Transcriptome and small RNA combined sequencing analysis of cold tolerance response in non-heading Chinese cabbage

Jin Wang
 Nanjing Agricultural University

Qinxue Zhang
 Nanjing Agricultural University

Xiong You
 Nanjing Agricultural University

Xilin Hou (✉ hxl@njau.edu.cn)
 Nanjing Agricultural University   https://orcid.org/0000-0002-9337-177X

Research article

Abstract

Background: Non-heading Chinese cabbage (Brassica rapa ssp. chinensis), as an important leaf vegetable grown worldwide. However, there is currently no enough transcriptome and small RNA combined sequencing analysis of cold tolerance, which hinders further functional genomics research.

Results: In this study, 63.43 Gb of clean data was obtained from the transcriptome analysis. The clean data of each sample reached 6.99 Gb, and the basic percentage of Q30 was 93.68% and above. The clean reads of each sample were sequence aligned with the designated reference genome (Brassica rapa, IVFCAASv1), and the efficiency of the alignment varied from 81.54% to 87.24%. According to the comparison results, 1,860 new genes were discovered, of which 1,613 were functionally annotated. Among them, 13 common differentially expressed genes were detected in all materials, including 7 up-regulated and 6 down-regulated. At the same time, we used quantitative real-time PCR to confirm the changes of these gene expression levels. In addition, we sequenced miRNA of the same material. Our findings revealed a total of 34,182,333 small RNA reads, 88,604,604 kinds of small RNA, among which the most common size was 24 nt. In all materials, the number of common differential miRNAs is 8. According to the corresponding relationship between miRNA and its target genes, we carried out Gene Ontology and Kyoto Encyclopedia of Genes and Genomes enrichment analysis on the set of target genes that each group of differentially expressed miRNAs. Through the analysis, it is found that the distribution of candidate target genes in different materials is different. We not only use transcriptome sequencing and small RNA sequencing, but also use experiment to prove that the expression level of differentially expressed genes are obtained by sequencing. Sequencing combined with experiments proved the mechanism of some differential gene expression levels after low temperature treatment.

Conclusions: In all, this study provides a resource for genetic and genomic research under abiotic stress in Pak-choi.

Background

One of the most common cultivated vegetables in China, non-heading Chinese cabbage (*Brassica rapa* ssp. *chinensis*), also known as Pak-choi, while its growth are often affected, from the surroundings by cold even freezing stresses [1, 2]. Low temperature, well known, as an important factor, affecting growth and reducing productivity [3-5]. However, the cold domestication of plants, which usually involves many biochemical and physiological changes, which is complicated and difficult to understand [4-7]. The metabolism and transcriptome of plants, which can be greatly affected by cold stress, after low temperature treatment, the expression levels of certain genes will also be regulated, as we all know, some related metabolic enzymes will be inhibited, and the degree of plant metabolism will be affected to some extent [4]. Nowadays, genetic, biochemical, and various sequencing methods, which have been applied to respond to many genes that can cope with corresponding environmental stress [8-10]. To better understand the response mechanisms of cold stress, many plants have been studied today, such as rice [11, 12], cotton [13], tomato [14], potato [15], muskmelons [16], and sugarcane [17-19]. However, the

response mechanisms of non-heading Chinese cabbage to cold stress, which is not very clear, and further joint research is needed.

Nowadays, transcriptome analysis, which has gradually become a useful and general tool for discovering genes in multiple stress pathways, including determining the expression patterns of related genes [20-22]. The related genes reported in plant secondary metabolism, which can be discovered based on experimental methods of functional genome sequencing [23, 24]. In addition, RNA sequencing technology, which is often used to obtain complete transcriptome information from different plants, such as tea tree, chlorophytum borivilianum, and atractylodes lancea, and provide better insights into transcription or post-transcriptional force, including regulation of essential genes, during secondary metabolite biosynthetic pathways [25-27]. Nowadays, transcriptome sequencing, which has been successfully used to detect expression levels of related genes in many organisms, such as rice [28], yeast [29, 30], sweet potato [31], and taxus [32]. With Illumina sequencing technology, read millions of sequences at a time, and map individual assembled genes into a reference transcriptome map for molecular annotation [33].

Studies have found that microRNA (miRNA), trans short interfering RNA (ta-siRNA), and heterochromatic short interfering RNA (hc-siRNA), all of which play important roles in different organisms [34]. Among them, miRNA is composed of about 22 nucleotides (nt), which is a type of endogenous small RNA, and usually also plays a negative role in regulating gene expression [35]. Many studies have shown that miRNAs, which are often involved in plant development, hormone signaling, and abiotic stress responses [36, 37]. Generally speaking, small interfering RNAs are processed from perfect double-stranded RNA, while miRNAs are derived from single-stranded RNA transcripts, which form an imperfect double-stranded stem-loop precursor structure [38-41]. On the whole, miRNA, which plays a vital role in various biological and metabolic processes of plant growth and development, such as biotic (or abiotic) stresses, which can also negatively regulate the expression of target genes, by inhibiting (cutting) target mRNA or other ways [35, 37, 42-46]. Therefore, it is important to identify miRNAs and their target genes (or miRNAs), which is essential for a better understanding of miRNA-mediated regulation of cold stress genes.

In this study, we compared the tolerance of two common and typical non-heading Chinese cabbage varieties, *Suzhouqing* (*BcL.1*) and *Sijiucaixin* (*BcL.2*), to cold stress. We found that *BcL.1* is more tolerant to cold stress compared with *BcL.2*. We used RNA-Seq for comprehensive characterization and explored the effects of low temperature on global change. We identified the some most important genes in the low temperature response, and discussed their regulatory networks under cold stress. Furthermore, we identified conserved and novel miRNAs and their potential target genes in non-heading Chinese cabbage, and explored their interactions. Quantitative real-time polymerase chain reaction (qRT-PCR), which was also used to assess the expression levels of common differentially expressed genes (DEGs), and identify those candidate genes involved in cold tolarence. This work might serve as a reference that the functional analysis of cold tolarence in non-heading Chinese cabbage.

Results

Quantity statistics and veen diagram of differentially expressed genes

To study the effects of temperature on plant growth, plants (*BcL.1* and *BcL.2*) were grown for up to 6 hours in environments of 25 °C and 4 °C. Except for materials and temperature, the other conditions remain the same. Then, we observed that low temperatures have an important effect on plant phenotype. Whether it is *BcL.1* or *BcL.2*, under 4 °C treatment, leaves of plant are more likely to shrink or even wither than 25 °C treatment (Figure 1). This result corresponds to previous reports that cold stress usually lowers the seedling vigor [12, 47], leaf atrophy, slows crop growth and ultimately reduces yield [48-50].

Afterward, to study the up-regulation and down-regulation of common genes shared by each group of treatments, we established a venn diagram of differentially expressed genes. Between G0 (*BcL.2-25* vs *BcL.2-4*) and G2 (*BcL.1-25* vs *BcL.1-4*) groups, a total of 313 common genes were up-regulated and 308 common genes were down-regulated (Figure 2A and 2B); while between G1 (*BcL.1-25* vs *BcL.2-25*) and G3 (*BcL.1-4* vs *BcL.2-4*) groups, a total of 344 common genes were found to be up-regulated, 117 common genes were down-regulated (Figure 2C and 2D); in all materials, a total of 7 common genes were up-regulated and 6 common genes were down-regulated (Figure 2E and 2F; Table S1). We speculated that these DEGs might help increase the potential application value of non-heading Chinese cabbage under cold stress.

Functional annotation and classification

Between the *BcL.1-25* and *BcL.1-4* groups, 6,208 DEGs ($p < 0.05$) were detected, including 3,639 up-regulated and 2,569 down-regulated genes (Figure 3C). The annotated unigenes were then assigned to Gene Othology (GO) terms for functional classification. Three main categories (biological process, molecular function, and cellular component) of GO classification were analyzed separately to investigate their functional distribution. To simplify the functional distribution of plants, we assigned the annotated sequences to GO-slim terms to obtain a "thin" version of classification [51]. Cellular process (GO:0009987, 3,589 genes) and metabolic process (GO:0008152, 3,424 genes) in the biological process, cell part (GO:0004464, 5,170 genes) and cells (GO:0005623, 5,169 genes) in the cellular component and binding activity (GO:0005488, 2,808 genes) and catalytic activity (GO:0003824, 2,279 genes) in the molecular function were the most representative level 2 GO terms in all three data sets, respectively (Figure 3A; Table S2). To further identify the active biochemical pathways, we mapped it to the reference canonical pathways, which in the Kyoto Encyclopedia of Genes and Genomes (KEGG). KEGG is thought to provide a basic platform for systematic analysis of gene function in terms of the network of gene products [52]. A total of 24,199 unigenes were annotated based on a BLASTX search of the KEGG database (Table S3): 263 biosynthesis pathways were predicted and classified into five categories. Of which, the ribosome pathway was the largest, containing 287 genes (287 out of 1,586, 18.10%) (Figure 3B; Figure S1A; Table S10). The annotated unigenes were categorized into different functional groups based on the Cluster of Orthologus Groups (COG) database (Table S14). 3,700 unigenes could be classified into 23 COG categories. Out of 3,700 unigenes, general function prediction only (679, 18.35%) was assigned into the COG category of general function prediction, which represented the largest functional group of the 23

COG categories, followed by translation, ribosomal structure and biogenesis (433, 11.70%), transcription (329, 8.90%), replication, recombination and repair (293, 7.92%) and signal transduction mechanisms (290, 7.84%) (Figure S2A).

Between the *BcL.2-25* and *BcL.2-4* groups, 964 DEGs ($p < 0.05$) were detected, including 437 up-regulated and 527 down-regulated genes (Figure 4C). Through GO enrichment stratification analysis, cellular process (GO:0009987, 552 genes) and metabolic process (GO:0008152, 534 genes) in the biological process, cell part (GO:0004464, 741 genes) and cells (GO:0005623, 741 genes) in the cellular component and binding activity (GO:0005488, 429 genes) and catalytic activity (GO:0003824, 433 genes) in the molecular function, which were the most representative level 2 GO terms, respectively (Figure 4A; Table S4). Through KEGG pathway enrichment analysis (Table S5), the protein processing in endoplasmic reticulum pathway was the largest, containing 26 genes (26 out of 251, 10.36%) (Figure 4B; Figure S1B; Table S11). Through COG classification of differentially expressed gene. Out of 554 unigenes (Table S15), general function prediction only (109, 19.68%) was assigned into the COG category of general function prediction, which represented the largest functional group of the 21 COG categories, followed by posttranslational modification, protein turnover, chaperones (59, 10.65%), amino acid transport and metabolism (49, 8.84%), carbohydrate transport and metabolism (45, 8.12%) and signal transduction mechanisms (35, 6.32%) (Figure S2B).

Between the *BcL.1-25* and *BcL.2-25* groups, 1,448 DEGs ($p < 0.05$) were detected, including 1,052 up-regulated and 396 down-regulated genes (Figure 5C). Through GO enrichment stratification analysis, cellular process (GO:0009987, 729 genes) and metabolic process (GO:0008152, 695 genes) in the biological process, cell part (GO:0004464, 988 genes) and cells (GO:0005623, 989 genes) in the cellular component and binding activity (GO:0005488, 556 genes) and catalytic activity (GO:0003824, 465 genes) in the molecular function were the most representative level 2 GO terms, respectively (Figure 5A; Table S6). Through KEGG pathway enrichment analysis (Table S7), the ribosome pathway was the largest, containing 96 genes (96 out of 366, 26.23%) (Figure 5B; Figure S1C; Table S12). Through COG classification of differentially expressed gene. Out of 681 unigenes (Table S16), general function prediction only (124, 18.21%) was assigned into the COG category of general function prediction, which represented the largest functional group of the 23 COG categories, followed by translation, ribosomal structure and biogenesis (109, 16.01%), transcription (51, 7.49%), replication, recombination and repair (50, 7.34%) and amino acid transport and metabolism (44, 6.46%) (Figure S2C).

Between the *BcL.1-4* and *BcL.2-4* groups, 1,479 DEGs ($p < 0.05$) were detected, including 854 up-regulated and 625 down-regulated genes (Figure 6C). Through GO enrichment stratification analysis, cellular process (GO:0009987, 750 genes) and metabolic process (GO:0008152, 697 genes) in the biological process, cell part (GO:0004464, 1,056 genes) and cells (GO:0005623, 1,057 genes) in the cellular component and binding (GO:0005488, 610 genes) and catalytic activity (GO:0003824, 536 genes) in the molecular function, which were the most representative level 2 GO terms, respectively (Figure 6A; Table S8). Through KEGG pathway enrichment analysis (Table S9) , the DNA replication pathway was the largest, containing 13 genes (13 out of 329, 3.95%) (Figure 6B; Figure S1D; Table S13). Through COG

classification of differentially expressed gene. Out of 681 unigenes (Table S17), general function prediction only (144, 17.98%) was assigned into the COG category of general function prediction, which represented the largest functional group of the 23 COG categories, followed by carbohydrate transport and metabolism (74, 9.24%), posttranslational modification, protein turnover, chaperones (67, 8.36%), replication, recombination and repair (65, 8.11%) and amino acid transport and metabolism (58, 7.24%) (Figure S2D).

Clustering and functional enrichment of DEGs in all treatments

Among them, 13 DEGs were detected in all treatments, including 7 up-regulated and 6 down-regulated. Some of the DEGs, which were involved in response to stress (GO:0006950) and stimulus (GO:0050896), as well as response to abiotic stress (GO:0009628), such as freezing (GO:0050826), cold (GO:0009409) and salt (GO:0009651) stress. Some of the DEGs response to growth hormone (GO:0060416) and water deprivation (GO:0009414) (Table 1S). In addition, we performed cluster analysis on all screened differentially expressed genes (Figure 3C; Figure 4C; Figure 5C; Figure 6C). Nearly all differentially expressed genes are up-regulated between *BcL.1-25* (N-25) and *BcL.1-4* (N-4) groups; between *BcL.2-25* (B-25) and *BcL.2-4* (B-4) groups, most of the genes were up-regulated, while the *BcL.2-25* (B-25) group showed more up-regulated genes; between *BcL.1-25* (N-25) and *BcL.2-25* (B-25) groups, the performance of up-regulated genes and down-regulated genes, which is very similar to that between *BcL.1-4* (N-4) and *BcL.2-4* (B-4) groups (Figure S3).

qRT-PCR validation of the candidate DEGs responsive to cold tolarence

To test the reliability of the transcriptome sequencing results, qRT-PCR analysis, which were used to perform. In this study, 13 common candidate DEGs, which were selected and detected in all treatments by qRT-PCR analysis (Figure 7). The results of transcriptome sequencing, which were compared with the results of qRT-PCR experiments. Our results showed that even if the fold changes in the expression levels of certain genes detected by transcriptome sequencing and qRT-PCR analysis did not match, but almost all expression levels analyzed by qRT-PCR, which were highly consistent with the transcriptome sequencing results. These results, which also confirmed the reliability of transcriptome sequencing data (Figure 7). Through qRT-PCR analysis, it was found that there was only one down-regulated gene (*Brassica_rapa_*newGene_1153), and its expression level was different from the RNA-Seq data (Figure 7).

Overview of small RNA sequencing data

In this study, these samples included C1 (*BcL.1-25*), C2 (*BcL.1-4*), C3 (*BcL.2-25*) and C4 (*BcL.2-4*) were collected, sequenced and analyzed. 34,182,333 total reads were generated, and 8,836,042 unique reads were isolated. After removing low-quality reads, the length distribution of the small RNAs (18-35 nt) revealed that a length of 24 nt, which was the most abundant class among both clean and unique reads in all groups (Figure 8; Table 2).

