Seed Science Research

cambridge.org/ssr

Full-length transcriptome and co-expression network analysis reveal molecular mechanisms of seed development in *Elymus sibiricus*

Yuying Zheng†, Xiaoshan Lin†, Wengang Xie and Wenxian Liu

State Key Laboratory of Herbage Improvement and Grassland Agro-Ecosystems, Lanzhou University, Key Laboratory of Grassland Livestock Industry Innovation, Ministry of Agriculture and Rural Affairs, Engineering Research Centre of Grassland Industry, Ministry of Education, College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou 730000, China

Research Paper

†Contributed equally to this work.

Cite this article: Zheng Y, Lin X, Xie W, Liu W (2024). Full-length transcriptome and co-expression network analysis reveal molecular mechanisms of seed development in *Elymus sibiricus*. *Seed Science Research* **34**, 17–32. https://doi.org/10.1017/S0960258524000084

Received: 14 November 2023
Revised: 25 January 2024
Accepted: 2 March 2024
First published online: 16 May 2024

Keywords:
E. sibiricus; full-length transcriptome; seed development; starch; sucrose

Corresponding authors:
Wengang Xie; Email: xiewg@lzu.edu.cn;
Wenxian Liu; Email: liuwx@lzu.edu.cn

Abstract

Grass seeds play a critical and fundamental role in grass breeding and production. *Elymus sibiricus* L. is a widespread Poaceae forage grass in northern Eurasia which is used for ecological restoration and forage production. Sucrose is the main source of substrate and energy required for starch synthesis in the seeds, so the hydrolysis of sucrose determines and influences starch synthesis and filling in the seeds, especially Poaceae. However, the process behind carbohydrate metabolism during *E. sibiricus* seed development remains unclear. This study addresses a significant gap in our understanding of the carbohydrate metabolism during seed development in *E. sibiricus* by employing full-length transcriptome sequencing across five developmental stages for the first time. Full-length transcriptome sequencing was performed on *E. sibiricus* seeds at five developmental stages (S5, S9, S15, S20, S25) to get better molecular insights. We identified 13,205 differentially expressed genes, with 7,471 up-regulated and 5,734 down-regulated. Through KEGG enrichment analysis, genes were enriched in 'starch and sucrose metabolism', 'photosynthetic-related' and 'hormone signal transduction' pathways. Gene ontology enrichment analysis showed that genes were enriched in the 'beta-amylase activity' term of molecular functions. In addition, top 21 transcription factor families were identified as involved in seed development. The homologous genes of *ABSCISIC ACID-INSENSITIVE 3* (*ABI3*), *NUCLEAR FACTOR-YB1* (*NF-YB1*), *STARCH SYNTHASE I* (*SSI*) were identified as candidate genes of seed development in *E. sibiricus*. Combined with physiological index, transcriptome analyses, weighted gene co-expression network analysis and real-time quantitative PCR, the mechanism of starch and sucrose content of seed development was revealed and ten hub genes were identified. Overall, this study provides the molecular bases to understand seed development and starch and sucrose metabolism at the different seed developmental stages in *E. sibiricus*.

Introduction

Seeds are crucial to protect and nourish the developing embryos that represent the next sporophytic generation in flowering plants (Pradhan et al., 2014). Seed quality is usually measured by germination ability or physicochemical attributes of seeds which is gaining greater importance in agriculture field due to its critical and fundamental role in plant breeding and production. High-quality seeds have been reported to initiate successful plant growth which results in abundant yield/production (Feng et al., 2019). High quality and high yield have always been the main goals of seed production, and only by deeply understanding the dynamic changes of seed development can this goal be better achieved. Generally, seed development consists of three stages: embryogenesis, seed formation, maturation and desiccation (Garg et al., 2017). A series of complex and dynamic developmental, biochemical and metabolic processes are involved in seed development including cell division and differentiation, carbohydrate, protein, cell wall, lipid, amino acid, hormone and secondary metabolite biosynthesis (Baud et al., 2002). Starch is the main energy storage substance in seed development, and sucrose is one of the main sources of energy (Aguirre et al., 2018). Sucrose hydrolysis affects starch synthesis and thus seed filling and size (Weber et al., 2005).

Understanding the regulatory mechanisms of seed formation is essential to identify the molecular basis of seed development. Transcript profiling of developing seeds has become a popular method to identify key candidate genes and their associated regulatory pathways that can be used as breeding tools to improve seed quality traits. Previously, storage compound synthesis and expression pattern of responsible genes have been reported in *Arabidopsis thaliana* during seed development (Baud et al., 2002; Peng and Weselake, 2011). In *A. thaliana* ecotype Wassilewskija, starch accumulates during the morphogenesis stage of seed development, with a subsequent decrease coinciding with sucrose accumulation during late

maturation (Baud et al., 2002). Transcription factors (TFs), such as *LEAFY COTYLEDON 1* (*LEC1*), *LEC2*, *ABSCISIC ACID-INSENSITIVE 3* (*ABI3*), *FUSCA 3* (*FUS3*) and others from the CCAAT, bZIP (basic leucine zipper) and bHLH (basic helix-loop-helix) families, form a complex regulatory network essential for *Arabidopsis* seed development (Le et al., 2010; Verma et al., 2022). Similarly, these seed-specific TFs were also reported in *Oryza sativa* which are involved in hormone response, cellular organization processes and metabolism regulation (Xue et al., 2012). In *Triticum aestivum*, the overexpression of *TaNAC100* increased starch content, seed size and 1000-grain weight. In the meanwhile, key starch synthesis-related genes *TaGBSSI* (*GRANULE BOUND STARCH SYNTHASE I*) and *TaSUS2* (*SUCROSE SYNTHASE 2*) respond to *TaNAC100* expression (Li et al., 2021). Barley is also an important source of carbohydrates in the diet. The known genes *NUCLEAR FACTOR-YB* (*NF-YBs*) are also found in *Hordeum vulgare*. *AGPase* (*Hvulgare_GLEAN_10033640* and *Hvulgare_GLEAN_10056301*), as well as *SBE2b* (*Hvulgare_GLEAN_10018352*) may specifically participate in starch biosynthesis during seed development (Tang et al., 2017). Studies on the molecular mechanism of seed development have also been carried out in a few forage grasses. In a study to alter the nutritional value of legume forage seeds, it was found that enzymes involved in methionine biosynthesis were significantly compartmentalized between seed tissues in *Medicago truncatula* (Gallardo et al., 2007). By performing transcriptome sequencing on *Leymus chinensis* seeds at 14, 28 and 42 days after pollination, a total of 18,927 differentially expressed genes (DEGs) and TFs *NAC48*, *WRKY80* and *C3H14* were involved in *L. chinensis* seed development, and thereby appeared to influence germination rate (Li et al., 2019). The molecular regulatory mechanisms and accumulation patterns of oils and fatty acids of *Artemisia sphaerocephala* seeds at seven different development stages after flowering were revealed by RNA-seq, and FUS3 and bHLH family TFs played a vital role in this process (Nan et al., 2021). There are few studies on genes involved in starch and sucrose metabolism during forage seed development.

Elymus sibiricus (Siberian wild rye) is an economically significant, perennial, cold-season, self-pollinating and allotetraploid forage grass, indigenous to northern Asia and integral to Qinghai-Tibetan Plateau's agriculture due to its high protein content and robust adaptability (Xie et al., 2015; Zhao et al., 2017). In this context, *E. sibiricus* can be widely employed in establishing sown grasslands to develop stock raising and participate in ecological restoration (Yan et al., 2007). Despite its agricultural value, *E. sibiricus* faces challenges in seed shattering and variety development, hindering its cultivation and application. Moreover, there is a lack of comprehensive molecular studies connecting transcriptomic profiles to seed developmental stages in *E. sibiricus* (Lei et al., 2014; Zhao et al., 2019). Previous studies only focused on the morphological changes, physiological mechanisms and biochemical processes during seed development in *E. sibiricus* (Zhao et al., 2017; Lei et al., 2020). Meanwhile, the transcript profiling was mainly reported in relation to seed shattering and flowering time (Xie et al., 2017; Zhang et al., 2019; Zheng et al., 2022). There are no studies showing a connecting transcript gene expression profiles to seed developmental stages in *E. sibiricus*. To address this, we have utilized an integrated sequencing approach that combines next-generation sequencing (NGS) with advanced long-read sequencing technologies to construct a comprehensive transcriptome for *E. sibiricus*. In absence of a published *E. sibiricus* genome annotation, our study performs full-length transcriptome sequencing across different seed developmental stages. By constructing a weighted gene co-expression network analysis (WGCNA), we aim to uncover the genes and molecular mechanisms that regulate seed development in *E. sibiricus*, providing novel insights and a valuable resource for future breeding efforts.

Materials and methods

Sample collection and physiological indices measurement

The wild *E. sibiricus* germplasm LQ06 was used for this study, which was collected from Luqu, Gansu province of China (latitude 34°05′39″ N, longitude 102°37′55″ E, elevation 3,380 m). The seeds were germinated on two layers of filter paper and placed in a controlled environment chamber set at 25°C and 8 h in the dark/16 h in the light. Thirty healthy seedlings were grown in a greenhouse and were transplanted to the experimental field at the Yuzhong campus of Lanzhou University, Gansu, China (latitude 35°34′ N, longitude 103°34′ E, elevation 1,720 m) when two tillering buds have developed. Spikelets that flowered simultaneously were tagged at full bloom, and the flowering date was recorded. Seeds from spikelets with comparable developmental stages were collected at 5, 9, 15, 20 and 25 days post-anthesis (DPA), hereinafter referred to as S5, S9, S15, S20 and S25, respectively. Each sample was collected in triplicate, flash-frozen in liquid nitrogen and stored at −80°C for further analysis. Immediately after harvest, samples were analysed for starch and soluble sugar content using Suzhou Comin Biochemical Test Kits according to the manufacturer's protocol.

RNA isolation, PacBio and Illumina library construction and sequencing

Total RNA was isolated using TRIzol® Reagent (Invitrogen® Thermo Fisher Scientific, Waltham, MA, USA). The purity and integrity of RNA were assessed by using a Nanodrop and 2100 Bioanalyzer (Agilent Technologies®, Santa Clara, CA, USA). Agarose gel electrophoresis was used to detect genomic DNA contamination. The purified RNA samples from the collected tissues were used for PacBio and Illumina library construction and sequencing.

Procedure of full-length transcriptome sequencing: (1) Full-length cDNA of mRNA was synthesized using SMARTer™ PCR cDNA Synthesis Kit (Clontech, Mountain View, CA, USA). (2) Amplification of full-length cDNA by PCR. (3) End repair of full-length cDNA. (4) Connect SMRT dumbbell connector. (5) Exonuclease digestion was performed to obtain sequencing library. (6) Library quality control. (7) After passing the on-machine sequencing library check, full-length transcriptome sequencing was performed using PacBio (Sequel II) instrument according to the target on-machine data volume. In this study, all samples were mixed for full-length transcriptome sequencing without replicate for comparative analysis.

Procedure of second-generation transcriptome sequencing: (1) Enrichment of eukaryotic mRNA with magnetic beads with Oligo (dT). (2) mRNA was randomly interrupted by Fragmentation Buffer. (3) mRNA was used as template to synthesize the first cDNA strand with six-base random hexamers, and then the second cDNA strand was synthesized by adding buffer, dNTPs, RNase H and DNA polymerase I. cDNA was purified by AMPure XP beads. (4) The purified double-stranded cDNA was

https://doi.org/10.1017/S0960258524000084 Published online by Cambridge University Press

then end-repaired, A-tail was added and sequencing joints were connected, and AMPure XP beads were used for fragment size selection. (5) cDNA library was obtained by PCR enrichment. (6) Library quality control. (7) After the on-machine sequencing library is checked, different libraries are pooled according to the target on-machine data amount and sequenced by Illumina NovaSeq 6000 platform. In this study, samples of five stages were used for second-generation transcriptome sequencing with three replicates.

PacBio Iso-Seq data processing and read correction

Raw sequencing data were processed using the SMRTlink4.0 software through the standard Iso-Seq protocol. The full-length non-chimeric and non-full-length (nFL) reads were obtained by filtering, polishing and sorting. Error correction was performed using an Illumina RNA-seq dataset (Chen et al., 2019). Considering the limitation of the construction of the cDNA library, the consensus sequence screened may be an nFL sequence due to the deletion of the 5′end sequence. Therefore, we merged the sequences where only 5′end exons differ, and the other exons are consistent. The longest sequence is taken as the final transcript sequence, and second-generation data are used to quantify and differentiate new transcript sequences.

Mapping reads from the PacBio library and annotation analysis

After sequencing quality control checks, low-quality conforming sequences were corrected using Illumina high-throughput sequencing results. Clean reads were separated from the raw data by removing adapter sequences and low-quality reads, the high-quality clean reads were then used as a reference for further transcriptome data analysis, and PacBio data were applied to modify the reference genome using Bowtie (v2.2.3). The gene expression levels were quantified by the RSEM software package (v1.3.1) (Li and Dewey, 2011) and were normalized by the FPKM method (Trapnell et al., 2010). All expressed transcript functions were annotated to eight public databases using BLAST software (v2.2.26): the NCBI non-redundant (Nr) protein, Protein Family (Pfam), SwissProt (which contains manually annotated and reviewed protein sequences), Clusters of Orthologous Groups of proteins (KOG/COG/eggNOG), GO and KEGG databases.

