Plant Mol Biol (2014) 86:425–442
DOI 10.1007/s11103-014-0238-1

Analysis of global gene expression profiles to identify differentially expressed genes critical for embryo development in *Brassica rapa*

Yu Zhang · Lifang Peng · Ya Wu · Yanyue Shen · Xiaoming Wu · Jianbo Wang

Received: 21 May 2014 / Accepted: 12 August 2014 / Published online: 12 September 2014

Abstract Embryo development represents a crucial developmental period in the life cycle of flowering plants. To gain insights into the genetic programs that control embryo development in *Brassica rapa* L., RNA sequencing technology was used to perform transcriptome profiling analysis of *B. rapa* developing embryos. The results generated 42,906,229 sequence reads aligned with 32,941 genes. In total, 27,760, 28,871, 28,384, and 25,653 genes were identified from embryos at globular, heart, early cotyledon, and mature developmental stages, respectively, and analysis between stages revealed a subset of stage-specific genes. We next investigated 9,884 differentially expressed genes with more than fivefold changes in expression and false discovery rate ≤0.001 from three adjacent-stage comparisons; 1,514, 3,831, and 6,633 genes were detected between globular and heart stage embryo libraries, heart stage and early cotyledon stage, and early cotyledon and mature stage, respectively. Large numbers of genes related to cellular process, metabolism process, response to stimulus, and biological process were expressed during the early and middle stages of embryo development. Fatty acid biosynthesis, biosynthesis of secondary metabolites, and photosynthesis-related genes were expressed predominantly in embryos at the middle stage. Genes for lipid metabolism and storage proteins were highly expressed in the middle and late stages of embryo development. We also identified 911 transcription factor genes that show differential expression across embryo developmental stages. These results increase our understanding of the complex molecular and cellular events during embryo development in *B. rapa* and provide a foundation for future studies on other oilseed crops.

Keywords *Brassica rapa* L. · Embryo development · Transcriptome · Transcription factor · RNA-Seq

Introduction

Embryo development is a critical developmental period in the life cycle of flowering plants. It begins with a single zygote cell and proceeds through a well-coordinated series of cell division and differentiation to form a mature embryo. In many dicotyledonous plants (including *Arabidopsis*), embryo development consists of two major phases: morphogenesis and maturation (Braybrook and Harada 2008; Park and Harada 2008). Morphogenesis involves establishment of the embryonic body plan with primary shoot and root apical meristem (SAM and RAM, respectively) at the opposite ends, as well as cotyledons and a hypocotyl. Maturation involves cell expansion and accumulation of important storage reserves (usually carbohydrates, storage proteins, and oils) (Braybrook and Harada 2008; Park and Harada 2008). Numerous studies using the model system *Arabidopsis* have explored the molecular frameworks that control these developmental processes, and several key genes and biologically active molecules have been identified (Lau et al. 2012; Wendrich and Weijers 2013).

Electronic supplementary material The online version of this article (doi:10.1007/s11103-014-0238-1) contains supplementary material, which is available to authorized users.

Y. Zhang · L. Peng · Y. Wu · Y. Shen · J. Wang (✉)
State Key Laboratory of Hybrid Rice, College of Life Sciences, Wuhan University, Wuhan 430072, China
e-mail: jbwang@whu.edu.cn

X. Wu
Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Oil Crops Research Institute of CAAS, Wuhan 430062, China

The auxin-dependent pathway initiates apical/basal axis establishment immediately after zygote division (Lau et al. 2012; Rademacher et al. 2012), and the expression of *WUSCHEL* initiates the formation of the SAM. The specification of the SAM is regulated through a positive–negative feedback loop involving the *WUS* and *CLAVATA* genes (Lau et al. 2012; Machida et al. 2013). During the globular stage, the root tissues and their stem cells are specified (Wendrich and Weijers 2013). One of the best-known genes that play a role in specification of root cell types is the *AUXIN RESPONSE FACTOR5/MONOPTEROS* (*MP*), which is known to be a key regulator in embryonic root specification. Mutations in the *MP* gene result in rootless seedlings (Perilli et al. 2012; Wendrich and Weijers 2013). Another gene involved in RAM initiation and root meristem maintenance is *PLETHORA* (*PLT*); *plt* mutants contain seedlings completely lacking the hypocotyl and root (Aida et al. 2004). *SHOOT MERISTEMLESS* (*STM*) expression is required to maintain the SAM, and shoot meristem initiation is consistently inhibited in *stm* mutants. The regulation of *STM* expression in SAM boundaries by the *CUP-SHAPED COTYLEDON* genes along with auxin transport results in separation of the cotyledon primordia (Bowman and Eshed 2000; Lau et al. 2012). When the cotyledon primordia begins to emerge in *Arabidopsis*, the embryonic organization shifts from radial to bilateral symmetry and the embryo proceeds from the global stage to heart stage. *PINOID* (*PID*) is known to play roles in cotyledon development and the *pid wag1 wag2* triple mutant and the *pid enp* double mutant lack cotyledons (Lau et al. 2012). The embryonic body plan is established during the torpedo stage. *LEAFY COTYLEDON 1* (*LEC1*), *LEC2*, *FUSCA3* (*FUS3*) and *ABSCISIC ACID INSENSITIVE 3* (*ABI3*), which are intimately associated with abscisic acid (ABA) and gibberellic acid (GA) signaling, induce the maturation characteristics and activate the expression of seed storage protein genes during the maturation phase (Braybrook and Harada 2008; Sreenivasulu and Wobus 2013).

Analyses of the above genes have led to the identification and characterization of several genes with key roles in plant embryo development. However, embryo development from the action of a coordinated network of genes is a complex process, and mutagenesis alone cannot be used to identify all genes that are potentially involved. To increase our understanding of embryonic development, transcriptomic analysis can be used to identify differentially expressed genes at the whole-genome level.

Brassica rapa L., a cultivated species of the Brassicaceae family, is important for agriculture and human nutrition. Similar to *Arabidopsis*, the cellularized endosperm acts as nourishing tissue consumed by the embryo during maturation events. The primary storage products (lipids and proteins) accumulate in the cotyledons of the mature embryos (Dong et al. 2004; Sreenivasulu and Wobus 2013). With the completion of the draft genome of *B. rapa* (http://brassicadb.org/brad/) (Wang et al. 2011), it becomes feasible to further examination of the transcriptional activity of all of the functional elements in the *B. rapa* genome (Tong et al. 2013). RNA-sequencing (RNA-Seq) is a novel, high-throughput, deep sequencing technology commonly used for genomic research that provides novel strategies to analyze the functional complexity of transcriptomes (Mortazavi et al. 2008). Compared with microarrays and other sequence-based approaches using DNA libraries, the RNA-Seq method provides information on genome-wide gene expression and has a low background signal, a large dynamic range of expression levels, more accurate quantification, and high levels of reproducibility (Mortazavi et al. 2008). RNA-Seq has been used to investigate transcriptomes during plant embryo development, such as rice (Xu et al. 2012), maize (Lu et al. 2013), and cotton (Yang et al. 2012). Recently, the cDNA-amplified fragment length polymorphism (cDNA-AFLP) technique (Li et al. 2011), the analysis of expressed sequence tags (Venglat et al. 2013), and microarrays (Basnet et al. 2013) have been used to examine the expression patterns of embryo or seed genes of *Brassica* to provide information on transcriptome dynamics during embryo development.

In our report, we present the transcriptome atlas at key embryo stages (globular stage, heart stage, early cotyledon stage, and mature stage) during the process of morphogenesis and storage accumulation using deep RNA sequencing (the Illumina RNA-Seq) to rapidly identify and analyze the global expression data of *B. rapa* embryos. These results provide a comprehensive view of the transcriptome for embryo development in *B. rapa*, which will increase our understanding of the complex molecular and cellular events in *B. rapa* embryo development and provide a foundation for future studies on embryo development in *B. rapa* and other oilseeds.

Materials and methods

Plant materials and RNA isolation

Plants of *Brassica rapa* L. (cv. Qingyuan) were grown in soil under ambient conditions (Wuhan, China). Individual flowers on the primary inflorescence were bagged to prevent pollen contamination before flowering and hand pollinated to allow the formation of fruits. Siliques at 9, 13, 20 and 35 days after pollination were harvested from 15 individual plants at embryo-comparable developmental stages according to morphological changes of embryonic development: globular embryo, heart stage, early cotyledon stage, and mature embryo (abbreviate as G, H, E and M in Figures). Embryos from the four stages were taken from

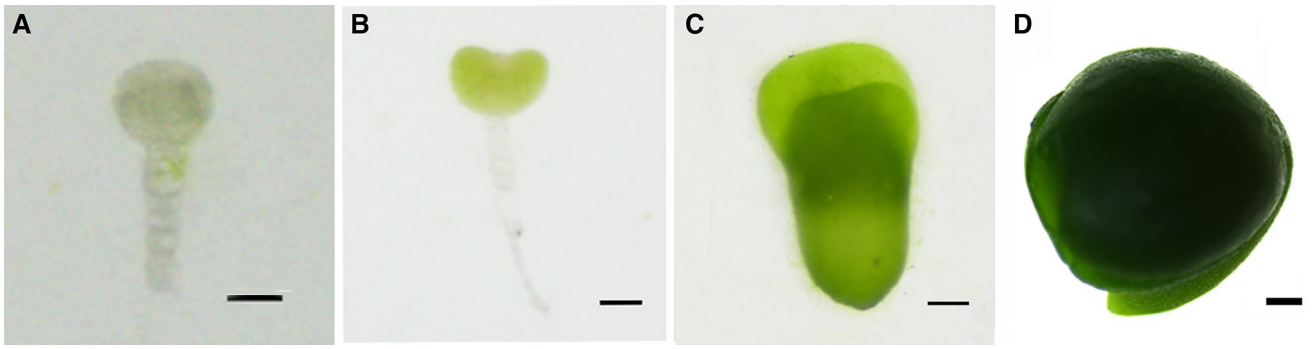

Fig. 1 Morphological changes occur during embryogenesis of *Brassica rapa* L. **a** Globular embryo; **b** Heart stage embryo; **c** Early cotyledon stage embryo; **d** Mature embryo. *Bars* represent 100 μm

developing seeds with microdissection needles under a dissection microscope (Olympus, Tokyo, Japan) (Fig. 1). The materials were immediately frozen in liquid nitrogen and kept at −80°C for total RNA extraction.

RNA sequencing library construction, Illumina sequencing and data processing

For Illumina sequencing, total RNAs were isolated from four-staged embryos using TRIzol Reagent (Invitrogen, Burlington, ON, Canada) according to the manufacturer's instructions and treated with RNase-free DNase I (Fermentas, Canada) to remove any genomic DNA. After accessing the yield and purity of RNA, the RNA sequencing libraries were constructed and sequenced by using Illumina HiSeq™2000. Illumina sequencing, data processing and mapping reads to reference genome were performed as previously described (Zhao et al. 2013).

Normalized expression levels of genes from RNA-Seq and gene annotation

For gene expression analysis, ERANGE software (version 4.0) (http://woldlab.caltech.edu/gitweb/) was used to calculate the normalized gene locus expression level by quantifying the number of reads that were mapped to the *B. rapa* genome sequences. The expression level of a gene from RNA-Seq was normalized by the RPKM (reads per kb per million mapped reads) method, which eliminates the influence of different gene length and sequence discrepancy on the calculation of gene expression (Mortazavi et al. 2008). Gene annotation including GO and KEGG annotation were also referred as Zhao et al. (2013).

