https://doi.org/10.17221/16/2023-CJGPB

Transcriptome profiling of *Sorghum bicolor* reveals cultivar-specific molecular signatures associated with starch and phenolic compounds biosyntheses and accumulation during sorghum grain development

*Li Wang[1], Derang Ni[2,3], Fan Yang[3], Lin Lin[3], Yubo Yang[3], Chongde Sun[4], Xingqian Ye[2], Jinping Cao[5], Xiangli Kong[4]**

[1]*Kweichow Moutai Group, Renhuai, P.R. China*
[2]*College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, P.R. China*
[3]*Kweichow Moutai Corporation Limited, Renhuai, P.R. China*
[4]*College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, P.R. China*
[5]*Rural Development Academy, Zhejiang University, Hangzhou, P.R. China*

**Corresponding author: xlkong@zju.edu.cn*

Citation: Wang L., Ni D.R., Yang F., Lin L., Yang Y.B., Sun C.D., Ye X.Q., Cao J.P., Kong X.L. (2023): Transcriptome profiling of *Sorghum bicolor* reveals cultivar-specific molecular signatures associated with starch and phenolic compounds biosyntheses and accumulation during sorghum grain development. Czech J. Genet. Plant Breed., 59: 235–252.

Abstract: Sorghum is an important crop, and starch and phenolic compounds are major and important components in the sorghum grain. However, the underlying critical genetic elements contributing to the rich portfolio of nutrients in sorghum grains are largely unknown. Transcriptomic methods were employed to characterize the expression patterns at five different grain developmental stages of Hongyingzi (an important brewing sorghum), and another two grain sorghums, Jinuoliang 1 and Hongliangfeng 1, for comparison. The uniquely expressed genes were identified at each developmental stage of Hongyingzi when compared with the other two sorghum cultivars. The co-regulated genes at different developmental stages and the regulatory network were determined; the determinant genes and single-nucleotide polymorphisms located at the promoters of these genes involved in starch and phenolic compounds biosynthetic pathways were also identified. These results will provide insights into the potential regulatory network and further contribute to the clarification of the key determinant genes involved in the biosyntheses of starch and phenolic compounds. Meanwhile, some new transcripts and genes were identified at five different developmental stages of grains of the three sorghum cultivars. Our work can provide impetus for further study of the genes responsible for the biosynthesis of starch and phenolic compounds in the sorghum grain, and pave a way for functional validation of a batch of potential genes and single-nucleotide polymorphisms proposed in current work.

Keywords: biosynthesis; phenolic compounds; sorghum; starch; transcriptome profiling

Abbreviations: AGPPs – ADP-glucose pyrophosphorylases; AS – alternative splicing; DEGs – differentially expressed genes; FDR – false-discovery rate; FPKM – fragments per kilobase of exon per million fragments mapped; GO – gene ontology; GWAS – genome-wide association analysis; HISAT – hierarchical indexing for spliced alignment of transcripts; NGS – next-generation sequencing; PAs – proanthocyanidins; PFK – 6-phosphofructokinase; RLPs – receptor-like proteins; SNPs – single-nucleotide polymorphisms; SPS – sucrose phosphate synthase; WGCNA – weighted correlation network analysis

© The authors. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

https://doi.org/10.17221/16/2023-CJGPB

Sorghum (*Sorghum bicolor* (L.) Moench) is cereal crop that is widely distributed in Africa, the Americas, Asia, Australia and Europe. Sorghum is one of the most important staple crops, and ranks fifth in cereal production behind maize, rice, wheat and barley. It is drought-tolerant, and is an important crop especially for semi-arid regions. Sorghum grain is a rich source of starch, and it also contains a considerable number of phenolic compounds, which are important for its utilization such as in functional foods. In China, sorghum grain is also one of the major raw materials employed for Baijiu (a traditional Chinese liquor) production, and starch and phenolic compounds are observed to determine the quality of Baijiu end-products to a large extent. Therefore, increasing attention has been devoted to study the accumulation of starch and phenolic compounds in sorghum grain, which is of great importance to understand the formation of grain quality (Ke et al. 2022; Zhou et al. 2022).

Starch is the most important component in the endosperm of sorghum grain, and is composed of two polymers of glucose in the form of amylose and amylopectin. Starch synthesis requires the function of a number of vital enzymes, primarily ADP-glucose pyrophosphorylases (AGPPs; *Shrunken2* (*Sh2*) and *Brittle2* (*Bt2*)), starch synthases (starch synthase-I (*SSI*), *Sugary2* (*su2*), starch synthase-IIb (*SSIIb*), *Dull 1* (*du1*) and *Waxy* (*wx1*)) and starch-branching (*Sbe*, starch-branching enzyme 1 (*sbe1*) and amylose extender 1 (*Ae1*)) and debranching (*Dbe*, *Sugary1* (*su1*) and *Pullulanase* (*zpu1*)) enzymes (James et al. 2003; Tetlow et al. 2004). The interactions and isoforms of these enzymes have been explored, further enhancing our understanding of starch synthesis (Tetlow et al. 2008). The genetic diversity and signatures of selection in all the predicted gene models for the storage starch synthesis pathway of *Sorghum bicolor* were studied, and the evolutionary forces acting on some starch synthesis genes were detected (Campbell et al. 2016). Effectively, many genes involved in the primary starch synthesis pathway showed a clear reduction in nucleotide diversity, and different evolutionary rates were identified upstream and downstream within the metabolic pathways (Campbell et al. 2016). Although notable progress has been achieved in understanding the genetic aspects of starch synthesis (Campbell et al. 2016; Boyles et al. 2017; Hu et al. 2019), the determinants of the yield and quality of starch and the regulatory network of the enzymes involved in starch metabolism remain unclear. Recently, the natural variation of sorghum grain quality traits was studied by genome-wide association analysis, and the underlying genetic loci were proposed (Chen et al. 2019; Kimani et al. 2020). The above-mentioned studies determined some correlations between nucleotide polymorphisms and starch yield. However, the biochemical pathway of storage starch synthesis regulation at the transcriptional level regarding all the reserve products accumulating in the developing endosperm has scarcely been determined. Transcriptomics studies using next-generation sequencing (NGS) provide an opportunity to complement these conclusions with studies conducted at the genome level.

Phenolics contribute to the varied applications of sorghum grains (de Morais Cardoso et al. 2017). Sorghum grain is a favoured raw material for liquor-making in China. The unique flavours of Baijiu were found to be largely contributed by the biological transformation of phenolic compounds, which highly accumulate in sorghum grains. The phenolic profiles of different sorghum genotypes are distinguishable (Shen et al. 2018; Xiong et al. 2019), but the underlying genetic mechanisms remain to be clarified. The phenolic compounds in sorghum are mainly composed of phenolic acids, 3-deoxyanthocyanidins and condensed tannins. The condensed tannins, also known as proanthocyanidins (PAs), are oligomers and polymers of flavan-3-ols. Combining quantitative trait locus mapping, meta-quantitative trait locus fine-mapping and association mapping, the nucleotide polymorphisms in the *Tannin1* (*Tan1*) gene, encoding a WD40 protein, were demonstrated to control the tannin biosynthesis in sorghum (Wu et al. 2012), and more alleles of *Tan1* were identified in another study (Wu et al. 2019). More work demonstrates that, on chromosome four, a hotspot region involving major effect markers linked with putative MYB–bHLH–WD40 complex genes involved in the biosynthesis of the polyphenol class of flavonoids was observed (Habyarimana et al. 2019), and SNPs located on *Tan1* and orthologs of *Zm1* and TRANSPARENT TESTA 16 (*TT16*) were also correlated with tannin content (Kimani et al. 2020). All these were proposed from genome-wide analyses but lack validation at the transcriptional level. Further studies on the expression profiles of these genes and extensions to the interactions of these genes encoding enzymes are needed.

Recently, transcriptomic analyses on broomcorn millet, chestnut and taro were conducted to explore

https://doi.org/10.17221/16/2023-CJGPB

the complex molecular mechanisms of quality formation by collecting the specimens at different developmental stages (Dong et al. 2021; Shi et al. 2021; Yang et al. 2023). The current work, aims to build comprehensive transcriptomes and address the question of how the synthetic pathways for starch and secondary metabolites, particularly phenolic compounds, are regulated at the transcriptional level in sorghum grain. Three sorghum cultivars at five developmental stages were used to identify the patterns of expression and determinants of these phenotypes (Figure 1) by NGS platforms with high throughput RNA-Sequencing (RNA-Seq) analysis approaches. Comparisons of the different stages of development and different sorghum genotypes can reveal that how these variations at the transcriptional level are correlated with starch and phenolic compounds metabolic pathways. These large transcriptomic resources will also provide valuable information for further studies on other phenotypes in these three sorghum genotypes and in related species, and finally contribute to genetic improvement and precise breeding.

MATERIAL AND METHODS

Plant materials, and measurements of starch and phenolic compounds. The three sorghum cultivars (i.e., X28, sorghum Hongyingzi; X35, sorghum Jinuoliang 1; and X38, sorghum Hongliangfeng 1) were planted in the agricultural experimental field of Zhejiang University in early April 2018; these three grain sorghum cultivars, which belong to the species of *Sorghum bicolor* (L.) Moench, were locally cultivated varieties originally from Guizhou Province, China. Seeds of the three sorghum cultivars were provided by the local institute of agricultural sciences in Guizhou Province. The standard agricultural guidelines were followed for optimum sorghum growth and development including irrigation, fertilizer application and pest control. All the grain samples of these three sorghum cultivars, collected at five developmental stages (S1 – early milky stage; S2 – middle to late milky stage; S3 – early to middle dough stage; S4 – late dough stage; S5 – maturation stage), were dehulled and ground. The starch content was determined according to the Ewers' method based on acid hydrolysis (Mitchell 1990). The modified method from a previous study was used to extract the free phenolics and bound phenolics (Adom & Liu 2002). The total phenolic contents (TPCs) were measured using the Folin–Ciocalteu assay adapted to 96-well microplates (Singleton et al. 1999; Margraf et al. 2015). A Waters system with a UV-vis detector was employed to detect the contents of both extractable and bound phenolic compounds. Chromatographic separation was performed using a reverse-phase column (ZORBAX SB-C18, 250 × 4.6 mm id., 5 µm) and a Waters UPLC (Waters Corp., Milford, USA).

