

# Transcriptome Profiling Reveals Spatial-temporal Dynamics of Gene Expression Essential for Soybean Seed Development

**Hengyou Zhang**

Henan Agricultural University

**Zhenbin Hu**

Donald Danforth Plant Science Center

**Yuming Yang**

Henan Agricultural University

**Xiaoqian Liu**

Henan Agricultural University

**Haiyan Lv**

Henan Agricultural University

**Bao-Hua Song**

Department of Biological Sciences

**Yong-qiang Charles An**

Donald Danforth Plant Science Center

**Zhimin Li**

Henan Agricultural University

**Dan Zhang** (✉ [zhangd@henau.edu.cn](mailto:zhangd@henau.edu.cn))

Collaborative Innovation Center of Henan Grain Crops, College of Agronomy, Henan Agricultural University, Zhengzhou 450002, China <https://orcid.org/0000-0001-6119-9533>

---

## Research article

**Keywords:** Soybean, Glycine max, seed development, transcriptome, spatial/temporal gene expression

**Posted Date:** December 18th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-128593/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** Seeds are the economic basis of oilseed crops, especially for soybean, thus far the most widely cultivated oilseed crop worldwide. Seed development is accompanied with a multitude of diverse cellular processes and revealing the underlying regulatory activities is critical for seed improvement.

**Results:** Here, we profiled transcriptomes of developing seeds (20, 25, 30, 40 days after flowering) representing key points of seed development from early to full development. We identified a set of highly-abundant genes and highlighted the importance of these genes to support nutrient accumulation and transcriptional regulation in developing seeds. We identified 8,925 differentially expressed genes that exhibited temporal expression patterns over the course and had expression specificities in distinct tissues including seeds and non-seed tissues (roots, stems, leaves). Genes with specificities to non-seed tissues have tissue-specialized roles while remain relatively low transcript abundance in developing seeds, exhibiting their supportive roles spatially for seed development. Co-expression network analysis identified several under-explored genes in soybean that bridge tissue-specific gene modules.

**Conclusions:** Our study provides a global view of gene activities and biological processes critical for seed formation in soybean and prioritizes a set of genes for further study. The results shed insight into the mechanism controlling seed development and storage reserves.

## 1. Introduction

Soybean [*Glycine max* (L.) Merr.] is the most widely cultivated oilseed crop worldwide, and the seeds serve as a major source of highly valuable protein and edible oil for human and animal consumption [1]. Soybean production (bushels per acre) nearly doubled since 1987 [2] and it accounts for approximately 68% and 27% of world plant protein, oil and meal production respectively ([www.fas.usda.gov/data/oilseeds](http://www.fas.usda.gov/data/oilseeds)), reflecting the key role of soybean seeds in feeding the world. It is estimated that 70% increase in soybean production is needed to meet the increasing demand for plant-based protein in the next decades [3, 4]. It is important to understand seed development with a goal for a continuous increase in seed production, maintaining its sustainable role as the major source provider of protein and oil worldwide.

Soybean seeds are composed of functionally distinct tissues, including the embryo, embryonic cotyledon, and seed coat. Those tissues are successively differentiated and grown during the maturation phase [5]. The filial embryo from the double fertilization eventually grows to the cotyledons that serve as a storage organ for food reserves mainly oil and protein. The seed coat is maternal tissue surrounding the embryo and cotyledon to serve for protection and nutrient delivery of photo-assemblies to support the developing embryo [5, 6]. Ontogenesis and development of these tissues are molecularly programmed and correlated with parallel changes of nutrients level in seeds. Seed storage reserves such as storage proteins and lipids are synthesized and accumulated during seed filling mainly from early to later maturation stages [7]. Overall, seed development is a complex developmental process accompanied with of a multitude of

molecular activities simultaneously occurring in coordinating tissues' development and storage reserve metabolism.

Thus far, several genes involved in seed nutrient accumulation such as bZIP123 involved lipid accumulation and GA20OX-associated seed weight and oil accumulation [8–11] and genes associated with seed development [12] have been identified. The conservation in some regulatory processes of seed development between soybean and model plant *Arabidopsis* was revealed [7, 13]. The homolog-based cloning approach has been successfully used in identifying related genes in soybean [14, 15]. However, it has been indicated that functions for many of soybean genes, with their *Arabidopsis* homologous genes having demonstrated roles in seed development and nutrient accumulation, have been modified in the palaeopolyploid soybean, such as *ABSCISIC ACID INSENSITIVE 3b* (*ABI3b*) functioning like LEAFY COTYLEDON 2 (*LEC2*) [16] and WRINKLED1 (*WRI1*) also associated with plant architecture other than its homologous role in affecting oil production [14]. Each of these major regulators such as LEC1, ABI3, ABA-RESPONSIVE ELEMENT BINDING PROTEIN3 (*AREB3*), BASIC LEUCINE ZIPPER67 (*bZIP67*) putatively regulated thousands of targets, as revealed by Chip-Seq assays [17]; meanwhile, these target gene sets involved in the control of diverse developmental processes by their mosaic combinations as coordinated by the regulators, and they functioned temporally but their regulated processes were partially overlapped spatially [17]. These studies suggested that soybean evolves a complex regulatory network during seed development that is, in part, different from the one in *Arabidopsis* by which thus far the current knowledge in plant seed development are mostly acquired [18, 19]. Hence regulatory mechanism of seed development in soybean remains largely to be explored. Continuous studies comprehensively investigating seed developmental processes are needed.

Sequencing-based transcriptome profiling (RNA-Seq) was proven an effective approach in dissecting regulatory mechanisms of complex traits [20] and has been applied to investigate the changes in gene expression during seed development in legumes and oilseed plants [5]. For example, transcriptomics has been used to reveal hub genes a set of regulatory genes putatively involved in soybean oil and protein accumulation [21, 22]. Here, we profiled transcriptomes of progressively developing seeds at 20, 25, 30, 40 days after flowering (DAF) to extend our knowledge of the molecular mechanism that controls seed development and nutrient production. We uncovered a variety of gene activities during the dynamic processes of seed development and identified multiple gene sets that exhibited spatial and temporal expressions during the period. The results provide a global view of gene activity essential to the developmental process in soybean, providing insight into the regulatory mechanism of forming a seed.

## 2. Materials And Methods

### 2.1 Plant growth and sample collection

Soybean variety Williams 82 was used through this study. Seeds were planted in Maozhuang Experimental Station of Henan Agricultural University in 2018 and planted in a three-row plot, with 200 cm row in length and 50-cm row spacing. A randomized block design was used for filed planting

design. Developing seeds at 20, 25, 30, and 40 DAF (days after flowering) were collected and flash-frozen with liquid nitrogen, and stored at -80 °C freezer for RNA extraction. For each time point, two-three developing seeds were sampled and pooled as one biological replicate, and three biological replicates were used per time point.

## 2.2 RNA isolation, library construction, and transcriptome sequencing

Total RNA was isolated using Trizol reagent (Invitrogen, CA, USA) following the manufacturer's protocol. Prior to library construction, total RNA was treated with RNase-free DNase I (New England Biolabs, Ipswich, MA, USA) to remove any contaminating genomic DNA. Library construction was performed following the protocol as previously described [23]. The quantified library was sequenced on the Illumina HiSeq 2500 sequencing platform (Illumina Inc, San Diego, CA, USA) at Biomarker Technologies (Beijing, China), and produced 200-bp paired-end reads.

## 2.3 Data processing and differential expression

Raw sequencing data quality were evaluated with FastQC v0.11.5 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) followed by removal of low quality reads (quality value < 20) using Trimmomatic v0.33 [24]. The trimmed reads were aligned to the Williams 82 soybean reference genome (Wm82.a2.v1) [25] with TopHat (v2.1.1) using the minimum intron size (-i) parameter 30 and the maximum intron size (-l) 15000. The numbers of reads mapped to annotated genes were counted using featureCounts [26]. Analysis to determine differentially expressed genes was performed using EdgeR [27], and genes with  $fdr \leq 0.01$  and fold change  $\geq 1.5$  were considered significantly differentially expressed between two conditions. Enrichment analyses for GO terms and KEGG pathways were performed at Gene Ontology (<http://geneontology.org/docs/go-enrichment-analysis/>) and KOBAS 3.0 (<http://kobas.cbi.pku.edu.cn/>), respectively, as previously described [28].

## 2.4 Quantitative real-time PCR

Expression of genes was examined with the same tissues for RNA-Seq using qRT-PCR as previously described [29]. Briefly, total RNA was isolated from the tissues using the RNA simple Total RNA Kit (TaKaRa, Japan), and 1 µg of RNA was treated with RNase-free DNase I (TaKaRa, Japan) before cDNA synthesis. The first strand of cDNA was synthesized using the SuperScript III First-Strand Synthesis System (Invitrogen, USA) following the manufacturer's instructions. Gene expression was determined using the Bio-Rad CFX96 Touch Real-Time PCR System (Bio-Rad, California, USA). The PCRs contained 5 µL of the first-strand cDNA, 0.5 µL of 10 µmol L<sup>-1</sup> gene-specific primers, and 10 µL of Real-Time PCR SYBR Mix using. The PCR conditions were as follows: 94 °C for 3 min and 40 cycles at 94 °C for 15 s and 60 °C for 15 s. The soybean *tubulin* gene (GenBank: AY907703.1) was amplified as an internal reference, and a negative control reaction was performed using water instead of cDNA. Three biological replicates per sample were used, and each reaction was performed in triplicate.