Screening of differentially expressed genes (DEGs)

The R package DEGseq was applied to identify DEGs with the random sampling model based on the read count for each gene at different developmental stages (Wang et al. 2010). The false discovery rate (FDR) was used to determine the threshold of *P* value (Benjamini and Yekutieli 2001). A combination of FDR $\leq$ 0.001 and the absolute value of $\log_2$ Ratio $\geq$ 1 were used as the threshold to judge the significance of gene expression difference. More stringent criteria with a smaller FDR and bigger fold-change value can be used to identify DEGs. In our study, we defined genes with fold change $\geq$5 and FDR $\leq$ 0.001 as DEGs. GO term enrichment was performed on the DEGs. To group DEGs with similar expression patterns, a hierarchical clustering was generated using the expression values from each library. The analysis was conducted using Cluster 3.0 software with Pearson correlation as the distance measure. The cluster tree contained distinct clusters which contained genes with a unique expression profile by visual inspection. For pathway analysis, we mapped all DEG with more than five folds differential representation to terms in the KEGG database, and then looked for significantly enriched pathway terms compared to the genome background. For a graphical overview of pathways of metabolism and regulation, we used the MapMan tool as described as Gao et al. (2013).

Quantitative real-time PCR (qRT-PCR)

To validate the genes from Illumina sequencing, nine randomly chosen genes (Bra001933, Bra026375, Bra038483, Bra028162, Bra016269, Bra010474, Bra017113, Bra000321 and Bra005425) were subjected to qRT-PCR. Primer sets were designed with the Primer5 software and the sequences of primers were listed in Supplemental Table 1. QRT-PCR was carried out using SYBR Green I as a fluorescent detection dye and they were performed on an ABI StepOne™ Real-Time PCR System (Applied Biosystems). *ACT2/7* (Girin et al. 2010) was used as internal reference control to standardize the results and statistical analysis was performed using the $2^{-\Delta\Delta CT}$ method. The reported values were

averages of four independent trials (two biological replicates and two technical replicates). All data were expressed as the mean SD after normalization.

Results

Illumina sequencing and data analysis

In this study, four cDNA libraries from globular, heart, early cotyledon, and mature-stage embryos, respectively (see Materials and methods section for detailed library information and Fig. 1), were constructed using total RNA from *B. rapa* L. (cv. Qingyuan) embryos, and a large amount of data was generated with the Illumina HiSeq 2000 platform. Based on Illumina sequencing, 43,752,120 sequence reads were obtained from the four libraries of *B. rapa* embryos. After removing the adapter sequences, duplicate sequences, ambiguous reads, and the low-quality reads, 7,034,705 (globular embryo), 11,727,228 (heart embryo), 12,270,433 (early cotyledon embryo), and 11,873,863 (mature embryo) clean reads were generated. Each library represented by more than 7 million reads, reaching the saturation level of gene identification and a tag density sufficient for the quantitative analysis of gene expression. The clean reads were mapped to the Brassica rapa Genome Database (http://brassica.org/brad/) with no more than two base mismatches allowed in the alignment using SOAPaligner/soap2 software. Of the total clean reads, 69.24 % matched to unique genomic locations, and the uniquely matched reads were used for gene expression analysis of each library. An overview of the mapped statistics is provided in Table 1. To evaluate the RNA-Seq data, we analyzed the distribution of gene coverage in each library, which is the percentage of a gene covered by reads. This value is determined as the ratio of the number of bases in a gene covered by unique mapping reads to the total number of bases in that gene, which evaluates the normality of the sequence data. Genes with more than 90 % of their sequence covered by reads were grouped into the most abundant category, accounting for 33, 38, 43, and 22 % of the matched genes in globular, heart, early cotyledon, and mature stage embryos, respectively. The second most abundant category was genes with 80–90 % sequence coverage, and the percentages of matched genes were similar among the other eight categories (Supplemental Fig. 1).

Gene expression and functional classification of all detected genes

Our RNA-Seq data generated 32,941 genes accounting for approximately 80 % (32,941/41,174) of the annotated genes in the *B. rapa* genome. To obtain statistical confirmation of gene expression among the developmental stages, we used the reads per kilobase per million mapped reads (RPKM) method to normalize the gene expression levels. This method allows gene expression levels to be compared both within and between samples (Marioni et al. 2008), which can avoid biases caused by different gene exon sizes and the different number of read sequences obtained from the four libraries. These uniquely mapped reads were used to calculate the gene RPKM values. Gene expression patterns during four embryo stages are summarized in Supplemental Table 2. In total, 27,760, 28,871, 28,384, and 25,653 genes (ranging from 100 to ≥2,000 bp) were identified from embryos in the globular, heart, early cotyledon, and mature stages, respectively. As shown in Table 2, the proportion of longer assembled sequences was higher among these genes, and the lengths of the 17,716 genes exceeded 1,000 bp (Table 2). A Venn diagram showed the number of genes uniquely expressed in each stage or genes that were shared between one or more other stages. Of these genes, 22,507 of the expressed genes were represented in all four stages and 3,694 were expressed specifically in a single stage. We also compared the number

Table 1 Summary of alignment statistics of RNA-Seq in four libraries of *B. rapa* developing embryos referring to *B. rapa* genome

	Globular	Heart	Early cotyledon	Mature
Total reads	7,034,705	11,727,228	12,270,433	11,873,863
Total mapped reads	5,410,756 (76.92 %)	8,935,253 (76.19 %)	9,716,679 (79.19 %)	10,375,239 (87.38 %)
Unique match	4,776,935 (67.91 %)	8,163,687 (69.61 %)	8,782,663 (71.58 %)	8,058,603 (67.87 %)
Multi-position match	633,821 (9.01 %)	771,566 (6.58 %)	934,016 (7.61 %)	2,316,636 (19.51 %)
Unmapped reads	1,623,949 (23.08 %)	2,791,975 (23.81 %)	2,553,754 (20.81 %)	1,498,624 (12.62 %)

Table 2 Distribution of the gene sequence length of all 32,941 genes detected in *B. rapa* embryo development via RNA-Seq technology

Gene length (bp)	Total number	Percentage
100–500	5,149	15.63
500–1,000	10,076	30.59
1,000–1,500	8,467	25.70
1,500–2,000	4,443	13.49
≥2,000	4,806	14.59
Total	32,941	100

For these genes, the gene sequence length distribution can be divided into five grades. The total number and the percentage of all genes are presented in this table. The proportion of longer assembled sequences was higher among these genes, and the lengths of the 17,716 genes exceeded 1,000 bp

of genes exclusively expressed in two adjacent embryonic stages. The number of genes simultaneously expressed in the early cotyledon embryo stage and mature embryo stage was lower than that expressed simultaneously in the other two adjacent embryos stages. Differences in the expression of co-expressed genes were of interest to determine how they change throughout embryo development. Moreover, specifically expressed genes were of interest because of their potential importance at the corresponding stage. The detailed statistical analysis of identified genes between all four samples in various combinations is shown in Fig. 2.

To facilitate the global analysis of gene expression, all predicted genes were assigned to different functional categories using Blast2GO (version 2.3.5) (http://www.blast2go.org/). The annotations were verified manually and integrated using gene ontology (GO) classification (http://www.geneontology.org). Of the 32,941 detected genes, 25,704 (78.03 %) had at least one GO annotation based on sequence similarity. We categorized these annotated genes according to the secondary classification of GO terms, assigning genes to 42 functional groups of the three main GO classification categories (Fig. 3). For the three main categories of GO classification, the assignments to "biological process" (61,146, 46.3 %) accounted for the majority of genes, followed by "cellular component" (42,848, 32.45 %) and "molecular function" (28,065, 21.25 %). Among them, the terms cellular process (GO: 0009987) and metabolic process (GO: 0008152), which included 13,277 genes (40.3 %) and 13,228 genes (40.2 %), respectively, were dominant in the biological process category, indicating that embryos from the four developmental stages were metabolically active. The terms "response to stimulus" (GO: 0050896; 7,249 genes; 22.0 %) and "developmental process" (GO: 0032502; 3,702; 11.2 %) were also common, suggestive of protective mechanisms against potential external and/or internal stresses during embryonic development. In the cellular component category, "cell" (GO: 000555623), "cell part" (GO: 0044464), and "organelle" (GO: 0043226) were prominent groups. In the molecular function category, the term "binding" (GO: 0005488; 13,193; 40.1 %) and "catalytic activity" (GO: 0003824; 11,325; 34.4 %) were most highly represented. Only a few genes were clustered in terms of "auxiliary transport protein activity", "metallochaperone activity", "electron carrier activity", "biological adhesion", and "cell killing".

Pathway-based analysis can be used to examine the biological functions and interactions of genes. To perform functional classification and pathway assignment of genes during *B. rapa* embryo development, we mapped the detected genes to reference canonical pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.ad.jp/kegg/). As a result, of the 32,941 detected genes, 22,060 were assigned to 128 KEGG

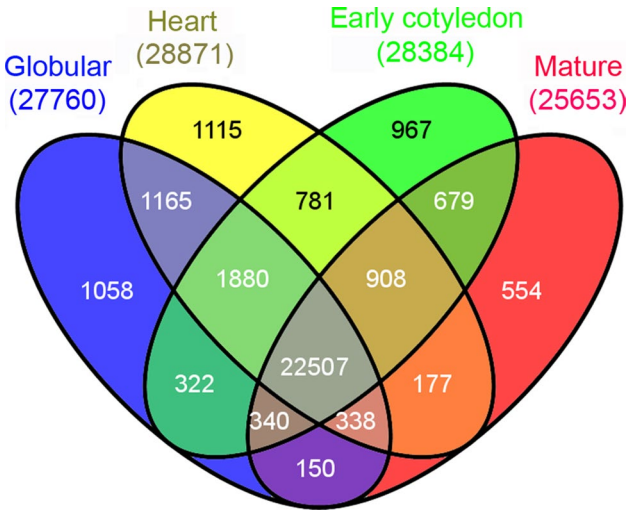

Fig. 2 Venn diagram showing the genes expressed in each of the four stages of *B. rapa* embryo development. Among these genes, 22,507 genes are expressed at all four developmental stages. The number of stage-specifically expressed genes is 1,058 (Globular), 1,115 (Heart), 967 (Early cotyledon) and 554 (Mature), respectively

pathways (Supplemental Table 3). Metabolism pathway was the largest category (4,953; 22.45 %), followed by biosynthesis of secondary metabolites (2,760; 12.51 %), plant–pathogen interaction (1,663; 7.54 %), and plant hormone signal transduction (1,540; 6.98 %). These results indicated that embryo development involves genetic and active metabolic processes, and the functional classification of KEGG provided valuable information for investigating specific processes, functions, and pathways during *B. rapa* embryo development.

