Gene expression atlas of embryo development in Arabidopsis

Gao, Peng; Xiang, Daoquan; Quilichini, Teagen D.; Venglat, Prakash; Pandey, Prashant K.; Wang, Edwin; Gillmor, C. Stewart; Datla, Raju

Publisher's version / Version de l'éditeur:
https://doi.org/10.1007/s00497-019-00364-x
Plant Reproduction, 2019-02-14

Plant Reproduction
https://doi.org/10.1007/s00497-019-00364-x

METHODS PAPER

Gene expression atlas of embryo development in *Arabidopsis*

Peng Gao[1,5] · Daoquan Xiang[1] · Teagen D. Quilichini[1] · Prakash Venglat[1,2] · Prashant K. Pandey[1] · Edwin Wang[3] · C. Stewart Gillmor[4] · Raju Datla[1,5]

Received: 16 November 2018 / Accepted: 1 February 2019

Abstract

Embryogenesis represents a critical phase in the life cycle of flowering plants. Here, we characterize transcriptome landscapes associated with key stages of embryogenesis by combining an optimized method for the isolation of developing *Arabidopsis* embryos with high-throughput RNA-seq. The resulting RNA-seq datasets identify distinct overlapping patterns of gene expression, as well as temporal shifts in gene activity across embryogenesis. Network analysis revealed stage-specific and multi-stage gene expression clusters and biological functions associated with key stages of embryo development. Methylation-related gene expression was associated with early- and middle-stage embryos, initiation of photosynthesis components with the late embryogenesis stage, and storage/energy-related protein activation with late and mature embryos. These results provide a comprehensive understanding of transcriptome programming in *Arabidopsis* embryogenesis and identify modules of gene expression corresponding to key stages of embryo development. This dataset and analysis are a unique resource to advance functional genetic analysis of embryo development in plants. AQ1

Keywords Embryogenesis · Embryo isolation · RNA-seq · Transcriptome · Bioinformatics · *Arabidopsis*

Introduction

Embryogenesis begins with fusion of male and female gametes to form the single-cell zygote. In *Arabidopsis*, the first division of the zygote is asymmetric and is followed by precisely oriented cell divisions which generate the major tissue lineages of the adult plant within the first five days of embryogenesis (Goldberg et al. 1994; Xiang et al. 2011a). Landmark events of embryogenesis include establishment of shoot and root meristems and provascular and ground tissue in early stages; growth of the embryo in middle stages; and activation of seed storage protein and seed dormancy programs in late embryogenesis (Braybrook and Harada 2008; Huh et al. 2007; Jenik et al. 2007; Moller et al. 2017; Yang et al. 2009).

Discovering the genetic mechanisms employed during plant embryogenesis requires accessing embryos at progressive stages of development. This is challenging due to their

Communicated by Dolf Weijers.

RNA sequencing raw data can be found in the Gene Expression Omnibus (GEO) under the accession number "GSE123010."

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00497-019-00364-x) contains supplementary material, which is available to authorized users.

✉ Raju Datla
raju.datla@nrc-cnrc.gc.ca

[1] Aquatic and Crop Resource Development, National Research Council Canada, 110 Gymnasium Place, Saskatoon, SK S7N 0W9, Canada

[2] Department of Plant Sciences and Crop Development Centre, University of Saskatchewan, 51 Campus Drive, Saskatoon, SK S7N 5A8, Canada

[3] Center for Health Genomics and Informatics, University of Calgary Cumming School of Medicine, Calgary, AB T2N 4N1, Canada

[4] Laboratorio Nacional de Genómica para la Biodiversidad (Langebio), Unidad de Genómica Avanzada, Centro de Investigación y de Estudios Avanzados del IPN (CINVESTAV-IPN), Irapuato, Guanajuato, Mexico

[5] Global Institute for Food Security, University of Saskatchewan, Saskatoon, SK S7N 4J8, Canada

Journal : Large 497 | Article No : 364 | Pages : 12 | MS Code : SPRE-D-18-00106 | Dispatch : 5-2-2019

small size, their enclosure within maternal tissue, and their close association with the endosperm and surrounding seed coat tissues (Xiang et al. 2011a). Laser capture microdissection (LCM), a technology to facilitate the precise excision of select tissues or cells, has been used to isolate embryos and larger embryonic tissue domains from *Arabidopsis* for transcriptional analysis (Belmonte et al. 2013; Casson et al. 2005; Kerk et al. 2003). To avoid contamination from adjoining cells, LCM requires high precision during tissue excision. In addition, transcript alterations or reductions due to damage caused by material fixation remain a concern for obtaining accurate transcriptome data from LCM isolates. Tools employing fluorescence or affinity tagging with fluorescence-activated cell/nuclei sorting (FACS/FANS), translating ribosome affinity purification (TRAP), and the isolation of nuclei tagged in specific cell types (INTACT), have the potential to aid embryo identification and improve isolation accuracy (Bonner et al. 1972; Deal and Henikoff 2010; Heiman et al. 2008; Zhang et al. 2008). These tag-labeling methods have been successfully applied in the sorting of protoplasts and select tissues in plants. For studies of embryogenesis, one disadvantage of FACS and INTACT analyses is that the ratio of maternal tissues to embryo target cells is enormous, making avoidance of maternal contamination difficult. Another drawback of tag-labeling methods is the requirement of genetic manipulation and transformation protocols for the species of interest and the time to generate these lines.

While the techniques mentioned above offer unique advantages and disadvantages for the isolation of embryos for transcriptome studies, isolation of embryos by precise hand dissection in conditions that maintain RNA integrity is a valuable approach with significantly lower cost (Xiang et al. 2011a, b). The use of living materials in manual dissections is minimally invasive and does not create fixation artifacts, making this method conducive to capturing the native transcriptome landscape. Furthermore, manual dissection does not require genetic manipulation of the species under study, making it an efficient and affordable approach to accessing embryo transcriptome data in a broad range of species.

Transcriptome analyses of embryo development in *Arabidopsis* have been conducted using Affymetrix Gene Chip and *Arabidopsis* oligonucleotide microarrays (Belmonte et al. 2013; Palovaara et al. 2017; Slane et al. 2014; Xiang et al. 2011a) and RNA-seq (Autran et al. 2011; Hsieh et al. 2011; Nodine and Bartel 2012; Pignatta et al. 2014). In comparison with microarray assays, RNA-seq offers several advantages, including improved sensitivity, access to alternative splicing information, and accurate detection of a broad range of expression data. In the present study, high-throughput RNA-seq analysis was performed using *Arabidopsis* embryos, from zygote to maturity, that were isolated by an optimized hand dissection method. Comprehensive transcriptome analysis across all key stages of embryogenesis revealed gene expression clusters associated with different stages of embryo development. Functional interrogation of these gene clusters identified characteristic biological pathways involved in the early, middle, late, and mature stages of embryo development. These results define gene expression patterns and coexpression networks associated with embryogenesis in *Arabidopsis* and provide a framework for hypothesis generation and functional validation of the molecular mechanisms governing embryogenesis.

Materials and methods

Plant growth and embryo isolation

Arabidopsis (Arabidopsis thaliana, ecotype Col-0) seeds were placed on ½ Murashige and Skoog (MS) plates for 4 days at 4 °C in the dark and then transferred to 24-hour light for 10 days. Seedlings were transferred to soil and grown under long-day conditions (16-h light and 8-h dark) with a constant temperature of 22 °C and 120–150 µmol m^{-2} s^{-1} light intensity. Timed and controlled self-pollinations were performed to ensure synchronized development of early-stage embryos in the ovules. For each pollination, the main inflorescence was pruned to remove all flowers and buds, except for the two oldest flower buds. The remaining flower buds were pollinated 4 h after emasculation using pollen from an open flower of the same plant.

Embryo isolation by hand dissection

Embryo isolation was performed as described in Xiang et al. (2011a, b), with some modifications and additional details outlined below. The approximate number of embryos harvested for each of the seven stages of embryo development defined by this study, was as follows: 100 zygote and 2 cell embryos (Z), 200 octant (O), 100 globular (G), 100 heart (H), 60 torpedo (T), 30 bent (B), and 30 mature (M) embryos. Z-, O-, and G-stage embryos were isolated at approximately 24, 60, and 72 h after pollination, respectively. The H- to M-stage embryos were isolated based on their morphology.