Identification of differentially expressed genes (DEGs) and transcription factors (TFs)

The expression level of transcripts was calculated by quantifying the reads according to the fragments per kilobase of transcript per million mapped reads (FPKM) values (Mortazavi et al., 2008). The transcript fold change was calculated by the formula of $\log_2$(FPKM treatment/FPKM control) using an MA plot-based method with the random sampling model via the R package DEGseq (Wang et al., 2010). The DEGs that conformed to both $|\text{fold change}| \geq 2$ and a false discovery rate < 0.01 conditions were used for subsequent analysis. Cluster analysis was performed and DEG expression patterns were assessed using the Biomarker (BMK) Cloud platform (Chang et al., 2017). The expression level (FPKM) of mutually expressed genes among different treatments was analysed using TBtools (Chen et al., 2020). Venn diagrams were generated using the Venny 2.1 tool (https://bioinfogp.cnb.csic.es/tools/venny/index.html). GO and KEGG pathway enrichment analyses of the DEGs were performed via the OmicShare Tools online platform. The potential TFs were identified by the PlantTFDB database with the default parameters (https://planttfdb.gao-lab.org/blast.php).

Establishment of the gene co-expression network

WGCNA was conducted using the R packages (Langfelder and Horvath 2008). Expression correlation coefficients were calculated for gene networks using the scale-free topology model (Wang et al., 2018). Then, the WGCNA modules (co-expression network) of eigengenes were identified and the networks correlated with physiological indices were identified with the criterion of stability correlation $P \leq 0.05$. The resulting networks were visualized with Cytoscape(v3.7.1) and MCODE (Bader and Hogue, 2003).

Real-time quantitative PCR (qRT-PCR) analysis

cDNA was synthesized from 500 ng of total RNA using ReverTra Ace® qPCR RT Master Mix (with DNase) (Toyobo Biotech Co., Ltd., Shanghai, China) according to the manufacturer's instructions. Ten sequences of hub genes derived from WGCNA were used to synthesize primers via a tool on the NCBI website (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). The qRT-PCR analysis was performed on a CFX 96 Real-Time PCR system (Bio-Rad) using 2xSG Fast qPCR Master Mix (Sangon Biotech, Shanghai, China). The qRT-PCR analyses were performed in three biological replicates with three technical replicates in 96-well plates. The thermal cycling conditions were as follows: 95°C for 3 min, followed by 40 cycles of 3 s at 95°C and 30 s at 60°C. The relative gene expression levels were calculated according to the $2^{-\Delta\Delta Ct}$ method (Arocho et al., 2006).

Results

Morphological changes at different seed development stages

Morphological changes in developing seeds were monitored from 5 DPA (S5) to 25 DPA (S25) (Fig. 1). After 5 days of flowering (S5), the seeds were short, shrivelled and bright green. After 9 days (S9), the seeds grew longer, bright green and gradually full. At 15 DPA (S15), the seed colour changed from green to yellow, the middle of the seed appeared brown and the seed became wide and full. At 20 DPA (S20), most of the seeds became reddish-brown in colour, while the seeds extended and the seed-coat

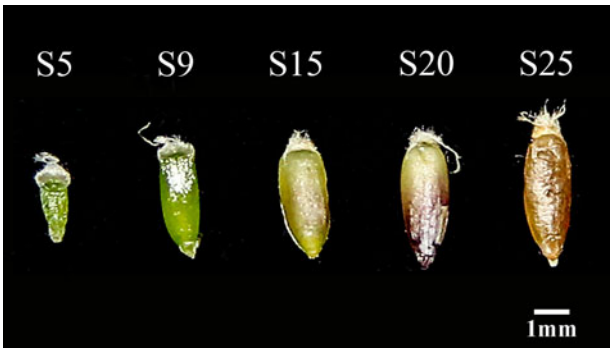


Figure. 1. Seed tissues at five different developmental stages (S5, S9, S15, S20 and S25 represent corresponding days of post-anthesis).

https://doi.org/10.1017/S0960258524000084 Published online by Cambridge University Press

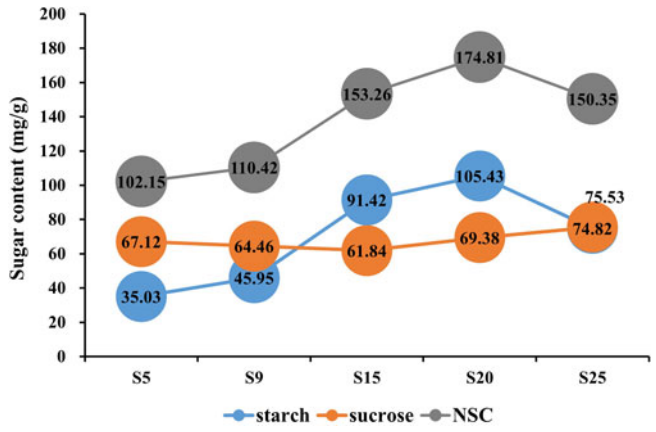


Figure. 2. Sugar content of different development stages.

wrinkled. At 25 DPA (S25), seeds turned brown and gradually lost water content. These results indicate that the seed colour of *E. sibiricus* transitions from bright green to brown as they mature, with a noticeable increase in size up until S25. In addition, the seed coat changed from crinkled to full to crinkled.

Change in sugar content at different seed development stages

In this study, the starch and sucrose content were determined by the standard curve method at five seed development stages. Non-structural carbohydrate (NSC) content was obtained by calculating the sum of starch and soluble sugar concentration. The sucrose content significantly decreased from 67.12 mg/g at 5 DPA (S5) to 61.84 mg/g at 15 DPA (S15), then remained relatively stable from 15 to 25 DPA (S15–S25). Starch content increased from 35.03 mg/g at 5 DPA (S5) to 105.43 mg/g at 20 DPA (S20), with the most rapid accumulation occurring between 9 and 15 DPA. After 20 DPA (S20), the starch content declined. The trend in NSC content mirrored that of the starch content (Fig. 2), suggesting that the period between 9 and 15 DPA is critical for seed development, characterized by sucrose decomposition and rapid starch accumulation in *E. sibiricus*.

High-throughput RNA-seq and global analysis of gene expression

A total of 105,508 full-length transcripts were generated from a library of 1–6 kb in length (Supplementary Fig. S1). The Illumina results were mapped to the PacBio library, and mapping ratios were calculated, as shown in Table 1. Among them, the Q30 values of the sequences in the 15 libraries reached 90%, and all of the total mapping ratios were greater than 60% (Table 1). In total, 82,438 redundant transcripts were annotated against eight public databases (the Nr, SwissProt, eggNOG, COG, KOG, Pfam, GO and KEGG databases) (Table 2). Among them, 78.13% of transcripts from SMRT were annotated in at least one database. The number of transcripts annotated in the eight databases ranged from 26,862 (25.46%, COG) to 81,986 (77.71%, Nr).

Differentially expressed genes (DEGs) during seed development

A total of 13,205 DEGs were identified at five stages of seed development. Among them, 6,196 (2,923 up, 3,273 down), 4,853 (3,423 up, 1,430 down), 746 (440 up, 306 down) and 1,410 (685 up, 725 down) DEGs were identified in S5 vs S9, S9 vs S15, S15 vs S20 and S20 vs S25, respectively (Fig. 3A–D). It appeared that transcription events of seed development occurred actively during S5–S15 and S20–S25. The Venn result showed nine overlap DEGs expressed in the overall comparison sets. The specific DEGs

Table 1. The Illumina high-throughput sequencing results and mapping ratio

	Clean data				Mapped reads (ratio)		
Sample names	Read sum (M)	Base sum (M)	GC (%)	Q30 (%)	Total mapped	Uniquely mapped	Multiple mapped
S5-1	31.37	9,374.18	56.42	92.85	67.74%	38.66%	26.82%
S5-2	32.48	9,711.71	56.45	93.51	69.23%	38.25%	27.88%
S5-3	27.28	8,146.86	56.37	93.82	68.38%	38.19%	27.59%
S9-1	29.03	8,680.15	54.63	92.09	66.02%	30.36%	27.41%
S9-2	21.07	6,307.19	54.83	92.24	67.44%	31.23%	28.88%
S9-3	23.29	6,961.42	55.36	92.31	66.89%	33.90%	27.13%
S15-1	25.74	7,689.36	53.71	92.08	68.97%	23.38%	32.85%
S15-2	19.92	5,911.14	51.76	91.41	63.54%	17.71%	29.03%
S15-3	30.38	9,059.35	53.73	91.40	66.96%	22.53%	32.50%
S20-1	29.42	8,778.24	54.31	90.01	64.67%	21.60%	30.66%
S20-2	23.65	7,046.25	53.80	90.12	63.65%	19.46%	31.39%
S20-3	24.73	7,393.29	54.51	88.65	58.99%	20.43%	27.62%
S25-1	31.92	9,525.74	54.90	91.10	66.43%	22.11%	35.23%
S25-2	29.11	8,693.97	54.97	90.77	64.51%	22.95%	31.48%
S25-3	24.89	7,434.04	53.98	90.69	63.12%	21.24%	30.94%

https://doi.org/10.1017/S0960258524000084 Published online by Cambridge University Press

Table 2. The transcripts functional annotation results

Anno_Database	Annotated_Number	Percentage (%)
COG_Annotation	26862	25.46
GO_Annotation	65471	62.05
KEGG_Annotation	29573	28.03
KOG_Annotation	40887	38.75
Pfam_Annotation	57745	54.73
Swissprot_Annotation	53008	50.24
eggNOG_Annotation	70405	66.73
nr_Annotation	81986	77.71
All_Annotated	82438	78.13

were 5,352 (48.3%), 3,906 (35.3%), 441(4%), 273(2.5%) in S5 vs S9, S9 vs S15, S15 vs S20 and S20 vs S25, respectively (Fig. 3E).

Hierarchical clustering of all 13,205 DEGs was performed, using the Pearson correlation method associated with average linkage clustering. As a result, eight clusters were identified for expression pattern analysis, namely cluster 1–8 (Fig. 4). On the whole, each cluster showed a peak in a certain stage. Cluster 1 and 6 contained high-expressed DEGs at S5. Cluster 3 contained high-expressed DEGs at S9. Cluster 2 contained high-expressed DEGs at S15. Cluster 4, 7 and 8 contained high-expressed DEGs at S20. Cluster 5 and 7 contained high-expressed DEGs at S25. The classification results were beneficial to the study of genes with the same expression pattern in seed development.

GO and KEGG enrichment analysis of DEGs

Gene ontology (GO) assignments were used to classify functions of the predicted seed developing DEGs in *E. sibiricus*. Among them, DEGs were categorized into three main divisions (biological process, cellular components and molecular functions). The top 20 enriched GO terms during the five seed development stages were identified (Fig. 5). All these terms were directly or indirectly related to seed development and starch and sucrose metabolism of *E. sibiricus*. The term 'vacuole' contained 335 DEGs belonging to cellular components. Vacuoles store carbohydrates, proteins and fats needed for growth, providing a source of nutrients for seed development (Shimada et al., 2018). The term 'xylan catabolic process' contained 55 DEGs belonging to biological process which is the chemical reactions and pathways resulting in the breakdown of xylan, a polymer containing a beta-1,4-linked D-xylose backbone. These DEGs of 'xylan catabolic process' were involved in seed maturation (Shen et al., 2022). The beta-amylase activity reaction was the hydrolysis of 1,4-alpha-glucosidic linkages in polysaccharides so as to remove successive maltose units from the non-reducing ends of the chains. This term involves in seed germination and maturation process (Gong et al., 2013). The term 'beta-amylase activity' contained 56 DEGs belonging to molecular functions and the expression values are shown in Fig. 6. The annotation results of the Pfam dataset indicated that these DEGs were from the *GLYCOSYL HYDROLASE FAMILY 14*, and *BETA-AMYLASE 1* (*BMY1*) was a representative gene in this family. The homologs of *BMY1*, *F01_transcript_191270* and *F01_transcript_98193* were down-regulation from S5 to S25. However, *F01_transcript_128398*, *F01_transcript_150576*, *F01_transcript_180023*, *F01_transcript_8657*, *F01_transcript_90591* and *F01_transcript_97370* were up-regulated.