## 2.5 Co-expression network construction

Differentially expressed genes were extracted from the co-expression network in SoyNet that was pre-constructed using transcriptomes from diverse tissues and seed compartments [30]. The co-expressed genes were re-constructed into modules using Molecular Complex Detection (MCODE) [31] using default parameters. The module network was visualized using Cytoscape [32].

### 3. Results And Discussion

## 3.1 Dramatic morphological and reserves changes during seed development

Whole soybean developing seeds at four sequential time points (20, 25, 30, 40 days after flowering (DAF)) that cover major stages of the seed development were sampled in triplicate (Fig. 1A). This period of seed development is coincident with seed growth stages from R5 (beginning seed) to R6 (full-size seed) of soybean development, which are morphologically comparable to early and mid-maturation stages as described in JY Lin, BH Le, M Chen, KF Henry, J Hur, TF Hsieh, PY Chen, JM Pelletier, M Pellegrini, RL Fischer, et al. [7]. It is also cooccurring with rapid accumulation of lipid and storage proteins [20, 33]. thus the seed size and seed weight increased significantly as it grew towards a fully developed seed (Fig. 1A). Two major storage reserves, protein and oil, in soybean seeds were quantified during this period. As shown in **Fig. 1B**, accumulation of storage oil was approximately 14.11% in seed at D20 and continuously increased to 17.85% at D40, with the greatest increase by 1.85% occurring in the period from D20 to D25 relative to the later periods at D25-D30 by 1.11% and D30-D40 by 0.78%. In contrast, as the seed matured, the storage protein content was 50.87% in seed at D20 and decreased to 45.83% (5.04% decrease) in seeds at D25, and continuously declined to 42.26% at D30 and to 41.77% at D40. Thus, oil content in the developing seeds increased more dramatically at early and less dramatically at the late period of seed filling, and vice versa for protein decrease. Reduction in the relative percentage of protein content over the period is due to its negative correlation with oil content in soybean seeds [34], on the contrary, its actual protein levels increased rapidly between 10–40 DAF [35]. Hence, this period of seed development with rapid morphological changes (such as size, weight) and major metabolic changes (such as oil, protein) should be governed by a variety of cellular processes/genes in seeds [17, 20].

## 3.2 Transcriptomic abundance shifts during seed development

An in-depth understanding of biological processes associated with seed development and reserves accumulation in soybean seeds is important for improving seed quality and yielding potential. Given the multiplicity of genes controlling seed development coupled with seed filling, time-course transcriptomic profiling of four-stage seeds that covered the representative period of seed development was carried out with an aim to obtain a global view of regulatory mechanism underlying this complex development process. In total, ~ 313.4 million sequencing reads from all sampled seeds were generated, with 26.1 million reads per sample (**Table S1**). Over 97.43% (25.4 million), on average, were kept after trimming the low-quality reads and used for read alignment. Uniquely mapped reads were used to

quantify gene expression abundance as normalized by Count Per Million (cpm). After removing reads/transcript < 5, we identified that, an average, 3,5000 (62.5%) out of 56,044 gene models were expressed in developing seeds. We observed that a gradient reduction (approximately 700 genes, 2% of total) in the number of expressed genes over the course of seed development. The reduction could be related to the shutdown in many transcription processes as the seeds grew toward dormancy [36] and it would be intriguing to determine if this reduction is associated with a programmed increase of methylation levels during soybean seed development [7].

Both clustering dendrogram analysis and multi-dimensional scale indicated that gene expression levels among the replicated samples were generally highly related relative to those between time-point samples (Fig. 2A, B), indicating the gene expression specificity to the time points. Based on transcript abundance, approximately 61.07% expressed gene were in the range of 0–10 cpm, 36.41% were in the range of 10–100 cpm, and only 2.53% of genes are highly expressed with cpm  $\geq$  100 (Fig. 2C). Interestingly, as seeds enlarged, the number of genes with an abundance  $\geq$  10 cpm generally decreased while the number of those with an abundance  $\leq$  10 cpm gradually increased. Among the expressed genes, 32,326 were expressed across all four-stage developing seeds. There were 942, 329, 200, and 546 genes that are specifically expressed at D20, D25, D30, and D40, respectively (Fig. 2D). These results suggested a temporal shift in transcript abundance for many highly expressed genes as seeds developed, which might be reflective of the foregoing morphological and/or metabolic changes during seeds development. All the raw sequencing data were deposited to the Sequence Read Archive of the National Center for Biotechnology Information (BioProject: PRJNA551573).

A cluster dendrogram analysis showed that replicates from each condition were closely clustered, supported by a similar separation pattern in a multidimensional scaling (MDS) plot of log-FC values. We further performed a qRT-PCR for randomly selected 16 differentially expressed genes (Table S4) to validate and a high correlation ( $r = 0.87$ ) between the RNA-Seq and qRT-PCR indicates the robustness of our RNA-Seq result.

### 3.3 Genes with high transcript abundance during seed development and filling

We next characterized the genes that were constitutively, highly expressed across all four time points. In total, 443 genes with mRNA abundance cpm  $\geq$  100 were identified [37] (Table S3). GO terms enrichment analysis indicated that biosynthetic processes (BP) essential for nutrients accumulation were enriched, such as amino acid metabolism, carbohydrate metabolism, glycolysis, and hexose metabolism (Table S3). Accordingly, this gene set was also enriched for molecular function (MF) terms of nutrient reservoir activity cellular component (CC) terms related to lipid storage body, endoplasmic reticulum (ER), ribosome where fatty acids were primarily stored and synthesized [38, 39]. If we relaxed the threshold to cpm  $\geq$  50, the number of newly defined highly expressed genes was increased by three folds to 1,386 genes (Table S3). Accordingly, the numbers of the expressed genes that fell in the majority of those enriched terms were approximately doubled or tripled, such as hexose metabolic process (8 to 22),

glycolysis (7 to 18) cellular nitrogen compound metabolic process (18 to 36), cellular amino acid biosynthetic process (15 to 32), Ribosomal proteins (47 to 132). It should be noted that no change in the numbers of the genes with the terms related to monolayer-surrounded lipid storage body (7), nutrient reservoir activity (11), and sulfur amino acid metabolic process (6 to 7) was observed, suggesting the importance of maintaining high transcript abundance of these genes for seed filling. Additional enriched GO terms for the 1,343 genes were identified, protein localization, glutamine family amino acid biosynthetic process, intracellular transport, gene silencing by RNA, vesicle-related components (COP-II, COP-I, coated vesicle, transport vesicle, ER to Golgi transport vesicle, membrane-bounded vesicle).

A close investigation of the 443 genes (cpm  $\geq$  100) revealed several genes with demonstrated roles in the accumulation of storage reserves, such as oil, protein, and carbohydrates. It is not surprising that some genes participating metabolism of protein and oil were identified. However, a many of the highly abundant genes during seed filling was rarely reported (**Table S3**), such as oil accumulation-related genes, *FAD2* (467.83–2000.69) catalyzing oleic acid accumulation [40] and OLEOSIN proteins (373.03–6190.57 cpm) encoding oil body surrounded lipid storage body [29, 39], metabolic genes for biosynthesis of leucine, thiamin (vitamin B1), and Ribosomal proteins. It is noted that genes encoding storages proteins such as cupin family proteins rank the highest transcript abundance, ranging 34000–100000 cpm, in developing seeds, followed by xylem bark cysteine peptidase 3 (*Glyma.08G116300* with 28399.81 cpm), trypsin inhibitor 1 (222.79 ~ 18559.51 cpm), and oleosin family proteins. Oleosins were highly abundant (373.03–6190.57 cpm). Those proteins in soybean appear to be redundant and serve as a surfactant of the oil body [41, 42]. A recent study on the oleosin on chr20 [29] indicates that it is likely regulated by LEC2 [15], by which oil production is enhanced through increasing oil body turnover. Investigation of the cis-regulatory sequences in the promoter region could be helpful to use this set of genes to achieve oil improvement. In contrast, a set of high-abundant trypsin inhibitor genes (222.79–18559.51 cpm) could be silenced by gene editing to inhibit their expression or mutate them for better seed quality because trypsin inhibitors have an antinutritional effect [43].

More attention could be given to nutrient transporters that mediate nutrient flux into the developing embryo, such as sugar transporter *SWEETs* (274.98–943.19 cpm), and proton pump interactor 1. *SUCROSE SYNTHASEs* (121.5–1186.21 cpm) controlling starch synthesis. Recently studies highlighted the essential role of SWEET proteins in transporting sugar from maternal to embryonic tissues to support the seed development and oil accumulation [12, 44] and indicated the agriculturally critical role in causing high oil accumulation in cultivated soybean seeds relative to low oil in wild soybean [11]. High abundances of these transporters together with those related studies suggested that the rate of sugar transport and sugar accumulation is important for nutrient reserves accumulation in seeds. Further investigation of downstream affected steps/components may help in identifying the major route of carbon flow for oil accumulation or carbon partitioning between oil and protein in soybean, and perhaps other plant species such as maize and rice [45, 46]. Several *LOX1* genes (541.27–16226.05 cpm) were in high transcript abundance here but the homologous proteins were not detected during seed filling in wheat and maize endosperm [47, 48], suggesting a lineage-specific role of LOX proteins in soybean seeds,

possibly related to seed storage protein and lipids such as unsaturated fatty acid [49, 50] or special tastes and odors [51].