Contamination of the embryo mRNA with seed coat- and endosperm-derived mRNA in the early-stage embryos (Z and O stages) is a major concern. To ensure that clean embryos were obtained for mRNA extraction, we performed the isolation of embryos from ovules in mini-petri dishes containing isolation buffer (4.8% sucrose solution + 0.1% RNAlater, Ambion Cat# AM7020). Two precise incisions were made at the micropylar end with needles (Fine Science Tools Cat. 10130-05) (Xiang et al. 2011a, b). This

Journal : Large 497 | Article No : 364 | Pages : 12 | MS Code : SPRE-D-18-00106 | Dispatch : 5-2-2019

approach enabled the separation of the micropylar region that houses the early-stage embryo from the seed coat and the subsequent isolation of the embryo within the micropyle from the maternal ovule tissue and endosperm cells. Isolated embryos from early stages were inspected under a microscope (Leica DMR), and only samples with no visible contamination from the ovule tissue or early endosperm nuclei/cells were collected for RNA isolation. Since the ovule soon after fertilization contains very few endosperm cells/nuclei, the risk of endosperm contamination is greater at later stages (after globular) which usually has dense endosperm cells associated with the embryos. Therefore, embryos at early and later stages of development (after globular) were carefully washed several times in isolation buffer to avoid contamination from surrounding tissues such as the endosperm. The washing steps were performed using mini-petri dishes filled with freshly prepared isolation buffer. After one single embryo was isolated from the ovule, it was transferred using pipette to a mini-petri dish, followed by gentle agitation to separate any attached debris. To remove remaining debris, each embryo was transferred to fresh isolation buffer in a mini-petri dish for at least three rounds of sequential washing steps. All washing steps were performed on ice and inspected with a dissecting Leica microscope. After all washing steps, embryos were transferred to Eppendorf tubes on dry ice using fine glass pipettes. Due to their diminutive size and slight variation in the timing of first division, precisely differentiating and separating was difficult for single-celled zygotes from zygotes which had divided to produce 2-cell embryos (i.e., with an apical embryo and basal suspensor cell), and the isolated zygote stage samples may also contain few 2-cell embryos.

To test the efficacy of our isolation procedure for obtaining clean embryos (without a significant contaminating endosperm or seed coat tissues), embryos were placed on a glass slide in large droplet of water, within the confines of a well (created by a Mini PAP pen, Invitrogen Cat# 008877) and coverslip and imaged using a Leica DMR equipped with a Microfire camera (Optronics, California).

RNA extraction, antisense RNA amplification, and high-throughput sequencing

Total RNA was extracted from embryos corresponding to the seven developmental stages described above, for two biological replicates at each stage, following the RNAqueous-Micro kit protocol (Ambion, Catalog# 1927). The isolated embryos' homogenization was performed using polypropylene pestles in 1.5-ml Eppendorf tubes after adding lysis buffer. The whole homogenization process was operated on ice. The quantity of RNA isolated from early-stage embryos was insufficient for direct library preparation and RNA-seq experiments. Therefore, the mRNA of all stages was similarly subjected to amplification and the resulting antisense RNA (aRNA) was used for RNA-seq experiments. The mRNA amplification was conducted according to the protocol provided in the MessageAmp aRNA kit (Ambion, Catalog# 1750). For RNA-seq profile analysis, Illumina mRNA-seq libraries were prepared using the TruSeq RNA kit (ver. 1, rev A) according to the manufacturer's instructions. For Illumina HiSeq 2000 sequencing (with an Illumina GA II instrument), four indexed libraries were pooled per sequencing lane, paired-end reads were aligned to the reference genome (TAIR10), and the reads for each gene were summarized for further analysis.

Differential gene expression (DEG) analysis

High-quality reads were filtered using the Trimmomatic tool (Bolger et al. 2014) with default parameters to remove sequencing adapters and low-quality reads. A quality check was performed with FastQC (Andrews 2010), and the filtered reads were mapped to the *Arabidopsis* genome TAIR10 using STAR (Dobin et al. 2013). Read counting was carried out with HTSeq (Anders et al. 2015). Analysis of differential expression genes (DEGs) was performed with DESeq 2 to normalize the raw reads and identify DEGs in pairwise comparisons between each stage of embryo development (Love et al. 2014). An experimental design based on the DESeq 2 R package was applied to raw read counts for all libraries with p value (adjusted) < 0.05, log2FoldChange > 1 and < -1 to reflect transcriptional regulation across all seven developmental stages, as shown in Supplementary Data S1. Raw data counts were normalized by library size and fit to a negative binomial model. Principal component analysis (PCA) was calculated with the built-in plotPCA function provided by the DESeq 2 tool. All DEGs were annotated by querying the open reading frame (ORF) sequences against the non-redundant protein database (with BLAST) using an e-value cutoff of 10^{-5} and reporting the maximum "hit" sequence per query. Differential expression of genes was determined to be significant if a log2 (fold change) of > 1 or < -1 and p value (adjusted) < 0.05 was obtained.

Gene coexpression analysis and heatmap generation

To identify genes that exhibit similar expression across all seven stages of embryo development, raw counts were normalized to the $\log_2$ scale, using the rlog function in DESeq 2. Sample-to-sample distance was calculated based on rlog matrix using the R package dist function. Unrooted hierarchical trees were generated using SplitsTree4 software with the R package hclust function "complete" method. Counts with rlog < 3 across all samples were removed. Groups of closely connected genes were identified by weighted

Journal : Large 497 | Article No : 364 | Pages : 12 | MS Code : SPRE-D-18-00106 | Dispatch : 5-2-2019

gene coexpression network analysis (WGCNA) with a soft thresholding power of 9 (Langfelder and Horvath 2008) and "unsigned" network model. Hierarchical clustering was performed based on the topological overlap matrix and cutting the resulting dendrogram with the dynamicTreeCut program with deepslipt = 4, minModuleSize = 30, MaxModuleSize = 5000 to obtain maximum dynamic gene clusters. Initial clusters with similar expression profiles were merged at cutheight = 0.25. Gene modules with stage-specific patterns of gene expression (specifically, patterns where gene expression was relatively high in one stage and low in the other six) were correlated with the stage-specific expression patterns in different embryo samples, based on significantly high Pearson's correlation coefficient values ($r > 0.8$, p value < 0.001; Fig. 2b). Based on the resulting adjacency matrix, we calculated the topological overlap, a concept defined for weighted networks, which is a robust and biologically meaningful measure of network interconnectedness (Langfelder and Horvath 2008). Heatmaps were generated by the R pheatmap package using rlog values with Z-score transformation.

Functional classification of transcripts based on gene ontology, KEGG, and MapMan pathway enrichment

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichments were performed using the topGO and KEGGprofile R packages, respectively. REVIGO analysis (http://revigo.irb.hr) was used to slim the enriched GO terms based on the "medium similarity" parameter. The MapMan tool (Thimm et al. 2004) was used to facilitate the assignment of identified DEGs into functional categories (bins). A MapMan mapping file that mapped the genes into bins via hierarchical ontologies through the searching of a variety of reference databases was generated using the Mercator tool (http://mapman.gabipd.org/web/guest/app/mercator). DEGs associated with seed development pathways from the seven stages of embryo development were categorized and visualized in ImageAnnotator module. GO term interact graphs were generated by Cytoscape software.