Analysis of known miRNA

To obtain the details of the miRNA matched on each sample, the above-mentioned reads, which mapped to the reference sequence are compared with the specified range of sequences in miRBase, including the secondary structure of the known miRNAs on the match. Information on the sequence, length, and number of occurrences of the miRNA in the present invention. When the miRNA develops into a mature body from the precursor, the process is completed by dicer digestion. The specificity of the cleavage site makes the miRNA mature sequence sequence the first base. There is a strong bias, so the first base distribution of miRNAs of different lengths is also carried out, in addition to the base distribution statistics of the miRNAs. As shown in Table 3, the number of miRNAs in the C4 group, which was the highest, 338,668; the number of miRNAs in the C1 group, which was the least, 129,932. However, the C3 group had the largest number of miRNAs in the 1,774 species; the C1 group had the least miRNA species, which was 1,244 species. We listed the secondary structure of the 10 known miRNAs on the match (Figure 9).

Predicted new miRNA

The signature hairpin structure of miRNA precursors, which can be used to predict new miRNAs. As shown in Table 4, the number of miRNAs was the highest in the C4 group, 103,663; the number of miRNAs was the least in the C1 group, 60,344. However, the C3 group had the largest number of miRNAs in the 2,985 species. The C1 group had the least miRNA species, which was 2,318 species. We listed the secondary structure of the 10 predicted new miRNA on the match (Figure 10).

Screening and identification of differential miRNAs

To test the reliability of the experimental results and the rationality of sample selection, the correlation analysis of gene expression levels between samples, which was carried out. R^2: The square of the pearson correlation coefficient is basically at 0.772-1 (Figure 11C), indicating that the similarity of expression patterns between samples is higher. By comparing the TPM density distribution of miRNA under different experimental conditions, so that the TPM distribution between different experimental conditions, which can be checked as a whole (Figure 11A; Figure 11B). Finally, by using volcano plots, to infer the overall distribution of differential miRNAs. Differential miRNAs, which were screened based on fold changes in levels and corrected significance levels (padj / qvalue). In the C2 and C1 groups, 20 differentially miRNAs, which were up-regulated and 16 differentially miRNAs were down-regulated; in the C4 and C3 groups, 31 differentially miRNA, whichs were up-regulated and 47 differentially miRNAs were down-regulated; in the C3 and C1 groups, 44 differentially miRNAs, which were up-regulated and 33 differentially miRNAs were down-regulated; in the C4 and C2 groups, 37 differentially miRNAs, which were up-regulated and 47 differentially miRNAs were down-regulated (Figure 12).

Cluster analysis of differential miRNAs

The clustering pattern of differential miRNA expression under different experimental conditions, which was determined by using differential miRNA cluster analysis. For each comparison combination, a set of differential miRNAs, which is obtained and will be used for hierarchical clustering analysis. The number

of miRNAs with high expression levels of C1, C2, and C3 was higher than that of the C4 group. In addition, the number of highly expressed miRNAs, which was highest in the C2 group, while C4 was the smallest. Some miRNAs in the C4 group have lower expression levels in other groups, but the highest expression in the C4 group, such as bra-miR9557-3p, bra-miR9557-5p, and so on (Figure 13).

Venn diagram of differential miRNAs

In order to more intuitively show the common and unique differences of each comparison combination. When the number of miRNAs, which is greater than or equal to 2 and less than or equal to 5, the number of differential miRNAs, which obtained by comparison in each group can be counted and plotted as a venn diagram (Figure 14). In Figure 14A, there are 17 common differential miRNAs; while in Figure 14B, there are 38 common differential miRNAs. In all combinations, 8 common differential miRNAs are shown in Figure 14C (Table 5; Table S18).

Enrichment analysis of differential miRNA candidate target genes

After obtaining the differentially expressed miRNAs between the groups, according to the correspondence between the miRNA and its target genes, we performed GO and KEGG enrichment analysis on the set of target genes of each group of differentially expressed miRNAs.

GO enrichment stratification analysis, between the C2 and C1 groups, single-organism cellular process (GO:0044763, 896 genes), membrane (GO:0016020, 747 genes), and protein binding (GO:0005515, 994 genes) were the most representative GO terms in biological process, cellular component, and molecular function, respectively (Figure 15A; Table S19); between the C4 and C3 groups, single-organism cellular process (GO:0044763, 1,748 genes), membrane (GO:0016020, 1,431 genes), and protein binding (GO:0005515, 1,973 genes) were the most representative GO terms (Figure S4A; Table S21); between the C3 and C1 groups, single-organism cellular process (GO:0044763, 1,841 genes), membrane (GO:0016020, 1,511 genes), and protein binding (GO:0005515, 2,041 genes), which were the most representative GO terms (Figure S5A; Table S23); between the C4 and C2 groups, single-organism cellular process (GO:0044763, 1,880 genes), membrane (GO:0016020, 1,515 genes), and ion binding (GO:0005515, 2,198 genes) were the most representative GO terms (Figure S6A; Table S25).

Through KEGG pathway enrichment analysis, between the C2 and C1 groups, the starch and sucrose metabolism pathway, which was the largest, 42 genes, followed by the plant-pathogen interaction pathway, 36 genes (Figure 15B; Table S20); between the C4 and C3 groups, the starch and sucrose metabolism pathway, which was the largest, 74 genes, followed by the plant-pathogen interaction pathway, 67 genes (Figure S4B; Table S22); between the C3 and C1 groups, the plant-pathogen interaction pathway, 68 genes, RNA transport pathway, also containing 68 genes, they are both the pathway with the most genes (Figure S5B; Table S24); between the C4 and C2 groups, the plant-pathogen interaction pathway, which was the largest, 86 genes, followed by the starch and sucrose metabolism pathway, 78 genes (Figure S6B; Table S26).

Discussion

As a convenient tool for transcriptome analysis, RNA-Seq, has been increasingly targeted at species that do not have access to genomic sequences [53-55]. Here, we used RNA-Seq to measure gene expression levels in non-heading Chinese cabbage different varieties and under low temperature treatment. To obtain all the DEGs from RNA-Seq data, the expression of all genes was analyzed depending on the RPKM. By comparing the datas from each groups, we found that between the *BcL.1-25* and *BcL.1-4* groups, the most differentially expressed genes (6,208 DEGs), which were enriched, and the number was almost 6 times that of the other groups, and the difference is the largest. At the same time, the ribosome pathway, which is worth mentioning, the most abundant genes (287 genes), which are also shown, far more than other groups. The results showed that cold stress could affect the genes involved in the expression of these pathways, previous reports have also detected these rich pathways, partially reflecting the credibility of our results [56, 57]. This also suggests that the ribosomal pathway, which might be involved in cold stress.

RNA-Seq searched for some low temperature-related differentially expressed genes. qRT-PCR, a practical method, often used to identify gene expression levels [58]. To confirm the reliability of the RNA-Seq results, we performed qRT-PCR experiment to identify them. According to qRT-PCR results, the expression patterns of all unigenes, which were consistent with the transcriptome sequencing data, showed that our experimental results were reliable.

small RNAs are short, non-coding RNAs, which usually 19-25 nt in length, and two protruding sizes of 21 and 24 nt, respectively [59]. In general, miRNA corresponds to the 21 nt class of small molecule RNA, research have also found that small RNAs shown a wide range of functions, including heterochromatin formation, gene silencing, and DNA methylation [60, 61]. By small RNA sequencing, the result shows that 24 nt length, which is the most abundant category, among pure and unique reads in all groups (Figure 8). This result was highly consistent with previous studies on *A. thaliana* [62], *Oriza sativa* [63], *Medicago truncatula* [64], and *Populus trichocarpa* [65].

Further, the known miRNAs are analyzed, and new miRNAs are predicted. We discover that these selected known miRNAs, which have at least two stem-loop structures that are obtained by self-folding (Figure 9A-9J), suggesting that there might be protein or chromosomal binding sites. Meanwhile, these predicted miRNAs, which have at least two stem-loop structures that are obtained by self-folding (Figure 10A-10J), suggesting that there might also be protein or chromosomal binding sites. The functions of these miRNAs and their target genes, which were comprehensively analyzed, might provide new insights into miRNA-mediated epigenetic control of Pak-choi under low temperature.

To better understand the functional roles of these predicted miRNA targets, the target genes, which were functionally annotated in biological processes, and the KEGG pathway, which was also used to describe the corresponding metabolic pathways. After low temperature treatment, most target genes, which were enriched in the metabolic pathways of starch and sucrose, which was basically consistent with the results of the above-mentioned transcriptome study. This indicated that a series of biosynthetic and

metabolic pathways, which might be induced, after low temperature treatment [66]. Here, a combination of transcriptome and small RNA sequencing, which was used to analyze the cold tolerance of non-heading Chinese cabbage. This study might promote further molecular regulation mechanisms of cold tolerance in Pak-choi.

Conclusions

In this study, a total of 63.43 Gb clean data, which was obtained from the transcriptome analysis. Based on the comparison results, a total of 1,860 new genes were discovered, 13 common DEGs, which were detected in all treatments, including 7 up-regulated and 6 down-regulated. Some of the DEGs, which were involved in response to stress, especially in terms of abiotic stress, such as freezing, cold and salt stresses. Meanwhile, we used qRT-PCR experiment to confirm changes in the expression levels of these genes. Further, we performed miRNA sequencing analysis on the same material. We found that the results revealed a total of 34,182,333 small RNA reads, a total of 88,604,604 kinds of small RNA, among which, the most common size was 24 nt. In all materials, the number of common differential miRNAs is 8. The number of known mature miRNAs and the number of precursors, which were 110 and 80, respectively; the number of predicted novel miRNA matures and the number of precursors, which were 75 and 84, respectively. According to GO and KEGG enrichment analysis, single-organism cellular process in the biological process, membrane in the cellular component, and protein binding in the molecular function were almost all the most representative level GO terms in all data sets; while the starch and sucrose metabolism pathway was almost all the largest, followed by the plant-pathogen interaction pathway. Additionally, our findings highlight the significance of cold signaling in Pak-choi, and might provide a foundation for subsequent research under abiotic stress in the future.

Methods

Plant growth environment and treatment conditions

The materials, which used in this study are non-heading Chinese cabbage varieties, *Suzhouqing* (*BcL.1*) and *Sijiucaixin* (*BcL.2*), provided by Nanjing Agricultural University, Cabbage System Biology Laboratory. Non-heading Chinese cabbage varieties, *Suzhouqing* (*BcL.1*) and *Sijiucaixin* (*BcL.2*), numbered: NHCC001 and NHCC008, separately, which are laboratory-specific and commonly used varieties, stored in an open herbarium in the laboratory. Seedlings, which were transferred into plastic pots containing a mixture of soil and vermiculite (volume ratio is 3 : 1), and cultured in a growth chamber under 16 h light (22 °C)/8 h dark (18 °C). After growing for about 30 days, the plants, which were grown in light incubators at 4 °C (*BcL.1-4*, *BcL.2-4*) and 25 °C (*BcL.1-25*, *BcL.2-25*) for up to 6 hours, respectively. The leaves of plants, which were harvested, immediately frozen in liquid nitrogen, and stored at -80°C for experimental used.

Transcriptome sequencing

Following the manufacturer's instructions to extract total RNA from the samples, by using RNAiso Plus reagent (Takara Bio, Dalian, China). RNA samples, which were examined, by using a spectrophotometer and electrophoresed on a 1% agarose gel. cDNA libraries construction and transcriptome sequencing, which were performed by the Biomarker Technologies (Beijing, China). The using Illumina HiSeq X Ten sequencing platform and Pe150 mode sequencing, all clean reads, which were subsequently mapped to the *Brassica rapa* reference genome sequence (*IVFCAASv1*) (http://brassicadb.org). The clean reads of each sample, which were sequence aligned with the designated reference genome. Gene expression analysis, which was performed, based on the comparison results; differentially expressed genes, which were identified, based on their expression levels in different samples, and their functional annotation and enrichment analysis were aslo performed [67].

Transcriptome assembly and functional annotation

The raw data of the transcriptome sequencing, which were purified by trimming adapters, removing reads containing poly-N, and rejecting the low-quality data (*quality value* $\leq$ 10 or unknown nucleotides larger than 5%) to get the clean reads. Meanwhile, the proportion of nucleotides with quality values greater than 30 (Q30) and GC content of the clean data were calculated. Then, all of the clean reads were assembled, using Trinity program [68]. Firstly, the certain short reads with overlap regions, which were assembled into longer contiguous sequences for each library. Then, the distance of different contigs, which was recognized, mapping the clean reads, based on the paired-end information, to obtain the sequence of the transcripts. Finally, the unigenes were obtained, performing the sequence of potential transcript to TGI Clustering tool [69]. The new genes discovered, which were performed with NR [70], Swiss-Prot [71], GO [72], COG [73], KOG [74], Pfam [75], and KEGG [76] databases, using BLAST [77] software. Using KOBAS2.0 [78] to obtain the KEGG orthology result of the new gene for sequence alignment. After predicting the amino acid sequence of the new gene, using HMMER [79] software to align with the Pfam database, in order to obtain the annotation information of the new gene.

Differentially expressed genes analysis

Following fragments per kilobase of exon per million fragments mapped reads (FPKM) method, the expression level of unigene, which was calculated [80]. The ratio of the FPKM values (using 0.001 instead of 0 if the FPKM was 0), which were taken as the fold-changes in the expression of each gene to identify DEGs between each groups. The false discovery rate (FDR) control method, which was used to identify the threshold of the *p-value* in multiple tests, in order to compute the significance of the difference in transcript abundance [81]. In this result, only fold change with |$\log_2$ (case_FPKM/control_FPKM)| $\geq$ 1, and an FDR $\leq$ 0.001, which were taken as the threshold for significantly differential expression. The $\log_2$-transformed FPKM value for DEGs was applied to generate heat map by MeV 4.7 [82]. Meanwhile, the DEGs, which were annotated with GO and KEGG databases.

Validation of DEGs with qRT-PCR

qRT-PCR, which was used to confirm the expression of common differentially expressed genes in Pak-choi. Total RNA was extracted from each sample. The first-strand cDNA, which was synthesized, using a PrimeScript™ II First Strand cDNA synthesis kit (Takara Bio, Dalian, China), according to the manufacturer's protocol. The primers were designed, using the software Premier 5.0, and listed in Table 1. The quantified expression levels of the tested genes were normalized against the housekeeping genes Cyclophilin 1 (CYP1) [83]. The qRT-PCR assays, which were performed with three biological, and technical replicates. Each reaction, which was performed in 20 μL reaction mixtures containing a diluted cDNA sample as template, SYBR Premix Ex Taq (2×) (TaKaRa, Kyoto, Japan) and gene-specific primers. Conditions for quantitative analysis were as follows: 95 °C for 3 min, then 40 cycles (95 °C 30 s, 60 °C 30 s), and 72 °C for 30 s. qRT-PCR, which was performed according to previous report [84]. The comparative Ct value method, which was adopted to analyze the relative gene expression, according to a previous analysis, RNA expression levels relative to actin gene, which were calculated as $2^{-\Delta\Delta CT}$ [84, 85].

small RNA sequencing

High-throughput sequencing (such as Illumina HiSeqTM2500 / MiSeq and other sequencing platforms) sequenced raw image data files, which are converted into sequenced reads by base calling analysis, we call them raw data or raw reads. As a result, it is stored in FASTQ (abbreviated as fq) file format, which contains the sequence information of read sequences and its corresponding sequencing quality information. Referring to the standard definition of miRNA [86, 87], the candidate target gene of miRNA, which was compared as the query sequence with the *Brassica rapa* database (http://brassicadb.org/brad/). The control samples and inoculation samples, which were mixed for small RNA libraries construction, respectively. According to the reported procedures, the construction of small RNA libraries, which were completed [88]. Three micrograms of total RNA per sample, which was used as input material for the small RNA library [89]. small RNA library construction and small RNA deep sequencing, which were proceeded following the detailed protocol provided by the genome sequencing company (Novogene, China).

Bioinformatic analysis of sequence data

Raw data, which were first processed through custom Perl and Python scripts to obtain clean data. The clean data, which were mapped to the reference sequence in miRBase21.0 by Bowtie [90], without mismatch to look for known miRNAs. Then, the other reads, which were integrated to predict novel miRNAs using the available miREvo [91] and miRDeep2 [92] software. The miRNA counts as well as base bias, which were identified by using custom scripts. Then, the miFam.dat (http://www.mirbase.org/ftp.shtml) was used to look for families of known miRNAs. The novel miRNA precursor was submitted to Rfam (http://rfam.sanger.ac.uk/search/) to look for Rfam families.

