Digital gene expression analysis during floral transition in pak choi (Brassica rapa subsp. chinensis)

Xueting Wang*, Hongxia Song*, Mengxia Sun*, Zhujun Zhu, Guoming Xing, Xiaoyong Xu, Meiyung Ga, Leiping Hou and Meilan Li

*College of Horticulture, Shanxi Agricultural University, Taigu, Shanxi, P. R. China; bKey Laboratory for Quality Improvement of Agricultural Products of Zhejiang Province, College of Agricultural and Food Science, Zhejiang A&F University, Hangzhou, P. R. China

Digital gene expression profiling technology was used to examine the changes in gene expression during floral transition in pak choi. A total of 1486 differentially expressed genes (DEGs) during floral transition were identified, of which 505 were upregulated and 981 were downregulated. Gene ontology enrichment analysis showed that the proteins encoded by DEGs were mainly located in eight cell regions including the apoplast, plant-type cell wall, chloroplast, etc. They had eight kinds of molecular functions such as transcription factor and oxidoreductase, and involved in 72 biological processes containing jasmonic acid and salicylic acid metabolism, glucose metabolism, the MAPK cascade, etc. Further pathway enrichment analysis showed that four metabolic pathways were significantly enriched by DEGs. DEGs that exhibited at least a 10-fold difference were analyzed in detail. Results showed that flavonoid metabolism and phenylpropanoid biosynthesis (or pathways) were significantly enriched and the corresponding gene number was the highest. Moreover, 15 genes among the highly expressed genes were involved in reproductive development, of which five homologues in Arabidopsis thaliana were directly related to floral transition, which were predicted to have a similar function in pak choi.

Introduction

The transition from vegetative growth to reproductive growth is not only relevant in the perpetuation of species, but also related closely to human life; this transition in higher plants is regulated by endogenous and exogenous signals [1–3]. With the completion of whole genome sequencing in the model plants Arabidopsis (Arabidopsis thaliana) and rice (Oryza sativa), the molecular mechanisms of flowering were also investigated, and this endeavour has achieved significant progress [4]. New genes were identified and intensively characterized [5–7], and new regulation pathways were discovered in succession. Wang et al. [8] revealed that the microRNA156-targeted SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) genes not only act downstream of FT/FD, but also define a separate endogenous flowering pathway. Currently, at least six pathways that control flowering have been identified and they are vernalization pathway, photoperiod pathway, autonomous pathway, gibberellin pathway, ambient temperature pathway, and ageing pathway [1,9].

Pak choi (Brassica rapa subsp. chinensis) belongs to the Brassicaceae family, and it is one of the more cultivated and consumed vegetables in China and other Asian countries. Pak choi was recently introduced in North America and Europe. The recent completion of whole genome sequencing and the constant enrichment of information on gene annotation in Chinese cabbage, which is the representative of the Brassica species [10,11], facilitates the gene comparison and screening for expression profile analysis of pak choi and provides an opportunity for genetic improvement and molecular design breeding among Brassica species. Although some studies on flower molecular regulation in Brassica species have been conducted [12–16], information is still lacking compared with the flowering molecular mechanism of the model plant A. thaliana. Therefore, further studies on the flowering mechanism of pak choi are necessary in B. rapa, as well as in breeding within Brassicaceae.

Digital gene expression (DGE) profiling technology is based on high-throughput sequencing technology and it is an effective method for the systematic investigation of the overall differences in gene expression at the transcriptome level. Given its high sensitivity and ability to detect splicing isoforms and somatic mutations, DGE has
been widely used in gene expression profiling and pathway studies [17]. In addition, because DGE does not require the genome reference sequence, this technology has also been applied to decode the genomes of several non-model organisms, thereby providing valuable information that elucidates gene functions in these organisms [18,19].

In the present study, DGE profiling was performed using Illumina sequencing technology to examine gene expression profile changes in the shoot apex before and after flower bud differentiation in pak choi. This study aimed to screen the differentially expressed genes (DEGs) and identify the pathways involved during floral transition in pak choi. The results may be used as a basis to improve the flowering molecular mechanism of B. rapa vegetables.

Materials and methods

Plant material and cultivation management

Pak choi variety ‘75’ is an easy-bolting inbred line provided by the Institute of Vegetables Research, Shanxi Academy of Agricultural Sciences.

Seeds were rinsed and soaked in water at 55 °C for 10 min and kept for 2 h at room temperature. The seeds were placed in Petri dishes with double filter paper on the bottom and then kept in a growth chamber at 25 °C and 16 h/d light until the radicle emerged through the seed coat. Thereafter, the seeds were placed in the refrigerator at 4 °C for 20 d in the dark. The filter paper was kept moist during the entire period to ensure seed germination [20]. After treatment, the seedlings were planted in the 50-hole pot filled with sphagnum peat/vermiculite substrate mixture (3:1, v/v), and standard agricultural practices were applied throughout the experiment.