Identification of differentially expressed genes

Analysis of differentially expressed genes (DEG) between samples can be accomplished using RNA-Seq. In this study, DEG was defined as the fold change of the normalized (RPKM) expression values, and was at least fivefold in either direction when the absolute value of the $\log_2$ ratio ≥ 2.322 and false discovery rate (FDR) ≤ 0.001. As a result, a set of 9,884 genes exhibited significant differential expression between two consecutive developmental stages (Supplemental Table 4). Comparing globular and heart stage embryo libraries, 1,514 genes were modulated, of which 687 genes were up-regulated and 827 genes were down-regulated, respectively (Fig. 4 and Supplemental Table 5). Compared with the heart stage, the early cotyledon stage embryo library had 2,182 up-regulated genes and 1,649 down-regulated genes (Fig. 4 and Supplemental Table 6). In the mature stage relative to the early cotyledon stage, 6,266 genes were down-regulated, while only 327 genes were up-regulated. The trend of gene expression was down-regulation, perhaps

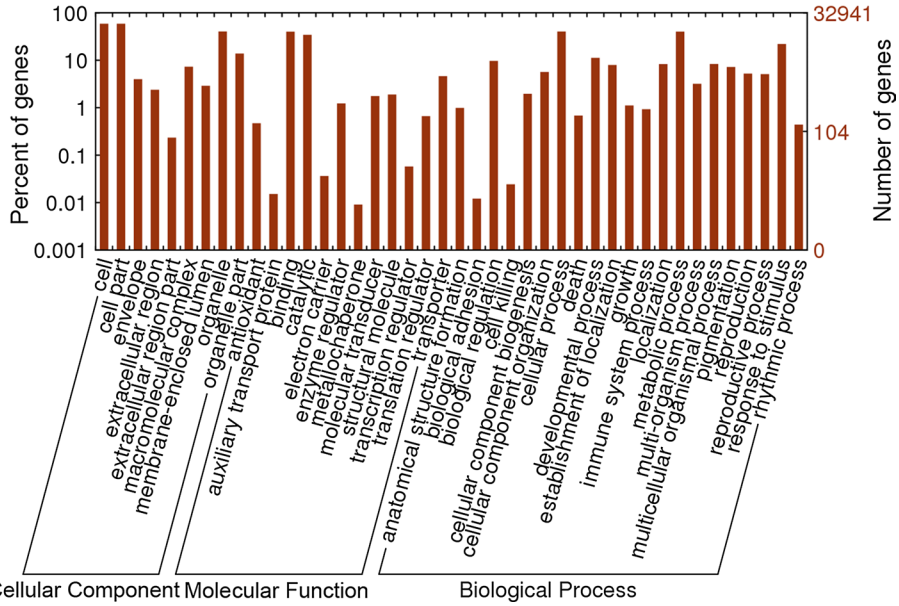

Fig. 3 Gene ontology classification of all detected genes. The results are summarized in three main GO categories (cellular component, molecular function and biological process) and 42 sub-categories. Cell (GO: 0005623), with 19,378 genes, are dominant in the main categories of cellular component. Binding (GO: 0005488) and cellular process (GO: 0009987) consisted of 13,193 and 13,277 genes, are dominant in the main categories of molecular function and biological process, respectively. The x-axis represents the name of the GO sub-categories. The right y-axis indicates the number of genes expressed in a given sub-category. The left y-axis indicates log(10) scale, the percent of a specific category of genes in that main category

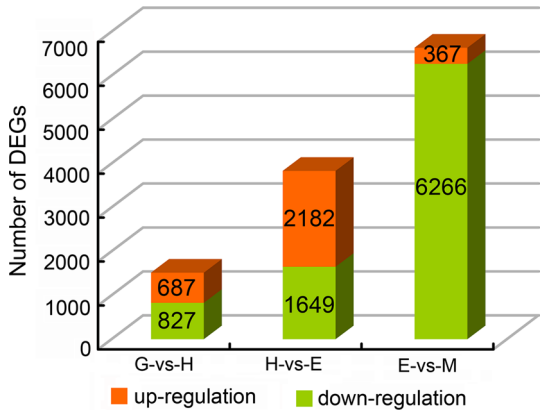

Fig. 4 Differentially expressed genes detected between two adjacent embryonic stages (G-vs-H, H-vs-E and E-vs-M). The numbers of up-regulated and down-regulated genes between two adjacent embryonic stages are revealed

reflecting the molecular mechanisms that control both the ceasing of embryo growth and entrance into the maturation phase (Fig. 4 and Supplemental Table 7).

Expression pattern analysis of DEGs

We performed hierarchical clustering of the 9,884 DEGs to examine the similarity and diversity of expression profiles using Cluster 3.0 software. The RPKM value was transformed as the binary logarithm. In Fig. 5, the transformed values were visualized in different colors, with red representing up-regulation and green representing down-regulation. The hierarchical clustering generated a global view of the gene expression pattern of DEGs, indicating that the majority of genes were up-regulated during the early and middle stages of embryo development (G-vs-H or/and H-vs-E), and down-regulated in E-vs-M. Correlations between embryos at the adjacent stages were investigated, showing that the globular to heart stage embryos had a closer relationship with heart stage to early cotyledon stage embryos in gene expression than with the embryos at other time points.

We then categorized 9,884 genes showing differential expression from the three comparisons into eight patterns (Fig. 6 and Supplemental Table 4) to identify similar expression patterns in all DEGs across the stages. Among the 9,884 DEGs, pattern 1 contained 331 (1 %) genes whose expression level was up-regulated continuously from the globular stage to mature stage. Pattern 2 consisted of 1,471 (15 %) genes whose expression level peaked at the globular embryo stage and was negatively modulated throughout all four developmental stages. These genes may contribute to embryo morphogenesis. Pattern 3 comprised 2,519 (25 %) genes up-regulated from the globular stage until the early cotyledon stage and was down-regulated

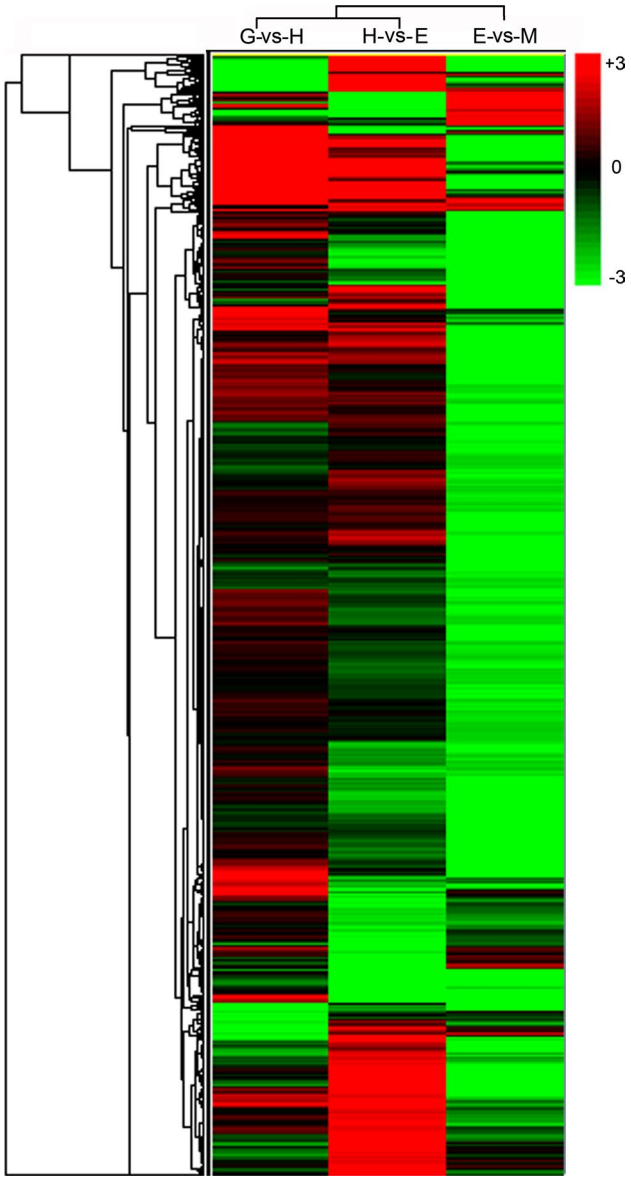

Fig. 5 Hierarchical cluster analysis of 9,884 DEGs based on log ratio RPKM data. The cluster display expression patterns for a subset of 9,884 DEGs in each of the three comparisons (G-vs-H, H-vs-E and E-vs-M). The *color key* represents RPKM normalized $\log_2$ transformed counts. *Red* represents up-regulation expression and *green* represents down-regulation expression. *Each column* represents an experimental condition (e.g. G-vs-H), *each row* represents a gene

thereafter. 442 (4 %) genes up-regulated from the globular to heart stage, down-regulated from the heart to early cotyledon stage, and then up-regulated from the early cotyledon to mature stage were defined as pattern 4. Pattern 5 contained the most genes (2,661; 27 %), with expression levels up-regulated from the globular stage, peaking at the heart stage, and then down-regulated until the mature stage. Genes that were down-regulated from the globular to heart stage and then up-regulated until the mature stage were classified into pattern 6, which consisted of 470 (5 %) genes. Pattern 7 contained 1,555 (16 %) genes that were down-regulated from the globular to heart stage and early cotyledon to mature stage and up-regulated from the heart to early cotyledon stage. Pattern 8 consisted of 435 (4 %) genes that were down-regulated from the globular to early cotyledon stage and up-regulated from the early cotyledon to mature stage (Fig. 6 and Supplemental Table 4). These expression patterns were suggestive of diverse and complex interactions during embryo development and regulation.

GO analysis of DEGs

To further examine the gene functional differences during *B. rapa* embryo development, DEGs were functionally annotated. All DEGs were mapped to GO terms in the database (http://www.geneontology.org/), calculating gene numbers for every term and then using a hypergeometric test to identify significantly enriched GO terms in DEGs compared to the genomic background. GO enrichment analysis for 9,884 genes is shown in Supplemental Table 8. Using GO functional classification analysis (WEGO), we categorized the DEGs between the globular and heart stage, heart and early cotyledon stage, and early cotyledon and mature according to the secondary classification of GO terms. Nine, 12, and 19 functional groups were in the cellular component, molecular function, and biological process category, respectively (Fig. 7a). In the biological process category, "cellular process" and "metabolic process" were the dominant groups, indicating that extensive metabolic activities occurred in *B. rapa* embryos during all four stages. In the molecular function category, "binding" and "catalytic processes" were highly represented, while in the cellular components category, the terms "cell" and "cell part" were the dominant categories (Fig. 7a). GO analysis of up-regulated and down-regulated DEGs in the three comparisons are shown in Fig. 7b–d, respectively. As shown in Fig. 7b, in the biological process category, a total of 11 GO categories showed statistically significant differences between globular and heart embryo, including "anatomical structural formation", "biological regulation", "metabolic process", and eight other subcategories. Although the number of down-regulated genes was higher than up-regulated genes in most functional groups, the number of up-regulated genes was dominant in the functional groups mentioned above between globular and heart (Fig. 7b). In Fig. 7c, analysis of GO terms in the DEGs in heart and early cotyledon showed that 10 subcategory biological processes, including "response to stimulus", "developmental process", "multicellular organismal process", and seven other subcategories, differed significantly in terms of gene percentage. In Fig. 7d, among the 19 functional groups of biological process from early cotyledon

