A peer-reviewed version of this preprint was published in PeerJ on 28 August 2019.

View the peer-reviewed version (peerj.com/articles/7528), which is the preferred citable publication unless you specifically need to cite this preprint.

https://doi.org/10.7717/peerj.7528
Comparative transcriptomics reveals the difference in early endosperm development between maize with different amylose contents

Zhou Jian Qu, Tu Shu Xu, Kang Xiao Tian, Ting Li, Cheng Li Wang, Yue Yu Zhong, Quan Ji Xue, Wei Dong Guo

College of Agronomy, Northwest A&F University, Yangling, Shaanxi, China

Corresponding Authors: Quan Ji Xue, Wei Dong Guo
Email address: xjq2934@163.com, gdwei@nwsuaf.edu.cn

The endosperm is a crucial organ for seeds that plays vital roles in supporting embryo development and determining seed weight and quality. Starch is the predominant storage carbohydrate of the endosperm and accounts for ~70% of the mature maize kernel weight. Nonetheless, because starch biosynthesis is a complex process that is orchestrated by multiple enzymes, the gene regulatory networks of starch biosynthesis, particularly amylose and amylopectin biosynthesis, have not been fully elucidated. Here, through high-throughput RNA sequencing, we developed a temporal transcriptome atlas of the endosperms of high-amylose maize and common maize at 5-, 10-, 15- and 20-day after pollination and found that 21,986 genes are involved in the programming of the high-amylose maize and common maize endosperm. A coexpression analysis identified multiple sequentially expressed gene sets that are closely correlated cellular and metabolic programs and provided valuable insight into the dynamic reprogramming of the transcriptome in common and high-amylose maize. In addition, a number of genes and transcription factors (TFs) were found to be strongly linked to starch synthesis, which might help elucidate the key mechanisms and regulatory networks underlying amylose and amylopectin biosynthesis. This study will aid the understanding of the spatiotemporal patterns and genetic regulation of endosperm development in different types of maize.
Comparative transcriptomics reveals the difference in early endosperm development between maize with different amylose contents

Jianzhou Qu\textsuperscript{1,2}, Shutu Xu\textsuperscript{1,2}, Xiaokang Tian\textsuperscript{1,2}, Ting Li\textsuperscript{1,2}, Licheng Wang\textsuperscript{1,2}, Yuyue Zhong\textsuperscript{1,2}, Jiquan Xue\textsuperscript{1,2,*}, Dongwei Guo\textsuperscript{1,2,*}

\textsuperscript{1}The key Laboratory of Biology and Genetics Improvement of Maize in Arid Area of Northwest Region, Ministry of Agriculture and Rural Affairs, College of Agronomy, Northwest A\&F University, Yangling 712100, Shaanxi, China;

\textsuperscript{2}Maize Engineering Technology Research Centre of Shaanxi Province, Yangling 712100, Shaanxi, China

* Corresponding author: xjq2934@163.com (Jiquan Xue), gdwei1973@126.com (Dongwei Guo)
Abstract
The endosperm is a crucial organ for seeds that plays vital roles in supporting embryo development and determining seed weight and quality. Starch is the predominant storage carbohydrate of the endosperm and accounts for ~70% of the mature maize kernel weight. Nonetheless, because starch biosynthesis is a complex process that is orchestrated by multiple enzymes, the gene regulatory networks of starch biosynthesis, particularly amylose and amylopectin biosynthesis, have not been fully elucidated. Here, through high-throughput RNA sequencing, we developed a temporal transcriptome atlas of the endosperms of high-amylose maize and common maize at 5-, 10-, 15- and 20-day after pollination and found that 21,986 genes are involved in the programming of the high-amylose maize and common maize endosperm. A coexpression analysis identified multiple sequentially expressed gene sets that are closely correlated cellular and metabolic programs and provided valuable insight into the dynamic reprogramming of the transcriptome in common and high-amylose maize. In addition, a number of genes and transcription factors (TFs) were found to be strongly linked to starch synthesis, which might help elucidate the key mechanisms and regulatory networks underlying amylose and amylopectin biosynthesis. This study will aid the understanding of the spatiotemporal patterns and genetic regulation of endosperm development in different types of maize.

Introduction
Maize (Zea mays) is one of the most important food and feed crop sources worldwide. The mature maize seed consists of the endosperm, embryo, and seed coat, and the endosperm occupies ~85% of the seed volume and is a critical tissue influencing the maize grain weight (Lopes and Larkins, 1993; Rousseau et al., 2015). The mature endosperm contains large amounts of carbohydrates and proteins that provide nutrients for the developing embryo and seedling, and starch is the main type of carbohydrate in maize endosperm, accounting for ~70% of the endosperm volume. Additionally, maize starch is one of the best-quality starches, with a purity that can reach 99.5%, and 80% of the world’s starch is obtained from maize (Zhang et al., 2013).

The maize endosperm begins to develop through double fertilization, and the process can be classified as a nuclear process. A time-series analysis of maize endosperm development has identified four phases: syncytial, cellularisation, differentiation and death (Berger 1999). The development of the maize endosperm is initiated by successive divisions of the triploid zygotic nucleus without cytokinesis, which creates a coenocyte that contains 128 to 512 nuclei during 0~3 days after pollination (DAP) (Olsen 2004; Leroux et al., 2014; Larkins 2017). Starting at approximately 4~6 DAP, the development of the endosperm from cellularization can include differentiation into four recognizable cell types: basal endosperm transfer layer cell, aleurone cell, embryo-surrounding region cell and central starchy endosperm cell (Olsen 2004). At approximately 8 DAP, subaleurone cells, conducting zone cells, and basal intermediate zone cells emerge and become cytologically identifiable (Larkins 2017). After that time (approximately 8~12 DAP), the endosperm undergoes rapid cell divisions, and the cells in the central portion of the endosperm gradually switch from a mitotic cell cycle to endoreduplication and start accumulating
substantial amounts of starch and storage proteins (Sabelli and Larkins, 2009; Becraft and Gutiérrez-Marcos, 2012). At 20~25 DAP, the cell division enters a late period, and the outer two cell layers of the endosperm develop into the aleurone and the subaleurone (Olsen 2004). During the late period of endosperm development, the maize endosperm undergoes maturation and desiccation, and cell death results in dormancy.

Gene expression is key for revealing the mechanisms underlying the development of the maize endosperm. A series of genetic experiments have identified a number of genes that play important roles in regulating endosperm development, such as maize *globby-1* and H3K27 methyltransferase, because mutations of these genes induce abnormal endosperm cellularization; moreover, the *globby-1* gene also regulates early cell proliferation and differentiation of the endosperm, whereas H3K27 methyltransferase increases coenocytic proliferation (Costa et al., 2003; Gehring and Satyaki, 2017). Additionally, the mutation or deletion of *APETALA2* results in delayed endosperm cellularization and increased seed size, whereas mutation of *AGAMOUS-LIKE62* results in precocious endosperm cellularization and reduced seed size (Kang et al., 2008; Orozco-Arroyo et al., 2015).