Transcription factors such as ABA INSENSITIVE 3 (ABI3; *Glyma.08G357600*; *Glyma.18G176100*) and its interacting regulator bZIP67 were identified in high abundance. High abundance of both regulators may be due to increasing needs for their involved cell growth and accumulation of storage reserves as seeds enlarged. Indeed, both regulators have functions partially in common and were essential for the initiation of gene expression changes for biological processes critical for seed development such as gibberellic acid signaling, photosynthesis and maturation related processes [16, 17, 52, 53]. Abundance of LEC2s (cpm = 0) were undetectable in this study, possibly because they are highly specific to certain seed compartments and barely detected in whole seeds [15]. The well-documented LEC1 homologous genes, a central regulator of seed development [17], were not in high abundance but showed decreased expression over the period. Other high-abundance regulators with demonstrated roles in *Arabidopsis* but were less explored in soybean seeds were also identified, such as PLASTID TRANSCRIPTIONALLY ACTIVE 9 (PTAC9), TANDEM CCCH ZINC FINGER PROTEIN 4 (TCZF4), circadian rhythm regulators including TIME FOR COFFEE (TIC), LATE ELONGATED HYPOCOTYL 1 (LHY1), CONSTANS-LIKE 4 (COL4). Determination of the function of the regulators and the cognate cis-regulatory elements may facilitate identification of seed regulatory networks that, in part, contribute to the regulation of genes temporally and spatially in seeds [17, 18, 54]. All highly expressed genes and the functional annotations were listed in **Table S2**.

### 3.4 Genes were temporally regulated over the seed development

To comprehensively identify genes involved in developmental processes during seed filling, we conducted a pairwise comparison of transcriptomic profiles among the different-stage seeds. We observed a gradient increase in the number of differentially expressed genes (DEGs) between the two neighboring time points as the seeds developed, with 523 in D25vsD20, 1,486 in D30vsD25, and 2,485 in D40vsD30 (Fig. 3). The numbers of DEGs increased when comparisons were made across non-adjacent time points (over 10 d interval), including 4,079 in D30 vs D20, 5,617 in D40 vs D25, and 7,670 in D40 vs D20. In total, the pairwise comparison identified 8,925 non-redundant DEGs that are required to program soybean seed development from early to full seed development (**Table S2**).

Clustering analysis was performed for all 8,925 DEGs based on expression and shown in a heat map (Fig. 4). The analysis classified the DEGs to seven groups (A to G) exhibiting different gene expression patterns. The difference in the temporal expression patterns suggested stage-specific roles of these grouped DEGs throughout seed development. Groups A and G represented the two largest clusters (2,182 and 4,945 DEGs, respectively) while exhibited contrast expression patterns, with the DEGs in cluster A displaying predominant and highest expression level at D40 and those in group G showing a continuous decrease in gene expression as seeds matured. Enrichment analysis for cluster A genes showed that D40-specific genes were enriched for cell communication (GO:0007154), transcription (GO:0006351), and the biological process enriched occurred in the cell periphery and plasma membrane (Fig. 4B). This stage

represented the later stage of producing and storing reserves and the active period of the shift to maturation in which multiple programming might be initiated, such as cessation of organ differentiation and cell division, preparation for seed desiccation [55]. A representative gene with specifically high expression at D40 is ABI3 (Fig. 4C) [16], which is an essential regulator in ABA (abscisic acid) signaling primarily involved in seed maturation and dormancy in *Arabidopsis* [56]. Its high abundance during D20-D30 and dramatic increase toward D40 suggested its critical role during seed development, especially in the later stage of maturation-related processes. Indeed, ABI3 was recently demonstrated that it is a key transcription factor modulating the transition from the morphogenesis to the maturation phase by combinatorial interaction with other seed transcription factors [17]. Further study of the co-expressed genes or binding sites [17, 21] would provide an opportunity to understand how the soybean seeds are regulated. Identification of the key seed development regulators (such as ABI3) indicated the robustness of our analysis and suggested that other seed development-associated genes identified in our study deserve further attention. For example, a similar expression pattern was also observed in two SUSs (Sucrose synthase) which involved in cellulose and starch biosynthesis in seeds [57].

Genes classified in cluster B (810) were highly expressed at D40 that has a similar expression pattern as cluster A, except for those genes that were up-regulated beginning at D30. Enriched GO terms for cluster B are transmembrane transport (GO:0055085) and nutrient reservoir activity (GO:0045735), suggesting that those biological processes involved in nutrient transport such as transporters for sugar and amino acids and nutrient accumulation were active beginning at D30 and maintained at D40. The developmental period of D30-D40 is approximately coincident with the middle-maturation stage which represented the major period of accumulation for storages such as protein and lipids [7]. Expression pattern analysis showed that, despite varying expression patterns observed, those genes involved in the foregoing biological processes showed obvious upregulation during seed development and filling (Fig. 4C), such as cupin proteins, oleosins, amino acid permeases (APPs), FAD2, nitrate transporters, Seed Maturation Proteins (SMPs). The stark upregulation in expression is undoubtedly reflective of the critical roles of the related biological processes for storage reserve accumulation. Transcription factors involved in seed development were also upregulated in the cluster B, such as AUS/IAA transcription factor, YABBY, uncharacterized PLATZ. Interestingly, the genes involved in flowering (TFLs, Tof12, PHYA) showed an upregulation-downregulation shift during the course, suggesting their expanded roles during seed development but have yet been explored. This finding was partially supported by a recent study where *Arabidopsis TLF1* not only functioned as a shoot identity gene but also served as a signaling molecule that is essential in determining seed size, and loss-of-function mutant *tlf1-20* producing larger developing seeds than wild type by regulating endosperm cellularization [58]. The expression patterns for those DEGs during the seed development were illustrated in **Fig. 4C**.

In contrast, top GO terms enriched for the genes in cluster G were mainly related to photosynthesis (GO:0015979) and microtubule-based process (GO:0007017). Down-regulation of these gene sets mainly reflected the reduction in the photosynthetic process as seeds developed, as evident by the downregulation of the bulk of genes encoding varying processes of photosynthesis, such as Chlorophyll A-B binding family proteins, high chlorophyll fluorescent proteins, and light-harvesting complex

photosystem II (**Table S1**). Other family genes exhibiting downregulation expression patterns were also identified, such as GA20Xs, motor proteins, LACS9s, stay green-like, LEC1-like genes (**Table S1**). The decrease in the expressions of these genes with diverse roles associated with phytohormones (such as GA) [59], chlorophyll breakdown during senescence [60], seed developmental processes [54], which might be corresponding to multiple processes of gradual cessation of cell division or commencement of maturation as seeds developed and matured. Clusters C-F comprises relative fewer numbers of DEGs that display relatively high expression levels in two of four selected time points, and no GO term was significantly enriched.

## 3.5 DEGs have tissue-specificity in expression and exhibit tissue-specific functionality

To better understand the change or the DEGs, we examined the spatial expression profiles of all the DEGs in different developmental tissues using the Harada-Goldberg soybean RNA-seq dataset (<http://seedgenenetwork.net>). This dataset contains 10 tissues of Williams 82, the same accession as used in our study, including 6 reproductive tissues (floral buds (FB), five-stage developing seeds including whole-mount globular (GLOB), heart (HRT), cotyledon (COT), early-maturation (EM), dry seed (DS)) and four vegetative tissues (leaves (LF), roots (RT), stems (STM), and seedlings (SDL)). This independent research is complementary to our study, and an integration of our result with spatial analysis using multiple tissues may provide additional insight into the mechanism.