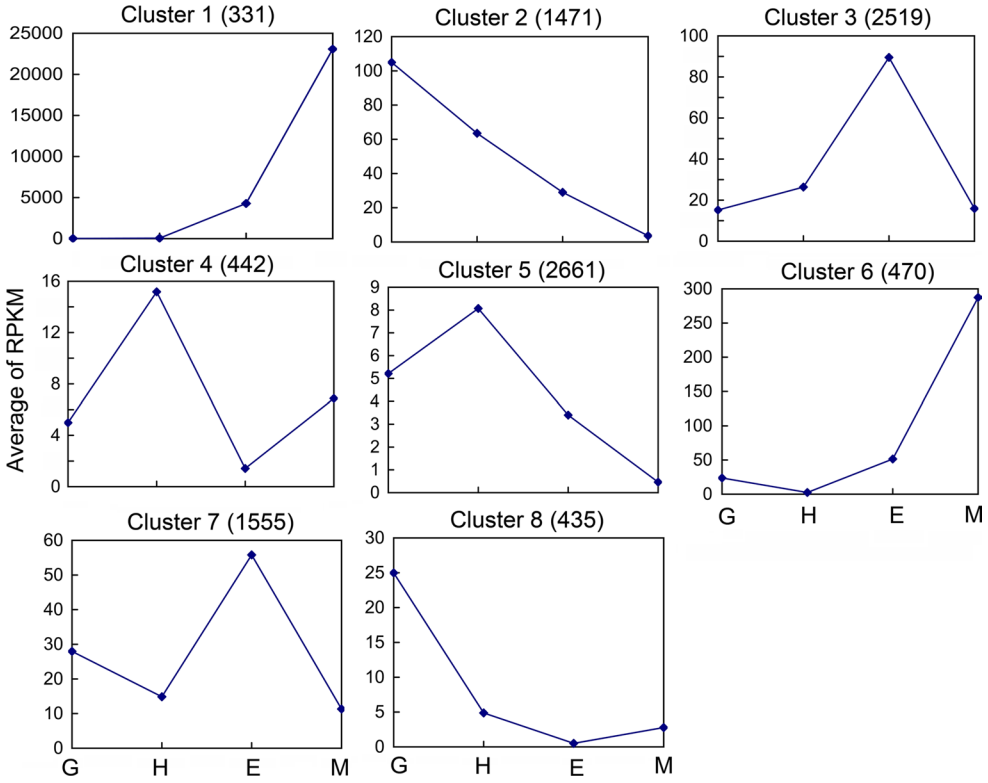

Fig. 6 The clustering of expression patterns of the 9,884 DEGs. The x-axis indicates the embryo developmental stage (*G* globular stage, *H* heart stage, *E* early cotyledon stage, *M* mature stage). The y-axis indicates the average gene-expression levels of genes in each cluster. The total numbers within a cluster are shown in parentheses

to mature, GO terms were predominantly associated with cellular process, metabolic process, biological regulation, cellular component organization, pigmentation, and reproduction. These GO annotations demonstrated that the *B. rapa* embryo expressed genes encoding diverse structural, regulatory, and metabolic proteins.

KEGG and MapMan pathway analysis of DEGs

Gene interactions play an important role in biological functions. To identify genes involved in metabolic or signal transduction pathways that were significantly enriched, we mapped all DEGs to reference canonical pathways in the KEGG (http://www.genome.ad.jp/kegg/). By comparing with the whole transcriptome background, all DEGs were assigned to 127 pathways (Supplemental Table 9). Ten, 30, and 13 pathways which were significantly enriched ($Q \leq 0.05$) with DEGs are found in the three comparisons, respectively. Starch and sucrose metabolism, pentose and glucuronate interconversions, and metabolic pathways were significantly enriched with DEGs in globular-vs-heart stage embryos. The most significantly enriched pathways were biosynthesis of secondary metabolites, metabolic pathway, and limonene and pinene degradation in heart-vs-early cotyledon stage embryos. Between the early cotyledon and mature stage, photosynthesis and photosynthesis-antenna proteins pathways were significantly enriched only with down-regulated DEGs; ribosome, circadian rhythm plant, and glyoxylate and dicarboxylate metabolism pathways were significantly enrichment with down-regulated and up-regulated genes. Based on RPKM, we identified eight significant expression types among all pathways (Fig. 8 and Supplemental Table 9). Pathways in type 7 were the most abundant group, containing 42 pathways that were down-regulated between the globular stage to heart stage and early cotyledon stage to mature stage, and up-regulated from the heart stage to early cotyledon stage. Pathways in type 3 were the second most abundant group, comprising 31 pathways that were up-regulated from the globular stage until early cotyledon stage and then down-regulated from the early cotyledon stage to mature stage. Type 2 had 24 down-regulated pathways, indicated by a negative slope during embryo development. Twenty-one pathways in type 5 were up-regulated from the globular stage to heart stage and down-regulated until the mature stage. We noticed that fatty acid (FA) biosynthesis, glucosinolate biosynthesis, sesquiterpenoid and triterpenoid biosynthesis, and sulfur metabolism were positively modulated throughout all four stages, and they were classified as type 1. Carotenoid biosynthesis, hypotaurine metabolism, other glycan degradation, and taurine pathways in type 8 were down-regulated from the globular stage to early cotyledon stage and up-regulated from the early cotyledon to mature stage. Type 4 contained genes involved in the cutin, suberine, and wax biosynthesis pathways, which were up-regulated between the globular and heart stage, down-regulated between the

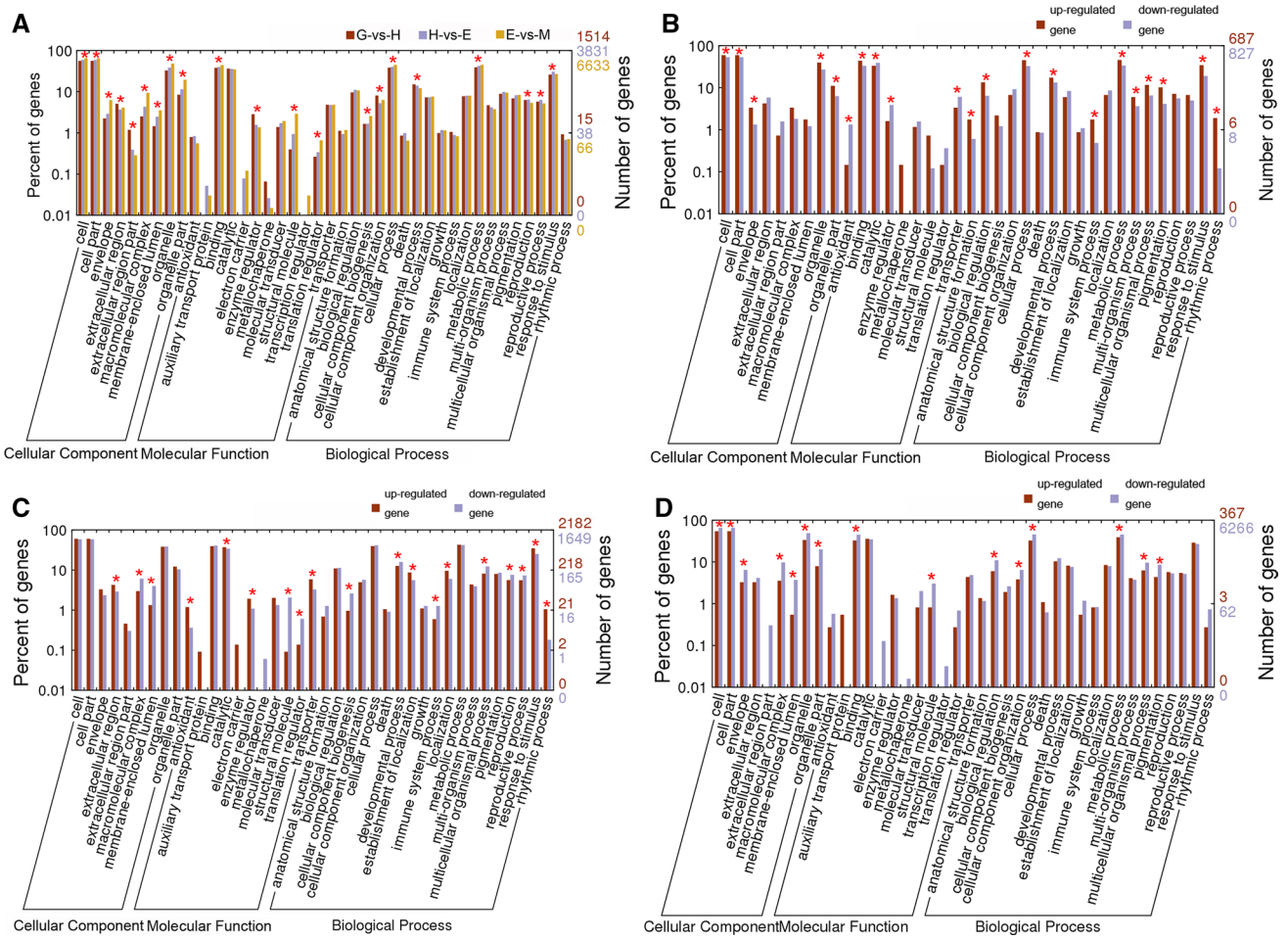

Fig. 7 GO classification of DEGs. The results are summarized in three main categories: biological process, molecular function and cellular component by GO analysis. **a** GO classification of all 9,884 DEGs among three contrasts (G-vs-H, H-vs-E and E-vs-M). **b** GO analysis of the up-regulated and down-regulated genes in G-vs-H. **c** GO analysis of the up-regulated and down-regulated genes in H-vs-E. **d** GO analysis of the up-regulated and down-regulated genes in E-vs-M. A statistically significant difference was determined if the *P* value based on gene percentage was below 0.05 ($*P \leq 0.05$)

heart and early cotyledon stage, and then up-regulated from the early cotyledon to mature stage. Finally, only the phenylpropanoid biosynthesis pathway (which belonged to type 6) was down-regulated from the globular stage to heart stage and then up-regulated until the mature stage (Fig. 8 and Supplemental Table 9). A summary of KEGG analyses is shown in Table 3. Numerous pathways were altered, providing a macroscopic view of the KEGG pathways analysis. Thus, the transcript profile varies substantially among the stages analyzed might correlate with the profound differences between physiological processes during *B. rapa* embryo development.