The accumulation and synthesis of reserve metabolites and endosperm development are inextricably linked. During endosperm development, aleurone cells form a major nutritive layer that stores proteins, lipids, non-starch carbohydrates, and mineral nutrients. In addition, even if all the endosperm cells undergo programmed cell death during endosperm maturation, the aleurone cells remain alive (Becraft and Gutierrez-Marcos 2012). Genetic studies have revealed that mutation of the naked endosperm (*nkd*) gene not only affects the fate and differentiation of aleurone cells but also alters the accumulation and synthesis of reserve metabolites by regulating widespread gene expression (Gontarek et al. 2016). For example, opaque-2 and prolamin box-binding factor are regulatory genes of naked endosperm1 and naked endosperm2, which play important roles in controlling multiple classes of zeins and show downregulated expression in aleurone cells and nkd mutants (Gontarek et al. 2016). The mutation of opaque-2 results in a starchy endosperm, but high expression of 27-kD γ-zein can convert the starchy endosperm into a vitreous phenotype (Wu and Messing 2010). Furthermore, opaque-2 and prolamin box-binding factor are involved in the regulation of carbon and nitrogen metabolism, pathogen and stress and alter the lysine content during endosperm development (Li et al. 2015; Morton et al. 2016). These factors also indirectly regulate the transcript and protein levels of SSIIa and SBEI in starch synthesis (Zhang et al. 2016).

In higher plants, the starch biosynthesis pathway is orchestrated by multiple enzymes, which mainly include ADP-glucose pyrophosphorylase (AGPase), granule-bound starch synthase (GBSS), soluble starch synthase (SS), starch branching enzyme (SBE), and starch debranching enzyme (DBE). Starch consists of two types of polysaccharides, namely, amylose and amylopectin, and the ratio of amylose to amylopectin determines the suitability of starch for particular end-uses, such as gelatinisation/gelation characteristics, solubility, the formation of
resistant starch, cooking, textural characteristics, and digestibility or nutritional values (Jane et al., 1999; Varavinit et al., 2003; Li et al., 2008; Witt et al., 2010; Zhu et al., 2011). Thus, the amylose content is frequently used to predict the properties of starch, and starch with a high amylose content is optimal for the manufacturing of photodissociative plastics and helps control serious “white pollution” (Jiang et al., 2010; Jiang et al., 2013). Many researchers have reported that the mutation of *amylose extender1* (*ae1*) yields maize kernels with higher amounts of amylose than those found in nonmutant kernels (Fisher et al., 1996; Kim et al., 1998). Additionally, the GBSSI activity of the *wx1* mutant in amylose synthesis is decreased (~5-95%), resulting in a higher level of amylopectin in the maize endosperm and pollen (Sprague et al., 1943; Nelson and Rines, 1962). These two mutants are currently used to create either high- or low-amylose maize and thus obtain altered starch properties and utility in corn breeding.

However, the gene network that links endosperm development with starch synthesis remains unclear. The development of high-throughput technology has resulted in a massive increase in the amount of RNA-Seq data available, and these data are highly reproducible, with few systematic discrepancies among technical replicates (Marioni et al., 2008). Moreover, coexpression analyses based on these large-scale data have been performed for the screening and identification of novel genes that might be involved in metabolic pathways (Saito et al., 2008; Higashi and Saito, 2013). These techniques and methods provide a novel approach for revealing the relationship between metabolism and development. Here, we utilized next-generation sequencing technology to profile the gene expression and function of the endosperms of maize with different amylose contents at 5 DAP, 10 DAP, 15 DAP and 20 DAP and thereby provide valuable insight into the relationship among starch metabolism, early maize endosperm development and other essential metabolisms.

### Materials & Methods

#### Plant materials and sequencing

The common maize (*Zea mays*. L) hybrid Shandan 609 (SD609) and the high-amylose maize hybrid high-amylose 68 (HS68) were grown under field conditions, at approximately 67,500 plants ha⁻¹, in the summer of 2013 in Yangling, Shaanxi Province, China. The ears were bagged prior to silk emergence and then subjected to manual self-pollination. Three biological replicates of the ears were harvested at 5, 10, 15 and 20 DAP. To ensure the integrity and specificity of the endosperm tissues and the absence of contamination by any other tissues, all slightly damaged endosperm tissues were discarded, frozen immediately in liquid nitrogen, and stored at -80°C prior to RNA extraction.

The total RNA of the endosperm from three biological replicates was extracted separately from each sample using the TRIzol-A+ Reagent (Invitrogen) following the manufacturer’s instructions. RNA sequencing libraries were constructed according to the standard Illumina protocol and were sequenced to generate 2×100-nucleotide paired-end reads with the Illumina HiSeq™ 2000 platform (Qu et al. 2016).
Sequence data from this article has been deposited in the National Center for Biotechnology Information Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under accession numbers SRP065059 and SRP149609.

Read mapping and expression analysis

RNA-Seq reads from each sample were aligned to the maize B73 reference genome (ZmB73_RefGen_v2; http://www.maizesequence.org/) with the Tophat2 (v2.09; http://ccb.jhu.edu/software/tophat/index.shtml) programme using Bowtie2 read alignment software (Trapnell et al., 2009; Langmead et al., 2009). The aligned reads obtained after Tophat2 mapping were subjected to transcript assembly using Cufflinks software (version 2.1.0; http://cole-trapnell-lab.github.io/cufflinks). The expression level was calculated in FPKM as described previously (Qu et al. 2016). The differentially expressed genes (DEGs) were identified through pairwise comparison using EBSeq software, and the P-values were adjusted using the Benjamini-Hochberg procedure to determine the false discovery rate (FDR) (Benjamini and Hochberg, 1995; Robinson et al., 2010). Only the results that met the criteria of FDR <0.01 and fold change ≥2 or ≤0.5 in FPKM between the two conditions were deemed to be significantly differentially expressed.

Gene coexpression network analysis

A gene coexpression network analysis was performed using the WGCNA R package (version 1.63; https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/) (Langfelder and Horvath, 2008). We selected all DEGs for the network analysis based on their variance across different samples. We subsequently calculated the Pearson correlation coefficients for each gene-gene comparison and then constructed an adjacency matrix of the connection strengths. The power β was optimized to 6 to adjust the scale-free property of the coexpression network and the sparsity of the connections between genes. Modules were defined as sets of genes with high topologic overlap. An adjacency matrix was used to calculate the topographical overlap matrix (TOM), and a TOM-based dissimilarity measure (1 – TOM) was used for hierarchical clustering. The cutreeDynamic function was used to define the gene modules corresponding to the branches of the resulting dendogram. Highly similar clusters were merged in the network using the mergeCloseModules function with a cutHeight value of 0.25. Each module was assigned a unique number. A NetworkAnalyzer plugin available in Cytoscape was used to calculate the relevant network parameters, such as the degree of connection (Assenov et al., 2008), which measures the number of incoming and outgoing edges of a node. A higher degree of connection indicates that the node has a higher number of edges; thus, this property can be used to identify whether a gene is a major hub of the network, which are genes that might represent major regulating genes in the pathway. The top 5% of genes in each module were identified as hub genes. The modules were visualized using Cytoscape (version 3.6.1; http://www.cytoscape.org/download.php), and the layout was set based on an edge-weighted spring-embedded layout (Shannon et al., 2003).

Statistical analyses
Functional enrichment analyses were performed using the KEGG database (https://www.genome.jp/kegg/) and agriGO (http://bioinfo.cau.edu.cn/agriGO/analysis.php; version 2.0) (Du et al., 2010). A gene ontology (GO) term was considered significantly enriched if the adjusted $p$-value was lower than 0.05 compared with the genome-wide background. The K-means clustering method, Euclidean distance and Calinski-Harabasz index were used to identify the gene expression patterns (Calinski et al., 1974). Additionally, all statistical analyses were performed using the R language (http://www.r-project.org/).