The DEGs exhibited clear tissue-specific (TS) patterns based on the expression similarity. This analysis grouped the DEGs into seven large (TS1-TS7) groups each containing over 400 genes and the rest three groups (not shown) comprising less than 30 genes (**Table S2**). The seven large groups were indicated in **Fig. 5A** and further investigated. Enrichment analysis for the seven TS groups reveals that none of the top five enriched GO terms were in common, indicating the tissue-specific roles. Interestingly, genes that were specifically expressed in developing seeds were divided into three groups, TS1 grouped genes that are highly expressed in three early-stage developing seeds, GLOB, HRT, and COT, TS2 in the EM-stage seeds, and TS3 in the later-stage seeds, DS. Approximately 2,647 (29.7%) of the DEGs were clustered to the TS group 1 (TS1) consisting of genes primarily expressed in GLOB-, HRT-, and COT-stage seeds and these genes showed continuous expression either up-regulation or downregulation. The majority (79.1%, 2,094 of 2,647) of the TS1 fell into the G group (down-regulation group) (Fig. 5A, 5C) exhibiting downregulation as seeds grew. Genes (585, 6.6% of total) in TS2 were predominantly expressed in the EM-stage seeds. 970 (10.9%) of DEGs were exclusively expressed in DS in TS7. This obvious spatial difference in expression among the different-stage seeds, suggesting the distinct roles of the two gene sets in seeds. Indeed, TS1 genes mainly involved in processes related to cell cycle and DNA replication, which might account for the early seed compartment morphogenesis and growth (Fig. 5B), and TS2 (EM) appeared to be more specialized in metabolic processes such as fatty acid biosynthesis and metabolism of carboxylic acid, and organic acid. In contrast, DS-specific genes were involved in the response to environmental stimuli, such as temperature, heat, alcohol, and plant hormones that regulate plant growth such as abscisic acid, and over half (67.7%) of this gene set were upregulated, (Fig. 5C). For example,

LEA14 (LATE EMBRYOGENESIS ABUNDANT 14) from TS7 involved in wounding and light stress as well as perhaps protection against desiccation [61, 62]. Thus, the DS-specific DEGs might be involved in preparing transcripts for dormancy and possibly seed germination once in a favorable condition. Recent studies indicated that expression for many of these seed-expressed genes were highly specified spatially in seeds, and the specificity represents the functional relevance. For example, a seed coat-specially expressed *SWEET39* involved in sugar delivery from maternal plant to developing embryo [63], endosperm-specific *TFL1* involve in endosperm cellularization by which affects seed size [58], and endosperm-exclusive *SWEET15* involved in seed development by regulating sugar delivery into developing embryo [12]. These studies suggested that many of the seed-specific genes identified here (**Table S2**) can be prioritized for detailed investigations to gain more comprehensive understanding, such as YABBA, GA3OX1, NF-YB6, NF-YA9, FAD8, KAS I/III, SWEETs, OLEOSINs, AATs (amino acid transporters), and uncharacterized LEC1-LIKE (Glyma.07G268100, Glyma.17G005600) that serves as a key regulator of fatty acid biosynthesis in *Arabidopsis* [64].

A total of 1357 (15.2%) DEGs were clustered in TS3 that were mainly expressed in leaves and were involved in many aspects of photosynthesis, and over 80% of them were downregulated as seeds matured (Fig. 5C). This result indicated that this gene set maintained the photosynthetic roles and was involved in seed developmental processes. It is likely that expression for many of these photosynthesis-related genes in seeds might be influenced by mosaic interactions of different transcription factors such as LEC1-ABI3-bZIP67-AREB3 [17, 54]. Genes that were specially expressed in RT and STM respectively shared many DEGs in common and were clustered in TS5, in agreement with the tissue functionality that are both involved in nutrient translocation [65], these genes mainly involved in carbohydrate and sugar metabolism and nutrients/ion transport to meet the large demand for nutrient uptake and transport (Fig. 5B). TS6 is the smallest group (400) that clustered the genes highly expressed in SDL and involved in processes such as oxidation-reduction. These results indicated that genes that have expression specificity in non-seed tissues such as LF, FB, and RT also represent its tissue-represented function in developing seeds (Fig. 5B). These genes remained the expression in relatively low abundance in developing seeds, suggesting their indispensable roles to maintain tissue-representative function for seed development. For example, AAP2 from TS6 (stem and leaf) affect levels of nitrogen and carbon assimilation and seed yield by altering xylem-phloem transfer of amino acid transfer [66]; FRD3 (MATE efflux family protein) from TS6 functions in root xylem for efficient iron uptake from the xylem into leaf cells [67]. Therefore, identification of these genes with non-seed tissue-specificities in developing seed was reflective of necessity of basic functionalities of the non-tissue genes in maintaining needed processes to support seed development. Further investigation of them may provide deep insight into their supportive roles during seed development.

### **3.6 Co-expression network inferred seed regulatory network**

We next extracted the co-expression connection of all differentially expressed genes from a global SoyNet database (<http://www.inetbio.org/soynet/>) that was constructed with transcriptomes from different seed compartments at different developmental stages along with other tissues [30, 54]. Only a small fraction

of the DEGs can be reconstructed into modules. In total, 134 modules were identified with the node (genes) ranging from 2 to 151. After color assignment based on the foregoing TS groups, 11 modules consisting of relatively tissue specific patterns were identified (Fig. 5D). It is interesting that many genes from TS1 (GLOB, HRT, COT) and TS3 (LF), rather than other TS groups, were respectively clustered together as individual TS-specific modules. Two modules specific for TS1 were identified. TCX2 (TESMIN-LIKE CXC2) controls stem cell division by regulating stem cell-type-specific networks [68]. It was slightly deviated from the main module, suggesting its role in orchestrating the connected genes of this module for a coordinated division of different cell types for seed development, in agreement with the function of GLOB/HRT/COT where ontogenesis of different seed compartments occur during this stage (<http://seedgenenetwork.net/>). Similarly, TDM1 plays important role in meiosis termination [69] and it appears to be extensively associated with genes in the TS1-dominated modules. We identified that the uncharacterized TKL2 (TRANSKETOLASE 2) might be a key connector bridging two modules comprising LF and GLOB/HRT/COT genes respectively. Likely, the link between RBCS3B (RUBISCO SMALL SUBUNIT 3B) and BLH1 (EMBRYO SAC DEVELOPMENT ARREST 29) connected the two individual modules. These results suggest a functional relevance of these modules in seed development and these genes that were under-explored in soybean deserved more attention.

## 4. Conclusions

In conclusion, we profiled the transcriptomes from developing seeds at the four representative stages and provided a global transcriptomic view on the developmental process of storage reserves accumulation and seed development. The results highlighted expression of high abundance of genes (storage genes, transporters, transcription factors) associated with soybean development and unveiled that seed formation is involved in a variety of tissues-specific genes with tissue-associated roles to facilitate the energy source transport, cell division, enlargement as well as synthesis of highly abundant storage proteins and oil. Our study provides a global view of how a large quantity of genes is coordinated during seed development could be a framework for further investigation. Certainly, how these region-specific and period-specific genes/processes play in seed formation remains to be determined.

## Declarations

### ***Ethics approval and consent to participate***

*Not applicable.*

### ***Consent for publication***

*Not applicable.*

### ***Competing interests***

The authors declare that they have no competing interests

## Funding

This work was supported by the National Natural Science Foundation of China (32072088), the Ministry of Science and Technology of China (2016YFD0100500), the key scientific and technological project of Henan Province (192102110023, 202102110005), the Key Scientific Research Projects of Higher Education Institutions in Henan Province (20A210017), and the Henan agricultural university science and technology innovation fund (KJCX2019C02). The study was carried out with the High-Performance Computing (HPC) cluster of the University Research Computing at the University of North Carolina at Charlotte.

## Author contributions

Conceived and designed the experiments: DZ and HZ. Performed the experiments: HZ, ZH, YY, and HL. Analyzed the data: HZ and HL. Contributed reagents/materials/analysis tools: ZL, YQA, and BS. Wrote the paper: HZ, DZ. All authors read and approved the final manuscript.

## Acknowledgements

*Not applicable.*

## Data Availability Statement

The RNA-seq data we mentioned in our manuscript has been uploaded to NCBI (<https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA551573>), (BioProject:PRJNA551573), (SRA Study: SRP212336).

## References

1. Anderson EJ, Ali MD, Beavis WD, Chen P, Clemente TE, Diers BW, Graef GL, Grassini P, Hyten DL, McHale LK *et al*: **Soybean [*Glycine max* (L.) Merr.] Breeding: History, improvement, production and future opportunities**. In: *Advances in Plant Breeding Strategies: Legumes*. Edited by Al-Khayri J, Jain S, Johnson D: Springer, Cham; 2019: 431-516.
2. USDA-NASS: **United State Department of Agriculture, National Agricultural Statistics Service**. [https://www.nass.usda.gov/Charts\\_and\\_Maps/Field\\_Crops/index.php](https://www.nass.usda.gov/Charts_and_Maps/Field_Crops/index.php). 2019.
3. Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM, Toulmin C: **Food security: The challenge of feeding 9 billion people**. *Science* 2010, **327**(5967):812-818.
4. Tilman D, Balzer C, Hill J, Befort BL: **Global food demand and the sustainable intensification of agriculture**. *P Natl Acad Sci USA* 2011, **108**(50):20260-20264.
5. Le BH, Wagmaister JA, Kawashima T, Bui AQ, Harada JJ, Goldberg RB: **Using genomics to study legume seed development**. *Plant Physiol* 2007, **144**(2):562-574.