RNA preparation and sequencing. The three sorghum cultivars were sampled at the S1–S5 developmental stages with triplicates for each sampling

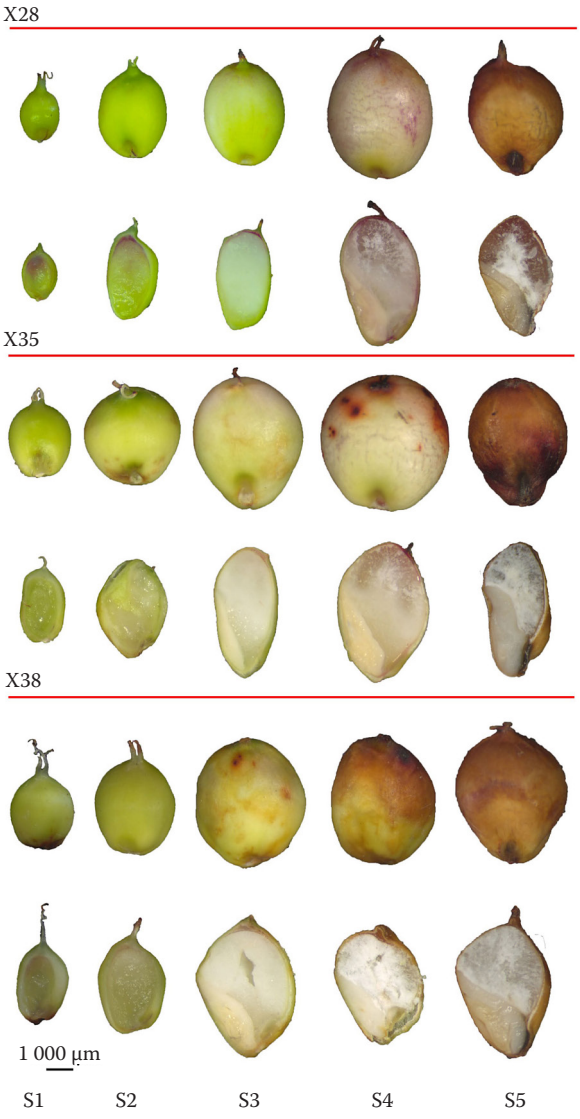

Figure 1. Morphological properties of three sorghum grains at five different developmental stages

S1 – early milky stage; S2 – middle to late milky stage; S3 – early to middle dough stage; S4 – late dough stage; S5 – maturation stage; X28 – Hongyingzi; X35 – Jinuoliang 1; X38 – Hongliangfeng 1

https://doi.org/10.17221/16/2023-CJGPB

point. In total, 45 samples were subjected to the transcriptomic study, and measurements of starch and phenolic compounds. After collection, the samples were stored in a −80 °C freezer until further processing. The total RNA was extracted separately from the frozen samples using the modified cetyl trimethyl ammonium bromide (CTAB) method with three biological replicates (Yin et al. 2008). The yield and quality of the RNA were determined using gel electrophoresis and a Nanodrop 2000 spectrophotometer (Thermo Scientific, Waltham, USA). mRNA was enriched using OligodT ligated with magnetic beads through the polyA tail on mRNA. Library construction, RNA sequencing and bioinformatic analyses were performed by the Beijing Genome Institute (BGI; Shenzhen, China). The libraries were sequenced using the BGISEQ-500 sequencing platform.

Bioinformatic analyses. Prior to mapping and assembly, the total reads were filtered to remove those with adapters, of low quality (mean Q < 10), with a number of N base calls (over 5% of the total nucleotide number), incomplete reads (< 50 bp), or repetitive reads with mutual information scores > 0.5, using Trimmomatic/0.36-Java-1.8.0_121 (Bolger et al. 2014). Both reads in a pair were removed if either one did not pass through the filters. On average, < 1% of all the reads failed to pass the filter. SOAPnuke was used to perform statistics on the results of the quality-control process (Chen et al. 2018). On average, across the libraries, 89% of the reads mapped uniquely to the reference genome. The published sorghum reference genome (*Sorghum_bicolor*, GCF_000003195.3_*Sorghum_bicolor*_NCBIv3) was retrieved from GenBank. Clean reads were then mapped onto the reference genome using the Hierarchical Indexing for Spliced Alignment of Transcripts (HISAT2) (Kim et al. 2015) or Bowtie2 (Langmead & Salzberg 2012) against the reference genes with default parameters. HISAT was developed using Burrows–Wheeler transform and Ferragina–Manzini (FM) to index at the whole genome scale and local genome regions to improve the accuracy of the alignments. RNA-Seq by expectation maximization (RSEM) was performed to calculate the abundance of reads mapped on genes or transcripts (Li & Dewey 2011). The fragments per kilobase of exon per million fragments mapped (FPKM) were used to evaluate the expression levels of genes or transcripts and followed the cutoffs: FPKM ≥ 0.5 and gene-level read counts ≥ 10, for differential gene expression analysis. To investigate the correlations of expression among samples, the Pearson coefficients were calculated based on the expression of all genes using *princomp* in R. The results were plotted as heatmaps to present similarities in expression patterns between two samples using *ggplot2* in the R platform.

Transcripts were reconstructed for each sample using StringTie with parameters (-f 0.3 -j 3 -c 5 -g 100 -s 10000 -p 8) (Pertea et al. 2015). Cuffmerge, from Cufflinks 2.0.2, was used to merge the transcripts from all the samples, and Cuffcompare was used to compare these merged transcripts with referenced annotations (Trapnell et al. 2012). The transcripts assigned to class code "u" (unknown, located on intergenic regions), "i" (located on introns of known genes), "o" (transcripts overlapped with exons of known genes) and "j" (new potential transcripts or fragments including at least one junction site consistent with the reference gene) were defined as new transcripts. CPC v0.9-r2 was used to predict the protein-coding potential of transcripts (Kong et al. 2007), and these transcripts were predicted and merged into the reference gene dataset for following studies.

The annotations on the transcripts included transcription factors (TFs) and plant resistance genes. For the TFs, the getorf of EMBOSS 6.5.7.0 (Rice et al. 2000) was used to detect the ORF of the UniGene (-minsize 150), and hmmsearch v3.0 (Eddy 1998) was employed with the default settings against the protein domain database of TFs. Plant TFDB 5.0 (http://planttfdb.cbi.pku.edu.cn) was referenced to assign the family of TFs. DIAMOND v0.8.31 (Buchfink et al. 2015) was conducted against (-evalue 1e-5 -outfmt 6-max-target-seqs 1-more-sensitive) the Plant Resistance Gene Database (PRGdb) (v2.0, http://prgdb.crg.eu/) (Sanseverino et al. 2010). The best hits were filtered with criteria of query coverage ≥ 50% and identity ≥ 40%. All the predicted genes from the transcripts were also functionally annotated using Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). The classifications of the GO and metabolic pathways from KEGG were assigned according to the classification systems from both databases. For the differentially expressed genes (DEGs) in each pair of comparisons, *phyper* was used to conduct gene set enrichment analyses in R, and the *P*-value was corrected using the FDR. *q*-values ≤ 0.05 were considered significant.

DESeq2 was employed to identify the DEGs among collected samples (Wang et al. 2010). Fisher's exact test and the likelihood ratio test were proposed to identify DEGs under the assumption that the

https://doi.org/10.17221/16/2023-CJGPB

number of reads coming from a gene (or transcript isoform) followed a binomial distribution (and could be approximated by a Poisson distribution) (Marioni et al. 2008; Bloom et al. 2009). A method based on the MA-plot was used to test the hypothesis and calculate *P*-values for each gene, which were adjusted to *q*-values for multiple testing corrections by the Benjamini–Hochberg and Storey–Tibshirani methods. As suggested by the tool, either a *P*-value or a false-discovery rate (FDR) threshold was adopted to identify differentially expressed genes.

Mfuzz v2.34.0, a tool developed from soft clustering, was used in the present work with parameters (-c 12, -m 1.25) to cluster genes that shared similar functions (Kumar & Futschik 2007). This benefited the discovery of regulatory genes, as genes showing similar expression patterns were generally responsible for the same bioprocess. In the present study, genes which showed similar expression patterns at different time points, were clustered together. To find clusters (modules) of correlated genes, such clusters were summarized, modules were related to one another and to external sample trains, and module membership measures were calculated. Weighted correlation network analysis (WGCNA) was used to conduct the gene co-expression network analysis with the parameters -GeneFrac and Threshold 0.5 (Langfelder & Horvath 2008).

rMATS, a statistical method for the robust and flexible detection of differential alternative splicing (AS) from replicated RNA-Seq data, was used to detect AS (Shen et al. 2014). Five types of AS were proposed to be detected using rMATS, including skipped exon (SE), alternative 5' splice sites (A5SSs), alternative 3' splice sites (A3SSs), mutually exclusive exons (MXEs) and retained introns (RIs). A hierarchical model to simultaneously account for the sampling uncertainty in individual replicates and variability among replicates was adopted by rMATS. The hypothesis-testing framework of rMATS included quantifying the expression abundance of AS in different samples, likelihood-ratio tests for *P*-values between pairs of samples on the IncLevel (inclusion level) and the Benjamini–Hochberg method to adjust *P*-values to FDRs. The Genomic Analysis ToolKit was used to detect SNPs and InDels in each sample after mapping onto the reference genome (McKenna et al. 2010). The HISAT2 was used firstly to map reads against the reference genome, and SNPs and InDels were filtered using HaplotypeCaller.

R and Python scripts were compiled for the statistics and plots performed in the present study, if not specified.

The RNA-Seq datasets analyzed in the present study were deposited in the NCBI Sequence Read Archive (SRA) database.

Statistical analysis. The data of starch content and total phenolics content of sorghum grain samples at different developmental stages were analyzed and plotted using Origin Software version 2022 (OriginLab, Northampton, USA). The data were expressed as means of three biological replicates, and the error bars in the graph represent standard deviations.

RESULTS AND DISCUSSION

Changes in starch and phenolic compounds during sorghum grain development. The development of sorghum grains was divided into five stages (Figure 1). The dynamic changes in the starch and total phenolic contents at the five developmental stages are shown in Figure 2. The starch content increased continuously during the whole developmental period and showed the highest value at S5 for all the three sorghum varieties. However, the highest value of total phenolic content was observed to be at S4. At the final stage (S5), the Hongyingzi sorghum grains showed the highest starch content, followed by Hongliangfeng 1 and Jinuoliang 1. The free phenolic contents were much higher than those of bound phenolics, the total phenolic contents showed a decrease in the maturing stage, and the Hongyingzi sorghum grains exhibited higher content of total phenolics than Jinuoliang 1 and Hongliangfeng 1.