From the identified differential developmental stages of the seeds and the associated DEGs group, further annotation was done by utilizing KEGG database. As a result, DEGs were categorized into three main divisions (metabolism, genetic information processing and environmental information processing). Similarly, the top 20 enriched KEGG pathways during the five seed developmental stages were identified (Fig. 7). In top 20 enriched KEGG pathways, the 'photosynthesis' pathway contained 115 DEGs. In the process of seed maturation, the genes involved in photosynthesis were down-regulated, photosynthesis in the seeds was inhibited, and the seeds turned yellow (Shen et al., 2022). The 'plant hormone signal transduction' pathway contained 190 DEGs. Many plant hormones are involved in the regulation of seed dormancy and germination, of which abscisic acid (ABA) and gibberellin are the two most important regulation hormones (Kozaki and Aoyanagi, 2022). The 'starch and sucrose metabolism' pathway contained 246 DEGs. According to annotation results, DEGs related with sucrose synthase, alpha-amylase, UDP-glucose 6-dehydrogenase, UDP-glucuronate 4-epimerase, UDP-glucuronic acid decarboxylase, starch synthase and beta-glucosidase were selected and the expression values are shown in Fig. 8.

Transcriptional networks controlling seed development

To identify the different co-expressed modules in seed development, a WGCNA with the DEGs was constructed. The minimum number of genes in each module was set to 100, and a 0.85 threshold was used to merge similar modules. In total, 19 distinct modules with various colours were ultimately identified (Fig. 9A). To identify the modules that were significantly associated with sugar content, each module was subjected to a correlation interaction analysis based on the *P*-value, the midnight blue module had a positive correlation with total starch ($r = 0.87$, $P = 0.0004$) and NSC content ($r = 0.88$, $P = 0.0004$) (Fig. 9B). The midnight blue module contained 1159 DEGs. For visualization, 64 hub genes interaction networks with scores >10 were obtained through MCODE (Supplementary Table S1). The colour and size of the circle were related to the score of the hub genes. Darker colours and larger circles indicated that the gene is strongly involved in sugar regulation (Fig. 10). The top ten hub genes were detected closely related to the sugar metabolism of seed development in *E. sibiricus*, including RING-H2 finger protein ATL8 (*F01_transcript_104493*), a member of SCP-2 sterol transfer family (*F01_transcript_142764*), protein TIFY 10A (*F01_transcript_170863*), ubiquitin-conjugating enzyme E2 34 (*F01_transcript_184778*), transmembrane emp24 domain-containing protein p24beta2 (*F01_transcript_30051*), ras-related protein Rab7 (*F01_transcript_30832*), ubiquitin-conjugating enzyme E2 11 (*F01_transcript_32363*), multiprotein-bridging factor 1a (*F01_transcript_33926*), eukaryotic translation initiation factor 5A-2 (*F01_transcript_34235*) and 60S acidic ribosomal protein P2B (*F01_transcript_38379*) (Table 3). In the meanwhile, the ten hub genes were used for qRT-PCR to validate their expression levels at five stages of seed development in *E. sibiricus*. The expression levels of the hub genes were generally consistent with the FPKM value by transcription sequencing, suggesting relative rationality and

https://doi.org/10.1017/S0960258524000084 Published online by Cambridge University Press

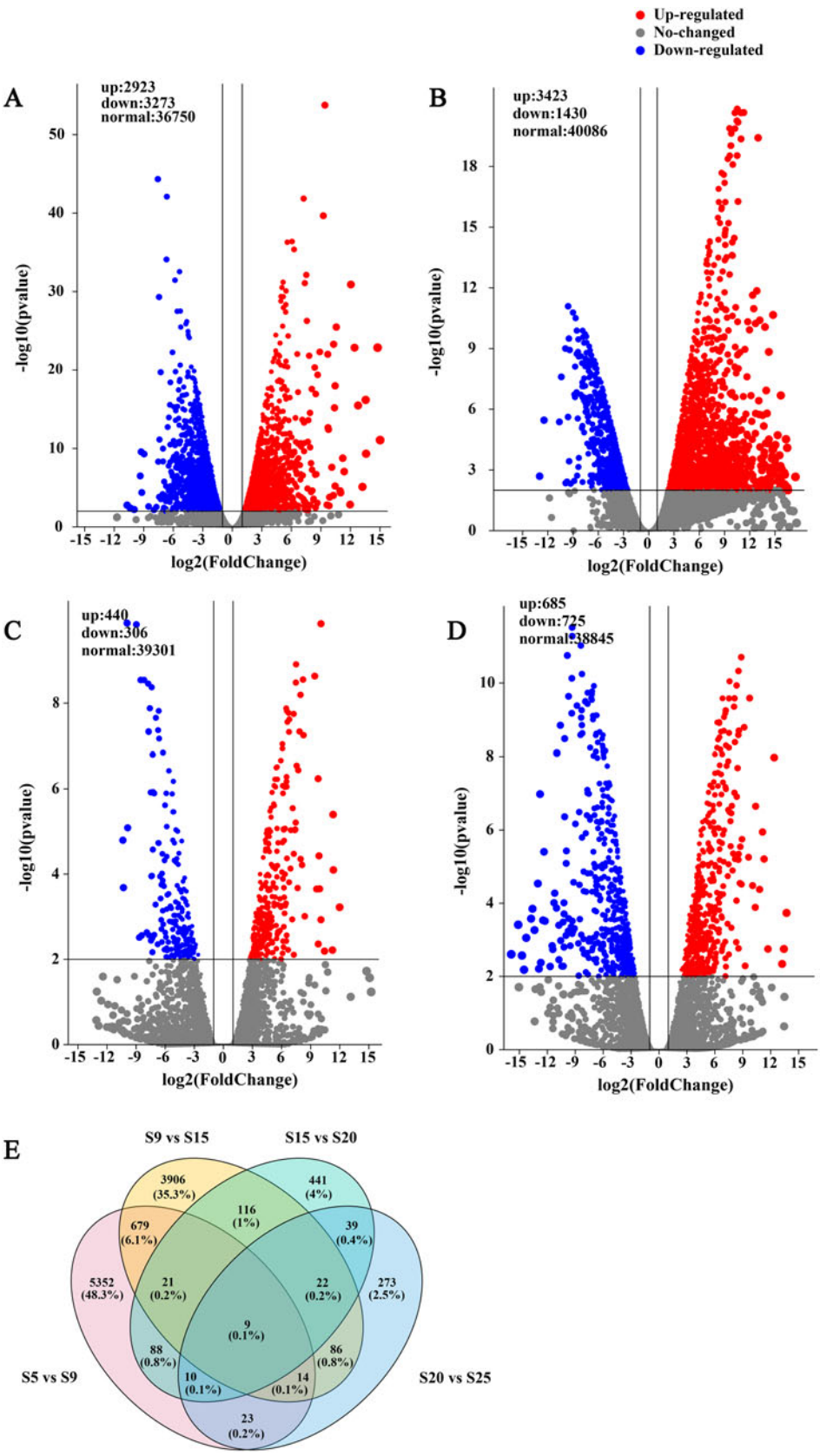


Figure. 3. Volcano plot and Venn diagram of DEGs in *E. sibiricus* seed embryos. (A) S5 vs S9. (B) S9 244 vs S15. (C) S15 vs S20. (D) S20 vs S25. (E) Venn diagram represents the number of overlapping DEGs 245 between S5, S9, S15, S20, S25.

accuracy of the transcriptome analysis in this study (Fig. 11). The expression levels of *F01_transcript_170863*, *F01_transcript_184778*, *F01_transcript_30051* and *F01_transcript_33926* were gradually rising during S5–S10 and decreasing during S20–S25. The expression of other hub genes increased during the whole experiment stage.

https://doi.org/10.1017/S0960258524000084 Published online by Cambridge University Press

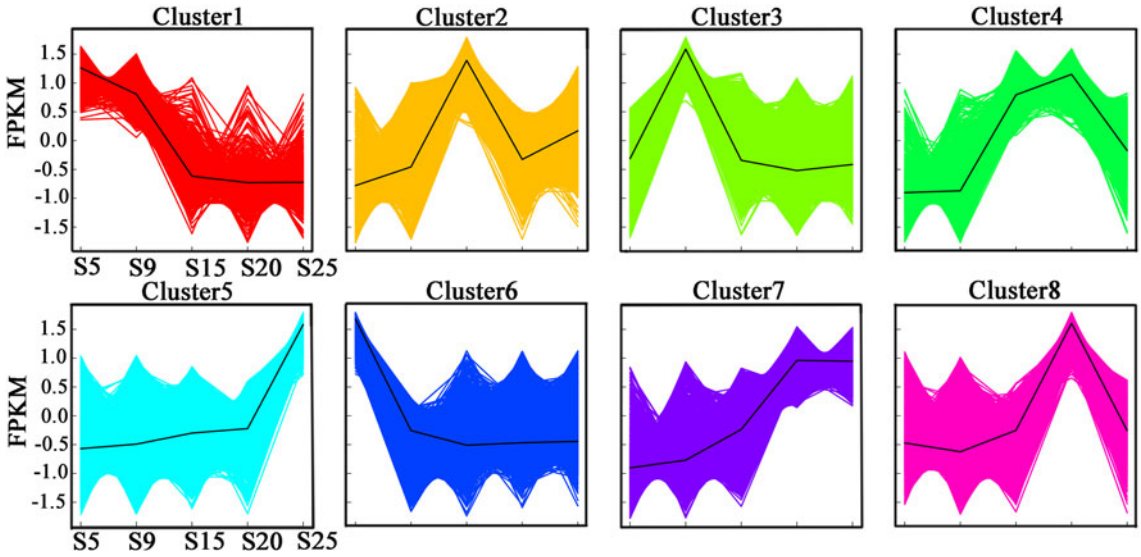
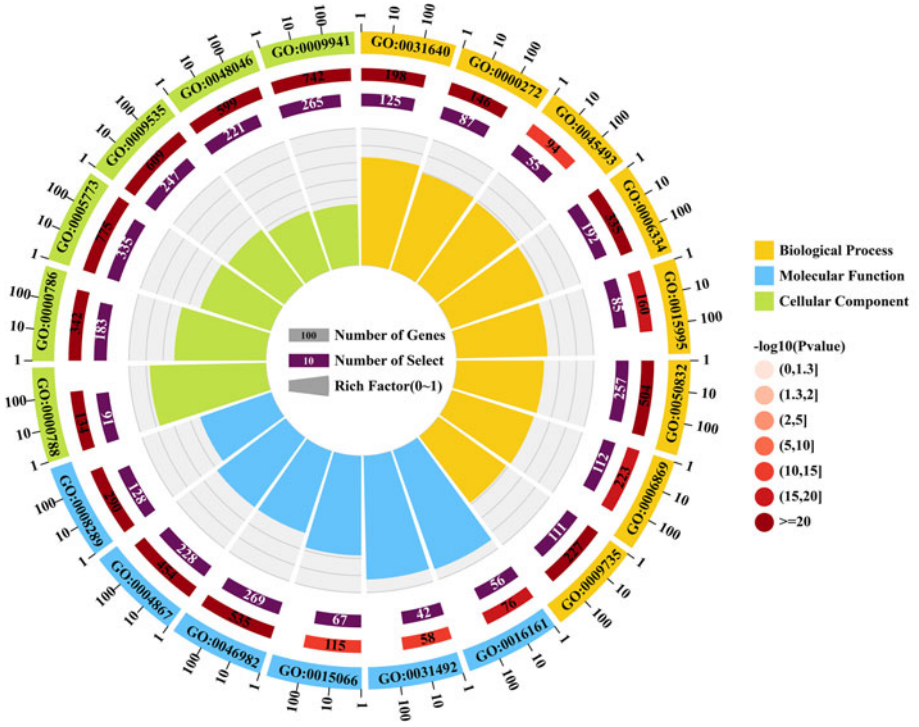


Figure. 4. The DEGs expression pattern analysis of five *E. sibiricus* developing stages

Identification of transcription factors (TFs) associated with seed development

TFs are powerful regulators in controlling gene expression in every aspect of plant growth and development (Riechmann and Ratcliffe, 2000; Franco-Zorrilla et al., 2014). To gain insight into the involvement of TFs and to determine how they regulate seed development in *E. sibiricus*, all the differentially expressed TF-encoding genes were identified. In total, 54 TF families were identified in the process of seed development, and the top 21 families were filtered out, including NAM, ATAF1/2, CUC1/2 (NAC (63)), APETALA2/ethylene responsive factor (AP2/ERF (51), bZIP (43), v-myb avian myeloblastosis viral oncogene homolog (MYB, 43), bHLH (38)) and C2H2 zinc finger proteins (C2H2, 37), WRKY (37), MYB-related (22), GAI, RGA, SCR (GRAS, 18), C3H zinc finger proteins (C3H, 17) and NF-YB (7) (Fig. 12). In plants, bZIP TFs regulate processes such as pathogen defence, light and stress signalling, seed maturation and flower development (Jakoby et al., 2002). According to annotation results, TFs involved in seed dormancy control and the homologs of *ABI5* in the bZIP family. In addition, TFs from C2C2-Dof and NF-YB also were concerned and the expression patterns are shown in Fig. 13.