6. Weber H, Borisjuk L, Wobus U: **Molecular physiology of legume seed development.** *Annu Rev Plant Biol* 2005, **56**:253-279.
7. Lin JY, Le BH, Chen M, Henry KF, Hur J, Hsieh TF, Chen PY, Pelletier JM, Pellegrini M, Fischer RL *et al*: **Similarity between soybean and *Arabidopsis* seed methylomes and loss of non-CG methylation does not affect seed development.** *Proceedings of the National Academy of Sciences of the United States of America* 2017, **114**(45):E9730-E9739.
8. Song QX, Li QT, Liu YF, Zhang FX, Ma B, Zhang WK, Man WQ, Du WG, Wang GD, Chen SY *et al*: **Soybean GmbZIP123 gene enhances lipid content in the seeds of transgenic *Arabidopsis* plants.** *J Exp Bot* 2013, **64**(14):4329-4341.
9. Lu X, Li QT, Xiong Q, Li W, Bi YD, Lai YC, Liu XL, Man WQ, Zhang WK, Ma B *et al*: **The transcriptomic signature of developing soybean seeds reveals the genetic basis of seed trait adaptation during domestication.** *Plant J* 2016, **86**(6):530-544.
10. Danzer J, Mellott E, Bui AQ, Le BH, Martin P, Hashimoto M, Perez-Lesher J, Chen M, Pelletier JM, Somers DA *et al*: **Down-regulating the expression of 53 soybean transcription factor genes uncovers a role for SPEECHLESS in initiating stomatal cell lineages during embryo development.** *Plant Physiol* 2015, **168**(3):1025-1035.
11. Wang S, Liu S, Wang J, Yokosho K, Zhou B, Yu Y, Liu Z, Frommer WB, Ma J, Chen L *et al*: **Simultaneous changes in seed size, oil content, and protein content driven by selection of SWEET homologues during soybean domestication.** *National Science Review* 2020:nwaa110.
12. Wang S, Yokosho K, Guo R, Whelan J, Ruan YL, Ma JF, Shou H: **The soybean sugar transporter GmSWEET15 mediates sucrose export from endosperm to early embryo.** *Plant Physiol* 2019:<https://doi.org/10.1104/pp.1119.00641>.
13. Jo L, Pelletier JM, Harada JJ: **Central role of the LEAFY COTYLEDON1 transcription factor in seed development.** *Journal of Integrative Plant Biology* 2019, **61**(5):564-580.
14. Guo W, Chen LM, Chen HF, Yang HL, You QB, Bao AL, Chen SL, Hao QN, Huang Y, Qiu DZ *et al*: **Overexpression of GmWRI1b in soybean stably improves plant architecture and associated yield parameters, and increases total seed oil production under field conditions.** *Plant Biotechnol J* 2020, **18**(8):1639-1641.
15. Manan S, Ahmad MZ, Zhang GY, Chen BB, Haq BU, Yang JH, Zhao J: **Soybean LEC2 regulates subsets of genes involved in controlling the biosynthesis and catabolism of seed storage substances and seed development.** *Frontiers in Plant Science* 2017, **8**:1604.
16. Zhang D, Zhao M, Li S, Sun L, Wang W, Cai C, Dierking EC, Ma J: **Plasticity and innovation of regulatory mechanisms underlying seed oil content mediated by duplicated genes in the palaeopolyploid soybean.** *Plant J* 2017, **90**(6):1120-1133.
17. Jo L, Pelletier JM, Hsu SW, Baden R, Goldberg RB, Harada JJ: **Combinatorial interactions of the LEC1 transcription factor specify diverse developmental programs during soybean seed development.** *Proceedings of the National Academy of Sciences of the United States of America* 2020, **117**(2):1223-1232.

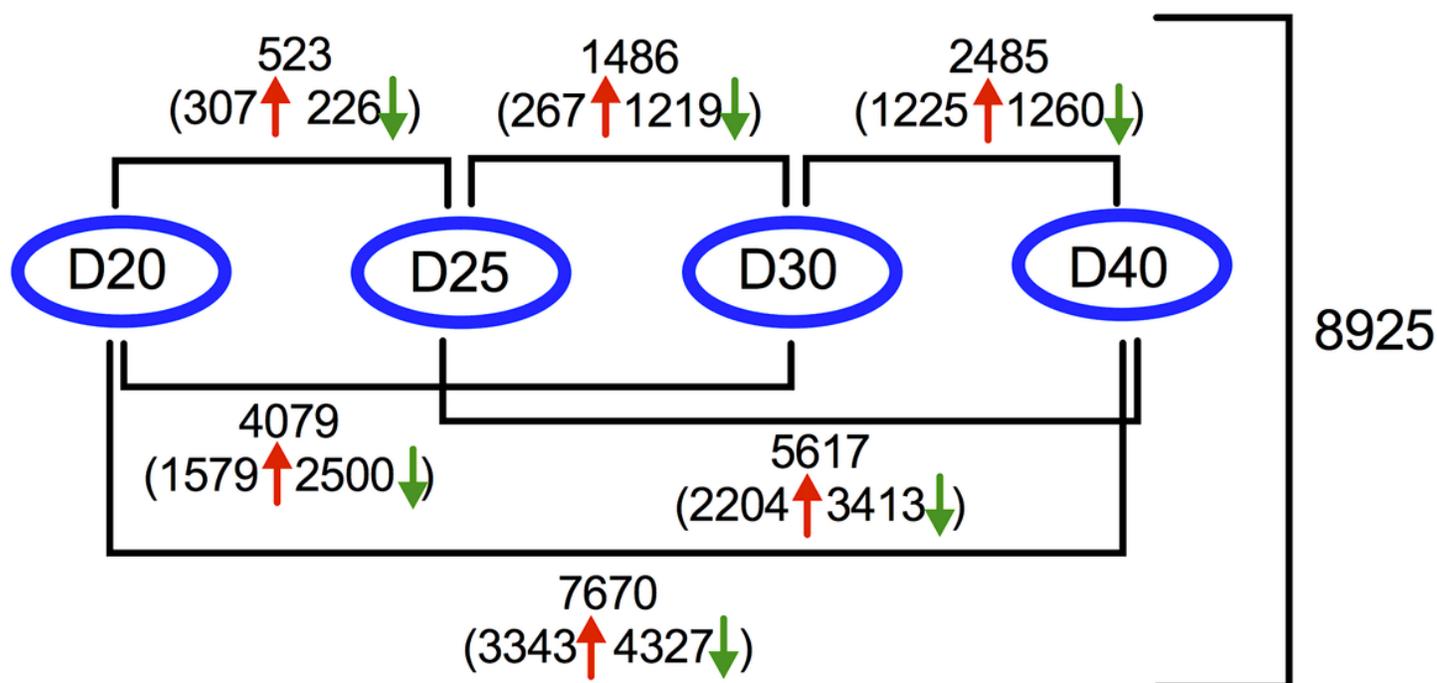
18. Le BH, Cheng C, Bui AQ, Wagmaister JA, Henry KF, Pelletier J, Kwong L, Belmonte M, Kirkbride R, Horvath S *et al*: **Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors.** *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**(18):8063-8070.
19. Garg R, Singh VK, Rajkumar MS, Kumar V, Jain M: **Global transcriptome and coexpression network analyses reveal cultivar-specific molecular signatures associated with seed development and seed size/weight determination in chickpea.** *Plant J* 2017, **91**(6):1088-1107.
20. Du J, Wang S, He C, Zhou B, Ruan YL, Shou H: **Identification of regulatory networks and hub genes controlling soybean seed set and size using RNA sequencing analysis.** *J Exp Bot* 2017, **68**(8):1955-1972.
21. Qi ZM, Zhang ZG, Wang ZY, Yu JY, Qin HT, Mao XR, Jiang HW, Xin DW, Yin ZG, Zhu RS *et al*: **Meta-analysis and transcriptome profiling reveal hub genes for soybean seed storage composition during seed development.** *Plant Cell and Environment* 2018, **41**(9):2109-2127.
22. Yang SN, Miao L, He JB, Zhang K, Li Y, Gai JY: **Dynamic Transcriptome Changes Related to Oil Accumulation in Developing Soybean Seeds.** *Int J Mol Sci* 2019, **20**(9):18.
23. Zhang D, Zhang H, Chu S, Li H, Chi Y, Triebwasser-Freese D, Lv H, Yu D: **Integrating QTL mapping and transcriptomics identifies candidate genes underlying QTLs associated with soybean tolerance to low-phosphorus stress.** *Plant Mol Biol* 2017, **93**(1-2):137-150.
24. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data.** *Bioinformatics* 2014, **30**(15):2114-2120.
25. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J *et al*: **Genome sequence of the palaeopolyploid soybean.** *Nature* 2010, **463**(7278):178-183.
26. Liao Y, Smyth GK, Shi W: **featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.** *Bioinformatics* 2014, **30**(7):923-930.
27. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.** *Bioinformatics* 2010, **26**(1):139-140.
28. Zhang HY, Kjemtrup-Lovelace S, Li CB, Luo Y, Chen LP, Song BH: **Comparative RNA-seq analysis uncovers a complex regulatory network for soybean cyst nematode resistance in wild soybean (*Glycine soja*).** *Sci Rep-Uk* 2017, **7**:9699.
29. Zhang D, Zhang H, Hu Z, Chu S, Yu K, Lv L, Yang Y, Zhang X, Chen X, Kan G *et al*: **Artificial selection on *GmOLEO1* contributes to the increase in seed oil during soybean domestication.** *PLoS Genet* 2019, **15**(7):e1008267.
30. Kim E, Hwang S, Lee I: **SoyNet: a database of co-functional networks for soybean *Glycine max*.** *Nucleic Acids Res* 2017, **45**(D1):D1082-D1089.
31. Bader GD, Hogue CW: **An automated method for finding molecular complexes in large protein interaction networks.** *BMC bioinformatics* 2003, **4**:2.
32. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B *et al*: **Integration of biological networks and gene expression data using**