Sampling

Flower bud differentiation during plant growth was graded based on B. rapa subsp. Pekinensis [21] and Arabidopsis [22]. On the 16th day after planting, the shoot apex started to differentiate into the flower bud (stage 1). Each apex was accurately determined to be in stage 1 using a stereomicroscope (OLYMPUS SZX16), the apex of six plants were taken and mixed into approximately 0.15 g of sample and were quickly frozen in liquid nitrogen, and then stored in −80 °C until RNA extraction. Plant apices were labelled as flower bud differentiation stage 1 (stage 1). The sample obtained on 15th day (when the plant had about seven true leaves and when the flower bud did not yet start to differentiate) was set as the control (CK).

Total RNA extraction and gene expression profile sequencing

Total RNA was extracted using QIAGEN RNeasy Plant Mini Kit (QIAGEN, 74903, Germany) with DNasel according to the suggested protocol. The integrity of total RNA was verified by running the samples on 1% agarose gel. The quality and quantity of the purified RNA were determined by measuring the absorbance at 260/280 nm (A260/A280) using a NANODROP 2000c spectrophotometer (Thermo Scientific, USA). High-throughput sequencing was performed by Biomarker Technologies (Beijing, China). Total RNA (6 μg) was enriched by binding to oligo(dT) magnetic adsorption beads. The mRNA was fragmented by adding fragmentation buffer. The RNA fragments then served as a template for first-strand cDNA synthesis using random hexamers and reverse transcriptase. Second-strand cDNA was synthesized using DNA polymerase I, dNTPs, and RNaseH. Second-strand DNA was then purified using a QiaQuick PCR extraction kit. The double-stranded cDNA was end-repaired, and a single ‘A’ base was added to the 3’-end of the cDNA fragments. Solexa adapters were then ligated to the ends of the cDNA fragments. Solexa adapters were then ligated to the ends of the cDNA fragments. Solexa adapters were then ligated to the ends of the cDNA fragments.

Analysis of gene expression and functional annotation

The raw sequences of the two datasets were evaluated. Raw reads were processed by remove reads containing adaptors and reads with more than 10% unknown bases. Low-quality reads were filtered out that each reads with a quality score of less than 10 was not more than half of the bases, and the clean reads consist of more than 80% base-pair with Q-value ≥30. The clean reads were aligned with TopHat software [23], and then were mapped with the reference gene sequence (http://brassicadb.org/brad/), allowing for up to 1 bp mismatch only. The mapped reads were classified as annotated. The false discovery rate (FDR) ≤ 0.01 and fold change (FC) ≥ 2 were used to determine the threshold of significant DEGs.

Gene function and pathway involved in were analyzed using the gene ontology (GO) (http://www.geneontology.org/) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg/) databases, and the
genes upregulated and downregulated by at least 10-fold were analyzed in detail.

**Quantitative real-time PCR**

To validate the DGE results, we assessed two randomly selected genes’ relative expression levels by quantitative real-time PCR (qPCR) analysis. For reverse transcription, the concentration of each RNA sample was adjusted to 500 ng/μL. Single-stranded cDNA was synthesized using PrimeScript® RT Reagent Kit for Perfect Real Time (TaKaRa, DRR037A, Japan) following the manufacturer’s instructions. Specific primers used for qPCR were designed with the online Primer 3 software, and the pak choi ACTIN gene was chosen as an internal reference gene. qPCR was carried out using SYBR® Premix Ex TaqTM (Tli RNaseH Plus) (TaKaRa, DRR820A, Japan) in a 25-μL reaction volume. The reaction mixture included 20 ng cDNA (2 μL), 8.5 μL ddH2O, 12.5 μL SYBR and 1 μL of each of the forward and reverse primers (10 μmol/L). The program was as follows: 30 s at 95 °C, followed by 40 PCR cycles of 5 s at 95 °C, 15 s at 55 °C and 20 s at 72 °C. The relative levels of gene expression were calculated using the comparative 2−ΔΔCT method.

**Data access**

The transcriptome sequencing data from this study have been deposited in the NCBI SRA database and are accessible through accession numbers SRP066286 (http://www.ncbi.nlm.nih.gov/sra).

**Results and discussion**

**Sequencing quality evaluation**

Read analysis showed that the raw reads were transformed into clean reads after removing the low-quality reads. The proportions of the clean reads in control and stage 1 sample were 99.90% and 99.96%, respectively, indicating that the quality of the cDNA library and sequences was high.