GO term Top 20

ID	Descrption
GO:0005773	vacuole
GO:0008289	lipid binding
GO:0009535	chloroplast thylakoid membrane
GO:0009735	response to cytokinin
GO:0009941	chloroplast envelope
GO:0048046	apoplast
GO:0050832	defense response to fungus
GO:0006334	nucleosome assembly
GO:0046982	protein heterodimerization activity
GO:0000786	nucleosome
GO:0004867	serine-type endopeptidase inhibitor activity
GO:0031640	killing of cells of other organism
GO:0000788	nuclear nucleosome
GO:0000272	polysaccharide catabolic process
GO:0016161	beta-amylase activity
GO:0006869	lipid transport
GO:0015995	chlorophyll biosynthetic process
GO:0015066	alpha-amylase inhibitor activity
GO:0031492	nucleosomal DNA binding
GO:0045493	xylan catabolic process

Figure. 5. The GO enrichment analysis of DEGs. The colour from the light pink to dark red module represents –log10 (*P*-value), the number on the module represents the number of genes enriched in the pathway and the number on the purple module represents the number of genes in the pathway. And the fan length on the pathway represents the enrichment factor.

https://doi.org/10.1017/S0960258524000084 Published online by Cambridge University Press

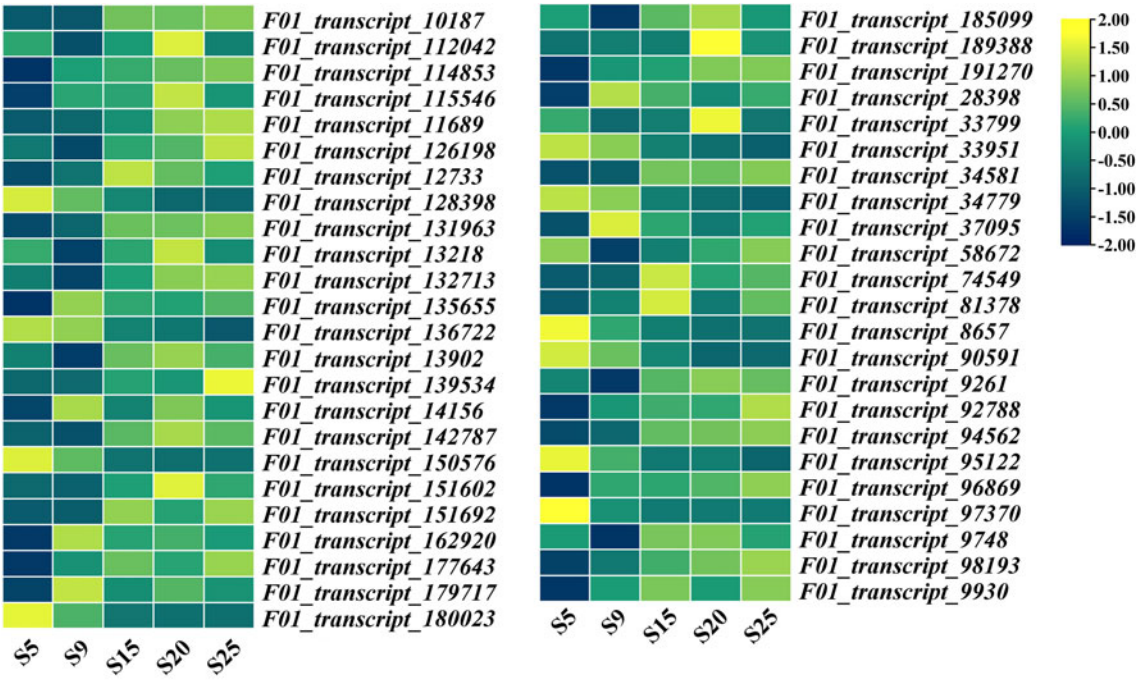


Figure. 6. The expression heatmap of DEGs involved in beta-amylase activity

Discussion

Full-length transcriptome sequencing is a useful tool for genes selecting

Full-length transcriptome sequencing has the advantages of comprehensively identifying alternative splicing, discovering more new genes and identifying more lncRNAs (Grabherr et al., 2011). In this study, we provided a comprehensive transcriptomic profile of the seed-developing network in different stages and identified a total of 105,508 full-length transcripts in 15 sample libraries and 13,205 DEGs at five stages. Among these transcripts, more than 78% were significantly similar (in terms of their sequence) to genes in public databases, the percentage of which was lower than drought research in *E. sibiricus* using full-length sequencing (96.75%) (Yu et al., 2023), but greater than a previous study in *E. sibiricus* using NGS (46.69%) (Xie et al., 2017). For

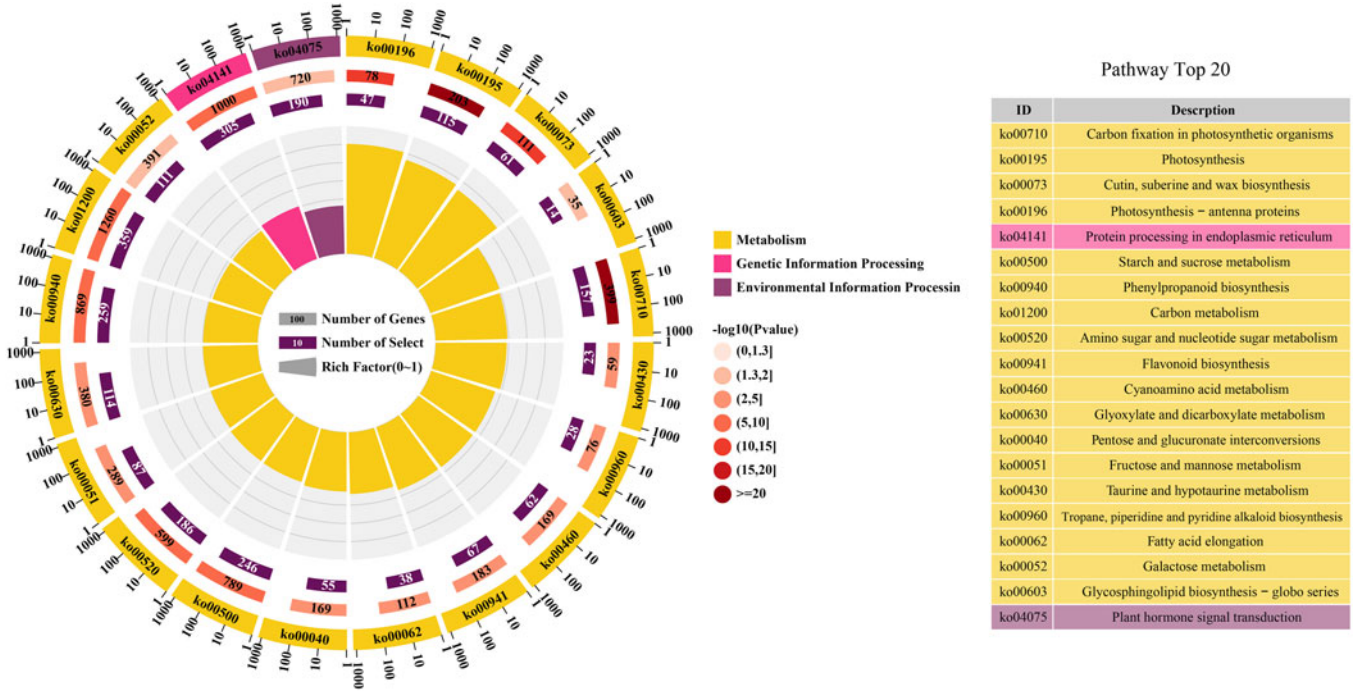


Figure. 7. The KEGG enrichment analysis of DEGs. The colour from the light pink to dark red module represents –log10 (*P*-value), the number on the module represents the number of genes enriched in the pathway and the number on the purple module represents the number of genes in the pathway. And the fan length on the pathway represents the enrichment factor.

https://doi.org/10.1017/S0960258524000084 Published online by Cambridge University Press

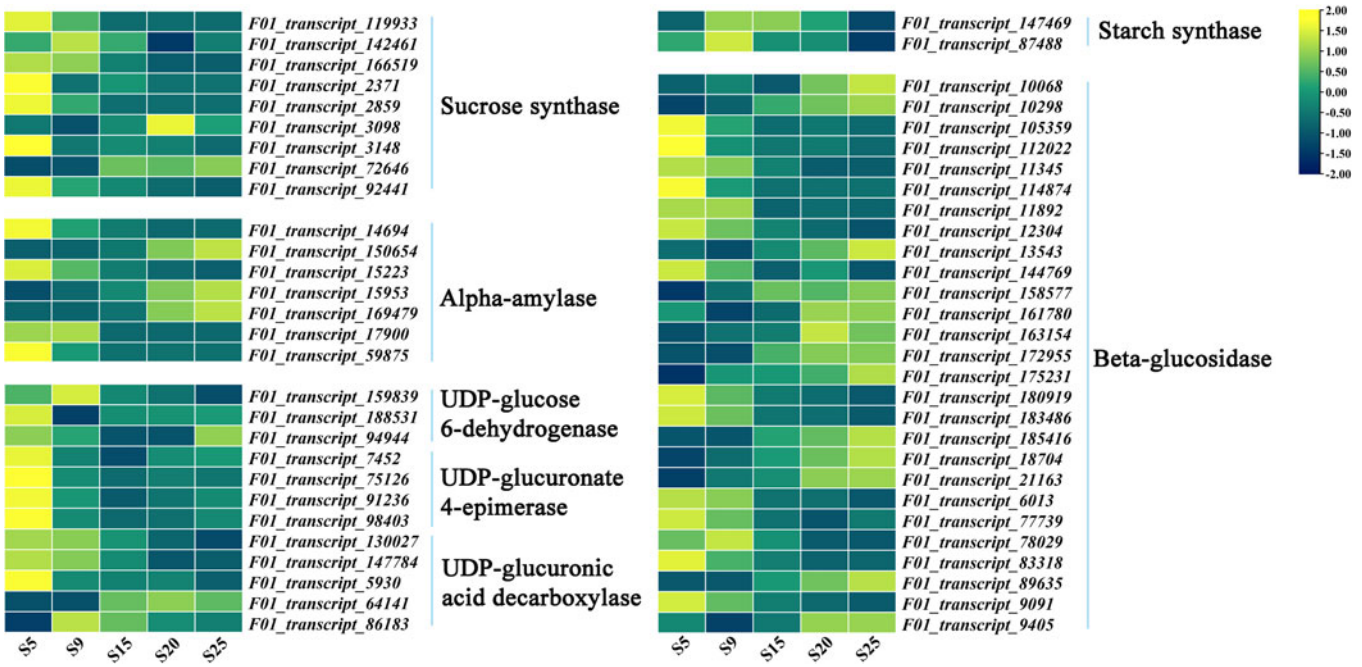


Figure. 8. The expression heatmap of DEGs involved in starch and sucrose metabolism

forage grasses, full-length transcriptomics has been widely used in recent years. A first full-length transcriptome database under 300 mM NaCl treatment at different time points was constructed for *Bromus inermis* L. revealing the molecular regulation response to salt stress (Li et al., 2022). In *Trifolium ambiguum* M., DEGs were identified in five tissues using full-length and NGS to provide insight in its growth and development (Yin et al., 2020). In *Pennisetum giganteum*, leaf and root tissues under room and chilling temperature were used for comprehensive full-length transcriptome analysis to explore the cold stress response mechanism (Li et al., 2020). However, there are no studies on the use of full-length sequencing to explore forage grass seed development. Our research opened a new horizon for the study of forage seeds.