Results

Transcriptome profiling of *Arabidopsis* embryogenesis

To obtain a comprehensive view of transcriptome profiles during embryogenesis in *Arabidopsis*, RNA-seq analysis was performed using isolated embryos at seven key stages of embryogenesis, referred to herein as the zygote (Z), octant (O), globular (G), heart (H), torpedo (T), bent (B), and mature (M) embryo. Embryo poly(A) + transcripts were used to generate libraries for Illumina high-throughput sequencing. In total, 257 million sequenced reads were obtained and aligned to the *Arabidopsis* TAIR 10 Genome. The corresponding numbers of mapped, unmapped reads, unique mapped reads, and multiple mapped reads from each sample are presented in Figure S1. Given the large number of samples in this study, relative relatedness and reproducibility among the two biological replicates were examined by principal component analysis (PCA) (Fig. 1a), which displayed a clear clustering of replicate samples isolated from the same stage of embryogenesis. Sample-to-sample distances were calculated based on a normalized read count transformed rlog matrix from all samples as shown in Fig. 1b. All replicate samples had highly correlated Pearson's coefficients demonstrating accurate isolation of embryos representing the specific stage of their development as well as the reproducibility of the corresponding RNA isolation and library preparation procedures. Moreover, the analysis showed higher degrees of correlation in neighboring stages of embryo development (i.e., Z to O, G to H, and T to B), except for the M stage, which was more distant (Fig. 1b). With replicates combined, the unrooted hierarchical tree showed the same trend as the sample-to-sample distance analysis, with four discrete groupings of related expression patterns across the seven stages of embryo development (Fig. 1c). Thus, four groupings for embryo development emerged, termed them as the early (Z and O), middle (G and H), late (T and B), and mature (M) embryo groups (Fig. 1c). These groupings are consistent with findings described in a previous microarray study by Xiang et al. (2011a, b) using comparable samples from similar embryo stages. Both studies revealed a clear separation of four distinct expression patterns from early to late embryogenesis.

DEG identification across whole embryo developmental stages

The study of differential gene expression provides insight into the putative activation and repression of pathways and processes associated with the complex process of embryogenesis. In total, 15,912 differentially expressed genes (DEGs) were identified through pairwise comparisons between consecutive stages of development with p value (adjusted) < 0.05, log2FoldChange > 1 and < -1. Each pairwise comparison identified a large proportion of DEGs, including 8436 in Z versus O; 5568 in O versus G; 4836 in G versus H; 4638 in H versus T; 3550 in T versus B; and 10,339 in B versus M. Among the DEGs identified, 101 genes appeared in six comparisons across all seven stages of embryo development (Data S2). Gene ontology (GO) analysis of these 101 genes revealed enrichment of the GO

Journal : Large 497 | Article No : 364 | Pages : 12 | MS Code : SPRE-D-18-00106 | Dispatch : 5-2-2019

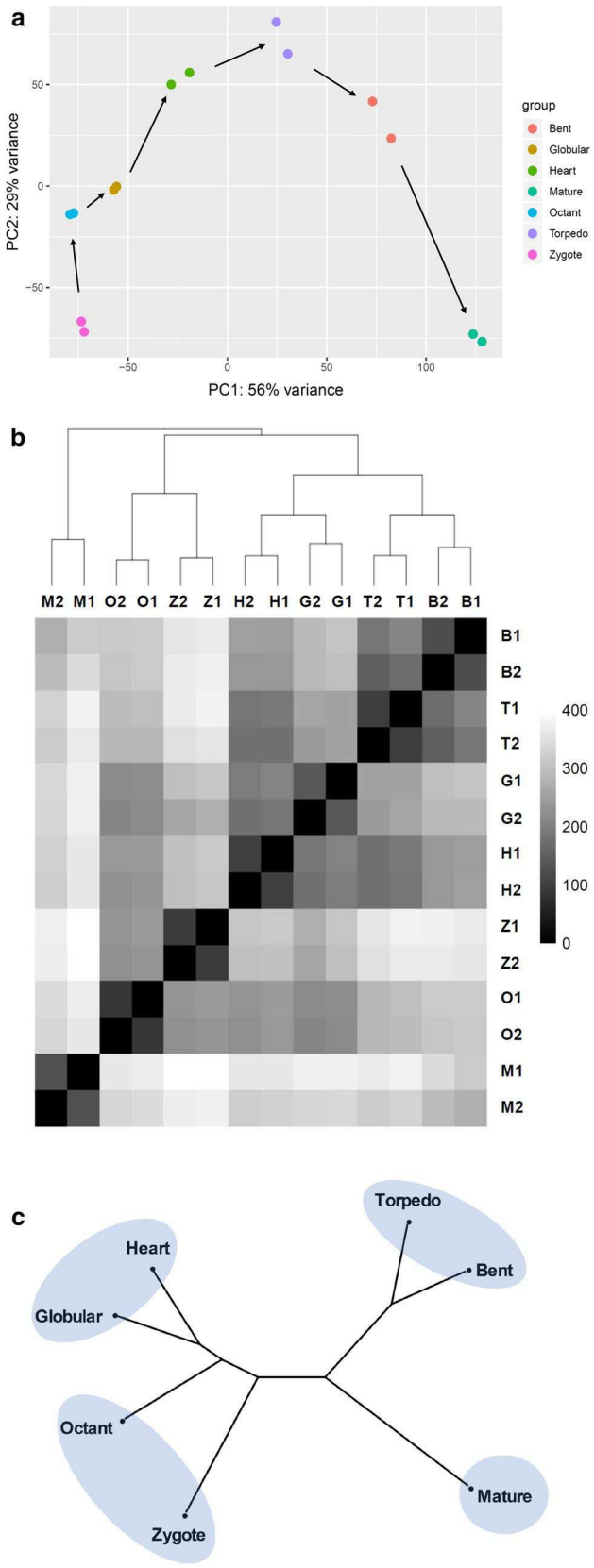

Fig. 1 Hierarchical clusters of embryo transcriptomes in different developmental stages. **a** Principal component analysis (PCA) of the transcriptomes for seven stages of embryo development. Biological replicates from the same embryo stage are represented by the same color. Arrows indicate the direction of embryo development. **b** Heatmap of sample distance from seven different embryo stages, with two replicates (numbered 1 and 2) per stage. The grayscale spectrum represents sample distance calculated by R dist function ranging from 0 (black) to 400 (white), indicating high to low correlations, respectively. Embryo stages include Z, zygote; O, octant; G, globular; H, heart; T, torpedo; B, bent; and M, mature. **c** Unrooted hierarchical tree generated with combined sample replicates using SplitsTree4 software. The node distance was calculated with the R hclust package "complete" method. The length of lines connecting stages of embryo development are proportional to relatedness

terms "post-embryonic development," "seed development," and "fruit development," suggesting fundamental roles for these gene products in different developmental phases of embryogenesis (Table 1).

Stage-specific gene coexpression networks

Since the 15,912 DEGs represent diverse functionalities in an array of biological processes, weighted gene coexpression network analysis (WGCNA) and dynamic hierarchical clustering approaches were used to define clusters, called modules, of coexpressed genes that follow specific patterns of expression (Fig. 2). By including genes with expression levels greater than or equal to ten counts in at least one sample from all seven developmental stages, a total of 26 modules were identified (Fig. 2a). We first defined the module eigengene (ME), a single value that represents the highest percent of variance in expression values for all module genes in a stage. Thus, the expression profiles of module genes among different stages can be summarized as the expression profile of MEs. Pearson's correlation coefficient values (r) between MEs and each stage were then used to determine the relationship between a module and a developmental stage (Fig. 2b). Collectively, we identified six "stage-specific" modules (see asterisks, Fig. 2b), defined as modules that significantly correlated with one stage of embryo development ($r > 0.8$ and p value < 0.001), correlating "MEblack" to zygote, "MEtan" to octant, "MElightgreen" to heart, "MEdarkgray" to torpedo, "MEdarkred" to bent, and "MEmagenta" to mature. No stage-specific module was identified for the globular stage, suggesting this important transition stage connecting early and late embryo development may be distinguished by an overlap of biological processes. Besides the stage-specific modules, we also identified seven "multi-stage" modules (see dashed outlines, Fig. 2b), defined by a constant dominant expression pattern in two or more consecutive stages of embryogenesis (r of each stage > 0.2, sum(r) > 0.8), including "MElightcyan," "MEgreenyellow," "MEsalmon," "MEyellow," "MEblue,"

Journal : Large 497 | Article No : 364 | Pages : 12 | MS Code : SPRE-D-18-00106 | Dispatch : 5-2-2019