General features of transcriptomic data. Triplicate analyses were conducted for five developmental stages of the three sorghum cultivars. In total, 45 samples were sequenced using the BGISEQ-500. On average, 6.97 Gb of data were generated for every sample. Correlation analysis was performed among 15 samples using the FPKM method, and most of the Pearson correlation coefficients (*r*) among the three replicates for each developmental stage were greater than 0.96, indicating that samples from each stage were available. Principal coordinate analysis also indicated the correlations of the replicates for each sampling point (Figure 3). The saturation of sequencing demonstrated that the sequencing depth was sufficient for all the samples; with an increase in the number of reads, no further transcripts would be identified.

With *Sorghum_bicolor* (GCF_000003195.3_Sorghum_bicolor_NCBIv3) serving as the reference

https://doi.org/10.17221/16/2023-CJGPB

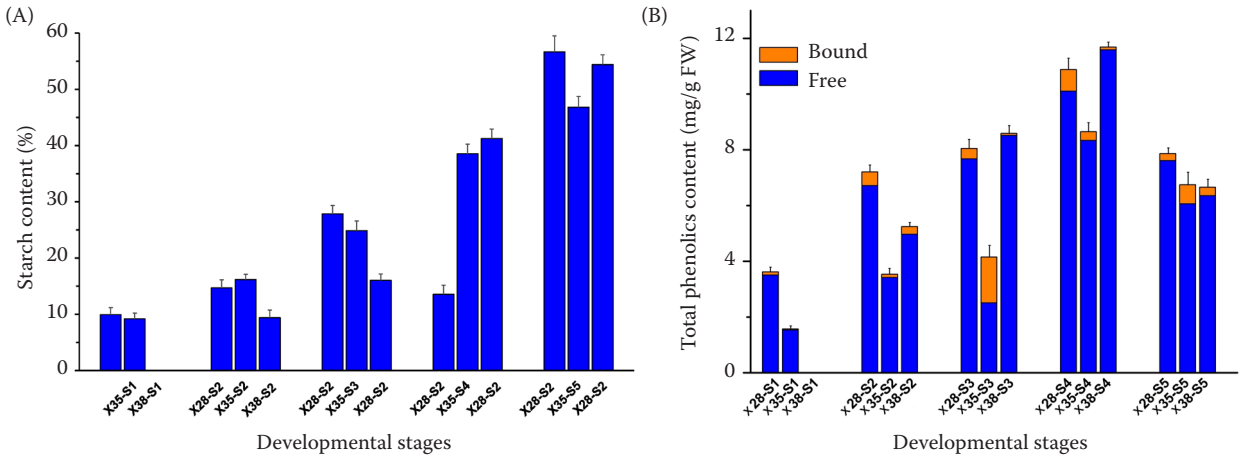

Figure 2. Contents of starch (A) and phenolics (B) at different stages of three sorghum cultivars
S1 – early milky stage; S2 – middle to late milky stage; S3 – early to middle dough stage; S4 – late dough stage; S5 – maturation stage; X28 – Hongyingzi; X35 – Jinuoliang 1; X38 – Hongliangfeng 1; the error bars indicate standard deviations

genome, the average ratio of aligned reads was 86.0%, and the average ratio of reads aligned against the gene database was 78.43%.

After predicting the genes for all the transcripts, a total of 31 503 UniGenes were identified, of which 29 765 were known, and 1 738 had not been reported before. In total, 24 726 new transcripts were detected, 13 238 of which were assigned to alternative-splicing isoforms of known annotated protein-encoding genes, 1 736 were assigned to transcripts possessing unknown protein-encoding genes, and the remaining 9 725 were assigned to long non-encoding RNAs.

A total of 1 590 transcription factors were identified, including MYB, bHLH, AP2-EREBP and MADS (Figure 4), which had been reported to be the major transcription factor families at the maturing stage of grain development. For instance, the regulatory gene *Zm1*, belonging to the Myb family, plays a critical role in the biosyntheses of tannins (Kimani et al. 2020). Other important transcription-factor-encoding genes, such as *myb*, are involved in regulating some important phenotypes (Scully et al. 2016).

The plant resistance genes were also identified in the transcripts in the present study. Among them,

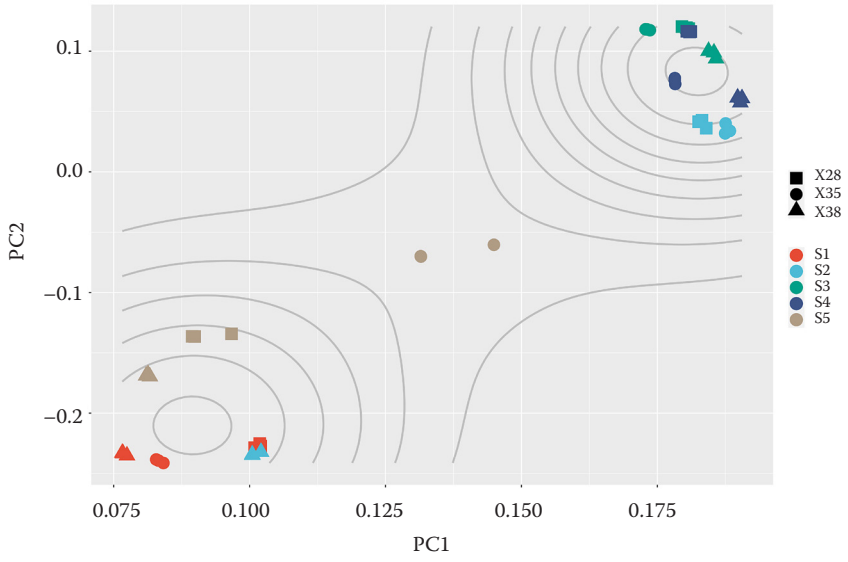

Figure 3. Principal coordinate analysis (PCoA) of gene expression at the different developmental stages of sorghum
S1 – early milky stage; S2 – middle to late milky stage; S3 – early to middle dough stage; S4 – late dough stage; S5 – maturation stage; X28 – Hongyingzi; X35 – Jinuoliang 1; X38 – Hongliangfeng 1

https://doi.org/10.17221/16/2023-CJGPB

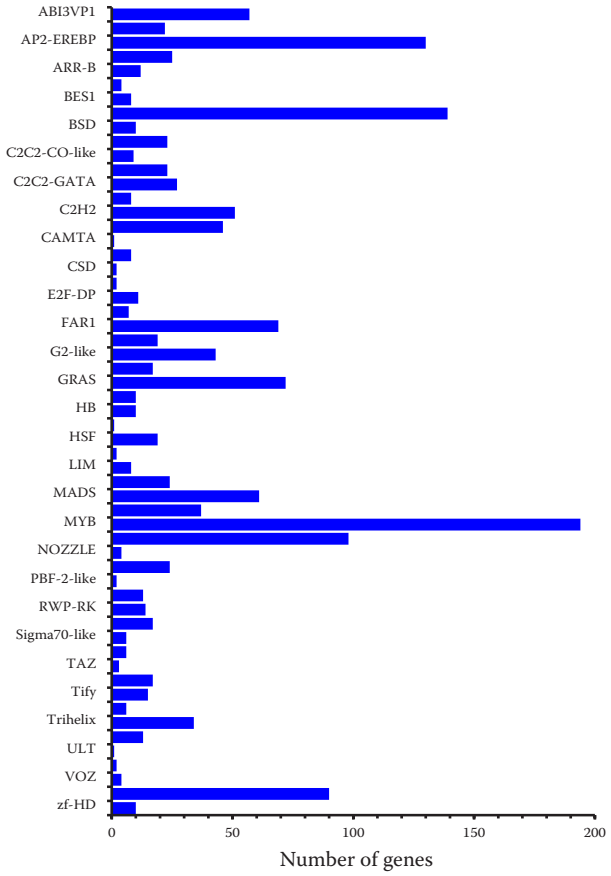

Figure 4. The classification of transcription regulators identified in all the transcripts of Hongyingzi (X28), Jinuoliang 1 (X35) and Hongliangfeng 1 (X38) at the five developmental stages

RLPs (receptor-like proteins), NL (including the NBS domain) and CNL (containing one central nucleotide-binding (NB) subdomain) were the most abundant (Figure 4).

Transcriptional profiles at the five developmental stages of three sorghum varieties. The numbers of differentially expressed genes (DEGs) between each pair of the typical sorghum variety (Hongyingzi) at different developmental stages are shown in Figure 5. Compared with the other two sorghum cultivars (i.e., Jinuoliang 1 and Hongliangfeng 1) at different developmental stages, higher numbers of DEGs were observed at different stage comparisons for Hongyingzi (Figure S1 in Electronic Supplementary Material (ESM)). Except for comparisons between the two sorghum cultivars and Hongyingzi at stage S1, the numbers of down-regulated DEGs were higher than those of up-regulated DEGs in all the comparisons. At stages S3, S4 and S5, totally of 7 409, 9 221 and 9 214 genes were down-regulated, respectively, compared with S1 for Hongyingzi. By contrast, the number of DEGs between cultivars at the same developmental stage was lower. At stage S5, the up-regulated and down-regulated DEGs between Hongyingzi and Hongliangfeng 1 were 1 026 and 1 892, respectively. Another prominent difference in current analysis was that the largest number of DEGs, 5 301 up-regulated genes, was observed at the stage S5 between Hongyingzi and Hongliangfeng 1, and 3 997 were observed between