Considering the differences in genetic background and limitation of the annotated information, approximately 85.08% of the clean reads could be mapped to the reference sequences in the control sample (Table 1), while percentage in stage 1 was approximately 84.25%. Most of them (62.58% of control and 63.50% of stage 1) perfectly matched the genome. The mismatches between all annotation reads and the genome were nearly less than 2 bp. In addition, multiple reads mapped to similar genes was low, and most of the annotated genes were only a single read (93.79% of control, 93.83% of stage 1). These results indicated that the homology between pak choi and the reference genome was relatively high, and the selected reference genome was good, which could meet the needs of subsequent analyses.

![Correlation Plot](image)

**Gene expression profile analysis before and after flower bud differentiation**

A total of 41,174 reference genes were selected for DEGs analysis. Reads that could be annotated produced 41,020 effective reference genes, accounting for 99.87% of the total number of reference genes. FDR value ≤0.001 and FC ≥2 were used as a threshold to identify significant DEGs. The number of DEGs between the stage 1 group and control group was 1486, of which 505 were upregulated and 981 were downregulated, and the number of downregulated genes was significantly higher than that of the upregulated genes (Figure 1).

Gene expression fold changes were highly uneven (Figure 2). A total of 1486 genes showed at least two-fold changes. Fold changes were mainly two- to five-folds,

![Figure 1](image)

**Table 1. Sequence comparison of samples with reference genome.**

|       | Total reads | Clean reads | Mapped reads | Perfect map | <2 bp mismatch | Unique mapped reads | Multiple mapped reads |
|-------|-------------|-------------|--------------|-------------|----------------|---------------------|-----------------------|
| CK    | 11,072,438  | 11,061,770  | 9,411,660 (85.08%) | 5,011,092 (53.24%) | 8.72% | 8,827,419 (93.79%) | 584,241 (6.21%) |
| Stage 1 | 10,709,754  | 10,695,085  | 9,010,384 (84.25%) | 4,820,829 (53.50%) | 8.74% | 8,454,786 (93.83%) | 555,598 (6.17%) |
thereby accounting for 88.1% (74.3%) of the total number of DEGs. Five- to ten-fold changes in DEGs accounted for 9.11% (16.0%). Twelve (81%) DEGs exhibited 10- to 30-fold changes, accounting for 2.38% (8.26%). The DEGs with expression of 30-fold or greater accounted for only 1.08%, of which only 2 were upregulated and 14 were downregulated at stage 1. These results showed that the number of DEGs with significant differences was low, and it significantly decreased with differences increasing during floral transition in pak choi. Thus, floral transition did not affect significantly most genes’ expression levels, which suggested that the function of DEGs with minor differences could not be ignored in analysis.

**GO functional analysis of DEGs**

GO functional enrichment analysis determines the main biological functions and explores the functional categories associated with the DEGs. Based on the overall result of DEG analysis, DEGs were annotated using GO analysis during floral transition in pak choi. The proteins encoded by DEGs were mainly located in the apoplast, plant cell wall, chloroplast thylakoid, thylakoid lumen, chloroplast thylakoid membrane, and other cell regions. They have transcription factors, oxidoreductase and other molecular functions, and involved in eight biological processes including jasmonic acid (JA) and salicylic acid (SA) metabolism, sugar metabolism, programmed cell death and MAPK cascade, and so on. Eighty-eight significantly enriched terms were detected by GO analysis at a corrected p-value <0.05. Figure 3 showed the top 30 significantly enriched terms, including those described as cellular components and molecular functions. Only the top 14 of 72 biological processes’ significantly enriched terms were listed because of limited space according to the corrected p-values.

Among the terms listed in cellular components, the apoplast (GO: 0048046) was the most enriched term, followed by plant-type cell wall (GO: 0009505) (Figure 3). The apoplast is the major pathway for nutrient transport, storage, and activation. The apoplast and plant-type cell wall do not only play a key role in intercellular communication and interactions, but the apoplast can also determine cell fate during cell differentiation [24]. In addition, the chloroplast thylakoid lumen and thylakoid membrane were shown to be enriched regions, both of which are sites for food production and energy conversion. Cellular localization analysis was also performed on 109 genes that were either upregulated or downregulated by at least 10-fold. These genes corresponded to the chloroplast thylakoid membrane and apoplast, indicating that plants make large amounts of nutrients through energy conversion and transport them *in vivo*. This result further explained that floral transition by itself is a complex process and requires large amounts of energy and nutrients.