**Real-time quantitative PCR**

Eleven DEGs were closely related to endosperm development and starch metabolite were selected to verify the RNA-Seq results and were thus subjected to quantitative real-time PCR (qRT-PCR) using SYBR Green I (Bio-Rad) and a CFX96 real-time PCR detection system (Bio-Rad, Hercules, USA). The gene-specific primers used in this assay were designed using Primer Premier 5.0 software (Premier, CAN) and are listed in Table S1, and the specificity of the primers was further verified using the non-redundant (Nr) database and Primer-BLAST (GenBank, NCBI). Each sample from each time point was divided into three biological replicates. The quantification was performed with the comparative CT method using Actin (gene ID: GRMZM2G082484) as the internal control. The amplification conditions were 95°C for 15 min, 40 cycles of 95°C for 10 s, 60°C for 20 s and 72°C for 30 s, and this step was followed by a melting curve programme (95°C for 10 s followed by an increase from 65°C to 95°C in 5-s increments of 0.5°C) to confirm the specificity of the PCR amplification. The relative expression levels were calculated as previously described (Livak et al., 2001).

**Isolation of starch and determination of the amylose content**

To eliminate other ingredient distractions, fat was removed using diethyl ether prior to the isolation of starch, and then, 200 g of corn flour was steeped in 1000 mL of distilled water. After the mixture was evenly stirred, the homogenate was passed three times through a stainless-steel sieve (150 μm/0.106 mm). During this step, the homogenate was required to pass through a higher-precision stainless-steel sieve to improve the purity of the isolated starch. The slurries were then poured into 50-mL centrifuge tubes and centrifuged at 2000 rpm for 20 min. This procedure was repeated three times. The supernatant in the centrifuge tube was then discarded, and the sediment was suspended in 40 mL of 0.5% (w/v) sodium hydroxide solution for at least 4 hours. The supernatant was discarded, and 40 mL of new 0.5% (w/v) sodium hydroxide solution was added. After the mixture was evenly stirred, the centrifuge tube was centrifuged at 4000 rpm for 10 min. These steps were repeated three times. Subsequently, 30 mL of absolute ethanol was added to the centrifuge tube, and the tube was centrifuged at 4000 rpm for 10 min. The final sediment was then oven-dried (40°C) or air-dried and ground using a mortar and pestle. The amylose content of the B73 starch was determined from the size-exclusion chromatography weight distributions of the debranched starch, which was calculated as the ratio of the area under the curve of the amylose branches (defined to have degree of polymerization P 100) to the area under the curve of the entire distribution (including both the amylopectin and amylopectin branches) (Fitzgerald et al., 2009).
The amylose content of the sequenced samples (SD609 and HS68) was determined using a two-wavelength iodine binding procedure, which was described in detail previously (Zhu et al., 2008). Two amylose determination methods were used for high-accuracy measurements.

Results

Overview of the RNA-Seq datasets

To obtain global transcriptome maps of the early endosperms of maize with different amylose contents, we performed a RNA-Seq analysis of the common maize SD609 (average amylose content, 27.70%) and the high-amylose maize HS68 (average amylose content, 50.76%) (Fig. 1A). Specifically, the Illumina HiSeq2000 platform was used to perform paired-end RNA-Seq of the endosperm tissues from the two maize plants at 5, 10, 15 and 20 DAP. We generated approximately 0.4 billion high-quality reads, and 95.34% of the mapped reads could be uniquely mapped to the B73 reference genome (Table S2). Additionally, the visualizations of the reads mapped to the maize genome sequences varied across the four stages, indicating that the transcriptome is dynamic during the early stage of maize endosperm development and that the RNA-Seq data were reliable for further analysis (Fig. S1).

Dynamic gene expression patterns and cellular function

To gain insight into the temporal and spatial gene expression patterns during the early stage of maize endosperm development, the gene expression levels at each stage of endosperm development in SD609 and HS68 were classified into five categories based on their FPKM values (Fig. 1B, Table S3). The highest level had the lowest number of expressed genes, i.e., approximately 2,000, whereas the lowest level had more expressed genes than the other levels, i.e., approximately 8,000 (Fig. 1). In particular, the difference in the overall number of expressed genes between SD609 and HS68 was only 326, and the numbers of expressed genes in the endosperm of B73 at 10 and 20 DAP were 22,494 and 22,358, respectively. Approximately 6000 extended (expressed) genes were detected in SD609 and HS68 at 10 DAP (SD609-29,205, HS68-28,421) and 20 DAP (SD609-28,723, HS68-22,358), respectively (Chen et al. 2014).

To further understand the divergence in expressed genes between SD609 and HS68 and reduce the influence of transcription noise, only genes with FPKM values ≥ 1 were used in the estimation of the number of expressed genes. In total, 21,986 expressed genes were found at all four stages of the early endosperms of SD609 and HS68 (Fig. 1C), and a GO enrichment analysis suggested that these common genes were mainly involved in cellular carbohydrate biosynthesis, cellular protein metabolism, cellular nitrogen compound biosynthesis, and other essential processes of endosperm development (Fig. S2). A total of 390 genes were expressed only during specific stages of endosperm development in SD609 and HS68, and most of these genes are related to the response to stress and regulate gene expression and cell development (Table S4). In particular, upstream of the starch metabolism pathway, sugary enhancer 1 (se1; GRMZM5G842214) was only detected at 5 DAP in HS68. However, previous genetic studies have uncovered that se1 is a recessive modifier of su1 that leads to a higher sugar content and an increased kernel moisture for longer post-harvest
periods (Ferguson et al. 1978; Carey et al. 1982). In addition, increased maltose concentrations
have also been detected in mature sel kernels (Ferguson et al. 1979; Tadmor et al. 1995). These
characteristics indicate that sel might be a regulator that affects the ratio of amylose to amylopectin
in the early stage of endosperm development in HS68.

Overview of gene expression patterns

To understand the expression patterns of genes in maize with different amylose contents, pairwise
comparisons were made between SD609 and HS68 at two specific developmental stages. After
genes with read counts of less than 10 at each stage were filtered out, a total of 15,922 DEGs were
identified based on the log₂-fold change (log₂FC) ≥1 or ≤-1 and FDR≤0.05. Of these, 1,193 DEGs
were specifically differentially expressed between SD609 and HS68, 7,812 DEGs were
consistently expressed among the different developmental stages of SD609 and HS68, and 4,905
and 2,000 specific DEGs were detected in SD609 and HS68, respectively (Fig. S3, Table S5). A
functional enrichment analysis indicated that 1,193 DEGs between SD609 and HS68 are mainly
involved in regulating the intracellular signalling cascade, response to hormone stimulus, system
development and fatty acid biosynthetic process (Fig. S4A). In addition, the material-specific
DEGs were also found to regulate the same biological processes, such as the intracellular signalling
cascade, fatty acid biosynthetic process, dephosphorylation, and response to external stimuli (Fig.
S4B, Table S6). Notably, only the material-specific DEGs from HS68 were found to be involved
in the response to water deprivation, and these DEGs in HS68, which can potentially be attributed
to the higher amylose content of this maize species compared with common maize, likely induced
the loose binding of starch grains and ultimately resulted in rapid water dispersal, as observed at
different stages of the development of these kernels compared with those of common maize (Kibar
et al. 2010; Ai and Jane 2018) (Figs. S4C-D). In contrast, the material-specific DEGs of SD609
were enriched in diverse biological processes, such as cell communication, defence response,
nucleosome assembly and response to oxidative stress (Fig. S4D). These results indicate the
existence of obvious metabolic differences between common maize and high-amylose maize in
the early stage of kernel development.