Venn diagrams of known miRNAs and novel miRNAs

Normalization formula (Normalized expression = Mapped read count/Total reads*1,000,000), which was used to estimated miRNA expression levels [93]. DEG seq R package, which was used to analyze the

differential expression of two samples with the criterion of Q < 0.01 and |log$_2$(fold-change)| > 1 [94].

Construction of degradome libraries

Target genes of candidate miRNAs, which were verified by degradome sequencing by using total RNA same to the RNA used for small RNA sequencing library construction, following the published parallel analysis of RNA Ends (PARE) protocol [95]. The data analysis was processed, following the procedure instructions (Novogene, China).

Target gene prediction and annotation for known and novel miRNAs

The psRobot_tarin psRobot, which was performed to predict target genes of miRNA [96]. To further explore the detailed molecular mechanism of miRNAs in Pak-choi response to cold stress, the target transcripts of differentially expressed miRNAs were analyzed by GO and KEGG functional snnotation suites. Subsequently, Revigo tool (http://revigo.irb.hr), which was implemented for enrichment analysis of the target genes. The KOBAS software, which was used to test the statistical enrichment of the target gene candidates in the KEGG pathways [97].

Abbreviations

COG: Cluster of Orthologus Groups; CYP1: Cyclophilin 1; DEGs: Differentially expressed genes; FC: Fold change; FDR: False discovery rate; GO: Gene ontology; hc-siRNA: heterochromatic short interfering RNA; KEGG: Kyoto Encyclopedia of Genes and Genomes; miRNA: microRNA; nt: nucleotides; qRT-PCR: quantitative real-time polymerase chain reaction; RNA-Seq: RNA sequencing; ta-siRNA: trans short interfering RNA

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets supporting the conclusions of this manuscript are included within the article and its additional files.

Competing interest

The authors declare no any competing interest.

Authors' contributions

JW completed the relevant experiments and wrote the paper. QZ, XY and XH modified and approved the manuscript. XY and XH interpreted the results and coordinated research. All authors have read and finalized the draft, agreeing that the paper will be submitted to this prestigious journal.

Authors' Information

[1] State Key Laboratory of Crop Genetics and Germplasm Enhancement / Key Laboratory of Biology and Germplasm Enhancement of Horticultural Crops in East China, Ministry of Agriculture / Engineering Research Center of Germplasm Enhancement and Utilization of Horticultural Crops, Ministry of Education, Nanjing Agricultural University, Nanjing 210095, China; 2017204023@njau.edu.cn (JW); 2017204027@njau.edu.cn (QZ); youx@njau.edu.cn (XY)

[2] College of Sciences, Nanjing Agricultural University, Nanjing 210095, China

[*] Corresponding author: hxl@njau.edu.cn (XH); Tel.: +86-025-8439-5917.

References

1. Yu S, Zhang F, Wang X, Zhao X, Zhang D, Yu Y, Xu J: Genetic diversity and marker-trait associations in a collection of Pak-choi (*Brassica rapa* L. ssp. *chinensis Makino*) Accessions. Genes & Genomics 2010, 32(5):419-428.

2. Li C, Du C, Zeng Y, Ma F, Shen Y, Xing Z, Zhou J: Two-dimensional visualization of nitrogen distribution in leaves of chinese cabbage (*Brassica rapa* subsp. *chinensis*) by the fourier transform infrared photoacoustic spectroscopy technique. Journal of agricultural and food chemistry 2016, 64(41):7696-7701.

3. Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K: *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. The Plant Cell 2003,

15(1):63-78.

4. Chinnusamy V, Zhu J, Zhu J-K: Cold stress regulation of gene expression in plants. Trends in plant science 2007, 12(10):444-451.

5. Krasensky J, Jonak C: Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. Journal of experimental botany 2012, 63(4):1593-1608.

6. Salinas J: Molecular mechanisms of signal transduction in cold acclimation. Plant Signal Transduction 2002, 38:116.

7. Kaplan F, Kopka J, Haskell DW, Zhao W, Schiller KC, Gatzke N, Sung DY, Guy CL: Exploring the temperature-stress metabolome of *Arabidopsis*. Plant physiology 2004, 136(4):4159-4168.

8. Kong X, Pan J, Zhang M, Xing X, Zhou Y, Liu Y, Li D, Li D: ZmMKK4, a novel group C mitogen-activated protein kinase kinase in maize (*Zea mays*), confers salt and cold tolerance in transgenic *Arabidopsis*. Plant, cell & environment 2011, 34(8):1291-1303.

9. Sobkowiak A, Jończyk M, Jarochowska E, Biecek P, Trzcinska-Danielewicz J, Leipner J, Fronk J, Sowiński P: Genome-wide transcriptomic analysis of response to low temperature reveals candidate genes determining divergent cold-sensitivity of maize inbred lines. Plant molecular biology 2014, 85(3):317-331.

10. Wang F, Qiu N, Ding Q, Li J, Zhang Y, Li H, Gao J: Genome-wide identification and analysis of the growth-regulating factor family in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). BMC genomics 2014, 15(1):807.

11. Hussain S, Khan F, Cao W, Wu L, Geng M: Seed priming alters the production and detoxification of reactive oxygen intermediates in rice seedlings grown under sub-optimal temperature and nutrient supply. Frontiers in plant science 2016, 7:439.

12. Wang W, Chen Q, Hussain S, Mei J, Dong H, Peng S, Huang J, Cui K, Nie L: Pre-sowing seed treatments in direct-seeded early rice: consequences for emergence, seedling growth and associated metabolic events under chilling stress. Scientific reports 2016, 6:19637.

13. Zhao J, Li S, Jiang T, Liu Z, Zhang W, Jian G, Qi F: Chilling stress—the key predisposing factor for causing Alternaria alternata infection and leading to cotton (*Gossypium hirsutum* L.) leaf senescence. PLoS One 2012, 7(4):e36126.

14. Starck Z, Niemyska B, Bogdan J, Tawalbeh RA: Response of tomato plants to chilling stress in association with nutrient or phosphorus starvation. Plant and soil 2000, 226(1):99-106.

15. Shinozaki K, Yamaguchi-Shinozaki K: Molecular responses to drought and cold stress. Current Opinion in Biotechnology 1996, 7(2):161-167.

16. Wang W, Vinocur B, Shoseyov O, Altman A: Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends in plant science 2004, 9(5):244-252.

17. Thakur P, Kumar S, Malik JA, Berger JD, Nayyar H: Cold stress effects on reproductive development in grain crops: an overview. Environmental and Experimental Botany 2010, 67(3):429-443.

18. Anjum SA, Xie X-y, Wang L-c, Saleem MF, Man C, Lei W: Morphological, physiological and biochemical responses of plants to drought stress. African journal of agricultural research 2011, 6(9):2026-2032.

19. Zhu J-J, Li Y-R, Liao J-X: Involvement of anthocyanins in the resistance to chilling-induced oxidative stress in *Saccharum officinarum* L. leaves. Plant physiology and biochemistry 2013, 73:427-433.

20. Martin JA, Wang Z: Next-generation transcriptome assembly. Nature Reviews Genetics 2011, 12(10):671.

21. Ward JA, Ponnala L, Weber CA: Strategies for transcriptome analysis in nonmodel plants. American Journal of Botany 2012, 99(2):267-276.

22. Hamilton JP, Robin Buell C: Advances in plant genome sequencing. The Plant Journal 2012, 70(1):177-190.

23. Dixon RA: Natural products and plant disease resistance. Nature 2001, 411(6839):843.

24. Goossens A, Häkkinen ST, Laakso I, Seppänen-Laakso T, Biondi S, De Sutter V, Lammertyn F, Nuutila AM, Söderlund H, Zabeau M: A functional genomics approach toward the understanding of secondary metabolism in plant cells. Proceedings of the National Academy of Sciences 2003, 100(14):8595-8600.

25. Kalra S, Puniya BL, Kulshreshtha D, Kumar S, Kaur J, Ramachandran S, Singh K: De novo transcriptome sequencing reveals important molecular networks and metabolic pathways of the plant, Chlorophytum borivilianum. PLoS One 2013, 8(12):e83336.

26. Li Z, Cheng Y, Cui J, Zhang P, Zhao H, Hu S: Comparative transcriptome analysis reveals carbohydrate and lipid metabolism blocks in *Brassica napus* L. male sterility induced by the chemical hybridization agent monosulfuron ester sodium. BMC genomics 2015, 16(1):206.

27. Devi K, Mishra SK, Sahu J, Panda D, Modi MK, Sen P: Genome wide transcriptome profiling reveals differential gene expression in secondary metabolite pathway of Cymbopogon winterianus. Scientific reports 2016, 6:21026.

28. Zhang G, Guo G, Hu X, Zhang Y, Li Q, Li R, Zhuang R, Lu Z, He Z, Fang X: Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. Genome research 2010, 20(5):646-654.

29. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M: The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 2008, 320(5881):1344-1349.

30. Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bähler J: Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature 2008, 453(7199):1239.

31. Wang Z, Fang B, Chen J, Zhang X, Luo Z, Huang L, Chen X, Li Y: De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweetpotato (*Ipomoea batatas*). BMC genomics 2010, 11(1):726.

32. Ge G, Xiao P, Zhang Y, Yang L: The first insight into the tissue specific taxus transcriptome via Illumina second generation sequencing. PLoS one 2011, 6(6):e21220.

33. Cheng Y, Liu J, Zhang H, Wang J, Zhao Y, Geng W: Transcriptome analysis and gene expression profiling of abortive and developing ovules during fruit development in hazelnut. PloS one 2015, 10(4):e0122072.

34. Axtell MJ: Classification and comparison of small RNAs from plants. Annual review of plant biology 2013, 64:137-159.

35. Voinnet O: Origin, biogenesis, and activity of plant microRNAs. Cell 2009, 136(4):669-687.

36. Chambers C, Shuai B: Profiling microRNA expression in *Arabidopsis* pollen using microRNA array and real-time PCR. BMC plant biology 2009, 9(1):87.

37. Jones-Rhoades MW, Bartel DP, Bartel B: MicroRNAs and their regulatory roles in plants. Annu Rev Plant Biol 2006, 57:19-53.

38. Llave C, Xie Z, Kasschau KD, Carrington JC: Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science 2002, 297(5589):2053-2056.

39. Khraiwesh B, Arif MA, Seumel GI, Ossowski S, Weigel D, Reski R, Frank W: Transcriptional control of gene expression by microRNAs. Cell 2010, 140(1):111-122.

40. Hao D, Chao M, Yin Z, Yu D: Genome-wide association analysis detecting significant single nucleotide polymorphisms for chlorophyll and chlorophyll fluorescence parameters in soybean (*Glycine max*) landraces. Euphytica 2012, 186(3):919-931.

41. Szittya G, Burgyán J: RNA interference-mediated intrinsic antiviral immunity in plants. In: Intrinsic Immunity. Springer; 2013: 153-181.

42. Bartel DP: MicroRNAs: genomics, biogenesis, mechanism, and function. cell 2004, 116(2):281-297.

43. Jones-Rhoades MW, Bartel DP: Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Molecular cell 2004, 14(6):787-799.

44. Mallory AC, Vaucheret H: Functions of microRNAs and related small RNAs in plants. Nature genetics 2006, 38(6s):S31.

45. Sunkar R, Kapoor A, Zhu J-K: Posttranscriptional induction of two Cu/Zn superoxide dismutase genes in *Arabidopsis* is mediated by downregulation of miR398 and important for oxidative stress tolerance. The Plant Cell 2006, 18(8):2051-2065.

46. Wu S, Huang S, Ding J, Zhao Y, Liang L, Liu T, Zhan R, He X: Multiple microRNAs modulate p21Cip1/Waf1 expression by directly targeting its 3' untranslated region. Oncogene 2010, 29(15):2302.

47. Kang HM, Saltveit ME: Chilling tolerance of maize, cucumber and rice seedling leaves and roots are differentially affected by salicylic acid. Physiologia Plantarum 2002, 115(4):571-576.

48. Cruz RPd, Milach SCK: Cold tolerance at the germination stage of rice: methods of evaluation and characterization of genotypes. Scientia Agricola 2004, 61(1):1-8.

49. Oliver SN, Dennis ES, Dolferus R: ABA regulates apoplastic sugar transport and is a potential signal for cold-induced pollen sterility in rice. Plant and Cell physiology 2007, 48(9):1319-1330.

50. Ruelland E, Vaultier M-N, Zachowski A, Hurry V: Cold signalling and cold acclimation in plants. Advances in botanical research 2009, 49:35-150.

51. Gao J, Luo M, Zhu Y, He Y, Wang Q, Zhang C: Transcriptome sequencing and differential gene expression analysis in Viola yedoensis Makino (Fam. Violaceae) responsive to cadmium (Cd) pollution. Biochemical and biophysical research communications 2015, 459(1):60-65.

52. Kanehisa M, Goto S: KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research 2000, 28(1):27-30.

53. Li X, Luo J, Yan T, Xiang L, Jin F, Qin D, Sun C, Xie M: Deep sequencing-based analysis of the Cymbidium ensifolium floral transcriptome. PloS one 2013, 8(12):e85480.

54. Liu Z, Ma L, Nan Z, Wang Y: Comparative transcriptional profiling provides insights into the evolution and development of the zygomorphic flower of Vicia sativa (*Papilionoideae*). PLoS One 2013, 8(2):e57338.

55. Soetaert SS, Van Neste CM, Vandewoestyne ML, Head SR, Goossens A, Van Nieuwerburgh FC, Deforce DL: Differential transcriptome analysis of glandular and filamentous trichomes in *Artemisia annua*. BMC plant biology 2013, 13(1):220.

56. Kreps JA, Wu Y, Chang H-S, Zhu T, Wang X, Harper JF: Transcriptome changes for *Arabidopsis* in response to salt, osmotic, and cold stress. Plant physiology 2002, 130(4):2129-2141.

57. Pang T, Ye C-Y, Xia X, Yin W: De novo sequencing and transcriptome analysis of the desert shrub, Ammopiptanthus mongolicus, during cold acclimation using Illumina/Solexa. BMC genomics 2013, 14(1):488.

58. Taylor S, Wakem M, Dijkman G, Alsarraj M, Nguyen M: A practical approach to RT-qPCR—publishing data that conform to the MIQE guidelines. Methods 2010, 50(4):S1-S5.

59. Kim B, Yu H-J, Park S-G, Shin JY, Oh M, Kim N, Mun J-H: Identification and profiling of novel microRNAs in the *Brassica rapa* genome based on small RNA deep sequencing. BMC plant biology 2012, 12(1):218.

60. Penterman J, Zilberman D, Huh JH, Ballinger T, Henikoff S, Fischer RL: DNA demethylation in the *Arabidopsis* genome. Proceedings of the National Academy of Sciences 2007, 104(16):6752-6757.

61. Lippman Z, Martienssen R: The role of RNA interference in heterochromatic silencing. Nature 2004, 431(7006):364.

62. Rajagopalan R, Vaucheret H, Trejo J, Bartel DP: A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. Genes & development 2006, 20(24):3407-3425.

63. Zhu Y-G, Sun G-X, Lei M, Teng M, Liu Y-X, Chen N-C, Wang L-H, Carey A, Deacon C, Raab A: High percentage inorganic arsenic content of mining impacted and nonimpacted Chinese rice. Environmental Science & Technology 2008, 42(13):5008-5013.

64. Szittya G, Moxon S, Santos DM, Jing R, Fevereiro MP, Moulton V, Dalmay T: High-throughput sequencing of Medicago truncatula short RNAs identifies eight new miRNA families. BMC genomics 2008, 9(1):593.

65. Puzey JR, Karger A, Axtell M, Kramer EM: Deep annotation of Populus trichocarpa microRNAs from diverse tissue sets. PloS one 2012, 7(3):e33034.

66. Hu Z, Fan J, Xie Y, Amombo E, Liu A, Gitau MM, Khaldun A, Chen L, Fu J: Comparative photosynthetic and metabolic analyses reveal mechanism of improved cold stress tolerance in bermudagrass by exogenous melatonin. Plant Physiology and Biochemistry 2016, 100:94-104.