In molecular function categories, the sequence-specific DNA binding transcription factor activity (GO: 0003700) was the most enriched term. Upon comparative analysis of the upregulated and downregulated genes of at least 10-fold, 14 genes were classified into nine classes of transcription factors. The MYB
transcription factors, including Bra037419, Bra017813, and Bra000453, were expressed the most, demonstrating expression changes of at least 20-fold. The MYB transcription factor family, one of the largest families of transcription factors in plants, regulates plant growth and metabolism, such as cell morphology and pattern formation, secondary metabolism, and biological and abiotic stress responses [25]. The MYB transcription factor also played a role in regulating anther development and pollen formation [26]. Another enriched item is phenylalanine ammonia-lyase (PAL), which participates in polyphenol (e.g., flavonoid and phenylpropanoid) synthesis, whereas, adenylyl-sulfate reductase (glutathione), phosphoadenylyl-sulfate reductase (thioredoxin) activity, and adenylyl-sulfate reductase are all categorized as reductant–oxidant enzymes. Thus, plants require high levels of energy during floral transition, and this finding was similar to that for enriched terms for cellular components.

In terms of biological process, the term JA-mediated signalling pathway (GO: 0009867) showed the largest number of annotated genes. High amounts of JA, which plays an important role in the formation of flowers, fruits, and seeds, are found in the reproductive organs of plants; many genes responding to JA exhibit high expression levels in flower and fruit development [27]. The present study found a large number of annotated genes for the JA-mediated signalling pathway, indicating that JA may be involved in reproductive organ development during floral transition in pak choi.

In the present study, the MAPK cascade contained 10 genes with at least 10-fold DEGs. The MAPK cascade is a protein kinase that is also known as protein phosphorylase. The MAPK cascade is activated by various biological and abiotic stresses, and it is also involved in hormone signal transduction and plant growth and development [28]. In this research, the MAPK cascade was found to be involved in floral transition in pak choi.

The term response to sucrose was significantly enriched in the biological process category, and it contained 17 genes with at least 10-fold DEGs. Sugar, which is the main product of photosynthesis, is essential in regulating plant metabolism and development, such as germination, flowering, ageing, and stress reaction. Sucrose significantly promotes flowering, for example, pineapple and rudbeckia stem tips accumulate large amounts of sucrose during floral transition [29,30]. The experimental results showed that sugar plays a key role in floral transition in pak choi.

SA biosynthesis was significantly enriched in the biological process category, and it contained 10 genes with

Figure 3. Functional categorization of DEGs during floral transition in pak choi-based gene ontology distribution. X-axis represents the number of differential genes of each GO subcategory. GO subcategories are on the Y-axis.
at least 10-fold DEGs. SA is a simple phenolic compound in plants. SA is a floral induction factor during flowering. Low SA concentrations promote flower bud differentiation, whereas high SA concentrations inhibit flower bud differentiation [31]. Treatment with a certain SA concentration suppresses the growth, delays the onset of flowering, reduces the number of flowers, and affects the consistency and corm plumpness of Freesia [32]. Moreover, low SA concentration promotes flower bud differentiation in cucumber [33]. These results suggested that SA may affect the flowering of pak choi. However, the mechanism of the effects of SA on plant flowering remains to be further studied.

**KEGG pathway analysis**

To identify the biological pathways involved in floral transition, 1486 genes were mapped to the referenced and canonical pathway in KEGG. A total of 88 KEGG pathways were assigned, of which 19 pathways showed significant differences between the control and stage 1 groups (p < 0.05) (Table 2). However, at a corrected p-value < 0.05, only four of these pathways were significantly enriched, namely, nitrogen metabolism, flavonoid biosynthesis, phenylpropanoid biosynthesis, and tryptophan metabolism.

A total of 109 genes that were either upregulated or downregulated by at least 10-fold were used in pathway enrichment analysis. The highest number of enriched genes was involved in the flavonoid biosynthetic and phenylpropanoid biosynthetic pathways. This result was consistent with DEGs significantly enriched in the GO biological process ‘flavonoid biosynthetic process’ and GO molecular function ‘phenylalanine ammonia-lyase activity.’ It suggested that flavonoids and phenylpropanoids played an important role in floral transition in pak choi. Phenylpropanoid metabolism involves a complex series of branched biochemical reactions. This process produces numerous phenolic substances in plants, and its metabolic branches vary among plant species; one of its main branches produces flavonoid compounds, including flavonoid alcohol, anthocyanin, and tannin [34]. Flavonoids are the main pigments in flowers, fruits, and seeds; they also protect plants against UV, prevent pathogenic microorganisms invasion, increase crop yield, adjust auxin transport, promote pollen germination, and are directly related to pollen fertility [35–39].