Table 1 Gene ontology (GO) terms associated with selected 101 DEGs

GO_acc	GO_term	Gene_ratio	Bg_ratio	FDR
GO:0009791	Post-embryonic development	11/101	705/37767	0.00047
GO:0048316	Seed development	9/101	530/37767	0.001
GO:0010154	Fruit development	9/101	557/37767	0.001
GO:0009793	Embryonic development ending in seed dormancy	7/101	465/37767	0.0062
GO:0048608	Reproductive structure development	10/101	978/37767	0.0062

For GO enrichment analysis, 101 DEGs identified from comparisons between every two consecutive stages of embryo development (for a total of six comparisons) were selected. The top five enriched GO terms and the corresponding gene ratio are indicated. Gene ratio is defined by the number of genes related to the GO term/the number of genes in the gene set. Bg_ratio (Background_ratio) is defined as the number of genes in *Arabidopsis* related to this GO term/total number of genes in *Arabidopsis* genome. *GO_acc* GO accession number. *FDR* false discovery rate, which is the rate of type I errors in null hypothesis testing when conducting multiple comparisons

"MEred," and "MEgreen." We combined these thirteen modules and named them M1–M13, based on their respective dominant expression stages during the sequential developmental progression from zygote to mature embryo. The color modules and corresponding module numbers are listed in Supplemental Data S3. An updated eigengene adjacency heatmap generated using the thirteen MEs and seven stages of embryo development highlights the association between development stages and the stage-specific (M1, 2, 5, 6, 9, 11) or multi-stage modules (M3, 4, 7, 8, 10, 12, 13) (Fig. 3a).

Cluster analysis of coexpression gene modules

To investigate transcriptional regulation across embryo development, all the genes represented by the thirteen modules were analyzed. First, a heatmap of stage-specific modules associated with the developmental stages M1, M2, M5, M6, M9, and M11 was generated. As shown in Fig. 2c, genes enriched or absent in specific stages of development were clustered with each corresponding module.

To investigate the putative functions associated with the stage-specific and multi-stage modules, we performed gene ontology (GO) enrichment analysis (Ashburner et al. 2000). Top GO terms in the biological process (BP, Fig. 3b) and cellular component (CC, Figure S2) categories were generated. No enrichment of BP GO terms was detected in M2, 3, 6, 9, 12 with a threshold setting of p value (adjusted) < 0.01, indicating no dominant biological process could be linked to these module-associated embryo stages. Gene lists and gene functional descriptions of all 13 modules are presented in Data S3. The majority of genes enriched in M1 are involved in developmental regulation and protein-targeting processes, supporting the importance of these genes and the associated processes in the zygote and early embryo establishment phase. Genes in M4 are highly expressed in the zygote and octant stages (Fig. 3a) and show GO term enrichment in cell wall modification and organism developmental processes. Gene activities involved in developmental processes in M1 and M4 are closely related to the embryonic patterning events that occur in early stages of embryonic development, as highlighted in a previous study (Jenik et al. 2007). M7 and M8 are multi-stage modules where high gene expression was associated specifically with early and middle stages of embryo development, from O to T and from Z to T, respectively. GO terms associated with these two modules were predominantly related to cell development. Genes with GO term enrichment in methylation were also observed, particularly in M8, supporting a significant epigenetic programming and its influence on early and middle stages of embryo development. M10 is a multi-stage module associated with late embryo stages of development, corresponding to T and B stages, in which GO enrichment analysis revealed the onset of the components associated with photosynthetic processes. M11 and M13, modules associated with late and mature embryo development, respectively, showed GO term enrichment in metabolic processes and storage reserve synthesis (storage proteins, lipids, fatty acid oxidation), suggesting that lipid metabolism is reprogrammed for seed dormancy before the formation of mature seeds. The interact graphs of GO terms in these modules were simulated (Fig. 3c–e, S4). Our results suggest that the top enriched GO terms from most gene modules are grouped to one cluster, indicating that stage-specific genes in the same module are primarily involved in similar biological processes. One exception to this is gene module M1. The top enriched GO terms from M1 are clustered in three different groups (Fig. S4a), revealing complex reprogramming after fertilization at the zygote stage of embryo development.

The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database resource for understanding the high-level functions of a biological system (Kanehisa and Goto 2000). The significantly enriched pathway ontology (PO) terms were phenylpropanoid biosynthesis in M1, ribosome in M8, photosynthesis in M10, and lipid metabolism in M13 (Fig. S3).

Journal : Large 497 | Article No : 364 | Pages : 12 | MS Code : SPRE-D-18-00106 | Dispatch : 5-2-2019