We next used the MapMan package (http://MapMan.gabipd.org) to investigate the pathways involved in *B. rapa* embryo development. The MapMan tool used input from several experts to curate specific biological processes using information from the Brassica Database (BRAD, http://brassicadb.org/brad/). Based on the classification of DEGs into 35 major pathways and 237 branch pathways, MapMan was originally used to visualize *B. rapa* embryo developmental pathways. An overview of DEGs between globular and mature stages in terms of cellular metabolism and regulation pathways is shown in Fig. 9. We customized the figure to depict the biological process of interest and displayed $\log_2$ RPKM-normalized expression counts onto pictorial diagrams. Based on the RPKM, 82 pathways were down-regulated in the mature stage compared to the globular stage. The downward trend in overall transcription is presented. Only 13 pathways, such as development (storage proteins), lipid metabolism (triacylglycerol synthesis), transport, redox, and ATP synthesis were up-regulated. The numbers of pathways in type 3 and 4 expression patterns were 34 and 3, respectively. A more detailed list of all DEGs corresponding to MapMan functional categories

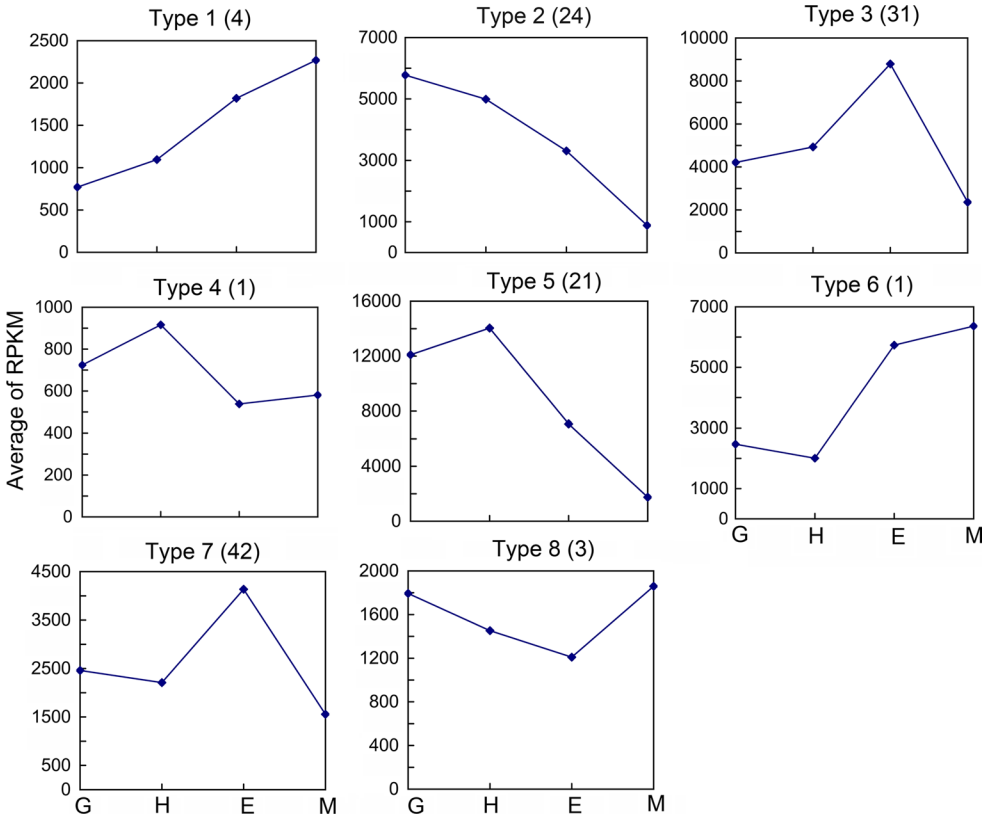

Fig. 8 Eight types of expression pattern of 127 KEGG pathways. The x-axis indicates the embryo developmental stage (*G* globular stage, *H* heart stage, *E* early cotyledon stage, *M* mature stage). The y-axis indicates the average expression levels of genes that belong to the KEGG functional classes in each type. The total numbers of KEGG functional class within a type are shown in parentheses

Table 3 The considerably changed KEGG pathways in the RNA-seq analysis

	Top 10 of expression (G + H+E + M)	Top 10 of number quantity	Top 10 of average value (expression/quantity)	Top 10 of up-regulated pathways (M-G)	Top 10 of down-regulated pathways (M-G)
1	Ribosome (520,638)	Metabolic pathways (2,229)	Plant hormone signal transduction (1,130.6)	Biosynthesis of unsaturated fatty acids (4,560)	Ribosome (−160,794)
2	Metabolic pathways (348,708.9)	Biosynthesis of secondary metabolites (1,112)	Ribosome biogenesis in eukaryotes (1,037)	Tryptophan metabolism (4,040.1)	Metabolic pathways (−44,039)
3	Biosynthesis of secondary metabolites (133,322.7)	Plant hormone signal transduction (600)	Photosynthesis- antenna proteins (540)	Phenylpropanoid biosynthesis (3,895.7)	RNA transport RNA (−20,668.5)
4	RNA transport RNA (60,966.4)	Plant-pathogen interaction (507)	Fatty acid elongation (475.9)	Fatty acid biosynthesis (3,385.3)	Spliceosome (−20,718.8)
5	Spliceosome (51,910.5)	Ribosome (502)	Biotin metabolism (384.4)	Cyanoamino acid metabolism (3,070.7)	Phagosome (−15,208.3)
6	Photosynthesis (48,599.5)	RNA transport RNA (285)	Phenylalanine metabolism (336.7)	Starch and sucrose metabolism (1,097.7)	Protein processing in endoplasmic reticulum (−12,376.3)
7	Protein processing in endoplasmic reticulum (48,599)	Starch and sucrose metabolism (275)	Histidine metabolism (306.2)	Glucosinolate biosynthesis (1,156.6)	Plant-pathogen interaction (−10,132.2)
8	Plant hormone signal transduction (35,618)	Spliceosome (222)	Carotenoid biosynthesis (299)	Sulfur metabolism (1,097.7)	Ribosome biogenesis in eukaryotes (−10,035.1)
9	Phagosome (34,683.5)	Protein processing in endoplasmic reticulum (214)	Non-homologous end-joining (282.5)	Galactose metabolism (571.7)	Proteasome (−8,710.9)
10	Oxidative phosphorylation (33,737.2)	Purine metabolism (170)	Protein export (275.7)	Taurine and hypotaurine metabolism (476.2)	Purine metabolism (−7,974.5)

All DEGs are assigned to 127 KEGG pathways. The top 10 of expression are the most abundant pathways according to the RPKM value sum of four libraries. Based on the RPKM, the top 10 of up-regulated pathways and the top 10 of down-regulated pathways can be determined. The figures in brackets indicate the RPKM value

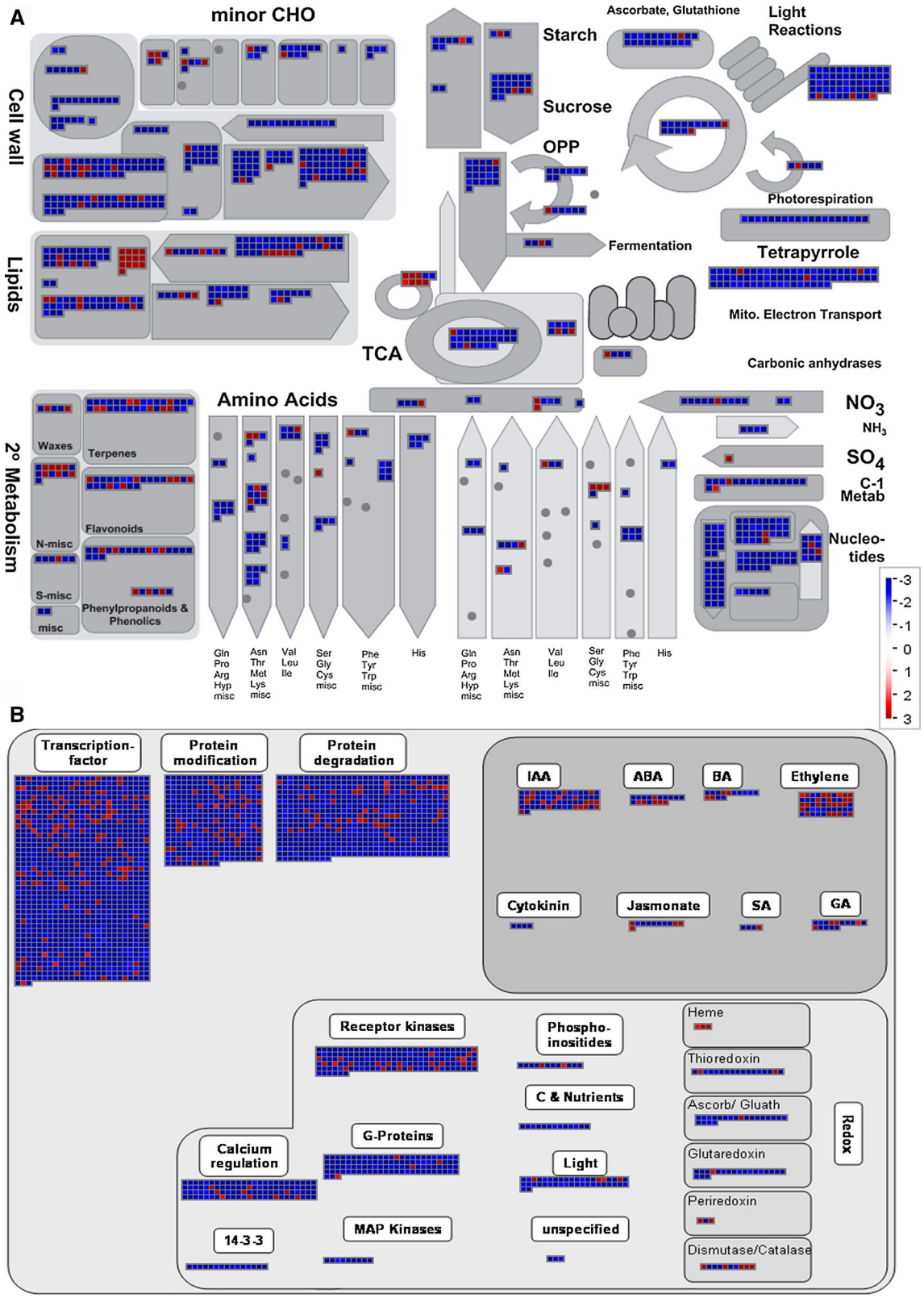

Fig. 9 MapMan overview of cellular metabolism (**a**) and regulation (**b**) showing all DEGs between globular and mature embryos. Individual genes are represented by *small squares*. The *color key* represents RPKM normalized $\log_2$ transformed counts. *Red* represents up-regulation and *blue* represents down regulation between globular and mature stage embryos

is provided in Supplemental Table 10. These visual annotations provide a valuable resource for the investigation of pathways in embryo development.

Transcription factor (TF) gene expression during *B. rapa* embryo development

To identify TF genes expressed during *B. rapa* embryo development, we queried *B. rapa* TF genes in the Plant and Transcription Factor Database (http://planttfdb.cbi.pku.edu.cn/index.php) and identified 2,502 putative TF genes out of 32,941 genes, which could be classified into 57 TF families. Based on RPKM, 911 genes of the 2,502 putative TFs showed significantly differential expression, accounting for 36.4 % of the total TF genes (Supplemental Table 11).

Based on RPKM, nine families were down-regulated throughout embryo development. The expression patterns of genes encoding the Nin-like family were classified as pattern 1, which were up-regulated throughout the entire developmental process. Eighteen families which were up-regulated from the globular until early cotyledon stage and then down-regulated at the mature stage, were classified as pattern 3. The LFY family was classified as pattern 4, which were only highly expressed during the globular and heart stage. YABBY (YAB) family and another 18 families were classified as pattern 5, which was up-regulated during the heart stage and then down-regulated until the mature stage. Nine families showed down-regulation from the globular to heart stage and up-regulation until the early cotyledon, followed again by down-regulation; this was classified as pattern 7 (Supplemental Table 12). Some TF families, such as the AP2, ARF, and MYB gene family will be discussed later in detail. A more in-depth analysis is required to determine how the above gene families are associated with the molecular and cellular changes that occur during *B. rapa* embryo development.