Key stage determination of seed development in E. sibiricus

According to morphological observations, three key seed developmental stages could be determined: embryogenesis before 5 DPA, seed formation between 5 and 15 DPA, followed by maturation and desiccation around 20 DPA. In rice seeds, cellularization of endosperm was completed before 6 DPA, seeds were formed at 6–10 DPA and embryos matured at 11–20 DPA (Xue et al., 2012). The seeds of soybean mature and increase dry weight at 20–30 DPA (Yao et al., 2023). In addition, seed shattering is a sign that seed development has entered the maturity stage. Our previous study found that seed shattering occurred at 28 days after heading in *E. sibiricus* (Xie et al., 2017). In the current study, the DEGs linked to development and cell death such as *F01_transcript_110107*,

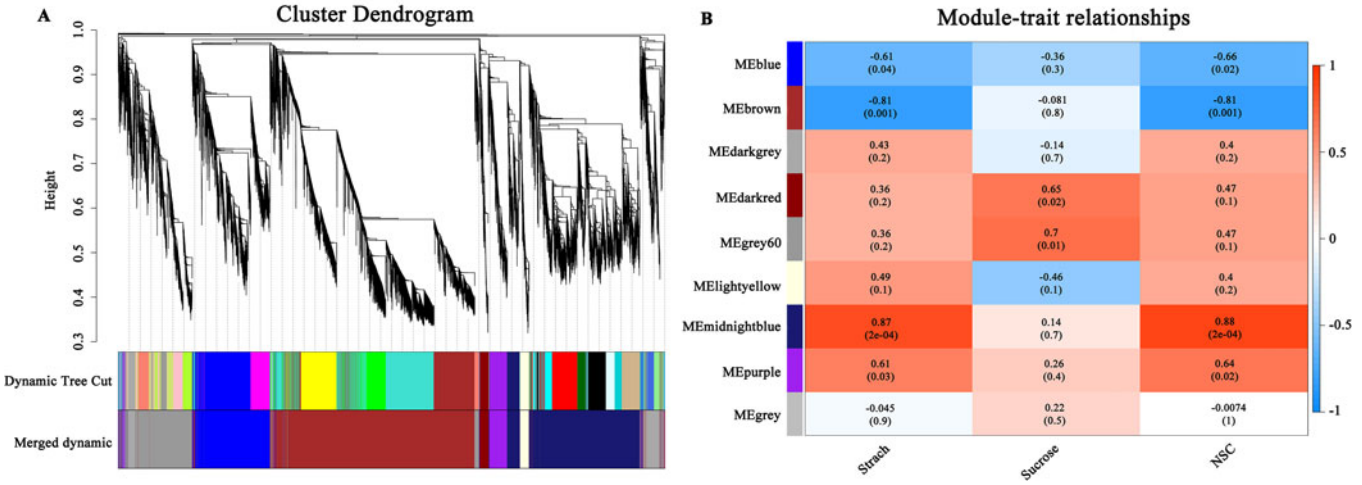


Figure. 9. Weighted gene co-expression network analysis (WGCNA). (A) Clustering dendrograms of genes. (B) Module correlations and corresponding *P*-values (in parentheses) of starch, sucrose and NSC. The colour scale on the right shows module–trait correlation from –1 (blue) to 1 (red).

https://doi.org/10.1017/S0960258524000084 Published online by Cambridge University Press

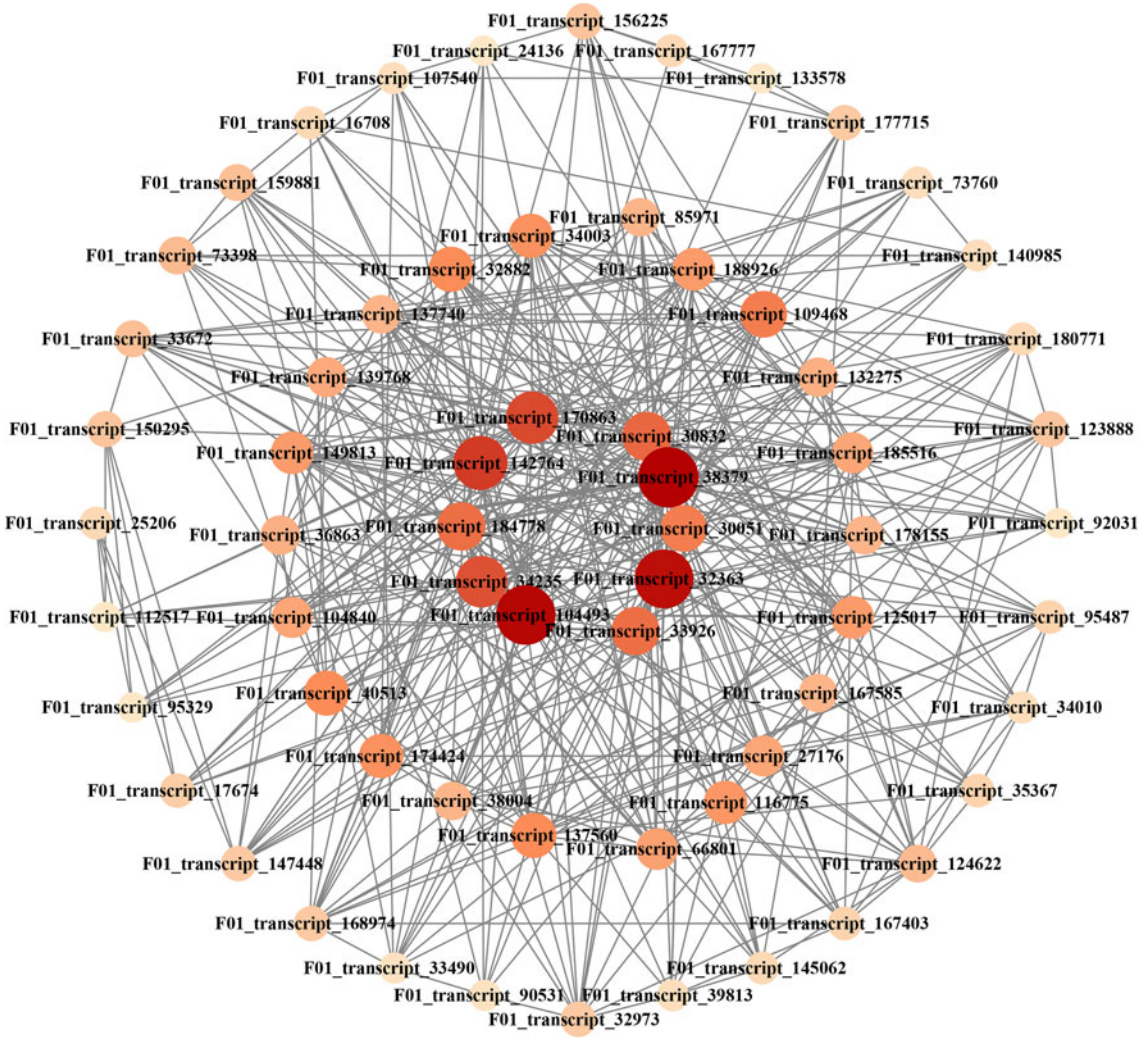


Figure. 10. Hub genes co-expression network analysis in midnight blue module

F01_transcript_18619 reached maximum expression at S20. Above all, the seed development time differences were due to the developmental difference of the species itself, the seed maturity stage of *E. sibiricus* was S20.

Key GO term and KEGG pathway involving in seed development

A key term 'beta-amylase activity' was obtained by GO enrichment. Beta-amylase activity increased with seed development to

Table 3. The annotation information of ten hub genes in midnight blue module based on database

Gene ID	Pfam annotation	Swissprot annotation
F01_transcript_104493	Zinc finger, C3HC4 type	RING-H2 finger protein ATL8
F01_transcript_142764	SCP-2 sterol transfer family	–
F01_transcript_170863	tify domain, divergent CCT motif	Protein TIFY 10A
F01_transcript_184778	Ubiquitin-conjugating enzyme	Ubiquitin-conjugating enzyme E2 34
F01_transcript_30051	emp24/gp25L/p24 family/GOLD	Transmembrane emp24 domain-containing protein p24beta2
F01_transcript_30832	Ras family, Gtr1/RagA G protein conserved region	Ras-related protein Rab7
F01_transcript_32363	Ubiquitin-conjugating enzyme	Ubiquitin-conjugating enzyme E2 11
F01_transcript_33926	Multiprotein bridging factor 1, helix-turn-helix	Multiprotein-bridging factor 1a
F01_transcript_34235	Eukaryotic elongation factor 5A hypusine, DNA-binding OB fold	Eukaryotic translation initiation factor 5A-2
F01_transcript_38379	60s acidic ribosomal protein	60S acidic ribosomal protein P2B

https://doi.org/10.1017/S0960258524000084 Published online by Cambridge University Press

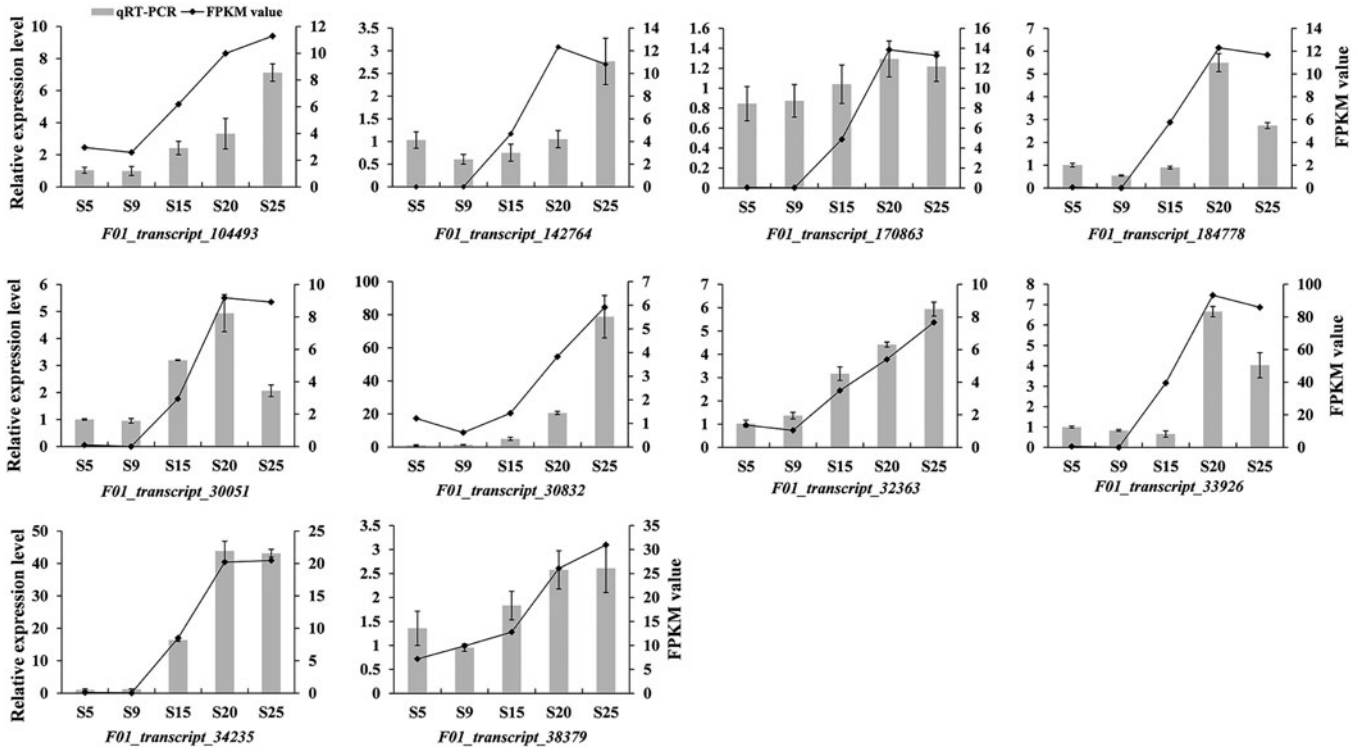


Figure. 11. qRT-PCR confirmation of ten hub genes

improve grain quality, and *BETA-AMYLASE 1* (*BMY1*) was a representative gene in this term (Gong et al., 2013). In this study, the homologs of *BMY1*, *F01_transcript_191270* and *F01_transcript_98193* were down-regulated between S5 and S25. However, *F01_transcript_128398*, *F01_transcript_150576*, *F01_transcript_180023*, *F01_transcript_8657*, *F01_transcript_90591* and *F01_transcript_97370* were up-regulated. Most transcripts were up-regulated, which is consistent with a previous study in barley. Vinje et al. analysed the expression of the *BMY1* gene in barley, which rose sharply between 5 and 17 DPA, peaking at 25 DPA (Vinje et al., 2019). The down-regulated expression trends of *F01_transcript_191270* and *F01_transcript_98193* may be related to intron polymorphism. Wu et al. found that intron 3 of *BMY1* was polymorphic, and the presence of three alleles had different effects on beta-amylase activity (Wu et al., 2022).

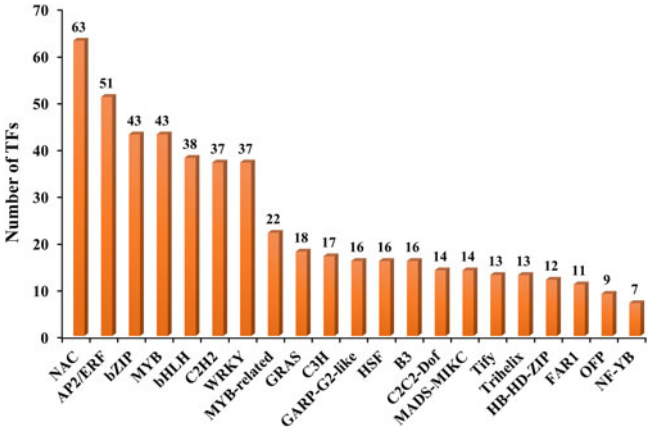


Figure. 12. Distribution of the top 21 TF families involved in seed development of *E. sibiricus*

Photosynthesis and hormone signal transduction pathways play significant roles in seed development of *E. sibiricus*. In this study, the annotation results of DEGs contained photosystem I reaction centre, photosystem II reaction centre and oxygen-evolving enhancer protein. Photosynthesis is a basic biological process that requires energy for plant growth and development. Seeds as non-foliar green tissues also can contribute energy (Simkin et al., 2020). In soybean, photosynthesis of the pod and seed can provide 13–14% of the energy for seed maturation (Cho et al., 2023). In *T. aestivum*, the contribution of seed photosynthesis to yield ranged from 12 to 42% (Maydup et al., 2010). However, photosynthesis is not necessary in the final stage of seed maturation. If chlorophyll degradation is not complete, it will lead to 'green seed problems' and affect the yield (Smolikova and Medvedev, 2016). In *Chenopodium quinoa*, the 'photosynthesis' pathway also was identified during seed maturation using proteomic analysis (Shen et al., 2022).

