Number of genes
283,485
28,48
2,4
0,0
M2-CK-VS-M2-T_up
M2-CK-VS-M2-T_down
metabolic process
cellular process
single-organism process
biological regulation
regulation of biological process
response to stimulus
localization
cellular component organization or biogenesis
developmental process
multicellular organismal process
negative regulation of biological process
reproductive process
signaling
multi-organism process
immune system process
reproduction
positive regulation of biological process
growth
locomotion
cell
cell part
organelle
membrane
membrane part
organelle part
macromolecular complex
cell junction
membrane-enclosed lumen
virion
extracellular matrix component
extracellular matrix
extracellular region
catalytic activity
binding
transporter activity
molecular function regulator
enzyme regulator activity
molecular transducer activity
protein binding transcription factor activity
antioxidant activity
biological_process
cellular_component
molecular_function


Top 20 of Pathway Enrichment
Valine, leucine and isoleucine degradation
Tyrosine metabolism
Tryptophan metabolism
Stilbenoid, diarylheptanoid and gingerol biosynthesis
Sesquiterpenoid biosynthesis
Phenylpropanoid biosynthesis
Phenylalanine metabolism
Metabolic pathways
Lysine degradation
Limonene and pinene degradation
Glycolysis / Gluconeogenesis
Glutathione metabolism
Glucosinolate biosynthesis
Flavonoid biosynthesis
Fatty acid metabolism
Diterpenoid biosynthesis
Biosynthesis of secondary metabolites
Amino sugar and nucleotide sugar metabolism
alpha-Linolenic acid metabolism
ABC transporters
Pathway
0.10
0.15
0.20
0.25
0.30
RichFactor
QValue
0.2
0.1
GeneNumber
50
100
150


Top 20 of Pathway Enrichment
Tyrosine metabolism
Stilbenoid, diarylheptanoid and gingerol biosynthesis
Pyruvate metabolism
Plant hormone signal transduction
Phenylpropanoid biosynthesis
Phenylalanine metabolism
Pentose and glucuronate interconversions
Nitrogen metabolism
Nicotinate and nicotinamide metabolism
Limonene and pinene degradation
Isoquinoline alkaloid biosynthesis
Glycerolipid metabolism
Fatty acid elongation in mitochondria
Cyanoamino acid metabolism
Butanoate metabolism
Biosynthesis of secondary metabolites
beta-Alanine metabolism
Amino sugar and nucleotide sugar metabolism
alpha-Linolenic acid metabolism
ABC transporters
Pathway
0.10
0.15
0.20
0.25
RichFactor
GeneNumber
25
50
75
100
QValue
0.6
0.4
0.2


Top 20 of Pathway Enrichment
Sulfur metabolism
Stilbenoid, diarylheptanoid and gingerol biosynthesis
Starch and sucrose metabolism
Selenocompound metabolism
Plant-pathogen interaction
Phenylpropanoid biosynthesis
Phenylalanine metabolism
Pentose and glucuronate interconversions
Metabolic pathways
Lysine degradation
Isoquinoline alkaloid biosynthesis
Glutathione metabolism
Glucosinolate biosynthesis
Galactose metabolism
Flavonoid biosynthesis
Fatty acid biosynthesis
Biosynthesis of secondary metabolites
Amino sugar and nucleotide sugar metabolism
alpha-Linolenic acid metabolism
ABC transporters
Pathway
0.2
0.3
0.4
0.5
RichFactor
GeneNumber
100
200
300
QValue
0.15
0.10
0.05


Top 20 of Pathway Enrichment
Vitamin B6 metabolism
Tyrosine metabolism
Stilbenoid, diarylheptanoid and gingerol biosynthesis
Sphingolipid metabolism
Plant-pathogen interaction
Plant hormone signal transduction
Phenylpropanoid biosynthesis
Phenylalanine metabolism
Pentose and glucuronate interconversions
Non-homologous end-joining
Monoterpenoid biosynthesis
Linoleic acid metabolism
Glutathione metabolism
Flavonoid biosynthesis
Fatty acid metabolism
Fatty acid elongation in mitochondria
DNA replication
Biosynthesis of unsaturated fatty acids
Biosynthesis of secondary metabolites
alpha-Linolenic acid metabolism
Pathway
0.05
0.10
0.15
0.20
RichFactor
GeneNumber
10
20
30
40
50
QValue
0.4
0.3
0.2
0.1

bioRxiv preprint doi: https://doi.org/10.1101/869263; this version posted December 10, 2019. The copyright holder has placed this preprint (which was not certified by peer review) in the Public Domain. It is no longer restricted by copyright. Anyone can legally share, reuse, remix, or adapt this material for any purpose without crediting the original authors.