We also identified a set of DEGs specific to each stage, which encode TFs from a variety of different families. During the globular stage, three specific TF mRNAs were identified, including Bra011631 (C2H2 family), Bra013891 (MIKC family), and Bra023551 (Nin-like family). During the heart stage, we identified one specific TF, Bra006741, which was a member of the MYB family. Five specific expressed genes were identified from all DEGs during the early cotyledon stage, including Bra000245 (C2H2 family), Bra004035 (CO-like family), Bra017425 (G2-like family), and two MYB family members, Bra006811 and Bra015029. Bra002159, a member of the ERF family, was identified as the mature stage-specific TF from the DEGs.

Validation of the RNA-Seq data using quantitative real-time polymerase chain reaction (qRT-PCR)

To assess the validity and reliability of our RNA-Seq data, qRT-PCR analysis was performed using gene-specific primers for nine randomly chosen genes across the embryo development stages (Supplemental Table 1). The results are shown in Supplemental Fig. 2. Of the nine genes, Bra001933, encoding a hypothetical protein involved in the cell part pathway showed a high expression pattern at the globular stage, suggesting that it may play an important role in early embryo development. Bra028162, encoding a seed-specific protein, Bra016269, encoding a 60S ribosomal protein and Bra005425, encoding chlorophyll *a*/*b* binding protein, were all expressed at high levels during the heart stage. Bra000321, encoding an acyl-carrier protein, showed two peaked of expression at heart stage and mature stage. The expression of Bra010474, encoding oleosin, was low at the globular stage and then increased from the heart stage and peaked at the mature stage. Bra038483 and Bra026375, encoding storage proteins, as well as Bra017113, encoding a lipid transfer protein, showed low expression levels in the early stages, with their expression peaking during the mature stage with lipid and storage protein accumulation. Overall, these results demonstrated that all nine genes exhibited the same expression profiles as obtained from the original RNA-Seq results.

Discussion

DEGs that may play critical roles for morphogenesis in *B. rapa* embryos

Morphogenesis in seed plants commences with highly stereotypical cell division sequences and is under tight transcriptional control during early embryogenesis (Lau et al. 2012). Although a set of genes considered to be important regulators of embryo patterning encoding TFs or components of signaling pathways has been implicated in this process (Spencer et al. 2007; Le et al. 2010; Lau et al. 2012; Wendrich and Weijers 2013), a large gap in understanding exists between the transcriptional control of embryo morphogenesis and the cellular execution (Belmonte et al. 2013; Sreenivasulu and Wobus 2013; Wendrich and Weijers 2013). We used our results to increase our understanding of the gene expression changes during embryo morphogenesis. For example, DEGs were significantly enriched for the GO term “anatomical structural formation” from the globular to heart stage, which was consistent with the result that GO terms associated with patterning events, such as determination of bilateral symmetry and abaxial cell fate specification, were enriched during *Arabidopsis*

embryo development (Belmonte et al. 2013). During the anatomical structural formation process, DEGs encoding the toxin-coregulated pilus (TCP), flotillin-like protein, class III homeodomain leucine zipper (HD-ZIPIII), XND, GROWTH-REGULATING FACTOR (GRF), WOX, GODEN, YAB, BREVIS RADIX, and APETALA 2 (AP2) were up-regulated during the heart stage compared to the globular stage. TCP TFs control the morphology of shoot lateral organs via the regulation of other genes (Koyama et al. 2007). Two DEGs were shown to encode TCP13 (Bra039158 and Bra001032) and one to encode TCP4 (Bra021586) belonging to the TCP family. GRF is required for the development of leaves, cotyledons, and the SAM and enhances seed oil production by regulating cell number and plant photosynthesis (Kim and Lee 2006; Liu et al. 2012). GRF7 (Bra022667) and GRF5 (Bra001532) may play a role in phyllome development and response to red or far-red light processes; however, this requires further study. Brassinosteroids (BRs) are known to stimulate growth by promoting cell division and differentiation, and control overall developmental programs leading to morphogenesis (Vanstraelen and Benkova 2012). BREVIS RADIX (*BRX*) encodes a key regulator of cell proliferation and elongation in the root, which has been implicated in the BR pathway as well as in the regulation of auxin-responsive gene expression (Mouchel et al. 2006). Bra035521 encoding BRX was up-regulation from the globular stage to heart stage. AP2 TFs are involved in embryo initiation and development, and redundantly control embryo patterning through interactions with other genes (Ouakfaoui et al. 2010). Bra000487, encoding a RAP2-7-like protein, a member of the AP2 family, showed high expression at the heart stage, but strongly decreased thereafter. Aintegumenta (ANT) (Bra011782; Bra017852) and ANT2 (Bra010610) were involved in anatomical structure morphogenesis. ANT-like proteins have been reported to regulate embryogenesis, meristem development, organ initiation, and growth (Horstman et al. 2014).

The second group of genes are involved in adaxial/abaxial axis specification between globular and heart stage embryos. Adaxial/abaxial polarity in plants is specified by interactions between genes that individually specify either adaxial or abaxial identity (Fukushima and Hasebe 2014). In *Arabidopsis*, adaxial identity is specified by HD-ZIPIII TFs and the TF ASYMMETRIC LEAVES2 (AS2), whereas abaxial identity is specified by KANADI (KAN), YAB, and AUXIN RESPONSE FACTOR (ARF) (Izhaki and Bowman 2007; Fukushima and Hasebe 2014). Four DEGs (Bra003309, Bra004364, Bra020236, and Bra039733) involved in adaxial/abaxial axis specification were up-regulated from the globular stage to heart stage. Bra003309 (encoding a putative male sterility protein) and Bra004364 (encoding CRC protein) belong to the YAB family. Bra020236 and Bra03973, both encoding hypothetical proteins, belong to the HD-ZIPIII family and LBD family, respectively. In our study, REVOLUTA (Bra002458) and ATHB (Bra021926) showed high expression levels during the globular stage, which is consistent with the function of HD-ZIP III at the central–apical domain of early embryos being required for the correct specification of the SAM and RAM (Izhaki and Bowman 2007). LBD6/AS2 functions in the specification of adaxial/abaxial organ polarity and negatively regulates expression of KNOX TF genes in lateral organs (Wu et al. 2008). Bra039733 encoding a hypothetical gene belonging to the LBD family may play a role in adaxial specification. In total, 10 YAB genes were identified as genes involved in adaxial/abaxial axis specification, and nine genes were up-regulated from the globular to heart stage embryos. Total RPKM was highest at the heart stage of the four embryo stages.

Stem cell niches are often found in specialized compartments, such as in the embryonic SAM and the RAM, which can usually be divided into several regions that contain cells with distinct properties (Bowman et al. 2000; Perilli et al. 2012). However, stem cell niche specification remains relatively uncharacterized (Wendrich and Weijers 2013). In our study, nine genes were identified that play a role in "stem cell division" (GO: 0017145) and "stem cell maintenance" (GO: 0019827). Based on the stem cell division pathway, Bra004403 (encoding a no apical meristem domain containing transcriptional regulator) expression level observed was higher at the globular stage and could affect meristem formation. For stem cell maintenance, one gene (Bra033698) encoding a zwille-like protein 2 and another three (Bra034764, Bra038707, and Bra001446) encoding retinoblastoma-related proteins were down-regulated from the globular stage to mature stage. In *Arabidopsis*, retinoblastoma-related genes were found to regulate stem cell maintenance in roots (Wildwater et al. 2005). Consistent with differentiation of the apex, zwille-like AGO 1 showed high expression levels during early embryogenesis (Carmell et al. 2002). These genes may be important stem cell factors for embryogenesis in *B. rapa*.

Since auxin plays critical roles in embryo patterning in *Arabidopsis* (Mun et al. 2010; Lau et al. 2012; Rademacher et al. 2012; Sreenivasulu and Wobus 2013), we analyzed the *B. rapa* transcriptome data sets for the expression of auxin-related genes across different stages of embryo development. Several families of auxin-regulated genes, including AUX/IAA family genes (Bra032521, Bra015298, and Bra021184), GH3 family genes (Bra039284, Bra039832, Bra004543, and Bra039186), and SAUR (Bra017676) were up-regulated from the globular to heart stage. Other than three auxin-regulated families, *ARF9-1* (Bra013748) was also up-regulated. ARF9 and redundant ARFs, as well as their inhibitor IAA10, reportedly act in suspensor cells to

mediate hypophysis specification and to prevent its transformation to embryo identity (Rademacher et al. 2012). The ARF gene family is known for their role in auxin-mediated responses (Mun et al. 2012). We detected 29 members of ARF, of which 12 were DEGs. ARF genes were highly expressed from the globular to heart stage (Supplemental Tables 11, 12), which correlates with the function of auxin in embryo patterning (Wendrich and Weijers 2013). Auxin plays essential roles in part through the action of the ARF5/MP TF and its auxin-labile inhibitor IAA12/BDL (Rademacher et al. 2012). Bra012190 is homologous to *ARF5/MP*, and its expression increased from the globular stage and peaked at the heart stage, suggesting that this gene plays a role in early embryo developmental events, consistent with the function of MP in embryonic root initiation and initiation of cotyledon primordia (Lau et al. 2012; Rademacher et al. 2012). BABY BOOM (BBM), a member of the AP2 family, was originally isolated as a marker for embryogenic cells in tissue culture (Boutilier et al. 2002). Bra013978, a homolog of *BBM*, showing high expression during the globular stage among the four stages, may play important roles in embryogenesis, which is consistent with the essential function of *BBM* in the establishment of embryonic morphogenesis (Boutilier et al. 2002).

MYB proteins are key factors in regulatory networks controlling cell fate and identity, development, metabolism, and responses to biotic and abiotic stresses (Dubos et al. 2010). Three DEGs specifically expressed at stages were members of the MYB family. Bra006741 was specifically expressed at the heart stage, while Bra006811 and Bra015029, both coding for hypothetical proteins, were specifically expressed during the early cotyledon stage, suggestive of a divergent role at specific stages. In addition to ARF, AP2, YABBY, MYB, and the HD-ZIPs family, another two families (WOX and WRKY) are well represented from the globular to early cotyledon stages (Supplemental Table 12), which are thought to play key regulatory functions in cell division and differentiation that takes place during early embryo development (Xiang et al. 2011; Basnet et al. 2013; Venglat et al. 2013; Jia et al. 2014). These genes, potentially involved in morphogenesis in *B. rapa* are yet to be functionally defined and fully understood, but will drive some interesting analyses in the future.

Expression changes of metabolism-related genes in *B. rapa* developing embryos

The embryo is the major storage organ in *B. rapa* seeds, and lipid can account for up to 50 % of the seed dry weight in the two cotyledons and the embryo axis (Li et al. 2011; Sreenivasulu and Wobus 2013; Venglat et al. 2014). Metabolic process and gene expression patterns during seed development have been documented for *Arabidopsis* as well as *B. napus* seeds/embryos (Dong et al. 2004; Yu et al. 2010; Troncoso-Ponce et al. 2011; Belmonte et al. 2013). However, little information exists on dynamic genetic regulation of lipid metabolism during embryo development in *B. rapa*. In our study, distinct gene expression patterns related to carbohydrate metabolism, lipid biosynthesis, and storage protein accumulation were documented.