In this study, nitrogen metabolism was the first significant enrichment pathway, which had two genes with at least 10-fold change in expression. Nitrogen metabolism is one of the basic physiological processes in plants. Plants generally absorb inorganic nitrogen compounds, such as ammonia salt or nitrate. Nitrate reductase and ammonia salts are directly involved in amino acid synthesis and transformation. The amino acid is then used as the main substrate in protein synthesis in the cell, and it becomes an important part of the plant, following protein modification, classification, transport, storage, and other relevant processes. Nitrogen affects the flowering of Boronia and Alive [40,41]. Zhang [42] also reported that nitrogen nutrition can adjust gene expression for the induction of flowering in A. thaliana, resulting in changes in floral induction. This result again showed the importance of nitrogen metabolism in floral transition in pak choi.

In the present study, tryptophan metabolism was the fourth among the KEGG pathways, showing significant

**Table 2. The top19 significantly enriched KEGG pathways and DEGs with at least 10-fold change.**

| Pathway                                       | DEGs with pathway annotation (ratio) | Corr. p-value | Pathway ID | DEGs with at least 10-fold |
|-----------------------------------------------|-------------------------------------|---------------|------------|---------------------------|
| Nitrogen metabolism                           | 18 (6.23%)                          | 7.26E–07      | ko00910    | Bra039113                 |
| Flavonoid biosynthesis                        | 13 (4.50%)                          | 7.19E–06      | ko00941    | Bra023441, Bra006224, Bra029211, Bra0609312 |
| Phenylpropanoid biosynthesis                  | 23 (7.96%)                          | 0.001007      | ko00940    | Bra004109, Bra023872, Bra028761, Bra031290 |
| Tryptophan metabolism                         | 11 (3.81%)                          | 0.043024      | ko00380    | Bra004837, Bra034674      |
| Circadian rhythm – plant                      | 8 (2.77%)                           | 0.30621       | ko04712    | Bra023441, Bra006224      |
| Carotenoid biosynthesis                       | 7 (2.42%)                           | 0.36192       | ko00906    | 0                       |
| Fatty acid metabolism                         | 9 (3.11%)                           | 0.42049       | ko00071    | 0                       |
| Flavone and flavonol biosynthesis             | 3 (1.04%)                           | 0.4286        | ko00944    | 0                       |
| Diterpenoid biosynthesis                      | 5 (1.73%)                           | 0.4836        | ko00904    | Bra009285, Bra022565, Bra08362 |
| Phenylalanine metabolism                      | 15 (5.19%)                          | 0.57319       | ko00360    | Bra004109                 |
| Carbon fixation in photosynthetic organisms   | 14 (4.84%)                          | 0.59899       | ko00710    | 0                       |
| Gynoacino acid metabolism                     | 7 (2.42%)                           | 0.82596       | ko00460    | 0                       |
| Glycerolipid metabolism                       | 9 (3.11%)                           | 0.86248       | ko00561    | 0                       |
| Pentose phosphate pathway                     | 9 (3.11%)                           | 1             | ko00030    | 0                       |
| Starch and sucrose metabolism                 | 16 (5.54%)                          | 1             | ko00500    | 0                       |
| Brassinosteroid biosynthesis                  | 3 (1.04%)                           | 1             | ko00905    | 0                       |
| alpha-Linolenic acid metabolism               | 6 (2.08%)                           | 1             | ko00592    | 0                       |
| Plant–pathogen interaction                    | 20 (6.92%)                          | 1             | ko04626    | Bra005104                 |
| Ubiquinone and other terpenoid-quinone biosynthesis | 5 (1.73%)                           | 1             | ko00130    | Bra004109                 |

Note: 289 DEGs with pathway annotation.
enrichment. Therefore, tryptophan metabolism may play an important role in floral transition. It is reported that tryptophan metabolism not only provides tryptophan, but also provides a large number of metabolic precursors for the synthesis of various secondary metabolites, such as IAA, phytoalexin, and alkaloids derived from indole and antheranilic acid. Tryptophan biosynthesis also plays a direct regulatory role on plant growth and development, pathogen defense response, and plant–insect interactions [43]. Tryptophan was recently found to promote flowering in *Pharbitis nil* [44].

Three genes with at least 10-fold change in expression were found in the diterpenoid biosynthesis pathways. Terpenoids are one of the main floral components, and flower fragrance plays an important role in plant reproduction [45]. Terpenoids do not only render fragrance to flowers, but also perform many other physiological and ecological functions, such as attracting pollinators, regulating plant growth and development, adjusting plant heat resistance and resisting photooxidation stress [46]. This study suggested that terpenoids are synthesized during floral transition in pak choi to prepare for flower formation and protect against bad environmental conditions.