In hormone signal transduction, ABA is the main hormone regulating seed development. The accumulation of ABA occurred in two stages, the early stage and the middle stage of seed development (Kozaki and Aoyanagi, 2022). The key gene *ABI3* is expressed during seed development and interacts with *ABI5* in relation to ABA signalling (Collin et al., 2021). In this study, the homologs of *ABI3* and *ABI5* were identified and the putative flow chart of *ABI3* and *ABI5* participating in the ABA signalling pathway is shown in Fig. 14. Expression of the homologs of *ABI3*, *F01_transcript_170113*, *F01_transcript_96236*, *F01_transcript_23598* and *F01_transcript_62821*, was up-regulated. The homologs of *ABI5* divided into two groups, up-regulated and down-regulated. Expression of *ABI3* and *ABI5* inhibits seed germination, while mutation of *ABI3* may reduce seed dormancy

https://doi.org/10.1017/S0960258524000084 Published online by Cambridge University Press

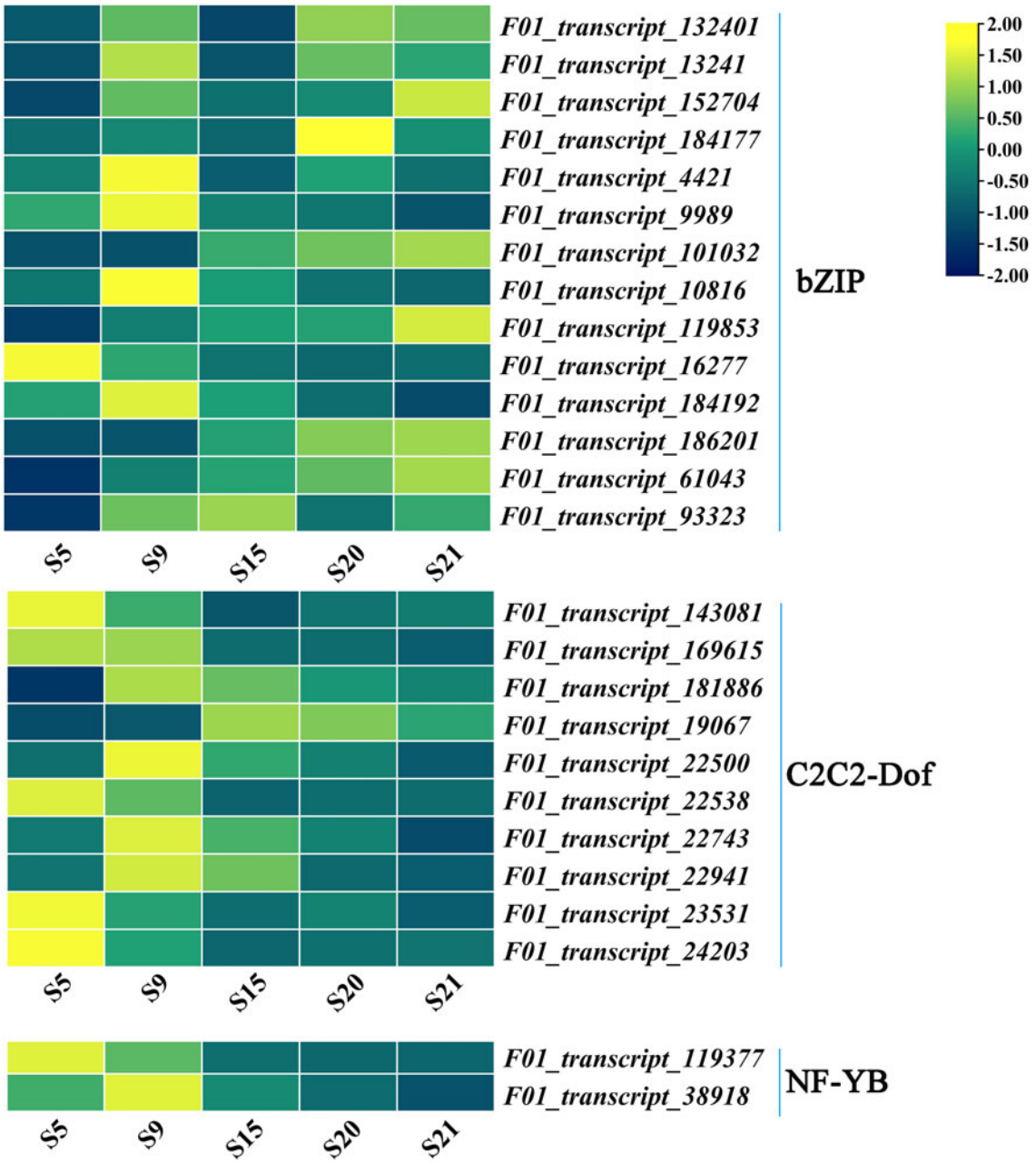


Figure. 13. The expression heatmap of TFs from bZIP, C2C2-Dof and NF-YB families.

(Verma et al., 2022). *ABI3* was up-regulated in *E. sibiricus*, and the seeds may mature into dormancy. The function of these two genes can be studied by seed dormancy and germination experiments in the future.

Transcript factors regulating seed development of E. sibiricus

In our dataset, about 50% of the TFs were found to be differentially expressed between different developmental stages. The TFs from the bZIP, C2C2-Dof and NF-YB families were found to be related with seed development in *E. sibiricus*. In the bZIP family, TFs involved in seed dormancy control and the homologs of *ABI5* were identified. The homologs of *ABI5* were consistent with what was identified in the ABA signalling pathway. TFs involved in seed dormancy control contained the homologs of *HISTONE BINDING PROTEIN-1B* (*HBP-1b*) and *TGACG-SEQUENCE-SPECIFIC DNA-BINDING PROTEIN* (*TGA*). In this study, the expression of *HBP-1b* homologs *F01_transcript_1184177* peaked at S20 and *F01_transcript_152704* peaked at S25. Studies on *HBP-1b* are scarce, the transcript and protein abundance of *HBP-1b* were significantly increased during embryo germination in rice (Sano et al., 2022). *TGA* plays an important role in plant defence but other functions have not been studied (Ullah et al., 2019). *TGAs* were also identified in studies of wheat seed storage protein gene regulators, but their functions were not verified (Luo et al., 2021).

Ten TFs were identified from the C2C2-dof family, five of which were down-regulated and five TFs were down-regulated after up-regulation. In rice, 30 Dof TFs were divided into four groups based on their expression pattern during seed

https://doi.org/10.1017/S0960258524000084 Published online by Cambridge University Press

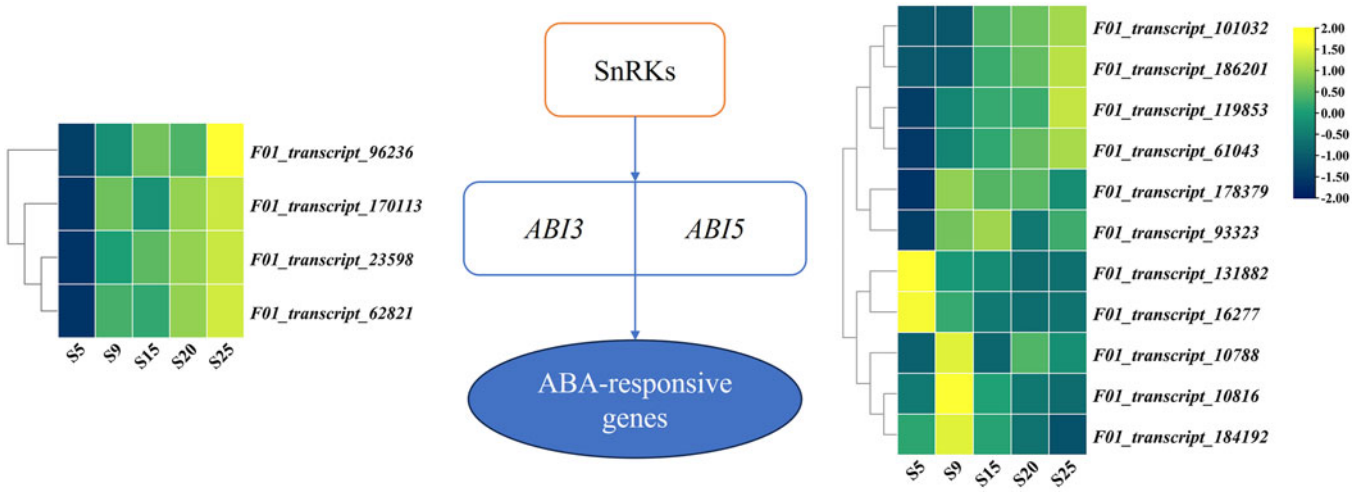


Figure. 14. The putative flow chart of *ABI3* and *ABI5* participated in ABA signalling pathway

development (Gaur et al., 2011). In the NF-YB family, the expression of *F01_transcript_119377* (*NF-YB3*) peaked at S5 and *F01_transcript_38918* (*NF-YB1*) peaked at S9. *OsNF-YB1* was shown to be involved in the regulation of seed size and also affected seed starch characteristics and grain filling rate in rice (Xu et al., 2016). Combined with physiological results, the content of starch in seeds increased rapidly in S9–S15, and *F01_transcript_38918* may be involved in regulation.

The mechanisms of starch and sucrose during seed development in E. sibiricus

During this study, starch and sucrose were the focus of attention. Starch is the main polysaccharide stored in seeds (Aguirre et al., 2018). Sucrose is the main product of photosynthesis, and the accumulation of storage compounds in seeds depends on their ability to import sucrose from their parent tissues (Weber et al., 2005). Sugar contents determination implied that starch content

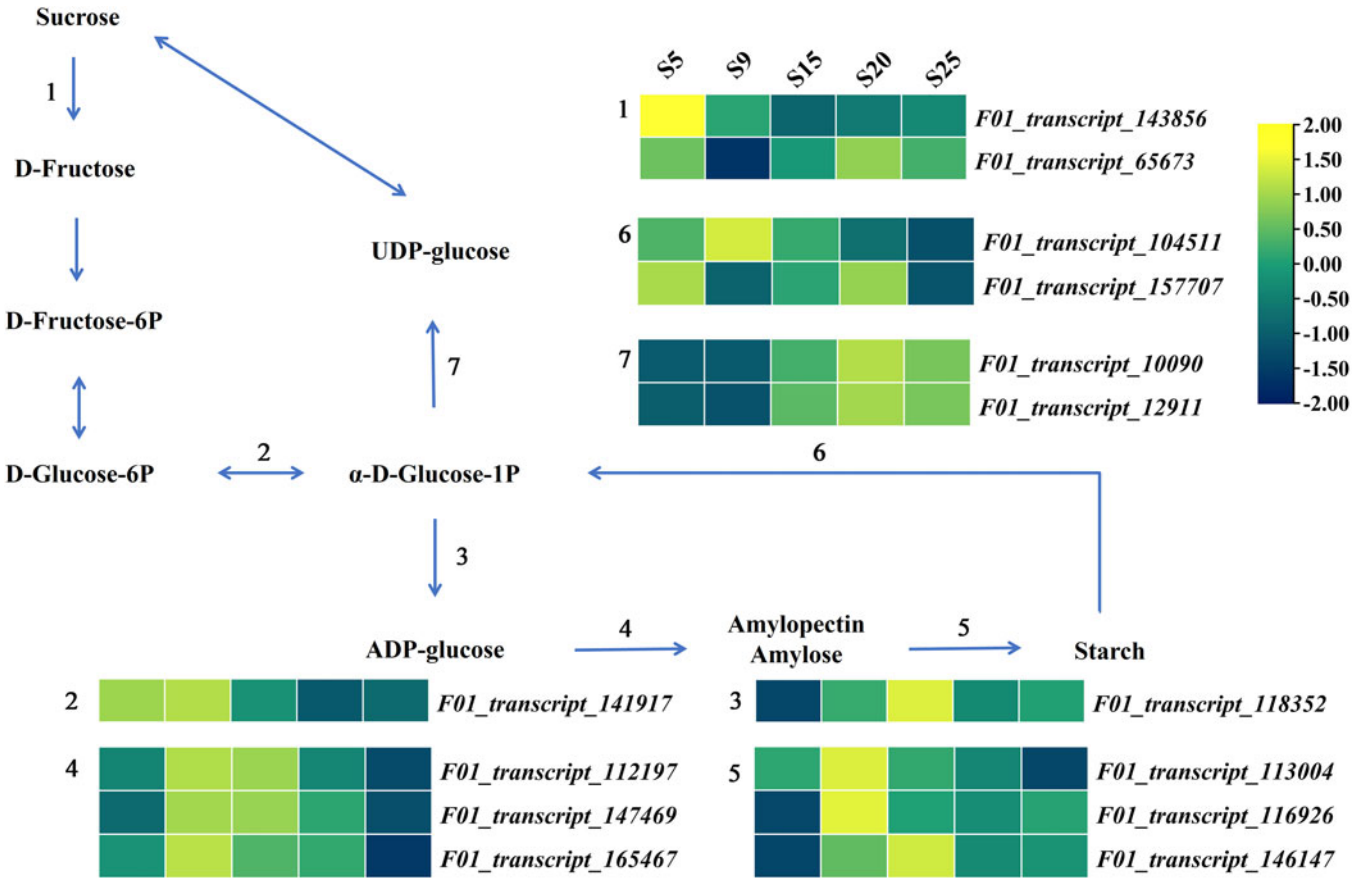


Figure. 15. The putative mechanism of starch and sucrose metabolism pathway

https://doi.org/10.1017/S0960258524000084 Published online by Cambridge University Press

increased from S5 to S20 and then decreased at S25, and sucrose content decreased from S5 to S15 and then increased. The variation range of sucrose content increased after an initial decline. In *T. aestivum*, the accumulation of starch was reported to gradually increase in the later stages of seed development, while sucrose is gradually consumed with the progress of seed development (Shewry et al., 2009; Weichert et al., 2010). In this study, the change of sugar content from S5 to S20 was consistent with *T. aestivum*, but the change at S25 was different. This may be due to a slower rate of starch accumulation or a blockage of sucrose conversion. In *triticale* (x *Triticosecale Wittmack*), the endosperm underwent programmed cell death at 21 DPA with a decrease in starch accumulation (Li et al., 2010). We also found the expression of genes that cause cell death at S20. In *H. vulgare*, programmed cell death was affected, resulting in impaired sucrose distribution and less starch accumulation (Radchuk et al., 2021).