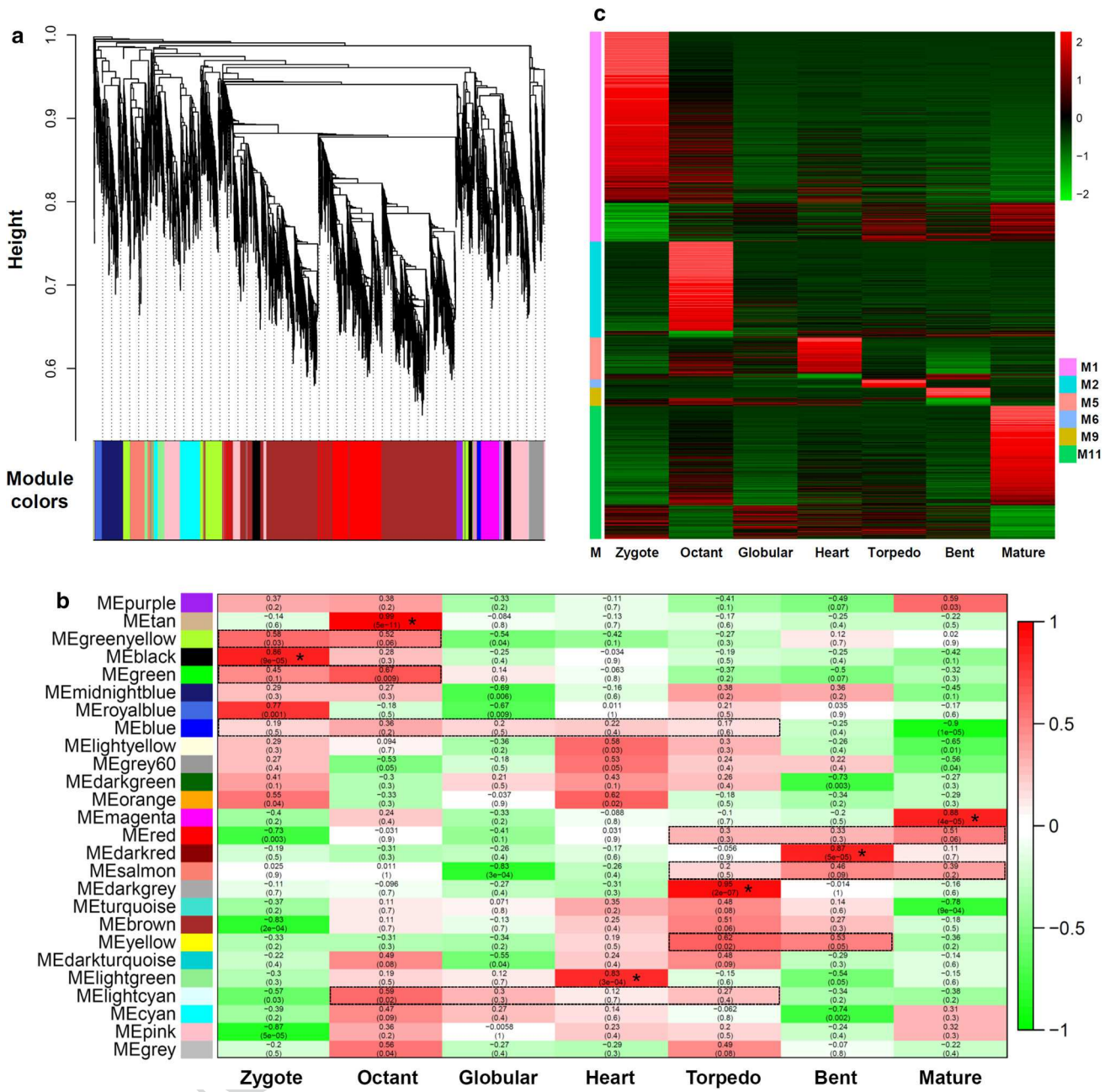

Fig. 2 Stage-specific gene modules identified by weighted gene coexpression network analysis (WGCNA). **a** Hierarchical cluster tree showing modules of coexpressed genes, with a total of 26 coexpressed gene modules identified by WGCNA dendrograms. The height (*y*-axis) indicates level of correlation. Colors represent the 26 different modules, with gray color indicating genes that could not be assigned to any module. **b** Heatmap of correlations between coexpressed gene modules and stages of embryo development, with numerical Pearson's correlation coefficient (top) and corresponding *p* values (bottom). The color scheme, from red through white to green, indicates the level of correlation, from high to low. Stage-specific modules (*) are identified by $r > 0.8$, $p < 0.001$. Associated consecutive stages in multi-stage modules are identified by *r* of each stage > 0.2, sum (*r*) > 0.8, and grouped in dashed boxes. **c** Heatmap of expression dynamics of gene members in six stage-specific modules, M1, 2, 5, 6, 9, 11. Genes in each module are associated with a color on the left of the heatmap. *Z*-score was applied for each row. Each module contains high expression (red) and low expression (green) genes in corresponding stages

These PO terms are similar to the results of GO term enrichment analysis, providing further support for the functional annotation of gene modules.

DEGs associated with seed developmental networks

In this study, an ImageAnnotator module from the MapMan visualization tool was used to define hierarchical functional

Journal : Large 497 | Article No : 364 | Pages : 12 | MS Code : SPRE-D-18-00106 | Dispatch : 5-2-2019

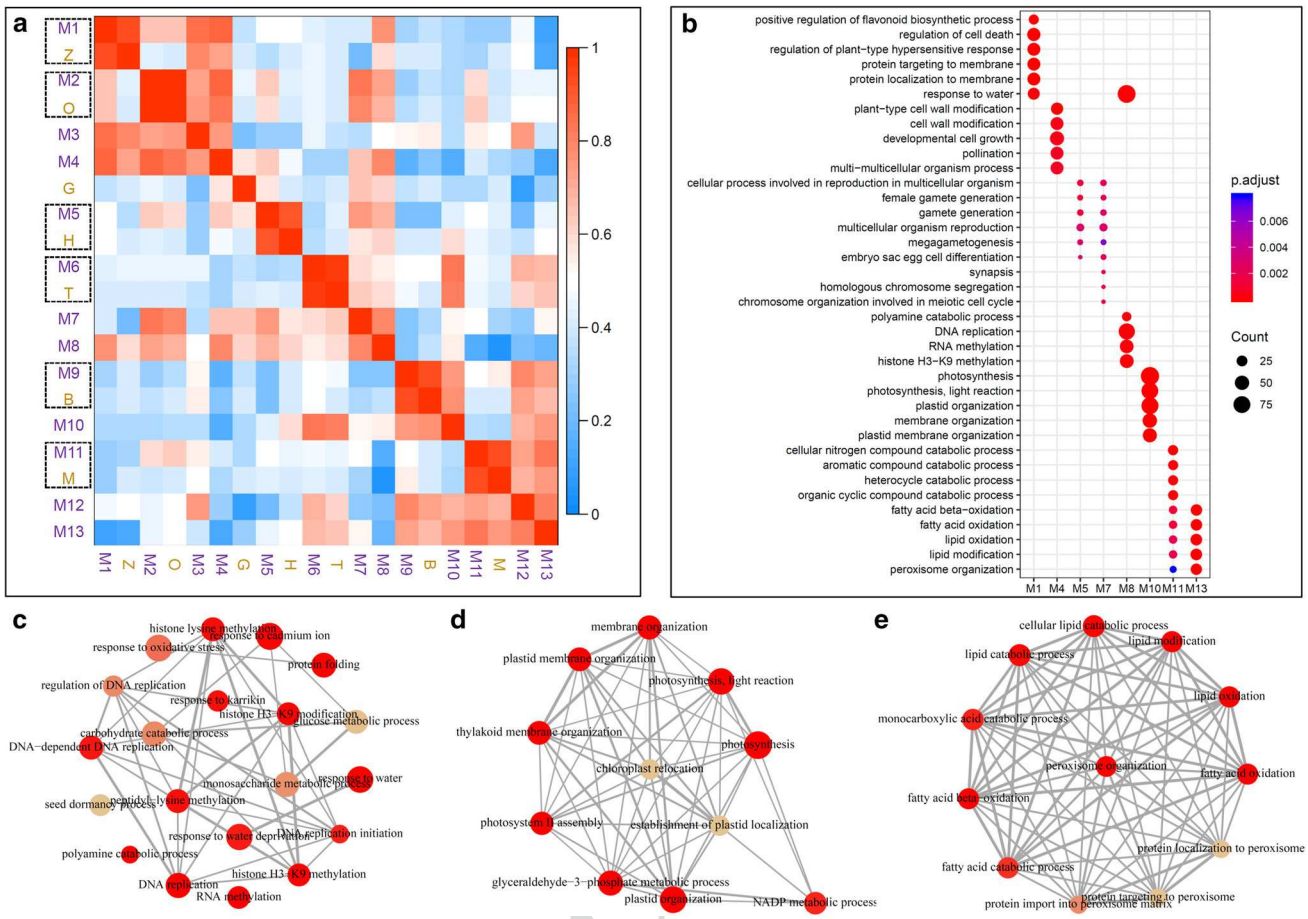

Fig. 3 Cluster analysis and functional annotation of stage-associated gene modules. **a** Eigengene adjacency heatmap of the consensus eigengene network that includes thirteen stage-associated modules and seven embryo development stages. The color scheme, from red through white to blue, indicates the level of correlation, from high to low. M1–13 represent thirteen stage-specific modules (purple text). Zygote—Z, octant—O, globular—G, heart—H, torpedo—T, bent—B, and mature—M represent the seven developmental stages (orange text). Related stage-specific gene modules and embryo development stages are grouped by a dashed black box. **b** A visual representation of GO enrichment for the DEGs with enriched GO terms in the biological process category (on the *y*-axis) plotted against each module (*x*-axis). No significant GO terms were enriched in M2, 3, 6, 9, 12. The GO terms with a significant p (<0.01) are shown. Dot size represents related gene count number. **c**–**e** Interact graphs of enriched GO terms in different modules. GO terms with p value (corrected) <0.01 were selected and applied to Cytoscape software to generate GO interact graphs. **c** M8 (zygote to torpedo), **d** M10 (heart and torpedo), **e** M13 (torpedo, bent, and mature). Colors from red to yellow indicate p value from low to high. Thickness of lines indicates the relative strength of the association between two GO terms

categories (bins) for all DEGs associated with embryogenesis. This module was downloaded from the MapManStore server including known *Arabidopsis* seed development-related genes and their assignment to corresponding bins. The outline of this module is shown in Figure S5. DEGs in different embryo developmental stages were assigned to different bins in this module, and their expression levels in comparison with the previous embryonic stage were visualized (Fig. S5a–f). The progression from Z to O was accompanied by an upregulation of most of the genes in this module. In the O and G stages, downregulation of almost all bins was observed, whereas from the G to T stage, the majority of genes in this module were predominantly upregulated. Finally, the majority of the seed development-associated genes are downregulated in the B and M stages. Altogether, the DEGs involved in this module followed this significant developmental stage-dependent expression pattern. Two typical examples representing this pattern include a hormone regulation signaling pathway and transcriptional factors involved in embryo development (Fig. S5). High expression peaks of most members in these two pathways are in O and T stages, which are closely connected to major events in embryogenesis, embryo patterning, and embryonic body plan establishment and elaboration.