A temporal expression pattern analysis of the shared DEGs between SD609 and HS68 showed that
although the same genes were differentially expressed in two maize materials, their expression
patterns were not synchronized and showed a single or multiple transition points during the
different developing stages. For instance, clusters 6, 11 and 15 exhibited significant differences in
expression transition (Fig. 2). In addition, although some clusters such as clusters 7, 8 and 12
showed synchronous expression fluctuation patterns, these showed only differences in their
expression levels (Fig. 2). The qRT-PCR patterns for all candidate genes were almost equivalent
to the RNA-Seq patterns (Fig. 3, Table S7). A stringent GO term enrichment analysis (P<0.0001,
FDR<0.01) of the shared DEGs revealed that the gene sets of most clusters involved in the
regulation of system development, reproductive process, and multicellular organismal
development (Fig. S5A, Table S6). Additionally, some clusters were highly correlated to specific
biological processes, such as negative regulation of transcription (cluster 5), glucan metabolic
process (cluster 6), regulation of catalytic activity (cluster 11) and nucleosome metabolic process (cluster 16) (Fig. S5A).

We further classified the material-specific DEGs based on the CH index and revealed no unique expression transition pattern for a specific material. A thorough analysis of the functional enrichment of these material-specific DEGs revealed that some unique expression patterns were closely related to key biological processes of endosperm development. For instance, cluster 9 of SD609 and cluster 3 of HS68, both of which are related to cell differentiation, showed the same expression patterns, and clusters 3 and 8 of SD609 and clusters 1 and 10 of HS68, all of which are involved in the regulation of organ development, exhibited the same synchronous expression transition. In addition, cluster 4 of SD609 and cluster 2 of HS68, which are related to negative regulation of biological process, also showed the same expression pattern (Figs. S5-6, Table S6). Interestingly, we found opposite expression patterns for clusters that are relevant to the same biological processes; for example, clusters 3 and 4 of SD609, which are related to negative regulation of biological processes, exhibited the opposite expression patterns (Figs. S5-6). In addition, this finding was also found between materials, such as cluster 9 of SD609 and cluster 3 of HS68, which are involved in the regulation of growth and presented opposite expression transitions (Figs. S5-6). These results indicate that our gene expression pattern sets provide a sufficient dynamic atlas to identify genes that function in key biological processes during early maize endosperm development and provide valuable insights for annotating the function of putative genes.

Identification of weighted gene coexpression network analysis (WGCNA) modules associated with starch synthesis-related enzymes

Starch metabolism is a complex physiological process due to its close connection with sucrose metabolism and regulation by dozens of enzymes (Fig. 4). Additionally, some key enzymes play crucial roles in regulating the synthesis of amylose and amylopectin and might also be crucial factors responsible for the differences in the amylose content between SD609 and HS68. To obtain an in-depth understanding of the potential regulatory mechanism and core gene networks responsible for the differences in the amylose content between SD609 and HS68, a WGCNA was performed using the 15,922 above-described DEGs, which resulted in the identification of 31 WGCNA modules (Fig. 5). We searched the core genes involved in starch synthesis in all the modules based on a previous study (Qu et al. 2018). In total, 19 of the 28 core genes associated with starch synthesis were differentially expressed and were distributed among 10 modules, including Plum (AGPLS1), Plum_1 (SSI1a), Steel_blue (AGPLS4), Blue (SBEIIb and SSIV), Dark_magenta (su1), Dark_orange (SBEI, SBEIII and SSIIIb-b), Dark_violet (SSIIb), Grey_60 (AGPLS2, AGPSS3, PUL and SSI), Maroon (SSIIIb-a) and Orange-red_3 (AGPSS1, GBSSI, GBSSI2 and SSIIIa) (Fig. 6A).

A GO enrichment analysis revealed that genes in each of the 10 modules were not only significantly enriched in specific biological processes but also involved in the regulation of
common pathways. For example, the Orange-red_3 module includes the most critical enzyme, GBSS, which regulates amylose synthesis, and other genes belonging to this module were specifically and significantly enriched in the lipid biosynthetic process, fatty acid biosynthetic process, and cell proliferation (Fig. 6B, Table S8). Therefore, the Orange-red_3 module might play an important role in regulating the balance between cell development and the synthesis of nutrients. The Grey_60 module also showed specific enrichment and appeared to be highly associated with the glycerolipid metabolic process, glycerophospholipid metabolic process and phospholipid metabolic process (Fig. 6B, Table S8). These biological processes and the starch biosynthesis process both utilize fructose-6-P as a substrate for further extension, which suggests that the Grey_60 module likely affects the relationships between sucrose decomposition and starch synthesis through substrate competitive inhibition. Additionally, it should be noted that the ten modules exhibited a high correlation and formed a cascade-based community, and a functional analysis revealed that genes in the ten modules were highly significantly enriched in the reproductive process (Fig. 6B, Fig. S7). This result indicates that the reproduction process might be the most fundamental biological process and could further affect other essential metabolism processes, including the starch synthesis process, by interacting with multiple genes belonging to different modules during the early stages of maize endosperm development.

Identification of hub genes and key networks associated with starch synthesis

Most maize agronomic traits and metabolic processes are complex, arising not from the action of a single gene but from interactions between combinations of genes. In each network module, we found that a sizable portion of the genes exhibited extremely high connectivity with other genes belonging to the ten modules constituting the distribution of the starch synthesis core genes. These highly connected genes are usually considered hub components of the network and have been shown to be biologically important in multiple previous applications. Therefore, we performed a NetworkAnalyzer analysis and found that 5% (918) of the genes from the ten modules had a higher degree of connection and could thus be identified as hub genes for further study (Table S9). A functional enrichment analysis of these hub genes indicated that many of these genes encode proteins functioning in chromatin and protein binding, oxidoreductase activity and transferase activity (transferring hexosyl and glycosyl groups). In addition, these hub genes mainly play an important role in the response to hormone stimulus, regulation of the cellular component size, system development, heterocycle catabolic process, and response to water deprivation (Table S9).

According to the gene classification results, 77 transcription factors of 918 hub genes were identified, and these belonged to distinct families, some of which (AP2/ERF, ARF, bHLH, MYB, WRKY, NAC, bZIP, and GRAS) have been reported to play roles in cell development and the synthesis and accumulation of starch in maize, rice, hulless barley, Arabidopsis, and other plants (Iwase et al., 2011; Licausi et al., 2013; Liu et al., 2015; Huang et al., 2016; Butardo et al., 2017; Cai et al., 2017; Tang et al., 2017; Xiao et al., 2018) (Table S9). In particular, only one core gene of starch synthesis (GRMZM2G158043, ZmPUL) was identified as a hub gene in the Grey_60 module, and this gene is mainly involved in the modification of excessively branched chains or
the removal of improper branches of amylopectin formed by branching enzymes to maintain the cluster structure of amylopectin. More remarkably, beta-glucosidase 18 and aldose 1-epimerase (encoded by GRMZM2G100452 and GRMZM2G031628, respectively) were also found in the Grey_60 module and are mainly related to carbohydrate transport and metabolism (Table S8). These results indicate that some hub genes of the Grey_60 module might play a greater role in regulating the synthesis of amylose and amylopectin. However, this finding raises the question of whether other core genes of starch biosynthesis are key factors that determine the contents of amylose and amylopectin.