- Cytoscape**. *Nature protocols* 2007, **2**(10):2366-2382.
33. Richie SW, Hanway JJ, Thomas HE, Benson GO: **How a soybean plant develops** Special Report **NO.53**. Ames, IA: Iowa State University of Science and Technology Cooperative Extension Service. 1996.
34. Assefa Y, Purcell LC, Salmeron M, Naeve S, Casteel SN, Kovacs P, Archontoulis S, Licht M, Below F, Kandel H *et al*: **Assessing variation in US soybean seed composition (protein and oil)**. *Frontiers in Plant Science* 2019, **10**:298.
35. Hill JE, Breidenbach RW: **Proteins of soybean seeds: II. Accumulation of the major protein components during seed development and maturation**. *Plant Physiol* 1974, **53**(5):741-751.
36. Considine MJ, Considine JA: **On the language and physiology of dormancy and quiescence in plants**. *J Exp Bot* 2016, **67**(11):3189-3203.
37. Feng N, Song G, Guan J, Chen K, Jia M, Huang D, Wu J, Zhang L, Kong X, Geng S *et al*: **Transcriptome profiling of wheat inflorescence development from spikelet initiation to floral patterning identified stage-specific regulatory genes**. *Plant Physiol* 2017, **174**(3):1779-1794.
38. Huang AHC: **Plant lipid droplets and their associated proteins: Potential for rapid advances**. *Plant Physiol* 2018, **176**(3):1894-1918.
39. Ohlrogge J, Browse J: **Lipid biosynthesis**. *Plant Cell* 1995, **7**(7):957-970.
40. Du C, Chen Y, Wang K, Yang Z, Zhao C, Jia Q, Taylor DC, Zhang M: **Strong co-suppression impedes an increase in polyunsaturated fatty acids in seeds overexpressing *FAD2***. *J Exp Bot* 2019, **70**(3):985-994.
41. Schmidt MA, Herman EM: **Suppression of soybean oleosin produces micro-oil bodies that aggregate into oil body/ER complexes**. *Molecular Plant* 2008, **1**(6):910-924.
42. Sarmiento C, Ross JHE, Murphy DJ: **Expression and subcellular targeting of a soybean oleosin in transgenic rapeseed. Implications for the mechanism of oil-body formation in seeds**. *Plant J* 1997, **11**(4):783-796.
43. Gillman JD, Kim WS, Krishnan HB: **Identification of a new soybean kunitz trypsin inhibitor mutation and its effect on bowman-birk protease inhibitor content in soybean seed**. *J Agric Food Chem* 2015, **63**(5):1352-1359.
44. Chen LQ, Qu XQ, Hou BH, Sosso D, Osorio S, Fernie AR, Frommer WB: **Sucrose efflux mediated by SWEET proteins as a key step for phloem transport**. *Science* 2012, **335**(6065):207-211.
45. Sosso D, Luo DP, Li QB, Sasse J, Yang JL, Gendrot G, Suzuki M, Koch KE, McCarty DR, Chourey PS *et al*: **Seed filling in domesticated maize and rice depends on SWEET-mediated hexose transport**. *Nat Genet* 2015, **47**(12):1489-1493.
46. Yang JL, Luo DP, Yang B, Frommer WB, Eom JS: **SWEET11 and 15 as key players in seed filling in rice**. *New Phytol* 2018, **218**(2):604-615.
47. Mechin V, Thevenot C, Le Guilloux M, Prioul JL, Damerval C: **Developmental analysis of maize endosperm proteome suggests a pivotal role for pyruvate orthophosphate dikinase**. *Plant Physiol*

- 2007, **143**(3):1203-1219.
48. Vensel WH, Tanaka CK, Cai N, Wong JH, Buchanan BB, Hurkman WJ: **Developmental changes in the metabolic protein profiles of wheat endosperm.** *Proteomics* 2005, **5**(6):1594-1611.
  49. Porta H, Rocha-Sosa M: **Plant lipoxygenases. Physiological and molecular features.** *Plant Physiol* 2002, **130**(1):15-21.
  50. Fischer AM, Dubbs WE, Baker RA, Fuller MA, Stephenson LC, Grimes HD: **Protein dynamics, activity and cellular localization of soybean lipoxygenases indicate distinct functional roles for individual isoforms.** *Plant J* 1999, **19**(5):543-554.
  51. King JM, Svendsen LK, Fehr WR, Narvel JM, White PJ: **Oxidative and flavor stability of oil from lipoxygenase-free soybeans.** *Journal of the American Oil Chemists Society* 1998, **75**(9):1121-1126.
  52. Kroj T, Savino G, Valon C, Giraudat J, Parcy F: **Regulation of storage protein gene expression in *Arabidopsis*.** *Development* 2003, **130**(24):6065-6073.
  53. Lara P, Onate-Sanchez L, Abraham Z, Ferrandiz C, Diaz I, Carbonero P, Vicente-Carbajosa J: **Synergistic activation of seed storage protein gene expression in *Arabidopsis* by ABI3 and two bZIPs related to OPAQUE2.** *J Biol Chem* 2003, **278**(23):21003-21011.
  54. Pelletier JM, Kwong RW, Park S, Le BH, Baden R, Cagliaria A, Hashimoto M, Munoz MD, Fischer RL, Goldberg RB *et al.*: **LEC1 sequentially regulates the transcription of genes involved in diverse developmental processes during seed development.** *Proceedings of the National Academy of Sciences of the United States of America* 2017, **114**(32):E6710-E6719.
  55. Angelovici R, Galili G, Fernie AR, Fait A: **Seed desiccation: a bridge between maturation and germination.** *Trends Plant Sci* 2010, **15**(4):211-218.
  56. Nambara E, Hayama R, Tsuchiya Y, Nishimura M, Kawaide H, Kamiya Y, Naito S: **The role of ABI3 and FUS3 loci in *Arabidopsis thaliana* on phase transition from late embryo development to germination.** *Dev Biol* 2000, **220**(2):412-423.
  57. Baroja-Fernandez E, Munoz FJ, Li J, Bahaji A, Almagro G, Montero M, Etxeberria E, Hidalgo M, Sesma MT, Pozueta-Romero J: **Sucrose synthase activity in the sus1/sus2/sus3/sus4 *Arabidopsis* mutant is sufficient to support normal cellulose and starch production.** *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**(1):321-326.
  58. Zhang B, Li CX, Li Y, Yu H: **Mobile TERMINAL FLOWER1 determines seed size in *Arabidopsis*.** *Nature Plants* 2020:1146–1157.
  59. Rieu I, Eriksson S, Powers SJ, Gong F, Griffiths J, Woolley L, Benlloch R, Nilsson O, Thomas SG, Hedden P *et al.*: **Genetic analysis reveals that C-19-GA 2-Oxidation is a major gibberellin inactivation pathway in *Arabidopsis*.** *Plant Cell* 2008, **20**(9):2420-2436.
  60. Park SY, Yu JW, Park JS, Li J, Yoo SC, Lee NY, Lee SK, Jeong SW, Seo HS, Koh HJ *et al.*: **The senescence-induced staygreen protein regulates chlorophyll degradation.** *Plant Cell* 2007, **19**(5):1649-1664.
  61. Bray EA: **Genes commonly regulated by water-deficit stress in *Arabidopsis thaliana*.** *J Exp Bot* 2004, **55**(407):2331-2341.