### DEGs related to reproductive growth during floral transition

A total of 15 genes among the highly expressed (FC > 10-fold) 109 genes (Table 3) were speculated to be related to reproductive growth. Bra022565 positively regulated flowering compared with the remaining 14 genes, which displayed downregulated expression levels during floral transition. By comparing these 15 genes with the floral genes published in reference genome, only Bra002788 has been known. Braun02788 was involved in the gibberellin pathway, negatively regulated flowering, exhibited downregulated expression during floral transition in pak choi, and performed a function that was consistent with that of the reference genome. A comparative analysis of the remaining 14 genes from *A. thaliana* showed that Bra022565 was upregulated by 11.77-fold. Its homologue in *A. thaliana* encodes GA20OX2, which belongs to the GA pathways and catalyzes several steps in the biosynthesis of GA by oxidizing a number of precursors, thereby promoting flowering. In addition, the homologue of Bra034674 in *A. thaliana* encodes a peroxisomal catalase, participates in photoperiodic response, and was highly expressed in bolts and leaves, and shows a circadian pattern in mRNA expression, in which mRNA levels are high in the early morning. The Bra027530 was speculated to suppress the meristem transition from vegetative to reproductive growth, and its homologue in *A. thaliana* encodes a member of the NAP subfamily of ABC transporters and is involved in anthocyanin and flavonoid biosynthesis, but without similar functions of switching vegetative growth to reproductive growth. Furthermore, the Bra003998 homologue in *A. thaliana* encodes a member of the NAC transcription factor gene family and regulates flower development and leaf senescence; this homologue is also expressed in floral primordia and is upregulated by AP3 and PI, its expression is also associated with leaf senescence. The Bra031255 homologue is a member of the zinc finger (B-box type) family protein, and it is involved in the light signalling pathway in *A. thaliana* and regulates flower development. The Bra004109 homologue in *A. thaliana* is involved in the formation of pollen exine, and it is associated with anthocyanin and flavonoid biosynthesis. The Bra033350 and Bra026606 homologues in *A. thaliana* are involved in embryonic development. Bra016250 is presumably related to post-embryonic development in pak choi, although its homologue is not functional in *A. thaliana*. The homologues of Bra036883, Bra033911, Bra001886, and Bra023798 in *A. thaliana* all encode early light-inducible protein and

| Gene ID | Stage1RPKM | CKRPKM | FC | Nr...annotation | Chinese cabbage floral genes | Arabidopsis gene or annotation |
|---------|-------------|--------|----|-----------------|-----------------------------|-------------------------------|
| Bra022565 | 4.64 | 0.25 | 11.77 | Gibberellin 20 oxidase 2 | – | GA20OX2 |
| Bra034674 | 2.46 | 0.08 | 8.41 | Catalase | – | CAT2 |
| Bra027530 | 9.51 | 0.08 | 111.65 | Hypothetical protein | – | ATPO1 |
| Bra002788 | 0.13 | 0.03 | 4.89 | Hypothetical protein | – | NAP |
| Bra003998 | 0.00 | 0.00 | 29.33 | Unnamed protein product | – | 4CL3 |
| Bra031255 | 0.52 | 0.03 | 17.00 | B-box 32 protein | – | LEA3 family protein |
| Bra004109 | 1.23 | 0.03 | 33.10 | Late embryogenesis abundant 3 family protein | – | MEE66 |
| Bra033350 | 4.89 | 0.03 | 157.67 | Late embryogenesis abundant 3 family protein | – | ATTBX32 |
| Bra026606 | 0.28 | 0.00 | 5.40 | Hypothetical protein | – | ATG7U11 |
| Bra016250 | 0.00 | 0.00 | 10.54 | Hypothetical protein | – | ELP1 |
| Bra001886 | 4.18 | 0.00 | 13.87 | Hypothetical protein | – | ELP1 |
| Bra033911 | 55.65 | 0.00 | 13.92 | ELIP | – | ELP1 |
| Bra0001886 | 42.00 | 0.00 | 13.92 | ELIP | – | ELP1 |
| Bra023798 | 122 | 0.00 | 13.92 | ELIP | – | ELP1 |
| Bra012600 | 12.10 | 0.00 | 13.92 | Unnamed protein product | – | TCP family transcription factor |
positively control seed germination. Furthermore, Bra012600 positively regulates development. These findings indicated that the homologues of Bra022565, Bra034674, Bra027530, Bra003998, and Bra031255 in A. *thaliana* were directly related to floral transition, which predicted that these genes may have similar functions in pak choi and need to be investigated further.

**Validation of DEGs by quantitative real-time PCR**

To validate the DGE data, gene Bra028599 and Bra029305 were selected for DGE analysis during floral transition; the results are shown in Figure 4. The results of qPCR showed these genes had the concordant direction of expression change with DGE, indicating that the results from DGE are believable.