67. Wang R, Xu S, Wang N, Xia B, Jiang Y, Wang R: Transcriptome analysis of secondary metabolism pathway, transcription factors, and transporters in response to methyl jasmonate in *Lycoris aurea*. Frontiers in plant science 2017, 7:1971.

68. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q: Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology 2011, 29(7):644.

69. Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B: TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 2003, 19(5):651-652.

70. Deng Y, Li J, Wu S, Zhu Y, Chen Y, He F: Integrated nr database in protein annotation system and its localization. Comput Eng 2006, 32(5):71-72.

71. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M: UniProt: the universal protein knowledgebase. Nucleic acids research 2004, 32(suppl_1):D115-D119.

72. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT: Gene ontology: tool for the unification of biology. Nature genetics 2000, 25(1):25.

73. Tatusov RL, Galperin MY, Natale DA, Koonin EV: The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic acids research 2000, 28(1):33-36.

74. Koonin EV, Fedorova ND, Jackson JD, Jacobs AR, Krylov DM, Makarova KS, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS: A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome biology 2004, 5(2):R7.

75. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J: Pfam: the protein families database. Nucleic acids research 2013, 42(D1):D222-D230.

76. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M: The KEGG resource for deciphering the genome. Nucleic acids research 2004, 32(suppl_1):D277-D280.

77. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research 1997, 25(17):3389-3402.

78. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li C-Y, Wei L: KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic acids research 2011, 39(suppl_2):W316-W322.

79. Eddy SR: Profile hidden Markov models. Bioinformatics (Oxford, England) 1998, 14(9):755-763.

80. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature methods 2008, 5(7):621.

81. Reiner A, Yekutieli D, Benjamini Y: Identifying differentially expressed genes using false discovery rate controlling procedures. Bioinformatics 2003, 19(3):368-375.

82. Howe EA, Sinha R, Schlauch D, Quackenbush J: RNA-Seq analysis in MeV. Bioinformatics 2011, 27(22):3209-3210.

83. Ma R, Xu S, Zhao Y, Xia B, Wang R: Selection and validation of appropriate reference genes for quantitative real-time PCR analysis of gene expression in *Lycoris aurea*. Frontiers in plant science 2016, 7:536.

84. Song X, Liu G, Huang Z, Duan W, Tan H, Li Y, Hou X: Temperature expression patterns of genes and their coexpression with LncRNAs revealed by RNA-Seq in non-heading Chinese cabbage. BMC genomics 2016, 17(1):297.

85. Pfaffl MW: A new mathematical model for relative quantification in real-time RT–PCR. Nucleic acids research 2001, 29(9):e45-e45.

86. Allen E, Xie Z, Gustafson AM, Carrington JC: microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell 2005, 121(2):207-221.

87. Schwab R, Palatnik JF, Riester M, Schommer C, Schmid M, Weigel D: Specific effects of microRNAs on the plant transcriptome. Developmental cell 2005, 8(4):517-527.

88. Sunkar R, Zhou X, Zheng Y, Zhang W, Zhu J-K: Identification of novel and candidate miRNAs in rice by high throughput sequencing. BMC plant biology 2008, 8(1):25.

89. Huang F, Wu X, Hou X, Shao S, Liu T: Vernalization can regulate flowering time through microRNA mechanism in *Brassica rapa*. Physiologia plantarum 2018, 164(2):204-215.

90. Langmead B, Trapnell C, Pop M, Salzberg SL: Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology 2009, 10(3):R25.

91. Wen H, Li J, Song T, Lu M, Kan P-Y, Lee MG, Sha B, Shi X: Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation. Journal of Biological Chemistry 2010, 285(13):9322-9326.

92. Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N: miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic acids research 2011, 40(1):37-52.

93. Zhou L, Chen J, Li Z, Li X, Hu X, Huang Y, Zhao X, Liang C, Wang Y, Sun L: Integrated profiling of microRNAs and mRNAs: microRNAs located on Xq27. 3 associate with clear cell renal cell carcinoma. PloS one 2010, 5(12):e15224.

94. Storey JD: The positive false discovery rate: a Bayesian interpretation and the q-value. The Annals of Statistics 2003, 31(6):2013-2035.

95. German MA, Luo S, Schroth G, Meyers BC, Green PJ: Construction of Parallel Analysis of RNA Ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. Nature protocols

2009, 4(3):356.

96. Wu H-J, Ma Y-K, Chen T, Wang M, Wang X-J: PsRobot: a web-based plant small RNA meta-analysis toolbox. Nucleic acids research 2012, 40(W1):W22-W28.

97. Mao X, Cai T, Olyarchuk JG, Wei L: Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. Bioinformatics 2005, 21(19):3787-3793.

Tables

TABLE 1. Primers used in the paper.

Primer name	Forward primer sequence (5'-3')
evm.TU.BraA01003362	AAACTTCCCAAATCTCAA
evm.TU.BraA03001657	AGGATGTGATAAGGTAAC
evm.TU.BraA06001385	CTAACATCATCGTTGAGTAT
evm.TU.BraA07000744	TGAAGGAGTGTTGGCATA
evm.TU.BraA09000675	CGCCGAGAATACTACCAT
evm.TU.BraA09001549	GCTTCTTCAACCATCATC
evm.TU.BraA10002276	ATTAAGGCTTACGCAATG
Brassica_rapa_newGene_1153	GGTAATAGGCGACTGGATA
evm.TU.BraA01000616	TCTTCTCCTGATGACTGT
evm.TU.BraA06001282	ATGGCAACGAATAGTGAGA
evm.TU.BraA08001363	GGAAGACTATACTATGACAATGG
evm.TU.BraA09000654	CGAGTTATCAGAGGCAATC
evm.TU.BraA10000511	CTTCCTAAGTTAGCCAATCT
qBcACTIN	GTTGCTATCCAGGCTGTTCT

TABLE 2. Type and quantity of miRNA.

Sample	Total reads	Total bases (bp)	Uniq reads	Uniq bases (bp)
C1	7404323	195732796	1708990	42044898
C2	9762950	244486883	2480460	59410849
C3	8345217	194120302	2669504	62223537
C4	8669843	207657265	1977088	47080337

Note: Sample: The sample ID; Total reads: the total number of sRNAs; Total bases (bp): The total length of the sRNA; Uniq reads: the type of sRNA; Uniq bases (bp): The total length of various sRNAs.

TABLE 3. Known miRNA alignment table for each sample.

Types	Total	C1	C2	C3	C4
Mapped mature	110	78	91	95	91
Mapped hairpin	80	73	75	77	77
Mapped uniq miRNA	6168	1244	1692	1774	1458
Mapped total miRNA	868276	129932	196191	203485	338668

TABLE 4. Statistical table of predicted new miRNA and comparison of miRNA of each sample.

Types	Total	C1	C2	C3	C4
Mapped mature	75	65	70	72	73
Mapped star	56	38	43	48	44
Mapped hairpin	84	78	80	81	81
Mapped uniq miRNA	10967	2318	2837	2985	2827
Mapped total miRNA	346364	60344	80423	101934	103663

TABLE 5. Number of combined difference (DIFF), up-regulated (UP) and down-regulated (DOWN) miRNAs in each groups.

Group	Diff	Up	Down
C2 vs C1	36	20	16
C4 vs C3	78	31	47
C3 vs C1	77	44	33
C4 vs C2	84	37	47

Supplementary Information

Additional file 1: Figure S1. Classification diagram of KEGG pathway types. **(A)** KEGG pathway types from C2 vs C1. **(B)** KEGG pathway types from C4 vs C3. **(C)** KEGG pathway types from C3 vs C1. **(D)** KEGG pathway types from C4 vs C2. X-axis, the number of genes in this pathway and their ratio to the total number of genes; Y-axis, the name of the KEGG metabolic pathway. C1: *BrL.1-25*, C2: *BrL.1-4*, C3: *BrL.2-25*, C4: *BrL.2-4*.

Additional file 2: Figure S2. COG classification of differentially expressed genes. **(A)** COG classification from C2 vs C1. **(B)** COG classification from C4 vs C3. **(C)** COG classification from C3 vs C1. **(D)** COG classification from C4 vs C2. X-axis, the specific content of each COG classification, Y-axis, the number of genes. C1: *BrL.1-25*, C2: *BrL.1-4*, C3: *BrL.2-25*, C4: *BrL.2-4*.

Additional file 3: Figure S3. Cluster analysis of differentially expressed genes. X-axis, name and clustering results of samples; Y-axis, differential gene and gene clustering results. Different columns in the figure represent different samples, and different rows represent different genes. The color represents the expression level $\log_{10}$ (FPKM + 0.000001) of the gene in the sample.

Additional file 4: Figure S4. Enrichment analysis of candidate target genes for differentially expressed miRNA between C4 and C3 groups. **(A)** Histogram of GO enrichment for candidate target genes from C4 vs C3. Three basic classifications of Go term (from left to right, BP: biological processes, CC: cellular components, MF: molecular functions). **(B)** KEGG enrichment scatter plot of candidate target genes from C4 vs C3. X-axis, the Rich factor; Y-axis, the name of the pathway. The size of the dots indicates the number of candidate target genes in this pathway, and the colors of the dots correspond to different Qvalue ranges. C1: *BrL.1-25*, C2: *BrL.1-4*, C3: *BrL.2-25*, C4: *BrL.2-4*.

Additional file 5: Figure S5. Enrichment analysis of candidate target genes for differentially expressed miRNA between C3 and C1 groups. **(A)** Histogram of GO enrichment for candidate target genes from C3 vs C1. Three basic classifications of Go term (from left to right, BP: biological processes, CC: cellular components, MF: molecular functions). **(B)** KEGG enrichment scatter plot of candidate target genes from C3 vs C1. X-axis, the Rich factor; Y-axis, the name of the pathway. The size of the dots indicates the number of candidate target genes in this pathway, and the colors of the dots correspond to different Qvalue ranges. C1: *BrL.1-25*, C2: *BrL.1-4*, C3: *BrL.2-25*, C4: *BrL.2-4*.

Additional file 6: Figure S6. Enrichment analysis of candidate target genes for differentially expressed miRNA between C4 and C2 groups. **(A)** Histogram of GO enrichment for candidate target genes from C4 vs C2. Three basic classifications of Go term (from left to right, BP: biological processes, CC: cellular components, MF: molecular functions). **(B)** KEGG enrichment scatter plot of candidate target genes from C4 vs C2. X-axis, the Rich factor; Y-axis, the name of the pathway. The size of the dots indicates the number of candidate target genes in this pathway, and the colors of the dots correspond to different Qvalue ranges. C1: *BrL.1-25*, C2: *BrL.1-4*, C3: *BrL.2-25*, C4: *BrL.2-4*.

Additional file 7: Table S1. Thirteen common differentially expressed genes (seven up-regulated and six down-regulated) in all materials.

Additional file 8: Table S2. GO enrichment analysis of differentially expressed genes between C2 and C1 groups.

Additional file 9: Table S3. KEGG enrichment analysis of differentially expressed genes between C2 and C1 groups.

Additional file 10: Table S4. GO enrichment analysis of differentially expressed genes between C4 and C3 groups.

Additional file 11: Table S5. KEGG enrichment analysis of differentially expressed genes between C4 and C3 groups.

Additional file 12: Table S6. GO enrichment analysis of differentially expressed genes between C3 and C1 groups.

Additional file 13: Table S7. KEGG enrichment analysis of differentially expressed genes between C3 and C1 groups.

Additional file 14: Table S8. GO enrichment analysis of differentially expressed genes between C4 and C2 groups.

Additional file 15: Table S9. KEGG enrichment analysis of differentially expressed genes between C4 and C2 groups.

Additional file 16: Table S10. KEGG enrichment pathway type of differentially expressed genes between C2 and C1 groups.

Additional file 17: Table S11. KEGG enrichment pathway type of differentially expressed genes between C4 and C3 groups.

Additional file 18: Table S12. KEGG enrichment pathway type of differentially expressed genes between C3 and C1 groups.

Additional file 19: Table S13. KEGG enrichment pathway type of differentially expressed genes between C4 and C2 groups.

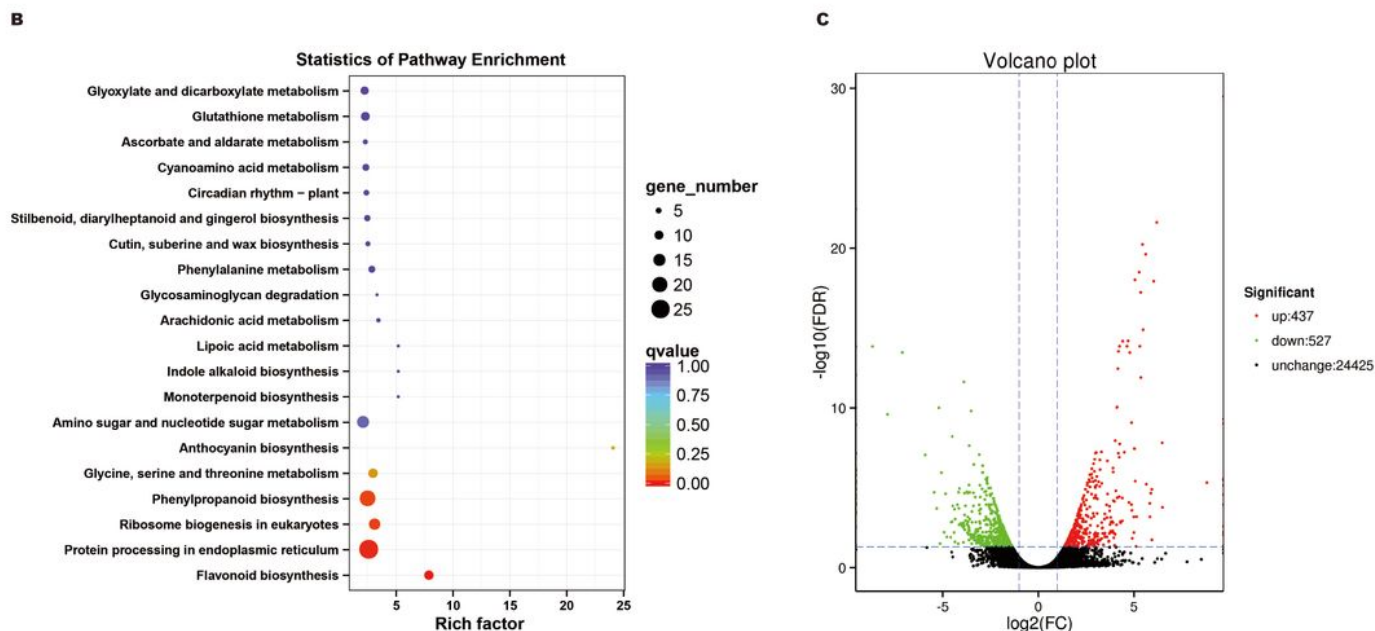**Additional file 20: Table S14.** COG classification statistics of differentially expressed genes between C2 and C1 groups.

Additional file 21: Table S15. COG classification statistics of differentially expressed genes between C4 and C3 groups.

Additional file 22: Table S16. COG classification statistics of differentially expressed genes between C3 and C1 groups.

Additional file 23: Table S17. COG classification statistics of differentially expressed genes between C4 and C2 groups.

Additional file 24: Table S18. Number of common differential miRNAs.

Additional file 25: Table S19. GO enrichment analysis of differential miRNA target genes between C2 and C1 groups.

Additional file 26: Table S20. KEGG enrichment analysis of differential miRNA target genes between C2 and C1 groups.

Additional file 27: Table S21. GO enrichment analysis of differential miRNA target genes between C4 and C3 groups.

Additional file 28: Table S22. KEGG enrichment analysis of differential miRNA target genes between C4 and C3 groups.

Additional file 29: Table S23. GO enrichment analysis of differential miRNA target genes between C3 and C1 groups.

Additional file 30: Table S24. KEGG enrichment analysis of differential miRNA target genes between C3 and C1 groups.