To further explore the roles of core genes of the starch synthesis pathway in regulating the contents of amylose and amylopectin, we extracted the coexpression network of 19 differentially expressed core genes of starch synthesis. The analysis revealed that starch biosynthesis is regulated by multiple cascade networks that fragment into more sub-networks, and some of the sub-networks of the core genes of starch synthesis, particularly the sub-networks centred on ZmPUL, were linked by one or more linking genes (Fig. 7). Other disjoint sub-networks might rely on more linking genes as bridges in the regulation of starch synthesis (Fig. 7). Additionally, the relationships of the genes in the community network were also further supported by those of the protein-protein interactions in the STRING database (https://string-db.org/) presented in a previous study (Qu et al. 2018). A rigorous functional enrichment analysis indicated that these linking genes play crucial roles in cell cycle regulation, replication, the MAPK cascade, energy metabolism, binding (including protein binding, nucleotide binding and ion binding), carbohydrate transport and metabolism, lipid transport and metabolic processes, amino acid transport and metabolism, and translation and posttranslational modification (Table S10). These results further indicate that the above biological processes might be intimately linked to starch synthesis, and the specific functions of core genes of starch synthesis might be affected by regulatory genes of these metabolism processes. Nevertheless, the detailed regulatory and interactive mechanisms between core genes of starch synthesis and genes of other key biological processes remain to be elucidated.

**Discussion**

As the major storage carbohydrate in plants, starch has attracted considerable attention as a significant food ingredient and biomaterial. Different contents of amylose and amylopectin in starch give rise to important differences in the properties and functionality of starches obtained from different plants. Nevertheless, only a few studies have examined the differences between corn materials with different amylose contents at the level of transcription. In this study, we performed in-depth transcriptomic surveys based on high-throughput RNA sequencing technology of time-series samples of endosperm tissue from the common maize SD609 and the high-amylose maize HS68. Our results showed that at least 28,378 and 28,144 genes are required to programme the development of the respective maize endosperms. In addition, a comparative analysis revealed that the difference in the overall number of expressed genes between SD609 and HS68 was only 326. These results will facilitate the identification of gene regulatory networks and essential metabolic processes that are important for the programming of endosperm development.
Confirmation of the phases of different processes is important for expounding the endosperm developmental stages. Gene expression regulation occurs at different levels and is one of the key players that offers a powerful atlas for uncovering the complexity of different processes during endosperm developmental stages. An expression pattern analysis of the specific DEGs of SD609 and HS68 in our dataset suggested that most of the temporal patterns are similar and that the only difference was in the expression level (Fig. S6). However, obvious differences in the transition points and expression patterns of the shared DEGs of SD609 and HS68 appeared at 10 and 15 DAP (Fig. 2). This result indicated that some developmental processes might not be synchronous between SD609 and HS68, and the biological functions of some synchronous or asynchronous expression clusters of SD609 and HS68 further support this hypothesis. Additionally, some synchronous expression pattern clusters between SD609 and HS68 mainly function in cell differentiation, multicellular organismal development, nucleosome metabolic processes and glucan metabolic processes (Fig. S5A, Table S6). These results agree with the morphogenetic characteristics of the endosperm. The endosperm grows rapidly from 8 to 12 DAP until it fills the entire seed cavity, and the maize endosperm cells gradually and asynchronously switch from a mitotic to an endoreduplication cell cycle. In addition, genes related to carbohydrate biosynthesis and energy reserve metabolic processes are upregulated at 10 DAP. Cell division continues until approximately 20 to 25 DAP in the external cell layer of the endosperm, and this layer then develops into the aleurone and subaleurone layers (Sabelli and Larkins, 2009; Leroux et al., 2014; Li et al. 2014).

The starch metabolism process is a metabolic process mainly based on energy reserves. In addition, starch biosynthesis is a complex and highly regulated process that requires coordinated activities among multiple enzymes, and different isozymes exert different effects (Jeon et al., 2010; Qu et al. 2018). We analysed core genes of starch synthesis and found that the main differences in the expression patterns of 28 core genes between SD609 and HS68 were observed at 10 and 15 DAP, and the expression transition patterns of some core genes agree with the reported expression patterns (Fig. 8) (Li et al. 2014). For example, GBSSI, a crucial enzyme that regulates the synthesis of amylose, showed higher expression levels at 10 (one step up) and 15 DAP (one step down) in HS68 compared with SD609, and this upregulated pattern agreed with that in maize B73 at 10 DAP and was further verified by qRT-PCR (Figs. 3, 8). In addition, other core genes, such as SBEIIb, AGPLS1 and AGPSS3, also shared a similar temporal pattern (downregulated-downregulated-upregulated), and the expression levels of AGPLS1 and AGPSS3 were higher in HS68 than in SD609. This finding indicated that AGPLS1 and AGPSS3 might be the most efficient partners that interact and polymerize into the native heterotetrameric enzyme structure in the high-amylase maize HS68. Although AGPSS also forms a homotetramer, it also shows defective properties in terms of catalysis and regulation, requires a higher concentration of 3-PGA for its activation and is more easily inhibited by Pi (Kavakli et al. 2001; Salamone et al. 2000). However, the expression level of SBEIIb was higher in SD609 than in HS68 at 5 and 10 DAP. In the starch synthesis process, SBEII contains the two isoforms, SBEIIa and SBEIIb, which play important roles in promoting the production of the short amylopectin chain (degree of polymerization of 6-
12 chains) and further impact the structure and phenotype of amylopectin, but the differences in
the expression of SBEIIb between HS68 and SD609 indicated that SBEIIb might be induced by
increasing the branch frequency and lengthening the glucan chains in clusters to improve the ratio
of amylopectin (Fig. 8) (Nishi et al. 2001; Liu et al., 2009).

We measured the amylose content in various maize materials, which could be ordered HS68
(50.76%) > SD609 (27.70%) > B73 (25.16%) based on a high to a low amylose level (Fig. 1A).
The combination of this information with the expression analysis of the core genes in starch
synthesis revealed that the amylose content of different maize materials might be affected by the
expression levels of certain core genes. Additionally, the synthesis patterns of amylose and
amylopectin might be regulated by a multienzyme complex (Fig. 5). For example, SSI, SSIa,
SBEI and SBEIIb showed the same expression pattern, which is consistent with the results of
earlier studies that showed the following: the SSI and SSII isoforms and either SBEIIa or SBEIIb
form a trimeric complex to regulate the starch metabolic pattern (Liu et al., 2012). Furthermore,
SBEIIb deficiencies have been shown to affect the binding of SSI and SBEI to starch granules in
the amylose extender (ae1.1) mutant, and an experiment involving mutations in the amylose
extender (ae1.2) also showed that SSI, SSIa, SBEI and SBEIIb form another multiple-enzyme
complex (Liu et al., 2009; Abe et al., 2014). In particular, obvious differences in the amylose
content and granule size have been between two ae mutations from two near-isogenic maize lines
(Liu et al., 2012).