**Conclusions**

Using DGE profiling technology, 1486 DEGs were identified during floral transition in pak choi. Among these DEGs, 505 were upregulated, whereas 981 were downregulated. The expression change fold of these genes was mostly 2–5. The proteins encoded by DEGs were mainly located in the apoplast, cell wall, chloroplast thylakoid and other parts, and they were widely involved in biological macromolecular metabolism, such as protein and sugar metabolism, hormone metabolism, secondary metabolism, and energy metabolism. The flavonoid metabolic pathway and phenylpropanoid synthesis pathway were mainly involved during floral transition. Fifteen genes (FC ≥ 10-fold) were involved in reproductive growth, among which the homologues of Bra022565, Bra027530, Bra034674, Bra003998, and Bra031255 in A. *thaliana* were directly related to flowering transition. It suggested that these genes may be directly related to flowering in pak choi. Functional analysis and validation of these genes are necessary so as to lay the foundation for further molecular research on the mechanism of floral transition in pak choi.

**Disclosure statement**

The authors have declared that no competing interests exist.

**Funding**

This work was supported by the Shanxi Natural Science Foundation [grant number 201701D121101]; the College and University Specific Scientific Research Fund [grant number 20131403110006]; the Agricultural Science and Technology Research Plan Project of Shanxi Province [grant number FT-201402-06], [grant number 20150311 010-2].

**References**

[1] Srikanth A, Schmid M. Regulation of flowering time: all roads lead to Rome. Cell Mol Life Sci. 2011;68(12):2013–2037.
[2] Meng XQ, Hou ZX, Zhang L. Progress on molecular and genetic mechanisms of floral control in higher plants. Curr Biotechnol. 2013;3:1–6.
[3] Song Y, Dou LD, Zhang HJ. Molecular and genetic mechanisms of control of floral induction in higher plants. Plant Physiol J. 2014;50:1459–1468.
[4] Song YH, Ito S, Imaizumi T. Flowering time regulation: photoperiod- and temperature-sensing in leaves. Trends Plant Sci. 2013;18:575–583.
[5] Liu L, Liu C, Hou X, et al. FTIP1 is an essential regulator required for florigen transport. Plos Biol. 2012;10: e1001313.
[6] Li CN, Zhou YY, Fan LM. A novel repressor of floral transition, MEE3, an abiotic stress regulated protein, functions as an activator of FLC by binding to its promoter in Arabidopsis. Environ Exp Bot. 2015;113:1–10.
[7] Milec Z, Valárik M, Bartoš J, et al. Can a late bloomer become an early bird? Tools for flowering time adjustment. Biotechnol Adv. 2014;32:200–214.

[8] Wang JW, Czech B, Weigel D. MiR156-regulated SPL transcription factors define an endogenous flowering pathway in Arabidopsis thaliana. Cell. 2009;138:738–749.

[9] Fornara F, de Montaigu A, Coupland G. SnapShot: control of flowering in Arabidopsis. Cell. 2010;141:550–553.

[10] The Brassica rapa Genome Sequencing Project Consortium. The genome of the mesopolyploid crop species Brassica rapa. Nat Genet. 2011;43:1035–1039.

[11] Cheng F, Liu SY, Wu J, et al. BRAD, the genetics and genomics database for Brassica. BMC Plant Biol. 2011;11:136.

[12] Lin SI, Wang JG, Poon SY, et al. Differential regulation of FLOWERING LOCUS C expression by vernalization in cabbage and Arabidopsis. Plant Physiol. 2005;137:1037–1048.

[13] Yuan YX, Wu J, Sun RF, et al. A naturally occurring splicing site mutation in the Brassica rapa FLC1 gene is associated with variation in flowering time. J Exp Bot. 2009;60:1299–1308.

[14] Zhuang J, Xiong AS, Peng RH, et al. Analysis of Brassica rapa ESTs: gene discovery and expression patterns of AP2/ERF family genes. Mol Biol Rep. 2010;37:2485–2492.

[15] Österberg MK, Shavorskaya O, Lascoux M, et al. Naturally occurring indel variation in the Brassica nigra CO1 gene is associated with variation in flowering time. Genetics. 2002;161:299–306.

[16] Li J, Gao G, Zhang T, et al. The putative phytocyanin genes in Chinese cabbage (Bra sissca rapa L.): genome-wide identification, classification and expression analysis. Mol Genet Genomics. 2013;288:1–20.

[17] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10:57–63.

[18] Li QJ, Wang LH, Zhan QW, et al. Sorghum bmr6 mutant analysis demonstrates that a shared MYB1 transcription factor binding site in the promoter links the expression of genes in related pathways. Func Integr Genomic. 2013;13:445–453.

[19] Xin HP, Zhu W, Wang LN, et al. Genome wide transcriptional profile analysis of Vitis amurensis and Vitis vinifera in response to cold stress. Plos One. 2013;8:1–14.