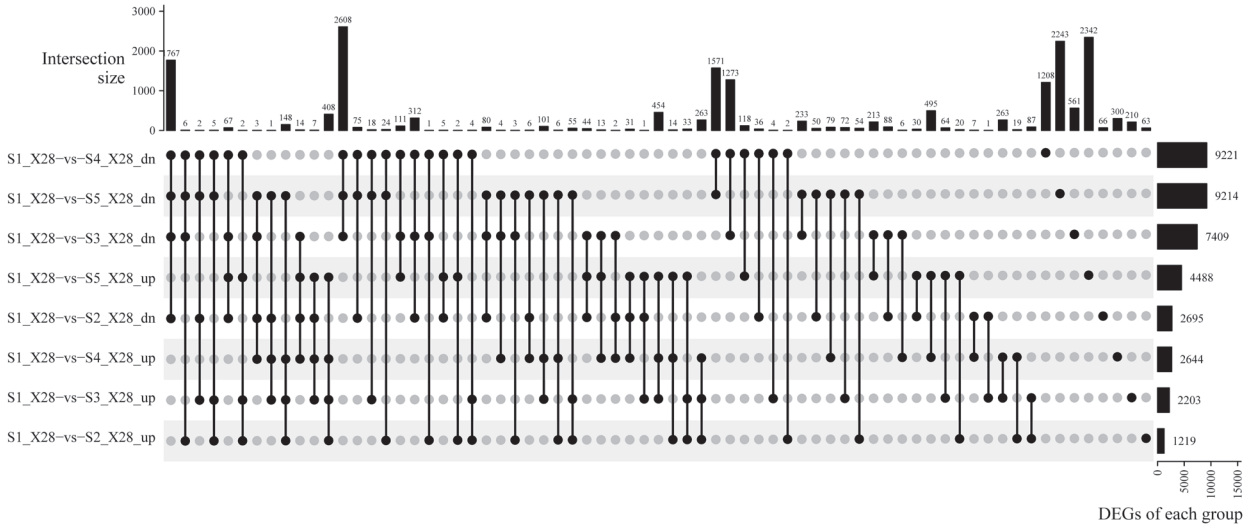

Figure 5. Comparison of differentially expressed genes (DEGs) between each pair of stages for sorghum cultivar X28 (Hongyingzi) S1 – early milky stage; S2 – middle to late milky stage; S3 – early to middle dough stage; S4 – late dough stage; S5 – maturation stage; connected solid black circles under the vertical bars represent specific pairs are involved in each intersection, while unconnected circles represent pairs observed exclusively in the corresponding database

https://doi.org/10.17221/16/2023-CJGPB

Hongyingzi and Jinuoliang 1. At S5, 6 066 DEGs had higher expression levels in Jinuoliang 1 than in Hongyingzi, and at stage S4, 4 901 and 4 613 DEGs were expressed more highly in Jinuoliang 1 and Hongliangfeng 1 than in Hongyingzi. The DEGs observed in this analysis indicate that the differentiations in starch and phenolic compounds between Hongyingzi and the other two grain sorghum cultivars were possibly attributable to these DEGs.

To screen the key genes responsible for the biosyntheses of starch and phenolic compounds in sorghum, a weighted correlation network analysis (WGCNA) was performed to determine the gene interaction network and regulatory network at each developmental stage between different sorghum cultivars at the system biology level. WGCNA discriminated 1 280 DEGs into 18 groups (Figure 6). Among these genes, the highest correlations were observed between 124 genes expressed at the S1 stage of sorghum Hongliangfeng 1 (Pearson correlation: 0.52~0.53), 46 genes at the S5 stage of Jinuoliang 1 (Pearson correlation: 0.50~0.53) and 42 genes at the S5 stage of Hongyingzi (Pearson correlation: 0.54~0.55). It is significant to recognize those major regulatory genes and the related network that contribute to the different developmental stages through identifying the co-expression patterns of these genes. For instance, *aceA* (LOC8084064, isocitrate lyase), co-regulated at the S5 stage of sorghum Hongliangfeng 1, was observed to show increasing expression from S1 to S5 in Hongyingzi, different from the high expression only at the S5 stage of Hongliangfeng 1. Further analyses demonstrate three unique SNPs (i.e., 2:69 490 251; 2:69 490 664; and 2:69 490 684). The reference genome is G, G- > C or deletion in these three cultivars were observed in Hongyingzi compared with sorghum grains Jinuoliang 1 and Hongliangfeng 1. Further studies on these grouped gene clusters may reveal insights into correlating phenotypes with genotypes.

The unique expression genes at each developmental stage of Hongyingzi. At the five stages, from milky to maturity, the number of characterized expression genes ranged from 159 to 354 (Figure 7). By contrast, the number of co-regulated genes at all stages was 24 793.

The DEGs at the five developmental stages for Hongyingzi were further investigated, and those showing similar expression patterns were clustered. Referring to time series analyses, 12 gene clusters were identified with similar and prominent expression patterns. For instance, 4 850 genes were assigned to Cluster 2. Genes in Cluster 2 were observed with higher expression levels at the stage of S1 and decreased in the following stages from S2 to S5. Stress

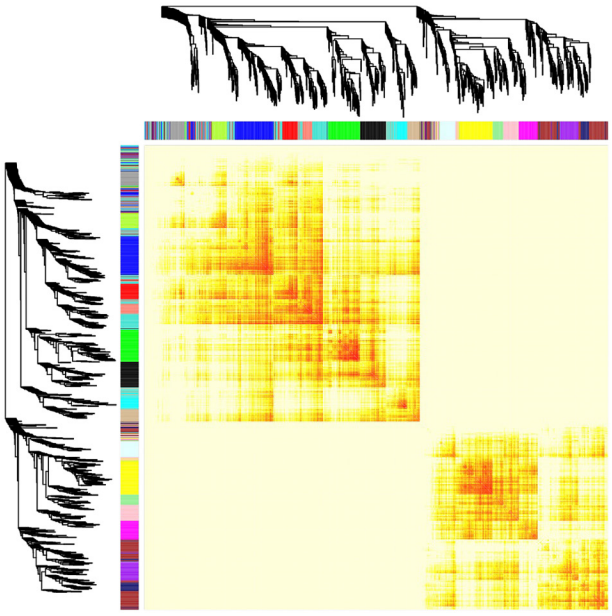

Figure 6. The weighted correlation network analysis (WGCNA) of differentially expressed genes (DEGs) at all stages of the three sorghum cultivars

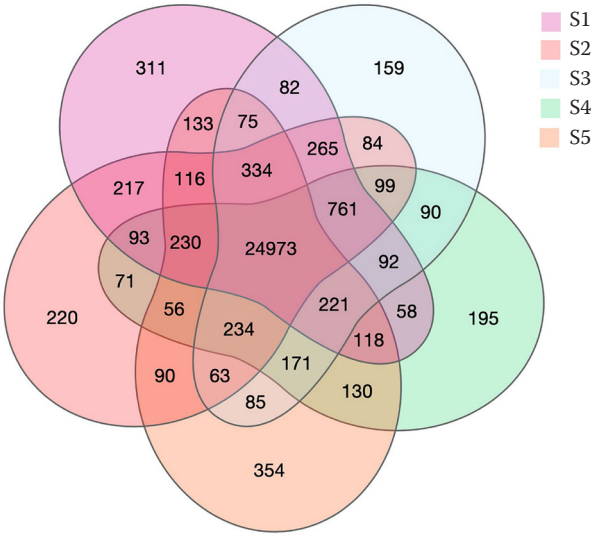

Figure 7. Overlap of differentially expressed genes (DEGs) at the developmental stages of S1–S5 (S1 – early milky stage; S2 – middle to late milky stage; S3 – early to middle dough stage; S4 – late dough stage; S5 – maturation stage) for sorghum Hongyingzi (X28)

https://doi.org/10.17221/16/2023-CJGPB

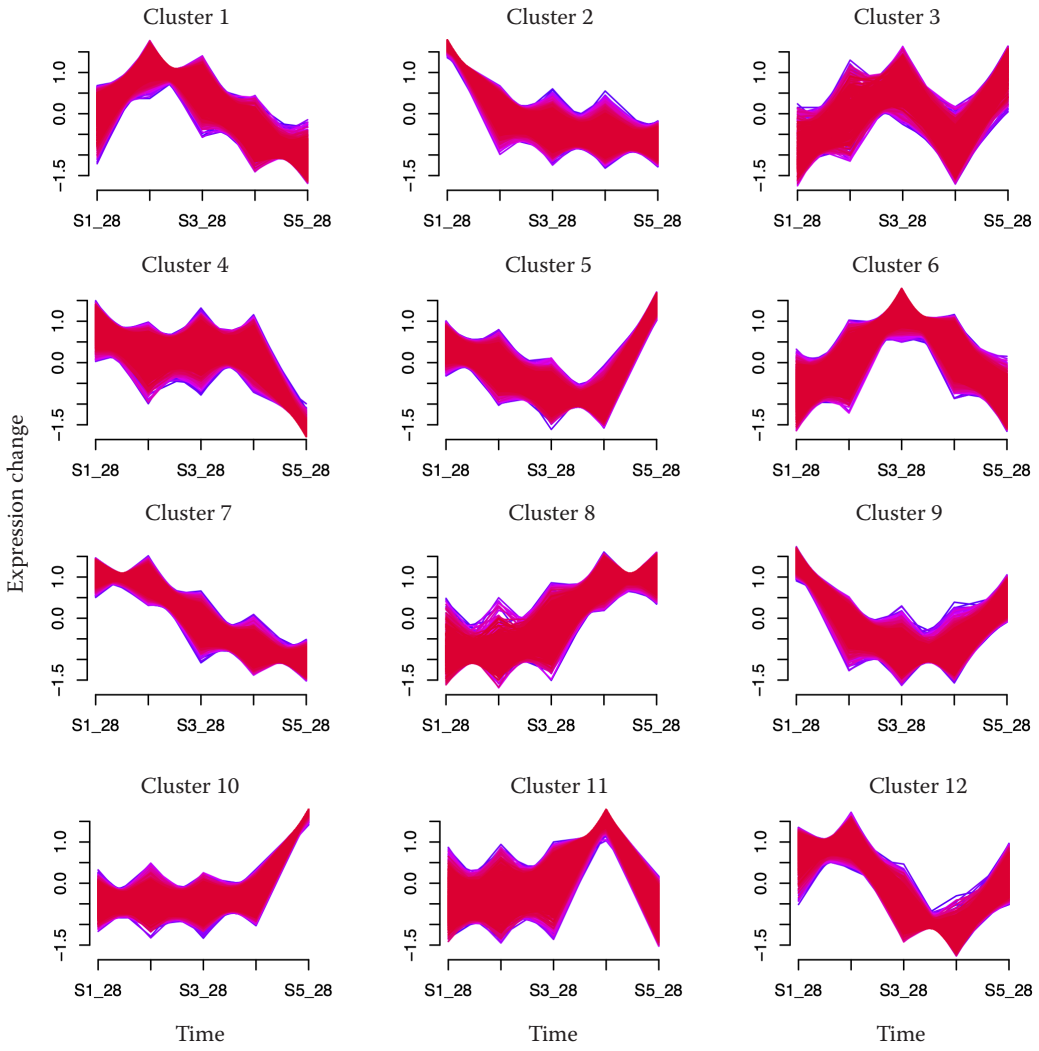

Figure 8. The clusters of co-regulated genes in sorghum Hongyingzi (X28)

response genes were enriched in Cluster 2, e.g., *ubiquitin C* (LOC110430698). Some other genes related to the biosyntheses of secondary metabolites were also observed in Cluster 2, e.g., *tan1* (LOC8076027), which has been demonstrated to play a critical role in the biosyntheses of tannins. The gene expression levels of Cluster 8, which includes 1 130 genes, were observed to increase, especially at stages S4 and S5. Genes assigned to Cluster 10 were only observed with high expression levels at stage S5 (Figure 8). These genes in Clusters 8 and 10 might mainly contribute to maturity in Hongyingzi.