Additional file 31: Table S25. GO enrichment analysis of differential miRNA target genes between C4 and C2 groups.

Additional file 32: Table S26. KEGG enrichment analysis of differential miRNA target genes between C4 and C2 groups.

Figures
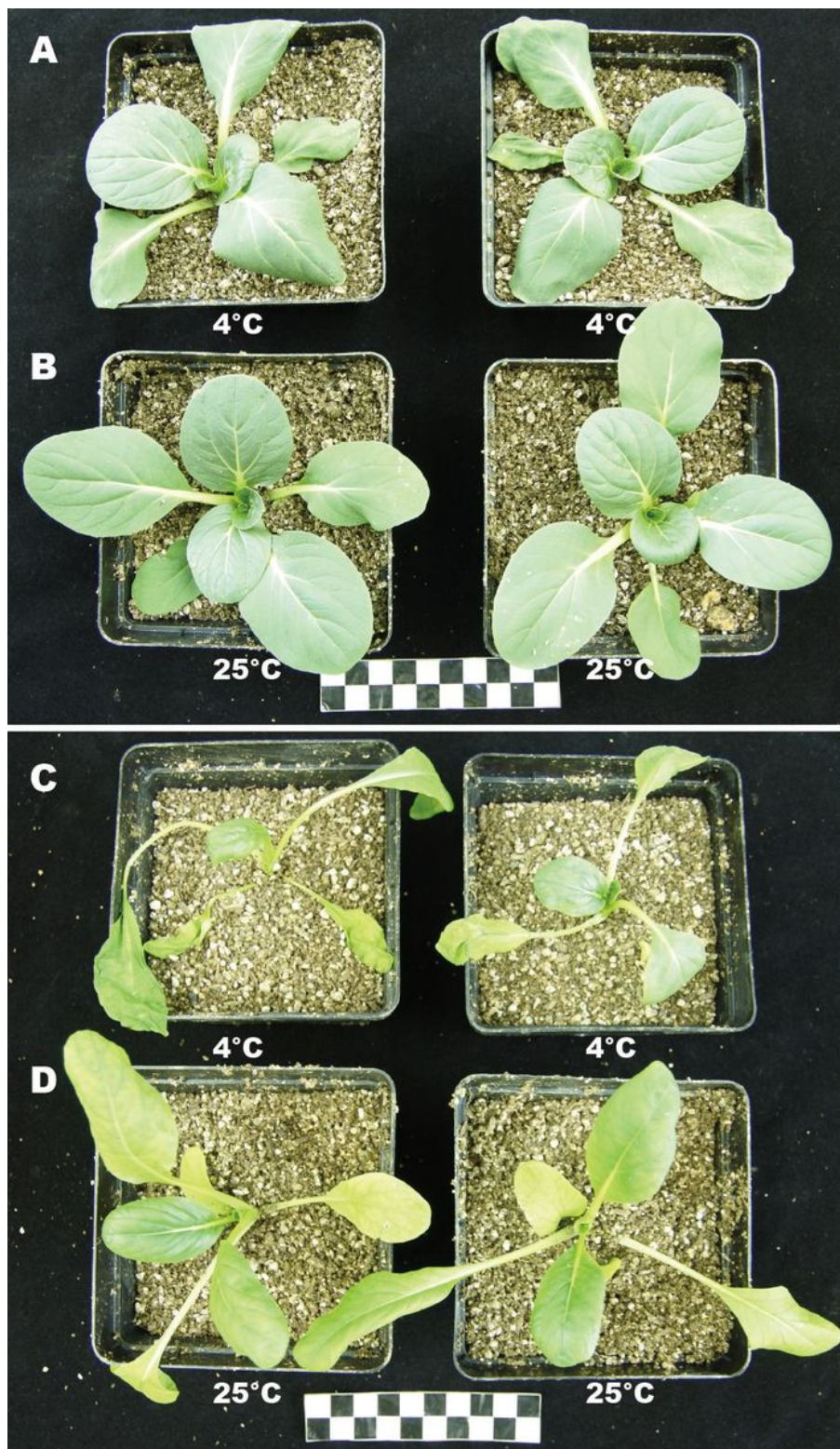
Figure 1

Phenotype of BrL.1 and BrL.2 under cold stress. Approximately 30-day-old seedlings were subjected to cold stress (4 °C for 6 hours), plants grown under normal condition (22 °C) for 6 hours were used as control. (A) BrL.1-4. (B) BrL.1-25. (C) BrL.2-4. (D) BrL.2-25.
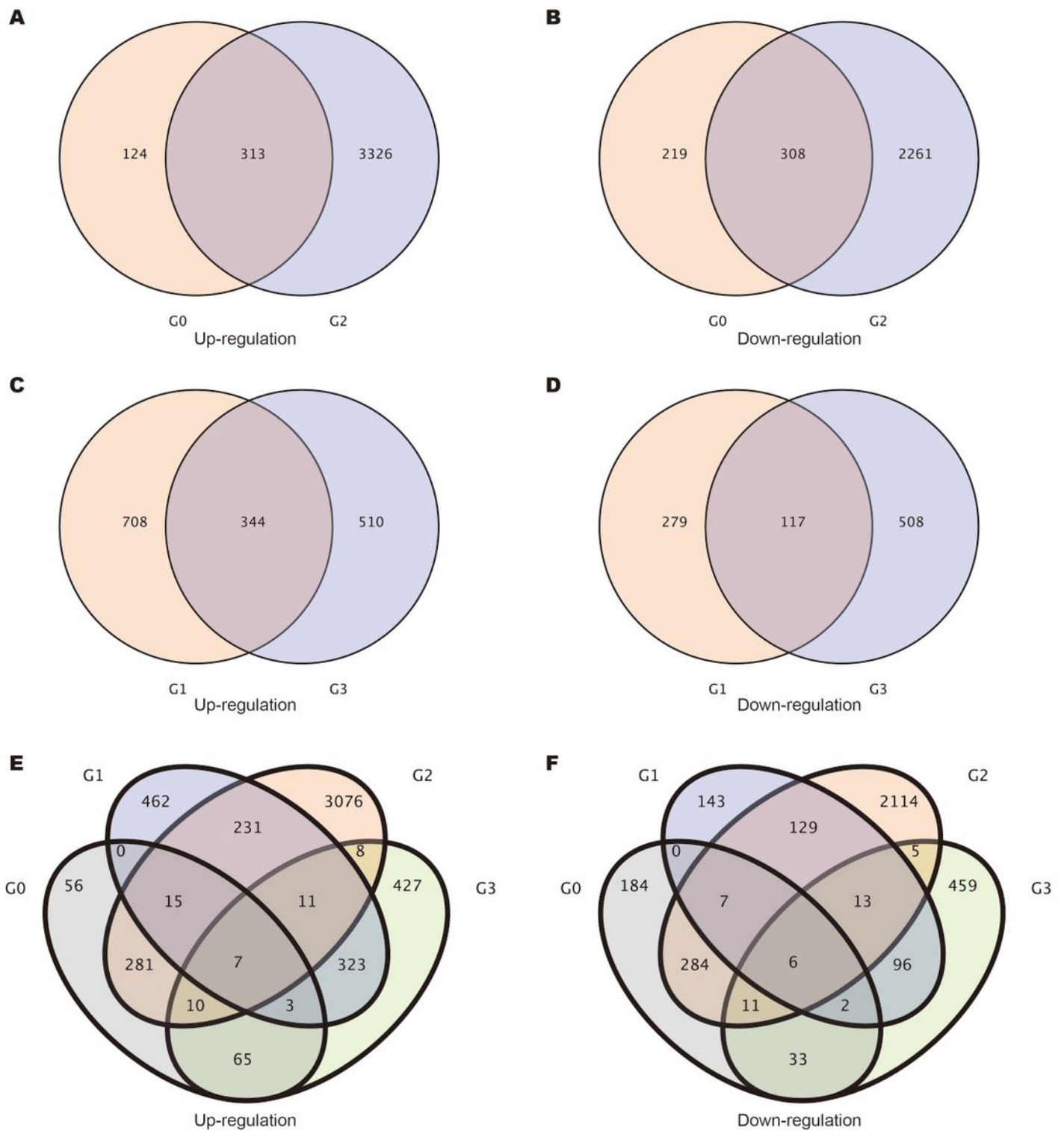
Figure 2

Statistics of the common genes and specific genes in different groups. (A, B) Venn diagram of up/down-regulated differentially expressed genes between G2 and G0 groups. (C, D) Venn diagram of up/down-regulated differentially expressed genes between G3 and G1 groups. (E, F) Venn diagram of up/down-regulated differentially expressed genes in all groups. G0: BrL.2-25 vs BrL.2-4; G1: BrL.1-25 vs BrL.2-25; G2: BrL.1-25 vs BrL.1-4; G3: BrL.1-4 vs BrL.2-4.
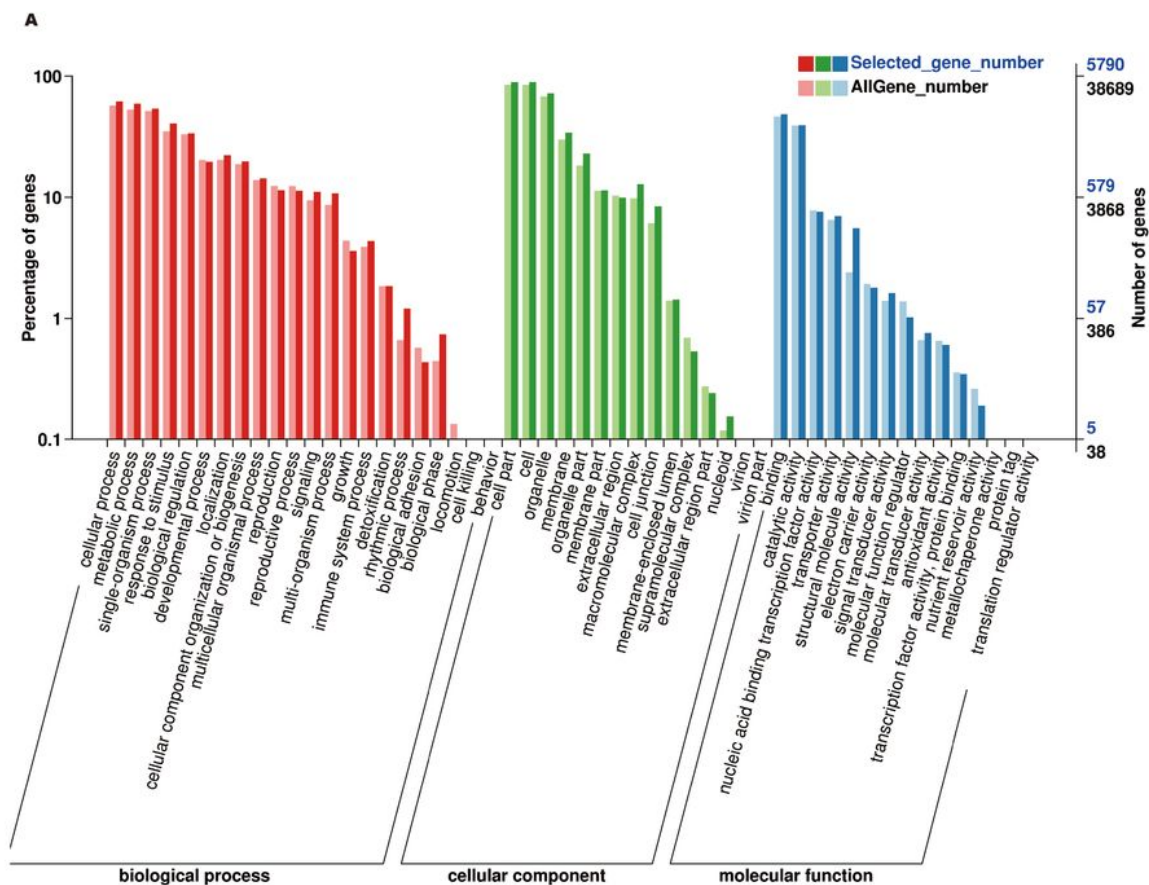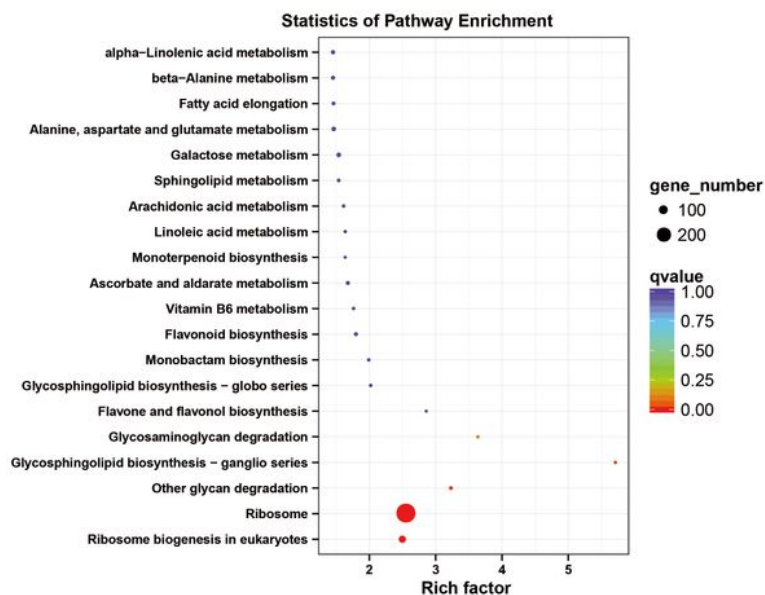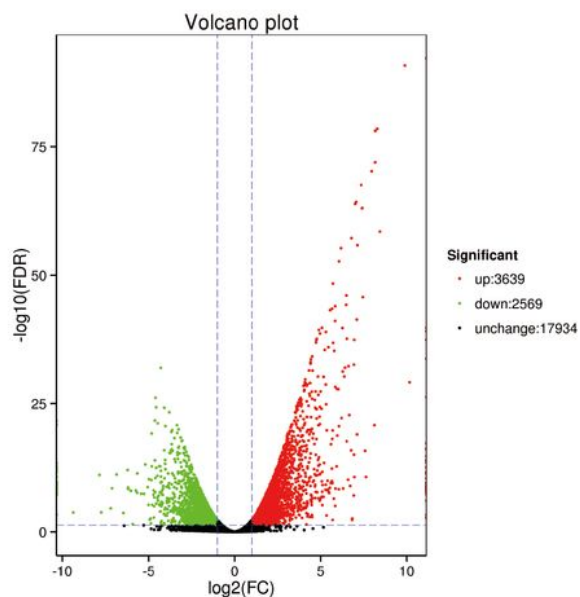
Figure 3