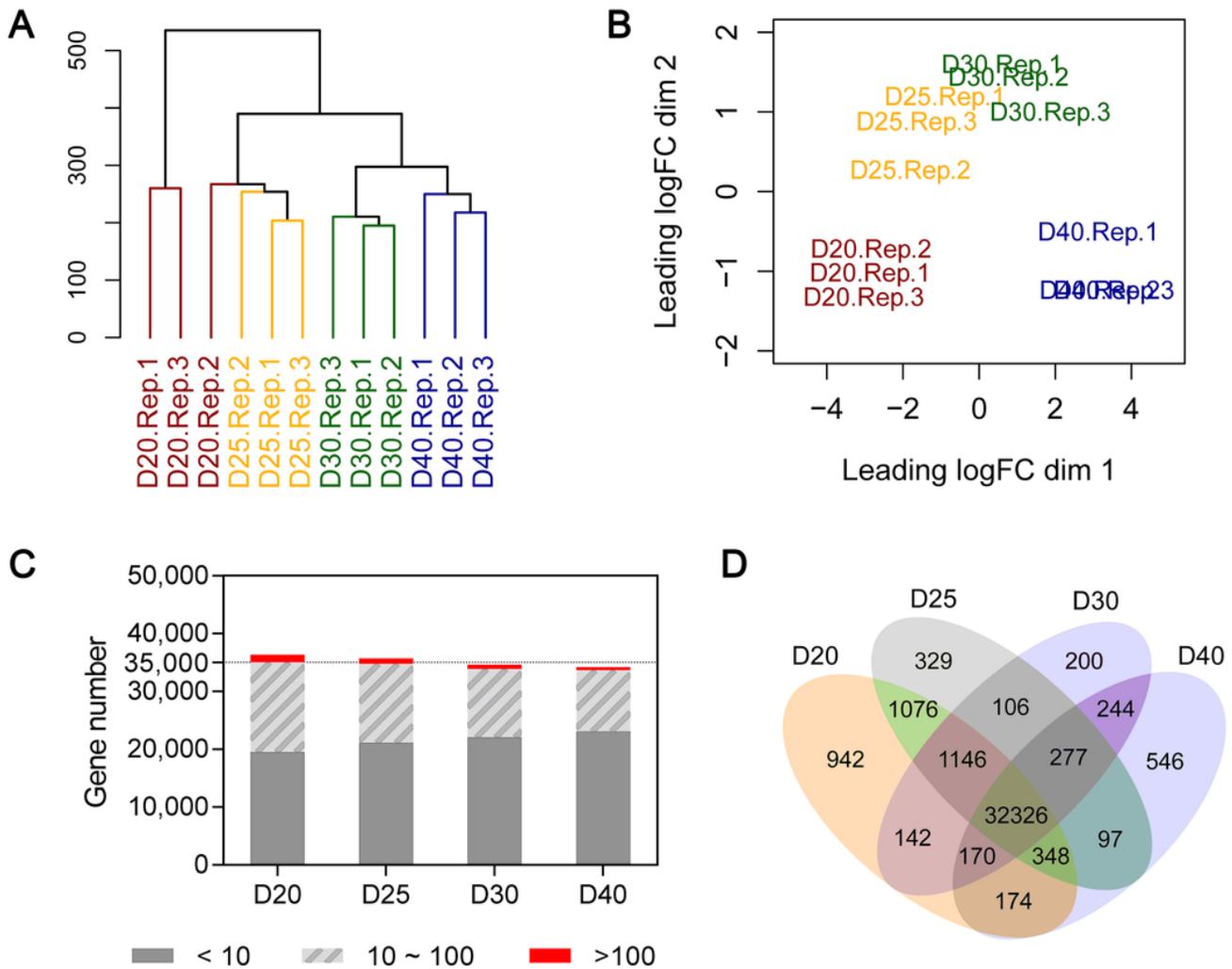
62. Dunaeva M, Adamska W: **Identification of genes expressed in response to light stress in leaves of *Arabidopsis thaliana* using RNA differential display.** *Eur J Biochem* 2001, **268**(21):5521-5529.
63. Miao L, Yang SN, Zhang K, He JB, Wu CH, Ren YH, Gai JY, Li Y: **Natural variation and selection in GmSWEET39 affect soybean seed oil content.** *New Phytol* 2020, **225**(4):1651-1666.
64. Tan HL, Yang XH, Zhang FX, Zheng X, Qu CM, Mu JY, Fu FY, Li JA, Guan RZ, Zhang HS *et al*: **Enhanced seed oil production in canola by conditional expression of *Brassica napus* LEAFY COTYLEDON1 and LEC1-LIKE in developing seeds.** *Plant Physiol* 2011, **156**(3):1577-1588.
65. Atkins CA, Smith PMC: **Translocation in legumes: Assimilates, nutrients, and signaling molecules.** *Plant Physiol* 2007, **144**(2):550-561.
66. Zhang L, Tan Q, Lee R, Trethewy A, Lee YH, Tegeder M: **Altered xylem-phloem transfer of amino acids affects metabolism and leads to increased seed yield and oil content in Arabidopsis.** *Plant Cell* 2010, **22**(11):3603-3620.
67. Rogers EE, Guerinot ML: **FRD3, a member of the multidrug and toxin efflux family, controls iron deficiency responses in Arabidopsis.** *Plant Cell* 2002, **14**(8):1787-1799.
68. Clark NM, Buckner E, Fisher AP, Nelson EC, Nguyen TT, Simmons AR, Balaguer MAD, Butler-Smith T, Sheldon PJ, Bergmann DC *et al*: **Stem-cell-ubiquitous genes spatiotemporally coordinate division through regulation of stem-cell-specific gene networks.** *Nature Communications* 2019, **10**:5574.
69. Cifuentes M, Jolivet S, Cromer L, Harashima H, Bulankova P, Renne C, Crismani W, Nomura Y, Nakagami H, Sugimoto K *et al*: **TDM1 regulation determines the number of meiotic divisions.** *PLoS Genet* 2016, **12**(2):e1005856.

## Figures



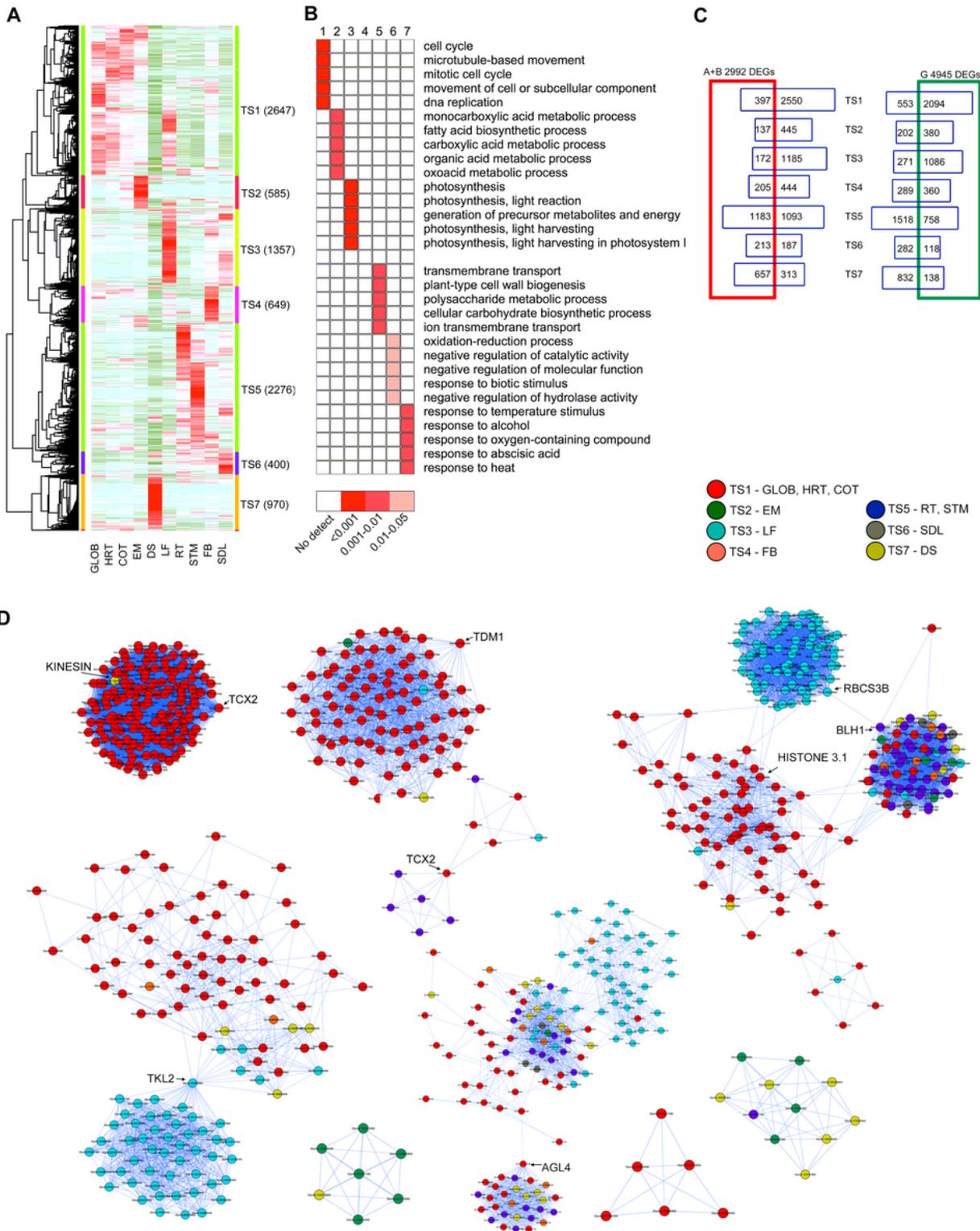
**Figure 1**

Number of differentially expressed genes estimated by pair-wise comparison of the four stages.