In contrast to most categories in the seed development molecular networks listed above, DEGs from six pathway bins showed distinct expression patterns (Fig. S5, Data S4). A heatmap of related genes' dynamic expression across

Journal : Large 497 | Article No : 364 | Pages : 12 | MS Code : SPRE-D-18-00106 | Dispatch : 5-2-2019

embryo development was generated (Figs. 4b–d, S5). Photosynthesis and Calvin Cycle category BIN genes exhibited high expression in T and B stages, which is identical to the GO and KEGG enrichment analysis results for the M10 gene module (Figs. 3b, 4b, S6a). The other four pathway bins, including most seed storage proteins, late embryogenesis abundant (LEA) proteins, triacylglycerol (TAG), and fermentation-related genes, were significantly upregulated in B and M stages (Figs. 4c, d, S6b, c). Associated genes in these four pathways are suggested to play important roles in regulating seed maturation processes such as seed dormancy and storage reserve accumulation.

The observations from seed development networks revealed unique gene expression enrichment profiles involved in late and mature embryo development stages. Similar enrichment, such as methylation-related pathways, was observed in early and middle stages of embryo development. A list of histone methylation-related genes in *Arabidopsis* was established (Data S5), and a heatmap of their expression patterns in embryo development was generated (Fig. 4a). Like the GO enrichment analysis, most of these genes were enriched in early and middle stages of embryo development, suggesting epigenetic reprogramming during these stages of embryogenesis. Stringent regulation of the transcript levels of these genes from methylation, photosynthesis, and storage protein-related pathways reveals their key roles in embryogenesis.

Discussion

Due to the small size of early embryos, and the potential for contamination of embryo samples by surrounding seed coat and endosperm tissues, transcriptomic analysis of early plant embryogenesis has been challenging. In this study, we optimized and validated a method for hand dissection of *Arabidopsis* embryos from zygotes to mature stages. Our protocol offers an affordable, reliable, and precise approach to obtaining plant embryos for transcriptome analysis. The embryo isolation method we developed has been successfully adopted in different species and applied in several studies (Armenta-Medina et al. 2017; Moller et al. 2017; Quint et al. 2012; Venglat et al. 2011; Xiang et al. 2011a, b). Previous transcriptome studies of the full developmental time course of *Arabidopsis* embryogenesis have used microarray-based technologies (Belmonte et al. 2013; Xiang et al. 2011a), which are not as quantitative for gene expression as RNA-seq studies, and have the disadvantage that transcript isoform variants like different mRNA splice products cannot be detected. Other studies of embryo transcriptomes have used RNA-seq, but focused only on a few stages of embryogenesis (Autran et al. 2011; Hsieh et al. 2011; Nodine and Bartel 2012; Pignatta et al. 2014). For these reasons, our RNA-seq analysis of the complete time course of *Arabidopsis* embryogenesis is comprehensive and a valuable resource for studies of embryogenesis in the widely studied model plant *Arabidopsis* and has potential to aid embryogenesis studies in other plant species. This dataset has already allowed us to discover gene expression patterns correlated with stage transitions throughout embryo development. Furthermore, through our application of RNA-seq technology to embryogenesis, the transcriptomic reprogramming associated with embryo development is presented in a high-throughput and quantitative manner. This study revealed novel gene expression patterns and their correlation with stage transitions in embryo development. This comprehensive description of gene activities during

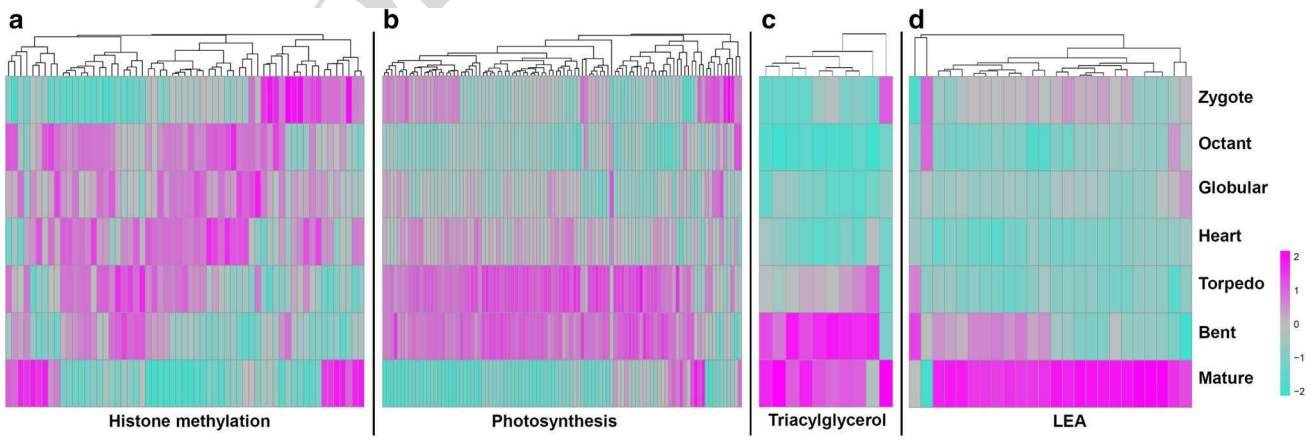

Fig. 4 Distinct expression trends of DEGs. Heatmaps of DEGs generated with the R pheatmap package. Rlog value of **a** histone methylation, **b** photosynthesis, **c** triacylglycerol (TAG), and **d** late embryogenesis abundant protein (LEA)-related DEGs were extracted, and the heatmaps were generated based on the *Z*-score transformed rlog value. Gene names are indicated to the right of the heatmap. Genes with similar expression patterns were clustered, and hierarchical trees were generated on the top of the heatmaps. The color scheme, from magenta through gray to cyan, indicates the level of correlation, from high to low

Journal : Large 497 | Article No : 364 | Pages : 12 | MS Code : SPRE-D-18-00106 | Dispatch : 5-2-2019

embryo development and the identification of stage-specific and multi-stage modules offer new insights and resources for future plant embryogenesis research.

Based on our DEG analysis, the transitions between every two consecutive stages of embryo development defined by this study are characterized by a significant transcriptome reprogramming (Data S1). The early and middle stages of embryo development are associated with embryonic patterning events (Jenik et al. 2007; Moller and Weijers 2009; Venglat et al. 2011), so the large shifts observed in expression and transcriptional regulation at these stages are not surprising. Consistent with our findings (Fig. S5, Data S1), previous reports indicated that early embryo patterning processes are promoted by auxin, the SHORT SUSPENSOR (SSP)/YODA (YDA)/WUSCHEL-LIKE HOMEOBOX (WOX) cascade, and miRNAs (Armenta-Medina et al. 2017; Bayer et al. 2009; Breuninger et al. 2008; Moller and Weijers 2009; Moller et al. 2017; Musielak and Bayer 2014; Nodine and Bartel 2010, 2012; Robert et al. 2018; Seefried et al. 2014; Ueda et al. 2011, 2017).

Studies in mammalian and plant systems have revealed an important role for epigenetic factors such as DNA methylation in regulating early development (Garcia-Aguilar et al. 2010; Rose and Klose 2014). Based on our DEG analysis of transcriptome datasets, DNA methylation is an important epigenetic factor observed in early and middle stages of embryo development, providing another proof for this conclusion (Fig. 3b). In mammals, the absence of DNA replication maintenance caused by methylation has been implicated in the developmental regulation of early embryos (Rose and Klose 2014). In this study, GO terms including DNA replication, RNA methylation, and histone 3 lysine 9 (H3-K9) methylation were enriched in gene module M8 (Fig. 3b, c), which represent the Z to T stages of embryo development (Fig. 3a), indicating that methylation-associated changes likely occur soon after fertilization. In plants, DNA methylation is found in three different sequence contexts: CG, CHG, and CHH. CHH methylation occurs through two pathways, the H3K9me2-dependent DNA methyltransferase CHROMOMETHYLASE 2 (CMT2) and RNA-directed DNA methylation (RdDM) (Jullien et al. 2012). Both CHH methylation pathways were enriched in M8 gene modules, which implicates a key role for this epigenetic modification in early and middle stages of embryo development, and is supported by a previous study that detected increasing CHH methylation in *Arabidopsis* embryo development (Jullien et al. 2012).