Based on the result of 'starch and sucrose metabolism' pathway, we constructed a putative mechanistic model of starch and sucrose metabolism in *E. sibiricus* (Fig. 15). In this model, 14 DEGs are involved, including two DEGs regulating the hydrolysis of sucrose to D-fructose, one DEG regulating the conversion of D-glucose-6P and α-D-glucose-1P, one DEG regulating the production of ADP-glucose, six genes regulating starch synthesis, two genes regulating the production of α-D-glucose-1P from starch and two genes regulating the production of UDP-glucose. Many studies have shown that ADP-glucose pyrophosphorylase mediates key control steps in the starch synthesis pathway through *SOLUBLE STARCH SYNTHASE* (SS) and *STARCH BRANCHING ENZYME* (*SBE*) genes (Wei et al., 2017; Ferrero et al., 2020; Prathap and Tyagi, 2020). In this study, *F01_transcript_147469*, *F01_transcript_112197* and *F01_transcript_1665467*, which are homologs of *WHEAT STARCH SYNTHASE I* (*SSI*) (Fujita et al., 2011), were highly expressed at S9–S15. *F01_transcript_113004* was highly expressed at S9 and *F01_transcript_1146147* was highly expressed at S15. Both are homologs of *STARCH BRANCHING ENZYME* I (*SBEI*) (Utsumi et al., 2022). The homolog of *SBEII*, *F01_transcript_116926*, was highly expressed at S9. It can be seen that S9–S15 is the stage of rapid starch accumulation. The pattern of gene expression corresponds to physiological results. Ten hub genes related with change of starch and sucrose were identified by WGCNA (Table 3). The highest correlation genes are *F01_transcript_104493*, *F01_transcript_142764*, *F01_transcript_32363 and F01_transcript_38379*. *F01_transcript_104493* is a homolog of *RING-H2 FINGER PROTEIN ATL8*. *ATL8* plays a role in the embryonic development of rice and Arabidopsis (Serrano et al., 2006). *F01_transcript_142764* is a member of the *SCP-2 STEROL TRANSFER* family. *F01_transcript_32363* is related to *UBIQUITIN-CONJUGATING ENZYME E2 11*. *F01_transcript_38379* may be the homolog of *60S ACIDIC RIBOSOMAL PROTEIN P2B*. However, none of these have been reported to be related to seed development. In this study, qRT-PCR was used to verify expression patterns of ten hub genes and, subsequently, these genes could be verified by overexpression, knockout and gene editing in *E. sibiricus* to confirm that sugar accumulation and metabolism in seeds is affected.

Conclusion

This study aimed to investigate the molecular mechanisms underlying seed development, with a focus on differences in starch and sucrose content. We utilized full-length transcriptome sequencing on *E. sibiricus* seeds from five developmental stages. Key Go term 'Beta-amylase activity', and Key pathways such as 'Starch and sucrose metabolism', 'photosynthetic-related' and 'hormone signal transduction' were identified. Additionally, specific genes including *F01_transcript_104493*, *F01_transcript_142764*, *F01_transcript_32363*, *F01_transcript_38379*, *F01_transcript_170113*, *F01_transcript_119377* and *F01_transcript_147469* were found to be potentially involved in seed development and sugar content accumulation. These pathways and genes were identified as potentially significant through annotation and WGCNA, and their roles were further supported by additional analyses. These findings not only advance our understanding of seed development in *E. sibiricus*, but also provide valuable information for future seed production and breeding programmes, potentially leading to improved crop yields and quality.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0960258524000084.

Data availability statement. The *E. sibiricus* transcriptome sequencing data have been deposited in the National Center for Biotechnology Information (NCBI), under accession number SRP471113: SRX22478183–SRX22478198.

Funding statement. This work was supported by earmarked fund for CARS (CARS-34), Chinese National Natural Science Foundation (32271748, 31971751), the Leading Scientist Project of Qinghai Province (2023-NK-147), Leading Scientist Project of Gansu Province (23ZDAK013), Gansu Provincial Science and Technology Major Projects (22ZD6NA007) and the Fundamental Research Fund for the Central Universities (lzujbky-2021-ct21).

Author contributions. W. X. Liu and W. G. Xie designed and managed the experiments and reviewed the manuscript. Y. Y. Zheng and X. S. Lin performed the experiments, analysed the data, made the figures and drafted the manuscript. All authors read and approved the final manuscript.

Competing interests. The authors declare that there is no conflict of interest.

References

Aguirre M, Kiegle E, Leo G and Ezquer I (2018) Carbohydrate reserves and seed development: an overview. *Plant Reproduction* **31**, 263–290. https://doi.org/10.1007/s00497-018-0336-3

Arocho A, Chen B, Ladanyi M and Pan Q (2006) Validation of the 2-DeltaDeltaCt calculation as an alternate method of data analysis for quantitative PCR of BCR-ABL P210 transcripts. *Diagnostic Molecular Pathology* **15**, 56–61. https://doi.org/10.1097/00019606-200603000-00009

Bader GD and Hogue CWV (2003) An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4**, 2. https://doi.org/10.1186/1471-2105-4-2

Baud S, Boutin JP, Miquel M, Lepiniec L and Rochat C (2002) An integrated overview of seed development in *Arabidopsis thaliana* ecotype WS. *Plant Physiology and Biochemistry* **40**, 151–160. https://doi.org/10.1016/s0981-9428(01)01350-x

Chang Y, Nguyen BH, Xie YJ, Xiao BZ, Tang N, Zhu WL, Mou TM and Xiong LZ (2017) Co-overexpression of the constitutively active form of *OsbZIP46* and ABA-activated protein kinase *SAPK6* improves drought and temperature stress resistance in Rice. *Frontiers in Plant Science* **8**, 1102. https://doi.org/10.3389/fpls.2017.01102

Chen XZ, Li JR, Wang XB, Zhong LT, Tang Y, Zhou XX, Liu YT, Zhan RT, Zheng H, Chen WW and Chen LK (2019) Full-length transcriptome sequencing and methyl jasmonate-induced expression profile analysis of genes related to patchoulol biosynthesis and regulation in Pogostemon cablin. *BMC Plant Biology* **19**, 266. https://doi.org/10.1186/s12870-019-1884-x

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y and Xia R (2020) TBtools: An integrative toolkit developed for interactive analyses of big

biological data. *Molecular Plant* **13**, 1194–1202. https://doi.org/10.1016/j.molp.2020.06.009

Cho YB, Stutz SS, Jones SI, Wang Y, Pelech EA and Ort DR (2023) Impact of pod and seed photosynthesis on seed filling and canopy carbon gain in soybean. *Plant Physiology* **193**, 966–979. https://doi.org/10.1093/plphys/kiad324

Collin A, Daszkowska-Golec A and Szarejko I (2021) Updates on the role of abscisic acid insensitive 5 (*ABI5*) and abscisic acid-responsive element binding factors (*ABFs*) in ABA signaling in different developmental stages in plants. *Cells* **10**, 1996. https://doi.org/10.3390/cells10081996

Feng L, Zhu SS, Liu F, He Y, Bao YD and Zhang C (2019) Hyperspectral imaging for seed quality and safety inspection: a review. *Plant Methods* **15**, 91. https://doi.org/10.1186/s13007-019-0476-y

Ferrero DML, Piattoni CV, Asencion Diez MD, Rojas BE, Hartman MD, Ballicora MA and Iglesias AA (2020) Phosphorylation of ADP-glucose pyrophosphorylase during wheat seeds development. *Frontiers in Plant Science* **11**, 1058. https://doi.org/10.3389/fpls.2020.01058

Franco-Zorrilla JM, Lopez-Vidriero I, Carrasco JL, Godoy M, Vera P and Solano R (2014) DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proceedings of the National Academy of Sciences of the USA* **111**, 2367–2372. https://doi.org/10.1073/pnas.1316278111

Fujita N, Satoh R, Hayashi A, Kodama M, Itoh R, Aihara S and Nakamura Y (2011) Starch biosynthesis in rice endosperm requires the presence of either *starch synthase I* or *IIIa*. *Journal of Experimental Botany* **62**, 4819–4831. https://doi.org/10.1093/jxb/err125

Gallardo K, Firnhaber C, Zuber H, Hericher D, Belghazi M, Henry C, Kuster H and Thompson R (2007) A combined proteome and transcriptome analysis of developing *Medicago truncatula* seeds. *Molecular & Cellular Proteomics* **6**, 2165–2179. https://doi.org/10.1074/mcp.M700171-MCP200

Garg R, Singh VK, Rajkumar MS, Kumar V and Jain M (2017) Global transcriptome and coexpression network analyses reveal cultivar-specific molecular signatures associated with seed development and seed size/weight determination in chickpea. *Plant Journal* **91**, 1088–1107. https://doi.org/10.1111/tpj.13621

Gaur VS, Singh US and Kumar A (2011) Transcriptional profiling and in silico analysis of Dof transcription factor gene family for understanding their regulation during seed development of rice *Oryza sativa* L. *Molecular Biology Reports* **38**, 2827–2848. https://doi.org/10.1007/s11033-010-0429-z

Gong X, Westcott S, Zhang XQ, Yan G, Lance R, Zhang G, Sun D and Li C (2013) Discovery of novel *BMY1* alleles increasing β-amylase activity in Chinese landraces and Tibetan wild barley for improvement of malting quality via MAS. *PLoS ONE* **8**, e72875. https://doi.org/10.1371/journal.pone.0072875

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N and Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644–652. https://doi.org/10.1038/nbt.1883

Jakoby M, Weisshaar B, Dröge-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T and Parcy F (2002) bZIP transcription factors in *Arabidopsis*. *Trends in Plant Science* **7**, 106–111. https://doi.org/10.1016/s1360-1385(01)02223-3

Kozaki A and Aoyanagi T (2022) Molecular aspects of seed development controlled by gibberellins and abscisic acids. *International Journal of Molecular Sciences* **23**, 1876. https://doi.org/10.3390/ijms23031876

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

Le BH, Cheng C, Bui AQ, Wagmaister JA, Henry KF, Pelletier J, Kwong L, Belmonte M, Kirkbride R, Horvath S, Drews GN, Fischer RL, Okamuro JK, Harada JJ and Goldberg RB (2010) Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *Proceedings of the National Academy of Sciences USA* **107**, 8063–8070. https://doi.org/10.1073/pnas.1003530107

Lei YT, Zhao YY, Yu F, Li Y and Dou QW (2014) Development and characterization of 53 polymorphic genomic-SSR markers in Siberian wildrye (*Elymus sibiricus* L.). *Conservation Genetics Resources* **6**, 861–864. https://doi.org/10.1007/s12686-014-0225-5

Lei X, Liu W, Zhao J, You M, Xiong C, Xiong Y, Xiong Y, Yu Q, Bai S and Ma X (2020) Comparative physiological and proteomic analysis reveals different involvement of proteins during artificial aging of Siberian wildrye seeds. *Plants (Basel)* **9**, 1370. https://doi.org/10.3390/plants9101370

Li B and Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323. https://doi.org/10.1186/1471-2105-12-323

Li CY, Li WH, Li C, Gaudet DA, Laroche A, Cao LP and Lu ZX (2010) Starch synthesis and programmed cell death during endosperm development in triticale (x Triticosecale Wittmack). *Journal of Integrative Plant Biology* **52**, 602–615. https://doi.org/10.1111/j.1744-7909.2010.00961.x

Li XX, Liu S, Yuan GX, Zhao PC, Yang WG, Jia JT, Cheng LQ, Qi DM, Chen SY and Liu GS (2019) Comparative transcriptome analysis provides insights into the distinct germination in sheepgrass (*Leymus chinensis*) during seed development. *Plant Physiology and Biochemistry* **139**, 446–458. https://doi.org/10.1016/j.plaphy.2019.04.007