Referring to the WGCNA, DEGs from the five developmental stages for sorghum Hongyingzi were subjected to analyses for correlation with phenotypes (Figure 9). At stages S1 and S5, a few gene clusters were shared. Among these genes, the grey cluster (grey60) showed the highest correlation index and included 42 genes (Pearson correlation: 0.54~0.55), in which most were also affiliated with Cluster 10 from time series analyses (Figure 9). Other correlated gene clusters, such as the green–yellow cluster, included 52 genes (Pearson correlation: 0.48~0.49), the red gene cluster (Pearson correlation: 0.45~0.49) included 72 genes and the green gene cluster (Pearson correlation: 0.35~0.36) included 90 genes (Figure 9).

To further understand the underlying relationships between DEGs and metabolism, the DEGs were annotated with KEGG pathways and subjected to enrichment analyses. The key aspects of metabolism differentiated among the developmental stages were expected to be determined. For each comparison, the GO enrichment with the lowest 20 Q-values are shown in each graph (Figure 10). Compared with the stage S1, the distinguishing metabolism was related to a photosynthesis–antenna protein in the other

https://doi.org/10.17221/16/2023-CJGPB

four stages. Other than photosynthesis-related genes, the starch and sucrose metabolism, and flavonoid biosynthesis-related genes were observed to show the most prominent metabolic differences when stages S1 and S4 were compared.

Expression comparison among three sorghum cultivars at the stage of maturation. The major stage of starch accumulation was observed to be S5, according to the measurements of starch at the different stages for the three sorghum cultivars: Hongyingzi, Jinuoliang 1 and Hongliangfeng 1. To understand the mechanism of starch biosynthesis, a comparison of the gene expression at stage S5 for these three sorghum cultivars was performed. At S5 stage, 25 813 expressed genes were observed in the three sorghum cultivars, but fewer uniquely expressed genes and a lower abundance of expression in Hongyingzi were detected (Figure 11A). The shared genes between Hongyingzi and the other two cultivars were fewer than those genes shared by Jinuoliang 1 and Hongliangfeng 1, indicating that a larger discrepancy between Hongyingzi and the other two sorghum grains than the difference between Jinuoliang 1 and Hongliangfeng 1.

According to the comparisons between sorghum Hongyingzi and the other two cultivars (Jinuoliang 1 and Hongliangfeng 1), 1 824 overlapped genes mainly contributed to the difference between them (Figure 11B). Among these 1 824 genes, although 29 genes are annotated with starch and sucrose metabolism, lower expression levels were observed in this study, indicating that they play minor roles in the accumulation of starch in the three sorghum cultivars. Interestingly, 213 genes were annotated with the secondary metabolic pathways, particularly biosynthesis pathways for phenolic compounds, suggesting that these genes are responsible for the large differences in these metabolic pathways between Hongyingzi and the other two sorghum cultivars (Figure 11C, D).

Biosynthesis of starch. Although numerous studies employing genome-wide association analysis (GWAS) were conducted on sorghum, and a batch of potential genes and SNPs that play important roles in the biosynthesis of starch has been predicted (Chen et al. 2019; Kimani et al. 2020), the transcriptional levels of these genes and related SNPs and their contributions to phenotype remain to be

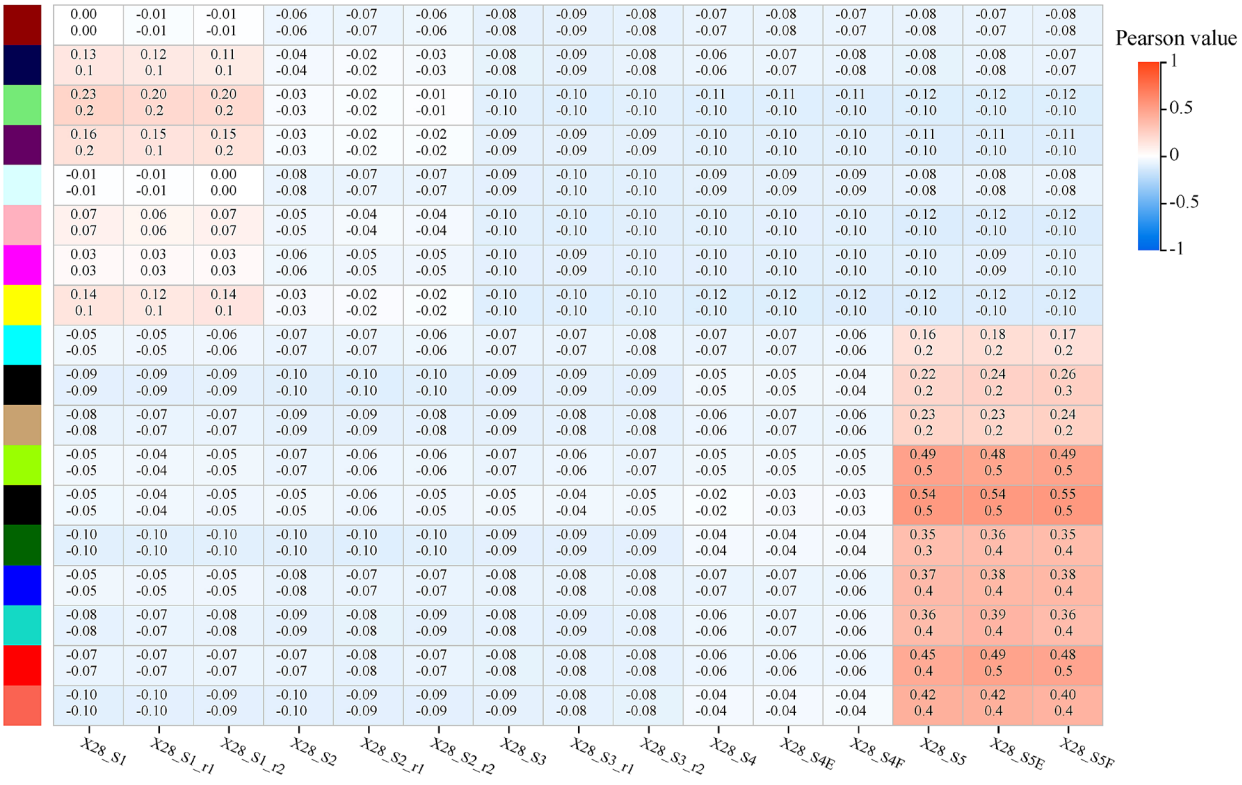

Figure 9. The correlations between co-regulated genes and phenotypes for sorghum Hongyingzi (X28)

https://doi.org/10.17221/16/2023-CJGPB

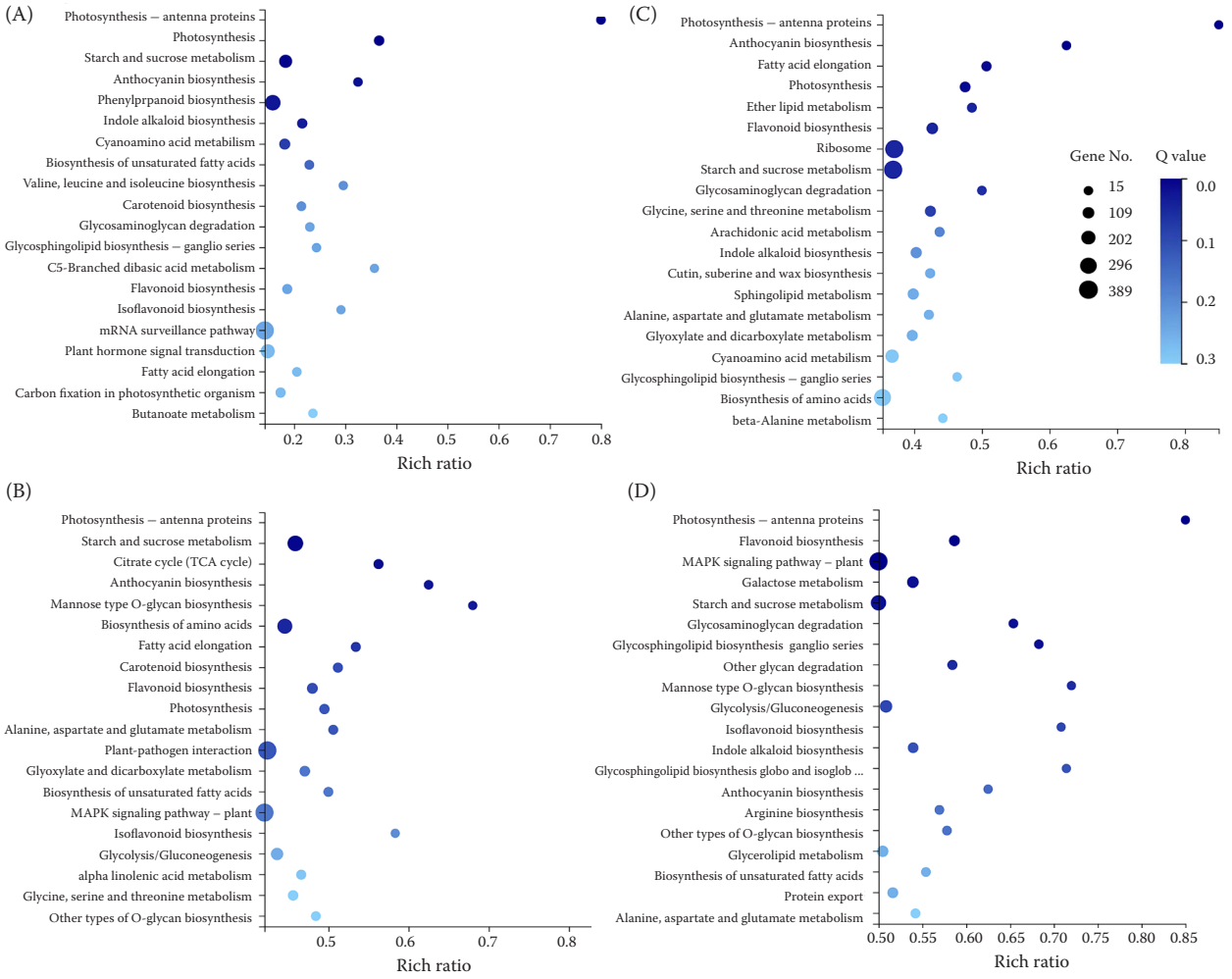

Figure 10. The Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathways of enriched differentially expressed genes (DEGs) for each pair of comparisons between different developmental stages: comparison of S1 and S2 (A), comparison of S1 and S3 (B), comparison of S1 and S4 (C), comparison of S1 and S5 (D)