Oilseeds transiently accumulate starch during the early phases of seed development, but have little or no starch at maturity (Yu et al. 2010; Xiang et al. 2011; Basnet et al. 2013). The expression of most genes involved in carbohydrate production, such as starch and sucrose metabolism, glycolysis/gluconeogenesis, citrate cycle, photosynthesis process, carbon fixation in photosynthetic organisms, and photosynthesis-antenna proteins pathways gradually increased after the globular stage, peaked at the early cotyledon stage, and sharply decreased during the mature stage. This was consistent with the peak in starch accumulation and preparing a carbon source for FA and protein synthesis (Santos-Mendoza et al. 2008; Basnet et al. 2013). This suggests that the expression of photosynthetic components is initiated early in embryo development, consistent with the roles in preparing for deposition of lipid and protein storage reserves (Yu et al. 2010; Venglat et al. 2013).

Lipid biosynthesis is a highly coordinated process that involves carbon metabolism, FA synthesis, and triacylglycerol (TAG) synthesis pathways (Yu et al. 2010; Xiang et al. 2011; Basnet et al. 2013). Genes involved in FA biosynthesis, FA elongation, FA metabolism, glycerolipid metabolism, and biosynthesis of unsaturated FAs were first expressed from the globular stage and were up-regulated at the early cotyledon embryo stage. This observation is in agreement with several other studies in which FA synthesis is active at the cotyledon stage because of their roles in preparing for TAG synthesis and accumulation of storage reserves (Santos-Mendoza et al. 2008; Yu et al. 2010). Based on KEGG pathway analysis, the pathways up-regulated to the greatest extent were biosynthesis of unsaturated FAs (4,560) in the embryo from the globular stage to mature stage. FA biosynthesis and starch and sucrose metabolism represented the top up-regulated pathways (Table 3). In addition, the MapMan analysis suggested that storage protein synthesis and lipid metabolism (TAG synthesis) pathways were up-regulated. The expression of several key enzymes in the pathways of FA modification and TAG, including Bra034777 (*FAD2-1*), Bra018348 (*BnaA.FAD3.a*), Bra022767 (*BnaA.FAD3.c*), Bra034635 (*FAE1.1*), and Bra036722 (*DGAT1*) (To et al., 2012), which are critical for determining the composition and/or quantity of seed storage oil in *Arabidopsis* (Sreenivasulu and Wobus 2013), may also have important roles in *B. rapa*. Storage protein, late embryogenesis abundant protein (LEA), and oleosins accumulate in mature stage embryos. Oleosins

are embedded within the phospholipid monolayer of oil bodies to prevent oil body coalescence. They have been reported to have specific functions in the dynamics of lipid accumulation in *Arabidopsis* (Miquel et al. 2014). Oleosin 1, 2, and 4 are three major oil body proteins (Miquel et al. 2014). In our study, oleosin S3-2 (Bra010474), oleosin S1-1 (Bra001002), oleosin S2-1 (Bra033021), oleosin S2-2 (Bra039071), and oleosin S4-4 (Bra025577) were the most highly expressed oleosins, and were expressed from the early cotyledon stage and peaked during the mature stage. Thus, the timing of oleosin gene expression coincides with lipid accumulation, suggestive of a conserved function of oleosins in affecting the lipid content (Joliveta et al. 2011; Liu et al. 2013; Miquel et al. 2014). Our results generated a global expression profile of lipid metabolism in *B. rapa*, providing information for further analyses to uncover the molecular mechanisms underlying FA synthesis and lipid deposition in *Brassica* embryos.

Storage during embryo development is associated with complex pathways that integrate information from genetic programs and from both hormonal and metabolic signals (Santos-Mendoza et al. 2008). The hormones ABA and GA interact with a network of TFs during embryo development, and a high ratio of ABA to GA promotes maturation via ABSCISIC ACID INSENSITIVE3 (ABI3) and ABI5 (Santos-Mendoza et al. 2008). ABI5 is involved in ABA signaling in *Arabidopsis*, which includes the regulation of a subset of LEA genes during both maturation and germination (Cheng et al. 2014). In our study, the increase in *ABI5* (Bra016953) gene expression also coincides with the expression pattern of LEA genes, displaying high expression at the early cotyledon stage. LEAFY COTYLEDON1 (LEC1), LEC2, FUSCA3 (FUS3), and ABI3 are four important embryo-specific TFs that regulate seed maturation (Basnet et al. 2013; Jia et al. 2014). The best-studied member of the NF-Y TFs is LEC1, encoding a homolog of the subunit of CAAT box binding factors, which plays a critical role in embryogenesis and seed maturation (Braybrook and Harada 2008). LEC2, FUS3, and ABI3 encode related TFs of the B3 domain family and are positively regulated by LEC1. LEC2 is thought to regulate WRI1, which in turn mediates glycolysis and FA biosynthesis (To et al. 2012; Sreenivasulu and Wobus 2013; Jia et al. 2014). BnWRI1, BnLEC1, and BnLIL are three key regulators in FA biosynthesis in canola (Yu et al. 2010). Based on sequence similarity, Bra012301 and Bra031356 were identified as homologs of LEC1, which were mapped to "embryo development ending in seed dormancy" (GO:0009793) and "positive regulation of lipid biosynthetic process" (GO:0046889). Bra007066 and Bra003178 were identified as homologs of WRI1, which are involved in the glycolysis and triglyceride metabolic process. Bra029365 was identified as a homolog of Transparent Testa16, which is thought to play multiple roles in plant development and is involved in lipid synthesis and embryo development in canola (Deng et al. 2012). Other positive regulators acting either at the same level or downstream of the LEC genes by binding to the promoters of genes encoding storage proteins and other genes involved in seed maturation have been identified: bZIP TFs that cooperate with the NF-Y complexes (ABI5, bZIP10, bZIP25, bZIP53, and bZIP67; Mendes et al. 2013; Cheng 2014), MYB TFs (AtMYB115; Dubos et al. 2010), and MADS box TFs (AGAMOUS-LIKE15; Zheng et al. 2009). Similarly, these genes are likely to play roles in regulating gene expression during embryo development in *B. rapa*.

Comparison of transcriptomic analysis for oilseed and *Arabidopsis* seeds/embryos

We performed a transcriptome analysis of embryo development in *B. rapa* and provide a genome-wide transcriptional landscape and characterization of the annotated transcripts in its embryos. Our analysis used successive stages of embryo development to determine the temporal expression pattern of each identified gene. In total, 32,941 genes were identified, which is similar to a recent profiling of *B. rapa* and *Brassica oleracea* tissue mRNAs using RNA-Seq analysis (Tong et al. 2013; Liu et al. 2014). In *B. rapa*, 32,335 genes were detected as expressed in at least one tissue (callus, root, leaf, stem, flower, or silique) (Tong et al. 2013), and most mRNAs detected during embryo development in this study are also present throughout the plant life cycle. Thus, a majority of expressed genes in the *B. rapa* genome are active and likely required at all developmental stages.

By analyzing the gene expression profiles in *Arabidopsis* seeds containing the zygote and globular, cotyledon, mature green, and post-mature green stage embryos using DNA microarray analyses, similar numbers of distinct mRNAs accumulated in seeds at the early developing stages, while mRNA levels decreased during the later stages (Le et al., 2010). Similar gene expression trends were observed in *B. rapa* developing embryos. In our study, GO term analysis supported this expression trend, and the expression of genes involved in cellular process, metabolic processes, biological regulation, cellular component organization, pigmentation, and reproduction were significantly down-regulated in mature relative to early cotyledon embryos. A comprehensive transcriptome analysis of gene activity was conducted in *Arabidopsis* embryos using a DNA microarray approach (Xiang et al., 2011). Cluster analysis demonstrated that the expression of genes associated with meristem and morphogenesis accumulated primarily during the globular and heart stages, while carbohydrate, FA, and storage protein synthesis-related genes were highly expressed during the torpedo and bent stages (Xiang et al.

2011). Some similarities and differences exist between our work and previous reports. For example, auxin pathway-related genes and many TF genes were up-regulated during the early stage (globular–heart period); response to stimuli, transporter activity, photosynthesis, and FA biosynthesis-related genes were expressed during the middle stage (heart–early cotyledon period); and lipid and storage protein metabolism-related genes were expressed during the mature stage. These results demonstrated that many pathways were similar in *Arabidopsis* and *B. rapa* during the globular and heart embryo stages, and a divergence in transcriptional programs occurs during the heart stage in embryo development.

Transcriptome profiling in *B. napus* and three other oilseed species revealed conserved expression patterns in FA biosynthesis regulation and distinct species-specific expression patterns for genes involved in synthesis of the glycerolipid pathway (Troncoso-Ponce et al. 2011). A comparative transcriptome analysis of developing seeds from two diverse *B. rapa* morphotypes, namely, a pakchoi (leafy-type) and a yellow sarson (oil-type), revealed that FA synthesis and elongation-related genes showed high expression at 18–25 days after pollination (DAP) and decreased thereafter, while triacylglycerol and FA desaturation biosynthesis processes increased during the late stages or early to middle stages of development, respectively (Basnet et al. 2013). Our analysis showed that genes involved in FA biosynthesis, FA elongation, FA metabolism, glycerolipid metabolism, and biosynthesis of unsaturated FAs were expressed from the globular stage and up-regulated during the early cotyledon embryo stage (20 DAP) and then decreased, while genes involved in storage protein development and triacylglycerol synthesis pathways were up-regulated, consistent with the results of Basnet et al. (2013) and observations on *B. napus* developing embryos (Troncoso-Ponce et al. 2011).

In summary, a comparative transcriptional expression profiling strategy between different developmental stages was used to identify a subset of genes that were differentially expressed. In addition, some potential regulators of embryo development, particularly TF families, which may play important roles during embryo development, were also identified. The identification of genes involved in *B. rapa* embryo development and the temporal expression pattern of each identified gene will increase our understanding of the genetic programs that control embryo development in these development processes and provide a foundation for future studies on the metabolism, growth, and differentiation of embryos in *B. rapa* and other oilseed crops.

Acknowledgments This work was supported by the State Key Basic Research and Development Plan of China (2013CB126900) and the National Natural Science Foundation of China (31370258).