Li Q, Xiang C, Xu L, Cui J, Fu S, Chen B, Yang S, Wang P, Xie Y, Wei M and Wang Z (2020) SMRT sequencing of a full-length transcriptome reveals transcript variants involved in C18 unsaturated fatty acid biosynthesis and metabolism pathways at chilling temperature in *Pennisetum giganteum*. *BMC Genomics* **21**, 52. https://doi.org/10.1186/s12864-019-6441-3

Li J, Xie L, Tian X, Liu S, Xu D, Jin H, Song J, Dong Y, Zhao D, Li G, Li Y, Zhang Y, Zhang Y, Xia X, He Z and Cao S (2021) *TaNAC100* acts as an integrator of seed protein and starch synthesis exerting pleiotropic effects on agronomic traits in wheat. *The Plant Journal* **108**, 829–840. https://doi.org/10.1111/tpj.15485

Li Q, Song J, Zhou Y, Chen Y, Zhang L, Pang Y and Zhang B (2022) Full-Length transcriptomics reveals complex molecular mechanism of salt tolerance in *Bromus inermis* L. *Frontiers in Plant Science* **13**, 917338. https://doi.org/10.3389/fpls.2022.917338

Luo G, Shen L, Zhao S, Li R, Song Y, Song S, Yu K, Yang W, Li X, Sun J, Wang Y, Gao C, Liu D and Zhang A (2021) Genome-wide identification of seed storage protein gene regulators in wheat through coexpression analysis. *The Plant Journal* **108**, 1704–1720. https://doi.org/https://doi.org/10.1111/tpj.15538

Maydup ML, Antonietta M, Guiamet JJ, Graciano C, Lopez JR and Tambussi EA (2010) The contribution of ear photosynthesis to grain filling in bread wheat (*Triticum aestivum* L.). *Field Crops Research* **119**, 48–58. https://doi.org/10.1016/j.fcr.2010.06.014

Mortazavi A, Williams BA, McCue K, Schaeffer L and Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**, 621–628. https://doi.org/10.1038/nmeth.1226

Nan S, Zhang L, Hu X, Miao X, Han X and Fu H (2021) Transcriptomic analysis reveals key genes involved in oil and linoleic acid biosynthesis during *Artemisia sphaerocephala* seed development. *International Journal of Molecular Sciences* **22**, 8369. https://doi.org/10.3390/ijms22168369

Peng FY and Weselake RJ (2011) Gene coexpression clusters and putative regulatory elements underlying seed storage reserve accumulation in *Arabidopsis*. *BMC Genomics* **12**, 286. https://doi.org/10.1186/1471-2164-12-286

Pradhan S, Bandhiwal N, Shah N, Kant C, Gaur R and Bhatia S (2014) Global transcriptome analysis of developing chickpea (*Cicer arietinum* L.) seeds. *Frontiers in Plant Science* **5**, 698. https://doi.org/10.3389/fpls.2014.00698

Prathap V and Tyagi A (2020) Correlation between expression and activity of ADP glucose pyrophosphorylase and starch synthase and their role in starch accumulation during grain filling under drought stress in rice. *Plant Physiology and Biochemistry* **157**, 239–243. https://doi.org/10.1016/j.plaphy.2020.10.018

Radchuk V, Tran V, Hilo A, Muszynska A, Gündel A, Wagner S, Fuchs J, Hensel G, Ortleb S, Munz E, Rolletschek H and Borisjuk L (2021) Grain filling in barley relies on developmentally controlled programmed cell

https://doi.org/10.1017/S0960258524000084 Published online by Cambridge University Press

death. *Communications Biology* **4**, 428. https://doi.org/10.1038/s42003-021-01953-1

Riechmann JL and Ratcliffe OJ (2000) A genomic perspective on plant transcription factors. *Current Opinion in Plant Biology* **3**, 423–434. https://doi.org/10.1016/s1369-5266(00)00107-2

Sano N, Lounifi I, Cueff G, Collet B, Clément G, Balzergue S, Huguet S, Valot B, Galland M and Rajjou L (2022) Multi-omics approaches unravel specific features of embryo and endosperm in rice seed germination. *Frontiers in Plant Science* **13**, 867263. https://doi.org/10.3389/fpls.2022.867263

Serrano M, Parra S, Alcaraz LD and Guzmán P (2006) The *ATL* gene family from *Arabidopsis thaliana* and *Oryza sativa* comprises a large number of putative ubiquitin ligases of the RING-H2 type. *Journal of Molecular Evolution* **62**, 434–445. https://doi.org/10.1007/s00239-005-0038-y

Shen ZJ, Xu SX, Huang QY, Li ZY, Xu YD, Lin CS and Huang YJ (2022) TMT proteomics analysis of a pseudocereal crop, quinoa (*Chenopodium quinoa* Willd.), during seed maturation. *Frontiers in Plant Science* **13**, 975073. https://doi.org/10.3389/fpls.2022.975073

Shewry PR, Underwood C, Wan YF, Lovegrove A, Bhandari D, Toole G, Mills ENC, Denyer K and Mitchell RAC (2009) Storage product synthesis and accumulation in developing grains of wheat. *Journal of Cereal Science* **50**, 106–112. https://doi.org/10.1016/j.jcs.2009.03.009

Shimada T, Takagi J, Ichino T, Shirakawa M and Hara-Nishimura I (2018) Plant vacuoles. *Annual Review of Plant Biology* **69**, 123–145. https://doi.org/10.1146/annurev-arplant-042817-040508

Simkin AJ, Faralli M, Ramamoorthy S and Lawson T (2020) Photosynthesis in non-foliar tissues: implications for yield. *The Plant Journal* **101**, 1001–1015. https://doi.org/10.1111/tpj.14633

Smolikova GN and Medvedev SS (2016) Photosynthesis in the seeds of chloroembryophytes. *Russian Journal of Plant Physiology* **63**, 1–12. https://doi.org/10.1134/S1021443715060163

Tang Y, Zeng X, Wang Y, Bai L, Xu Q, Wei Z, Yuan H and Nyima T (2017) Transcriptomics analysis of hulless barley during grain development with a focus on starch biosynthesis. *Functional & Integrative Genomics* **17**, 107–117. https://doi.org/10.1007/s10142-016-0537-5

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ and Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* **28**, 511–U174. https://doi.org/10.1038/nbt.1621

Ullah I, Magdy M, Wang L, Liu M and Li X (2019) Genome-wide identification and evolutionary analysis of TGA transcription factors in soybean. *Scientific Reports* **9**, 11186. https://doi.org/10.1038/s41598-019-47316-z

Utsumi Y, Utsumi C, Tanaka M, Takahashi S, Okamoto Y, Ono M, Nakamura Y and Seki M (2022) Suppressed expression of *starch branching enzyme 1 and 2* increases resistant starch and amylose content and modifies amylopectin structure in cassava. *Plant Molecular Biology* **108**, 413–427. https://doi.org/10.1007/s11103-021-01209-w

Verma S, Attuluri VPS and Robert HS (2022) Transcriptional control of *Arabidopsis* seed development. *Planta* **255**, 90. https://doi.org/10.1007/s00425-022-03870-x

Vinje MA, Walling JG, Henson CA and Duke SH (2019) Comparative gene expression analysis of the β-amylase and hordein gene families in the developing barley grain. *Gene* **693**, 127–136. https://doi.org/10.1016/j.gene.2018.12.041

Wang LK, Feng ZX, Wang X, Wang XW and Zhang XG (2010) DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **26**, 136–138. https://doi.org/10.1093/bioinformatics/btp612

Wang N, Wang R, Wang R and Chen S (2018) Transcriptomics analysis revealing candidate networks and genes for the body size sexual dimorphism of Chinese tongue sole (*Cynoglossus semilaevis*). *Functional & Integrative Genomics* **18**, 327–339. https://doi.org/10.1007/s10142-018-0595-y

Weber H, Borisjuk L and Wobus U (2005) Molecular physiology of legume seed development. *Annual Review of Plant Biology* **56**, 253–279. https://doi.org/10.1146/annurev.arplant.56.032604.144201

Wei X, Jiao G, Lin H, Sheng Z, Shao G, Xie L, Tang S, Xu Q and Hu P (2017) *GRAIN INCOMPLETE FILLING 2* regulates grain filling and starch synthesis during rice caryopsis development. *Journal of Integrative Plant Biology* **59**, 134–153. https://doi.org/10.1111/jipb.12510

Weichert N, Saalbach I, Weichert H, Kohl S, Erban A, Kopka J, Hause B, Varshney A, Sreenivasulu N, Strickert M, Kumlehn J, Weschke W and Weber H (2010) Increasing sucrose uptake capacity of wheat grains stimulates storage protein synthesis. *Plant Physiology* **152**, 698–710. https://doi.org/10.1104/pp.109.150854

Wu XJ, Yue WH, Cai KF, Wang H, Zeng FR and Wang JM (2022) Single-nucleotide polymorphisms in *BMY1* intron III alleles conferring the genotypic variations in β-amylase activity under drought stress between Tibetan wild and cultivated barley. *Agronomy-Basel* **12**, 1737. https://doi.org/10.3390/agronomy12081737

Xie WG, Zhao XH, Zhang JQ, Wang YR and Liu WX (2015) Assessment of genetic diversity of Siberian wild rye (*Elymus sibiricus* L.) germplasms with variation of seed shattering and implication for future genetic improvement. *Biochemical Systematics and Ecology* **58**, 211–218. https://doi.org/10.1016/j.bse.2014.12.006

Xie WG, Zhang JC, Zhao XH, Zhang ZY and Wang YR (2017) Transcriptome profiling of *Elymus sibiricus*, an important forage grass in Qinghai-Tibet plateau, reveals novel insights into candidate genes that potentially connected to seed shattering. *BMC Plant Biology* **17**, 78. https://doi.org/10.1186/s12870-017-1026-2

Xu JJ, Zhang XF and Xue HW (2016) Rice aleurone layer specific *OsNF-YB1* regulates grain filling and endosperm development by interacting with an ERF transcription factor. *Journal of Experimental Botany* **67**, 6399–6411. https://doi.org/10.1093/jxb/erw409

Xue LJ, Zhang JJ and Xue HW (2012) Genome-wide analysis of the complex transcriptional networks of rice developing seeds. *PLoS ONE* **7**, e31081. https://doi.org/10.1371/journal.pone.0031081

Yan J, Bai S, Ma X, Gan Y and Zhang J (2007) Genetic diversity of *Elymus sibiricus* and its breeding in China. *Chinese Bulletin of Botany* **24**, 226–231.

Yao Y, Xiong E, Qu X, Li J, Liu H, Quan L, Lu W, Zhu X, Chen M, Li K, Chen X, Lian Y, Lu W, Zhang D, Zhou X, Chu S and Jiao Y (2023) WGCNA and transcriptome profiling reveal hub genes for key development stage seed size/oil content between wild and cultivated soybean. *BMC Genomics* **24**, 494. https://doi.org/10.1186/s12864-023-09617-6

Yin X, Yi K, Zhao Y, Hu Y, Li X, He T, Liu J and Cui G (2020) Revealing the full-length transcriptome of Caucasian clover rhizome development. *BMC Plant Biology* **20**, 429. https://doi.org/10.1186/s12870-020-02637-4

Yu Q, Xiong Y, Su X, Xiong Y, Dong Z, Zhao J, Shu X, Bai S, Lei X, Yan L and Ma X (2023) Integrating full-length transcriptome and RNA sequencing of Siberian wildrye (*Elymus sibiricus*) to reveal molecular mechanisms in response to drought stress. *Plants (Basel)* **12**, 2719. https://doi.org/10.3390/plants12142719

Zhang ZY, Xie WG, Zhao YQ, Zhang JC, Wang N, Ntakirutimana F, Yan JJ and Wang YR (2019) EST-SSR marker development based on RNA-sequencing of *E. sibiricus* and its application for phylogenetic relationships analysis of seventeen *Elymus* species. *BMC Plant Biology* **19**, 235. https://doi.org/10.1186/s12870-019-1825-8

Zhao XH, Xie WG, Zhang JC, Zhang ZY and Wang YR (2017) Histological characteristics, cell wall hydrolytic enzymes activity and candidate genes expression associated with seed shattering of *Elymus sibiricus* accessions. *Frontiers in Plant Science* **8**, 606. https://doi.org/10.3389/fpls.2017.00606

Zhao YQ, Zhang JC, Zhang ZY and Xie WG (2019) *Elymus nutans* genes for seed shattering and candidate gene-derived EST-SSR markers for germplasm evaluation. *BMC Plant Biology* **19**, 102. https://doi.org/10.1186/s12870-019-1691-4

Zheng Y, Wang N, Zhang Z, Liu W and Xie W (2022) Identification of flowering regulatory networks and hub genes expressed in the leaves of *Elymus sibiricus* L. using comparative transcriptome analysis. *Frontiers in Plant Science* **13**, 877908. https://doi.org/10.3389/fpls.2022.877908