S1 – early milky stage; S2 – middle to late milky stage; S3 – early to middle dough stage; S4 – late dough stage; S5 –maturation stage

validated. For instance, a strong association between sucrose phosphate synthase (SPS) (LOC8071544, Sobic.005G089600) and phenotype was proposed in a previous study (Kimani et al. 2020). However, low expression of SPS was detected in three sorghum cultivars in the present study, possibly suggesting a minor role for these genetic elements in starch biosynthesis. 6-Phosphofructokinase (PFK) (LOC8065903, Sobic.009G071800) has been demonstrated to be involved in multiple metabolic pathways including the starch biosynthesis process (Chen et al. 2019), and it was also proposed to be correlated with the biosyntheses of tannins in another GWAS study (Kimani et al. 2020). However, in the present study, PFK was detected with a decreasing trend of expression during the five developmental stages of sorghum Hongyingzi and with low levels at all the developmental stages. All of these phenomena emphasize the importance of transcriptional studies on genetic determinants underlying starch biosynthesis in Hongyingzi.

Our chemical analyses indicate that the greatest accumulation of starch occurred at stages S4 and S5 for the three sorghum cultivars, leading us to focus on the study of starch biosynthesis at these two stages. In the present study, for the five developmental stages of sorghum Hongyingzi, 535 genes assigned to the starch and sucrose biosynthesis pathway were detected and differentiated in terms of the expression levels at the different stages. Among these 535 genes, LOC8062300 (glucan endo-1,3-

https://doi.org/10.17221/16/2023-CJGPB

beta-glucosidase GII) was observed to present the highest expression level at stage S5 for the three sorghum cultivars, especially for Hongyingzi. This suggests this gene plays an important role in the biosynthesis of starch. Further study did not support the existence of SNPs in this gene for sorghum Hongyingzi, indicating there was no amino acid change in the enzyme and its regulatory factors might play key roles in its expression level (Table 1). LOC8062300 was assigned to the green cluster in WCGNA and Cluster 10 in the time series analysis of the S1–S5 stages for sorghum Hongyingzi (Figure 8). These genes clustered with LOC8062300 provide insights into the regulatory factors that interact with it. One bHLH transcription factor, *myc2* (LOC8084341), was expressed at a high level only at stage S5 for these three sorghum cultivars, indicating its potential role in regulating genes within the same cluster.

In addition, LOC8068390 (*waxy*, Sobic.010G022600, granule-bound starch synthase) showed the highest expression level at stage S4 for sorghum Hongyingzi and was also highly expressed at S3 and S4 stages for the other two sorghum cultivars. This waxy enzyme catalyzes one of the enzymatic steps of starch synthesis, and is responsible for the synthesis of amylose and building the final structure of amylopectin. This suggests that *waxy* might be a key gene that discriminates the compositions of starch for these three sorghum cultivars.

To completely visualize the metabolic pathways and related key DEGs of starch and sucrose, a KEGG pathway map was generated with DEGs between the S1 and S5 stages in Hongyingzi (Figure 12).

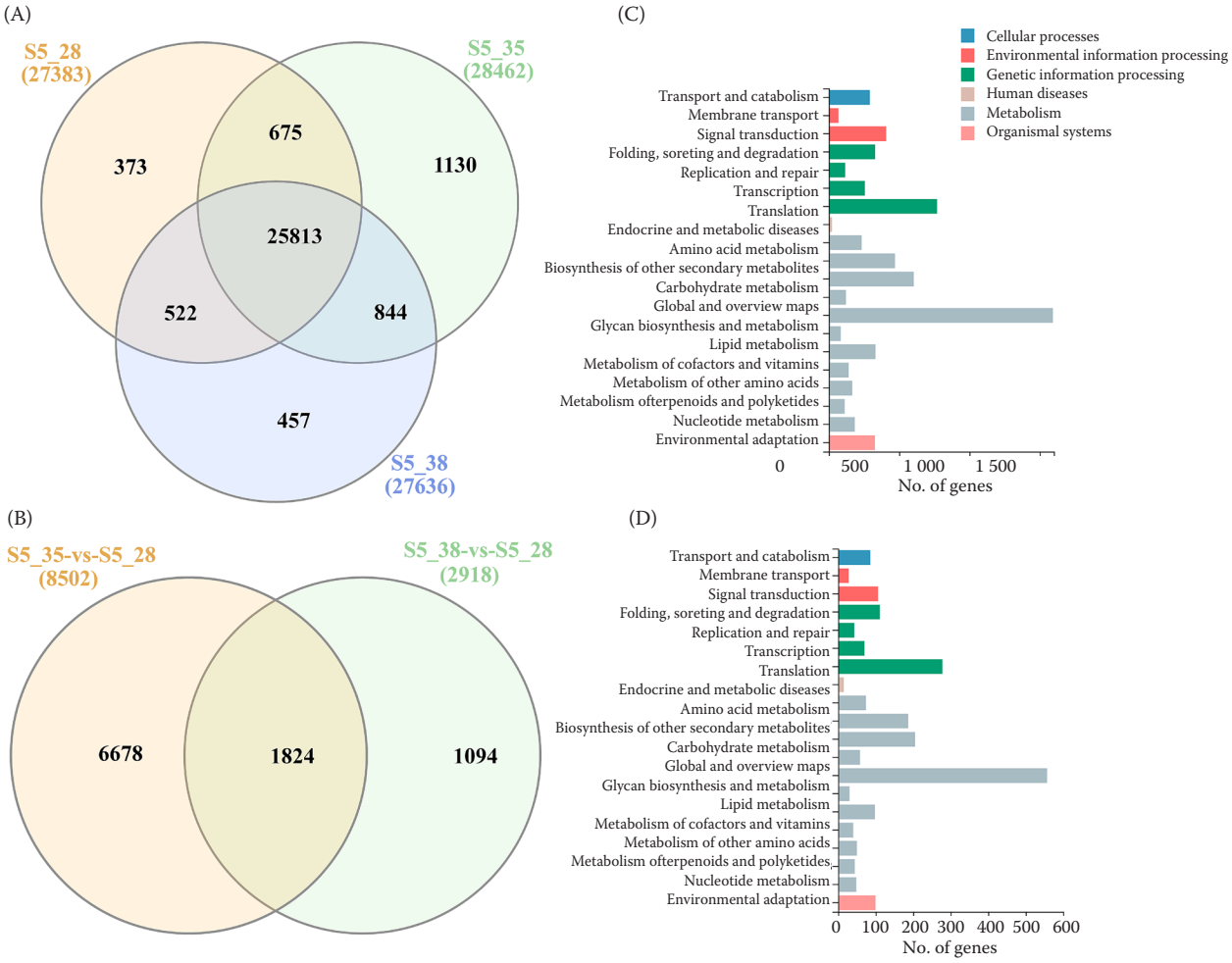

Figure 11. The comparison of metabolic pathways assigned for differentially expressed genes (DEGs) between Hongyingzi (X28) and Jinuoliang 1 (X35) and Hongliangfeng 1 (X38) at S5 stage (maturation stage): Venn diagram of transcripts shared by different grains at S5 stage (A), Venn diagram of DEGs shared by two comparisons between different grains at S5 stage (B), assigned metabolic pathways of DEGs within two comparisons, X28 vs X35 (C) and X28 vs X38 (D) at S5 stage

https://doi.org/10.17221/16/2023-CJGPB

Table 1. The representative genes associated with the starch biosynthesis pathway in sorghum grains

Gene	Chr	Cluster	Gene locus	Unique SNPs of Hongyingzi (X28)	Coordinates on reference genome	Annotations	Metabolic pathway
speD, AMD1	4	10	LOC8059669, Sobic.004G211866	4: 56 151 023 (A- > G)	56150894-56153875	S-adenosylmethionine decarboxylase	arginine, proline, cysteine and methionine metabolism
COX6B	4	7	LOC8072449, Sobic.004G211833	N/A	56145061-56149972	cytochrome c oxidase subunit 6b	oxidative phosphorylation
SPS	5	2	LOC8071544, Sobic.005G089600	5: 12 855 792(C- > A) 5: 12 855 900(G- > A) 12 856 057(A- > C) 12 856 533(C- > T) 12 856 741(C- > G) 12 856 818(G- > T) 12 857 271(C- > T)	12853735-12859673	sucrose–phosphate synthase	starch and sucrose metabolism
menA	1	9	LOC8081915, Sobic.001G475600	N/A	74788395-74791633	1,4-dihydroxy-2-naphthoate octaprenyltransferase	ubiquinone and terpene biosynthesis pathway
IRAK4	5	4	LOC8076902, Sobic.005G084100	N/A	11311899-11370725	interleukin-1 receptor-associated kinase 4	multiple metabolic pathways
	6	3	LOC110436578, Sobic.006G036500.1	N/A	12447139-12448051	peptidyl-prolyl cis-trans-isomerase CYP38	
pfkA	2	2	LOC8065903, Sobic.009G071800	8 622 289(G- > C) 8 622 301(G- > A) 8 622 862(C- > G) 8 623 482(A- > C)	8622004-8624299	ATP-dependent 6-phosphofructokinase 6	multiple metabolic pathways including the pentose phosphate and glycolytic pathways
waxy	10	11	LOC8068390, Sobic.010G022600	N/A	1860943-1865281	granule-bound starch synthase	starch and sucrose metabolism
GN4	2	10	LOC8062300	N/A	72684105-72685501	glucan endo-1,3-beta-glucosidase 4	starch and sucrose metabolism

N/A – absent or not detected; Chr – chromosome; SNPs – single-nucleotide polymorphisms

Most of the key enzymes in the metabolic pathway for starch and sucrose were expressed more highly at stage S5 than S1.