In addition to the above factors that affect starch synthesis, we revealed that each differentially
expressed core gene involved in starch synthesis can be classified into a potential interaction
module containing multiple genes (Fig. 7), including members of multiple TF families. For
example, ZmEREB156 (GRMZM2G421033), a member of the AP2/EREBP TF family that
positively modulates the starch biosynthetic gene SSIIIa via the synergistic effect of sucrose and
ABA, exhibited two transition points (upregulated-downregulated-upregulated) in SD609 and
HS68 (Huang et al., 2016). GRAS20 (GRMZM2G023872) belongs to the GRAS TF family and
can reduce fractions of long-branched starch chains, and its overexpression leads to the formation
of a chalky region in the ventral endosperm with a decreased starch content. In addition, compared
with wild-type plants, GRAS20 transgenic seeds exhibit an altered starch granule morphology,
which leads to defective agronomic characteristics (Cai et al., 2017). In our study, GRAS20
showed different expression patterns in the two maize materials with different amylose contents,
i.e., upregulated-downregulated-upregulated (SD609) and upregulated-upregulated-upregulated
(HS68), which suggests that GRAS20 might be involved in the regulation of amylose synthesis.
Moreover, some other members of the TF family have also been identified to play a role in
regulating starch synthesis, but these were not detected in the core gene modules of starch
synthesis. For example, O2, an endosperm-specific bZIP TF that affects the protein levels of SSIIa,
SSIII and SBEI, showed a high expression level and exhibited consistent expression patterns
between the RNA-Seq and qRT–PCR results in our study but was not present in the core gene
modules (Zhang et al., 2016). Because five bZIP TFs were detected in the core gene modules of
starch synthesis, one possible reason to explain the absence of O2 is that it might affect other gene(s) to regulate starch synthesis, which is in line with the results of a previous study (Zhang et al., 2016); alternatively, O2 might have been replaced by other bZIP TF(s) members (Table S11).

Zeins are abundant storage proteins in most cereals and constitute an important factor that affects seed quality. According to their structural differences, zeins are subdivided into four subclasses: α-zeins, which are the major subclass (71–85% of total zeins), followed by γ-zeins (10–20%), and β- and δ-zeins (1–5% each) (Esen, 1987; Xu and Messing, 2008). In our study, we detected 30 zein genes, including 25 α-zeins, one β-zein, three γ-zeins and one δ-zein, and most of these showed a high expression level in SD609 and HS68 at the early stage of endosperm development (Fig. 8).

Compared with the gene expression sets of maize B73, we found obvious differences in the number of expressed zein genes and expression patterns among SD609, HS68 and B73 (Chen et al. 2014). Thus, we speculate that β-, γ-, and δ-zeins and some of the α-zeins are likely essential zeins during early maize endosperm development. Previous studies support the hypothesis that the smallest zein-containing protein bodies, such as β- and γ-zeins, are observed in subaleurone cells and play a role in α- and δ-zein retention in the rough endoplasmic reticulum, whereas α-zeins and some δ-zeins are detected in subaleurone and starchy endosperm cells (Coleman et al., 1996; Bagga et al., 1997; Woo et al., 2001). Additionally, an analysis of the interactions between zeins revealed that 50-, 27-, and 16-kD γ-zeins and 15-kD β-zein exhibit strong interactions, and strong interactions were also detected between α- and δ-zeins and the 16-kD γ-zein and the 15-kD β-zein, respectively (Kim et al. 2002). Moreover, many studies have noted that all zein classes must be present at correct stoichiometric ratios to ensure proper protein body formation and the resulting vitreous endosperm, and specific zein mutations during zein synthesis not only alter the protein body shape and size but also influence the endosperm texture and cause opacity (Guo et al. 2013; Holding 2014; Holding et al. 2007; Kim et al. 2004; Wang et al. 2011; Wang et al. 2012). These results further suggested that the expression differences in α-, β-, γ- and δ-zein-encoding genes are likely to affect the development of the maize endosperm. Interestingly, previous studies revealed that O2 encodes and mainly regulates the expression of α- and β-zein genes by recognizing the O2 box in their promoters. Furthermore, O2-heterodimerizing proteins, O2 and the prolamine box-binding factor are master regulators of zein synthesis by acting in an additive and synergistic mode (Schmidt et al., 1990; Schmidt et al., 1992; Zhang et al., 2015). Additionally, it is worth noting that a recent study revealed that O2 and the prolamine box-binding factor also affect the protein levels of SSIIa, SSIII and SBEI (Zhang et al., 2016), which implies that O2 might be a bridge that connects starch synthesis and zein synthesis or could potentially regulate the levels of other TF(s) or mRNA regulatory factor(s) and thereby affect the functions of the core genes involved in starch synthesis and zein genes. Nonetheless, the detailed regulation mechanism through which zein proteins affects amylose/amylopectin biosynthesis remains to be determined, but this study provides key transcriptional-level clues that will aid the further elucidation of the regulatory relationships between amylose biosynthesis and zein-encoding genes.

In summary, this study offer an essential step to identified gene atlas that are specifically expressed in common maize and high-amylose maize. Temporal patterns of gene expression between
common maize endosperm and high-amylose maize endosperm offer an initial step toward understanding endosperm development and key metabolic alternations. In particular, coexpression network of core genes of starch synthesis provided new clues to explore regulatory factors of starch synthesis, in addition, this part of researches can be integrated with cytology and molecular biology will help to shed light on our understanding of endosperm development and regulation mechanism of starch synthesis.

Additional Information and Declarations

Competing Interests

The authors declare there are no competing interests.

Author contributions statement

Dongwei Guo, Jiquan Xue and Jianzhou Qu conceived and designed the research study. Jianzhou Qu planned and performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables. Shutu Xu, Xiaokang Tian, Ting Li, Licheng Wang and Yuyue Zhong contributed to revising the paper.

Funding

This work was supported by the Shaanxi Province Science and Technology Innovation Coordination Project (2015KTZDNY01-01-01) and the Natural Key Research and Development Program of China (2017YFD0300304).

References


hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res 45, D1040-D1045. doi: 10.1093/nar/gkw982


structures. Carbohydrate Polymers 74, 396-404. doi: 10.1016/j.carbpol.2008.03.012


Orozco-Arroyo G, Paolo D, Ezquer I, Colombo L. 2015. Networks controlling seed size in...


factors O2 and PBF network the regulation of protein and starch synthesis. PNAS 113, 10842-10847. doi: 10.1073/pnas.1613721113


**Figure legends**

**Figure 1** Overview of the amylose content and distribution of the expressed genes. A, Amylose contents of SD609, HS68 and B73. B, Distribution of transcripts at five expression levels based on the FPKM values. The vertical axis represents the number of transcripts, and the horizontal axis represents the developmental stages of the endosperm of SD609 and HS68. C, Venn diagram of the 21,986 genes detected among the four developmental stages of the SD609 and HS68 endosperms.

**Figure 2** Expression patterns of the DEGs in the developing maize endosperm. (A) Twelve clusters were characterized by the fluctuating expression of gene sets at 5, 10, 15 and 20 DAP in SD609. (B) Eleven clusters were characterized by the fluctuating expression of gene sets at 5, 10, 15 and 20 DAP in HS68. The up- and downregulated gene sets are staggered or depicted consecutively during the developmental stages of the maize endosperm. The scaled expression levels of the DEGs are provided on the y-axis, the developmental stages are shown on the x-axis, the coloured lines represent the individual gene expression clusters, and the trend in the expression of each gene set is depicted as a black line. “n” represents the number of DEGs.

**Figure 3** Analysis of the expression patterns of candidate genes by qRT-PCR and RNA-Seq. The y-axis shows the mRNA levels, and the scale on the right indicates the gene expression level based on the RNA-Seq results. The scale on the left shows the relative gene expression levels based on the qRT-PCR results. The x-axis indicates the day of endosperm sampling after pollination. The letters correspond to the genes. The red lines correspond to the RNA-Seq data, and the green bar graph corresponds to the qRT-PCR data.