Enrichment analysis of differentially expressed genes between C2 and C1 groups. (A) GO classification statistics of differentially expressed genes. It mainly includes 3 branches: biological process, cellular component, and molecular function. (B) KEGG pathway enrichment statistics of differentially expressed genes. (C) Volcano plot statistics of the number of differentially expressed genes. C1: BrL.1-25, C2: BrL.1-4.
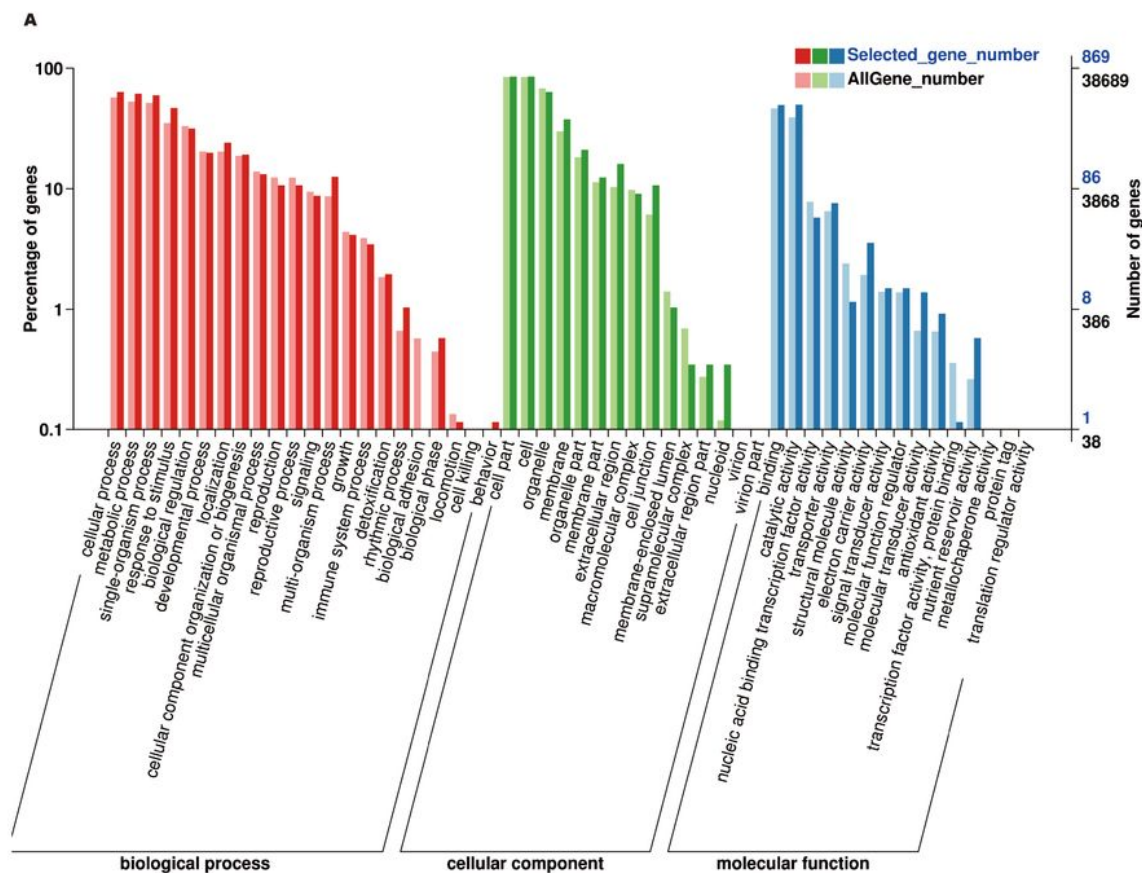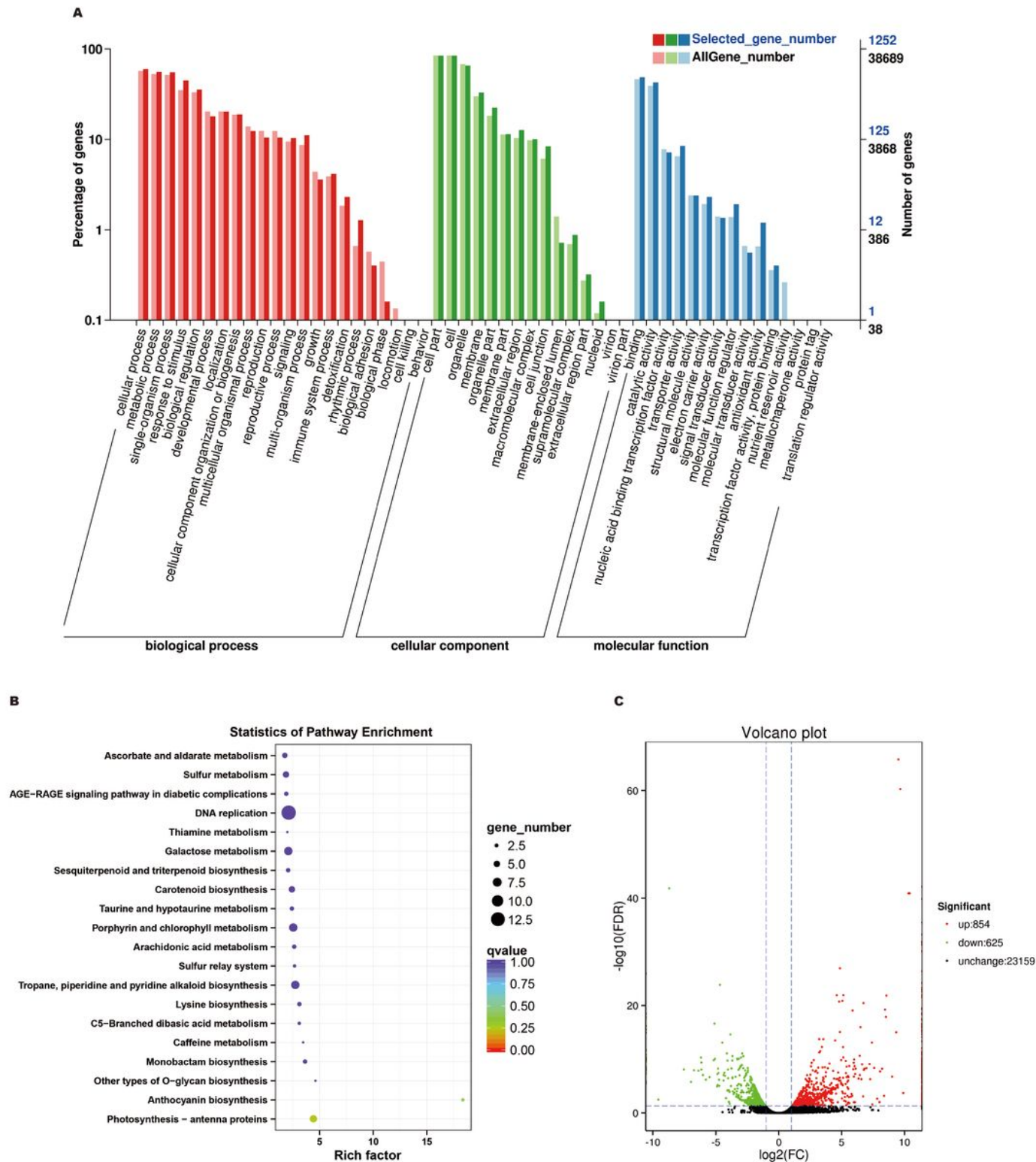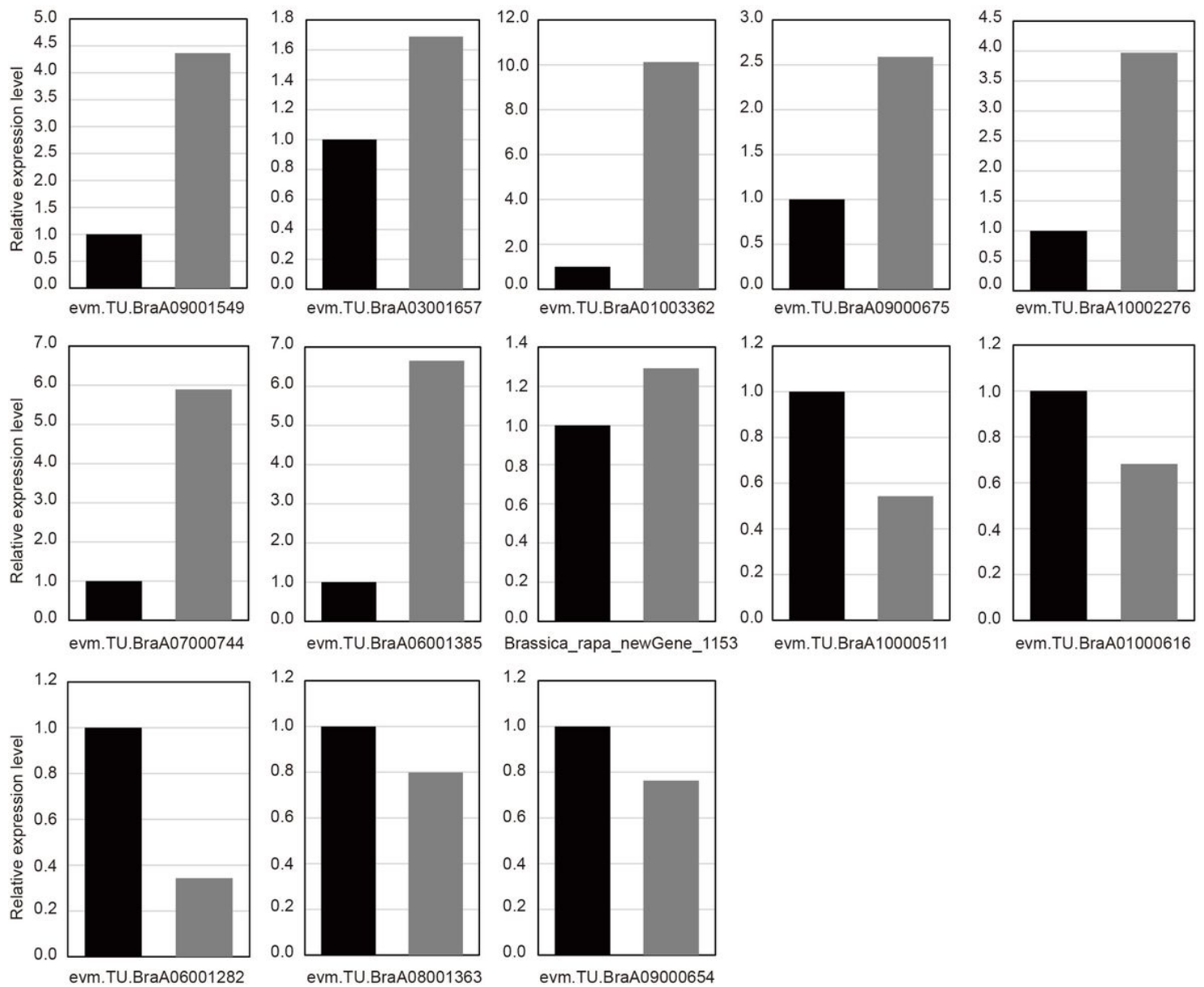
Figure 4

Enrichment analysis of differentially expressed genes between C4 and C3 groups. (A) GO classification statistics of differentially expressed genes. It mainly includes 3 branches: biological process, cellular component, and molecular function. (B) KEGG pathway enrichment statistics of differentially expressed genes. (C) Volcano plot statistics of the number of differentially expressed genes. C3: BrL.2-25, C4: BrL.2-4.
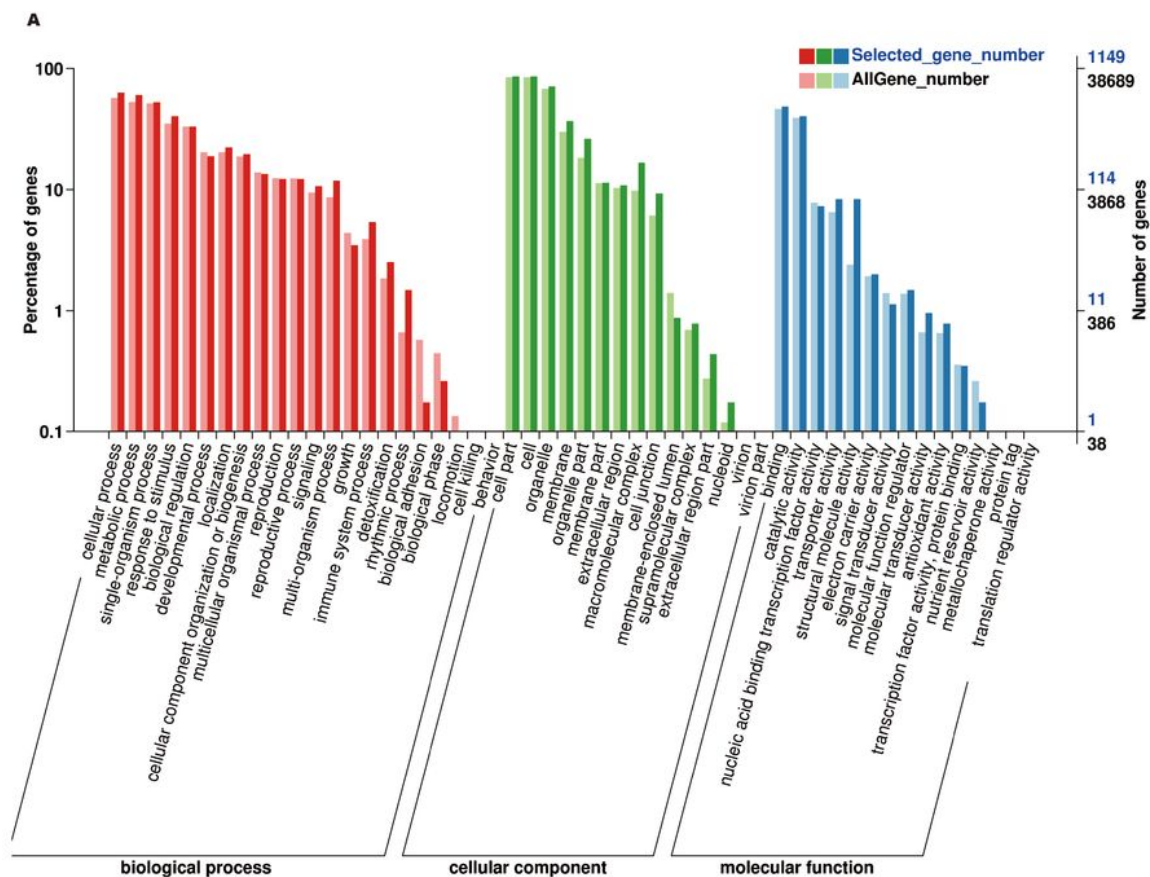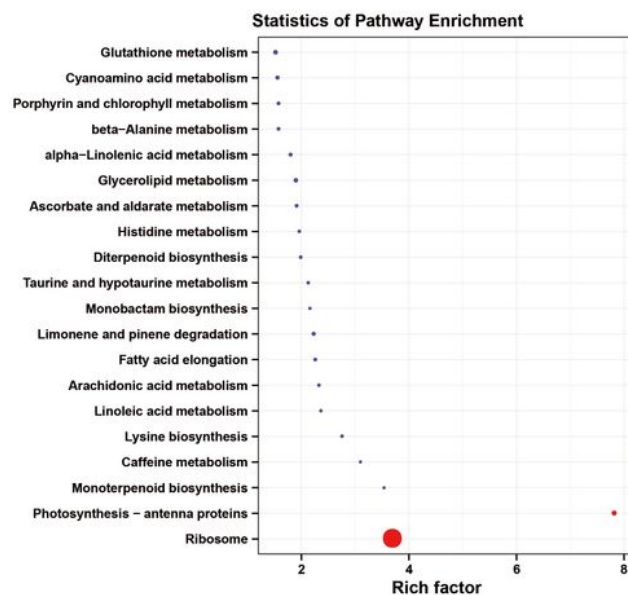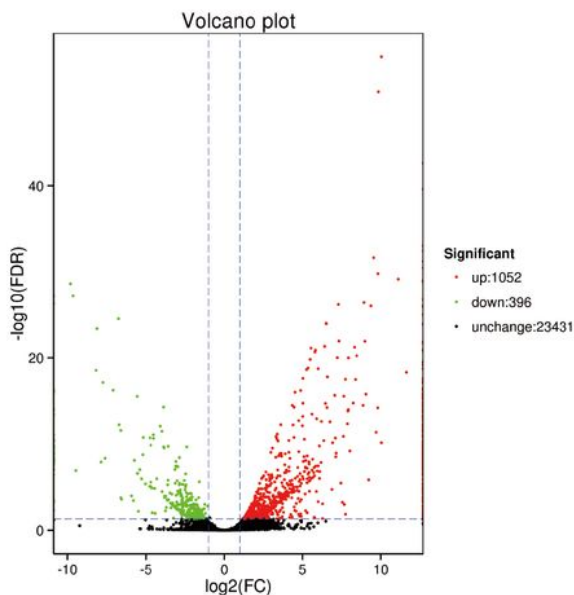
Figure 5

Enrichment analysis of differentially expressed genes between C3 and C1 groups. (A) GO classification statistics of differentially expressed genes. It mainly includes 3 branches: biological process, cellular component, and molecular function. (B) KEGG pathway enrichment statistics of differentially expressed genes. (C) Volcano plot statistics of the number of differentially expressed genes. C1: BrL.1-25, C3: BrL.2-25.