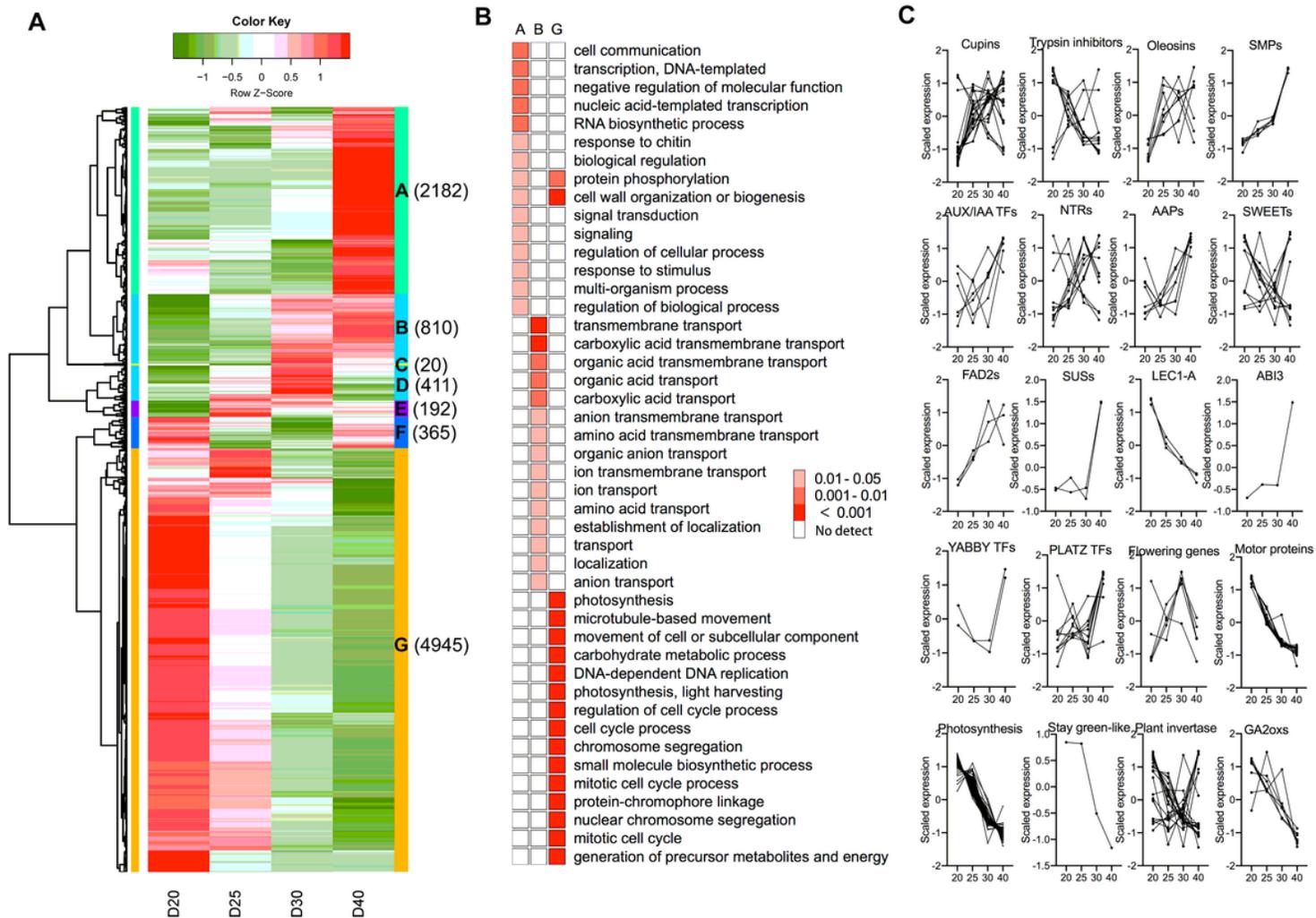
**Figure 1**

Gene expression profiles during seed filling. (A) Cluster dendrogram of gene expression profiles among biological replicates and four stages. (B) Multi-dimensional scaling plot of the collected samples across the four stages. (C) The number of genes expressed during each of the four stages. (D) A Venn diagram illustrating the number of expressed genes shared between or specific to the four different stages.



**Figure 1**

Tissue-specific expression patterns of all DEGs. (A) Clustering of the DEGs in different tissues globular- (GLOB), heart- (HRT), cotyledon- (COT), early maturation- (EM) stage seeds, leaves (LF), roots (RT), stem (STM), floral bud (FB) and seedlings (SDL). (B) Top five GO terms (biological process) enriched for the clustered genes. (C) Venn diagrams for TS clustered genes with clustered genes identified in A and G in Figure 4. (D) Co-expression modules re-constructed by DEGs and node color were assigned by tissues.



**Figure 1**

Temporal expression patterns of all DEGs. (A) Clustering of gene expression patterns across four time points during seed development. (B) GO terms enriched for the large clusters A, B, and G. (C) Expression patterns for selected genes.

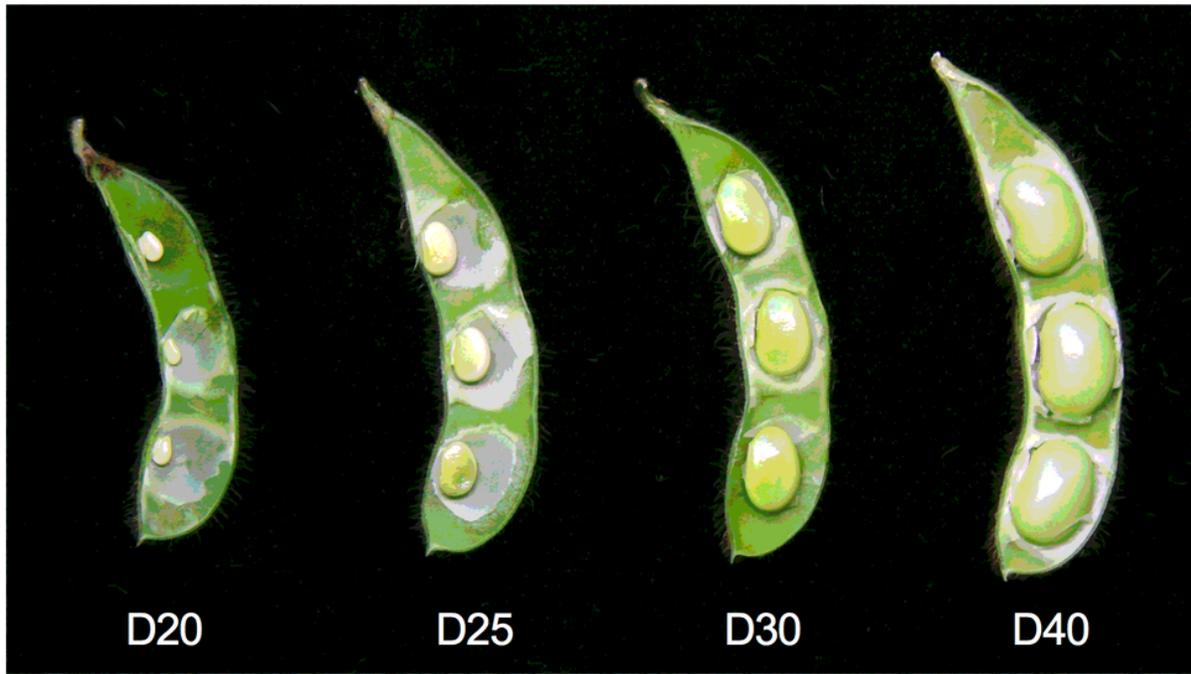
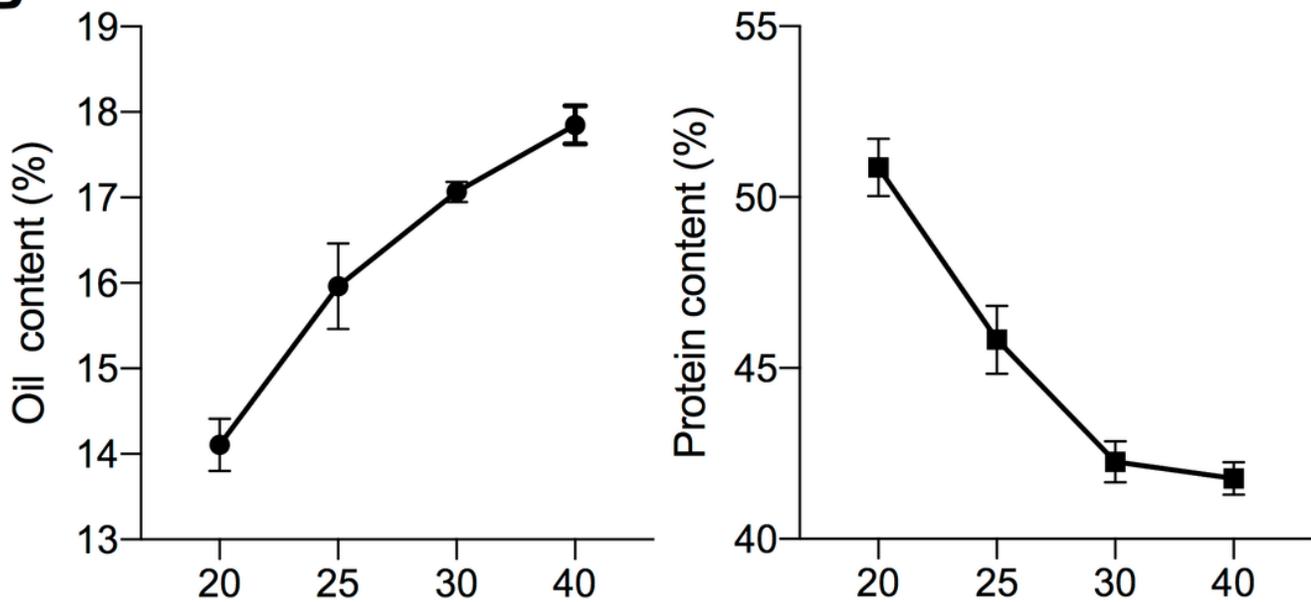
**A****B**

Figure 1

Morphology of developing seeds at 20, 25, 30, and 40 days after flowering (DAF) (A) and oil and protein changes during the selected stages (B).

## Supplementary Files

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

- [SupplementalTablesWm82Seeds.xlsx](#)