**Figure 4** Sucrose and starch biosynthetic pathways in non-photosynthetic cells. (A) In non-photosynthetic cells, sucrose is degraded via hydrolysis into hexoses (glucose and fructose) through a reason catalysed by invertase (INV) or reversibly converted into fructose and uridine diphosphate glucose (UDPG) through a reaction catalysed by sucrose synthase (SUS). Fructose is transformed into fructose-6-phosphate (Fructose-6-P) by fructokinase (FRK) and is further metabolized to G-6-P by glucose-6-phosphate isomerase (PGI). Glucose is then transformed into glucose-6-phosphate (G-6-P) by hexokinase (HXK) and translocated into the amyloplasts via phosphate translators. UDPG is further metabolized to glucose-1-phosphate (G-1-P) by the action of UDPG pyrophosphorylase (UGPase). G-1-P, which can also be obtained from the phosphoglucomutase (PGM)-mediated transformation of G-6-P, serves as a precursor for the ADPG pyrophosphorylase (AGPase)-catalysed formation of adenosine diphosphate glucose
Both G-1-P and G-6-P are translocated into the amyloplasts via phosphate translators, whereas ADPG is translocated via ADPG transporters. The content of ADP-glucose is also negatively regulated by Nudix hydrolases (NUDT), which break down ADP-glucose linked to starch biosynthesis, and starch biosynthesis subsequently occurs in the amyloplast. Starch can be chemically classified into two homopolymers: amylose and amylopectin. Amylose is an almost linear α-1,4 glucan molecule synthesized by AGPase and granule-bound starch synthase (GBSS), whereas amylopectin is a highly branched glucan achieved by a coordinated series of enzymatic reactions involving AGPase, soluble starch synthase (SSS), starch branching enzyme (SBE), and starch debranching enzyme (DBE) (Qu et al. 2018). In particular, GBSS is mainly responsible for the synthesis of amylose and long amylopectin chains. Enzymes that catalyse specific reactions are shown in italics. The heat map shows the expression levels of genes encoding the corresponding enzyme, and the scale bar represents the gene expression value (log$_{10}$ (FPKM+1)).

**Figure 5 Modules from weighted gene coexpression network analysis (WGCNA).** Each module was given an arbitrary colour-based name and is summarized by a metric known as the module eigengene (ME), which is the first principal component of the module (i.e., the axis capturing the majority of variation in the expression in the module). Dendrogram and heat map showing the relationships among the modules. The values are based on the linear regression coefficients of determination, which were obtained using the module eigengene values from the WGCNA analysis. Higher regression scores indicate more similar expression patterns between modules or increased predictive power of the main effects on module expression in the dendrogram (closer clustering) and heat map (deeper red).

**Figure 6 Relationship and functional annotation of the modules.** A, Relationships between different modules. Only the correlation values between different modules that are greater than 0.5 are shown. B, Gene ontology enrichment from genes in ten modules. Only biological processes with a P value <0.0001 and FDR<0.01 are shown.

**Figure 7 Starch biosynthesis community networks.** The nodes correspond to genes, and the edges represent coexpression links. Only those with a weight value greater than 0.5 remain. This network includes 2804 DEGs and was subdivided into 19 core gene modules. The core genes of starch synthesis are marked with a red colour and labels, and each core gene module is marked with background colours. The nodes between core gene modules are marked with a pink colour, and other correlated genes are differentiated by different colours.

**Figure 8 Expression map of starch synthesis core genes and zein genes.** Heat map showing the expression patterns of starch synthesis core genes and zein genes in the SD609 and HS68 endosperms at the different developmental stages. The scale bar shows the normalized FPKM values.

**Supplementary figure and table legends**

**Supplementary figure 1 Reads coverage of the SD609 and HS68 endosperms at the four developmental stages.** From outside to inside of Circos image representing reads of 5, 10, 15 and 20 DAP mapped to maize reference genome sequences. Of which, red colour and green colour represents normalized reads number of SD609 and HS68, respectively.
Supplementary figure 2 Functional enrichment analysis of the shared genes among the four stages of endosperm development in SD609 and HS68. (A) The enriched GO terms “biological process”, (B) “cellular component”, and (C) “molecular function” categories are shown.

Supplementary figure 3 Number of DEGs identified in SD609 and HS68. A total of 1,193 DEGs were differentially expressed between SD609 and HS68, 7,813 DEGs were consistently expressed at the different development stages of SD609 and HS68, and 4,914 and 2,002 DEGs were specifically detected in SD609 and HS68, respectively.

Supplementary figure 4 Functional enrichment analysis of DEGs at different developmental stages. (A) Functional annotation of 1,193 DEGs between SD609 and HS68. (B) Shared biological processes regulated by the material-specific DEGs of SD609 and HS68. (C) Specific biological processes of HS68 that are regulated by the material-specific DEGs of SD609 and HS68. (D) Specific biological processes of SD609 that are regulated by the material-specific DEGs of SD609 and HS68. Only biological processes with P values<0.0001 and FDR<0.01 are shown.

Supplementary figure 5 Functional enrichment analysis of the DEG sets of the coexpression clusters. Only biological processes with a P value<0.0001 and FDR<0.01 are shown. (B) Functional analysis of coexpression clusters of SD609. (C) Functional analysis of coexpression clusters of HS68. (A) Functional analysis of shared DEGs in coexpression clusters between SD609 and HS68.

Supplementary figure 6 Expression patterns of shared DEGs between SD609 and HS68. (A) Sixteen clusters were characterized by the fluctuating expression of gene sets at 5, 10, 15 and 20 DAP in both SD609 and HS68. The up- and downregulated gene sets are staggered or depicted consecutively during the developmental stages of the maize endosperm, and the same DEG sets with the same or different expression patterns in two different maize materials are shown. The scaled expression levels of the DEGs are provided on the y-axis, the developmental stages are shown on the x-axis, the coloured lines represent the individual gene expression clusters, and the trend in the expression of each gene set is depicted as a black line. “n” represents the number of DEGs.

Supplementary figure 7 Relationship between module eigengenes. The diagonals show the distribution. The lower left section shows a bivariate scatterplot with a fitting line, and the upper right section shows the correlation coefficient and the significance level.

Supplementary table 1 List of primers used for the qRT-PCR gene detection analysis.
Supplementary table 2 Statistical analysis of the sequenced and mapped reads of SD609 and HS68.
Supplementary table 3 Information on the expressed genes of SD609 and HS68.
Supplementary table 4 Functional enrichment analysis of stage-specific gene expression.
Supplementary table 5 Information of DEGs from SD609 and HS68.
Supplementary table 6 GO enrichment analysis of the DEGs of SD609 and HS68.
Supplementary table 7 qRT-PCR analysis of candidate genes.
Supplementary table 8 Functional annotation of the WGCNA modules.
Supplementary table 9 Functional annotation of the hub genes.
Supplementary table 10 Starch core gene networks.
Figure 1 Overview of the amylose content and distribution of the expressed genes

A, Amylose contents of SD609, HS68 and B73. B, Distribution of transcripts at five expression levels based on the FPKM values. The vertical axis represents the number of transcripts, and the horizontal axis represents the developmental stages of the endosperm of SD609 and HS68. C, Venn diagram of the 21,986 genes detected among the four developmental stages of the SD609 and HS68 endosperms.
Expression patterns of the DEGs in the developing maize endosperm