References

Aida M, Beis D, Heidstra R, Willemsen V, Blilou I, Galinha C, Nussaume L, Noh YS, Amasino R, Scheres B (2004) The *PLETHORA* genes mediate patterning of the *Arabidopsis* root stem cell niche. Cell 119:109–120

Basnet RK, Moreno-Pachon N, Lin K, Bucher J, Visser RG, Maliepaard C, Bonnema G (2013) Genome-wide analysis of coordinated transcript abundance during seed development in different *Brassica rapa* morphotypes. BMC Genomics 14:840

Belmonte MF, Kirkbride RC, Stone SL, Pelletiera JM, Bui AQ, Yeung EC, Hashimoto M, Fei J, Haradaa CM, Munoz MD, Le BH, Drews GN, Brady SM, Goldberg RB, Harada JJ (2013) Comprehensive developmental profiles of gene activity in regions and subregions of the *Arabidopsis* seed. Proc Natl Acad Sci USA 110:E435–E444

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Stat 29:1165–1188

Boutilier K, Offringa R, Sharma VK, Kieft H, Ouellet T, Zhang L, Hattori J, Liu CM, van Lammeren AA, Miki BL, Custers JB, Campagne MM (2002) Ectopic expression of *BABY BOOM* triggers a conversion from vegetative to embryonic growth. Plant Cell 14:1737–1749

Bowman JL, Eshed Y (2000) Formation and maintenance of the shoot apical meristem. Trends Plant Sci 5:110–115

Braybrook SA, Harada JJ (2008) LECs go crazy in embryo development. Trends Plant Sci 13:624–630

Carmell MA, XuanZ Zhang MQ, Hannon GJ (2002) The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Gene Dev 16:2733–2742

Cheng ZJ, Zhao XY, Shao XX, Wang F, Zhou C, Liu YG, Zhang Y, Zhang XS (2014) Abscisic acid regulates early seed development in *Arabidopsis* by ABI5-mediated transcription of *SHORT HYPOCOTYL UNDER BLUE1*. Plant Cell 26:1053–1068

Deng W, Chen G, Peng F, Truksa M, Snyder CL, Weselake RJ (2012) *Transparent testa16* plays multiple roles in plant development and is involved in lipid synthesis and embryo development in canola. Plant Physiol 160:978–989

Dong J, Keller WA, Yan W, Georges F (2004) Gene expression at early stages of *Brassica napus* seed development as revealed by transcript profiling of seed-abundant cDNAs. Planta 218:483–491

Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L (2010) MYB transcription factors in *Arabidopsis*. Trends Plant Sci 15:573–581

Fukushima K, Hasebe M (2014) Adaxial-abaxial polarity: the developmental basis of leaf shape diversity. Genesis 52:1–18

Gao Y, Xu H, Shen YY, Wang JB (2013) Transcriptomic analysis of rice (*Oryza sativa*) endosperm using the RNA-Seq technique. Plant Mol Biol 81:363–378

Girin T, Stephenson P, Goldsack CM, Kempin SA, Perez A, Pires N, Sparrow PA, Wood TA, Yanofsky MF, Ostergaard L (2010) Brassicaceae *INDEHISCENT* genes specify valve margin cell fate and repress replum formation. Plant J 63:329–338

Horstman A, Willemsen V, Boutilier K, Heidstra R (2014) AINTEGUMENTA-LIKE proteins: hubs in a plethora of networks. Trends Plant Sci 19:146–157

Izhaki A, Bowman JL (2007) KANADI and class III HD-Zip gene families regulate embryo patterning and modulate auxin flow during embryogenesis in *Arabidopsis*. Plant Cell 19:495–508

Jia H, Suzuki M, McCarty DR (2014) Regulation of the seed to seedling developmental phase transition by the LAFL and VAL transcription factor networks. Dev Biol 3:135–145

Joliveta P, Boularda C, Bellamyb A, Valotc B, d'Andréaa S, Zivyc M, Nesib N, Chardota T (2011) Oil body proteins sequentially

accumulate throughout seed development in *Brassica napus*. J Plant Physiol 168:2015–2020

Kim JH, Lee BH (2006) GROWTH-REGULATING FACTOR4 of *Arabidopsis thaliana* is required for development of leaves, cotyledons, and shoot apical meristem. J Plant Biol 49:463–468

Koyama T, Furutani M, Tasaka M, Ohme-Takagi M (2007) TCP transcription factors control the morphology of shoot lateral organs via negative regulation of the expression of boundary-specific genes in *Arabidopsis*. Plant Cell 19:473–484

Lau S, Slane D, Herud O, Kong J, Jurgens G (2012) Early embryogenesis in flowering plants: setting up the basic body pattern. Annu Rev Plant Biol 63:483–506

Le BH, Cheng C, Bui AQ, Wameister JA, Henry KF, Pelletier J, Kwong L, Belmonte M, Kirkbride R, Horvath S, Drews GN, Fischer RL, Okamuro JK, Harada JJ, Goldberg RB (2010) Global analysis of gene activity during *Arabidopsis* seed development and identification of seed-specific transcription factors. Proc Natl Acad Sci USA 170:8063–8070

Li WL, Zhao Q, Tao P, Wang JB (2011) Identification of genes differentially expressed during embryogenesis in *Brassica campestris* L. Acta Physiol Plant 34:669–681

Liu J, Hua W, Yang HL, Zhan GM, Li RJ, Deng LB, Wang XF, Liu GH, Wang HZ (2012) The *BnGRF2* gene (*GRF2-like* gene from *Brassica napus*) enhances seed oil production through regulating cell number and plant photosynthesis. J Exp Bot 63:3727–3740

Liu WX, Liu HL, Qu LQ (2013) Embryo-specific expression of soybean oleosin altered oil body morphogenesis and increased lipid content in transgenic rice seeds. Theor Appl Genet 126:2289–2297

Liu SY, Liu YM, Yang XH, Tong CB, Edwards D et al (2014) The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nat Genet 5:3930

Lu X, Chen D, Shu D, Zhang Z, Wang W, Klukas C, Chen LL, Fan Y, Chen M, Zhang C (2013) The differential transcription network between embryo and endosperm in the early developing maize seed. Plant Physiol 162:440–455

Machida Y, Fukaki H, Araki T (2013) Plant meristems and organogenesis: the new era of plant developmental research. Plant Cell Physiol 54:295–301

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-Seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res 18:1509–1517

Mendes A, Kelly AA, Erp VH, Shaw E, Powers SJ, Kurup S, Eastmond PJ (2013) bZIP67 regulates the omega-3 fatty acid content of *Arabidopsis* seed oil by activating *FATTY ACID DESATURASE3*. Plant Cell 25:3104–3116

Miquel M, Trigui G, d'Andréa S, Kelemen Z, Baud S, Berger A, Deruyffelaere C, Trubuil A, Lepiniec L, Dubreucq B (2014) Specialization of oleosins in oil body dynamics during seed development in *Arabidopsis* seeds. Plant Physiol 164:1866–1878

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5:621–628

Mouchel CF, Osmont KS, Hardtke CS (2006) *BRX* mediates feedback between brassinosteroid levels and auxin signalling in root growth. Nature 443:458–461

Mun JH, Yu HJ, Shin JY, Oh M, Hwang HJ, Chung H (2012) Auxin response factor gene family in *Brassica rapa*: genomic organization, divergence, expression, and evolution. Mol Genet Genomics 287:765–784

Ouakfaoui SE, Schnell J, Abdeen A, Colville A, Labbé H, Han S, Baum B, Laberge S, Miki B (2010) Control of somatic embryogenesis and embryo development by AP2 transcription factors. Plant Mol Biol 74:313–326

Park S, Harada JJ (2008) *Arabidopsis* embryogenesis. Methods Mol Biol 427:3–16

Perilli S, Di Mambro R, Sabatini S (2012) Growth and development of the root apical meristem. Curr Opin Plant Biol 15:17–23

Rademacher EH, Lokerse AS, Schlereth A, Llavata-Peris CI, Bayer M, Kientz M, Rios AF, Borst JW, Lukowitz W, Jürgens G, Weijers D (2012) Different auxin response machineries control distinct cell fates in the early plant embryo. Dev Cell 22:211–222

Santos-Mendoza M, Dubreucq B, Baud S, Parcy F, Caboche M, Lepiniec L (2008) Deciphering gene regulatory networks that control seed development and maturation in *Arabidopsis*. Plant J 54:608–620

Spencer MWB, Grene R, Lindsey K (2007) Transcriptional profiling of the *Arabidopsis* embryo. Plant Physiol 143:924–940

Sreenivasulu N, Wobus U (2013) Seed-development programs: a systems biology-based comparison between dicots and monocots. Annu Rev Plant Biol 64:189–217

To A, Joubes J, Barthole G, Lecureuil A, Scagnelli A, Jasinski S, Lepiniec L, Baud S (2012) WRINKLED transcription factors orchestrate tissue-specific regulation of fatty acid biosynthesis in *Arabidopsis*. Plant Cell 24:5007–5023

Tong CB, Wang XW, Yu JY, Wu J, Li WS, Huang JY, Dong CH, Hua W, Liu SY (2013) Comprehensive analysis of RNA-seq data reveals the complexity of the transcriptome in *Brassica rapa*. BMC Genomics 14:689

Troncoso-Ponce MA, Kilaru A, Cao X, Durrett TP, Fan J, Jensen JK, Thrower NA, Pauly M, Wilkerson C, Ohlrogge JB (2011) Comparative deep transcriptional profiling of four developing oilseeds. Plant J 68:1014–1027

Vanstraelen M, Benkova E (2012) Hormonal interactions in the regulation of plant development. Annu Rev Cell Dev Biol 28:463–487

Venglat P, Xiang D, Yang H, Wan L, Tibiche C, Ross A, Wang E, Selvaraj G, Datla R (2013) Gene expression profiles during embryo development in *Brassica napus*. Plant Breed 132:514–522

Venglat P, Xiang D, Wang E, Datla R (2014) Genomics of seed development: challenges and opportunities for genetic improvement of seed traits in crop plants. Biocatal Agric Biotechnol 3:24–30

Wang LK, Feng ZX, Wang X, Wang XW, Zhang XG (2010) DEGseq: an R package for identifying differentially expressed genes from RNA-Seq data. Bioinformatics 26:136–138

Wang XW, Wang HZ, Wang J, Sun RF, Wu J, Liu SY et al (2011) The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet 43:1035–1039

Wendrich JR, Weijers D (2013) The *Arabidopsis* embryo as a miniature morphogenesis model. New Phytol 199:14–25

Wildwater M, Campilho A, Perez-Perez JM, Heidstra R, Blilou I, Korthout H, Chatterjee J, Mariconti L, Gruissem W, Scheres B (2005) The *RETINOBLASTOMA-RELATED* gene regulates stem cell maintenance in *Arabidopsis* roots. Cell 123:1337–1349

Wu G, Lin W, Huang T, Poethig RS, Springer PS, Kerstetter RA (2008) KANADI1 regulates adaxial-abaxial polarity in *Arabidopsis* by directly repressing the transcription of *ASYMMETRIC LEAVES2*. Proc Natl Acad Sci USA 105:16392–16397

Xiang DQ, Venglat P, Tibiche C, Yang H, Risseeuw E, Cao YG, Basic V, Cloutier M, Keller W, Wang E, Selvaraj G, Datla R (2011) Genome-wide analysis reveals gene expression and metabolic network dynamics during embryo development in *Arabidopsis*. Plant Physiol 156:346–356

Xu H, Gao Y, Wang JB (2012) Transcriptomic analysis of rice (*Oryza sativa*) developing embryos using the RNA-Seq technique. PLoS ONE 7:e30646

Yang XY, Zhang XL, Yuan DJ, Jin FY, Zhang YC, Xu J (2012) Transcript profiling reveals complex auxin signalling pathway and transcription regulation involved in dedifferentiation and redifferentiation during somatic embryogenesis in cotton. BMC Plant Biol 12:110

Yu B, Gruber M, Khachatourians GG, Hegedus DD, Hannoufa A (2010) Gene expression profiling of developing *Brassica napus*

seed in relation to changes in major storage compounds. Plant Sci 178:381–389

Zhao Q, Zou J, Meng JL, Mei SY, Wang JB (2013) Tracing the transcriptomic changes in synthetic trigenomic allohexaploids of *Brassica* using an RNA-Seq approach. PLoS ONE 8:e68883

Zheng Y, Ren N, Wang H, Stromberg AJ, Perry SE (2009) Global identification of targets of the *Arabidopsis* MADS domain protein AGAMOUS-Like 15. Plant Cell 21:2563–2577