Many angiosperm seeds contain chlorophyllous embryos (called chloroembryos) (Yakovlev and Zhukova 1980) that contain components of the photosynthetic machinery, despite being embedded in ovule tissues within a highly osmotic environment (Puthur and Saradhi 2004). Our study revealed a significant enrichment of photosynthesis-related genes and pathways in M10 gene modules, representing late-stage H, T, and B embryos (Figs. 3b, d, 4b). In the M10 module, genes with annotated roles in plastid membrane organization were also enriched and are likely to be related to chloroplast biogenesis (Fig. 3b). These findings are consistent with recent research on *LEAFY COTYLEDON* 1 (*LEC1*). At early stages of seed development, *LEC1* regulates distinct gene sets to mediate the temporal transition between chloroplast biogenesis and photosynthesis (Pelletier et al. 2017). In our study, we also observed late embryo-enriched expression of *CHLOROPLAST STEM-LOOP BINDING PROTEIN 41 kDa* (*CSP41B*, AT1G09340) and *GERANYL(GERANYL)DIPHOSPHATE SYNTHASE 11* (*GGPPS11*, AT4G36810), two genes involved in photosynthesis and the Calvin Cycle (Figs. 4b, S5a, Data S1). Previous functional analysis of these genes has shown that they are expressed during late stages of embryo development and are required for chlorophyll production and embryo development (Ruiz-Sola et al. 2016). Thus, our transcriptome-based observations complement functional analysis of the importance of photosynthesis in late stages of embryo development.

Active photosynthesis in late-stage embryos contributes oxygen and energy supplies (ATP/NADPH), which might be an important resource supporting subsequent seed development and lipid synthesis (Ohlrogge and Browse 1995). The storage of lipids and fatty acids in the final stages of seed development is critical for supporting the energy demands of germination (Baud and Lepiniec 2010; Borek and Ratajczak 2010), and recent metabolic and transcriptional analysis demonstrated the importance of the embryo and endosperm in lipid and fatty acid biosynthesis (Han et al. 2017; Troncoso-Ponce et al. 2016). We observed that genes associated with lipid synthesis were enriched in the late embryogenesis (the M13 gene module) (Fig. 3b). Functional analysis of genes in this module should reveal the different roles of the embryo in storage of lipids and fatty acids during late embryonic development. The enrichment of triacylglycerols (TAGs), one of the main forms of energy storage in living organisms, was supported by gene expression patterns in B- and M-stage embryos (Fig. 4c). Similar trends were observed in other energy-related gene categories, such as fermentation-related genes (Fig. S6c). Late embryogenesis abundant (LEA) proteins accumulate at high levels during late stages of embryo development and are associated with dehydration tolerance (Dure and Galau 1981; Tunnacliffe and Wise 2007). In this study, most LEA proteins belong to mature-stage-specific gene modules and showed enrichment specifically in the mature-stage embryos (Fig. 4d).

Journal : Large 497 | Article No : 364 | Pages : 12 | MS Code : SPRE-D-18-00106 | Dispatch : 5-2-2019

Conclusion

This study presents an optimized method for characterizing the transcriptome of developing *Arabidopsis* embryos from zygote to maturity. RNA-seq analysis supported a significant transcriptional reprogramming during embryogenesis. A comprehensive examination of gene activities defined stage-specific and multi-stage gene modules and coexpression gene networks. Methylation, initiation of photosynthesis, and storage/energy-related protein activation were identified as three signature gene activities associated with early/middle, late, and mature stages of embryo development, respectively. The dynamic transcriptome and genes associated with different metabolic pathways in embryogenesis were revealed, providing a landscape of gene expression for ongoing efforts to unravel the genetic regulation of plant embryogenesis.

Author contribution statement RD and DX designed and coordinated the study. DX performed experiments. PG, DX, and TDQ performed data analysis, prepared the figures, and wrote the manuscript. EW participated in bioinformatics data analysis. PV and PKP assisted with data analysis. CSG edited the manuscript and contributed to interpretation of results. All authors read and approved the final manuscript.

Acknowledgements This work was funded by the Aquatic and Crop Resource Development Research Division of the National Research Council of Canada (ACRD manuscript #56424). We thank Dr. Wentao Zhang for reviewing the manuscript and providing suggestions for its improvement.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Anders S, Pyl PT, Huber W (2015) HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169. https://doi.org/10.1093/bioinformatics/btu638

AQ2 Andrews S (2010) FASTQC. A quality control tool for high throughput sequence data. (unpublished, open source: http://www.bioinformatics.babraham.ac.uk/projects/fastqc)

Armenta-Medina A, Lepe-Soltero D, Xiang D, Datla R, Abreu-Goodger C, Gillmor CS (2017) *Arabidopsis* thaliana miRNAs promote embryo pattern formation beginning in the zygote. Dev Biol 431:145–151. https://doi.org/10.1016/j.ydbio.2017.09.009

Ashburner M et al (2000) Gene ontology: tool for the unification of biology. Gene Ontol Consort Nat Genet 25:25–29. https://doi.org/10.1038/75556

Autran D et al (2011) Maternal epigenetic pathways control parental contributions to *Arabidopsis* early embryogenesis. Cell 145:707–719. https://doi.org/10.1016/j.cell.2011.04.014

Baud S, Lepiniec L (2010) Physiological and developmental regulation of seed oil production. Prog Lipid Res 49:235–249. https://doi.org/10.1016/j.plipres.2010.01.001

Bayer M, Nawy T, Giglione C, Galli M, Meinnel T, Lukowitz W (2009) Paternal control of embryonic patterning in *Arabidopsis* thaliana. Science 323:1485–1488. https://doi.org/10.1126/science.1167784

Belmonte MF et al (2013) Comprehensive developmental profiles of gene activity in regions and subregions of the *Arabidopsis* seed. Proc Natl Acad Sci USA 110:E435–E444. https://doi.org/10.1073/pnas.1222061110

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bonner WA, Hulett HR, Sweet RG, Herzenberg LA (1972) Fluorescence activated cell sorting. Rev Sci Instrum 43:404–409

Borek S, Ratajczak L (2010) Storage lipids as a source of carbon skeletons for asparagine synthesis in germinating seeds of yellow lupine *Lupinus luteus* L. J Plant Physiol 167:717–724. https://doi.org/10.1016/j.jplph.2009.12.010

Braybrook SA, Harada JJ (2008) LECs go crazy in embryo development. Trends Plant Sci 13:624–630. https://doi.org/10.1016/j.tplants.2008.09.008

Breuninger H, Rikirsch E, Hermann M, Ueda M, Laux T (2008) Differential expression of WOX genes mediates apical-basal axis formation in the *Arabidopsis* embryo. Dev Cell 14:867–876. https://doi.org/10.1016/j.devcel.2008.03.008

Casson S, Spencer M, Walker K, Lindsey K (2005) Laser capture microdissection for the analysis of gene expression during embryogenesis of *Arabidopsis*. Plant J: Cell Mol Biol 42:111–123. https://doi.org/10.1111/j.1365-313X.2005.02355.x

Deal RB, Henikoff S (2010) A simple method for gene expression and chromatin profiling of individual cell types within a tissue. Dev Cell 18:1030–1040. https://doi.org/10.1016/j.devcel.2010.05.013

Dobin A et al (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. https://doi.org/10.1093/bioinformatics/bts635

Dure L, Galau GA (1981) Developmental biochemistry of cottonseed embryogenesis and germination: XIII. REGULATION OF BIOSYNTHESIS OF PRINCIPAL STORAGE PROTEINS. Plant Physiol 68:187–194

Garcia-Aguilar M, Michaud C, Leblanc O, Grimanelli D (2010) Inactivation of a DNA methylation pathway in maize reproductive organs results in apomixis-like phenotypes. Plant Cell 22:3249–3267. https://doi.org/10.1105/tpc.109.072181

Goldberg RB, de Paiva G, Yadegari R (1994) Plant embryogenesis: zygote to seed. Science 266:605–614. https://doi.org/10.1126/science.266.5185.605