Figure 2 Expression patterns of the DEGs in the developing maize endosperm. (A) Twelve clusters were characterized by the fluctuating expression of gene sets at 5, 10, 15 and 20 DAP in SD609. (B) Eleven clusters were characterized by the fluctuating expression of gene sets at 5, 10, 15 and 20 DAP in HS68. The up- and downregulated gene sets are staggered or depicted consecutively during the developmental stages of the maize endosperm. The scaled expression levels of the DEGs are provided on the y-axis, the developmental stages are shown on the x-axis, the coloured lines represent the individual gene expression clusters, and the trend in the expression of each gene set is depicted as a black line. “n” represents the number of DEGs.
Cluster 1 (250)
Cluster 2 (335)
Cluster 3 (438)
Cluster 4 (599)
Cluster 5 (441)
Cluster 6 (442)
Cluster 7 (416)
Cluster 8 (722)
Cluster 9 (340)
Cluster 10 (705)
Cluster 11 (603)
Cluster 12 (508)
Cluster 13 (737)
Cluster 14 (654)
Cluster 15 (338)
Cluster 16 (284)

Scaled Expression Level

5 10 15 20 5 10 15 20 5 10 15 20 5 10 15 20 5 10 15 20 5 10 15 20

SD609 HS68 SD609 HS68 SD609 HS68 SD609 HS68

DAP DAP DAP DAP
**Figure 3** Analysis of the expression patterns of candidate genes by qRT-PCR and RNA-Seq

The y-axis shows the mRNA levels, and the scale on the right indicates the gene expression level based on the RNA-Seq results. The scale on the left shows the relative gene expression levels based on the qRT-PCR results. The x-axis indicates the day of endosperm sampling after pollination. The letters correspond to the genes. The red lines correspond to the RNA-Seq data, and the green bar graph corresponds to the qRT-PCR data.
Sucrose and starch biosynthetic pathways in non-photosynthetic cells

**Figure 4 Sucrose and starch biosynthetic pathways in non-photosynthetic cells.** (A) In non-photosynthetic cells, sucrose is degraded via hydrolysis into hexoses (glucose and fructose) through a reaction catalysed by invertase (INV) or reversibly converted into fructose and uridine diphosphate glucose (UDPG) through a reaction catalysed by sucrose synthase (SUS). Fructose is transformed into fructose-6-phosphate (Fructose-6-P) by fructokinase (FRK) and is further metabolized to G-6-P by glucose-6-phosphate isomerase (PGI). Glucose is then transformed into glucose-6-phosphate (G-6-P) by hexokinase (HXK) and translocated into the amyloplasts via phosphate translators. UDPG is further metabolized to glucose-1-phosphate (G-1-P) by the action of UDPG pyrophosphorylase (UGPase). G-1-P, which can also be obtained from the phosphoglucomutase (PGM)-mediated transformation of G-6-P, serves as a precursor for the ADPG pyrophosphorylase (AGPase)-catalysed formation of adenosine diphosphate glucose (ADPG). Both G-1-P and G-6-P are translocated into the amyloplasts via phosphate translators, whereas ADPG is translocated via ADPG transporters. The content of ADP-glucose is also negatively regulated by Nudix hydrolases (NUDT), which break down ADP-glucose linked to starch biosynthesis, and starch biosynthesis subsequently occurs in the amyloplast. Starch can be chemically classified into two homopolymers: amylose and amyllopectin. Amylose is an almost linear α-1,4 glucan molecule synthesized by AGPase and granule-bound starch synthase (GBSS), whereas amyllopectin is a highly branched glucan achieved by a coordinated series of enzymatic reactions involving AGPase, soluble starch synthase (SSS), starch branching enzyme (SBE), and starch debranching enzyme (DBE) (Qu et al. 2018). In particular, GBSS is mainly responsible for the synthesis of amylose and long amyllopectin chains. Enzymes that catalyse specific reactions are shown in italics. The heat map shows the expression levels of genes encoding the corresponding enzyme, and the
scale bar represents the gene expression value (\(\log_{10} (\text{FPKM}+1)\)).
Sucrose $\xrightarrow{\text{SUS}}$ UDP $\xrightarrow{\text{INV}}$ Glucose-6-P

UDP $\xrightarrow{\text{UDPG}}$ Fructose $\xrightarrow{\text{FRK}}$ Glucose-6-P

Fructose $\xrightarrow{\text{HXK}}$ Glucose-6-P

Glucose-6-P $\xrightarrow{\text{PGI}}$ Glucose-1-P

Glucose-1-P $\xrightarrow{\text{PGM}}$ Glucose-1-P

PPi $\xrightarrow{\text{AGPase}}$ ADP-glucose

UTP $\xrightarrow{\text{UGPase}}$ Fructose-6-P

Sucrose $\xrightarrow{\text{SUS}}$ H$_2$O

Amylopectin $\xrightarrow{\text{DBE}}$ Amylose

Starch granule $\xrightarrow{\text{SBE}}$ Amylopectin

Cytoplasm $\xrightarrow{\text{PGI}}$ UTP

AGPase $\xrightarrow{\text{AGPase}}$ ADP-glucose

ATP $\xrightarrow{\text{AGPase}}$ ADP-glucose

PPi $\xrightarrow{\text{AGPase}}$ ADP-glucose

AGPSS1, AGPSS2, AGPSS3, AGPSS4

GBSSI, GBSSII

AGPLS1, AGPLS2, AGPLS3, AGPLS4

NUDT1, NUDT2

PUL, SIIa, SIIb, SIIc, SIIIa, SIIIb-a, SIIIb-b, SIIIc, SIV, SSV

SU1, ISAII, ISAIIa, ISAIIb, ISAIIIa, ISAIIIb

SSIIa, SSIIb, SSIIIa, SSIIIb–a, SSIIIb–b, SSIIIc, SSV

SBEI, SBEIIa, SBEIIb, SBEIII

SU1, ISAII, ISAIIa, ISAIIb, ISAIIIa, ISAIIIb
Modules from weighted gene coexpression network analysis (WGCNA)

Each module was given an arbitrary colour-based name and is summarized by a metric known as the module eigengene (ME), which is the first principal component of the module (i.e., the axis capturing the majority of variation in the expression in the module). Dendrogram and heat map showing the relationships among the modules. The values are based on the linear regression coefficients of determination, which were obtained using the module eigengene values from the WGCNA analysis. Higher regression scores indicate more similar expression patterns between modules or increased predictive power of the main effects on module expression in the dendrogram (closer clustering) and heat map (deeper red).
Relationship and functional annotation of the modules

**Figure 6 Relationship and functional annotation of the modules.** A, Relationships between different modules. Only the correlation values between different modules that are greater than 0.5 are shown. B, Gene ontology enrichment from genes in ten modules. Only biological processes with a P value <0.0001 and FDR<0.01 are shown.
Figure 7 Starch biosynthesis community networks. The nodes correspond to genes, and the edges represent coexpression links. Only those with a weight value greater than 0.5 remain. This network includes 2804 DEGs and was subdivided into 19 core gene modules. The core genes of starch synthesis are marked with a red colour and labels, and each core gene module is marked with background colours. The nodes between core gene modules are marked with a pink colour, and other correlated genes are differentiated by different colours.
**Figure 8** (on next page)

Expression map of starch synthesis core genes and zein genes

**Figure 8 Expression map of starch synthesis core genes and zein genes.** Heat map showing the expression patterns of starch synthesis core genes and zein genes in the SD609 and HS68 endosperms at the different developmental stages. The scale bar shows the normalized FPKM values.