[20] Song HX, Fu C, Yin LH, et al. Cloning and expression analysis of LEAFY homologue in Pak Choi (B rassica a rapa subsp. chinensis). Biotechnol Biotechnol Equip. 2015;29:1035–1042.

[21] Li SX, Li SJ. Brassica campestris L. ssp. chinensis flower bud differentiation and leaf head formation. Sci Agric Sin. 1964:636–40.

[22] Smyth DR, Bowman JL, Meyerowitz EM. Early flower development in Arabidopsis. Plant Cell. 1990;2:755–767.

[23] Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;9:1105–1111.

[24] Strauss E. When wall can talk, plant biologists listen. Science. 1998;282:28–29.

[25] Jin H, Martin C. Multifunctionality and diversity within the plant MYB-gene family. Plant Mol Biol. 1999;41:577–585.

[26] Zhang H, Liang WQ, Yang XJ, et al. Carbon Starved Anther Encodes a MYB domain protein that regulates sugar partitioning required for rice pollen development. Plant Cell. 2010;22:672–689.

[27] Berger S, Bell E, Sadka A, et al. Arabidopsis thaliana Atvsp is homologous to soybean VspA and VspB, genes encoding vegetative storage protein acid phosphatases and is regulated similarly by methyl jasmonate, wounding, sugars, light and phosphate. Plant Mol Biol. 1995;27:933–942.

[28] Xu, J. Zhang S. Mitogen-activated protein kinase cascades in signaling plant growth and development. Trends Plant Sci. 2015;20:56–64.

[29] Madhusudanan KN, Nandakumar S. Carbohydrate changes in shoot tip and subtending leaves during ontogenetic development of pineapple. Z Pflanzenphysiol. 1983;110:429–438.

[30] Ohto M, Onai K, Furukawa Y, et al. Effects of sugar on vegetative development and floral transition in Arabidopsis. Plant Physiol. 2001;127:252–261.

[31] Martínez C, Pons E, Prats G, et al. Salicylic acid regulates flowering time and links defence responses and reproductive development. Plant J. 2004;37:209–217.

[32] Liu YY, Yu FM, Li YJ, et al. The effects of salicylic acid, boric acid and KH2PO4 treatment on Freesia refracta growth and development. J Hebei Normal Univ Sci Technol. 2004;18:68–72.

[33] Ren HX, Chen X, Zhao XJ, et al. Effects of nitrogen nutrition and salicylic acid on floral bud initiation in cucumber cotyledon culture. Acta Hortic Sin. 1999;26:105–109.

[34] Weisshaar B, Jenkins GL. Phenylpropanoid biosynthesis and its regulation. Curr Opin Plant Biol. 1998;1:251–257.

[35] Dooner HK, Robbins TP, Jorgensen RA. Genetic and developmental control of anthocyanin biosynthesis. Annu Rev Genet. 1991;25:173–199.

[36] Koes RE, Quattrocchio F, Mol JNM. The flavonoid biosynthetic pathway in plants: function and evolution. BioEssays. 1994;16:123–132.

[37] Dixon RA, Paiva NL. Stress-induced phenylpropanoid metabolism. Plant Cell. 1995;7:1085–1097.

[38] Taylor LP, Grotewold E. Flavonoids as developmental regulators. Curr Opin Plant Biol. 2005;8:317–323.

[39] Treutter D. Significance of flavonoids in plant resistance and enhancement of their biosynthesis. Plant Biol. 2005;7:581–591.

[40] Reddy KS, Menary RC. Vegetative growth, flowering and leaf nutrient concentration of boronia as affected by nitrogen level and form. Sci Hortic. 1989;40:335–344.

[41] Fernandez-Escobar R, Ortiz-Urquiza A, Prado B, et al. Nitrogen status influence on olive tree flower quality and ovule longevity. Environ Exp Bot. 2008;64:113–119.

[42] Zhang YR. The study of nitrogen nutrition regulation of flowering induction related gene expression in Arabidopsis thaliana [dissertation]. Hangzhou: Zhejiang University; 2011.

[43] Ouyang J, Li JY. Progress in plant tryptophan and indole-3-acetic acid biosynthesis. Adv Biol Eng. 1998;18:2–11.

[44] Wada KC, Yamada M, Shiraya T, et al. Salicylic acid and the flowering gene FLOWERING LOCUS T homolog are involved in poor-nutrition stress-induced flowering of Pharbitis nil. J Plant Physiol. 2010;167:447–452.

[45] Pichersky E, Dudareva N. Scent engineering: toward the goal of controlling how flowers smell. Trends Biotechnol. 2007;25:105–110.

[46] Van Schie CC, Haring MA, Schuurink RC. Regulation of terpenoid and benzenoid production in flowers. Curr Opin Plant Biol. 2006;9:203–208.