Han C, Zhen S, Zhu G, Bian Y, Yan Y (2017) Comparative metabolome analysis of wheat embryo and endosperm reveals the dynamic changes of metabolites during seed germination. Plant Physiol Biochem 115:320–327. https://doi.org/10.1016/j.plaphy.2017.04.013

Heiman M et al (2008) A translational profiling approach for the molecular characterization of CNS cell types. Cell 135:738–748. https://doi.org/10.1016/j.cell.2008.10.028

Hsieh TF et al (2011) Regulation of imprinted gene expression in *Arabidopsis* endosperm. Proc Natl Acad Sci USA 108:1755–1762. https://doi.org/10.1073/pnas.1019273108

Huh JH, Bauer MJ, Hsieh TF, Fischer R (2007) Endosperm gene imprinting and seed development. Curr Opin Genet Dev 17:480–485. https://doi.org/10.1016/j.gde.2007.08.011

Jenik PD, Gillmor CS, Lukowitz W (2007) Embryonic patterning in *Arabidopsis* thaliana. Annu Rev Cell Dev Biol 23:207–236. https://doi.org/10.1146/annurev.cellbio.22.011105.102609

Jullien PE, Susaki D, Yelagandula R, Higashiyama T, Berger F (2012) DNA methylation dynamics during sexual reproduction in *Arabidopsis* thaliana. Curr Biol 22:1825–1830. https://doi.org/10.1016/j.cub.2012.07.061

Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30

Kerk NM, Ceserani T, Tausta SL, Sussex IM, Nelson TM (2003) Laser capture microdissection of cells from plant tissues. Plant Physiol 132:27–35. https://doi.org/10.1104/pp.102.018127

Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinf 9:559. https://doi.org/10.1186/1471-2105-9-559

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8

Moller B, Weijers D (2009) Auxin control of embryo patterning. Cold Spring Harb Perspect Biol 1:a001545. https://doi.org/10.1101/cshperspect.a001545

Moller BK et al (2017) Auxin response cell-autonomously controls ground tissue initiation in the early *Arabidopsis* embryo. Proc Natl Acad Sci USA 114:E2533–e2539. https://doi.org/10.1073/pnas.1616493114

Musielak TJ, Bayer M (2014) YODA signalling in the early *Arabidopsis* embryo. Biochem Soc Trans 42:408–412. https://doi.org/10.1042/bst20130230

Nodine MD, Bartel DP (2010) MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. Genes Dev 24:2678–2692. https://doi.org/10.1101/gad.1986710

Nodine MD, Bartel DP (2012) Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. Nature 482:94–97. https://doi.org/10.1038/nature10756

Ohlrogge J, Browse J (1995) Lipid biosynthesis. Plant Cell 7:957–970. https://doi.org/10.1105/tpc.7.7.957

Palovaara J et al (2017) Transcriptome dynamics revealed by a gene expression atlas of the early *Arabidopsis* embryo. Nat Plants 3:894–904. https://doi.org/10.1038/s41477-017-0035-3

Pelletier JM et al (2017) LEC1 sequentially regulates the transcription of genes involved in diverse developmental processes during seed development. Proc Natl Acad Sci USA 114:E6710–E6719. https://doi.org/10.1073/pnas.1707957114

Pignatta D, Erdmann RM, Scheer E, Picard CL, Bell GW, Gehring M (2014) Natural epigenetic polymorphisms lead to intraspecific variation in *Arabidopsis* gene imprinting. eLife 3:e03198. https://doi.org/10.7554/eLife.03198

Puthur JT, Saradhi PP (2004) Developing embryos of Sesbania sesban have unique potential to photosynthesize under high osmotic environment. J Plant Physiol 161:1107–1118. https://doi.org/10.1016/j.jplph.2004.03.002

Quint M, Drost HG, Gabel A, Ullrich KK, Bonn M, Grosse I (2012) A transcriptomic hourglass in plant embryogenesis. Nature 490:98–101. https://doi.org/10.1038/nature11394

Robert HS et al (2018) Maternal auxin supply contributes to early embryo patterning in *Arabidopsis*. Nat Plants 4:548–553. https://doi.org/10.1038/s41477-018-0204-z

Rose NR, Klose RJ (2014) Understanding the relationship between DNA methylation and histone lysine methylation. Biochem Biophys Acta 1839:1362–1372. https://doi.org/10.1016/j.bbagrm.2014.02.007

Ruiz-Sola MA, Barja MV, Manzano D, Llorente B, Schipper B, Beekwilder J, Rodriguez-Concepcion M (2016) A single *Arabidopsis* gene encodes two differentially targeted geranylgeranyl diphosphate synthase isoforms. Plant physiology 172:1393–1402. https://doi.org/10.1104/pp.16.01392

Seefried WF, Willmann MR, Clausen RL, Jenik PD (2014) Global regulation of embryonic patterning in *Arabidopsis* by MicroRNAs. Plant Physiol 165:670–687. https://doi.org/10.1104/pp.114.240846

Slane D et al (2014) Cell type-specific transcriptome analysis in the early *Arabidopsis* thaliana embryo. Development 141:4831–4840. https://doi.org/10.1242/dev.116459

Thimm O et al (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J: Cell Mol Biol 37:914–939

Troncoso-Ponce MA, Barthole G, Tremblais G, To A, Miquel M, Lepiniec L, Baud S (2016) Transcriptional activation of two delta-9 palmitoyl-ACP desaturase genes by MYB115 and MYB118 Is critical for biosynthesis of omega-7 monounsaturated fatty acids in the endosperm of *Arabidopsis* seeds. Plant Cell 28:2666–2682. https://doi.org/10.1105/tpc.16.00612

Tunnacliffe A, Wise MJ (2007) The continuing conundrum of the LEA proteins. Die Naturwissenschaften 94:791–812. https://doi.org/10.1007/s00114-007-0254-y

Ueda M, Zhang Z, Laux T (2011) Transcriptional activation of *Arabidopsis* axis patterning genes WOX8/9 links zygote polarity to embryo development. Dev Cell 20:264–270. https://doi.org/10.1016/j.devcel.2011.01.009

Ueda M et al (2017) Transcriptional integration of paternal and maternal factors in the *Arabidopsis* zygote. Genes Dev 31:617–627. https://doi.org/10.1101/gad.292409.116

Venglat P et al (2011) Gene expression analysis of flax seed development. BMC Plant Biol 11:74. https://doi.org/10.1186/1471-2229-11-74

Xiang D et al (2011a) Genome-wide analysis reveals gene expression and metabolic network dynamics during embryo development in *Arabidopsis*. Plant Physiol 156:346–356. https://doi.org/10.1104/pp.110.171702

Xiang D et al (2011b) POPCORN functions in the auxin pathway to regulate embryonic body plan and meristem organization in *Arabidopsis*. Plant Cell 23:4348–4367. https://doi.org/10.1105/tpc.111.091777

Yakovlev MS, Zhukova GY (1980) Chlorophyll in embryos of angiosperm seeds, a review. Bot Not 133:323–336

Yang H, Xiang D, Venglat S, Cao Y, Wang E, Selvaraj G, Datla R (2009) PolA2 is required for embryo development in *Arabidopsis*. Botany 87:626–634. https://doi.org/10.1139/B09-028

Zhang C, Barthelson RA, Lambert GM, Galbraith DW (2008) Global characterization of cell-specific gene expression through fluorescence-activated sorting of nuclei. Plant Physiol 147:30–40. https://doi.org/10.1104/pp.107.115246

Journal : Large 497	Article No : 364	Pages : 12	MS Code : SPRE-D-18-00106	Dispatch : 5-2-2019

Journal:	**497**
Article:	**364**

Author Query Form

Please ensure you fill out your response to the queries raised below and return this form along with your corrections

Dear Author

During the process of typesetting your article, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the 'Author's response' area provided below

Query	Details Required	Author's Response
AQ1	Please check and confirm that the article note is correctly identified.	
AQ2	Kindly provide accessed date for the reference Andrews (2010).	