Determinants of the biosyntheses of phenolic compounds. Phenolic acids mainly include 3-deoxyanthocyanidins and condensed tannins. The flavonoids are one class of polyphenols, and their biosyntheses are widely regulated by transcription factors such as a tripartite complex composed of MYB, bHLH and WD40. Condensed tannins are the major phenolic acids. Classical genetic analyses revealed that the biosyntheses of tannins are regulated by B1 and B2. The most recent studies suggest that *tan1* (Sobic.004G280800, LOC8076027) and *tan2* (Sobic.002G076600, LOC8069098, bHLH-containing regulatory protein) are B1 and B2 representative genes, respectively (Wu et al. 2012; 2019). During the developmental stages from S1 to S5 for sorghum Hongyingzi, the expression levels of *tan1* and *tan2* were observed to decrease, while the contents of tannins increased, with a peak at stage S4, and decreased at the maturation stage of S5. This indicates that the biosyntheses of tannins occurred and they accumulated during the stages S1 to S4,

https://doi.org/10.17221/16/2023-CJGPB

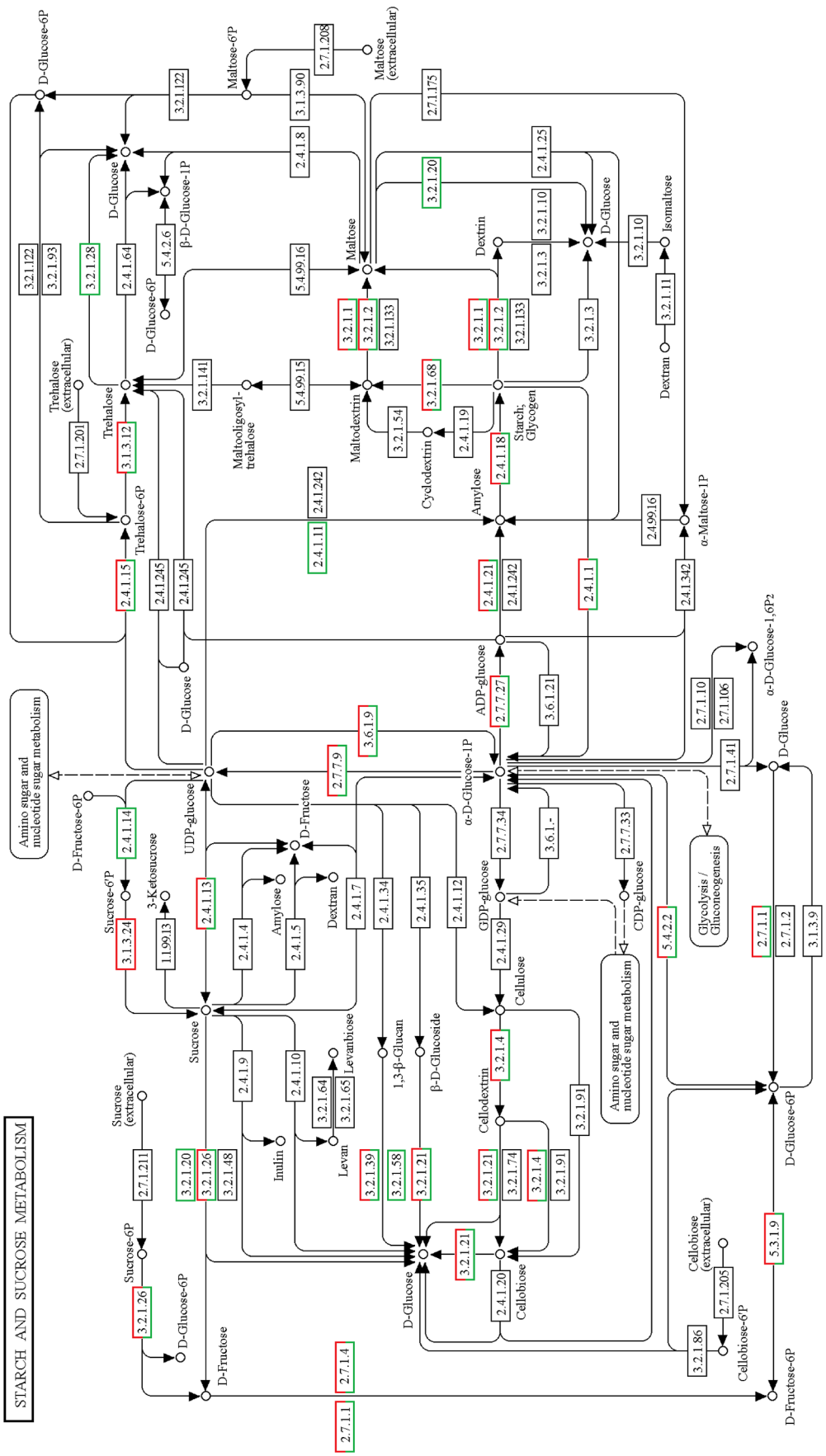

Figure 12. The differentially expressed genes (DEGs) assigned to the starch and sucrose metabolic pathway between stages of S1 (early milky stage) and S5 (maturation) for sorghum Hongyingzi (X28)Green box represents down-regulation and red box represents up-regulation

https://doi.org/10.17221/16/2023-CJGPB

in contrast with the biosynthesis of starch at S5. Some unique SNPs in *tan1* and *tan2* were identified in Hongyingzi compared with the two sorghum cultivars (Jinuoliang 1 and Hongliangfeng 1) (Table 2). The contributions of these SNPs to phenotypes need to be validated in future work.

Additionally, recent studies have revealed two SPNs on chromosome 4, which are located at the genes of *zm1* (LOC8073887, Sobic.004G273800) and *AGL6* (LOC8076029, Sobic.004G281000), respectively (Kimani et al. 2020) (Table 2). Both SNPs were correlated with the contents of tannins with high confidence. Four SNPs were identified in *AGL6*, while no mutation was detected in *zm1*. *AGL6* is a homolog of *TT16* (TRANSPARENT TESTA 16), which is a MADS-box transcription factor (Nesi et al. 2002), and is required for the accumulation of flavonoids such as proanthocyanidins (PAs) in *Arabidopsis* (Xu et al. 2017). Other genes, such as *6-phosphofructokinase I* (*pfkA*) (LOC8065903, Sobic.009G071800), are involved in multiple metabolic pathways including the pathway responsible for the biosynthesis of flavonoids. Some of them show unique SNPs in sorghum Hongyingzi, compared

Table 2. Synthetic genes identified in the present study for tannins in sorghum grains

Gene	Chr	Cluster	Gene locus tag	Unique SNPs in Hongyingzi (X28)	Coordinates on reference genome	Annotation	Metabolic pathway
tan1	4	2	LOC8076027, Sobic.004G280800	4: 62 316 361(C- > T)	62315375-62318806	protein transparent testa glabra 1	tannin biosynthesis
tan2	2	2	LOC8069098, Sobic.002G076600	2: 7 979 912(T- > C)	7975774-7985168	transcription factor MYC2	
AGL6 /TT16	4	2	LOC8076029, Sobic.004G281000	4: 62 326 334(A- > G) 4: 62 329 389(A- > G) 4: 62 329 540(C- > G) 4: 62 330 752(G- > A)	62324420-62331117	MADS-box transcription factor	
menB	4	2	LOC8074648, Sobic.004G044200	4: 3 633 686(T- > C)	7975774-7985168	1,4-dihydroxy-2-naphthoyl-CoA synthase, peroxisomal/naphthoate synthase	ubiquinone and other terpenoid-quinone biosynthesis
zm1	4	2	LOC8073887, Sobic.004G273800	N/A	61765945-61767253	Myb transcription factor	
	4	2	LOC8073888, Sobic.004G273900	N/A	61770743-61771753	superoxide	phenylpropanoid biosynthesis
	5	11	LOC8064408, Sobic.005G110600	N/A	35003426-35005235	chitinase-3-like protein 1	amino sugar and nucleotide sugar metabolism
HPSE	8	4	LOC8068132, Sobic.008G141700	N/A	57279402-57283616	heparanase	glycosaminoglycan degradation
	4	4	LOC8073181, Sobic.004G273600	N/A	61735730-61738644	RNA recognition motif	
purM	9	9	LOC8055712, Sobic.009G072000	8 632 510(G- > A) 8 632 751(C- > T) 8 634 594(G- > C)	8631815-8634660	phosphoribosylformylglycinamidine cyclo-ligase	purine metabolism
pfkA	2	2	LOC8065903, Sobic.009G071800	8 622 289(G- > C) 8 622 301(G- > A) 8 622 862(C- > G) 8 623 482(A- > C)	8622004-8624299	ATP-dependent 6-phosphofructokinase 6	multiple metabolic pathways including the pentose phosphate and glycolytic pathways

N/A – absent or not detected; Chr – chromosome; SNPs – single-nucleotide polymorphisms

https://doi.org/10.17221/16/2023-CJGPB

with the other two sorghum cultivars (Jinuoliang 1 and Hongliangfeng 1) (Table 2).

The genes involved in the biosyntheses of tannins were mainly grouped into Cluster 2, and this provides insights into the novel and key genes responsible for the biosyntheses of tannins. Within the genes of this cluster, 73 myb- or bHLH-class transcription-factor-encoding genes were identified, and these genes warrant future studies to discover novel transcription factors and to clarify aspects of metabolism underlying the biosyntheses of tannins.

CONCLUSION

New transcripts and genes were identified at five different developmental stages of grains for three sorghum cultivars using transcriptomics methods in the present study. Our work may provide impetus for further study of the genes responsible for the biosyntheses of starch and phenolic compounds in sorghum. The genes were characterized at each developmental stage of grains, and some of them were unique in Hongyingzi, compared with the other two sorghum cultivars. This revealed the differences among the sorghum grains in the co-regulated gene patterns at different developmental stages. The critical genes and SNPs related to the biosynthetic pathways for starch and phenolic compounds were further discussed. A few co-regulated gene clusters in multiple developmental stages identified from the present study provide insights into the potential regulatory network and further contribute to the clarification of the key determinant genes involved in the biosyntheses of starch and phenolic compounds. The effects of SNPs located at genes that are related to the biosyntheses of phenolic compounds need to be validated through molecular genetic bioassays.

Data availability statement. The data supporting our findings are presented in additional files, and all transcriptomics data has been deposited in GenBank under PRJNA713525. https://dataview.ncbi.nlm.nih.gov/object/PRJNA713525?reviewer=er2u4sa3or6rii86i4ahiihv2 t

REFERENCES

Adom K.K., Liu R.H. (2002): Antioxidant activity of grains. Journal of Agricultural and Food Chemistry, 50: 6182–6187.