Figure 6

Enrichment analysis of differentially expressed genes between C4 and C2 groups. (A) GO classification statistics of differentially expressed genes. It mainly includes 3 branches: biological process, cellular component, and molecular function. (B) KEGG pathway enrichment statistics of differentially expressed genes. (C) Volcano plot statistics of the number of differentially expressed genes. C2: BrL.1-4, C4: BrL.2-4.

Figure 7

Verification of common differentially expressed genes by qRT-PCR. 13 common DEGs were chosen for qRT-PCR validation. The relative expression levels of each genes were expressed as the fold change between BrL.1-25 (black column) and BrL.1-4 (gray column).
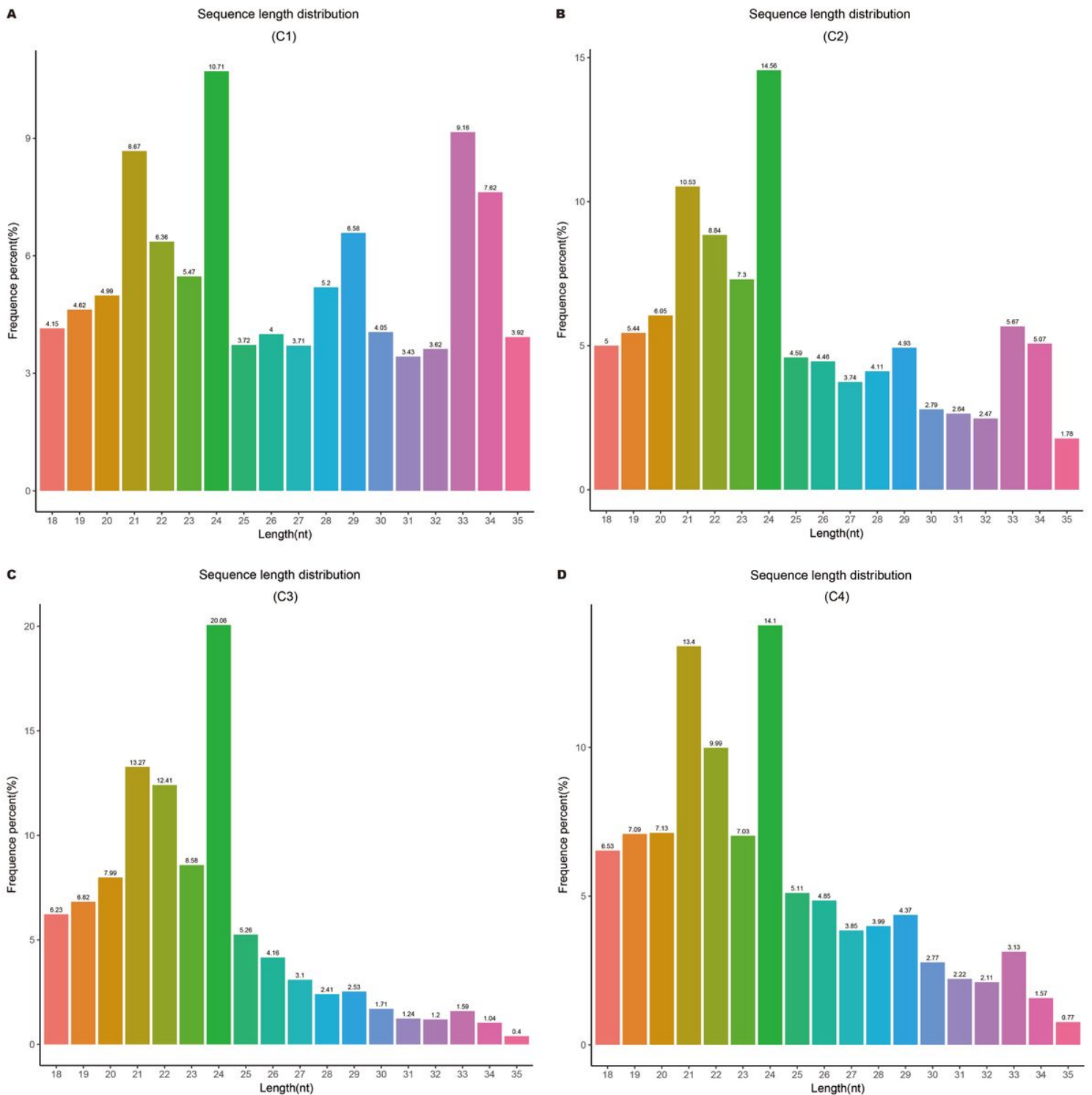
Figure 8

Length distribution and abundance of small RNAs in four libraries from Pak-choi. (A) Frequence percentage (%) and length of sRNA distribution in C1 (BrL.1-25) group. (B) Frequence percentage (%) and length of sRNA distribution in C2 (BrL.1-4) group. (C) Frequence percentage (%) and length of sRNA distribution in C3 (BrL.2-25) group. (D) Frequence percentage (%) and length of sRNA distribution in C4 (BrL.2-4) group.
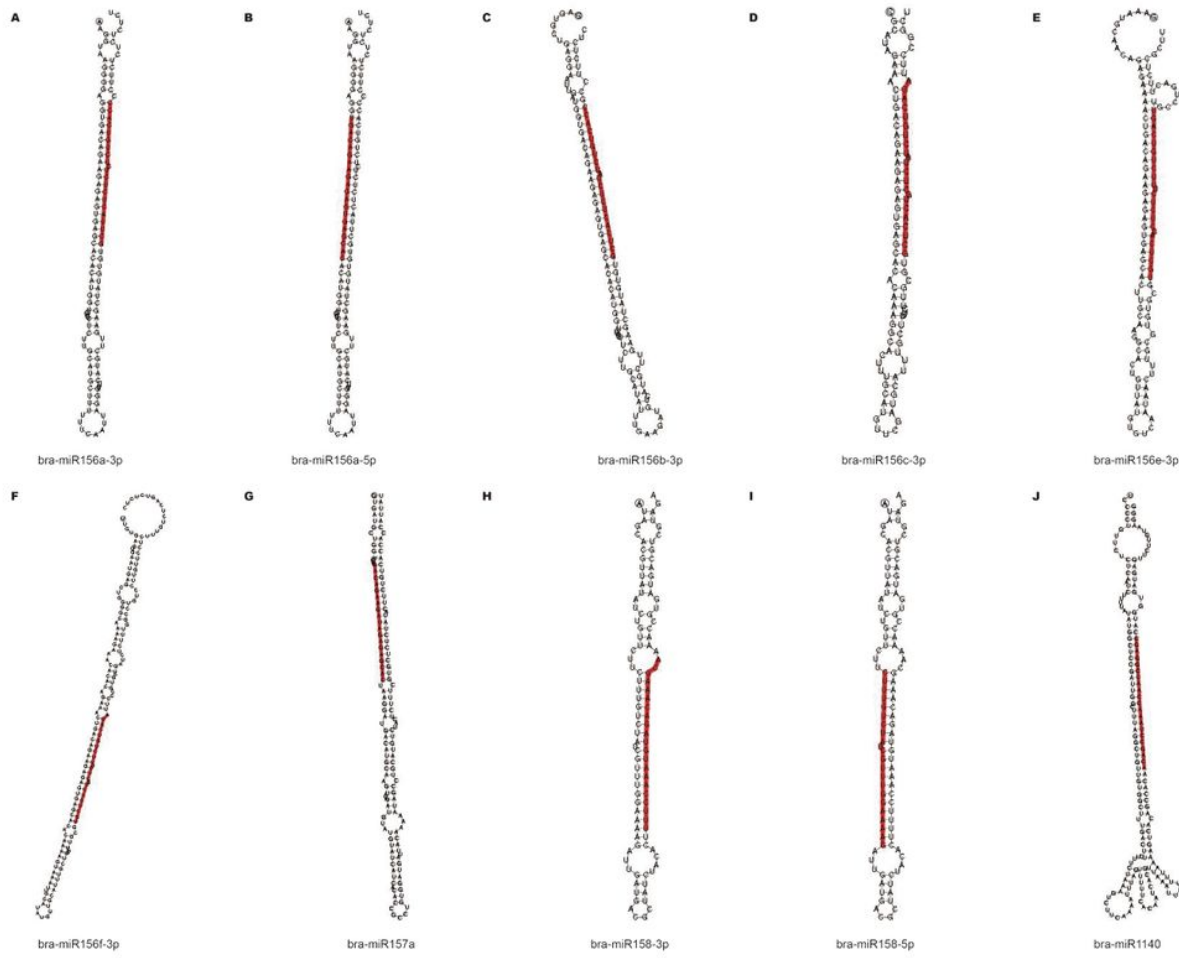
Figure 9

Secondary structure of the known miRNA on the match. (A-J) Secondary structure of 10 known miRNAs on the match from Pak-choi. The entire sequence is miRNA precursor, and the red highlight is where the mature sequence is located.
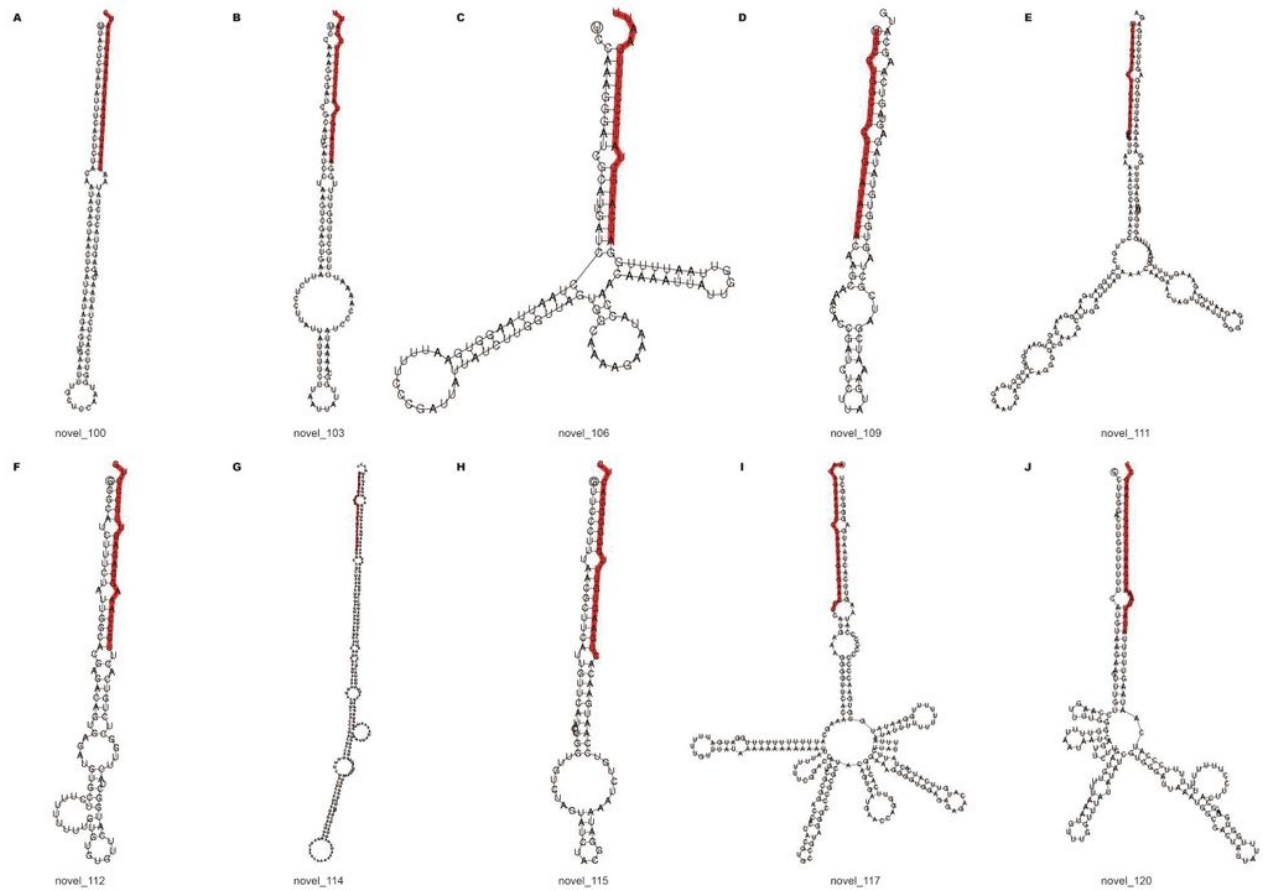
Figure 10

Secondary structure of predicted new miRNA. (A-J) Secondary structure of 10 predicted new miRNA. The entire sequence is miRNA precursor, and the red highlight is where the mature sequence is located.
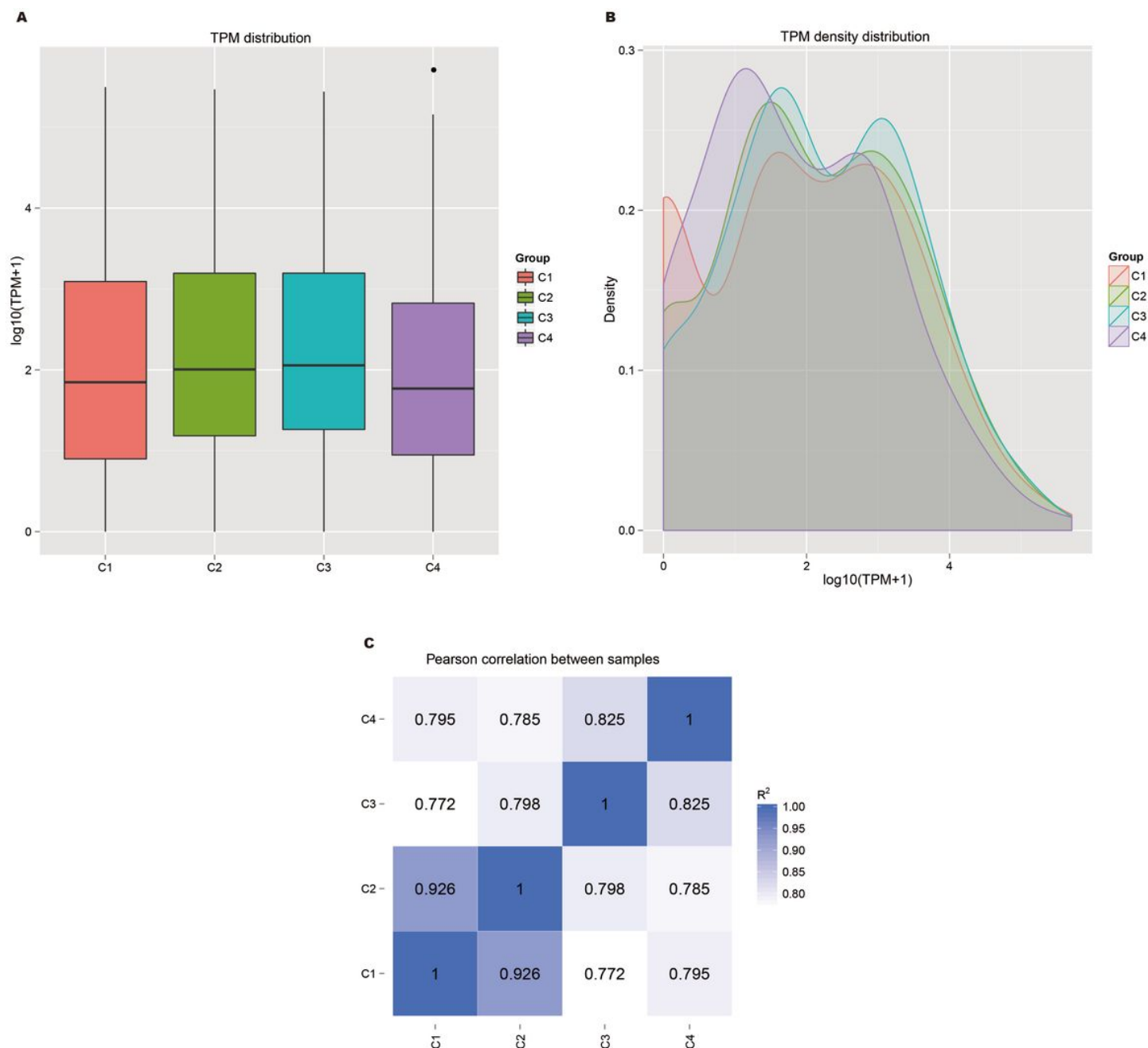
Figure 11

Gene expression patterns and correlations in each sample. (A) Box plot of TPM distribution in all groups. (B) TPM density distribution diagram of miRNA expression levels. (C) Pearson correlation of miRNA expression between samples. C1: BrL.1-25, C2: BrL.1-4, C3: BrL.2-25, C4: BrL.2-4.
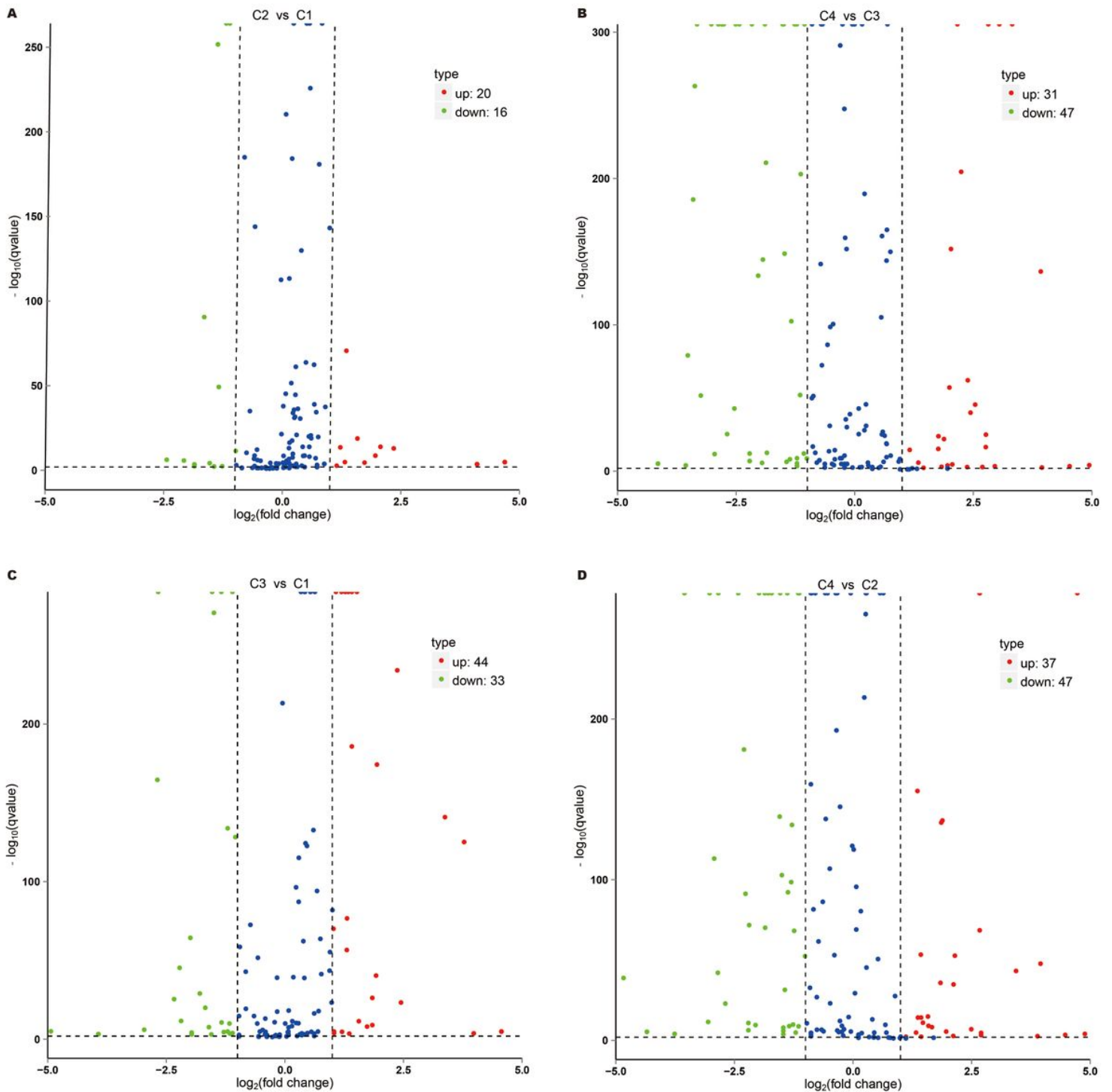
Figure 12

Volcano plot of differential miRNA. (A) differential miRNA from C2 vs C1. (B) differential miRNA from C4 vs C3. (C) differential miRNA from C3 vs C1. (D) differential miRNA from C4 vs C2. X-axis, miRNA expression fold change in different experimental groups/different control groups; Y-axis, statistical significance of the change in the expression level of miRNA. Red dots indicate significantly differential up-regulated miRNAs, green dots indicate significantly differential down-regulated miRNAs, blue dots

indicate those without significantly differential miRNAs. C1: BrL.1-25, C2: BrL.1-4, C3: BrL.2-25, C4: BrL.2-4.

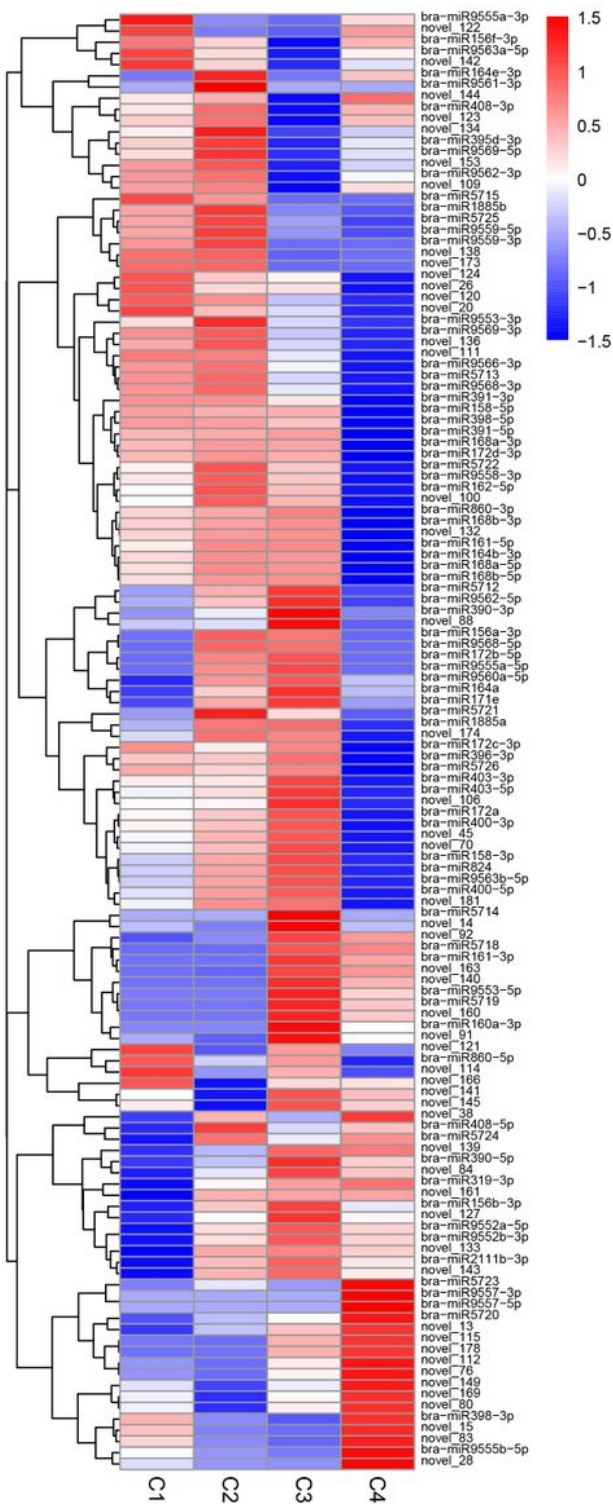Cluster analysis of differentially expressed sRNA

Figure 13

Cluster analysis of differentially expressed miRNA. The above figure is the overall hierarchical clustering chart. Clustering is performed by the log10 (TPM + 1) value. Red indicates high expression miRNA and blue indicates low expression miRNA.
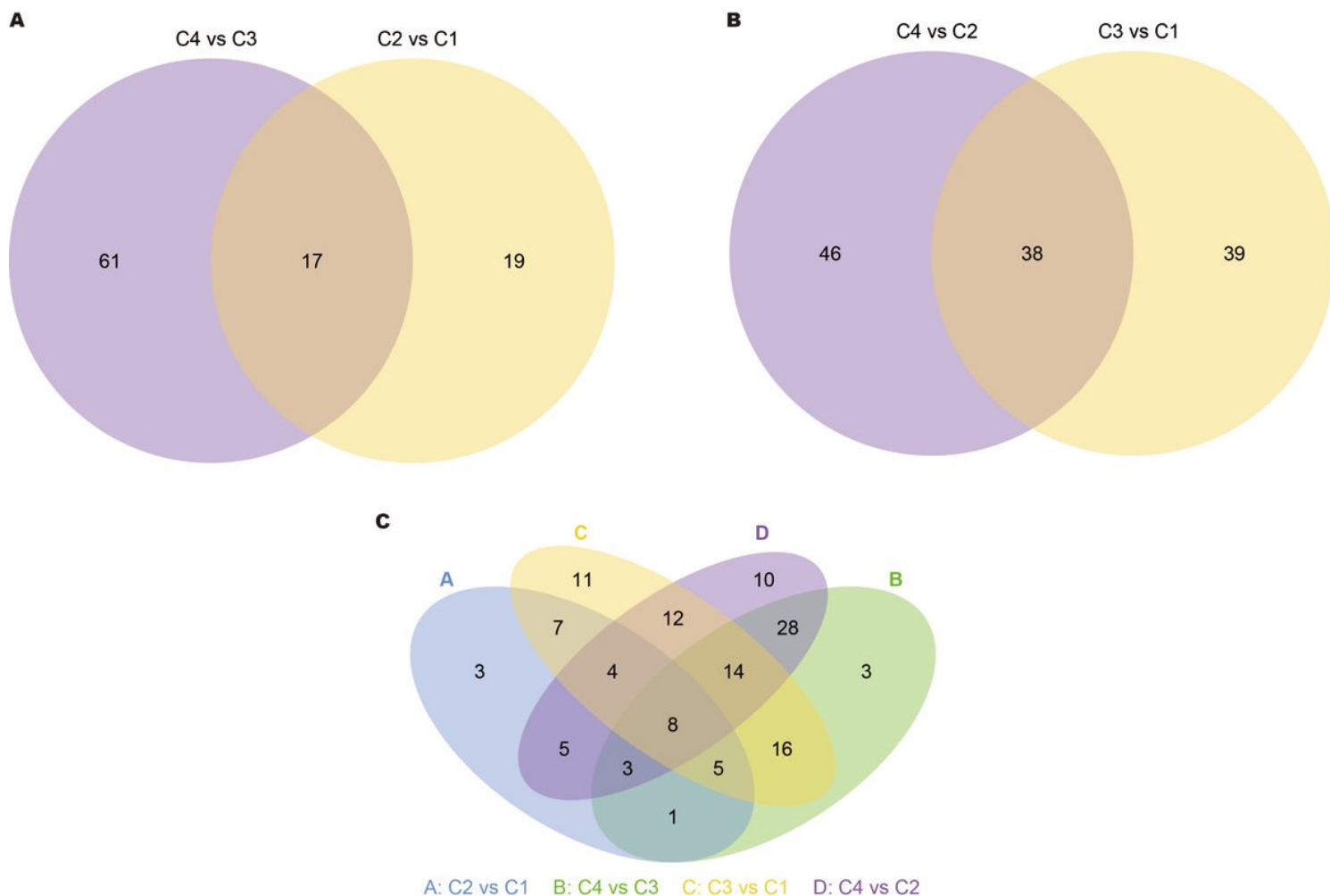
Figure 14

Venn diagram of differential miRNA. The large circles represent each comparison combinations, and the sum of the number in each large circle represents the total number of differential miRNAs from the comparison combinations, and the overlapping portions of the circles represent the number of common differential miRNAs between the combinations.
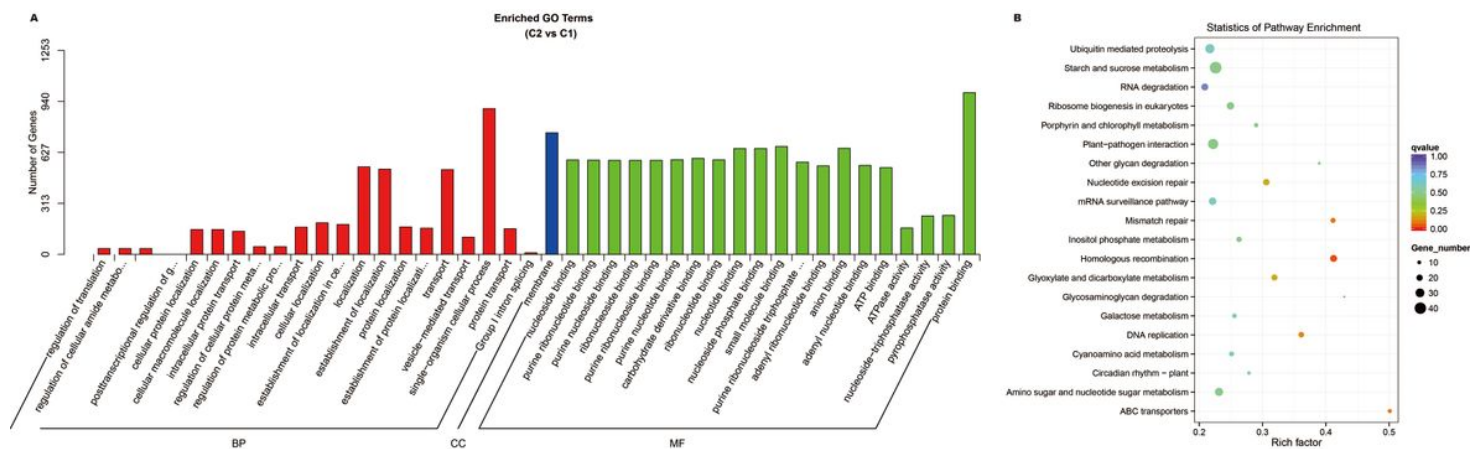

Figure 15

Enrichment analysis of candidate target genes for differentially expressed miRNA between C2 and C1 groups. (A) Histogram of GO enrichment for candidate target genes from C2 vs C1. Three basic classifications of Go term (from left to right, BP: biological processes, CC: cellular components, MF: molecular functions). (B) KEGG enrichment scatter plot of candidate target genes from C2 vs C1. X-axis, the Rich factor; Y-axis, the name of the pathway. The size of the dots indicates the number of candidate target genes in this pathway, and the colors of the dots correspond to different Qvalue ranges. C1: BrL.1-25, C2: BrL.1-4, C3: BrL.2-25, C4: BrL.2-4.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TABLES14.xls
- TABLES18.xls
- TABLES17.xls
- TABLES16.xls
- TABLES15.xls
- TABLES11.xls
- TABLES19.xls
- TABLES22.xls
- TABLES21.xls
- TABLES20.xls
- TABLES26.xls
- TABLES25.xls
- TABLES24.xls
- TABLES23.xls
- TABLES9.xls
- TABLES12.xls
- TABLES3.xls
- TABLES13.xls
- TABLES5.xls
- TABLES6.xls
- TABLES7.xls
- TABLES8.xls
- TABLES2.xls
- TABLES10.xls

- FIGURES2.tif
- FIGURES5.tif
- FIGURES6.tif
- TABLES1.xls
- FIGURES3.tif
- FIGURES4.tif
- FIGURES1.tif
- TABLES4.xls