Bloom J.S., Khan Z., Kruglyak L., Singh M., Caudy A.A. (2009): Measuring differential gene expression by short read sequencing: Quantitative comparison to 2-channel gene expression microarrays. BMC Genomics, 10: 221.

Bolger A.M., Lohse M., Usadel B. (2014): Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics, 30: 2114–2120.

Boyles R.E., Pfeiffer B.K., Cooper E.A., Rauh B.L., Zielinski K.J., Myers M.T., Brenton Z., Rooney W.L., Kresovich S. (2017): Genetic dissection of sorghum grain quality traits using diverse and segregating populations. Theoretical and Applied Genetics, 130: 697–716.

Buchfink B., Xie C., Huson D.H. (2015): Fast and sensitive protein alignment using DIAMOND. Nature Methods, 12: 59–60.

Campbell B.C., Gilding E.K., Mace E.S., Tai S., Tao Y., Prentis P.J., Thomelin P., Jordan D.R., Godwin I.D. (2016): Domestication and the storage starch biosynthesis pathway: Signatures of selection from a whole sorghum genome sequencing strategy. Plant Biotechnology Journal, 14: 2240–2253.

Chen B.R., Wang C.Y., Wang P., Zhu Z.X., Xu N., Shi G.S., Yu M., Wang N., Li J.H., Hou J.M., Li S.J., Zhou Y.F., Gao S.J., Lu X.C., Huang R.D. (2019): Genome-wide association study for starch content and constitution in sorghum (*Sorghum bicolor* (L.) Moench). Journal of Integrative Agriculture, 18: 2446–2456.

Chen Y., Chen Y., Shi C., Huang Z., Zhang Y., Li S., Li Y., Ye J., Yu C., Li Z., Zhang X., Wang J., Yang H., Fang L., Chen Q. (2018): SOAPnuke: A MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience, 7: 1–6.

de Morais Cardoso L., Pinheiro S.S., Martino H.S., Pinheiro-Sant'Ana H.M. (2017): Sorghum (*Sorghum bicolor* L.): Nutrients, bioactive compounds, and potential impact on human health. Critical Reviews in Food Science and Nutrition, 57: 372–390.

Dong W., He F., Jiang H., Liu L., Qiu Z. (2021): Comparative transcriptome sequencing of taro corm development with a focus on the starch and sucrose metabolism pathway. Frontiers in Genetics, 12: 771081.

Eddy S.R. (1998): Profile hidden Markov models. Bioinformatics, 14: 755–763.

Habyarimana E., Dall'Agata M., De Franceschi P., Baloch F.S. (2019): Genome-wide association mapping of total antioxidant capacity, phenols, tannins, and flavonoids in a panel of *Sorghum bicolor* and *S. bicolor* × *S. halepense* populations using multi-locus models. PLoS ONE, 14: e0225979.

Hu Z., Olatoye M.O., Marla S., Morris G.P. (2019): An integrated genotyping-by-sequencing polymorphism map for over 10,000 sorghum genotypes. Plant Genome, 12: 180044.

James M.G., Denyer K., Myers A.M. (2003): Starch synthesis in the cereal endosperm. Current Opinion in Plant Biology, 6: 215–222.

https://doi.org/10.17221/16/2023-CJGPB

Ke F., Zhang K., Li Z., Wang J., Zhang F., Wu H., Zhang Z., Lu F., Wang Y., Duan Y., Liu Z., Zou J., Zhu K. (2022): Transcriptomic analysis of starch accumulation patterns in different glutinous sorghum seeds. Scientific Reports, 12: 11133.

Kim D., Langmead B., Salzberg S.L. (2015): HISAT: A fast spliced aligner with low memory requirements. Nature Methods, 12: 357–360.

Kimani W., Zhang L.M., Wu X.Y., Hao H.Q., Jing H.C. (2020): Genome-wide association study reveals that different pathways contribute to grain quality variation in sorghum (*Sorghum bicolor*). BMC Genomics, 21: 112.

Kong L., Zhang Y., Ye Z.Q., Liu X.Q., Zhao S.Q., Wei L., Gao G. (2007): CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Research, 35: W345–349.

Kumar L., Futschik E.M. (2007): Mfuzz: A software package for soft clustering of microarray data. Bioinformation, 2: 5–7.

Langfelder P., Horvath S. (2008): WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics, 9: 559.

Langmead B., Salzberg S.L. (2012): Fast gapped-read alignment with Bowtie 2. Nature Methods, 9: 357–359.

Li B., Dewey C.N. (2011): RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics, 12: 323.

Margraf T., Karnopp A.R., Rosso N.D., Granato D. (2015): Comparison between Folin-Ciocalteu and Prussian blue assays to estimate the total phenolic content of juices and teas using 96-well microplates. Journal of Food Science, 80: C2397–2403.

Marioni J.C., Mason C.E., Mane S.M., Stephens M., Gilad Y. (2008): RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. Genome Research, 18: 1509–1517.

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., DePristo M.A. (2010): The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20: 1297–1303.

Mitchell G.A. (1990): Methods of starch analysis. Starch-Stärke, 42: 131–134.

Nesi N., Debeaujon I., Jond C., Stewart A.J., Jenkins G.I., Caboche M., Lepiniec L. (2002): The TRANSPARENT TESTA16 locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. Plant Cell, 14: 2463–2479.

Pertea M., Pertea G.M., Antonescu C.M., Chang T.C., Mendell J.T., Salzberg S.L. (2015): StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature Biotechnology, 33: 290–295.

Rice P., Longden I., Bleasby A. (2000): EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics, 16: 276–277.

Sanseverino W., Roma G., De Simone M., Faino L., Melito S., Stupka E., Frusciante L., Ercolano M.R. (2010): PRGdb: A bioinformatics platform for plant resistance gene analysis. Nucleic Acids Research, 38: D814–821.

Scully E.D., Gries T., Sarath G., Palmer N.A., Baird L., Serapiglia M.J., Dien B.S., Boateng A.A., Ge Z., Funnell-Harris D.L., Twigg P., Clemente T.E., Sattler S.E. (2016): Overexpression of SbMyb60 impacts phenylpropanoid biosynthesis and alters secondary cell wall composition in *Sorghum bicolor*. Plant Journal, 85: 378–395.

Shen S., Park J.W., Lu Z.X., Lin L., Henry M.D., Wu Y.N., Zhou Q., Xing Y. (2014): rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proceedings of the National Academy of Sciences of the USA, 111: E5593–5601.

Shen S., Huang R., Li C., Wu W., Chen H., Shi J., Chen S., Ye X. (2018): Phenolic compositions and antioxidant activities differ significantly among sorghum grains with different applications. Molecules, 23: 1203.

Shi L., Wang J., Liu Y., Ma C., Guo S., Lin S., Wang J. (2021): Transcriptome analysis of genes involved in starch biosynthesis in developing Chinese chestnut (*Castanea mollissima* Blume) seed kernels. Scientific Reports, 11: 3570.

Singleton V.L., Orthofer R., Lamuela-Raventós R.M. (1999): Analysis of total phenols and other oxidation substrates and antioxidants by means of Folin-Ciocalteu reagent. Methods in Enzymology, 229: 152–178.

Tetlow I.J., Morell M.K., Emes M.J. (2004): Recent developments in understanding the regulation of starch metabolism in higher plants. Journal of Experimental Botany, 55: 2131–2145.

Tetlow I.J., Beisel K.G., Cameron S., Makhmoudova A., Liu F., Bresolin N.S., Wait R., Morell M.K., Emes M.J. (2008): Analysis of protein complexes in wheat amyloplasts reveals functional interactions among starch biosynthetic enzymes. Plant Physiology, 146: 1878–1891.

Trapnell C., Roberts A., Goff L., Pertea G., Kim D., Kelley D.R., Pimentel H., Salzberg S.L., Rinn J.L., Pachter L. (2012): Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols, 7: 562–578.

Wang L., Feng Z., Wang X., Wang X., Zhang X. (2010): DEGseq: An R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics, 26: 136–138.

https://doi.org/10.17221/16/2023-CJGPB

Wu Y., Li X., Xiang W., Zhu C., Lin Z., Wu Y., Li J., Pandravada S., Ridder D.D., Bai G., Wang M.L., Trick H.N., Beane S.R., Tuinstra M.R., Tesso T.T., Yu J. (2012): Presence of tannins in sorghum grains is conditioned by different natural alleles of Tannin1. Proceedings of the National Academy of Sciences of the USA, 109: 10281–10286.

Wu Y., Guo T., Mu Q., Wang J., Li X., Wu Y., Tian B., Wang M.L., Bai G., Perumal R., Trick H.N., Bean S.R., Dweikat I.M., Tuinstra M.R., Morris G., Tesso T.T., Yu J., Li X. (2019): Allelochemicals targeted to balance competing selections in African agroecosystems. Nature Plants, 5: 1229–1236.

Xiong Y., Zhang P., Warner R.D., Fang Z. (2019): Sorghum grain: From genotype, nutrition, and phenolic profile to its health benefits and food applications. Comprehensive Reviews in Food Science and Food Safety, 18: 2025–2046.

Xu W., Bobet S., Le Gourrierec J., Grain D., De Vos D., Berger A., Salsac F., Kelemen Z., Boucherez J., Rolland A., Mouille G., Routaboul J.M., Lepiniec L., Dubos C. (2017): TRANSPARENT TESTA 16 and 15 act through different mechanisms to control proanthocyanidin accumulation in *Arabidopsis* testa. Journal of Experimental Botany, 68: 2859–2870.

Yang Q., Yuan Y., Liu J., Han M., Li J., Jin F., Feng B. (2023): Transcriptome analysis reveals new insights in the starch biosynthesis of non-waxy and waxy broomcorn millet (*Panicum miliaceum* L.). International Journal of Biological Macromolecules, 230: 123155.

Yin H., Chen L., Li X.Y., Chen Q.M., Yi M.F. (2008): Analysis and improvement of high-quality RNA extraction in leaves of lily. Journal of China Agricultural University, 13: 41–45. (in Chinese)

Zhou Y., Lv J., Yu Z., Wang Z., Li Y., Li M., Deng Z., Xu Q., Cui F., Zhou W. (2022): Integrated metabolomics and transcriptomic analysis of the flavonoid regulatory networks in *Sorghum bicolor* seeds. BMC Genomics, 23: 619.

Received: February 14, 2023
Accepted: April 11, 2023
Published online: June 1, 2023